# HARMONIC ANALYSIS, THE TRACE FORMULA, AND SHIMURA VARIETIES

**Clay Mathematics Proceedings**
Volume 4

# HARMONIC ANALYSIS, THE TRACE FORMULA, AND SHIMURA VARIETIES

Proceedings of the Clay Mathematics Institute
2003 Summer School, The Fields Institute
Toronto, Canada, June 2–27, 2003

**James Arthur**
**David Ellwood**
**Robert Kottwitz**
Editors

**American Mathematical Society**
**Clay Mathematics Institute**

2000 *Mathematics Subject Classification*. Primary 11–02; Secondary 11F70, 11F72, 11F85, 11G18, 14G35, 22E35, 22E50, 22E55.

Cover art developed with the assistance of Bill Casselman.

ISBN 0-8218-3844-X (alk. paper)


Published by the American Mathematical Society, Providence, RI,
for the Clay Mathematics Institute, Cambridge, MA.
Printed in the United States of America.


∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at http://www.ams.org/
Visit the Clay Mathematics Institute home page at http://www.claymath.org/

10 9 8 7 6 5 4 3 2 1      10 09 08 07 06 05

# Contents

# Preface

This volume is based on lectures given at the fourth Clay Mathematics Institute Summer School entitled "Harmonic Analysis, the Trace Formula, and Shimura Varieties." It was held at the Fields Institute in Toronto, Canada, from June 2 to June 27, 2003.

The main goal of the School was to introduce graduate students and young mathematicians to three broad and interrelated areas in the theory of automorphic forms. Much of the volume is comprised of the articles of Arthur, Kottwitz, and Milne. Although these articles are based on lectures given at the school, the authors have chosen to go well beyond what was discussed there, in order to provide both a sense of the underlying structure of the subject and a working knowledge of some of its techniques. They were written to be self-contained in some places, and to be used in conjunction with given references in others. We hope the volume will convey the depth and beauty of this challenging field, in which there yet remains so much to be discovered—perhaps some of it by you, the reader!

The theory of automorphic forms is formulated in terms of reductive algebraic groups. This is sometimes a serious obstacle for mathematicians whose background does not include Lie groups and Lie algebras. The monograph is by no means intended to exclude such mathematicians, even though the theory of reductive groups was an informal prerequisite for the Summer School. Some modest familiarity with the language of algebraic groups is often sufficient, at least to get started. For this reason, we have generally resisted the temptation to work with specific matrix groups. The short article of Murnaghan contains a summary of some of the basic properties of reductive algebraic groups that are used elsewhere in the monograph.

Much of the modern theory of automorphic forms is governed by two fundamental problems that are at the heart of the Langlands program. One is Langlands' principle of functoriality. The other is the general analogue of the Shimura-Taniyama-Weil conjecture on modular elliptic curves. (See [**A**] and [**L**, §2].) These problems are among the deepest questions in mathematics. It is premature to try to guess what various techniques will play a role in their ultimate resolution. However, the trace formula and the theory of Shimura varieties are both likely to be an essential part of the story. They have already been used to establish significant special cases.

The trace formula has perhaps been more closely identified with the first problem. Special cases of functoriality arise naturally from the conjectural theory of endoscopy, in which a comparison of trace formulas would be used to characterize the internal structure of the automorphic representations of a given group. (See [**Sh**] for a discussion of the first case to be investigated.) Likewise, Shimura varieties are usually associated with the second problem. As higher dimensional analogues of modular curves, they are attached by definition to certain reductive groups. In many cases, it has been possible to establish reciprocity laws between $\ell$-adic Galois representations on their cohomology groups and automorphic representations of the corresponding reductive groups. These laws can be formulated as an explicit formula for the zeta function of a Shimura variety in terms of automorphic $L$-functions. (See [**K**] for a discussion of the rough form such a formula is expected to take. The word "rough" should be taken seriously, given the current limitations of our understanding.)

The work of Wiles that led to a proof of Fermat's Last Theorem suggests that the two problems are inextricably linked. This is already apparent in the reciprocity laws that have been established for Shimura varieties. Indeed, the conjectural formula for the zeta function of a general Shimura variety requires the theory of endoscopy even to state. Moreover, the proof of these reciprocity laws requires a comparison of the (automorphic) trace formula with an ($\ell$-adic) Lefschetz trace formula. Some of the most striking parts of the argument are in the comparison of the various terms in the two formulas. The most sophisticated Shimura varieties for which there are complete results are the so-called Picard modular surfaces. (See [**LR**], especially the summary on pp. 255–302.) Picard modular surfaces are attached to unitary groups in three variables. It is no coincidence that the theory of endoscopy has also been established for these groups, thereby yielding a classification of their automorphic representations [**R**].

There is some discussion of these problems in the articles of Arthur and Milne. However, the articles of both Arthur and Milne really are intended as introductions, despite their length. The theory of endoscopy, and the automorphic description of zeta functions of Shimura varieties, are at the forefront of present day research. They are for the most part beyond the scope of this monograph.

The local terms in the trace formula are essentially analytic objects. They include the invariant orbital integrals and irreducible characters that are the basis for Harish-Chandra's theory of local harmonic analysis. They also include weighted orbital integrals and weighted characters, objects that arose for the first time with the trace formula. The article of Kottwitz is devoted to the general study of these terms at $p$-adic places. It is a largely self-contained course, which covers many of Harish-Chandra's basic results in invariant harmonic analysis, as well as their weighted, noninvariant analogues.

The article of DeBacker focuses on the phenomenon of homogeneity in invariant harmonic analysis at $p$-adic places. It concerns quantitative forms of some of the basic theorems of $p$-adic harmonic analysis, such as Howe's finiteness theorem and Harish-Chandra's local character expansion. The article also explains how homogeneity enters into Waldspurger's analysis of stability for linear combinations of nilpotent orbital integrals.

There are subtle questions concerning the terms in the trace formula that go beyond those treated by Kottwitz and DeBacker. The most basic of these is known as the fundamental lemma, even though it is still largely conjectural.[1] The article by Hales contains a precise statement of the conjecture and some remarks on progress toward a general proof. The fundamental lemma occupies a unique place in the theory. It is a critical ingredient in the comparison of trace formulas that is part of the theory of endoscopy. It has an equally indispensable role in the comparison of (automorphic and $\ell$-adic) trace formulas needed to establish reciprocity laws for Shimura varieties.

Some Shimura varieties are projective, which is to say that they are compact as complex varieties. They correspond to reductive groups over $\mathbb{Q}$ that are anisotropic. The trace formula in this case simplifies considerably. It reduces to the Selberg trace formula for compact quotient. On the other hand, the arithmetic geometry of such varieties is still very rich. In particular, the comparison of individual terms in the

[1] Moreover, the term *lemma* is ultimately a gross understatement.

two kinds of trace formulas is of major interest. There is a great deal left to be done, but it is in this case that there has been the most progress.

If the Shimura variety is not projective, the comparison is more sophisticated. It has to be based on the relationship between $L^2$-cohomology and intersection cohomology, conjectured by Zucker, and established by Saper and Stern, and Looijenga. The article of Goresky describes several compactifications of open Shimura varieties and their relations with associated cohomology groups. Goresky's article also serves as an introduction to work of Goresky and MacPherson, in which weighted cohomology complexes on the reductive Borel-Serre compactification are used to obtain a Lefschetz formula for the intersection cohomology of the Baily-Borel compactification. According to Zucker's conjecture, this last formula is equivalent to the relevant form of the automorphic trace formula. There remains the important open problem of establishing a corresponding $\ell$-adic Lefschetz formula that can be compared with either one of these two formulas.

The reciprocity laws proved for Picard modular surfaces in [**LR**] apply to places of good reduction. The same restriction has been implicit in our discussion of other Shimura varieties. In the final analysis, one would like to establish reciprocity laws between $\ell$-adic Galois representations and automorphic representations that apply to all places. The theory of Shimura varieties at places of bad reduction is considerably less developed, although there has certainly been progress. The article of Haines is a survey of recent work in this direction, concentrating on the case of level structures of parahoric type. It also touches upon the problem of comparing the automorphic trace formula with the Lefschetz formula, now in the context of bad reduction.

The article of Sarnak concerns the classical Ramanujan conjecture for modular forms and its higher dimensional analogues. Langlands has shown that the generalized Ramanujan conjecture is a consequence of the principle of functoriality. Conversely, it is possible that the generalized Ramanujan conjecture could play a critical role in the study of those cases of functoriality that are not part of the theory of endoscopy. Sarnak describes the present state of the conjecture and discusses various techniques that have been successfully applied to special cases.

We have tried to present the contents of the monograph from a unified perspective. Our description has been centered around two fundamental problems that are the essential expression of the Langlands program. The two problems ought to be treated as signposts, which give direction to current work, but which point to destinations that will not be reached in the foreseeable future. The reader is free to draw whatever inspiration from them his or her temperament permits. In any case, many of the questions discussed in the various articles here are of great interest in their own right. In point of fact, there is probably too much in the monograph for anyone to learn in a limited period of time. Perhaps the best strategy for a beginner would be to start with one or two articles of special interest, and try to master them.

As we have mentioned, participants were encouraged to bring a prior understanding of the basic properties of algebraic groups. The theory of reductive groups is rooted in the structure of complex semisimple Lie algebras, for which [**Se**] and [**H**] are good references. As for algebraic groups themselves, a familiarity with many of the topics in [**B**] or [**Sp**] is certainly desirable, though perhaps not essential.

Participants were also assumed to have some knowledge of number theory. The main theorem of class field theory is reviewed without proof in the article of Milne. A complete treatment can be found in [**CF**]. Tate's article on global class field theory in this reference contains a particularly good introduction to the theory. The thesis of Tate, reprinted as a separate article in [**CF**], is also recommended for its introduction to adeles and its construction of the basic abelian automorphic $L$-functions.

A reader might also want to consult other general articles in automorphic forms. A good introductory reference to the general theory of automorphic forms is the proceedings of the Edinburgh instructional conference [**BK**].

This Clay Mathematics Institute Summer School could not have taken place without the efforts of many people. We deeply appreciate the role of the Clay Mathematics Institute in making this summer school possible, and thank Vida Salahi in particular for the care and attention she exercised in bringing the volume to its final form. We are most grateful to the staff of the Fields Institute, who did such a superb job of making the School run smoothly. We are equally indebted to all the lecturers, not only for agreeing to take part in the School, but also for providing the texts collected in this volume. Last, but surely not least, we would like to thank the participants, whose enthusiastic response made it all worthwhile.

James Arthur, David Ellwood, Robert Kottwitz.
August, 2005.

## References


[A] J. Arthur, *The principle of functoriality*, Bull. Amer. Math. Soc. **40** (2002), 39–53

[BK] T.N. Baily and A.W. Knapp, *Representation Theory and Automorphic Forms*, Proc. Sympos. Pure Math. **61**, Amer. Math. Soc., 1996.

[B] A. Borel, *Linear Algebraic Groups*, Benjamin, 1969.

[CF] J. Cassels and A. Fröhlich, *Algebraic Number Theory*, Thompson, 1967.

[H] J. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, 1972.

[K] R. Kottwitz, *Shimura varieties and $\lambda$-adic representations*, in *Automorphic Forms, Shimura Varieties, and L-functions*, vol. I, Academic Press, 1990, 161–209.

[L] R. Langlands, *Automorphic representations, Shimura varieties, and motives. Ein Märchen*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 205–246.

[LR] R. Langlands and D. Ramakrishnan, *The Zeta Functions of Picard Modular Surfaces*, Les Publications CRM, Montreal, 1992.

[R] J. Rogawski, *Automorphic Representations of Unitary Groups in Three Variables*, Ann. of Math. Studies **123**, Princeton Univ. Press, 1990.

[Se] J.-P. Serre, *Complex Semisimple Lie Algebras*, Springer-Verlag, 1987.

[Sh] D. Shelstad, *Notes on L-indistinguishability (based on a lecture by R.P. Langlands)*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 185–192.

[Sp] T. Springer, *Linear Algebraic Groups*, Birkhäuser, 1981.

## Summer School Lecture Courses: June 2–20, 2003

Introduction to the trace formula
J. Arthur, June 2–20

Introduction to Shimura varieties
J. Milne, June 2–20

Background from algebraic groups
F. Murnaghan, June 2–6

Harmonic analysis on reductive groups and Lie algebras
R. Kottwitz, June 9–20

## Advanced Short Courses: June 23–27, 2003

An introduction to homogeneity with applications
S. DeBacker

Geometry and topology of compactifications of modular varieties
M. Goresky

Bad reduction of Shimura varieties
T. Haines

An introduction to the fundamental lemma
T. Hales

Analytic aspects of automorphic forms
P. Sarnak

Clay Mathematics Proceedings
Volume 4, 2005

# An Introduction to the Trace Formula

James Arthur

Contents



Supported in part by NSERC Discovery Grant A3483.

## Foreword

These notes are an attempt to provide an entry into a subject that has not been very accessible. The problems of exposition are twofold. It is important to present motivation and background for the kind of problems that the trace formula is designed to solve. However, it is also important to provide the means for acquiring some of the basic techniques of the subject. I have tried to steer a middle course between these two sometimes divergent objectives. The reader should refer to earlier articles [**Lab2**], [**Lan14**], and the monographs [**Sho**], [**Ge**], for different treatments of some of the topics in these notes.

I had originally intended to write fifteen sections, corresponding roughly to fifteen lectures on the trace formula given at the Summer School. These sections comprise what has become Part I of the notes. They include much introductory material, and culminate in what we have called the coarse (or unrefined) trace formula. The coarse trace formula applies to a general connected, reductive algebraic group. However, its terms are too crude to be of much use as they stand.

Part II contains fifteen more sections. It has two purposes. One is to transform the trace formula of Part I into a refined formula, capable of yielding interesting information about automorphic representations. The other is to discuss some of the applications of the refined formula. The sections of Part II are considerably longer and more advanced. I hope that a familiarity with the concepts of Part I will allow a reader to deal with the more difficult topics in Part II. In fact, the later sections still include some introductory material. For example, §16, §22, and §27 contain heuristic discussions of three general problems, each of which requires a further refinement of the trace formula. Section 26 contains a general introduction to Langlands' principle of functoriality, to which many of the applications of the trace formula are directed.

We begin with a discussion of some constructions that are part of the foundations of the subject. In §1 we review the Selberg trace formula for compact quotient. In §2 we introduce the ring $\mathbb{A} = \mathbb{A}_F$ of adeles. We also try to illustrate why adelic algebraic groups $G(\mathbb{A})$, and their quotients $G(F)\backslash G(\mathbb{A})$, are more concrete objects than they might appear at first sight. Section 3 is devoted to examples related to §1 and §2. It includes a brief description of the Jacquet-Langlands correspondence between quaternion algebras and $GL(2)$. This correspondence is a striking example of the kind of application of which the trace formula is capable. It also illustrates the need for a trace formula for noncompact quotient.

In §4, we begin the study of noncompact quotient. We work with a general algebraic group $G$, since this was a prerequisite for the Summer School. However, we have tried to proceed gently, giving illustrations of a number of basic notions. For example, §5 contains a discussion of roots and weights, and the related objects needed for the study of noncompact quotient. To lend Part I an added appearance of simplicity, we work over the ground field $\mathbb{Q}$, instead of a general number field $F$.

The rest of Part I is devoted to the general theme of truncation. The problem is to modify divergent integrals so that they converge. At the risk of oversimplifying matters, we have tried to center the techniques of Part I around one basic result, Theorem 6.1. Corollary 10.1 and Theorem 11.1, for example, are direct corollaries of Theorem 6.1, as well as essential steps in the overall construction. Other results in Part I also depend in an essential way on either the statement of Theorem 6.1 or a key aspect of its proof. Theorem 6.1 itself asserts that a truncation of the function

$$K(x,x) = \sum_{\gamma \in G(\mathbb{Q})} f(x^{-1}\gamma x), \qquad f \in C_c^\infty\big(G(\mathbb{A})\big),$$

is integrable. It is the integral of this function over $G(\mathbb{Q})\backslash G(\mathbb{A})$ that yields a trace formula in the case of compact quotient. The integral of its truncation in the general case is what leads eventually to the coarse trace formula at the end of Part I.

After stating Theorem 6.1 in §6, we summarize the steps required to convert the truncated integral into some semblance of a trace formula. We sketch the proof of Theorem 6.1 in §8. The arguments here, as well as in the rest of Part I, are both geometric and combinatorial. We present them at varying levels of generality. However, with the notable exception of the review of Eisenstein series in §7, we have tried in all cases to give some feeling for what is the essential idea. For example, we often illustrate geometric points with simple diagrams, usually for the special case $G = SL(3)$. The geometry for $SL(3)$ is simple enough to visualize, but often complicated enough to capture the essential point in a general argument. I am indebted to Bill Casselman, and his flair for computer graphics, for the diagrams. The combinatorial arguments are used in conjunction with the geometric arguments to eliminate divergent terms from truncated functions. They rely ultimately on that simplest of cancellation laws, the binomial identity

$$\sum_{F \subset S} (-1)^{|F|} = \begin{cases} 0, & \text{if } S \neq \emptyset, \\ 1, & \text{if } S = \emptyset, \end{cases}$$

which holds for any finite set $S$ (Identity 6.2).

The parallel sections §11 and §15 from the later stages of Part I anticipate the general discussion of §16–21 in Part II. They provide refined formulas for "generic" terms in the coarse trace formula. These formulas are explicit expressions, whose local dependence on the given test function $f$ is relatively transparent. The first problem of refinement is to establish similar formulas for all of the terms. Because the remaining terms are indexed by conjugacy classes and representations that are singular, this problem is more difficult than any encountered in Part I. The solution requires new analytic techniques, both local and global. It also requires extensions of the combinatorial techniques of Part I, which are formulated in §17 as properties of $(G, M)$-families. We refer the reader to §16–21 for descriptions of the various results, as well as fairly substantial portions of their proofs.

The solution of the first problem yields a refined trace formula. We summarize this new formula in §22, in order to examine why it is still not satisfactory. The problem here is that its terms are not invariant under conjugation of $f$ by elements in $G(\mathbb{A})$. They are in consequence not determined by the values taken by $f$ at irreducible characters. We describe the solution of this second problem in §23. It yields an invariant trace formula, which we derive by modifying the terms in the refined, noninvariant trace formula so that they become invariant in $f$.

In §24–26 we pause to give three applications of the invariant trace formula. They are, respectively, a finite closed formula for the traces of Hecke operators on certain spaces, a term by term comparison of invariant trace formulas for general linear groups and central simple algebras, and cyclic base change of prime order for $GL(n)$. It is our discussion of base change that provides the opportunity to review Langlands' principle of functoriality.

The comparisons of invariant trace formulas in §25 and §26 are directed at special cases of functoriality. To study more general cases of functoriality, one requires a third refinement of the trace formula.

The remaining problem is that the terms of the invariant trace formula are not stable as linear forms in $f$. Stability is a subtler notion than invariance, and is part of Langlands' conjectural theory of endoscopy. We review it in §27. In §28 and §29 we describe the last of our three refinements. This gives rise to a stable trace formula, each of whose terms is stable in $f$. Taken together, the results of §29 can be regarded as a stabilization process, by which the invariant trace formula is decomposed into a stable trace formula, and an error term composed of stable trace formulas on smaller groups. The results are conditional upon the fundamental lemma. The proofs, conditional as they may be, are still too difficult to permit more than passing comment in §29.

The general theory of endoscopy includes a significant number of cases of functoriality. However, its avowed purpose is somewhat different. The principal aim of the theory is to analyze the internal structure of representations of a given group. Our last application is a broad illustration of what can be expected. In §30 we describe a classification of representations of quasisplit classical groups, both local and global, into packets. These results depend on the stable trace formula, and the fundamental lemma in particular. They also presuppose an extension of the stabilization of §29 to twisted groups.

As a means for investigating the general principle of functoriality, the theory of endoscopy has very definite limitations. We have devoted a word after §30 to some recent ideas of Langlands. The ideas are speculative, but they seem also to represent the best hope for attacking the general problem. They entail using the trace formula in ways that are completely new.

These notes are really somewhat of an experiment. The style varies from section to section, ranging between the technical and the discursive. The more difficult topics typically come in later sections. However, the progression is not always linear, or even monotonic. For example, the material in §13–§15, §19–§21, §23, and §25 is no doubt harder than much of the broader discussion in §16, §22, §26, and §27. The last few sections of Part II tend to be more discursive, but they are also highly compressed. This is the price we have had to pay for trying to get close to the frontiers. The reader should feel free to bypass the more demanding passages, at least initially, in order to develop an overall sense of the subject.

It would not have been possible to go very far by insisting on complete proofs. On the other hand, a survey of the results might have left a reader no closer to acquiring any of the basic techniques. The compromise has been to include something representative of as many arguments as possible. It might be a sketch of the general proof, a suggestive proof of some special case, or a geometric illustration by a diagram. For obvious reasons, the usual heading "PROOF" does not appear in the notes. However, each stated result is eventually followed by a small box

$\Box$, when the discussion that passes for a proof has come to an end. This ought to make the structure of each section more transparent. My hope is that a determined reader will be able to learn the subject by reinforcing the partial arguments here, when necessary, with the complete proofs in the given references.

# Part I. The Unrefined Trace Formula

## 1. The Selberg trace formula for compact quotient

Suppose that $H$ is a locally compact, unimodular topological group, and that $\Gamma$ is a discrete subgroup of $H$. The space $\Gamma\backslash H$ of right cosets has a right $H$-invariant Borel measure. Let $R$ be the unitary representation of $H$ by right translation on the corresponding Hilbert space $L^2(\Gamma\backslash H)$. Thus,

$$\big(R(y)\phi\big)(x) = \phi(xy), \qquad \phi \in L^2(\Gamma\backslash H),\ x, y \in H.$$

It is a fundamental problem to decompose $R$ explicitly into irreducible unitary representations. This should be regarded as a theoretical guidepost rather than a concrete goal, since one does not expect an explicit solution in general. In fact, even to state the problem precisely requires the theory of direct integrals.

The problem has an obvious meaning when the decomposition of $R$ is discrete. Suppose for example that $H$ is the additive group $\mathbb{R}$, and that $\Gamma$ is the subgroup of integers. The irreducible unitary representations of $\mathbb{R}$ are the one dimensional characters $x \to \mathrm{e}^{\lambda x}$, where $\lambda$ ranges over the imaginary axis $i\mathbb{R}$. The representation $R$ decomposes as direct sum over such characters, as $\lambda$ ranges over the subset $2\pi i\mathbb{Z}$ of $i\mathbb{R}$. More precisely, let $\widehat{R}$ be the unitary representation of $\mathbb{R}$ on $L^2(\mathbb{Z})$ defined by

$$\big(\widehat{R}(y)c\big)(n) = \mathrm{e}^{2\pi i n y}c(n), \qquad c \in L^2(\mathbb{Z}).$$

The correspondence that maps $\phi \in L^2(\mathbb{Z}\backslash\mathbb{R})$ to its set of Fourier coefficients

$$\widehat{\phi}(n) = \int_{\mathbb{Z}\backslash\mathbb{R}} \phi(x)\mathrm{e}^{-2\pi i n x}\mathrm{d}x, \qquad n \in \mathbb{Z},$$

is then a unitary isomorphism from $L^2(\mathbb{Z}\backslash\mathbb{R})$ onto $L^2(\mathbb{Z})$, which intertwines the representations $R$ and $\widehat{R}$. This is of course the Plancherel theorem for Fourier series.

The other basic example to keep in mind occurs where $H = \mathbb{R}$ and $\Gamma = \{1\}$. In this case the decomposition of $R$ is continuous, and is given by the Plancherel theorem for Fourier transforms. The general intuition that can inform us is as follows. For arbitrary $H$ and $\Gamma$, there will be some parts of $R$ that decompose discretely, and therefore behave qualitatively like the theory of Fourier series, and others that decompose continuously, and behave qualitatively like the theory of Fourier transforms.

In the general case, we can study $R$ by integrating it against a test function $f \in C_c(H)$. That is, we form the operator

$$R(f) = \int_H f(y)R(y)\mathrm{d}y$$

on $L^2(\Gamma\backslash H)$. We obtain

$$\begin{aligned}\big(R(f)\phi\big)(x) &= \int_H \big(f(y)R(y)\phi\big)(x)\mathrm{d}y \\ &= \int_H f(y)\phi(xy)\mathrm{d}y \\ &= \int_H f(x^{-1}y)\phi(y)dy \\ &= \int_{\Gamma\backslash H} \Big(\sum_{\gamma\in\Gamma} f(x^{-1}\gamma y)\Big)\phi(y)\mathrm{d}y,\end{aligned}$$

for any $\phi \in L^2(\Gamma\backslash H)$ and $x \in H$. It follows that $R(f)$ is an integral operator with kernel

$$K(x,y) = \sum_{\gamma\in\Gamma} f(x^{-1}\gamma y), \qquad x,y \in \Gamma\backslash H. \tag{1.1}$$

The sum over $\gamma$ is finite for any $x$ and $y$, since it may be taken over the intersection of the discrete group $\Gamma$ with the compact subset

$$x\,\mathrm{supp}(f)y^{-1}$$

of $H$.

For the rest of the section, we consider the special case that $\Gamma\backslash H$ is *compact*. The operator $R(f)$ then acquires two properties that allow us to investigate it further. The first is that $R$ decomposes discretely into irreducible representations $\pi$, with finite multiplicities $m(\pi, R)$. This is not hard to deduce from the spectral theorem for compact operators. Since the kernel $K(x,y)$ is a continuous function on the compact space $(\Gamma\backslash H)\times(\Gamma\backslash H)$, and is hence square integrable, the corresponding operator $R(f)$ is of Hilbert-Schmidt class. One applies the spectral theorem to the compact self adjoint operators attached to functions of the form

$$f(x) = (g * g^*)(x) = \int_H g(y)\overline{g(x^{-1}y)}\mathrm{d}y, \qquad g \in C_c(H).$$

The second property is that for many functions, the operator $R(f)$ is actually of trace class, with

$$\mathrm{tr}\,R(f) = \int_{\Gamma\backslash H} K(x,x)\mathrm{d}x. \tag{1.2}$$

If $H$ is a Lie group, for example, one can require that $f$ be smooth as well as compactly supported. Then $R(f)$ becomes an integral operator with smooth kernel on the compact manifold $\Gamma\backslash H$. It is well known that (1.2) holds for such operators.

Suppose that $f$ is such that (1.2) holds. Let $\{\Gamma\}$ be a set of representatives of conjugacy classes in $\Gamma$. For any $\gamma \in \Gamma$ and any subset $\Omega$ of $H$, we write $\Omega_\gamma$ for the centralizer of $\gamma$ in $\Omega$. We can then write

$$\begin{aligned}
\operatorname{tr}\big(R(f)\big) &= \int_{\Gamma\backslash H} K(x,x)\mathrm{d}x \\
&= \int_{\Gamma\backslash H} \sum_{\gamma\in\Gamma} f(x^{-1}\gamma x)\mathrm{d}x \\
&= \int_{\Gamma\backslash H} \sum_{\gamma\in\{\Gamma\}} \sum_{\delta\in\Gamma_\gamma\backslash\Gamma} f(x^{-1}\delta^{-1}\gamma\delta x)\mathrm{d}x \\
&= \sum_{\gamma\in\{\Gamma\}} \int_{\Gamma_\gamma\backslash H} f(x^{-1}\gamma x)\mathrm{d}x \\
&= \sum_{\gamma\in\{\Gamma\}} \int_{H_\gamma\backslash H} \int_{\Gamma_\gamma\backslash H_\gamma} f(x^{-1}u^{-1}\gamma u x)\mathrm{d}u\,\mathrm{d}x \\
&= \sum_{\gamma\in\{\Gamma\}} \operatorname{vol}(\Gamma_\gamma\backslash H_\gamma) \int_{H_\gamma\backslash H} f(x^{-1}\gamma x)\mathrm{d}x.
\end{aligned}$$

These manipulations follow from Fubini's theorem, and the fact that for any sequence $H_1 \subset H_2 \subset H$ of unimodular groups, a right invariant measure on $H_1\backslash H$ can be written as the product of right invariant measures on $H_2\backslash H$ and $H_1\backslash H_2$ respectively. We have obtained what may be regarded as a geometric expansion of $\operatorname{tr}\big(R(f)\big)$ in terms of conjugacy classes $\gamma$ in $\Gamma$. By restricting $R(f)$ to the irreducible subspaces of $L^2(\Gamma\backslash H)$, we obtain a spectral expansion of $R(f)$ in terms of irreducible unitary representations $\pi$ of $H$.

The two expansions $\operatorname{tr}\big(R(f)\big)$ provide an identity of linear forms

$$\sum_\gamma a_\Gamma^H(\gamma) f_H(\gamma) = \sum_\pi a_\Gamma^H(\pi) f_H(\pi), \tag{1.3}$$

where $\gamma$ is summed over (a set of representatives of) conjugacy classes in $\Gamma$, and $\pi$ is summed over (equivalence classes of) irreducible unitary representatives of $H$. The linear forms on the geometric side are invariant orbital integrals

$$f_H(\gamma) = \int_{H_\gamma\backslash H} f(x^{-1}\gamma x)\mathrm{d}x, \tag{1.4}$$

with coefficients

$$a_\Gamma^H(\gamma) = \operatorname{vol}(\Gamma_\gamma\backslash H_\gamma),$$

while the linear forms on the spectral side are irreducible characters

$$f_H(\pi) = \operatorname{tr}\big(\pi(f)\big) = \operatorname{tr}\Big(\int_H f(y)\pi(y)\mathrm{d}y\Big), \tag{1.5}$$

with coefficients

$$a_\Gamma^H(\pi) = m(\pi, R).$$

This is the Selberg trace formula for compact quotient.

We note that if $H = \mathbb{R}$ and $\Gamma = \mathbb{Z}$, the trace formula (1.3) reduces to the Poisson summation formula. For another example, we could take $H$ to be a finite group and $f(x)$ to be the character $\operatorname{tr}\pi(x)$ of an irreducible representation $\pi$ of $H$. In this case, (1.3) reduces to a special case of the Frobenius reciprocity theorem, which applies to the trivial one dimensional representation of the subgroup $\Gamma$ of $H$. (A minor extension of (1.3) specializes to the general form of Frobenius reciprocity.)

Some of Selberg's most striking applications of (1.3) were to the group $H = SL(2,\mathbb{R})$ of real, $(2 \times 2)$-matrices of determinant one. Suppose that $X$ is a compact Riemann surface of genus greater than 1. The universal covering surface of $X$ is then the upper half plane, which we identify as usual with the space of cosets $SL(2,\mathbb{R})/SO(2,\mathbb{R})$. (Recall that the compact orthogonal group $K = SO(2,\mathbb{R})$ is the stabilizer of $\sqrt{-1}$ under the transitive action of $SL(2,\mathbb{R})$ on the upper half plane by linear fractional transformations.) The Riemann surface becomes a space of double cosets

$$X = \Gamma\backslash H/K,$$

where $\Gamma$ is the fundamental group of $X$, embedding in $SL(2,\mathbb{R})$ as a discrete subgroup with compact quotient. By choosing left and right $K$-invariant functions $f \in C_c^\infty(H)$, Selberg was able to apply (1.3) to both the geometry and analysis of $X$.

For example, closed geodesics on $X$ are easily seen to be bijective with conjugacy classes in $\Gamma$. Given a large positive integer $N$, Selberg chose $f$ so that the left hand side of (1.3) approximated the number $g(N)$ of closed geodesics of length less than $N$. An analysis of the corresponding right hand side gave him an asymptotic formula for $g(N)$, with a sharp error term. Another example concerns the Laplace-Beltrami operator $\Delta$ attached to $X$. In this case, Selberg chose $f$ so that the right hand side of (1.3) approximated the number $h(N)$ of eigenvalues of $\Delta$ less than $N$. An analysis of the corresponding left hand side then provided a sharp asymptotic estimate for $h(N)$.

The best known discrete subgroup of $H = SL(2,\mathbb{R})$ is the group $\Gamma = SL(2,\mathbb{Z})$ of unimodular integral matrices. In this case, the quotient $\Gamma\backslash H$ is not compact. The example of $\Gamma = SL(2,\mathbb{Z})$ is of special significance because it comes with the supplementary operators introduced by Hecke. Hecke operators include a family of commuting operators $\{T_p\}$ on $L^2(\Gamma\backslash H)$, parametrized by prime numbers $p$, which commute also with the action of the group $H = SL(2,\mathbb{R})$. The families $\{c_p\}$ of simultaneous eigenvalues of Hecke operators on $L^2(\Gamma\backslash H)$ are known to be of fundamental arithmetic significance. Selberg was able to extend his trace formula (1.3) to this example, and indeed to many other quotients of rank 1. He also included traces of Hecke operators in his formulation. In particular, he obtained a finite closed formula for the trace of $T_p$ on any space of classical modular forms.

Selberg worked directly with Riemann surfaces and more general locally symmetric spaces, so the role of group theory in his papers is less explicit. We can refer the reader to the basic articles [**Sel1**] and [**Sel2**]. However, many of Selberg's results remain unpublished. The later articles [**DL**] and [**JL**, §16] used the language of group theory to formulate and extend Selberg's results for the upper half plane.

In the next section, we shall see how to incorporate the theory of Hecke operators into the general framework of (1.1). The connection is through adele groups, where Hecke operators arise in a most natural way. Our ultimate goal is to describe a general trace formula that applies to any adele group. The modern role of such a trace formula has changed somewhat from the original focus of Selberg. Rather than studying geometric and spectral data attached to a given group in isolation, one tries to compare such data for different groups. In particular, one would like to establish reciprocity laws among the fundamental arithmetic data associated to Hecke operators on different groups.

## 2. Algebraic groups and adeles

Suppose that $G$ is a connected reductive algebraic group over a number field $F$. For example, we could take $G$ to be the multiplicative group $GL(n)$ of invertible $(n \times n)$-matrices, and $F$ to be the rational field $\mathbb{Q}$. Our interest is in the general setting of the last section, with $\Gamma$ equal to $G(F)$. It is easy to imagine that this group could have arithmetic significance. However, it might not be at all clear how to embed $\Gamma$ discretely into a locally compact group $H$. To do so, we have to introduce the adele ring of $F$.

Suppose for simplicity that $F$ equals the rational field $\mathbb{Q}$. We have the usual absolute value $v_\infty(\cdot) = |\cdot|_\infty$ on $\mathbb{Q}$, and its corresponding completion $\mathbb{Q}_{v_\infty} = \mathbb{Q}_\infty = \mathbb{R}$. For each prime number $p$, there is also a $p$-adic absolute value $v_p(\cdot) = |\cdot|_p$ on $\mathbb{Q}$, defined by

$$|t|_p = p^{-r}, \qquad t = p^r ab^{-1},$$

for integers $r$, $a$ and $b$ with $(a,p) = (b,p) = 1$. One constructs its completion $\mathbb{Q}_{v_p} = \mathbb{Q}_p$ by a process identical to that of $\mathbb{R}$. As a matter of fact, $|\cdot|_p$ satisfies an enhanced form of the triangle inequality

$$|t_1 + t_2|_p \leq \max\left\{|t_1|_p, |t_2|_p\right\}, \qquad t_1, t_2 \in \mathbb{Q}.$$

This has the effect of giving the compact "unit ball"

$$\mathbb{Z}_p = \left\{t_p \in \mathbb{Q}_p : \ |t_p|_p \leq 1\right\}$$

in $\mathbb{Q}_p$ the structure of a *subring* of $\mathbb{Q}_p$. The completions $\mathbb{Q}_v$ are all locally compact fields. However, there are infinitely many of them, so their direct product is not locally compact. One forms instead the restricted direct product

$$\begin{aligned} \mathbb{A} = \prod_v^{\text{rest}} \mathbb{Q}_v \quad &= \quad \mathbb{R} \times \prod_p^{\text{rest}} \mathbb{Q}_p = \mathbb{R} \times \mathbb{A}_{\text{fin}} \\ &= \quad \left\{t = (t_v) : \ t_p = t_{v_p} \in \mathbb{Z}_p \text{ for almost all } p\right\}. \end{aligned}$$

Endowed with the natural direct limit topology, $\mathbb{A} = \mathbb{A}_\mathbb{Q}$ becomes a locally compact ring, called the adele ring of $\mathbb{Q}$. The diagonal image of $\mathbb{Q}$ in $\mathbb{A}$ is easily seen to be discrete. It follows that $H = G(\mathbb{A})$ is a locally compact group, in which $\Gamma = G(\mathbb{Q})$ embeds as a discrete subgroup. (See [**Tam2**].)

A similar construction applies to a general number field $F$, and gives rise to a locally compact ring $\mathbb{A}_F$. The diagonal embedding

$$\Gamma = G(F) \subset G(\mathbb{A}_F) = H$$

exhibits $G(F)$ as a discrete subgroup of the locally compact group $G(\mathbb{A}_F)$. However, we may as well continue to assume that $F = \mathbb{Q}$. This represents no loss of generality, since one can pass from $F$ to $\mathbb{Q}$ by restriction of scalars. To be precise, if $G_1$ is the algebraic group over $\mathbb{Q}$ obtained by restriction of scalars from $F$ to $\mathbb{Q}$, then $\Gamma = G(F) = G_1(\mathbb{Q})$, and $H = G(\mathbb{A}_F) = G_1(\mathbb{A})$.

We can define an *automorphic representation* $\pi$ of $G(\mathbb{A})$ informally to be an irreducible representation of $G(\mathbb{A})$ that "occurs in" the decomposition of $R$. This definition is not precise for the reason mentioned in §1, namely that there could be a part of $R$ that decomposes continuously. The formal definition [**Lan6**] is in fact quite broad. It includes not only irreducible unitary representations of $G(\mathbb{A})$ in the continuous spectrum, but also analytic continuations of such representations.

The introduction of adele groups appears to have imposed a new and perhaps unwelcome level of abstraction onto the subject. The appearance is illusory. Suppose for example that $G$ is a simple group over $\mathbb{Q}$. There are two possibilities: either $G(\mathbb{R})$ is noncompact (as in the case $G = SL(2)$), or it is not. If $G(\mathbb{R})$ is noncompact, the adelic theory for $G$ may be reduced to the study of of arithmetic quotients of $G(\mathbb{R})$. As in the case $G = SL(2)$ discussed at the end of §1, this is closely related to the theory of Laplace-Beltrami operators on locally symmetric Riemannian spaces attached to $G(\mathbb{R})$. If $G(\mathbb{R})$ is compact, the adelic theory reduces to the study of arithmetic quotients of a $p$-adic group $G(\mathbb{Q}_p)$. This in turn is closely related to the spectral theory of combinatorial Laplace operators on locally symmetric hypergraphs attached to the Bruhat-Tits building of $G(\mathbb{Q}_p)$.

These remarks are consequences of the theorem of strong approximation. Suppose that $S$ is a finite set of valuations of $\mathbb{Q}$ that contains the archimedean valuation $v_\infty$. For any $G$, the product

$$G(\mathbb{Q}_S) = \prod_{v \in S} G(\mathbb{Q}_v)$$

is a locally compact group. Let $K^S$ be an open compact subgroup of $G(\mathbb{A}^S)$, where

$$\mathbb{A}^S = \big\{ t \in \mathbb{A} : \ t_v = 0, \ v \in S \big\}$$

is the ring theoretic complement of $\mathbb{Q}_S$ in $\mathbb{A}$. Then $G(F_S)K^S$ is an open subgroup of $G(\mathbb{A})$.

Theorem 2.1. (a) *(Strong approximation) Suppose that $G$ is simply connected, in the sense that the topological space $G(\mathbb{C})$ is simply connected, and that $G'(\mathbb{Q}_S)$ is noncompact for every simple factor $G'$ of $G$ over $\mathbb{Q}$. Then*

$$G(\mathbb{A}) = G(\mathbb{Q}) \cdot G(\mathbb{Q}_S) K^S.$$

(b) *Assume only that $G'(\mathbb{Q}_S)$ is noncompact for every simple quotient $G'$ of $G$ over $\mathbb{Q}$. Then the set of double cosets*

$$G(\mathbb{Q}) \backslash G(\mathbb{A}) / G(\mathbb{Q}_S) K^S$$

*is finite.*

For a proof of (a) in the special case $G = SL(2)$ and $S = \{v_\infty\}$, see [**Shim**, Lemma 6.15]. The reader might then refer to [**Kne**] for a sketch of the general argument, and to [**P**] for a comprehensive treatment. Part (b) is essentially a corollary of (a). □

According to (b), we can write $G(\mathbb{A})$ as a disjoint union

$$G(\mathbb{A}) = \coprod_{i=1}^{n} G(\mathbb{Q}) \cdot x^i \cdot G(\mathbb{Q}_S) K^S,$$

for elements $x^1 = 1, x^2, \dots, x^n$ in $G(\mathbb{A}^S)$. We can therefore write

$$\begin{aligned} G(\mathbb{Q}) \backslash G(\mathbb{A}) / K^S &= \coprod_{i=1}^{n} \big( G(\mathbb{Q}) \backslash G(\mathbb{Q}) \cdot x^i \cdot G(\mathbb{Q}_S) K^S / K^S \big) \\ &\cong \coprod_{i=1}^{n} \big( \Gamma_S^i \backslash G(\mathbb{Q}_S) \big), \end{aligned}$$

for discrete subgroups

$$\Gamma_S^i = G(\mathbb{Q}_S) \cap \big(G(\mathbb{Q}) \cdot x^i K^S (x^i)^{-1}\big)$$

of $G(\mathbb{Q}_S)$. We obtain a $G(\mathbb{Q}_S)$-isomorphism of Hilbert spaces

$$L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})/K^S\big) \cong \bigoplus_{i=1}^n L^2\big(\Gamma_S^i \backslash G(\mathbb{Q}_S)\big). \tag{2.1}$$

The action of $G(\mathbb{Q}_S)$ on the two spaces on each side of (2.1) is of course by right translation. It corresponds to the action by right convolution on either space by functions in the algebra $C_c\big(G(\mathbb{Q}_S)\big)$. There is a supplementary convolution algebra, the Hecke algebra $\mathcal{H}\big(G(\mathbb{A}^S), K^S\big)$ of compactly supported functions on $G(\mathbb{A}^S)$ that are left and right invariant under translation by $K^S$. This algebra acts by right convolution on the left hand side of (2.1), in a way that clearly commutes with the action of $G(\mathbb{Q}_S)$. The corresponding action of $\mathcal{H}\big(G(\mathbb{A}^S), K^S\big)$ on the right hand side of (2.1) includes general analogues of the operators defined by Hecke on classical modular forms.

This becomes more concrete if $S = \{v_\infty\}$. Then $\mathbb{A}^S$ equals the subring $\mathbb{A}_{\mathrm{fin}} = \{t \in \mathbb{A} : \ t_\infty = 0\}$ of "finite adeles" in $\mathbb{A}$. If $G$ satisfies the associated noncompactness criterion of Theorem 2.1(b), and $K_0$ is an open compact subgroup of $G(\mathbb{A}_{\mathrm{fin}})$, we have a $G(\mathbb{R})$-isomorphism of Hilbert spaces

$$L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})/K_0\big) \cong \bigoplus_{i=1}^n L^2\big(\Gamma^i \backslash G(\mathbb{R})\big),$$

for discrete subgroups $\Gamma^1, \dots, \Gamma^n$ of $G(\mathbb{R})$. The Hecke algebra $\mathcal{H}\big(G(\mathbb{A}_{\mathrm{fin}}), K_0\big)$ acts by convolution on the left hand side, and hence also on the right hand side.

Hecke operators are really at the heart of the theory. Their properties can be formulated in representation theoretic terms. Any automorphic representation $\pi$ of $G(\mathbb{A})$ can be decomposed as a restricted tensor product

$$\pi = \bigotimes_v \pi_v, \tag{2.2}$$

where $\pi_v$ is an irreducible representation of the group $G(\mathbb{Q}_v)$. Moreover, for every valuation $v = v_p$ outside some finite set $S$, the representation $\pi_p = \pi_{v_p}$ is *unramified*, in the sense that its restriction to a suitable maximal compact subgroup $K_p$ of $G(\mathbb{Q}_p)$ contains the trivial representation. (See [**F**]. It is known that the trivial representation of $K_p$ occurs in $\pi_p$ with multiplicity at most one.) This gives rise to a maximal compact subgroup $K^S = \prod_{p \notin S} K_p$, a Hecke algebra

$$\mathcal{H}^S = \bigotimes_{p\notin S} \mathcal{H}_p = \bigotimes_{p\notin S} \mathcal{H}\big(G(\mathbb{Q}_p), K_p\big)$$

that is actually abelian, and an algebra homomorphism

$$c(\pi^S) = \bigotimes_{p\notin S} c(\pi_p) : \ \mathcal{H}^S = \bigotimes_{p\notin S} \mathcal{H}_p \ \longrightarrow \ \mathbb{C}. \tag{2.3}$$

Indeed, if $v^S = \bigotimes_{p \notin S} v_p$ belongs to the one-dimensional space of $K^S$-fixed vectors for the representation $\pi^S = \bigotimes_{p \notin S} \pi_p$, and $h^S = \bigotimes_{p \notin S} h_p$ belongs to $\mathcal{H}^S$, the vector

$$\pi^S(h^S)v^S = \bigotimes_{p \notin S} \big(\pi_p(h_p)v_p\big)$$

equals

$$c(\pi^S, h^S)v^S = \bigotimes_{p \notin S} \big(c(\pi_p, h_p)v_p\big).$$

This formula defines the homomorphism (2.3) in terms of the unramified representation $\pi^S$. Conversely, for any homomorphism $\mathcal{H}^S \to \mathbb{C}$, it is easy to see that there is a unique unramified representation $\pi^S$ of $G(\mathbb{A}^S)$ for which the formula holds.

The decomposition (2.2) actually holds for general irreducible representations $\pi$ of $G(\mathbb{A})$. In this case, the components can be arbitrary. However, the condition that $\pi$ be automorphic is highly rigid. It imposes deep relationships among the different unramified components $\pi_p$, or equivalently, the different homomorphisms $c(\pi_p) : \mathcal{H}_p \to \mathbb{C}$. These relationships are expected to be of fundamental arithmetic significance. They are summarized by Langlands's principle of functoriality [**Lan3**], and his conjecture that relates automorphic representations to motives [**Lan7**]. (For an elementary introduction to these conjectures, see [**A28**]. We shall review the principle of functoriality and its relationship with unramified representations in §26.) The general trace formula provides a means for analyzing some of the relationships.

The group $G(\mathbb{A})$ can be written as a direct product of the real group $G(\mathbb{R})$ with the totally disconnected group $G(\mathbb{A}_{\mathrm{fin}})$. We define

$$C_c^\infty\big(G(\mathbb{A})\big) = C_c^\infty\big(G(\mathbb{R})\big) \otimes C_c^\infty\big(G(\mathbb{A}_{\mathrm{fin}})\big),$$

where $C_c^\infty\big(G(\mathbb{R})\big)$ is the usual space of smooth, compactly supported functions on the Lie group $G(\mathbb{R})$, and $C_c^\infty\big(G(\mathbb{A}_{\mathrm{fin}})\big)$ is the space of locally constant, compactly supported, complex valued functions on the totally disconnected group $G(\mathbb{A}_{\mathrm{fin}})$. The vector space $C_c^\infty\big(G(\mathbb{A})\big)$ is an algebra under convolution, which is of course contained in the algebra $C_c\big(G(\mathbb{A})\big)$ of continuous, compactly supported functions on $G(\mathbb{A})$.

Suppose that $f$ belongs to $C_c^\infty\big(G(\mathbb{A})\big)$. We can choose a finite set of valuations $S$ satisfying the condition of Theorem 2.1(b), an open compact subgroup $K^S$ of $G(\mathbb{A}^S)$, and an open compact subgroup $K_{0,S}$ of the product

$$G(\mathbb{Q}_S^\infty) = \prod_{v \in S - \{v_\infty\}} G(\mathbb{Q}_v)$$

such that $f$ is bi-invariant under the open compact subgroup $K_0 = K_{0,S}K^S$ of $G(\mathbb{A}_{\mathrm{fin}})$. In particular, the operator $R(f)$ vanishes on the orthogonal complement of $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})/K^S\big)$ in $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$. We leave the reader the exercise of using (1.1) and (2.1) to identify $R(f)$ with an integral operator with smooth kernel on a finite disjoint union of quotients of $G(\mathbb{R})$.

Suppose, in particular, that $G(\mathbb{Q})\backslash G(\mathbb{A})$ happens to be compact. Then $R(f)$ may be identified with an integral operator with smooth kernel on a compact manifold. It follows that $R(f)$ is an operator of trace class, whose trace is given by

(1.2). The Selberg trace formula (1.3) is therefore valid for $f$, with $\Gamma = G(\mathbb{Q})$ and $H = G(\mathbb{A})$. (See [**Tam1**].)

## 3. Simple examples

We have tried to introduce adele groups as gently as possible, using the relations between Hecke operators and automorphic representations as motivation. Nevertheless, for a reader unfamiliar with such matters, it might take some time to feel comfortable with the general theory. To supplement the discussion of §2, and to acquire some sense of what one might hope to obtain in general, we shall look at a few concrete examples.

Consider first the simplest example of all, the case that $G$ equals the multiplicative group $GL(1)$. Then $G(\mathbb{Q}) = \mathbb{Q}^*$, while

$$G(\mathbb{A}) = \mathbb{A}^* = \big\{x \in \mathbb{A} : \ |x| \neq 0, \ |x_p|_p = 1 \text{ for almost all } p\big\}$$

is the multiplicative group of *ideles* for $\mathbb{Q}$. If $N$ is a positive integer with prime factorization $N = \prod_p p^{\mathrm{e}_p(N)}$, we write

$$K_N = \big\{k \in G(\mathbb{A}_{\mathrm{fin}}) = \mathbb{A}^*_{\mathrm{fin}} : \ |k_p - 1|_p \leq p^{-\mathrm{e}_p(N)} \text{ for all } p\big\}.$$

A simple exercise for a reader unfamiliar with adeles is to check directly that $K_N$ is an open compact subgroup of $\mathbb{A}^*_{\mathrm{fin}}$, that any open compact subgroup $K_0$ contains $K_N$ for some $N$, and that the abelian group

$$G(\mathbb{Q})\backslash G(\mathbb{A})/G(\mathbb{R})K_N = \mathbb{Q}^*\backslash\mathbb{A}^*/\mathbb{R}^*K_N$$

is finite. The quotient $G(\mathbb{Q})\backslash G(\mathbb{A}) = \mathbb{Q}^*\backslash\mathbb{A}^*$ is not compact. This is because the mapping

$$x \longrightarrow |x| = \prod_v |x_v|_v, \qquad x \in \mathbb{A}^*,$$

is a continuous surjective homomorphism from $\mathbb{A}^*$ to the multiplicative group $(\mathbb{R}^*)^0$ of positive real numbers, whose kernel

$$\mathbb{A}^1 = \big\{x \in \mathbb{A} : \ |x| = 1\big\}$$

contains $\mathbb{Q}^*$. The quotient $\mathbb{Q}^*\backslash\mathbb{A}^1$ is compact. Moreover, we can write the group $\mathbb{A}^*$ as a canonical direct product of $\mathbb{A}^1$ with the group $(\mathbb{R}^*)^0$. The failure of $\mathbb{Q}^*\backslash\mathbb{A}^*$ to be compact is therefore entirely governed by the multiplicative group $(\mathbb{R}^*)^0$ of positive real numbers.

An irreducible unitary representation of the abelian group $GL(1,\mathbb{A}) = \mathbb{A}^*$ is a homomorphism

$$\pi : \ \mathbb{A}^* \longrightarrow U(1) = \big\{z \in \mathbb{C}^* : \ |z| = 1\big\}.$$

There is a free action

$$s : \ \pi \longrightarrow \pi_s(x) = \pi(x)|x|^s, \qquad s \in i\mathbb{R},$$

of the additive group $i\mathbb{R}$ on the set of such $\pi$. The orbits of $i\mathbb{R}$ are bijective under the restriction mapping from $\mathbb{A}^*$ to $\mathbb{A}^1$ with the set of irreducible unitary representations of $\mathbb{A}^1$. A similar statement applies to the larger set of irreducible (not necessarily unitary) representations of $\mathbb{A}^*$, except that one has to replace $i\mathbb{R}$ with the additive group $\mathbb{C}$.

Returning to the case of a general group over $\mathbb{Q}$, we write $A_G$ for the largest central subgroup of $G$ over $\mathbb{Q}$ that is a $\mathbb{Q}$-split torus. In other words, $A_G$ is $\mathbb{Q}$-isomorphic to a direct product $GL(1)^k$ of several copies of $GL(1)$. The connected component $A_G(\mathbb{R})^0$ of 1 in $A_G(\mathbb{R})$ is isomorphic to the multiplicative group $\big((\mathbb{R}^*)^0\big)^k$, which in turn is isomorphic to the additive group $\mathbb{R}^k$. We write $X(G)_{\mathbb{Q}}$ for the additive group of homomorphisms $\chi: \ g \to g^\chi$ from $G$ to $GL(1)$ that are defined over $\mathbb{Q}$. Then $X(G)_{\mathbb{Q}}$ is a free abelian group of rank $k$. We also form the real vector space

$$\mathfrak{a}_G = \mathrm{Hom}_{\mathbb{Z}}\big(X(G)_{\mathbb{Q}}, \mathbb{R}\big)$$

of dimension $k$. There is then a surjective homomorphism

$$H_G: \ G(\mathbb{A}) \longrightarrow \mathfrak{a}_G,$$

defined by

$$\big\langle H_G(x), \chi\big\rangle = \big|\log(x^\chi)\big|, \qquad\qquad x \in G(\mathbb{A}), \ \chi \in X(G)_{\mathbb{Q}}.$$

The group $G(\mathbb{A})$ is a direct product of the normal subgroup

$$G(\mathbb{A})^1 = \big\{x \in G(\mathbb{A}) : \ H_G(x) = 0\big\}$$

with $A_G(\mathbb{R})^0$.

We also have the dual vector space $\mathfrak{a}_G^* = X(G)_{\mathbb{Q}} \otimes_{\mathbb{Z}} \mathbb{R}$, and its complexification $\mathfrak{a}_{G,\mathbb{C}}^* = X(G)_{\mathbb{Q}} \otimes \mathbb{C}$. If $\pi$ is an irreducible unitary representation of $G(\mathbb{A})$ and $\lambda$ belongs to $i\mathfrak{a}_G^*$, the product

$$\pi_\lambda(x) = \pi(x)\mathrm{e}^{\lambda(H_G(x))}, \qquad\qquad x \in G(\mathbb{A}),$$

is another irreducible unitary representation of $G(\mathbb{A})$. The set of associated $i\mathfrak{a}_G^*$-orbits is in bijective correspondence under the restriction mapping from $G(\mathbb{A})$ to $G(\mathbb{A})^1$ with the set of irreducible unitary representations of $G(\mathbb{A})^1$. A similar assertion applies the larger set of irreducible (not necessary unitary) representations, except that one has to replace $i\mathfrak{a}_G^*$ with the complex vector space $\mathfrak{a}_{G,\mathbb{C}}^*$.

In the case $G = GL(n)$, for example, we have

$$A_{GL(n)} = \left\{\begin{pmatrix} z & & 0 \\ & \ddots & \\ 0 & & z \end{pmatrix} : \ z \in GL(1)\right\} \cong GL(1).$$

The abelian group $X\big(GL(n)\big)_{\mathbb{Q}}$ is isomorphic to $\mathbb{Z}$, with canonical generator given by the determinant mapping from $GL(n)$ to $GL(1)$. The adelic group $GL(n,\mathbb{A})$ is a direct product of the two groups

$$GL(n,\mathbb{A})^1 = \big\{x \in GL(n,\mathbb{A}) : \ |\det(x)| = 1\big\}$$

and

$$A_{GL(n)}(\mathbb{R})^0 = \left\{\begin{pmatrix} r & & 0 \\ & \ddots & \\ 0 & & r \end{pmatrix} : \ r \in (\mathbb{R}^*)^0\right\}.$$

In general, $G(\mathbb{Q})$ is contained in the subgroup $G(\mathbb{A})^1$ of $G(\mathbb{A})$. The group $A_G(\mathbb{R})^0$ is therefore an immediate obstruction to $G(\mathbb{Q})\backslash G(\mathbb{A})$ being compact, as indeed it was in the simplest example of $G = GL(1)$. The real question is then whether the quotient $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is compact. When the answer is affirmative, the discussion above tells us that the trace formula (1.3) can be applied. It holds for $\Gamma = G(\mathbb{Q})$ and $H = G(\mathbb{A})^1$, with $f$ being the restriction to $G(\mathbb{A})^1$ of a function in $C_c^\infty\big(G(\mathbb{A})\big)$.

The simplest nonabelian example that gives compact quotient is the multiplicative group

$$G = \{x \in A : \ x \neq 0\}$$

of a quaternion algebra over $\mathbb{Q}$. By definition, $A$ is a four dimensional division algebra over $\mathbb{Q}$, with center $\mathbb{Q}$. It can be written in the form

$$A = \bigl\{x = x_0 + x_1 i + x_2 j + x_3 k : \ x_\alpha \in \mathbb{Q}\bigr\},$$

where the basis elements 1, $i$, $j$ and $k$ satisfy

$$ij = -ji = k, \quad i^2 = a, \quad j^2 = b,$$

for nonzero elements $a, b \in \mathbb{Q}^*$. Conversely, for any pair $a, b \in \mathbb{Q}^*$, the $\mathbb{Q}$-algebra defined in this way is either a quaternion algebra or is isomorphic to the matrix algebra $M_2(\mathbb{Q})$. For example, if $a = b = -1$, $A$ is a quaternion algebra, since $A \otimes_{\mathbb{Q}} \mathbb{R}$ is the classical Hamiltonian quaternion algebra over $\mathbb{R}$. On the other hand, if $a = b = 1$, the mapping

$$x \longrightarrow x_0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + x_1 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

is an isomorphism from $A$ onto $M_2(\mathbb{Q})$. For any $A$, one defines an automorphism

$$x \longrightarrow \bar{x} = x_0 - x_1 i - x_2 j - x_3 k$$

of $A$, and a multiplicative mapping

$$x \longrightarrow N(x) = x\bar{x} = x_0 - a x_1^2 - b x_2^2 + ab x_3^2$$

from $A$ to $\mathbb{Q}$. If $N(x) \neq 0$, $x^{-1}$ equals $N(x)^{-1}\bar{x}$. It follows that $x \in A$ is a unit if and only if $N(x) \neq 0$.

The description of a quaternion algebra $A$ in terms of rational numbers $a, b \in \mathbb{Q}^*$ has the obvious attraction of being explicit. However, it is ultimately unsatisfactory. Among other things, different pairs $a$ and $b$ can yield the same algebra $A$. There is a more canonical characterization in terms of the completions $A_v = A \otimes_{\mathbb{Q}} \mathbb{Q}_v$ at valuations $v$ of $\mathbb{Q}$. If $v = v_\infty$, we know that $A_v$ is isomorphic to either the matrix ring $M_2(\mathbb{R})$ or the Hamiltonian quaternion algebra over $\mathbb{R}$. A similar property holds for any other $v$. Namely, there is exactly one isomorphism class of quaternion algebras over $\mathbb{Q}_v$, so there are again two possibilities for $A_v$. Let $V$ be the set of valuations $v$ such that $A_v$ is a quaternion algebra. It is then known that $V$ is a finite set of *even* order. Conversely, for any nonempty set $V$ of even order, there is a unique isomorphism class of quaternion algebras $A$ over $\mathbb{Q}$ such that $A_v$ is a quaternion algebra for each $v \in V$ and a matrix algebra $M_2(\mathbb{Q}_v)$ for each $v$ outside $V$.

We digress for a moment to note that this characterization of quaternion algebras is part of a larger classification of reductive algebraic groups. The general classification over a number field $F$, and its completions $F_v$, is a beautiful union of class field theory with the structure theory of reductive groups. One begins with a group $G_s^*$ over $F$ that is split, in the sense that it has a maximal torus that splits over $F$. By a basic theorem of Chevalley, the groups $G_s^*$ are in bijective correspondence with reductive groups over an algebraic closure $\overline{F}$ of $F$, the classification of which reduces largely to that of complex semisimple Lie algebras. The general group $G$ over $F$ is obtained from $G_s^*$ by twisting the action of the Galois group $\mathrm{Gal}(\overline{F}/F)$ by automorphisms of $G_s^*$. It is a two stage process. One first constructs an "outer twist" $G^*$ of $G_s^*$ that is quasisplit, in the sense that it has a Borel subgroup that is defined over $F$. This is the easier step. It reduces to a knowledge of the group of outer automorphisms of $G_s^*$, something that is easy to describe in terms of the general structure of reductive groups. One then constructs an "inner twist" $G \xrightarrow{\psi} G^*$, where $\psi$ is an isomorphism such that for each $\sigma \in \mathrm{Gal}(\overline{F}/F)$, the composition

$$\alpha(\sigma) = \psi \circ \sigma(\psi)^{-1}$$

belongs to the group $\mathrm{Int}(G^*)$ of inner automorphisms of $G^*$. The role of class field theory is to classify the functions $\sigma \to \alpha(\sigma)$. More precisely, class field theory allows us to characterize the equivalence classes of such functions defined by the Galois cohomology set

$$H^1\big(F, \mathrm{Int}(G^*)\big) = H^1\big(\mathrm{Gal}(\overline{F}/F), \mathrm{Int}(G)^*(\overline{F})\big).$$

It provides a classification of the finite sets of local inner twists $H^1\big(F_v, \mathrm{Int}(G_v^*)\big)$, and a characterization of the image of the map

$$H^1\big(F, \mathrm{Int}(G^*)\big) \hookrightarrow \prod_v H^1\big(F, \mathrm{Int}(G_v^*)\big)$$

in terms of an explicit generalization of the parity condition for quaternion algebras. The map is injective, by the Hasse principle for the adjoint group $\mathrm{Int}(G^*)$. Its image therefore classifies the isomorphism classes of inner twists $G$ of $G^*$ over $F$.

In the special case above, the classification of quaternion algebras $A$ is equivalent to that of the algebraic groups $A^*$. In this case, $G^* = G_s^* = GL(2)$. In general, the theory is not especially well known, and goes beyond what we are assuming for this course. However, as a structural foundation for the Langlands program, it is well worth learning. A concise reference for a part of the theory is [**Ko5**, §1-2].

Let $G$ be the multiplicative group of a quaternion algebra $A$ over $\mathbb{Q}$, as above. The restriction of the norm mapping $N$ to $G$ is a generator of the group $X(G)_{\mathbb{Q}}$. In particular,

$$G(\mathbb{A})^1 = \big\{x \in G(\mathbb{A}) : \ |N(x)| = 1\big\}.$$

It is then not hard to see that the quotient $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is compact. (The reason is that $G$ has no proper parabolic subgroup over $\mathbb{Q}$, a point we shall discuss in the next section.) The Selberg trace formula (1.3) therefore holds for $\Gamma = G(\mathbb{Q})$, $H = G(\mathbb{A})^1$, and $f$ the restriction to $G(\mathbb{A})^1$ of a function in $C_c^\infty\big(G(\mathbb{A})\big)$. If $\Gamma(G)$ denotes the set of conjugacy classes in $G(\mathbb{Q})$, and $\Pi(G)$ is the set of equivalence classes of automorphic representations of $G$ (or more properly, restrictions to $G(\mathbb{A})^1$ of automorphic representations of $G(\mathbb{A})$), we have

$$\sum_{\gamma \in \Gamma(G)} a^G(\gamma) f_G(\gamma) = \sum_{\pi \in \Pi(G)} a^G(\pi) f_G(\pi), \qquad f \in C_c^\infty\big(G(\mathbb{A})\big), \tag{3.1}$$

for the volume $a^G(\gamma) = a_\Gamma^H(\gamma)$, the multiplicity $a^G(\pi) = a_\Gamma^H(\pi)$, the orbital integral $f_G(\gamma) = f_H(\gamma)$, and the character $f_G(\pi) = f_H(\pi)$. Jacquet and Langlands gave a striking application of this formula in §16 of their monograph [**JL**].

Any function in $C_c^\infty\big(G(\mathbb{A})\big)$ is a finite linear combination of products

$$f = \prod_v f_v, \qquad f_v \in C_c^\infty\big(G(\mathbb{Q}_v)\big).$$

Assume that $f$ is of this form. Then $f_G(\gamma)$ is a product of local orbital integrals $f_{v,G}(\gamma_v)$, where $\gamma_v$ is the image of $\gamma$ in the set $\Gamma(G_v)$ of conjugacy classes in $G(\mathbb{Q}_v)$, and $f_G(\pi)$ is a product of local characters $f_{v,G}(\pi_v)$, where $\pi_v$ is the component of $\pi$ in the set $\Pi(G_v)$ of equivalence classes of irreducible representations of $G(\mathbb{Q}_v)$. Let $V$ be the even set of valuations $v$ such that $G$ is not isomorphic to the group $G^* = GL(2)$ over $\mathbb{Q}_v$. If $v$ does not belong to $V$, the $\mathbb{Q}_v$-isomorphism from $G$ to $G^*$ is determined up to inner automorphisms. There is consequently a canonical bijection $\gamma_v \to \gamma_v^*$ from $\Gamma(G_v)$ to $\Gamma(G_v^*)$, and a canonical bijection $\pi_v \to \pi_v^*$ from $\Pi(G_v)$ to $\Pi(G_v^*)$. One can therefore define a function $f_v^* \in C_c^\infty(G_v^*)$ for every $v \notin V$ such that

$$f^*_{v,G^*}(\gamma_v^*) = f_{v,G}(\gamma_v)$$

and

$$f^*_{v,G^*}(\pi_v^*) = f_{v,G}(\pi_v),$$

for every $\gamma_v \in \Gamma(G_v)$ and $\pi_v \in \Pi(G_v)$. This suggested to Jacquet and Langlands the possibility of comparing (3.1) with the trace formula Selberg had obtained for the group $G^* = GL(2)$ with noncompact quotient.

If $v$ belongs to $V$, $G(\mathbb{Q}_v)$ is the multiplicative group of a quaternion algebra over $\mathbb{Q}_v$. In this case, there is a canonical bijection $\gamma_v \to \gamma_v^*$ from $\Gamma(G_v)$ onto the set $\Gamma_{\mathrm{ell}}(G_v^*)$ of semisimple conjugacy classes in $G^*(\mathbb{Q}_v)$ that are either central, or do not have eigenvalues in $\mathbb{Q}_v$. Moreover, there is a global bijection $\gamma \to \gamma^*$ from $\Gamma(G)$ onto the set of semisimple conjugacy classes $\gamma^* \in \Gamma(G^*)$ such that for every $v \in V$, $\gamma_v^*$ belongs to $\Gamma_{\mathrm{ell}}(G_v^*)$. For each $v \in V$, Jacquet and Langlands assigned a function $f_v^* \in C_c^\infty\big(G^*(\mathbb{Q}_v)\big)$ to $f_v$ such that

$$f^*_{v,G^*}(\gamma_v^*) = \begin{cases} f_{v,G}(\gamma_v), & \text{if } \gamma_v^* \in \Gamma_{\mathrm{ell}}(G_v^*), \\ 0, & \text{otherwise,} \end{cases} \tag{3.2}$$

for every (strongly) regular class $\gamma_v^* \in \Gamma_{\mathrm{reg}}(G_v^*)$. (An element is *strongly regular* if its centralizer is a maximal torus. The strongly regular orbital integrals of $f_v^*$ are known to determine the value taken by $f_v^*$ at any invariant distribution on $G^*(\mathbb{Q}_v)$.) This allowed them to attach a function

$$f^* = \prod_v f_v^*$$

in $C_c^\infty\big(G^*(\mathbb{A})\big)$ to the original function $f$. They then observed that

$$f^*_{G^*}(\gamma^*) = \begin{cases} f_G(\gamma), & \text{if } \gamma^* \text{ is the image of } \gamma \in \Gamma(G), \\ 0, & \text{otherwise,} \end{cases} \tag{3.3}$$

for any class $\gamma^* \in \Gamma(G^*)$.

It happens that Selberg's formula for the group $G^* = GL(2)$ contains a number of supplementary terms, in addition to analogues of the terms in (3.1). However, Jacquet and Langlands observed that the local vanishing conditions (3.2) force all of the supplementary terms to vanish. They then used (3.3) to deduce that the remaining terms on the geometric side equaled the corresponding terms on the geometric side of (3.1). This left only a spectral identity

$$\sum_{\pi \in \Pi(G)} m(\pi, R)\mathrm{tr}\big(\pi(f)\big) = \sum_{\pi^* \in \Pi(G^*)} m(\pi^*, R^*_{\mathrm{disc}})\mathrm{tr}\big(\pi^*(f^*)\big), \tag{3.4}$$

where $R^*_{\mathrm{disc}}$ is the subrepresentation of the regular representation of $G^*(\mathbb{A})^1$ on $L^2\big(G^*(\mathbb{Q})\backslash G^*(\mathbb{A})^1\big)$ that decomposes discretely. By setting $f = f_S f^S$, for a fixed finite set $S$ of valuations containing $V \cup \{v_\infty\}$, and a fixed function $f_S \in C_c^\infty\big(G(\mathbb{Q}_S)\big)$, one can treat (3.4) as an identity of linear forms in a variable function $f^S$ belonging to the Hecke algebra $\mathcal{H}(G^S, K^S)$. Jacquet and Langlands used it to establish an injective global correspondence $\pi \to \pi^*$ of automorphic representations, with $\pi_v^* = \pi_v$ for each $v \notin V$. They also obtained an injective local correspondence $\pi_v \to \pi_v^*$ of irreducible representations for each $v \in V$, which is compatible with the global correspondence, and also the local correspondence $f_v \to f_v^*$ of functions. Finally, they gave a simple description of the images of both the local and global correspondences of representations.

The Jacquet-Langlands correspondence is remarkable for both the power of its assertions and the simplicity of its proof. It tells us that the arithmetic information carried by unramified components $\pi_p$ of automorphic representations $\pi$ of $G(\mathbb{A})$, whatever form it might take, is included in the information carried by automorphic representations $\pi^*$ of $G^*(\mathbb{A})$. In the case $v_\infty \notin V$, it also implies a correspondence between spectra of Laplacians on certain compact Riemann surfaces, and discrete spectra of Laplacians on noncompact surfaces. The Jacquet-Langlands correspondence is a simple prototype of the higher reciprocity laws one might hope to deduce from the trace formula. In particular, it is a clear illustration of the importance of having a trace formula for noncompact quotient.

## 4. Noncompact quotient and parabolic subgroups

If $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is not compact, the two properties that allowed us to derive the trace formula (1.3) fail. The regular representation $R$ does not decompose discretely, and the operators $R(f)$ are not of trace class. The two properties are closely related, and are responsible for the fact that the integral (1.2) generally diverges. To see what goes wrong, consider the case that $G = GL(2)$, and take $f$ to be the restriction to $H = G(\mathbb{A})^1$ of a nonnegative function in $C_c^\infty\big(G(\mathbb{A})\big)$. If the integral (1.2) were to converge, the double integral

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_{\gamma \in G(\mathbb{Q})} f(x^{-1}\gamma x)\mathrm{d}x$$

would be finite. Using Fubini's theorem to justify again the manipulations of §1, we would then be able to write the double integral as

$$\sum_{\gamma \in \{G(\mathbb{Q})\}} \mathrm{vol}\big(G(\mathbb{Q})_\gamma \backslash G(\mathbb{A})^1_\gamma\big) \int_{G(\mathbb{A})^1_\gamma \backslash G(\mathbb{A})^1} f(x^{-1}\gamma x)\mathrm{d}x.$$

As it happens, however, the summand corresponding to $\gamma$ is often infinite.

Sometimes the volume of $G(\mathbb{Q})_\gamma \backslash G(\mathbb{A})^1_\gamma$ is infinite. Suppose that $\gamma = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}$, for a pair of distinct elements $\gamma_1$ and $\gamma_2$ in $\mathbb{Q}^*$. Then

$$G_\gamma = \left\{ \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix} : \ y_1, y_2 \in GL(1) \right\} \cong GL(1) \times GL(1),$$

so that

$$G(\mathbb{A})^1_\gamma \cong \big\{(y_1, y_2) \in (\mathbb{A}^*)^2 : \ |y_1||y_2| = 1\big\},$$

and

$$G(\mathbb{Q})_\gamma \backslash G(\mathbb{A})^1_\gamma \cong (\mathbb{Q}^*\backslash \mathbb{A}^1) \times (\mathbb{Q}^*\backslash \mathbb{A}^*) \cong (\mathbb{Q}^*\backslash \mathbb{A}^1)^2 \times (\mathbb{R}^*)^0.$$

An invariant measure on the left hand quotient therefore corresponds to a Haar measure on the abelian group on the right. Since this group is noncompact, the quotient has infinite volume.

Sometimes the integral over $G(\mathbb{A})^1_\gamma \backslash G(\mathbb{A})^1$ diverges. Suppose that $\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then

$$G(\mathbb{A})_\gamma = \left\{ \begin{pmatrix} z & y \\ 0 & z \end{pmatrix} : \ y \in \mathbb{A}, \ z \in \mathbb{A}^* \right\}$$

The computation of the integral

$$\int_{G(\mathbb{A})^1_\gamma \backslash G(\mathbb{A})^1} f(x^{-1}\gamma x)\mathrm{d}x = \int_{G(\mathbb{A})_\gamma \backslash G(\mathbb{A})} f(x^{-1}\gamma x)\mathrm{d}x$$

is a good exercise in understanding relations among the Haar measures $d^*a$, $du$ and $dx$ on $\mathbb{A}^*$, $\mathbb{A}$, and $G(\mathbb{A})$, respectively. One finds that the integral equals

$$\int_{G_\gamma(\mathbb{A}) \backslash P_0(\mathbb{A})} \int_{P_0(\mathbb{A}) \backslash G(\mathbb{A})} f(k^{-1}p^{-1}\gamma pk) d_\ell p \mathrm{d}k,$$

where $P_0(\mathbb{A})$ is the subgroup of upper triangular matrices

$$\left\{ p = \begin{pmatrix} a^* & u \\ 0 & b^* \end{pmatrix} : \ a^*, b^* \in \mathbb{A}^*, \ u \in \mathbb{A} \right\},$$

with left Haar measure

$$d_\ell p = |a^*|^{-1} da^* db^* du,$$

and $dk$ is a Borel measure on the compact space $P_0(\mathbb{A}) \backslash G(\mathbb{A})$. The integral then reduces to an expression

$$c(f) \prod_p (1 - p^{-1})^{-1} = c(f) \left( \sum_{n=1}^{\infty} \frac{1}{n} \right),$$

where

$$c(f) = c_0 \int_{P_0(\mathbb{A}) \backslash G(\mathbb{A})} \int_{\mathbb{A}} f\left( k^{-1} \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} k \right) \mathrm{d}u \mathrm{d}k,$$

for a positive constant $c_0$. In particular, the integral is generally infinite.

Observe that the nonconvergent terms in the case $G = GL(2)$ both come from conjugacy classes in $GL(2, \mathbb{Q})$ that intersect the parabolic subgroup $P_0$ of upper triangular matrices. This suggests that rational parabolic subgroups are responsible for the difficulties encountered in dealing with noncompact quotient. Our suspicion is reinforced by the following characterization, discovered independently by Borel and Harish-Chandra [**BH**] and Mostow and Tamagawa [**MT**]. For a general group $G$ over $\mathbb{Q}$, the quotient $G(\mathbb{Q}) \backslash G(\mathbb{A})^1$ is noncompact if and only if $G$ has a proper parabolic subgroup $P$ defined over $\mathbb{Q}$.

We review some basic properties of parabolic subgroups, many of which are discussed in the chapter [**Mur**] in this volume. We are assuming now that $G$ is a general connected reductive group over $\mathbb{Q}$. A *parabolic subgroup* of $G$ is an algebraic subgroup $P$ such that $P(\mathbb{C}) \backslash G(\mathbb{C})$ is compact. We consider only parabolic subgroups $P$ that are defined over $\mathbb{Q}$. Any such $P$ has a Levi decomposition $P = MN_P$, which is a semidirect product of a reductive subgroup $M$ of $G$ over $\mathbb{Q}$ with a normal unipotent subgroup $N_P$ of $G$ over $\mathbb{Q}$. The *unipotent radical* $N_P$ is uniquely determined by $P$, while the *Levi component* $M$ is uniquely determined up to conjugation by $P(\mathbb{Q})$.

Let $P_0$ be a fixed minimal parabolic subgroup of $G$ over $\mathbb{Q}$, with a fixed Levi decomposition $P_0 = M_0N_0$. Any subgroup $P$ of $G$ that contains $P_0$ is a parabolic subgroup that is defined over $\mathbb{Q}$. It is called a *standard parabolic subgroup* (relative to $P_0$). The set of standard parabolic subgroups of $G$ is finite, and is a set of representatives of the set of all $G(\mathbb{Q})$-conjugacy classes of parabolic subgroups over $\mathbb{Q}$. A standard parabolic subgroup $P$ has a *canonical* Levi decomposition $P = M_PN_P$, where $M_P$ is the unique Levi component of $P$ that contains $M_0$. Given $P$, we can form the central subgroup $A_P = A_{M_P}$ of $M_P$, the real vector space $\mathfrak{a}_P = \mathfrak{a}_{M_P}$, and the surjective homomorphism $H_P = H_{M_P}$ from $M_P(\mathbb{A})$ onto $\mathfrak{a}_P$. In case $P = P_0$, we often write $A_0 = A_{P_0}$, $\mathfrak{a}_0 = \mathfrak{a}_{P_0}$ and $H_0 = H_{P_0}$.

In the example $G = GL(n)$, one takes $P_0$ to be the Borel subgroup of upper triangular matrices. The unipotent radical $N_0$ of $P_0$ is the subgroup of unipotent upper triangular matrices. For the Levi component $M_0$, one takes the subgroup of diagonal matrices. There is then a bijection

$$P \longleftrightarrow (n_1, \dots, n_p)$$

between standard parabolic subgroups $P$ of $G = GL(n)$ and partitions $(n_1, \dots, n_p)$ of $n$. The group $P$ is the subgroup of block upper triangular matrices associated to $(n_1, \dots, n_p)$. The unipotent radical of $P$ is the corresponding subgroup

$$N_P = \left\{ \begin{pmatrix} \underline{I_{n_1}|} & & * \\ & \ddots & \\ 0 & & |\overline{I_{n_p}} \end{pmatrix} \right\}$$

of block unipotent matrices, the canonical Levi component is the subgroup

$$M_P = \left\{ m = \begin{pmatrix} \underline{m_1|} & & 0 \\ & \ddots & \\ 0 & & |\overline{m_p} \end{pmatrix} : \ m_i \in GL(n_i) \right\}$$

of block diagonal matrices, while

$$A_P = \left\{ a = \begin{pmatrix} \underline{a_1 I_{n_1}|} & & 0 \\ & \ddots & \\ 0 & & |\overline{a_p I_{n_p}} \end{pmatrix} : \ a_i \in GL(1) \right\}.$$

Naturally, $I_k$ stands here for the identity matrix of rank $k$. The free abelian group $X(M_P)_{\mathbb{Q}}$ attached to $M_P$ has a canonical basis of rational characters

$$\chi_i : \ m \longrightarrow \det(m_i), \qquad m \in M_P, \ 1 \leq i \leq p.$$

We are free to use the basis $\frac{1}{n_1}\chi_1, \dots, \frac{1}{n_p}\chi_p$ of the vector space $\mathfrak{a}_P^*$, and the corresponding dual basis of $\mathfrak{a}_P$, to identify both $\mathfrak{a}_P^*$ and $\mathfrak{a}_P$ with $\mathbb{R}^p$. With this interpretation, the mapping $H_P$ takes the form

$$H_P(m) = \left( \frac{1}{n_1} \log|\det m_1|, \dots, \frac{1}{n_p} \log|\det m_p| \right), \qquad m \in M_P(\mathbb{A}).$$

It follows that

$$H_P(a) = \big( \log|a_1|, \dots, \log|a_p| \big), \qquad a \in A_P(\mathbb{A}).$$

For general $G$, we have a variant of the regular representation $R$ for any standard parabolic subgroup $P$. It is the regular representation $R_P$ of $G(\mathbb{A})$ on $L^2\big(N_P(\mathbb{A})M_P(\mathbb{Q})\backslash G(\mathbb{A})\big)$, defined by

$$\big(R_P(y)\phi\big)(x) = \phi(xy), \qquad \phi \in L^2\big(N_P(\mathbb{A})M_P(\mathbb{Q})\backslash G(\mathbb{A})\big),\ x, y \in G(\mathbb{A}).$$

Using the language of induced representations, we can write

$$R_P = \mathrm{Ind}_{N_P(\mathbb{A})M_P(\mathbb{Q})}^{G(\mathbb{A})}(1_{N_P(\mathbb{A})M_P(\mathbb{Q})}) \cong \mathrm{Ind}_{P(\mathbb{A})}^{G(\mathbb{A})}(1_{N_P(\mathbb{A})} \otimes R_{M_P}),$$

where $\mathrm{Ind}_K^H(\cdot)$ denotes a representation of $H$ induced from a subgroup $K$, and $1_K$ denotes the trivial one dimensional representation of $K$. We can of course integrate $R_P$ against any function $f \in C_c^\infty\big(G(\mathbb{A})\big)$. This gives an operator $R_P(f)$ on the Hilbert space $L^2\big(N_P(\mathbb{A})M_P(\mathbb{Q})\backslash G(\mathbb{A})\big)$. Arguing as in the special case $R = R_G$ of §1, we find that $R_P(f)$ is an integral operator with kernel

$$K_P(x,y) = \int_{N_P(\mathbb{A})} \sum_{\gamma \in M_P(\mathbb{Q})} f(x^{-1}\gamma n y)\mathrm{d}n, \qquad x, y \in N_P(\mathbb{A})M_P(\mathbb{Q})\backslash G(\mathbb{A}). \tag{4.1}$$

We have seen that the diagonal value $K(x,x) = K_G(x,x)$ of the original kernel need not be integrable over $x \in G(\mathbb{Q})\backslash G(\mathbb{A})^1$. We have also suggested that parabolic subgroups are somehow responsible for this failure. It makes sense to try to modify $K(x,x)$ by adding correction terms indexed by proper parabolic subgroups $P$. The correction terms ought to be supported on some small neighbourhood of infinity, so that they do not affect the values taken by $K(x,x)$ on some large compact subset of $G(\mathbb{Q})\backslash G(\mathbb{A})^1$. The diagonal value $K_P(x,x)$ of the kernel of $R_P(f)$ provides a natural function for any $P$. However, $K_P(x,x)$ is invariant under left translation of $x$ by the group $N_P(\mathbb{A})M_P(\mathbb{Q})$, rather than $G(\mathbb{Q})$. One could try to rectify this defect by summing $K_P(\delta x, \delta x)$ over elements $\delta$ in $P(\mathbb{Q})\backslash G(\mathbb{Q})$. However, this sum does not generally converge. Even if it did, the resulting function on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ would not be supported on a small neighbourhood of infinity. The way around this difficulty will be to multiply $K_P(x,x)$ by a certain characteristic function on $N_P(\mathbb{A})M_P(\mathbb{Q})\backslash G(\mathbb{A})$ that is supported on a small neighbourhood of infinity, and which depends on a choice of maximal compact subgroup $K$ of $G(\mathbb{A})$.

In case $G = GL(n)$, the product

$$K = O(n,\mathbb{R}) \times \prod_p GL(n,\mathbb{Z}_p)$$

is a maximal compact subgroup of $G(\mathbb{A})$. According to the Gramm-Schmidt orthogonalization lemma of linear algebra, we can write

$$GL(n,\mathbb{R}) = P_0(\mathbb{R})O(n,\mathbb{R}).$$

A variant of this process, applied to the height function

$$\|v\|_p = \max\{|v_i|_p :\ 1 \le i \le n\}, \qquad v \in \mathbb{Q}_p^n,$$

on $\mathbb{Q}_p^n$ instead of the standard inner product on $\mathbb{R}^n$, gives a decomposition

$$GL(n,\mathbb{Q}_p) = P_0(\mathbb{Q}_p)GL(n,\mathbb{Z}_p),$$

for any $p$. It follows that $GL(n,\mathbb{A})$ equals $P_0(\mathbb{A})K$.

These properties carry over to our general group $G$. We choose a suitable maximal compact subgroup

$$K = \prod_v K_v, \qquad K_v \subset G(\mathbb{Q}_v),$$

of $G(\mathbb{A})$, with $G(\mathbb{A}) = P_0(\mathbb{A})K$ [**Ti**, (3.3.2), (3.9], [**A5**, p. 9]. We fix $K$, and consider a standard parabolic subgroup $P$ of $G$. Since $P$ contains $P_0$, we obtain a decomposition

$$G(\mathbb{A}) = P(\mathbb{A})K = N_P(\mathbb{A})M_P(\mathbb{A})K = N_P(\mathbb{A})M_P(\mathbb{A})^1 A_P(\mathbb{R})^0 K.$$

We then define a continuous mapping

$$H_P : \ G(\mathbb{A}) \ \longrightarrow \ \mathfrak{a}_P$$

by setting

$$H_P(nmk) = H_{M_P}(m), \qquad n \in N_P(\mathbb{A}),\ m \in M_P(\mathbb{A}),\ k \in K.$$

We shall multiply the kernel $K_P(x,x)$ by the preimage under $H_P$ of the characteristic function of a certain cone in $\mathfrak{a}_P$.

## 5. Roots and weights

We have fixed a minimal parabolic subgroup $P_0$ of $G$, and a maximal compact subgroup $K$ of $G(\mathbb{A})$. We want to use these objects to modify the kernel function $K(x,x)$ so that it becomes integrable. To prepare for the construction, as well as for future geometric arguments, we review some properties of roots and weights.

The restriction homomorphism $X(G)_{\mathbb{Q}} \to X(A_G)_{\mathbb{Q}}$ is injective, and has finite cokernel. If $G = GL(n)$, for example, the homomorphism corresponds to the injection $z \to nz$ of $\mathbb{Z}$ into itself. We therefore obtain a canonical linear isomorphism

$$\mathfrak{a}_P^* = X(M_P)_{\mathbb{Q}} \otimes \mathbb{R} \ \xrightarrow{\sim} \ X(A_P)_{\mathbb{Q}} \otimes \mathbb{R}. \tag{5.1}$$

Now suppose that $P_1$ and $P_2$ are two standard parabolic subgroups, with $P_1 \subset P_2$. There are then $\mathbb{Q}$-rational embeddings

$$A_{P_2} \subset A_{P_1} \subset M_{P_1} \subset M_{P_2}.$$

The restriction homomorphism $X(M_{P_2})_{\mathbb{Q}} \to X(M_{P_1})_{\mathbb{Q}}$ is injective. It provides a linear injection $\mathfrak{a}_{P_2}^* \hookrightarrow \mathfrak{a}_{P_1}^*$ and a dual linear surjection $\mathfrak{a}_{P_1} \mapsto \mathfrak{a}_{P_2}$. We write $\mathfrak{a}_{P_1}^{P_2} \subset \mathfrak{a}_{P_1}$ for the kernel of the latter mapping. The restriction homomorphism $X(A_{P_1})_{\mathbb{Q}} \to X(A_{P_2})_{\mathbb{Q}}$ is surjective, and extends to a surjective mapping from $X(A_{P_1})_{\mathbb{Q}} \otimes \mathbb{R}$ to $X(A_{P_2})_{\mathbb{Q}} \otimes \mathbb{R}$. It thus provides a linear surjection $\mathfrak{a}_{P_1}^* \mapsto \mathfrak{a}_{P_2}^*$, and a dual linear injection $\mathfrak{a}_{P_2} \hookrightarrow \mathfrak{a}_{P_1}$. Taken together, the four linear mappings yield split exact sequences

$$0 \ \longrightarrow \ \mathfrak{a}_{P_2}^* \ \rightleftarrows \ \mathfrak{a}_{P_1}^* \ \longrightarrow \ \mathfrak{a}_{P_1}^*/\mathfrak{a}_{P_2}^* \ \longrightarrow \ 0$$

and

$$0 \ \longrightarrow \ \mathfrak{a}_{P_1}^{P_2} \ \longrightarrow \ \mathfrak{a}_{P_1} \ \rightleftarrows \ \mathfrak{a}_{P_2} \ \longrightarrow \ 0$$

of real vector spaces. We may therefore write

$$\mathfrak{a}_{P_1} = \mathfrak{a}_{P_2} \oplus \mathfrak{a}_{P_1}^{P_2}$$

and

$$\mathfrak{a}_{P_1}^* = \mathfrak{a}_{P_2}^* \oplus (\mathfrak{a}_{P_1}^{P_2})^*.$$

For any $P$, we write $\Phi_P$ for the set of roots of $(P, A_P)$. We also write $\mathfrak{n}_P$ for the Lie algebra of $N_P$. Then $\Phi_P$ is a finite subset of nonzero elements in $X(A_P)_{\mathbb{Q}}$ that parametrizes the decomposition

$$\mathfrak{n}_P = \bigoplus_{\alpha \in \Phi_P} \mathfrak{n}_\alpha$$

of $\mathfrak{n}_P$ into eigenspaces under the adjoint action

$$\mathrm{Ad}: \ A_P \longrightarrow GL(\mathfrak{n}_P)$$

of $A_P$. By definition,

$$\mathfrak{n}_\alpha = \big\{X_\alpha \in \mathfrak{n}_P : \ \mathrm{Ad}(a)X_\alpha = a^\alpha X_\alpha, \ a \in A_P\big\},$$

for any $\alpha \in \Phi_P$. We identify $\Phi_P$ with a subset of $\mathfrak{a}_P^*$ under the canonical mappings

$$\Phi_P \subset X(A_P)_{\mathbb{Q}} \subset X(A_P)_{\mathbb{Q}} \otimes \mathbb{R} \simeq \mathfrak{a}_P^*.$$

If $H$ belongs to the subspace $\mathfrak{a}_G$ of $\mathfrak{a}_P$, $\alpha(H) = 0$ for each $\alpha \in \Phi_P$, so $\Phi_P$ is contained in the subspace $(\mathfrak{a}_P^G)^*$ of $\mathfrak{a}_P^*$. As is customary, we define a vector

$$\rho_P = \frac{1}{2} \sum_{\alpha \in \Phi_P} (\dim \, \mathfrak{n}_\alpha)\alpha$$

in $(\mathfrak{a}_P^G)^*$. We leave the reader to check that left and right Haar measures on the group $P(\mathbb{A})$ are related by

$$d_\ell p = \mathrm{e}^{2\rho(H_P(p))} d_r p, \qquad\qquad p \in P(\mathbb{A}).$$

In particular, the group $P(\mathbb{A})$ is not unimodular, if $P \neq G$.

We write $\Phi_0 = \Phi_{P_0}$. The pair

$$(V, R) = \big((\mathfrak{a}_{P_0}^G)^*, \Phi_0 \cup (-\Phi_0)\big)$$

is a root system [**Ser2**], for which $\Phi_0$ is a system of positive roots. We write $W_0 = W_0^G$ for the Weyl group of $(V, R)$. It is the finite group generated by reflections about elements in $\Phi_0$, and acts on the vector spaces $V = (\mathfrak{a}_{P_0}^G)^*$, $\mathfrak{a}_0^* = \mathfrak{a}_{P_0}^*$, and $\mathfrak{a}_0 = \mathfrak{a}_{P_0}$. We also write $\Delta_0 \subset \Phi_0$ for the set of simple roots attached to $\Phi_0$. Then $\Delta_0$ is a basis of the real vector space $(\mathfrak{a}_0^G)^* = (\mathfrak{a}_{P_0}^G)^*$. Any element $\beta \in \Phi_0$ can be written uniquely

$$\beta = \sum_{\alpha \in \Delta_0} n_\alpha \alpha,$$

for nonnegative integers $n_\alpha$. The corresponding set

$$\Delta_0^\vee = \{\alpha^\vee : \ \alpha \in \Delta_0\}$$

of simple coroots is a basis of the vector space $\mathfrak{a}_0^G = \mathfrak{a}_{P_0}^G$. We write

$$\widehat{\Delta}_0 = \{\varpi_\alpha : \ \alpha \in \Delta_0\}$$

for the set of simple weights, and

$$\widehat{\Delta}_0^\vee = \{\varpi_\alpha^\vee : \ \alpha \in \Delta_0\}$$

for the set of simple co-weights. In other words, $\widehat{\Delta}_0$ is the basis of $(\mathfrak{a}_0^G)^*$ dual to $\Delta_0^\vee$, and $\widehat{\Delta}_0^\vee$ is the basis of $\mathfrak{a}_0^G$ dual to $\Delta_0$.

Standard parabolic subgroups are parametrized by subsets of $\Delta_0$. More precisely, there is an order reversing bijection $P \leftrightarrow \Delta_0^P$ between standard parabolic subgroups $P$ of $G$ and subsets $\Delta_0^P$ of $\Delta_0$, such that

$$\mathfrak{a}_P = \{H \in \mathfrak{a}_0 : \ \alpha(H) = 0, \ \alpha \in \Delta_0^P\}.$$

For any $P$, $\Delta_0^P$ is a basis of the space $\mathfrak{a}_{P_0}^P = \mathfrak{a}_0^P$. Let $\Delta_P$ be the set of linear forms on $\mathfrak{a}_P$ obtained by restriction of elements in the complement $\Delta_0 - \Delta_0^P$ of $\Delta_0^P$ in $\Delta_0$. Then $\Delta_P$ is bijective with $\Delta_0 - \Delta_0^P$, and any root in $\Phi_P$ can be written uniquely as a nonnegative integral linear combination of elements in $\Delta_P$. The set $\Delta_P$ is a basis of $(\mathfrak{a}_P^G)^*$. We obtain a second basis of $(\mathfrak{a}_P^G)^*$ by taking the subset

$$\widehat{\Delta}_P = \{\varpi_\alpha : \ \alpha \in \Delta_0 - \Delta_0^P\}$$

of $\widehat{\Delta}_0$. We shall write

$$\Delta_P^\vee = \{\alpha^\vee : \ \alpha \in \Delta_P\}$$

for the basis of $\mathfrak{a}_P^G$ dual to $\widehat{\Delta}_P$, and

$$\widehat{\Delta}_P^\vee = \{\varpi_\alpha^\vee : \ \alpha \in \Delta_P\}$$

for the basis of $\mathfrak{a}_P^G$ dual to $\Delta_P$. We should point out that this notation is not standard if $P \neq P_0$. For in this case, a general element $\alpha \in \Delta_P$ is not part of a root system (as defined in [**Ser2**]), so that $\alpha^\vee$ is not a coroot. Rather, if $\alpha$ is the restriction to $\mathfrak{a}_P$ of the simple root $\beta \in \Delta_0 - \Delta_0^P$, $\alpha^\vee$ is the projection onto $\mathfrak{a}_P$ of the coroot $\beta^\vee$.

We have constructed two bases $\Delta_P$ and $\widehat{\Delta}_P$ of $(\mathfrak{a}_P^G)^*$, and corresponding dual bases $\widehat{\Delta}_P^\vee$ and $\Delta_P^\vee$ of $\mathfrak{a}_P^G$, for any $P$. More generally, suppose that $P_1 \subset P_2$ are two standard parabolic subgroups. Then we can form two bases $\Delta_{P_1}^{P_2}$ and $\widehat{\Delta}_{P_1}^{P_2}$ of $(\mathfrak{a}_{P_1}^{P_2})^*$, and corresponding dual bases $(\widehat{\Delta}_{P_1}^{P_2})^\vee$ and $(\Delta_{P_1}^{P_2})^\vee$ of $\mathfrak{a}_{P_1}^{P_2}$. The construction proceeds in the obvious way from the bases we have already defined. For example, $\Delta_{P_1}^{P_2}$ is the set of linear forms on the subspace $\mathfrak{a}_{P_1}^{P_2}$ of $\mathfrak{a}_{P_1}$ obtained by restricting elements in $\Delta_0^{P_2} - \Delta_0^{P_1}$, while $\widehat{\Delta}_{P_1}^{P_2}$ is the set of linear forms on $\mathfrak{a}_{P_1}^{P_2}$ obtained by restricting elements in $\widehat{\Delta}_{P_1} - \widehat{\Delta}_{P_2}$. We note that $P_1 \cap M_{P_2}$ is a standard parabolic subgroup of the reductive group $M_{P_2}$, relative to the fixed minimal parabolic subgroup $P_0 \cap M_{P_2}$. It follows from the definitions that

$$\mathfrak{a}_{P_1 \cap M_{P_2}} = \mathfrak{a}_{P_1}, \quad \mathfrak{a}_{P_1 \cap M_{P_2}}^{M_{P_2}} = \mathfrak{a}_{P_1}^{P_2}, \quad \Delta_{P_1 \cap M_{P_2}} = \Delta_{P_1}^{P_2},$$

and

$$\widehat{\Delta}_{P_1 \cap M_{P_2}} = \widehat{\Delta}_{P_1}^{P_2}.$$

Consider again the example of $G = GL(n)$. Its Lie algebra is the space $M_n$ of $(n \times n)$-matrices, with the Lie bracket

$$[X, Y] = XY - YX,$$

and the adjoint action

$$\mathrm{Ad}(g) : \ X \longrightarrow gXg^{-1}, \qquad g \in G, \ X \in M_n,$$

of $G$. The group

$$A_0 = \left\{ a = \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix} : \ a_i \in GL(1) \right\}$$

acts by conjugation on the Lie algebra

$$\mathfrak{n}_0 = \mathfrak{n}_{P_0} = \left\{ \begin{pmatrix} 0 & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & \ddots & * \\ 0 & & & 0 \end{pmatrix} \right\}$$

of $N_{P_0}$, and

$$\Phi_0 = \{\beta_{ij} :\ a \longrightarrow a_i a_j^{-1},\ i < j\}.$$

As linear functionals on the vector space

$$\mathfrak{a}_0 = \left\{ u :\ \begin{pmatrix} u_1 & & 0 \\ & \ddots & \\ 0 & & u_n \end{pmatrix} :\ u_i \in \mathbb{R} \right\},$$

the roots $\Phi_0$ take the form

$$\beta_{ij}(u) = u_i - u_j, \qquad i < j.$$

The decomposition of a general root in terms of the subset

$$\Delta_0 = \{\beta_i = \beta_{i,i+1} :\ 1 \leq i \leq n-1\},$$

of simple roots is given by

$$\beta_{ij} = \beta_i + \cdots + \beta_{j-1}, \qquad i < j.$$

The set of coroots equals

$$\Phi_0^\vee = \big\{\beta_{ij}^\vee = e_i - e_j = (\overbrace{\underbrace{0, \ldots, 0, 1}_{i}, 0, \ldots, 0, -1}^{j}, 0, \ldots, 0) :\ i < j\big\},$$

where we have identified $\mathfrak{a}_0$ with the vector space $\mathbb{R}^n$, equipped with the standard basis $\mathrm{e}_1, \ldots, \mathrm{e}_n$. The simple coroots form the basis

$$\Delta_0^\vee = \{\beta_i^\vee = \mathrm{e}_i - \mathrm{e}_{i+1} :\ 1 \leq i \leq n-1\}$$

of the subspace

$$\mathfrak{a}_0^G = \{u \in \mathbb{R}^n :\ \sum u_i = 0\}.$$

The simple weights give the dual basis

$$\widehat{\Delta}_0 = \{\varpi_i :\ 1 \leq i \leq n-1\},$$

where

$$\varpi_i(u) = \frac{n-i}{n}(u_1 + \cdots + u_i) - \Big(\frac{i}{n}\Big)(u_{i+1} + \cdots + u_n).$$

The Weyl group $W_0$ of the root system for $GL(n)$ is the symmetric group $S_n$, acting by permutation of the coordinates of vectors in the space $\mathfrak{a}_0 \cong \mathbb{R}^n$. The dot product on $\mathbb{R}^n$ give a $W$-invariant inner product $\langle \cdot, \cdot \rangle$ on both $\mathfrak{a}_0$ and $\mathfrak{a}_0^*$. It is obvious that

$$\langle \beta_i, \beta_j \rangle \leq 0, \qquad i \neq j.$$

We leave to the reader the exercise of showing that

$$\langle \varpi_i, \varpi_j \rangle \geq 0, \qquad 1 \leq i,\ j \leq n-1.$$

Suppose that $P \subset GL(n)$ corresponds to the partition $(n_1, \dots, n_p)$ of $n$. The general embedding $\mathfrak{a}_P \hookrightarrow \mathfrak{a}_0$ we have defined corresponds to the embedding

$$t \longrightarrow \big(\underbrace{t_1, \dots, t_1}_{n_1}, \underbrace{t_2, \dots, t_2}_{n_2}, \dots, \underbrace{t_p, \dots, t_p}_{n_p}\big), \qquad t \in \mathbb{R}^p,$$

of $\mathbb{R}^p$ into $\mathbb{R}^n$. It follows that

$$\Delta_0^P = \{\beta_i : \ i \neq n_1 + \cdots + n_k, \ 1 \leq k \leq p-1\}.$$

Since $\Delta_P$ is the set of restrictions to $\mathfrak{a}_P \subset \mathfrak{a}_0$ of elements in the set

$$\Delta_0 - \Delta_0^P = \{\beta_{n_1}, \beta_{n_1+n_2}, \dots\},$$

we see that

$$\Delta_P = \{\alpha_i : \ t \to t_i - t_{i+1}, \ 1 \leq i \leq p-1, \ t \in \mathbb{R}^p\}.$$

The example of $G = GL(n)$ provides algebraic intuition. It is useful for readers less familiar with general algebraic groups. However, the truncation of the kernel also requires geometric intuition. For this, the example of $G = SL(3)$ is often sufficient.

The root system for $SL(3)$ is the same as for $GL(3)$. In other words, we can identify $\mathfrak{a}_0$ with the two dimensional subspace

$$\{u \in \mathbb{R}^3 : \ \sum u^i = 0\}$$

of $\mathbb{R}^3$, in which case

$$\Delta_0 = \{\beta_1, \beta_2\} \subset \Phi_0 = \{\beta_1, \beta_2, \beta_1 + \beta_2\},$$

in the notation above. We can also identify $\mathfrak{a}_0$ isometrically with the two dimension Euclidean plane. The singular (one-dimensional) hyperplanes, the coroots $\Phi_0^\vee$, and the simple coweights $(\widehat{\Delta}^0)^\vee$ are then illustrated in the familiar Figures 5.1 and 5.2.

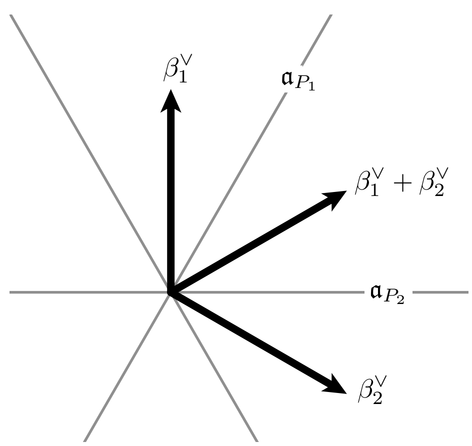



FIGURE 5.1. The two simple coroots $\beta_1^\vee$ and $\beta_2^\vee$ are orthogonal to the respective subspaces $\mathfrak{a}_{P_2}$ and $\mathfrak{a}_{P_1}$ of $\mathfrak{a}_0$. Their inner product is negative, and they span an obtuse angled cone.

There are four standard parabolic subgroups $P_0$, $P_1$, $P_2$, and $G$, with $P_1$ and $P_2$ being the maximal parabolic subgroups such that $\Delta_0^{P_1} = \{\beta_2\}$ and $\Delta_0^{P_2} = \{\beta_1\}$.

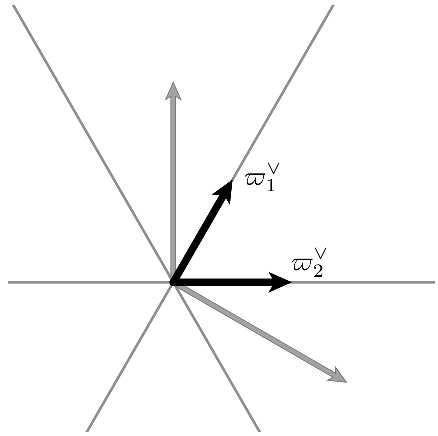



FIGURE 5.2. The two simple coweights $\varpi_1^\vee$ and $\varpi_2^\vee$ lie in the respective subspaces $\mathfrak{a}_{P_1}$ and $\mathfrak{a}_{P_2}$. Their inner product is positive, and they span an acute angled cone.

## 6. Statement and discussion of a theorem

Returning to the general case, we can now describe how to modify the function $K(x,x)$ on $G(\mathbb{Q})\backslash G(\mathbb{A})$. For a given standard parabolic subgroup $P$, we write $\tau_P$ for the characteristic function of the subset

$$\mathfrak{a}_P^+ = \{t \in \mathfrak{a}_P : \ \alpha(t) > 0, \ \alpha \in \Delta_P\}$$

of $\mathfrak{a}_P$. In the case $G = SL(3)$, this subset is the open cone generated by $\varpi_1^\vee$ and $\varpi_2^\vee$ in Figure 5.2 above. We also write $\widehat{\tau}_P$ for the characteristic function of the subset

$$\{t \in \mathfrak{a}_P : \ \varpi(t) > 0, \ \varpi \in \widehat{\Delta}_P\}$$

of $\mathfrak{a}_P$. In case $G = SL(3)$, this subset is the open cone generated by $\beta_1^\vee$ and $\beta_2^\vee$ in Figure 5.1.

The truncation of $K(x,x)$ depends on a parameter $T$ in the cone $\mathfrak{a}_0^+ = \mathfrak{a}_{P_0}^+$ that is suitably regular, in the sense that $\beta(T)$ is large for each root $\beta \in \Delta_0$. For any given $T$, we define

(6.1)
$$k^T(x) = k^T(x,f) = \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} K_P(\delta x, \delta x)\widehat{\tau}_P\big(H_P(\delta x) - T\big).$$

This is the modified kernel, on which the general trace formula is based. A few remarks might help to put it into perspective.

One has to show that for any $x$, the sum over $\delta$ in (6.1) may be taken over a finite set. In the case $G = SL(2)$, the reader can verify the property as an exercise in reduction theory for modular forms. In general, it is a straightforward consequence [**A3**, Lemma 5.1] of the Bruhat decomposition for $G$ and the construction by Borel and Harish-Chandra of an approximate fundamental domain for $G(\mathbb{Q})\backslash G(\mathbb{A})$. (We shall recall both of these results later.) Thus, $k^T(x)$ is given by a double sum over $(P,\delta)$ in a finite set. It is a well defined function of $x \in G(\mathbb{Q})\backslash G(\mathbb{A})$.

Observe that the term in (6.1) corresponding to $P = G$ is just $K(x,x)$. In case $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is compact, there are no proper parabolic subgroups $P$ (over

$\mathbb{Q}$). Therefore $k^T(x)$ equals $K(x,x)$ in this case, and the truncation operation is trivial. In general, the terms with $P \neq G$ represent functions on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ that are supported on some neighbourhood of infinity. Otherwise said, $k^T(x)$ equals $K(x,x)$ for $x$ in some large compact subset of $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ that depends on $T$.

Recall that $G(\mathbb{A})$ is a direct product of $G(\mathbb{A})^1$ with $A_G(\mathbb{R})^0$. Observe also that $k^T(x)$ is invariant under translation of $x$ by $A_G(\mathbb{R})^0$. It therefore suffices to study $k^T(x)$ as a function of $x$ in $G(\mathbb{Q})\backslash G(\mathbb{A})^1$.

Theorem 6.1. *The integral*

$$J^T(f) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k^T(x,f)\mathrm{d}x \tag{6.2}$$

*converges absolutely.*

Theorem 6.1 does not in itself provide a trace formula. It is really just a first step. We are giving it a central place in our discussion for two reasons. The statement of the theorem serves as a reference point for outlining the general strategy. In addition, the techniques required to prove it will be an essential part of many other arguments.

Let us pause for a moment to outline the general steps that will take us to the end of Part I. We shall describe informally what needs to be done in order to convert Theorem 6.1 into some semblance of a trace formula.

**Step 1.** *Find spectral expansions for the functions $K(x,y)$ and $k^T(x)$ that are parallel to the geometric expansions (1.1) and (6.1).*

This step is based on Langlands's theory of Eisenstein series. We shall describe it in the next section.

**Step 2.** *Prove Theorem 6.1.*

We shall sketch the argument in §8.

**Step 3.** *Show that the function*

$$T \longrightarrow J^T(f),$$

*defined a priori for points $T \in \mathfrak{a}_0^+$ that are highly regular, extends to a polynomial in $T \in \mathfrak{a}_0$.*

This step allows us to define $J^T(f)$ for any $T \in \mathfrak{a}_0$. It turns out that there is a canonical point $T_0 \in \mathfrak{a}_0$, depending on the choice of $K$, such that the distribution $J(f) = J^{T_0}(f)$ is independent of the choice of $P_0$ (though still dependent of the choice of $K$). For example, if $G = GL(n)$ and $K$ is the standard maximal compact subgroup of $GL(n,\mathbb{A})$, $T_0 = 0$. We shall discuss these matters in §9, making full use of Theorem 6.1.

**Step 4.** *Convert the expansion (6.1) of $k^T(x)$ in terms of rational conjugacy classes into a geometric expansion of $J(f) = J^{T_0}(f)$.*

We shall give a provisional solution to this problem in §10, as a direct corollary of the proof of Theorem 6.1.

**Step 5.** *Convert the expansion of $k^T(x)$ in §7 in terms of automorphic representations into a spectral expansion of $J(f) = J^{T_0}(f)$.*

This problem turns out to be somewhat harder than the last one. We shall give a provisional solution in §14, as an application of a truncation operator on functions on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$.

We shall call the provisional solutions we obtain for the problems of Steps 4 and 5 the coarse geometric expansion and the coarse spectral expansion, following [**CLL**]. The identity of these two expansions can be regarded as a first attempt at a general trace formula. However, because the terms in the two expansions are still of an essentially global nature, the identity is of little use as it stands. The general problem of refining the two expansions into more tractible local terms will be left until Part II. In order to give some idea of what to expect, we shall deal with the easiest terms near the end of Part I.

In §11, we will rewrite the geometric terms attached to certain semisimple conjugacy classes in $G(\mathbb{Q})$. The distributions so obtained are interesting new linear forms in $f$, known as *weighted orbital integrals.* In §15, we will rewrite the spectral terms attached to certain induced cuspidal automorphic representations of $G(\mathbb{A})$. The resulting distributions are again new linear forms in $f$, known as *weighted characters.* This will set the stage for Part II, where one of the main tasks will be to write the entire geometric expansion in terms of weighted orbital integrals, and the entire spectral expansion in terms of weighted characters.

There is a common thread to Part I. It is the proof of Theorem 6.1. For example, the proofs of Corollary 10.1, Theorem 11.1, Proposition 12.2 and parts (ii) and (iii) of Theorem 14.1 either follow directly from, or are strongly motivated by, the proof of Theorem 6.1. Moreover, the actual assertion of Theorem 6.1 is the essential ingredient in the proofs of Theorems 9.1 and 9.4, as well as their geometric analogues in §10 and their spectral analogues in §14. We have tried to emphasize this pattern in order to give the reader some overview of the techniques.

The proof of Theorem 6.1 itself has both geometric and analytic components. However, its essence is largely combinatorial. This is due to the cancellation in (6.1) implicit in the alternating sum over $P$. At the heart of the proof is the simplest of all cancellation laws, the identity obtained from the binomial expansion of $\big(1+(-1)\big)^n$.

IDENTITY 6.2. *Suppose that $S$ is a finite set. Then*

$$\sum_{F\subset S}(-1)^{|S|-|F|}=\begin{cases}1 & \text{if } S=\emptyset,\\ 0 & \text{otherwise.}\end{cases} \tag{6.3}$$

□

## 7. Eisenstein series

Eisenstein series are responsible for the greatest discrepancy between what we need and what we can prove here. Either of the two main references [**Lan5**] or [**MW2**] presents an enormous challenge to anyone starting to learn the subject. Langlands's survey article [**Lan1**] is a possible entry point. For the trace formula, one can usually make do with a statement of the main theorems on Eisenstein series. We give a summary, following [**A2**, §2].

The role of Eisenstein series is to provide a spectral expansion for the kernel $K(x,y)$. In general, the regular representation $R$ of $G(\mathbb{A})$ on $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ does not decompose discretely. Eisenstein series describe the continuous part of the spectrum.

We write $R_{G,\mathrm{disc}}$ for the restriction of the regular representation of $G(\mathbb{A})^1$ to the subspace $L^2_{\mathrm{disc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ of $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ that decomposes discretely.

Since $G(\mathbb{A})$ is a direct product of $G(\mathbb{A})^1$ with $A_G(\mathbb{R})^0$, we can identify $R_{G,\mathrm{disc}}$ with the representation of $G(\mathbb{A})$ on the subspace $L^2_{\mathrm{disc}}\big(G(\mathbb{Q})A_G(\mathbb{R})^0\backslash G(\mathbb{A})\big)$ of $L^2\big(G(\mathbb{Q})A_G(\mathbb{R})^0\backslash G(\mathbb{A})\big)$ that decomposes discretely. For any point $\lambda \in \mathfrak{a}^*_{G,\mathbb{C}}$, the tensor product

$$R_{G,\mathrm{disc},\lambda}(x) = R_{G,\mathrm{disc}}(x)\mathrm{e}^{\lambda(H_G(x))}, \qquad x \in G(\mathbb{A}),$$

is then a representation of $G(\mathbb{A})$, which is unitary if $\lambda$ lies in $i\mathfrak{a}^*_G$.

We have assumed from the beginning that the invariant measures in use satisfy any obvious compatibility conditions. For example, if $P$ is a standard parabolic subgroup, it is easy to check that the Haar measures on the relevant subgroups of $G(\mathbb{A})$ can be chosen so that

$$\begin{aligned}
&\int_{G(\mathbb{A})} f(x)\mathrm{d}x \\
&= \int_K \int_{P(\mathbb{A})} f(pk)\mathrm{d}_\ell p\mathrm{d}k \\
&= \int_K \int_{M_P(\mathbb{A})} \int_{N_P(\mathbb{A})} f(mnk)\mathrm{d}n\mathrm{d}m\mathrm{d}k \\
&= \int_K \int_{M_P(\mathbb{A})^1} \int_{A_P(\mathbb{R})^0} \int_{N_P(\mathbb{A})} f(mank)\mathrm{d}n\mathrm{d}a\mathrm{d}m\mathrm{d}k,
\end{aligned}$$

for any $f \in C^\infty_c\big(G(\mathbb{A})\big)$. We are assuming implicitly that the Haar measures on $K$ and $N_P(\mathbb{A})$ are normalized so that the spaces $K$ and $N_P(\mathbb{Q})\backslash N_P(\mathbb{A})$ each have volume 1. The Haar measure $\mathrm{d}x$ on $G(\mathbb{A})$ is then determined by Haar measures $\mathrm{d}m$ and $\mathrm{d}a$ on the groups $M_P(\mathbb{A})^1$ and $A_P(\mathbb{R})^0$. We write $\mathrm{d}H$ for the Haar measure on $\mathfrak{a}_P$ that corresponds to $\mathrm{d}a$ under the exponential map. We then write $\mathrm{d}\lambda$ for the Haar measure on $i\mathfrak{a}^*_P$ that is dual to $\mathrm{d}H$, in the sense that

$$\int_{i\mathfrak{a}^*_P} \int_{\mathfrak{a}_P} h(H)\mathrm{e}^{-\lambda(H)}\mathrm{d}H\mathrm{d}\lambda = h(0),$$

for any function $h \in C^\infty_c(\mathfrak{a}_P)$.

Suppose that $P$ is a standard parabolic subgroup of $G$, and that $\lambda$ lies in $\mathfrak{a}^*_{P,\mathbb{C}}$. We write

$$y \longrightarrow \mathcal{I}_P(\lambda, y), \qquad y \in G(\mathbb{A}),$$

for the induced representation

$$\mathrm{Ind}^{G(\mathbb{A})}_{P(\mathbb{A})}(I_{N_P(\mathbb{A})} \otimes R_{M_P,\mathrm{disc},\lambda})$$

of $G(\mathbb{A})$ obtained from $\lambda$ and the discrete spectrum of the reductive group $M_P$. This representation acts on the Hilbert space $\mathcal{H}_P$ of measurable functions

$$\phi:\ N_P(\mathbb{A})M_P(\mathbb{Q})A_P(\mathbb{R})^0\backslash G(\mathbb{A}) \longrightarrow \mathbb{C}$$

such that the function

$$\phi_x:\ m \longrightarrow \phi(mx), \qquad m \in M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1,$$

belongs to $L^2_{\mathrm{disc}}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$ for any $x \in G(\mathbb{A})$, and such that

$$\|\phi\|^2 = \int_K \int_{M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1} |\phi(mk)|^2\mathrm{d}m\mathrm{d}k < \infty.$$

For any $y \in G(\mathbb{A})$, $\mathcal{I}_P(\lambda, y)$ maps a function $\phi \in \mathcal{H}_P$ to the function

$$\big(\mathcal{I}_P(\lambda, y)\phi\big)(x) = \phi(xy)\mathrm{e}^{(\lambda+\rho_P)(H_P(xy))}\mathrm{e}^{-(\lambda+\rho_P)(H_P(x))}.$$

We have put the twist by $\lambda$ into the operator $\mathcal{I}_P(\lambda, y)$ rather than the underlying Hilbert space $\mathcal{H}_P$, in order that $\mathcal{H}_P$ be independent of $\lambda$. Recall that the function $\mathrm{e}^{\rho_P(H_P(\cdot))}$ is the square root of the modular function of the group $P(\mathbb{A})$. It is included in the definition in order that the representation $\mathcal{I}_P(\lambda)$ be unitary whenever the inducing representation is unitary, which is to say, whenever $\lambda$ belongs to the subset $i\mathfrak{a}_P^*$ of $\mathfrak{a}_{P,\mathbb{C}}^*$.

Suppose that

$$R_{M_P,\mathrm{disc}} \cong \bigoplus_{\pi} \pi \cong \bigoplus_{\pi} \Big( \bigotimes_{v} \pi_v \Big)$$

is the decomposition of $R_{M_P,\mathrm{disc}}$ into irreducible representations $\pi = \bigotimes\limits_v \pi_v$ of $M_P(\mathbb{A})/A_P(\mathbb{R})^0$. The induced representation $\mathcal{I}_P(\lambda)$ then has a corresponding decomposition

$$\mathcal{I}_P(\lambda) \cong \bigoplus_{\pi} \mathcal{I}_P(\pi_\lambda) \cong \bigoplus_{\pi} \Big( \bigotimes_{v} \mathcal{I}_P(\pi_{v,\lambda}) \Big)$$

in terms of induced representations $\mathcal{I}_P(\pi_{v,\lambda})$ of the local groups $G(\mathbb{Q}_v)$. This follows from the definition of induced representation, and the fact that

$$\mathrm{e}^{\lambda(H_{M_P}(m))} = \prod_{v} \mathrm{e}^{\lambda(H_{M_P}(m_v))},$$

for any point $m = \prod\limits_v m_v$ in $M_P(\mathbb{A})$. If $\lambda \in i\mathfrak{a}_P^*$ is in general position, all of the induced representations $\mathcal{I}_P(\pi_{v,\lambda})$ are irreducible. Thus, if we understand the decomposition of the discrete spectrum of $M_P$ into irreducible representations of the local groups $M_P(\mathbb{Q}_v)$, we understand the decomposition of the generic induced representations $\mathcal{I}_P(\lambda)$ into irreducible representations of the local groups $G(\mathbb{Q}_v)$.

The aim of the theory of Eisenstein series is to construct intertwining operators between the induced representations $\mathcal{I}_P(\lambda)$ and the continuous part of the regular representation $R$ of $G(\mathbb{A})$. The problem includes being able to construct intertwining operators among the representations $\mathcal{I}_P(\lambda)$, as $P$ and $\lambda$ vary. The symmetries among pairs $(P, \lambda)$ are given by the Weyl sets $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ of Langlands. For a given pair $P$ and $P'$ of standard parabolic subgroups, $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ is defined as the set of distinct linear isomorphisms from $\mathfrak{a}_P \subset \mathfrak{a}_0$ onto $\mathfrak{a}_{P'} \subset \mathfrak{a}_0$ obtained by restriction of elements in the Weyl group $W_0$. Suppose, for example that $G = GL(n)$. If $P$ and $P'$ correspond to the partitions $(n_1, \dots, n_p)$ and $(n_1', \dots, n_{p'}')$ of $n$, the set $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ is empty unless $p = p'$, in which case

$$W(\mathfrak{a}_P, \mathfrak{a}_{P'}) \cong \{s \in S_p : \ n_i' = n_{s(i)}, \ 1 \le i \le p\}.$$

In general, we say that $P$ and $P'$ are *associated* if the set $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ is nonempty. We would expect a pair of induced representations $\mathcal{I}_P(\lambda)$ and $\mathcal{I}_{P'}(\lambda')$ to be equivalent if $P$ and $P'$ belong to the same associated class, and $\lambda' = s\lambda$ for some element $s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'})$.

The formal definitions apply to any elements $x \in G(\mathbb{A})$, $\phi \in \mathcal{H}_P$, and $\lambda \in \mathfrak{a}_{M,\mathbb{C}}^*$. The associated Eisenstein series is

$$E(x, \phi, \lambda) = \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} \phi(\delta x)\mathrm{e}^{(\lambda+\rho_P)(H_P(\delta x))}. \tag{7.1}$$

If $s$ belongs to $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$, the operator

$$M(s,\lambda): \mathcal{H}_P \longrightarrow \mathcal{H}_{P'}$$

that intertwines $\mathcal{I}_P(\lambda)$ with $\mathcal{I}_{P'}(s\lambda)$ is defined by

$$\big(M(s,\lambda)\phi\big)(x) = \int \phi(w_s^{-1}nx)\mathrm{e}^{(\lambda+\rho_P)(H_P(w_s^{-1}nx)}\mathrm{e}^{(-s\lambda+\rho_{P'})(H_{P'}(x))}\mathrm{d}n, \tag{7.2}$$

where the integral is taken over the quotient

$$N_{P'}(\mathbb{A}) \cap w_s N_P(\mathbb{A}) w_s^{-1} \backslash N_{P'}(\mathbb{A}),$$

and $w_s$ is any representative of $s$ in $G(\mathbb{Q})$. A reader so inclined could motivate both definitions in terms of finite group theory. Each definition is a formal analogue of a general construction by Mackey [**Ma**] for the space of intertwining operators between two induced representations $\mathrm{Ind}_{H_1}^H(\rho_1)$ and $\mathrm{Ind}_{H_2}^H(\rho_2)$ of a finite group $H$.

It follows formally from the definitions that

$$E\big(x, \mathcal{I}_P(\lambda, y)\phi, \lambda\big) = E(xy, \phi, \lambda)$$

and

$$M(s,\lambda)\mathcal{I}_P(\lambda, y) = \mathcal{I}_{P'}(s\lambda, y)M(s,\lambda).$$

These are the desired intertwining properties. However, (7.1) and (7.2) are defined by sums and integrals over noncompact spaces. They do not generally converge. It is this fact that makes the theory of Eisenstein series so difficult.

Let $\mathcal{H}_P^0$ be the subspace of vectors $\phi \in \mathcal{H}_P$ that are $K$-finite, in the sense that the subset

$$\{\mathcal{I}_P(\lambda, k)\phi : \ k \in K\}$$

of $\mathcal{H}_P$ spans a finite dimensional space, and that lie in a finite sum of irreducible subspaces of $\mathcal{H}_P$ under the action $\mathcal{I}_P(\lambda)$ of $G(\mathbb{A})$. The two conditions do not depend on the choice of $\lambda$. Taken together, they are equivalent to the requirement that the function

$$\phi(x_\infty x_{\mathrm{fin}}), \qquad\qquad x_\infty \in G(\mathbb{R}), \ x_{\mathrm{fin}} \in G(\mathbb{A}_{\mathrm{fin}}),$$

be locally constant in $x_{\mathrm{fin}}$, and smooth, $K_{\mathbb{R}}$-finite and $\mathcal{Z}_\infty$-finite in $x_\infty$, where $\mathcal{Z}_\infty$ denotes the algebra of bi-invariant differential operators on $G(\mathbb{R})$. The space $\mathcal{H}_P^0$ is dense in $\mathcal{H}_P$.

For any $P$, we can form the chamber

$$(\mathfrak{a}_P^*)^+ = \big\{\Lambda \in \mathfrak{a}_P^* : \ \Lambda(\alpha^\vee) > 0, \ \alpha \in \Delta_P\big\}$$

in $\mathfrak{a}_P^*$.

Lemma 7.1 (Langlands). *Suppose that $\phi \in \mathcal{H}_P^0$ and that $\lambda$ lies in the open subset*

$$\big\{\lambda \in \mathfrak{a}_{P,\mathbb{C}}^* : \ \mathrm{Re}(\lambda) \in \rho_P + (\mathfrak{a}_P^*)^+\big\}$$

*of $\mathfrak{a}_{P,\mathbb{C}}^*$. Then the sum (7.1) and integral (7.2) that define $E(x,\phi,\lambda)$ and $\big(M(s,\lambda)\phi\big)(x)$ both converge absolutely to analytic functions of $\lambda$.* □

For spectral theory, one is interested in points $\lambda$ such that $\mathcal{I}_P(\lambda)$ is unitary, which is to say that $\lambda$ belongs to the real subspace $i\mathfrak{a}_P^*$ of $\mathfrak{a}_{P,\mathbb{C}}^*$. This is outside the domain of absolute convergence for (7.1) and (7.2). The problem is to show that the functions $E(x,\phi,\lambda)$ and $M(s,\lambda)\phi$ have analytic continuation to this space. The following theorem summarizes Langlands' main results on Eisenstein series.

THEOREM 7.2 (Langlands). (a) *Suppose that* $\phi \in \mathcal{H}_P^0$. *Then* $E(x, \phi, \lambda)$ *and* $M(s, \lambda)\phi$ *can be analytically continued to meromorphic functions of* $\lambda \in \mathfrak{a}_{P,\mathbb{C}}^*$ *that satisfy the functional equations*

$$E\big(x, M(s,\lambda)\phi, s\lambda\big) = E(x, \phi, \lambda) \tag{7.3}$$

*and*

$$M(ts, \lambda) = M(t, s\lambda)M(s, \lambda), \qquad t \in W(\mathfrak{a}_{P'}, \mathfrak{a}_{P''}). \tag{7.4}$$

*If* $\lambda \in i\mathfrak{a}_P^*$, *both* $E(x, \phi, \lambda)$ *and* $M(s, \lambda)$ *are analytic, and* $M(s, \lambda)$ *extends to a unitary operator from* $\mathcal{H}_P$ *to* $\mathcal{H}_{P'}$.

(b) *Given an associated class* $\mathcal{P} = \{P\}$, *define* $\widehat{L}_{\mathcal{P}}$ *to be the Hilbert space of families of measurable functions*

$$F = \{F_P : \; i\mathfrak{a}_P^* \longrightarrow \mathcal{H}_P, \; P \in \mathcal{P}\}$$

*that satisfy the symmetry condition*

$$F_{P'}(s\lambda) = M(s, \lambda)F_P(\lambda), \qquad s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'}),$$

*and the finiteness condition*

$$\|F\|^2 = \sum_{P \in \mathcal{P}} n_P^{-1} \int_{i\mathfrak{a}_P^*} \|F_P(\lambda)\|^2 d\lambda < \infty,$$

*where*

$$n_P = \sum_{P' \in \mathcal{P}} |W(\mathfrak{a}_P, \mathfrak{a}_{P'})|$$

*for any* $P \in \mathcal{P}$. *Then the mapping that sends* $F$ *to the function*

$$\sum_{P \in \mathcal{P}} n_P^{-1} \int_{i\mathfrak{a}_P^*} E\big(x, F_P(\lambda), \lambda\big) \mathrm{d}\lambda, \qquad x \in G(\mathbb{A}),$$

*defined whenever* $F_P(\lambda)$ *is a smooth, compactly supported function of* $\lambda$ *with values in a finite dimensional subspace of* $\mathcal{H}_P^0$, *extends to a unitary mapping from* $\widehat{L}_{\mathcal{P}}$ *onto a closed* $G(\mathbb{A})$*-invariant subspace* $L^2_{\mathcal{P}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ *of* $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$. *Moreover, the original space* $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ *has an orthogonal direct sum decomposition*

$$L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big) = \bigoplus_{\mathcal{P}} L^2_{\mathcal{P}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big). \tag{7.5}$$

□

Theorem 7.2(b) gives a qualitative description of the decomposition of $R$. It provides a finite decomposition

$$R = \bigoplus_{\mathcal{P}} R_{\mathcal{P}},$$

where $R_{\mathcal{P}}$ is the restriction of $R$ to the invariant subspace $L^2_{\mathcal{P}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ of $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$. It also provides a unitary intertwining operator from $R_{\mathcal{P}}$ onto the representation $\widehat{R}_{\mathcal{P}}$ of $G(\mathbb{A})$ on $\widehat{L}_{\mathcal{P}}$ defined by

$$\big(\widehat{R}_{\mathcal{P}}(y)F\big)_P(\lambda) = \mathcal{I}_P(\lambda, y)F_P(\lambda), \qquad F \in \widehat{L}^2_{\mathcal{P}}, \; P \in \mathcal{P}.$$

The theorem is thus compatible with the general intuition we retain from the theory of Fourier series and Fourier transforms.

Let $\mathcal{B}_P$ be an orthonormal basis of the Hilbert space $\mathcal{H}_P$. We assume that every $\phi \in \mathcal{B}_P$ lies in the dense subspace $\mathcal{H}_P^0$. It is a direct consequence of Theorem 7.2 that the kernel

$$K(x,y) = \sum_{\gamma \in G(\mathbb{Q})} f(x^{-1}\gamma y), \qquad f \in C_c^\infty\big(G(\mathbb{A})\big),$$

of $R(f)$ also has a formal expansion

$$\sum_P n_P^{-1} \int_{i\mathfrak{a}_P^*} \sum_{\phi \in \mathcal{B}_P} E\big(x, \mathcal{I}_P(\lambda, f)\phi, \lambda\big)\overline{E(y,\phi,\lambda)} \mathrm{d}\lambda \tag{7.6}$$

in terms of Eisenstein series. A reader to whom this assertion is not clear might consider the analogous assertion for the case $H = \mathbb{R}$ and $\Gamma = \{1\}$. If $f$ belongs to $C_c^\infty(\mathbb{R})$, the spectral expansion

$$K(x,y) = f(-x+y) = \frac{1}{2\pi i}\int_{i\mathbb{R}} \pi_\lambda(f) \mathrm{e}^{\lambda x}\overline{e^{\lambda y}}\mathrm{d}\lambda, \qquad f \in C_c^\infty(\mathbb{R}),$$

of the kernel of $R(f)$, in which

$$\pi_\lambda(f) = \int_{\mathbb{R}} f(u)\mathrm{e}^{\lambda u}\mathrm{d}u,$$

is just the inverse Fourier transform of $f$.

In the case of Eisenstein series, one has to show that the spectral expansion of $K(x,y)$ converges in order to make the formal argument rigorous. In general, it is not feasible to estimate $E(x,\phi,\lambda)$ as a function of $\lambda \in i\mathfrak{a}_P^*$. What saves the day is the following simple idea of Selberg, which exploits only the underlying functional analysis.

One first shows that $f$ may be written as a finite linear combination of convolutions $h_1 * h_2$ of functions $h_i \in C_c^r\big(G(\mathbb{A})\big)$, whose archimedean components are differentiable of arbitrarily high order $r$. An application of the Holder inequality to the formal expansion (7.6) establishes that it is enough to prove the convergence in the special case that $f = h_i * h_i^*$, where $h_i^*(x) = \overline{h_i(x^{-1})}$, and $x = y$. The integrand in (7.6) is then easily seen to be nonnegative. In fact, the double integral over $\lambda$ and $\phi$ can be expressed as an increasing limit of nonnegative functions, each of which is the kernel of the restriction of $R(f)$ to an invariant subspace. Since this limit is bounded by the nonnegative function

$$K_i(x,x) = \sum_{\gamma \in G(\mathbb{Q})} (h_i * h_i^*)(x^{-1}\gamma x),$$

the integral converges. (See [**A3**, p. 928–934].)

There is also a spectral expansion for the kernel

$$K_Q(x,y) = \int_{N_Q(\mathbb{A})} \sum_{\gamma \in M_Q(\mathbb{Q})} f(x^{-1}\gamma n y)\mathrm{d}n$$

of $R_Q(f)$, for any standard parabolic subgroup $Q$. One has only to replace the multiplicity $n_P = n_P^G$ and the Eisenstein series $E(x,\phi,\lambda) = E_P^G(x,\phi,\lambda)$ in (7.6) by their relative analogues $n_P^Q = n_{M_Q \cap P}$ and

$$E_P^Q(x,\phi,\lambda) = \sum_{\delta \in P(\mathbb{Q})\backslash Q(\mathbb{Q})} \phi(\delta x)\mathrm{e}^{(\lambda+\rho_P)(H_P(\delta x))},$$

for each $P \subset Q$. Since $P\backslash Q = M_Q \cap P \backslash M_Q$, the analytic continuation of $E_P^Q(x, \phi, \lambda)$ follows from Theorem 7.2(a), with $(M_Q, M_Q \cap P)$ in place of $(G, P)$. The spectral expansion of $K_Q(x, y)$ is

$$\sum_{P \subset Q} (n_P^Q)^{-1} \int_{i\mathfrak{a}_P^*} \sum_{\phi \in \mathcal{B}_P} E_P^Q\big(x, \mathcal{I}_P(\lambda, f)\phi, \lambda\big)\overline{E_P^Q(y, \phi, \lambda)} \mathrm{d}\lambda.$$

If we substitute this formula into (6.1), we obtain a spectral expansion for the truncated kernel $k^T(x)$. The two expansions of $k^T(x)$ ultimately give rise to two formulas for the integral $J^T(f)$. They are thus the source of the trace formula.

## 8. On the proof of the theorem

Theorem 6.1 represents a significant step in the direction of a trace formula. It is time now to discuss its proof. We shall outline the main argument, proving as much as possible. There are some lemmas whose full justification will be left to the references. However, in these cases we shall try to give the basic geometric idea behind the proof.

Suppose that $T_1$ belongs to the real vector space $\mathfrak{a}_0$, and that $\omega$ is a compact subset of $N_{P_0}(\mathbb{A})M_{P_0}(\mathbb{A})^1$. The subset

$$\begin{aligned}\mathcal{S}^G(T_1) &= \mathcal{S}^G(T_1, \omega)\\ &= \big\{x = pak : \ p \in \omega, \ a \in A_0(\mathbb{R})^0, \ k \in K, \ \beta\big(H_{P_0}(a) - T_1\big) > 0, \ \beta \in \Delta_0\big\}\end{aligned}$$

of $G(\mathbb{A})$ is called the *Siegel set* attached to $T_1$ and $\omega$. The inequality in the definition amounts to the assertion that

$$\tau_{P_0}\big(H_{P_0}(x) - T_1\big) = \tau_{P_0}\big(H_{P_0}(a) - T_1\big) = 1.$$

For example, if $G = SL(3)$, the condition is that the point $H_{P_0}(x)$ in the two dimensional vector space $\mathfrak{a}_0$ lies in the open cone in Figure 8.1.

THEOREM 8.1 (Borel, Harish-Chandra). *One can choose $T_1$ and $\omega$ so that*

$$G(\mathbb{A}) = G(\mathbb{Q})\mathcal{S}^G(T_1, \omega).$$

This is one of the main results in the foundational paper [**BH**] of Borel and Harish-Chandra. It was formulated in the adelic terms stated here in [**Bor1**]. The best reference might be the monograph [**Bor2**]. □

From now on, $T_1$ and $\omega$ are to be fixed as in Theorem 8.1. Suppose that $T \in \mathfrak{a}_0$ is a truncation parameter, in the earlier sense that $\beta(T)$ is large for each $\beta \in \Delta_0$. We then form the truncated Siegel set

$$\mathcal{S}^G(T_1, T) = \mathcal{S}^G(T_1, T, \omega) = \big\{x \in \mathcal{S}^G(T_1, \omega) : \ \varpi\big(H_{P_0}(x) - T\big) \leq 0, \ \varpi \in \widehat{\Delta}_0\big\}.$$

For example, if $G = SL(3)$, $\mathcal{S}^G(T_1, T)$ is the set of elements $x \in \mathcal{S}^G(T_1)$ such that $H_{P_0}(x)$ lies in the relatively compact subset of $\mathfrak{a}_0$ illustrated in Figure 8.2.

We write $F^G(x, T)$ for the characteristic function in $x$ of the projection of $\mathcal{S}^G(T_1, T)$ onto $G(\mathbb{Q})\backslash G(\mathbb{A})$. Since $G(\mathbb{A})^1 \cap \mathcal{S}^G(T_1, T)$ is compact, $F^G(\cdot, T)$ has compact support on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$, and is invariant under translation by $A_G(\mathbb{R})^0$.

More generally, suppose that $P$ is a standard parabolic subgroup. We define the sets $\mathcal{S}^P(T_1) = \mathcal{S}^P(T_1, \omega)$ and $\mathcal{S}^P(T_1, T) = \mathcal{S}^P(T_1, T, \omega)$ and the characteristic function $F^P(x, T)$ exactly as above, but with $\Delta_{P_0}$, $\widehat{\Delta}_{P_0}$ and $G(\mathbb{Q})\backslash G(\mathbb{A})$ replaced by

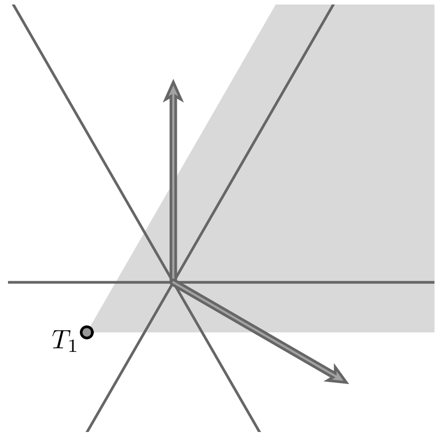



FIGURE 8.1. The shaded region is the projection onto $\mathfrak{a}_0$ of a Siegel set for $G = SL(3)$. It is the translate of the open cone $\mathfrak{a}_{P_0}^+$ by a point $T_1 \in \mathfrak{a}_0$. If $T_1$ is sufficiently regular in the negative cone $(-\mathfrak{a}_{P_0}^+)$, the Siegel set is an approximate fundamental domain.

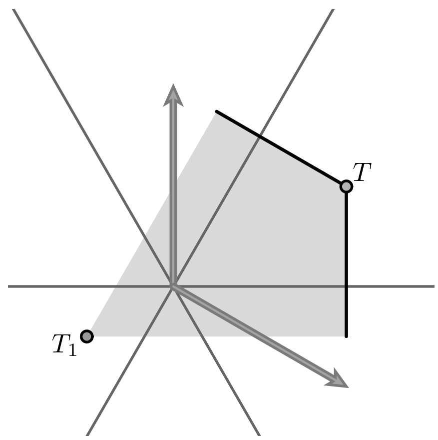



FIGURE 8.2. The shaded region represents a truncation of the Siegel set at a point $T \in \mathfrak{a}_{P_0}^+$. The image of the truncated Siegel set in $SL(3,\mathbb{Q})\backslash SL(3,\mathbb{A})$ is compact.

$\Delta_{P_0}^P$, $\widehat{\Delta}_{P_0}^P$ and $P(\mathbb{Q})\backslash G(\mathbb{A})$ respectively. In particular, $F^P(x,T)$ is the characteristic function of a subset of $P(\mathbb{Q})\backslash G(\mathbb{A})$. More precisely, if

$$x = nmak, \qquad n \in N_P(\mathbb{A}),\ m \in M_P(\mathbb{A})^1,\ a \in A_P(\mathbb{R})^0,\ k \in K,$$

then

$$F^P(x,T) = F^P(m,T) = F^{M_P}(m,T).$$

LEMMA 8.2. *For any $x \in G(\mathbb{A})$, we have*

$$\sum_P \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} F^P(\delta x, T)\tau_P\big(H_P(\delta x)\big) = 1.$$

In case $G = SL(2)$, the lemma follows directly from classical reduction theory, as we shall see in Figure 8.3 below. The general proof is established from properties of finite dimensional $\mathbb{Q}$-rational representations of $G$. (See [**A3**, Lemma 6.4], a result that is implicit in Langlands monograph, for example in [**Lan5**, Lemma 2.12].)

Lemma 8.2 can be restated geometrically in terms of the subsets

$$G_P(T) = \{x \in P(\mathbb{Q})\backslash G(\mathbb{A}) : \ F^P(x,T) = 1, \ \tau_P\big(H_P(x) - T\big) = 1\}$$

of $P(\mathbb{Q})\backslash G(\mathbb{A})$. The lemma asserts that for any $P$, the projection of $P(\mathbb{Q})\backslash G(\mathbb{A})$ onto $G(\mathbb{Q})\backslash G(\mathbb{A})$ maps $G_P(T)$ *injectively* onto a subset $\overline{G}_P(T)$ of $G(\mathbb{Q})\backslash G(\mathbb{A})$, and that $G(\mathbb{Q})\backslash G(\mathbb{A})$ is a disjoint union over $P$ of the sets $\overline{G}_P(T)$. Otherwise said, $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ has a partition parametrized by the set of standard parabolic subgroups, which separates the problem of noncompactness from the topological complexity of $G(\mathbb{Q})\backslash G(\mathbb{A})^1$. The subset corresponding to $P = G$ is compact but topologically complex, while the subset corresponding to $P = P_0$ is topologically simple but highly noncompact. The subset corresponding to a group $P \notin \{P_0, G\}$ is mixed, being a product of a compact set of intermediate complexity with a simple set of intermediate degree of noncompactness. The partition of $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is, incidentally, closely related to the compactification of this space defined by Borel and Serre.

Consider the case that $G = SL(2)$. If $K$ is the standard maximal compact subgroup of $SL(2,\mathbb{A})$, Theorem 2.1(a) tells us that

$$SL(2,\mathbb{Q})\backslash SL(2,\mathbb{A})/K \cong SL(2,\mathbb{Z})\backslash SL(2,\mathbb{R})/SO(2) \cong SL(2,\mathbb{Z})\backslash\mathcal{H},$$

where $\mathcal{H} \cong SL(2,\mathbb{R})/SO(2)$ is the upper half plane. Since they are right $K$-invariant, the two sets $\overline{G}_P(T)$ in this case may be identified with subsets of $SL(2,\mathbb{Z})\backslash\mathcal{H}$, which we illustrate in Figure 8.3. The darker region in the figure represents the standard fundamental domain for $SL(2,\mathbb{Z})$ in $\mathcal{H}$. Its intersection with the lower bounded rectangle equals $\overline{G}_G(T)$, while its intersection with the upper unbounded rectangle equals $\overline{G}_{P_0}(T)$. The larger unbounded rectangle represents a Siegel set, and its associated truncation. These facts, together with Lemma 8.2, follow in this case from a basic fact from classical reduction theory. Namely, if $\gamma \in SL(2,\mathbb{Z})$ and $z \in \mathcal{H}$ are such that the $y$-coordinates of both $z$ and $\gamma z$ are greater than $\mathrm{e}^T$, then $\gamma$ is upper triangular.

For another example, consider the case that $G = SL(3)$. In this case there are four sets, corresponding to the four standard parabolic subgroups $P_0$, $P_1$, $P_2$ and $G$. In Figure 8.4, we illustrate the partition of $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ by describing the corresponding partition of the image in $\mathfrak{a}_0$ of the Siegel set $\mathcal{S}(T_1)$. □

Lemma 8.2 is a critical first step in the proof of Theorem 6.1. We shall actually apply it in a slightly different form. Suppose that $P_1 \subset P$. Then

$$P_1\backslash P = (P_1 \cap M_P)N_P\backslash M_P N_P \cong P_1 \cap M_P\backslash M_P.$$

We write $\tau_{P_1}^P = \tau_{P_1\cap M_P}$ and $\widehat{\tau}_{P_1}^P = \widehat{\tau}_{P_1\cap M_P}$. We shall regard these two functions as characteristic functions on $\mathfrak{a}_0$ that depend only on the projection of $\mathfrak{a}_0$ onto $\mathfrak{a}_{P_1}^{P_2}$, relative to the decomposition

$$\mathfrak{a}_0 = \mathfrak{a}_0^{P_1} \oplus \mathfrak{a}_{P_1}^{P_2} \oplus \mathfrak{a}_{P_2}.$$

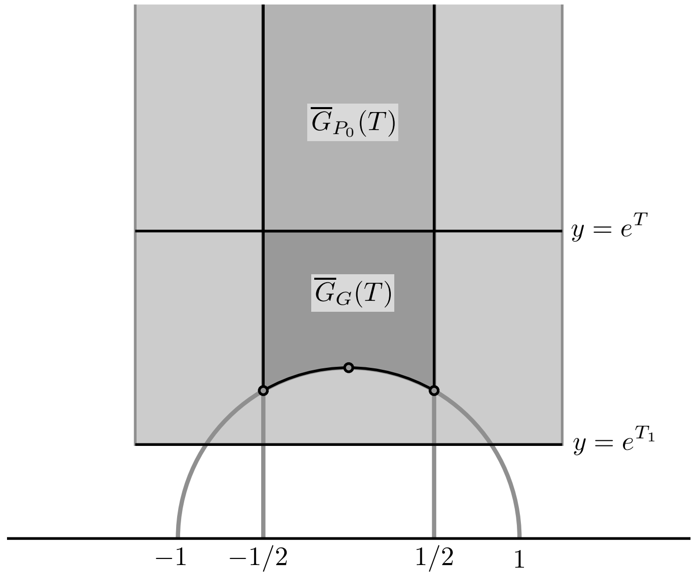



FIGURE 8.3. An illustration for $\mathcal{H} = SL(2,\mathbb{R})/SO(2,\mathbb{R})$ of a standard fundamental domain and its truncation at a large positive number $T$, together with the more tractible Siegel set and its associated truncation.

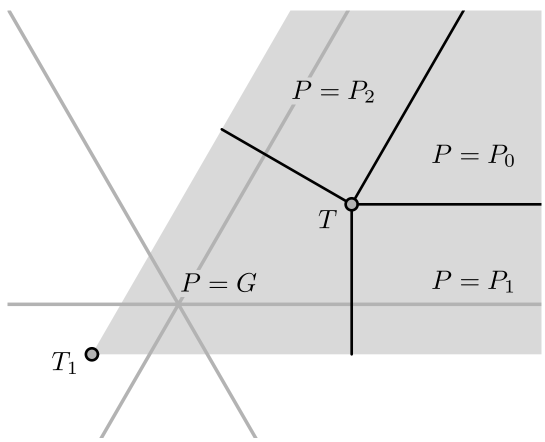



FIGURE 8.4. A partition of the region in Figure 8.1 into four sets, parametrized by the four standard parabolic subgroups $P$ of $SL(3)$. The set corresponding to $P = P_0$ is the truncated region in Figure 8.2.

If $P$ is fixed, we obtain the identity

$$\sum_{\{P_1 : P_1 \subset P\}} \sum_{\delta_1 \in P_1(\mathbb{Q})\backslash P(\mathbb{Q})} F^{P_1}(\delta_1 x, T)\tau_{P_1}^P\big(H_{P_1}(\delta_1 x) - T\big) = 1 \tag{8.1}$$

by applying Lemma 8.2 to $M_P$ instead of $G$, noting at the same time that

$$F^{P_1}(y,T) = F^{M_{P_1}}(m,T)$$

and

$$H_{P_1}(y) = H_{M_{P_1}}(m),$$

for any point

$$y = nmk, \qquad\qquad n \in N_P(\mathbb{A}),\ m \in M_P(\mathbb{A}),\ k \in K.$$

We can now begin the proof of Theorem 6.1. We write

$$\begin{aligned} k^T(x) &= \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} K_P(\delta x, \delta x)\widehat{\tau}_P\big(H_P(\delta x) - T\big) \\ &= \sum_P (-1)^{\dim(A_P/A_G)} \sum_\delta \Big( \sum_{P_1 \subset P} \sum_{\delta_1 \in P_1(\mathbb{Q})\backslash P(\mathbb{Q})} F^{P_1}(\delta_1\delta x, T)\tau_{P_1}^P\big(H_{P_1}(\delta_1\delta x) - T\big)\Big) \\ &\qquad\qquad \cdot \widehat{\tau}_P\big(H_P(\delta x) - T\big)K_P(\delta x, \delta x), \end{aligned}$$

by substituting (8.1) into the definition of $k^T(x)$. We then write

$$K_P(\delta x, \delta x) = K_P(\delta_1\delta x, \delta_1\delta x)$$

and

$$\widehat{\tau}_P\big(H_P(\delta x) - T\big) = \widehat{\tau}_P\big(H_P(\delta_1\delta x) - T\big),$$

since both functions are left $P(\mathbb{Q})$-invariant. Combining the double sum over $\delta$ and $\delta_1$ into a single sum over $\delta \in P_1(\mathbb{Q})\backslash G(\mathbb{Q})$, we write $k^T(x)$ as the sum over pairs $P_1 \subset P$ of the product of $(-1)^{\dim(A_P/A_G)}$ with

$$\sum_{\delta \in P_1(\mathbb{Q})\backslash G(\mathbb{Q})} F^{P_1}(\delta x, T)\tau_{P_1}^P\big(H_{P_1}(\delta x) - T\big)\widehat{\tau}_P\big(H_P(\delta x) - T\big)K_P(\delta x, \delta x).$$

The next step is to consider the product

$$\tau_{P_1}^P\big(H_{P_1}(\delta x) - T\big)\widehat{\tau}_P\big(H_P(\delta x) - T\big) = \tau_{P_1}^P(H_1)\widehat{\tau}_P(H_1),$$

for the vector

$$H_1 = H_{P_1}(\delta x) - T_{P_1}$$

in $\mathfrak{a}_{P_1}$. (We have written $T_{P_1}$ for the projection of $T$ onto $\mathfrak{a}_{P_1}$.) We claim that

$$\tau_{P_1}^P(H_1)\widehat{\tau}_P(H_1) = \sum_{\{P_2, Q: P \subset P_2 \subset Q\}} (-1)^{\dim(A_{P_2}/A_Q)}\tau_{P_1}^Q(H_1)\widehat{\tau}_Q(H_1),$$

for fixed groups $P_1 \subset P$. Indeed, for a given pair of parabolic subgroups $P \subset Q$, the set of $P_2$ with $P \subset P_2 \subset Q$ is bijective with the collection of subsets $\Delta_P^{P_2}$ of $\Delta_P^Q$. Since

$$(-1)^{\dim(A_{P_2}/A_Q)} = (-1)^{|\Delta_P^Q| - |\Delta_P^{P_2}|},$$

the claim follows from Identity 6.2. We can therefore write

$$\tau_{P_1}^P(H_1)\widehat{\tau}_P(H_1) = \sum_{\{P_2: P_2 \supset P\}} \sigma_{P_1}^{P_2}(H_1), \tag{8.2}$$

where

$$\sigma_{P_1}^{P_2}(H_1) = \sum_{\{Q: Q \supset P_2\}} (-1)^{\dim(A_{P_2}/A_Q)}\tau_{P_1}^Q(H_1)\widehat{\tau}_Q(H_1).$$

Lemma 8.3. *Suppose that $P_1 \subset P_2$, and that*

$$H_1 = H_1^2 + H_2, \qquad\qquad H_1^2 \in \mathfrak{a}_{P_1}^{P_2},\ H_2 \in \mathfrak{a}_{P_2}^{G},$$

*is a point in the space $\mathfrak{a}_{P_1}^G = \mathfrak{a}_{P_1}^{P_2} \oplus \mathfrak{a}_{P_2}^G$. The function $\sigma_{P_1}^{P_2}(H_1)$ then has the following properties.*

(a) *$\sigma_{P_1}^{P_2}(H_1)$ equals 0 or 1.*

(b) *If $\sigma_{P_1}^{P_2}(H_1) = 1$, then $\tau_{P_1}^{P_2}(H_1^2) = 1$, and $\|H_2\| \leq c\|H_1^2\|$, for a positive constant $c$ that depends only on $P_1$ and $P_2$.*

The proof of Lemma 8.3 is a straightforward analysis of roots and weights. It is based on the intuition gained from the example of $G = SL(3)$, $P_1 = P_0$, and $P_2$ a (standard) maximal parabolic subgroup. For the general case, we refer the reader to Lemma 6.1 of [**A3**], which gives an explicit description of the function $\sigma_{P_1}^{P_2}$ from which the conditions (a) and (b) are easily inferred. In the case of the example, $Q$ is summed over the set $\{P_2, G\}$, and we obtain a difference

$$\sigma_{P_1}^{P_2}(H_1) = \sigma_{P_0}^{P_2}(H_1) = \tau_{P_0}^{P_2}(H_1)\widehat{\tau}_{P_2}(H_1) - \tau_{P_0}(H_1)$$

of two characteristic functions. The first characteristic function is supported on the open cone generated by the vectors $\beta_1^\vee$ and $\varpi_2^\vee$ in Figure 8.5. The second characteristic function is supported on the open cone generated by $\varpi_1^\vee$ and $\varpi_2^\vee$. The difference $\sigma_{P_1}^{P_2}(H_1)$ is therefore the characteristic function of the half open cone generated by $\beta_1^\vee$ and $\varpi_1^\vee$, the region shaded in Figure 8.5. It is obvious that this function satisfies the conditions (i) and (ii).

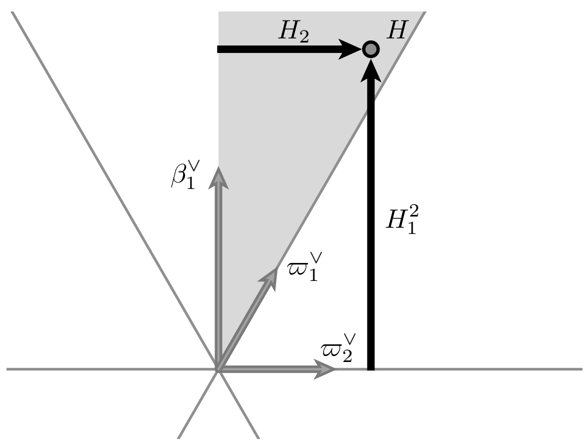



Figure 8.5. The shaded region is the complement in the upper right hand quadrant of the acute angled cone spanned by $\varpi_1^\vee$ and $\varpi_2^\vee$. It represents the support of the characteristic function $\sigma_{P_1}^{P_2}(H_1)$ attached to $G = SL(3)$, $P_1 = P_0$ minimal, and $P_2$ maximal. This function has compact support in the horizontal component $H_2$ of $H_1$, and semi-infinite support in the vertical component $H_1^2$.

$\square$

We have established that $k^T(x)$ equals

$$\sum_{P_1\subset P}(-1)^{\dim(A_P/A_G)}\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})}F^{P_1}(\delta x,T)\cdot$$
$$\cdot\Big(\sum_{\{P_2:P_2\supset P\}}\sigma_{P_1}^{P_2}\big(H_{P_1}(\delta x)-T\big)\Big)K_P(\delta x,\delta x).$$

Therefore $k^T(x)=k^T(x,f)$ has an expansion

$$\sum_{P_1\subset P_2}\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})}F^{P_1}(\delta x,T)\sigma_{P_1}^{P_2}\big(H_{P_1}(\delta x)-T\big)k_{P_1,P_2}(\delta x), \tag{8.3}$$

where $k_{P_1,P_2}(x)=k_{P_1,P_2}(x,f)$ is the value at $y=x$ of the alternating sum

$$\begin{aligned}K_{P_1,P_2}(x,y)&=\sum_{\{P:P_1\subset P\subset P_2\}}(-1)^{\dim(A_P/A_G)}K_P(x,y)\\&=\sum_P(-1)^{\dim(A_P/A_G)}\sum_{\gamma\in M_P(\mathbb{Q})}\int_{N_P(\mathbb{A})}f(x^{-1}\gamma ny)\mathrm{d}n.\end{aligned}\tag{8.4}$$

The function

$$\chi^T(x)=\chi^T_{P_1,P_2}(x)=F^{P_1}(x,T)\sigma_{P_1}^{P_2}\big(H_{P_1}(x)-T\big)$$

takes values 0 or 1. We can therefore write

$$|k^T(x)|\le\sum_{P_1\subset P_2}\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})}\chi^T(\delta x)|k_{P_1,P_2}(\delta x)|.$$

It follows that

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1}|k^T(x)|\mathrm{d}x\le\sum_{P_1\subset P_2}\int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1}\chi^T(x)|k_{P_1,P_2}(x)|\mathrm{d}x.\tag{8.5}$$

Suppose that the variable of integration $x\in P_1(\mathbb{Q})\backslash G(\mathbb{A})^1$ on the right hand side of this inequality is decomposed as

$$x=p_1a_1k,\tag{8.6}$$

and

$$H_{P_1}(a_1)=H_1^2+H_2,\qquad H_1^2\in\mathfrak{a}_{P_1}^{P_2},\ H_2\in\mathfrak{a}_{P_2}^G,\tag{8.7}$$

where $p_1\in P_1(\mathbb{Q})\backslash M_{P_1}(\mathbb{A})^1N_{P_1}(\mathbb{A})$, $a_1\in A_{P_1}(\mathbb{R})^0\cap G(\mathbb{A})^1$, and $k\in K$. The integrand is then compactly supported in $p_1$, $k$ and $H_2$. We need only study its behaviour in $H_1^2$, for points $H_1^2$ with $\tau_{P_1}^{P_2}(H_1^2-T)>0$. This is the heart of the proof. It is where we exploit the cancellation implicit in the alternating sum over $P$.

We claim that the sum over $\gamma\in M_P(\mathbb{Q})$ in the formula for

$$k_{P_1,P_2}(x)=K_{P_1,P_2}(x,x)$$

can be restricted to the subset $P_1(\mathbb{Q})\cap M_P(\mathbb{Q})$ of $M_P(\mathbb{Q})$. More precisely, given standard parabolic subgroups $P_1\subset P\subset P_2$, a point $T\in\mathfrak{a}_0^+$ with $\beta(T)$ large (relative to the support of $f$) for each $\beta\in\Delta_0$, and a point $x\in P_1(\mathbb{Q})\backslash G(\mathbb{A})^1$ with $\chi^T(x)\ne0$, we claim that

$$\int_{N_P(\mathbb{A})}f(x^{-1}\gamma nx)\mathrm{d}n=0,$$

for any element $\gamma$ in the complement of $P_1(\mathbb{Q})$ in $M_P(\mathbb{Q})$.

Consider the example that $G = SL(2)$, $P_1 = P_0$, and $P = P_2 = G$. Then $N_P = N_G = \{1\}$. Suppose that $\gamma$ belongs to the set

$$M_P(\mathbb{Q}) - P_1(\mathbb{Q}) = G(\mathbb{Q}) - P_0(\mathbb{Q}).$$

Then $\gamma$ is of the form $\begin{pmatrix} * & * \\ c & * \end{pmatrix}$, for some element $c \in \mathbb{Q}^*$. Suppose that $x$ is such that $\chi^T(x) \neq 0$. Then

$$x = p_1 a_1 k, \qquad p_1 = \begin{pmatrix} u_1 & * \\ 0 & u_1^{-1} \end{pmatrix},\ a_1 = \begin{pmatrix} \mathrm{e}^r & 0 \\ 0 & \mathrm{e}^{-r} \end{pmatrix},\ k \in K,$$

for an element $u_1 \in \mathbb{A}^*$ with $|u_1| = 1$ and a real number $r$ that is large. We see that

$$\begin{aligned}
&\int_{N_P(\mathbb{A})} f(x^{-1}\gamma n x)\mathrm{d}n = f(x^{-1}\gamma x) \\
&\quad = f\left(k^{-1}\begin{pmatrix} \mathrm{e}^r & 0 \\ 0 & \mathrm{e}^{-r} \end{pmatrix}^{-1}\begin{pmatrix} u_1 & 0 \\ 0 & u_1^{-1} \end{pmatrix}^{-1}\begin{pmatrix} * & * \\ c & * \end{pmatrix}\begin{pmatrix} u_1 & * \\ 0 & u_1^{-1} \end{pmatrix}\begin{pmatrix} \mathrm{e}^r & 0 \\ 0 & \mathrm{e}^{-r} \end{pmatrix} k\right) \\
&\quad = f\left(k^{-1}\begin{pmatrix} * & * \\ u_1^2\mathrm{e}^{2r}c & * \end{pmatrix} k\right).
\end{aligned}$$

Since $f$ is compactly supported, and $|u_1^2\mathrm{e}^{2r}c| = \mathrm{e}^{2r}$ is large, the last expression vanishes. The claim therefore holds in the special case under consideration.

The claim in general is established on p. 944 of [**A3**]. Taking it now for granted, we can then replace the sum over $M_P(\mathbb{Q})$ in the expression for $k_{P_1,P_2}(x)$ by a sum over $P_1(\mathbb{Q}) \cap M_P(\mathbb{Q})$. But $P_1(\mathbb{Q}) \cap M_P(\mathbb{Q})$ equals $M_{P_1}(\mathbb{Q})N_{P_1}^P(\mathbb{Q})$, where $N_{P_1}^P = N_{P_1} \cap M_P$ is the unipotent radical of the parabolic subgroup $P_1 \cap M_P$ of $M_P$. We may therefore write $k_{P_1,P_2}(x)$ as

$$\sum_{\{P: P_1 \subset P \subset P_2\}} (-1)^{\dim(A_P/A_G)} \sum_{\mu \in M_{P_1}(\mathbb{Q})} \sum_{\nu \in N_{P_1}^P(\mathbb{Q})} \int_{N_P(\mathbb{A})} f(x^{-1}\mu\nu n x)\mathrm{d}n.$$

Now the restriction of the exponential map

$$\exp:\ \mathfrak{n}_{P_1} = \mathfrak{n}_{P_1}^P \oplus \mathfrak{n}_P \ \longrightarrow\ N_{P_1} = N_{P_1}^P N_P$$

is an isomorphism of algebraic varieties over $\mathbb{Q}$, which maps the Haar measure $\mathrm{d}x_1$ on $\mathfrak{n}_{P_1}(\mathbb{A})$ to the Haar measure $\mathrm{d}n_1$ on $N_{P_1}(\mathbb{A})$. This allows us to write $k_{P_1,P_2}(x)$ as

$$\sum_{\mu \in M_{P_1}(\mathbb{Q})} \Big( \sum_{P: P_1 \subset P \subset P_2\}} (-1)^{\dim(A_P/A_G)} \sum_{\zeta \in \mathfrak{n}_{P_1}^P(\mathbb{Q})} \int_{\mathfrak{n}_P(\mathbb{A})} f\big(x^{-1}\mu\exp(\zeta + X)x\big)\mathrm{d}X\Big).$$

There is one more operation to be performed on our expression for $k_{P_1,P_2}(x)$. We shall apply the Poisson summation formula for the locally compact abelian group $\mathfrak{n}_{P_1}^P(\mathbb{A})$ to the sum over the discrete cocompact subgroup $\mathfrak{n}_{P_1}^P(\mathbb{Q})$. We identify $\mathfrak{n}_{P_1}^P$ with $\dim(\mathfrak{n}_{P_1}^P)$-copies of the additive group by choosing a rational basis of root vectors. We can then identify $\mathfrak{n}_{P_1}^P(\mathbb{A})$ with its dual group by means of the standard bilinear form $\langle\cdot,\cdot\rangle$ on $\mathbb{A}^{\dim(\mathfrak{n}_{P_1}^P)}$ and a nontrivial additive character $\psi$ on $\mathbb{A}/\mathbb{Q}$. We obtain an expression

$$\sum_{\mu}\sum_{P}(-1)^{\dim(A_P/A_G)}\sum_{\xi\in\mathfrak{n}_{P_1}^{P}(\mathbb{Q})}\int_{\mathfrak{n}_{P_1}(\mathbb{A})} f\big(x^{-1}\mu\exp(X_1)x\big)\psi\big(\langle\xi,X_1\rangle\big)\mathrm{d}X_1$$

for $k_{P_1,P_2}(x)$. But $\mathfrak{n}_{P_1}^{P}(\mathbb{Q})$ is contained in $\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})$, for any $P$ with $P_1\subset P\subset P_2$. As $P$ varies, certain summands will occur more than once, with differing signs. This allows us at last to effect the cancellation given by the alternating sum over $P$. Set

$$\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})'=\big\{\xi\in\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q}):\ \xi\notin\mathfrak{n}_{P_1}^{P}(\mathbb{Q}),\ \text{for any } P\subsetneq P_2\big\}.$$

It then follows from Identity 6.2 that $k_{P_1,P_2}(x)$ equals

(8.8)

$$(-1)^{\dim(A_{P_2}/A_G)}\sum_{\mu\in M_{P_1}(\mathbb{Q})}\ \sum_{\xi\in\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})'}\left(\int_{\mathfrak{n}_{P_1}(\mathbb{A})} f(x^{-1}\mu\exp X_1x)\psi\big(\langle\xi,X_1\rangle\big)\mathrm{d}X_1\right).$$

We have now obtained an expression for $k_{P_1,P_2}(x,x)$ that will be rapidly decreasing in the coordinate $H_1^2$ of $x$, relative to the decompositions (8.6) and (8.7). The main reason is that the integral

$$h_{x,\mu}(Y_1)=\int_{\mathfrak{n}_{P_1}(\mathbb{A})} f(x^{-1}\mu\exp X_1x)\psi\big(\langle Y_1,X_1\rangle\big)\mathrm{d}X_1$$

is a Schwartz-Bruhat function of $Y_1\in\mathfrak{n}_{P_1}(\mathbb{A})$. This function varies smoothly with $x\in G(\mathbb{A})$, and is finitely supported in $\mu\in M_{P_1}(\mathbb{Q})$, independently of $x$ in any compact set.

We substitute the formula (8.8) for $k_{P_1,P_2}(x)$ into the right hand side of (8.5), and then decompose the integral over $x$ according to the (8.6). We deduce that the the integral

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1}|k^T(x)|\mathrm{d}x$$

is bounded by a constant multiple of

$$\sum_{P_1\subset P_2}\ \sum_{\mu\in M_{P_1}(\mathbb{Q})}\ \sum_{\xi\in\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})'}\ \sup_y\int\big|h_{y,\mu}\big(\mathrm{Ad}(a_1)\xi\big)\big|\mathrm{d}a_1, \tag{8.9}$$

where the integral is taken over the set of elements $a_1$ in $A_{P_1}(\mathbb{R})^0\cap G(\mathbb{A})^1$ with $\sigma_{P_1}^{P_2}\big(H_{P_1}(a_1)-T\big)=1$, and the supremum is taken over the compact subset of elements

$$y=a_1^{-1}p_1a_1k,\qquad p_1\in P_1(\mathbb{Q})\backslash M_{P_1}(\mathbb{A})^1N_{P_1}(\mathbb{A}),\ a\in A_{P_1}(\mathbb{R})^0\cap G(\mathbb{A})^1,\ k\in K,$$

in $G(\mathbb{A})^1$ with $F^{P_1}(p_1,T)=\sigma_{P_1}^{P_2}\big(H_{P_1}(a_1)-T\big)=1$. We have used two changes of variables of integration here, with complementary Radon-Nikodym derivatives, which together have allowed us to write

$$\mathrm{d}X_1\mathrm{d}x=\mathrm{d}(a_1^{-1}X_1a_1)\mathrm{d}p_1\mathrm{d}a_1\mathrm{d}k,\qquad\qquad x=p_1a_1k.$$

The mapping $\mathrm{Ad}(a_1)$ in (8.9) acts by dilation on $\xi$. We leave the reader to show that this property implies that (8.9) is finite, and hence that the integral of $|k^T(x)|$ converges. (See [**A3**, Theorem 7.1].) This completes our discussion of the proof of Theorem 6.1. □

We have seen that Lemma 8.3 is an essential step in the proof of Theorem 6.1. There is a particularly simple case of this lemma that is important for other combinatorial arguments. It is the identity

$$\sum_{\{P:P_1\subset P\}}(-1)^{\dim(A_{P_1}/A_P)}\tau_{P_1}^P(H_1)\widehat{\tau}_P(H_1)=\begin{cases}0, & \text{if } P_1\neq G,\\ 1, & \text{if } P_1=G,\end{cases}\tag{8.10}$$

obtained by setting $P_2=P_1$. The identity holds for any standard parabolic subgroup $P_1$ and any point $H_1\in\mathfrak{a}_{P_1}$. Indeed, the left hand side of (8.10) equals $\sigma_{P_1}^{P_1}(H_1)$, so the identity follows from condition (ii) of Lemma 8.3.

There is also a parallel identity

$$\sum_{\{P:P_1\subset P\}}(-1)^{\dim(A_{P_1}/A_P)}\widehat{\tau}_{P_1}^P(H_1)\tau_P(H_1)=\begin{cases}0, & \text{if } P_1\neq G,\\ 1, & \text{if } P_1=G,\end{cases}\tag{8.11}$$

related by inversion to (8.10). To see this, it is enough to consider the case that $P_1$ is proper in $G$. One can then derive (8.11) from (8.10) by evaluating the expression

$$\sum_{\{P,Q:P_1\subset P\subset Q\}}(-1)^{\dim(A_P/A_Q)}\widehat{\tau}_{P_1}^P(H_1)\tau_P^Q(H_1)\widehat{\tau}_Q(H_1)$$

as two different iterated sums. For if one takes $Q$ to index the inner sum, and assumes inductively that (8.11) holds whenever $G$ is replaced by a proper Levi subgroup, one finds that the expression equals the sum of $\widehat{\tau}_{P_1}(H_1)$ with the left hand side of (8.11). On the other hand, by taking the inner sum to be over $P$, one sees from (8.10) that the expression reduces simply to $\widehat{\tau}_{P_1}(H_1)$. It follows that the left hand side of (8.11) vanishes, as required. In the case that $G=SL(3)$ and $P_1=P_0$ is minimal, the reader can view the left hand side of (8.11) (or of (8.10)) as an algebraic sum of four convex cones, formed in the obvious way from Figure 5.1. In general, (8.11) is only one of several identities that can be deduced from (8.10). We shall describe these identities, known collectively as Langlands' combinatorial lemma, in §17.

## 9. Qualitative behaviour of $J^T(f)$

Theorem 6.1 allows us to define the linear form

$$J^T(f)=J^{G,T}(f)=\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1}k^T(x,f)\mathrm{d}x,\qquad f\in C_c^\infty\big(G(\mathbb{A})\big),$$

on $C_c^\infty\big(G(\mathbb{A})\big)$. We are still a long way from converting the geometric and spectral expansions of $k^T(x,f)$ to an explicit trace formula. We put this question aside for the moment, in order to investigate two qualitative properties of $J^T(f)$.

The first property concerns the behaviour of $J^T(f)$ as a function of $T$.

THEOREM 9.1. *For any $f\in C_c^\infty\big(G(\mathbb{A})\big)$, the function*

$$T\longrightarrow J^T(f),$$

*defined for $T\in\mathfrak{a}_0^+$ sufficiently regular, is a polynomial in $T$ whose degree is bounded by the dimension of $\mathfrak{a}_0^G$.*

We shall sketch the proof of Theorem 9.1. Let $T_1$ be a fixed point in $\mathfrak{a}_0$ with $\beta(T_1)$ large for every $\beta \in \Delta_0$, and let $T \in \mathfrak{a}_0$ be a variable point with $\beta(T - T_1) > 0$ for each $\beta$. It would be enough to show that the function

$$T \longrightarrow J^T(f) - J^{T_1}(f) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \big(k^T(x) - k^{T_1}(x)\big)\mathrm{d}x$$

is a polynomial in $T$. If we substitute the definition (6.1) for the two functions in the integrand, we see that the only terms in the resulting expression that depend on $T$ and $T_1$ are differences of characteristic functions

$$\widehat{\tau}_P\big(H_P(\delta x) - T\big) - \widehat{\tau}_P\big(H_P(\delta x) - T_1\big).$$

We need to compare the supports of these two functions. We shall do so by expanding the first function in terms of analogues of the second function for smaller groups.

Suppose that $H$ and $X$ range over points in $\mathfrak{a}_0^G$. We define functions

$$\Gamma'_P(H, X), \qquad P \supset P_0,$$

inductively on $\dim(A_P/A_G)$ by setting

$$\widehat{\tau}_P(H - X) = \sum_{\{Q : Q \supset P\}} (-1)^{\dim(A_Q/A_G)} \widehat{\tau}_P^Q(H) \Gamma'_Q(H, X), \tag{9.1}$$

for any $P$. Since the summand with $Q = P$ equals the product of $(-1)^{\dim(A_P/A_G)}$ with $\Gamma'_P(H, X)$, (9.1) does indeed give an inductive definition of $\Gamma'_P(H, X)$ in terms of functions $\Gamma'_Q(H, X)$ with $\dim(A_Q/A_G)$ less than $\dim(A_P/A_G)$. It follows inductively from the definition that $\Gamma'_P(H, X)$ depends only on the projections $H_P$ and $T_P$ of $H$ and $T$ onto $\mathfrak{a}_P^G$.

LEMMA 9.2. (a) *For any $X$ and $P$, the function*

$$H \longrightarrow \Gamma'_P(H, X), \qquad H \in \mathfrak{a}_P^G,$$

*is compactly supported.*

(b) *The function*

$$X \longrightarrow \int_{\mathfrak{a}_P^G} \Gamma'_P(H, X)\mathrm{d}H, \qquad X \in \mathfrak{a}_P^G,$$

*is a homogeneous polynomial of degree equal to* $\dim(\mathfrak{a}_P^G)$.

Once again, we shall be content to motivate the lemma geometrically in some special cases. For the general case, we refer the reader to [**A5**, Lemmas 2.1 and 2.2].

The simplest case is when $\mathfrak{a}_P^G$ is one-dimensional. Suppose for example that $G = SL(3)$ and $P = P_1$ is a maximal parabolic subgroup. Then $Q$ is summed over the set $\{P_1, G\}$. Taking $X$ to be a fixed point in positive chamber in $\mathfrak{a}_P^G$, we see that $H \to \Gamma'_P(H, X)$ is the difference of characteristic functions of two open half lines, and is hence the characteristic function of the bounded half open interval in Figure 9.1.

Suppose that $G = SL(3)$ and $P = P_0$. Then $Q$ is summed over the set $\{P_0, P_1, P_2, G\}$, where $P_1$ and $P_2$ are the maximal parabolic subgroups represented in Figure 5.1. If $X$ is a fixed point in the positive chamber $\mathfrak{a}_0^+$ in $\mathfrak{a}_P^G = \mathfrak{a}_0$, we can describe the summands in (9.1) corresponding to $P_1$ and $P_2$ with the help of

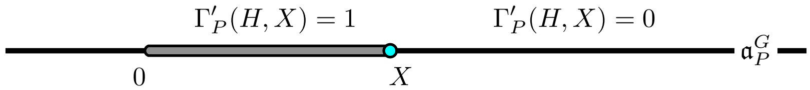



FIGURE 9.1. The half open, bounded interval represents the support of a characteristic function $\Gamma'_P(H,X)$ of $H$, for a maximal parabolic subgroup $P \subset G$. It is the complement of one open half line in another.

Figure 9.1. We see that the function $H \to \Gamma'_P(H,X)$ is a signed sum of characteristic functions of four regions, two obtuse cones and two semi-infinite rectangles. Keeping track of the signed contribution of each region in Figure 9.2, we see that $\Gamma'_P(H,X)$ is the characteristic function of the bounded shaded region in the figure. It is clear that the area of this figure is a homogeneous polynomial of degree 2 in the coordinates of $X$.

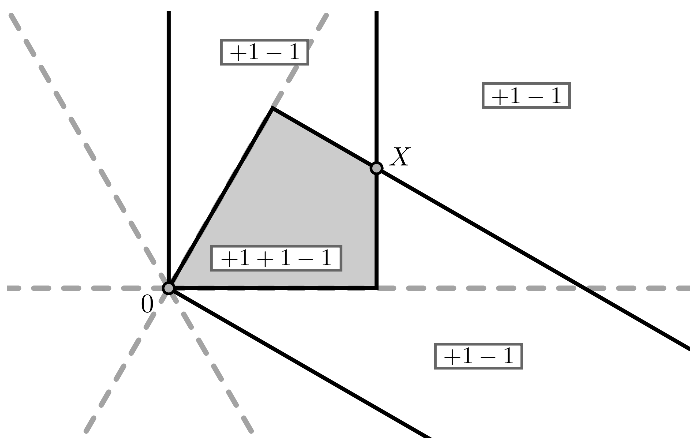



FIGURE 9.2. The bounded shaded region represents the support of the characteristic function $\Gamma'_P(H,X)$ of $H$, for the minimal parabolic subgroup $P = P_0$ of $SL(3)$. It is an algebraic sum of four unbounded regions, the two obtuse angled cones with vertices $0$ and $X$, and the two semi-infinite rectangles defined by $0$ and the projections of $X$ onto the two spaces $\mathfrak{a}_{P_1}$ and $\mathfrak{a}_{P_2}$.

$\square$

Let us use Lemma 9.2 to prove Theorem 9.1. We set $H = H_P(\delta x) - T_1$ and $X = T - T_1$. Then $H - X$ equals $H_P(\delta x) - T$, and the expansion (9.1) is

$$\widehat{\tau}_P\big(H_P(\delta x)-T\big) = \sum_{Q \supset P} (-1)^{\dim(A_Q/A_G)} \widehat{\tau}_P^Q\big(H_P(\delta x)-T_1\big)\Gamma'_Q\big(H_P(\delta x)-T_1, T-T_1\big).$$

Substituting the right hand side of this formula into the definition of $J^T(f)$, we obtain

$$\begin{aligned}
J^T(f) &= \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})} \sum_{Q\supset P} (-1)^{\dim(A_Q/A_G)} C(\delta x)\mathrm{d}x \\
&= \sum_Q \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_{P\subset Q} (-1)^{\dim(A_P/A_Q)} \sum_{\delta\in Q(\mathbb{Q})\backslash G(\mathbb{Q})} \sum_{\eta\in P(\mathbb{Q})\backslash Q(\mathbb{Q})} C(\eta\delta x)\mathrm{d}x \\
&= \sum_Q \int_{Q(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_{P\subset Q} (-1)^{\dim(A_P/A_Q)} \sum_{\eta\in P(\mathbb{Q})\cap M_Q(\mathbb{Q})\backslash M_Q(\mathbb{Q})} C(\eta x)\mathrm{d}x,
\end{aligned}$$

where

$$C(y) = K_P(y,y)\widehat{\tau}_P^Q\big(H_P(y)-T_1\big)\Gamma'_Q\big(H_Q(y)-T_1, T-T_1\big).$$

We are going to make a change of variables in the integral over $x$ in $Q(\mathbb{Q})\backslash G(\mathbb{A})^1$. Since the expression we ultimately obtain will be absolutely convergent, this change of variables, as well as the ones above, will be justified by Fubini's theorem.

We write $x = n_Q m_Q a_Q k$, for variables $n_Q, m_Q, a_Q$ and $k$ in $N_Q(\mathbb{Q})\backslash N_Q(\mathbb{A})$, $M_Q(\mathbb{Q})\backslash M_Q(\mathbb{A})^1$, $A_Q(\mathbb{R})^0 \cap G(\mathbb{A})^1$, and $K$ respectively. The invariant measures are then related by

$$\mathrm{d}x = \delta_Q(a_Q)\mathrm{d}n_Q\mathrm{d}m_Q\mathrm{d}a_Q\mathrm{d}k.$$

The three factors in the product $C(\eta x)$ become

$$\begin{aligned}
\Gamma'_Q\big(H_Q(\eta x)-T_1, T-T_1\big) &= \Gamma'_Q\big(H_Q(x)-T_1, T-T_1\big) = \Gamma'_Q\big(H_Q(a_Q)-T_1, T-T_1\big), \\
\widehat{\tau}_P^Q\big(H_P(\eta x)-T_1\big) &= \widehat{\tau}_P^Q\big(H_P(\eta m_Q)-T_1\big),
\end{aligned}$$

and

$$K_P(\eta x, \eta x) = \int_{N_P(\mathbb{A})} \sum_{\gamma\in M_P(\mathbb{Q})} f(k^{-1}m_Q^{-1}a_Q^{-1}n_Q^{-1}\eta^{-1}\cdot\gamma n\cdot\eta n_Q a_Q m_Q k)\mathrm{d}n.$$

In this last integrand, the element $\eta$ normalizes the variables $n_Q$ and $a_Q$ without changing the measures. The same is true of the element $\gamma$. We can therefore absorb both variables in the integral over $n$. Since

$$\delta_Q(a_Q)\mathrm{d}n = \mathrm{d}(a_Q^{-1}n_Q^{-1}n n_Q a_Q),$$

the product of $\delta_Q(a_Q)$ with $K_P(\eta x, \eta x)$ equals

$$\int_{N_P(\mathbb{A})} \sum_{\gamma\in M_P(\mathbb{Q})} f(k^{-1}m_Q^{-1}\eta^{-1}\cdot\gamma n\cdot\eta m_Q k)\mathrm{d}n. \tag{9.2}$$

The original variable $n_Q$ has now disappeared from all three factors, so we may as well write

$$\mathrm{d}n = \mathrm{d}(n^Q n_Q) = \mathrm{d}n^Q\mathrm{d}n_Q, \qquad n^Q\in N_P^Q(\mathbb{A}),\ n_Q\in N_Q(\mathbb{A}),$$

for the decomposition of the measure in $N_P(\mathbb{A})$. The last expression (9.2) is the only factor that depends on the original variable $k$. Its integral over $k$ equals

$$\begin{aligned}
&\int_K \int_{N_Q(\mathbb{A})} \int_{N_P^Q(\mathbb{A})} \sum_{\gamma \in M_P(\mathbb{Q})} f(k^{-1} m_Q^{-1} \eta^{-1} \cdot \gamma n^Q n_Q \cdot \eta m_Q k) \mathrm{d}n^Q \mathrm{d}n_Q \mathrm{d}k \\
&= \int_{N_P^Q(\mathbb{A})} \sum_{\gamma \in M_P(\mathbb{Q})} \int_K \int_{N_Q(\mathbb{A})} f(k^{-1} m_Q^{-1} \eta^{-1} \cdot \gamma n^Q \cdot \eta m_Q n_Q k) \mathrm{d}n_Q \mathrm{d}k \mathrm{d}n^Q \\
&= \int_{N_P^Q(\mathbb{A})} \sum_{\gamma \in M_P(\mathbb{Q})} f_Q(m_Q^{-1} \eta^{-1} \cdot \gamma n^Q \cdot \eta m_Q) \mathrm{d}n^Q \\
&= K_{P \cap M_Q}(\eta m_Q, \eta m_Q),
\end{aligned}$$

where

$$f_Q(m) = \delta_Q(m)^{\frac{1}{2}} \int_K \int_{N_Q(\mathbb{A})} f(k^{-1} m n_Q k) \mathrm{d}n_Q \mathrm{d}k, \qquad m \in M_Q(\mathbb{A}),$$

and $K_{P\cap M_Q}(\cdot,\cdot)$ is the induced kernel (4.1), but with $G$, $P$, and $f$ replaced by $M_Q$, $P \cap M_Q$, and $f_Q$ respectively. We have used the facts that

$$\mathrm{d}n_Q = \mathrm{d}\big((\eta m_Q)^{-1} n_Q (\eta m_Q)\big),$$

for $\eta$ and $m_Q$ as above, and that

$$\delta_Q(m) = \mathrm{e}^{2\rho_Q(H_Q(m))} = 1,$$

when $m = \gamma$ lies in $M_Q(\mathbb{Q})$. The correspondence $f \to f_Q$ is a continuous linear mapping from $C_c^\infty\big(G(\mathbb{A})\big)$ to $C_c^\infty\big(M_Q(\mathbb{A})\big)$. It was introduced originally by Harish-Chandra to study questions of descent.

We now collect the various terms in the formula for $J^T(f)$. We see that $J^T(f)$ equals the sum over $Q$ and the integral over $m_Q$ in $M_Q(\mathbb{Q})\backslash M_Q(\mathbb{A})^1$ of the product of

$$\sum_{P \subset Q} (-1)^{\dim(A_P/A_Q)} \sum_{\eta \in P(\mathbb{Q}) \cap M_Q(\mathbb{Q}) \backslash M_Q(\mathbb{Q})} K_{P\cap M_Q}(\eta m_Q, \eta m_Q) \widehat{\tau}_P^Q\big(H_P(\eta m_Q) - T_1\big)$$

with the factor

$$\begin{aligned}
p_Q(T_1, T) &= \int_{A_Q(\mathbb{R})^0 \cap G(\mathbb{R})^1} \Gamma_Q'\big(H_Q(a_Q) - T_1, T - T_1\big) \mathrm{d}a \\
&= \int_{\mathfrak{a}_Q^G} \Gamma_Q'(H - T_1, T - T_1) \mathrm{d}H.
\end{aligned}$$

By Lemma 9.2, the last factor is a polynomial in $T$ of degree equal to $\dim(\mathfrak{a}_Q^G)$. To analyze the first factor, we note that

$$\dim(A_P/A_Q) = \dim(A_{P\cap M_Q}/A_{M_Q})$$

and

$$\widehat{\tau}_P^Q\big(H_P(\eta m_Q) - T_1\big) = \widehat{\tau}_{P\cap M_Q}^{M_Q}\big(H_{P\cap M_Q}(\eta m_Q) - T_1\big),$$

and that the mapping $P \to P \cap M_Q$ is a bijection from the set of standard parabolic subgroups $P$ of $G$ with $P \subset Q$ onto the set of standard parabolic subgroups of $M_Q$. The first factor therefore equals the analogue $k^{T_1}(m_Q, f_Q)$ for $T_1$, $m_Q$ and $f_Q$ of the truncated kernel $k^T(x,f)$. Its integral over $m_Q$ equals $J^{M_Q,T_1}(f_Q)$. We conclude that

$$J^T(f) = \sum_{Q\supset P_0} J^{M_Q,T_1}(f_Q)p_Q(T_1,T). \tag{9.3}$$

Therefore $J^T(f)$ is a polynomial in $T$ whose degree is bounded by the dimension of $\mathfrak{a}_0^G$. This completes the proof of Theorem 9.1. $\square$

Having established Theorem 9.1, we are now free to define $J^T(f)$ at any point $T$ in $\mathfrak{a}_0$. We could always set $T = 0$. However, it turns out that there is a better choice in general. The question is related to the choice of minimal parabolic subgroup $P_0$.

We write $\mathcal{P}(M_0)$ for the set of (minimal) parabolic subgroups of $G$ with Levi component $M_0$. The mapping

$$s \longrightarrow sP_0 = w_s P_0 w_s^{-1}, \qquad s \in W_0,$$

is then a bijection from $W_0$ to $\mathcal{P}(M_0)$. We recall that $w_s$ is a representative of $s$ in $G(\mathbb{Q})$. If $G = GL(n)$, we can take $w_s$ to be a permutation matrix, an element in $G(\mathbb{Q})$ that also happens to lie in the standard maximal compact subgroup $K$ of $G(\mathbb{A})$. In general, however, $s$ might require a separate representative $\widetilde{w}_s$ in $K$. The quotient $w_s^{-1}\widetilde{w}_s$ does belong to $M_0(\mathbb{A})$, so the point

$$H_{P_0}(w_s^{-1}) = H_{M_0}(w_s^{-1}\widetilde{w}_s)$$

in $\mathfrak{a}_0$ is independent of the choice of $P_0$. By arguing inductively on the length of $s \in W_0$, one shows that there is a unique point $T_0 \in \mathfrak{a}_0^G$ such that

$$H_{P_0}(w_s^{-1}) = T_0 - s^{-1}T_0, \tag{9.4}$$

for every $s \in W_0$. (See [**A5**, Lemma 1.1].) In the case that $G$ equals $GL(n)$ and $K$ is the standard maximal compact subgroup of $GL(n,\mathbb{A})$, $T_0 = 0$.

Proposition 9.3. *The linear form*

$$J(f) = J^G(f), \qquad f \in C_c^\infty\big(G(\mathbb{A})\big),$$

*defined as the value of the polynomial*

$$J^T(f) = J^{G,T}(f)$$

*at $T = T_0$, is independent of the choice of $P_0 \in \mathcal{P}(M_0)$.*

The proof of Proposition 9.3 is a straightforward exercise. If $T \in \mathfrak{a}_0$ is highly regular relative to $P_0$, $sT$ is highly regular relative to the group $P_0' = sP_0$ in $\mathcal{P}(M_0)$. The mapping

$$P \;\longrightarrow\; P' = sP = w_s P w_s^{-1}, \qquad P \supset P_0,$$

is a bijection between the relevant families $\{P \supset P_0\}$ and $\{P' \supset P_0'\}$ of standard parabolic subgroups. For any $P$, the mapping $\delta \to \delta' = w_s\delta$ is a bijection from $P(\mathbb{Q})\backslash G(\mathbb{Q})$ onto $P'(\mathbb{Q})\backslash G(\mathbb{Q})$. It follows from the definitions that

$$\begin{aligned}
\widehat{\tau}_P\big(H_P(\delta x) - T\big) &= \widehat{\tau}_{P'}\big(sH_P(w_s^{-1}\delta' x) - sT\big)\\
&= \widehat{\tau}_{P'}\big(sH_P(\widetilde{w}_s^{-1}\delta' x) + sH_{P_0}(w_s^{-1}) - sT\big)\\
&= \widehat{\tau}_{P'}\big(H_{P'}(\delta' x) - (sT - sT_0 + T_0)\big).
\end{aligned}$$

Comparing the definition (6.1) of the truncated kernel with its analogue for $P_0' = sP_0$, we see that

$$J_{P_0}^T(f) = J_{sP_0}^{sT-sT_0+T_0}(f),$$

where the subscripts indicate the minimal parabolic subgroups with respect to which the linear forms have been defined. Each side of this identity extends to a polynomial function of $T \in \mathfrak{a}_0$. Setting $T = T_0$, we see that the linear form

$$J(f) = J^{T_0}_{P_0}(f) = J^{T_0}_{sP_0}(f)$$

is indeed independent of the choice of $P_0$. (See [**A5**, p. 18–19].) $\square$

The second qualitative property of $J^T(f)$ concerns its behaviour under conjugation by $G(\mathbb{A})$. A *distribution* on $G(\mathbb{A})$ is a linear form $I$ on $C_c^\infty\big(G(\mathbb{A})\big)$ that is continuous with respect to the natural topology. The distribution is said to be *invariant* if

$$I(f^y) = I(f), \qquad\qquad f \in C_c^\infty\big(G(\mathbb{A})\big),\ y \in G(\mathbb{A}),$$

where

$$f^y(x) = f(yxy^{-1}).$$

The proof of Theorem 6.1 implies that $f \to J^T(f)$ is a distribution if $T \in \mathfrak{a}^+_{P_0}$ is sufficiently regular. Since $J^T(f)$ is a polynomial in $T$, its coefficients are also distributions. In particular, $f \to J(f)$ is a distribution on $G(\mathbb{A})$, which is independent of the choice of $P_0 \in \mathcal{P}(M_0)$. We would like to compute its obstruction to being invariant.

Consider a point $y \in G(\mathbb{A})$, a function $f \in C_c^\infty\big(G(\mathbb{A})\big)$, and a highly regular point $T \in \mathfrak{a}_0^+$. We are interested in the difference $J^T(f^y) - J^T(f)$.

To calculate $J^T(f^y)$, we have to replace the factor

$$K_P(\delta x, \delta x) = \sum_{\gamma \in M_P(\mathbb{Q})} \int_{N_P(\mathbb{A})} f(x^{-1}\delta^{-1}\gamma n \delta x)\mathrm{d}n$$

in the truncated kernel (6.1) by the expression

$$\sum_{\gamma \in M_P(\mathbb{Q})} \int_{N_P(\mathbb{A})} f^y(x^{-1}\delta^{-1}\gamma n \gamma x)\mathrm{d}n = K_P(\delta x y^{-1}, \delta x y^{-1}).$$

The last expression is invariant under translation of $y$ by the central subgroup $A_G(\mathbb{R})^0$. We may as well therefore assume that $y$ belongs to the subgroup $G(\mathbb{A})^1$ of $G(\mathbb{A})$. With this condition, we can make a change of variables $x \to xy$ in the integral over $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ that defines $J^T(f^y)$. We see that $J^T(f^y)$ equals

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Big( \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} K_P(\delta x, \delta x) \widehat{\tau}_P\big(H_P(\delta x y) - T\big)\Big)\mathrm{d}x.$$

If $\delta x = nmak$, for elements $n$, $m$, $a$, and $k$ in $N_P(\mathbb{A})$, $M_P(\mathbb{A})^1$, $A_P(\mathbb{R})^0 \cap G(\mathbb{A})^1$, and $K$ respectively, set $k_P(\delta x) = k$. We can then write

$$\begin{aligned} \widehat{\tau}_P\big(H_P(\delta x y) - T\big) &= \widehat{\tau}_P\big(H_P(a) + H_P(ky) - T\big) \\ &= \widehat{\tau}_P\big(H_P(\delta x) - T + H_P(k_P(\delta x) y)\big). \end{aligned}$$

The last expression has an expansion

$$\sum_{Q \supset P} (-1)^{\dim(A_Q/A_G)} \widehat{\tau}^Q_P\big(H_P(\delta x) - T\big)\Gamma'_Q\big(H_P(\delta x) - T, -H_P(k_P(\delta x)y)\big)$$

given by (9.1), which we can substitute into the formula above for $J^T(f^y)$.

The discussion now is identical to that of the proof of Theorem 9.1. Set

$$u'_Q(k,y) = \int_{\mathfrak{a}_Q^G} \Gamma'_Q\big(H, -H_Q(ky)\big)\mathrm{d}H, \qquad k \in K,$$

and

$$f_{Q,y}(m) = \delta_Q(m)^{\frac{1}{2}} \int_K \int_{N_Q(\mathbb{A})} f(k^{-1}mnk)u'_Q(k,y)\mathrm{d}n\mathrm{d}k, \qquad m \in M_Q(\mathbb{A}).$$

The transformation $f \to f_{Q,y}$ is a continuous linear mapping from $C_c^\infty\big(G(\mathbb{A})\big)$ to $C_c^\infty\big(M_Q(\mathbb{A})\big)$, which varies smoothly with $y \in G(\mathbb{A})$, and depends only on the image of $y$ in $G(\mathbb{A})^1$. The proof of Theorem 9.1 then leads directly to the following analogue

$$J^T(f^y) = \sum_{Q \supset P_0} J^{M_Q,T}(f_{Q,y}) \tag{9.5}$$

of (9.3). Since we have taken $K_Q = K \cap M_Q(\mathbb{A})$ as maximal compact subgroup of $M_Q(\mathbb{A})$, $H_{P_0}(w_s^{-1})$ equals $H_{P_0 \cap M_Q}(w_s^{-1})$ for any $s$ in the subgroup $W_0^M$ of $W_0 = W_0^G$. The canonical point $T_0 \in \mathfrak{a}_0^G$, defined for $G$ by (9.4), therefore projects onto the canonical point in $\mathfrak{a}_0^Q$ attached to $M_Q$. Setting $T = T_0$ in (9.5), we obtain the following result.

Theorem 9.4. *The distribution $J$ satisfies the formula*

$$J(f^y) = \sum_{Q \supset P_0} J^{M_Q}(f_{Q,y})$$

*for conjugation of $f \in C_c^\infty\big(G(\mathbb{A})\big)$ by $y \in G(\mathbb{A})$.* □

## 10. The coarse geometric expansion

We have constructed a distribution $J$ on $G(\mathbb{A})$ from the truncated kernel $k^T(x) = k^T(x,f)$. The next step is to transform the geometric expansion for $k^T(x)$ into a geometric expansion for $J(f)$. The problem is more subtle than it might first appear. This is because the truncation $k^T(x)$ of $K(x,x)$ is not completely compatible with the decomposition of $K(x,x)$ according to conjugacy classes. The difficulty comes from those conjugacy classes in $G(\mathbb{Q})$ that are particular to the case of noncompact quotient, namely the classes that are not semisimple.

In this section we shall deal with the easy part of the problem. We shall give a geometric expansion of $J(f)$ into terms parametrized by semisimple conjugacy classes in $G(\mathbb{Q})$. The proof requires only minor variations of the discussion of the last two sections.

Recall that any element $\gamma$ in $G(\mathbb{Q})$ has a *Jordan decomposition* $\gamma = \mu\nu$. It is the unique decomposition of $\gamma$ into a product of a semisimple element $\mu = \gamma_s$ in $G(\mathbb{Q})$, with a unipotent element $\nu = \gamma_u$ in $G(\mathbb{Q})$ that commutes with $\gamma_s$. We define two elements $\gamma$ and $\gamma'$ in $G(\mathbb{Q})$ to be $\mathcal{O}$-equivalent if their semisimple parts $\gamma_s$ and $\gamma'_s$ are $G(\mathbb{Q})$-conjugate. We then write $\mathcal{O} = \mathcal{O}^G$ for the set of such equivalence classes. A class $\mathfrak{o} \in \mathcal{O}$ is thus a union of conjugacy classes in $G(\mathbb{Q})$.

The set $\mathcal{O}$ is in obvious bijection with the semisimple conjugacy classes in $G(\mathbb{Q})$. We shall say that a semisimple conjugacy class in $G(\mathbb{Q})$ is *anisotropic* if it does not intersect $P(\mathbb{Q})$, for any $P \subsetneq G$. Then $\gamma \in G(\mathbb{Q})$ represents an anisotropic class if and only if $A_G$ is the maximal $\mathbb{Q}$-split torus in the connected centralizer $H$ of $\gamma$ in

$G$. (Such classes were called elliptic in [**A3**, §2]. However, the term elliptic is better reserved for semisimple elements $\gamma$ in $G(\mathbb{Q})$ such as 1, for which $A_G$ is the maximal split torus in the *center* of $H$.) We can define an *anisotropic rational datum* to be an equivalence class of pairs $(P, \alpha)$, where $P \subset G$ is a standard parabolic subgroup, and $\alpha$ is an anisotropic conjugacy class in $M_P(\mathbb{Q})$. The equivalence relation is just conjugacy, which for standard parabolic subgroups is given by the Weyl sets $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ of §7. In other words, $(P', \alpha')$ is equivalent to $(P, \alpha)$ if $\alpha = w_s^{-1}\alpha' w_s$ for some element $s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'})$. The mapping that sends $\{(P, \alpha)\}$ to the conjugacy class of $\alpha$ in $G(\mathbb{Q})$ is a bijection onto the set of semisimple conjugacy classes in $G(\mathbb{Q})$. We therefore have a canonical bijection from the set of anisotropic rational data and our set $\mathcal{O}$. Anisotropic rational data will not be needed for the constructions of this section. We mention them in order to be able to recognize the formal relations between these constructions and their spectral analogues in §12.

In case $G = GL(n)$, the classes $\mathcal{O}$ are related to basic notions from linear algebra. The Jordan decomposition is given by Jordan normal form. Two elements $\gamma$ and $\gamma'$ in $GL(n, \mathbb{Q})$ are $\mathcal{O}$-equivalent if and only if they have the same set of complex eigenvalues (with multiplicity). This is the same as saying that $\gamma$ and $\gamma'$ have the same characteristic polynomial. The set $\mathcal{O}$ of equivalence classes in $GL(n, \mathbb{Q})$ is thus bijective with the set of rational monic polynomials of degree $n$ with nonzero constant term. If $\mathfrak{o} \in \mathcal{O}$ is an equivalence class, the intersection $\mathfrak{o} \cap P(\mathbb{Q})$ is empty for all $P \neq G$ if and only if the characteristic polynomial of $\mathfrak{o}$ is irreducible over $\mathbb{Q}$. This is the condition that $\mathfrak{o}$ consist of a single anisotropic conjugacy class in $G(\mathbb{Q})$. A general equivalence class $\mathfrak{o} \in \mathcal{O}$ consists of only one conjugacy class if and only if the elements in $\mathfrak{o}$ are all semisimple, which in turn is equivalent to saying that the characteristic polynomial of $\mathfrak{o}$ has distinct irreducible factors over $\mathbb{Q}$. We leave the reader to verify these properties from linear algebra.

If $G$ is arbitrary, we have a decomposition

$$K(x, x) = \sum_{\mathfrak{o} \in \mathcal{O}} K_{\mathfrak{o}}(x, x), \tag{10.1}$$

where

$$K_{\mathfrak{o}}(x, x) = \sum_{\gamma \in \mathfrak{o}} f(x^{-1}\gamma x).$$

More generally, we can write

$$K_P(x, x) = \sum_{\gamma \in M_P(\mathbb{Q})} \int_{N_P(\mathbb{A})} f(x^{-1}\gamma n x) \mathrm{d}n = \sum_{\mathfrak{o} \in \mathcal{O}} K_{P,\mathfrak{o}}(x, x)$$

for any $P$, where

$$K_{P,\mathfrak{o}}(x, x) = \sum_{\gamma \in M_P(\mathbb{Q}) \cap \mathfrak{o}} \int_{N_P(\mathbb{A})} f(x^{-1}\gamma n x) \mathrm{d}n.$$

We therefore have a decomposition

$$k^T(x) = \sum_{\mathfrak{o} \in \mathcal{O}} k^T_{\mathfrak{o}}(x) \tag{10.2}$$

of the truncated kernel, where

$$\begin{aligned} k^T_{\mathfrak{o}}(x) &= k^T_{\mathfrak{o}}(x,f) \\ &= \sum_P (-1)^{\dim(A_P/A_G)} \int_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})} K_{P,\mathfrak{o}}(\delta x,\delta x)\widehat{\tau}_P\big(H_P(\delta x)-T\big). \end{aligned}$$

The following extension of Theorem 6.1 can be regarded as a corollary of its proof.

COROLLARY 10.1. *The double integral*

$$\sum_{\mathfrak{o}\in\mathcal{O}} \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k^T_{\mathfrak{o}}(x,f)\mathrm{d}x \tag{10.3}$$

*converges absolutely.*

The proof of Corollary 10.1 is in fact identical to the proof of Theorem 6.1 sketched in §8, but for one point. The discrepancy arises when we apply the Poisson summation formula to the lattice $\mathfrak{n}^P_{P_1}(\mathbb{Q})$, for standard parabolic subgroups $P_1 \subset P$. To do so, we require a sum over the lattice, or what amounts to the same thing, a sum over elements $\nu \in N^P_{P_1}(\mathbb{Q})$. In the proof of Theorem 6.1, we recall that such a sum arose from the property

$$P_1(\mathbb{Q}) \cap M_P(\mathbb{Q}) = M_{P_1}(\mathbb{Q})N^P_{P_1}(\mathbb{Q}).$$

That it also occurs in treating a class $\mathfrak{o}\in\mathcal{O}$ is a consequence of the parallel property

$$P_1(\mathbb{Q}) \cap M_P(\mathbb{Q}) \cap \mathfrak{o} = \big(M_{P_1}(\mathbb{Q})\cap\mathfrak{o}\big)N^P_{P_1}(\mathbb{Q}). \tag{10.4}$$

This is in turn a consequence of the first assertion of the next lemma.

LEMMA 10.2. *Suppose that $P\supset P_0$, $\gamma\in M(\mathbb{Q})$, and $\phi\in C_c\big(N_P(\mathbb{A})\big)$. Then*

$$\sum_{\delta\in N_P(\mathbb{Q})_{\gamma_s}\backslash N_P(\mathbb{Q})}\ \sum_{\eta\in N_P(\mathbb{Q})_{\gamma_s}} \phi(\gamma^{-1}\delta^{-1}\gamma\eta\delta) = \sum_{\nu\in N_P(\mathbb{Q})} \phi(\nu)$$

*and*

$$\int_{N_P(\mathbb{A})_{\gamma_s}\backslash N_P(\mathbb{A})} \int_{N_P(\mathbb{A})_{\gamma_s}} \phi(\gamma^{-1}n_1^{-1}\gamma n_2 n_1)\mathrm{d}n_2\mathrm{d}n_1 = \int_{N_P(\mathbb{A})}\phi(n)\mathrm{d}n,$$

*where $N_P(\cdot)_{\gamma_s}$ denotes the centralizer of $\gamma_s$ in $N_P(\cdot)$.*

The proof of Lemma 10.2 is a typical change of variable argument for unipotent groups. The first assertion represents a decomposition of a sum over $N_P(\mathbb{Q})$, while the second is the corresponding decomposition of an adelic integral over $N_P(\mathbb{A})$. (See [**A3**, Lemmas 2.1 and 2.2].) $\square$

The first assertion of the lemma implies that $P(\mathbb{Q})\cap\mathfrak{o}$ equals $\big(M_P(\mathbb{Q})\cap\mathfrak{o}\big)N_P(\mathbb{Q})$. If we apply it to the pair $(M_P, P_1\cap M_P)$ in place of $(G,P)$, we obtain the required relation (10.4). We then obtain Corollary 10.1 by following step by step the proof of Theorem 6.1. (Theorem 7.1 of [**A3**] was actually stated and proved directly for the functions $k^T_{\mathfrak{o}}(x)$ rather than their sum $k^T(x)$.) $\square$

Once we have Corollary 10.1, we can apply Fubini's theorem to double integral (10.3). We obtain an absolutely convergent expansion

$$J^T(f) = \sum_{\mathfrak{o}\in\mathcal{O}} J^T_{\mathfrak{o}}(f),$$

whose terms are given by absolutely convergent integrals

$$J_{\mathfrak{o}}^T(f) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k_{\mathfrak{o}}^T(x,f)\mathrm{d}x, \qquad \mathfrak{o} \in \mathcal{O}. \tag{10.5}$$

The behaviour of $J_{\mathfrak{o}}^T(f)$ as a function of $T$ is similar to that of $J^T(f)$. We have only to apply the proof of Theorem 9.1 to the absolutely convergent integral (10.5). This tells us that for any $f \in C_c^\infty\big(G(\mathbb{A})\big)$ and $\mathfrak{o} \in \mathcal{O}$, the function

$$T \longrightarrow J_{\mathfrak{o}}^T(f),$$

defined for $T \in \mathfrak{a}_0^+$ sufficiently regular in a sense that is independent of $\mathfrak{o}$, is a polynomial in $T$ of degree bounded by the dimension of $\mathfrak{a}_0^G$. We can therefore define $J_{\mathfrak{o}}^T(f)$ for all values of $T \in \mathfrak{a}_0$ by its polynomial extension. We then set

$$J_{\mathfrak{o}}(f) = J_{\mathfrak{o}}^{T_0}(f), \qquad \mathfrak{o} \in \mathcal{O},$$

for the point $T_0 \in \mathfrak{a}_0^G$ given by (9.4). The proof of Proposition 9.3 tells us that $J_{\mathfrak{o}}(f)$ is independent of the choice of minimal parabolic subgroup $P_0 \in \mathcal{P}(M_0)$.

The distributions $J_{\mathfrak{o}}(f) = J_{\mathfrak{o}}^G(f)$ can sometimes be invariant, though they are not generally so. To see this, we apply the proof of Theorem 9.4 to the absolutely convergent integral (10.5). For any $Q \supset P_0$ and $h \in C_c^\infty\big(M_Q(\mathbb{A})\big)$, set

$$J_{\mathfrak{o}}^{M_Q}(h) = \sum_{\mathfrak{o}_Q} J_{\mathfrak{o}_Q}^{M_Q}(h), \qquad \mathfrak{o} \in \mathcal{O},$$

where $\mathfrak{o}_Q$ ranges over the finite preimage of $\mathfrak{o}$ in $\mathcal{O}^{M_Q}$ under the obvious mapping of $\mathcal{O}^{M_Q}$ into $\mathcal{O} = \mathcal{O}^G$. We then obtain the variance property

$$J_{\mathfrak{o}}(f^y) = \sum_{Q \supset P_0} J_{\mathfrak{o}}^{M_Q}(f_{Q,y}), \qquad \mathfrak{o} \in \mathcal{O},\ y \in G(\mathbb{A}), \tag{10.6}$$

in the notation of Theorem 9.4. Observe that $\mathfrak{o}$ need not lie in the image the map $\mathcal{O}^{M_Q} \to \mathcal{O}$ attached to any proper parabolic subgroup $Q \subsetneq G$. This is so precisely when $\mathfrak{o}$ is *anisotropic*, in the sense that it consists of a single anisotropic (semisimple) conjugacy class. It is in this case that the distribution $J_{\mathfrak{o}}(f)$ is invariant.

The expansion of $J^T(f)$ in terms of distributions $J_{\mathfrak{o}}^T(f)$ extends by polynomial interpolation to all values of $T$. Setting $T = T_0$, we obtain an identity

$$J(f) = \sum_{\mathfrak{o} \in \mathcal{O}} J_{\mathfrak{o}}(f), \qquad f \in C_c^\infty\big(G(\mathbb{A})\big), \tag{10.7}$$

of distributions. This is what we will call the coarse geometric expansion. The distributions $J_{\mathfrak{o}}(f)$ for which $\mathfrak{o}$ is anisotropic are to be regarded as general analogues of the geometric terms in the trace formula for compact quotient.

## 11. Weighted orbital integrals

The summands $J_{\mathfrak{o}}(f)$ in the coarse geometric expansion of $J(f)$ were defined in global terms. We need ultimately to describe them more explicitly. For example, we would like to have a formula for $J_{\mathfrak{o}}(f)$ in which the dependence on the local components $f_v$ of $f$ is more transparent. In this section, we shall solve the problem for "generic" classes $\mathfrak{o} \in \mathcal{O}$. For such classes, we shall express $J_{\mathfrak{o}}(f)$ as a weighted orbital integral of $f$.

We fix a class $\mathfrak{o} \in \mathcal{O}$, which for the moment we take to be arbitrary. Recall that

$$K_{P,\mathfrak{o}}(x,y) = \sum_{\gamma \in M_P(\mathbb{Q}) \cap \mathfrak{o}} \int_{N_P(\mathbb{A})} f(x^{-1}\gamma n y)\mathrm{d}n,$$

for any $P \supset P_0$. Lemma 10.2 provides a decomposition of the integral over $N_P(\mathbb{A})$ onto a double integral. We define a modified function

$$\widetilde{K}_{P,\mathfrak{o}}(x,y) = \sum_{\gamma \in M_P(\mathbb{Q}) \cap \mathfrak{o}} \sum_{\eta \in N_P(\mathbb{Q})_{\gamma_s} \backslash N_P(\mathbb{Q})} \int_{N_P(\mathbb{A})_{\gamma_s}} f(x^{-1}\eta^{-1}\gamma n \eta y)\mathrm{d}n \tag{11.1}$$

by replacing the outer adelic integral of the lemma with a corresponding sum of rational points. We then define a modified kernel $\widetilde{k}^T_{\mathfrak{o}}(x) = \widetilde{k}^T_{\mathfrak{o}}(x,f)$ by replacing the function $K_{P,\mathfrak{o}}(\delta x, \delta x)$ in the formula for $k^T_{\mathfrak{o}}(x)$ with the modified function $\widetilde{K}_{P,\mathfrak{o}}(\delta x, \delta x)$. That is,

$$\widetilde{k}^T_{\mathfrak{o}}(x,f) = \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(\mathbb{Q}) \backslash G(\mathbb{Q})} \widetilde{K}_{P,\mathfrak{o}}(\delta x, \delta x)\widehat{\tau}_P\big(H_P(\delta x) - T\big).$$

THEOREM 11.1. *If $T \in \mathfrak{a}^+_{P_0}$ is highly regular, the integral*

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \widetilde{k}^T_{\mathfrak{o}}(x,f)\mathrm{d}x \tag{11.2}$$

*converges absolutely, and equals $J^T_{\mathfrak{o}}(f)$.*

The proof of Theorem 11.1 is again similar to that of Theorem 6.1, or rather its modification for the class $\mathfrak{o}$ discussed in §10. Copying the formal manipulations from the first half of §8, we write

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \widetilde{k}^T_{\mathfrak{o}}(x)\mathrm{d}x = \sum_{P_1 \subset P_2} \int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \chi^T(x)\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x)\mathrm{d}x, \tag{11.3}$$

where $\chi^T(x)$ is as in (8.5), and

$$\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x) = \sum_{\{P: P_1 \subset P \subset P_2\}} (-1)^{\dim(A_P/A_G)} \widetilde{K}_{P,\mathfrak{o}}(x,x).$$

To justify these manipulations, we have to show that for any $P_1 \subset P_2$, the integral

$$\int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \chi^T(x)|\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x)|\mathrm{d}x \tag{11.4}$$

is finite. This would also establish the absolute convergence assertion of the theorem.

We estimate the integral (11.4) as in the second half of §8. We shall be content simply to mention the main steps. The first is to show that if $T$ is sufficiently regular and $\chi^T(x) \neq 0$, the summands in the formula for $\widetilde{K}_{P,\mathfrak{o}}(x,x)$ vanish for elements $\gamma$ in the complement of $P_1(\mathbb{Q}) \cap M_P(\mathbb{Q}) \cap \mathfrak{o}$ in $M_P(\mathbb{Q}) \cap \mathfrak{o}$. The next step is to write $P_1(\mathbb{Q}) \cap M_P(\mathbb{Q}) \cap \mathfrak{o}$ as a product $\big(M_{P_1}(\mathbb{Q}) \cap \mathfrak{o}\big)N^P_{P_1}(\mathbb{Q})$, by appealing to Lemma 10.2. We then have to apply Lemma 10.2 again, with $(M_P, P_1 \cap M_P)$ in place of $(G,P)$, to the resulting sum over $(\mu,\nu)$ in the product of $M_{P_1}(\mathbb{Q}) \cap \mathfrak{o}$ with $N^P_{P_1}(\mathbb{Q})$. This yields a threefold sum, one of which is taken over the set

$$N^P_{P_1}(\mathbb{Q})_{\mu_s} = \exp\big(\mathfrak{n}^P_{P_1}(\mathbb{Q})_{\mu_s}\big),$$

where $\mathfrak{n}_{P_1}^P(\mathbb{Q})_{\mu_s}$ denotes the centralizer of $\mu_s$ in the Lie algebra $\mathfrak{n}_{P_1}^P(\mathbb{Q})$. The last step is to apply the Poisson summation formula to the lattice $\mathfrak{n}_{P_1}^P(\mathbb{Q})_{\mu_s}$ in $\mathfrak{n}_{P_1}^P(\mathbb{A})_{\mu_s}$. The resulting cancellation from the alternating sum over $P$ then yields a formula for $\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x)$ analogous to the formula (8.8) for $k_{P_1,P_2}(x)$. Namely, $\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x)$ equals the product of $(-1)^{\dim(A_{P_2}/A_G)}$ with the sum over $\mu \in M_{P_1}(\mathbb{Q}) \cap \mathfrak{o}$ of

$$\sum_{\eta \in N_{P_1}(\mathbb{Q})_{\mu_s} \backslash N_{P_1}(\mathbb{Q})} \sum_{\xi \in \mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})'_{\mu_s}} \Big( \int_{\mathfrak{n}_{P_1}(\mathbb{A})_{\mu_s}} f\big(x^{-1}\eta^{-1}\mu \exp(X_1)\eta x\big)\psi\big(\langle \xi, X_1\rangle\big)\mathrm{d}X_1\Big),$$

where $\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})'_{\mu_s}$ is the intersection of $\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})_{\mu_s}$ with the set $\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})'$ in (8.8). The convergence of the integral (11.4) is then proved as at the end of §8. (See [**A3**, p. 948–949].)

Once we have shown that the integrals (11.4) are finite, we know that the identity (11.3) is valid. The remaining step is to compare it with the corresponding identity

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k_{\mathfrak{o}}^T(x)\mathrm{d}x = \sum_{P_1 \subset P_2} \int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \chi^T(x) k_{P_1,P_2,\mathfrak{o}}(x)\mathrm{d}x,$$

which we obtain by modifying the proof of Theorem 6.1 as in the last section.

Suppose that $P_1 \subset P_2$ are fixed. We can then write

$$\begin{aligned}&\int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \chi^T(x)\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x)\mathrm{d}x \\ &= \int_{M_{P_1}(\mathbb{Q})N_{P_1}(\mathbb{A})\backslash G(\mathbb{A})^1} \chi^T(x)\Big(\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} \widetilde{k}_{P_1,P_2,\mathfrak{o}}(n_1 x)\mathrm{d}n_1\Big)\mathrm{d}x,\end{aligned}$$

since $\chi^T(x)$ is left $N_{P_1}(\mathbb{A})$-invariant. The integral of $\widetilde{k}_{P_1,P_2,\mathfrak{o}}(n_1 x)$ over $n_1$ is equal to the sum over pairs

$$(P, \mu), \qquad\qquad P_1 \subset P \subset P_2,\ \mu \in M_{P_1}(\mathbb{Q}) \cap \mathfrak{o},$$

of the product of the sign $(-1)^{\dim(A_P/A_G)}$ with the expression

$$\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} \sum_{\eta \in N_P(\mathbb{Q})_{\mu_s}\backslash N_P(\mathbb{Q})} \Big(\int_{N_P(\mathbb{A})_{\mu_s}} f(x^{-1}n_1^{-1}\eta^{-1}\mu n \eta n_1 x)\mathrm{d}n\Big)\mathrm{d}n_1.$$

If we replace the variable $n_1$ by $\nu n_1$, and then integrate over $\nu$ in $N_P(\mathbb{Q})\backslash N_P(\mathbb{A})$, we can change the sum over $\eta$ to an integral over $\nu$ in $N_P(\mathbb{Q})_{\mu_s}\backslash N_P(\mathbb{A})$. Since the resulting integrand is invariant under left translation of $\nu$ by elements in the larger group $N_P(\mathbb{A})_{\mu_s}$, we can in fact integrate $\nu$ over $N_P(\mathbb{A})_{\mu_s}\backslash N_P(\mathbb{A})$. We can thus change the sum of $\eta$ in the expression to an adelic integral over $\nu$. Applying Lemma 10.2 to the resulting double integral over $\nu$ and $n$, we see that the expression equals

$$\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} \int_{N_P(\mathbb{A})} f(x^{-1}n_1^{-1}\mu n n_1 x)\mathrm{d}n\mathrm{d}n_1.$$

The signed sum over $(P, \mu)$ of this last expression equals

$$\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} k_{P_1,P_2,\mathfrak{o}}(n_1 x)\mathrm{d}n_1.$$

We conclude that

$$\begin{aligned}
&\int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \chi^T(x)\widetilde{k}_{P_1,P_2,\mathfrak{o}}(x)\mathrm{d}x \\
&= \int_{M_{P_1}(\mathbb{Q})N_{P_1}(\mathbb{A})\backslash G(\mathbb{A})^1} \chi^T(x)\Big(\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} k_{P_1,P_2,\mathfrak{o}}(n_1 x)\mathrm{d}n_1\Big)\mathrm{d}x \\
&= \int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \chi^T(x) k_{P_1,P_2,\mathfrak{o}}(x)\mathrm{d}x.
\end{aligned}$$

We have shown that the summands corresponding to $P_1 \subset P_2$ in the two identities are equal. It follows that

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \widetilde{k}^T_{\mathfrak{o}}(x)\mathrm{d}x = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k^T_{\mathfrak{o}}(x)\mathrm{d}x = J^T_{\mathfrak{o}}(f).$$

This is the second assertion of Theorem 11.1. □

The formula (11.2) for $J^T_{\mathfrak{o}}(f)$ is better suited to computation. As an example, we consider the special case that the class $\mathfrak{o} \in \mathcal{O}$ consists entirely of semisimple elements. Then $\mathfrak{o}$ is a semisimple conjugacy class in $G(\mathbb{Q})$, and for any element $\gamma \in \mathfrak{o}$, the centralizer $G(\mathbb{Q})_\gamma$ of $\gamma = \gamma_s$ contains no nontrivial unipotent elements. In particular, the group $N_P(\mathbb{Q})_{\gamma_s} = N_P(\mathbb{Q})_\gamma$ attached to any $P$ is trivial. It follows that

$$\widetilde{K}_{P,\mathfrak{o}}(x,x) = \sum_{\gamma \in M_P(\mathbb{Q})\cap\mathfrak{o}} \sum_{\eta \in N_P(\mathbb{Q})} f(x^{-1}\eta^{-1}\gamma\eta x).$$

To proceed, we need to characterize the intersection $M_P(\mathbb{Q}) \cap \mathfrak{o}$.

In §7, we introduced the Weyl set $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})$ attached to any pair of standard parabolic subgroups $P_1$ and $P_1'$. Suppose that $P_1$ is fixed. If $P$ is any other standard parabolic subgroup, we define $W(P_1; P)$ to be the set of elements $s$ in the union over $P_1' \subset P$ of the sets $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})$ such that $s^{-1}\alpha > 0$ for every root $\alpha$ in the subset $\Delta^P_{P_1'}$ of $\Delta_{P_1'}$. In other words, $s^{-1}\alpha$ belongs to the set $\Phi_{P_1}$ for every such $\alpha$. Suppose for example that $G = GL(n)$, and that $P_1$ corresponds to the partition $(\nu_1, \ldots, \nu_{p_1})$ of $n$. We noted in §7 that each of the sets $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})$ is identified with a subset of the symmetric group $S_{p_1}$. The union over $P_1'$ of these sets is identified with the full group $S_{p_1}$. If $P$ corresponds to the partition $(n_1, \ldots, n_p)$ of $n$, $W(P_1; P)$ becomes the set of elements $s \in S_{p_1}$ such that $(\nu_{s(1)}, \ldots, \nu_{s(p_1)})$ is finer than $(n_1, \ldots, n_p)$, and such that $s^{-1}(i) < s^{-1}(i+1)$, for any $i$ that is not of the form $n_1 + \cdots + n_k$ for some $k$.

The problem is simpler if we impose a second condition on $\mathfrak{o}$. Suppose that $(P_1, \alpha_1)$ represents the anisotropic rational datum attached to $\mathfrak{o}$ in the last section, and that $\gamma_1$ belongs to the anisotropic conjugacy class $\alpha_1$ in $M_{P_1}(\mathbb{Q})$. Then $\gamma_1$ represents the semisimple conjugacy class in $\mathfrak{o}$. We know that the group $H$, obtained by taking the connected component of 1 in the centralizer of $\gamma_1$ in $G$, is contained in $M_{P_1}$. For $H$ would otherwise have a proper parabolic subgroup over $\mathbb{Q}$, and $H(\mathbb{Q})$ would contain a nontrivial unipotent element, contradicting the condition that $\mathfrak{o}$ consist entirely of semisimple elements. The group $H(\mathbb{Q})$ is of finite index in $G(\mathbb{Q})_\gamma$. We shall say that $\mathfrak{o}$ is *unramified* if $G(\mathbb{Q})_\gamma$ is also contained in $M_{P_1}$. This is equivalent to asking that the stabilizer of the conjugacy class $\alpha_1$ in $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1})$ be equal to $\{1\}$. In the case $G = GL(n)$, the condition is automatically satisfied, since any centralizer is connected.

Assume that $\mathfrak{o}$ is unramified, and that $(P_1, \alpha_1)$ and $\gamma_1 \in \alpha_1$ are fixed as above. The condition that $\mathfrak{o}$ be unramified implies that if $(P_1', \alpha_1')$ is any other representative of the anisotropic rational datum of $\mathfrak{o}$, there is a unique element in $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})$ that maps $\alpha_1$ to $\alpha_1'$. Suppose that $P$ is any standard parabolic subgroup and that $\gamma$ is an element in $M_P \cap \mathfrak{o}$. It follows easily from this discussion that $\gamma$ can be expressed uniquely in the form

$$\gamma = \mu^{-1} w_s \gamma_1 w_s^{-1} \mu, \qquad s \in W(P_1; P),\ \mu \in M_P(\mathbb{Q})_{w_s\gamma_1 w_s^{-1}} \backslash M_P(\mathbb{Q}),$$

where as usual,

$$M_P(\mathbb{Q})_{w_s\gamma_1 w_s^{-1}} = M_{P, w_s\gamma_1 w_s^{-1}}(\mathbb{Q})$$

is the centralizer of $w_s\gamma_1 w_s^{-1}$ in $M_P(\mathbb{Q})$. (See [**A3**, p. 950].)

Having characterized the intersection $M_P(\mathbb{Q}) \cap \mathfrak{o}$, we can write

$$\begin{aligned}
&\widetilde{K}_{P,\mathfrak{o}}(x,x) \\
&= \sum_{s \in W(P_1;P)} \sum_{\mu} \sum_{\eta \in N_P(\mathbb{Q})} f(x^{-1}\eta^{-1}\mu^{-1} w_s\gamma_1 w_s^{-1}\mu\eta x) \\
&= \sum_{s} \sum_{\pi} f(x^{-1}\pi^{-1} w_s \gamma_1 w_s^{-1} \pi x),
\end{aligned}$$

where $\mu$ and $\pi$ are summed over the right cosets of $M_P(\mathbb{Q})_{w_s\gamma w_s^{-1}}$ in $M_P(\mathbb{Q})$ and $P(\mathbb{Q})$ respectively. Therefore $\widetilde{k}_{\mathfrak{o}}^T(x)$ equals the expression

$$\begin{aligned}
&\sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} \widetilde{K}_{P,\mathfrak{o}}(\delta x, \delta x) \widehat{\tau}_P\big(H_P(\delta x) - T\big) \\
&= \sum_P (-1)^{\dim(A_P/A_G)} \sum_{s \in W(P_1;P)} \sum_{\delta} f(x^{-1}\delta^{-1} w_s\gamma_1 w_s^{-1}\delta x) \widehat{\tau}_P\big(H_P(\delta x) - T\big),
\end{aligned}$$

where $\delta$ is summed over the right cosets of $M_P(\mathbb{Q})_{w_s\gamma_1 w_s^{-1}}$ in $G(\mathbb{Q})$. Set $\delta_1 = w_s^{-1}\delta$. Since

$$w_s^{-1}\big(M_P(\mathbb{Q})_{w_s\gamma_1 w_s^{-1}}\big) w_s = G(\mathbb{Q})_{\gamma_1} = M_{P_1}(\mathbb{Q})_{\gamma_1},$$

we obtain

$$\begin{aligned}
&\widetilde{k}_{\mathfrak{o}}^T(x) \\
&= \sum_P (-1)^{\dim(A_P/A_G)} \sum_{s \in W(P_1;P)} \sum_{\delta_1} f(x^{-1}\delta_1^{-1}\gamma_1\delta_1 x) \widehat{\tau}_P\big(H_P(w_s\delta_1 x) - T\big) \\
&= \sum_{\delta_1} f(x^{-1}\delta_1^{-1}\gamma_1\delta_1 x) \psi^T(\delta_1 x),
\end{aligned}$$

where $\delta_1$ is summed over right cosets of $M_{P_1}(\mathbb{Q})_{\gamma_1}$ in $G(\mathbb{Q})$, and

$$\begin{aligned}
\psi^T(y) = \psi^T_{P_1}(y) &= \sum_P (-1)^{\dim(A_P/A_G)} \sum_{s \in W(P_1;P)} \widehat{\tau}_P\big(H_P(w_s y) - T\big) \\
&= \sum_{P_1'} \sum_{s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})} \sum_{\{P : s \in W(P_1;P)\}} (-1)^{\dim(A_P/A_G)} \widehat{\tau}_P\big(H_{P_1'}(w_s y) - T\big).
\end{aligned}$$

Therefore

$$\begin{aligned} J_{\mathfrak{o}}^T(f) &= \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \widetilde{k}_{\mathfrak{o}}^T(x,f)\mathrm{d}x \\ &= \int_{M_{P_1}(\mathbb{Q})_{\gamma_1}\backslash G(\mathbb{A})^1} f(x^{-1}\gamma x)\psi^T(x)\mathrm{d}x. \end{aligned}$$

The convergence of the second integral follows from the convergence of the first integral (Theorem 11.1), and the fact (implied by Lemma 11.2 below) that the function $\chi_T$ is nonnegative.

We can write

$$M_{P_1}(\mathbb{Q})_{\gamma_1}\backslash G(\mathbb{A})^1 \cong \big(M_{P_1}(\mathbb{Q})_{\gamma_1}\backslash M_{P_1}(\mathbb{A})^1_{\gamma_1}\big) \times \big(M_{P_1}(\mathbb{A})^1_{\gamma_1}\backslash G(\mathbb{A})^1\big),$$

where $M_{P_1}(\mathbb{A})^1_{\gamma_1}$ is the centralizer of $\gamma_1$ in the group $M_{P_1}(\mathbb{A})^1$. Since the centralizer of $\gamma_1$ in $M_{P_1}(\mathbb{A})$ equals its centralizer $G(\mathbb{A})_{\gamma_1}$ in $G(\mathbb{A})$, we can also write

$$M_{P_1}(\mathbb{A})^1_{\gamma_1}\backslash G(\mathbb{A})^1 \cong \big(A_{P_1}(\mathbb{R})^0 \cap G(\mathbb{R})^1\big) \times \big(G(\mathbb{A})_{\gamma_1}\backslash G(\mathbb{A})\big).$$

In the formula for $J_{\mathfrak{o}}^T(f)$ we have just obtained, we are therefore free to decompose the variable of integration as

$$x = may, \quad m \in M_{P_1}(\mathbb{Q})_{\gamma_1}\backslash M_{P_1}(\mathbb{A})^1_{\gamma_1},\ a \in A_{P_1}(\mathbb{R})^0 \cap G(\mathbb{R})^1,\ y \in G(\mathbb{A})_{\gamma_1}\backslash G(\mathbb{A}).$$

Then $f(x^{-1}\gamma_1 x) = f(y^{-1}\gamma_1 y)$ and $\psi^T(x) = \psi^T(ay)$. Therefore

$$J_{\mathfrak{o}}^T(f) = \mathrm{vol}\big(M_{P_1}(\mathbb{Q})_{\gamma_1}\backslash M_{P_1}(\mathbb{A})^1_{\gamma_1}\big) \int_{G(\mathbb{A})_{\gamma_1}\backslash G(\mathbb{A})} f(y^{-1}\gamma_1 y) v_{P_1}^T(y)\mathrm{d}y, \tag{11.5}$$

where

$$v_{P_1}^T(y) = \int_{A_{P_1}(\mathbb{R})^0\cap G(\mathbb{R})^1} \psi^T(ay)\mathrm{d}a = \int_{\mathfrak{a}_{P_1}^G} \psi^T(\exp H \cdot y)\mathrm{d}H.$$

It remains to evaluate the function $v_{P_1}^T(y)$.

For any parabolic subgroup $Q \supset P_0$ and any point $\Lambda \in \mathfrak{a}_Q^*$, define $\varepsilon_Q(\Lambda)$ to be the sign $+1$ or $-1$ according to whether the number of roots $\alpha \in \Delta_Q$ with $\Lambda(\alpha^\vee) \leq 0$ is even or odd. Let

$$H \longrightarrow \phi_Q(\Lambda, H), \qquad\qquad H \in \mathfrak{a}_Q,$$

be the characteristic function of the set of $H$ such that for any $\alpha \in \Delta_Q$, $\varpi_\alpha(H) > 0$ if $\Lambda(\alpha^\vee) \leq 0$, and $\varpi_\alpha(H) \leq 0$ if $\Lambda(\alpha^\vee) > 0$. These functions were introduced by Langlands [**Lan1**], and are useful for studying certain convex polytopes. We apply them to the discussion above by taking $Q = P_1'$ and $\Lambda = s\Lambda_1$, for an element $s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})$ and a point $\Lambda_1$ in the chamber

$$(\mathfrak{a}_{P_1}^*)^+ = \big\{\Lambda_1 \in \mathfrak{a}_{P_1}^* :\ \Lambda_1(\alpha^\vee) > 0,\ \alpha \in \Delta_{P_1}\big\}.$$

Suppose that $s$ belongs to any one of the sets $W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})$. We claim that for any point $H' \in \mathfrak{a}_{P_1'}$, the expression

$$\sum_{\{P: s\in W(P_1;P)\}} (-1)^{\dim(A_P/A_G)}\widehat{\tau}_P(H') \tag{11.6}$$

that occurs in the definition of $\psi^T(y)$ equals

$$\varepsilon_{P_1'}(s\Lambda_1)\phi_{P_1'}(s\Lambda_1, H'), \qquad\qquad \Lambda_1 \in (\mathfrak{a}_{P_1}^*)^+. \tag{11.7}$$

To see this, define a parabolic subgroup $P^s \supset P_1'$ by setting

$$\Delta_{P_1'}^{P^s} = \{\alpha \in \Delta_{P_1'} :\ s^{-1}\alpha > 0\}.$$

The element $s$ then lies in $W(P_1; P)$ if and only if $P_1' \subset P \subset P^s$. The expression (11.6) therefore equals

$$\sum_{\{P: P_1' \subset P \subset P^s\}} (-1)^{\dim(A_P/A_G)} \widehat{\tau}_P(H').$$

If we write the projection of $H'$ onto $\mathfrak{a}_{P_1'}^G$ in the form

$$\sum_{\alpha} c_\alpha \alpha^\vee, \qquad \alpha \in \Delta_{P_1'},\ c_\alpha \in \mathbb{R},$$

we can apply (6.3) to the alternating sum over $P$. We see that the expression equals the sign $\varepsilon_{P_1'}(s\Lambda_1)$ if $H'$ lies in the support of the function $\phi_{P_1'}(s\Lambda_1, H')$, and vanishes otherwise. The claim is therefore valid.

The function $\psi^T(\exp H \cdot y)$ equals

$$\sum_{P_1'} \sum_{s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})} \sum_{\{P: s \in W(P_1; P)\}} (-1)^{\dim(A_P/A_G)} \widehat{\tau}_P\big(sH + H_{P_1'}(w_s y) - T_{P_1'}\big),$$

where $T_{P_1'}$ is the projection of $T$ onto $\mathfrak{a}_{P_1'}$. This in turn equals

$$\sum_{P_1'} \sum_{s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'})} \varepsilon_{P_1'}(s\Lambda_1) \phi_{P_1'}\big(s\Lambda_1, sH + H_{P_1'}(w_s y) - T_{P_1'}\big), \tag{11.8}$$

by what we have just established. Now as a function $H \in \mathfrak{a}_{P_1}^G$, (11.8) would appear to be complicated. It is not! One shows in fact that (11.8) equals the characteristic function of the projection onto $\mathfrak{a}_{P_1}^G$ of the convex hull of

$$\big\{Y_s = s^{-1}\big(T_{P_1'} - H_{P_1'}(w_s y)\big) :\ s \in W(\mathfrak{a}_{P_1}, \mathfrak{a}_{P_1'}),\ P_1' \supset P_0\big\}.$$

The proof of this fact [**A1**, Lemma 3.2] uses elementary properties of convex hulls and a combinatorial lemma of Langlands [**A1**, §2]. We shall discuss it in greater generality later, in §17. In the meantime, we shall illustrate the property geometrically in the special case that $G = SL(3)$.

Assume for the moment then that $G = SL(3)$ and $P_1 = P_0$. In this case, the signed sum of characteristic functions

$$\phi_{P_1'}\big(s\Lambda_1, sH + H_{P_1'}(w_s y) - T\big) = \phi_{P_1'}\big(s\Lambda_1, s(H - Y_s)\big), \qquad H \in \mathfrak{a}_{P_1} = \mathfrak{a}_{P_1}^G,$$

is over elements $s$ parametrized by the symmetric group $S_3$. We have of course the simple roots $\Delta_{P_1} = \{\alpha_1, \alpha_2\}$, and the basis $\{\alpha_1^\vee, \alpha_2^\vee\}$ of $\mathfrak{a}_{P_1}$ dual to $\widehat{\Delta}_{P_1}$. Writing

$$s(H - Y_s) = t_1\alpha_1^\vee + t_2\alpha_2^\vee, \qquad t_i \in \mathbb{R},$$

we see that $\phi_{P_1'}\big(s\Lambda_1, s(H - Y_s)\big)$ is the characteristic function of the affine cone

$$\big\{H = Y_s + t_1 s^{-1}(\alpha_1^\vee) + t_2 s^{-1}(\alpha_2^\vee) :\ t_i > 0 \text{ if } s^{-1}(\alpha_i) < 0;\ t_i \leq 0 \text{ if } s^{-1}(\alpha_i) > 0\big\}.$$

In Figure 11.1, we plot the six vertices $\{Y_s\}$, the associated six cones, and the signs

$$\varepsilon_{P_1'}(s\Lambda_1) = (-1)^{|\{i: s^{-1}(\alpha_i) < 0\}|}, \qquad s \in S_3,$$

by which the corresponding characteristic functions have to be multiplied. We then observe that the signs cancel in every region of the plane except the convex hull of the set of vertices.

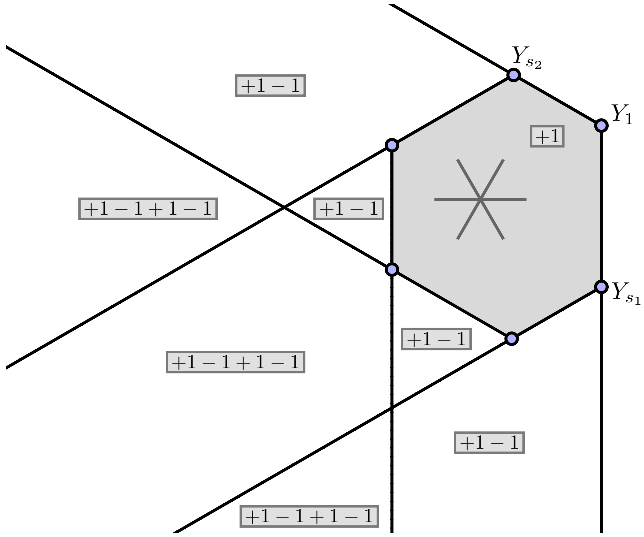



FIGURE 11.1. The shaded region is the convex hull of six points $\{Y_s\}$ in the two dimensional vector space $\mathfrak{a}_0$ attached to $SL(3)$. It is a signed sum of six cones, with vertices at each of the six points.

Returning to the general case, we take for granted the assertion that (11.8) is equal to the characteristic function of the convex hull. Then $v^T_{P_1}(y)$ equals the volume of the given convex hull. In particular, the manipulations used to derive the formula (11.5) for $J^T_{\mathfrak{o}}(f)$ are justified. Observe that

$$\begin{aligned} Y_s &= s^{-1}\big(T_{P_1'} - H_{P_1'}(w_s y)\big) \\ &= s^{-1}\big(T_{P_1'} - H_{P_1'}(\widetilde{w}_s y) - H_{P_1'}(w_s \widetilde{w}_s^{-1})\big) \\ &= s^{-1}\big(T_{P_1'} - H_{P_1'}(\widetilde{w}_s y) - (T_0)_{P_1'} + s(T_0)_{P_1}\big). \end{aligned}$$

When $T = T_0$, the point $Y_s$ equals

$$-s^{-1}H_{P_1'}(\widetilde{w}_s y) + (T_0)_{P_1}.$$

The point $(T_0)_{P_1}$ is independent of $s$, and consequently represents a fixed translate of the convex hull. Since it has no effect on the volume, it may be removed from consideration.

We have established the following result, which we state with $P$ and $\gamma$ in place of $P_1$ and $\gamma_1$.

THEOREM 11.2. *Suppose that $\mathfrak{o} \in \mathcal{O}$ is an unramified class, with anisotropic rational datum represented by a pair $(P, \alpha)$. Then*

$$J_{\mathfrak{o}}(f) = \mathrm{vol}\big(M_P(\mathbb{Q})_\gamma \backslash M_P(\mathbb{A})^1_\gamma\big) \int_{G(\mathbb{A})_\gamma \backslash G(\mathbb{A})} f(x^{-1}\gamma x) v_P(x) \mathrm{d}x, \tag{11.9}$$

*where $\gamma$ is any element in the $M_P(\mathbb{Q})$-conjugacy class $\alpha$, and $v_P(x)$ is the volume of the projection onto $\mathfrak{a}_P^G$ of the convex hull of*

$$\big\{ -s^{-1}H_{P'}(\widetilde{w}_s x) : \ s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'}), \ P' \supset P_0 \big\}.$$

□

## 12. Cuspidal automorphic data

We shall temporarily put aside the finer analysis of the geometric expansion in order to develop the spectral side. We are looking for spectral analogues of the geometric results we have already obtained. In this section, we introduce a set $\mathfrak{X}$ that will serve as the analogue of the set $\mathcal{O}$ of §10. Its existence is a basic consequence of Langlands' theory of Eisenstein series.

A function $\phi$ in $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ is said to be *cuspidal* if

$$\int_{N_P(\mathbb{A})} \phi(nx)\mathrm{d}n = 0, \tag{12.1}$$

for every $P \neq G$ and almost every $x \in G(\mathbb{A})^1$. This condition is a general analogue of the vanishing of the constant term of a classical modular form, which characterizes space of cusp forms. The subspace $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ of cuspidal functions in $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ is closed and invariant under right translation by $G(\mathbb{A})^1$. The following property of this subspace is one of the foundations of the subject.

Theorem 12.1 (Gelfand, Piatetski-Shapiro). *The space $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ decomposes under the action of $G(\mathbb{A})^1$ into a discrete sum of irreducible representations with finite multiplicities. In particular, $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ is a subspace of $L^2_{\mathrm{disc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$.*

The proof is similar to that of the discreteness of the decomposition of $R$, in the case of compact quotient. For if $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is compact, there are no proper parabolic subgroups, by the criterion of Borel and Harish-Chandra, and $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ equals $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$. In general, one combines the vanishing condition (12.1) with the approximate fundamental domain of Theorem 8.1 to show that for any $f \in C_c^\infty\big(G(\mathbb{A})^1\big)$, the restriction $R_{\mathrm{cusp}}(f)$ of $R(f)$ to $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ is of Hilbert-Schmidt class. In particular, if $f(x) = \overline{f(x^{-1})}$, $R_{\mathrm{cusp}}(f)$ is a compact self-adjoint operator. One then uses the spectral theorem to show that $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ decomposes discretely. See [**Lan5**] and [**Har4**]. □

The theorem provides a $G(\mathbb{A})^1$-invariant orthogonal decomposition

$$L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big) = \bigoplus_\sigma L^2_{\mathrm{cusp},\sigma}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big),$$

where $\sigma$ ranges over irreducible unitary representations of $G(\mathbb{A})^1$, and $L^2_{\mathrm{cusp},\sigma}\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ is $G(\mathbb{A})^1$-isomorphic to a finite number of copies of $\sigma$. We define a *cuspidal automorphic datum* to be an equivalence class of pairs $(P,\sigma)$, where $P \subset G$ is a standard parabolic subgroup of $G$, and $\sigma$ is an irreducible representation of $M_P(\mathbb{A})^1$ such that the space $L^2_{\mathrm{cusp},\sigma}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$ is nonzero. The equivalence relation is defined by the conjugacy, which for standard parabolic groups is given by the Weyl sets $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$. In other words, $(P',\sigma')$ is equivalent to $(P,\sigma)$ if there is an element $s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ such that the representation

$$s^{-1}\sigma' :\ m \ \longrightarrow\ \sigma'(w_s m w_s^{-1}), \qquad\qquad m \in M_P(\mathbb{A})^1,$$

of $M_P(\mathbb{A})^1$ is equivalent to $\sigma$. We write $\mathfrak{X} = \mathfrak{X}^G$ for the set of cuspidal automorphic data $\chi = \{(P,\sigma)\}$.

Cuspidal functions do not appear explicitly in Theorem 7.2, but they are an essential ingredient of Langlands's proof. For example, they give rise to a decomposition

$$L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big) = \bigoplus_{\mathcal{P}} L^2_{\mathcal{P}\text{-cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big), \tag{12.2}$$

which is based on cuspidal automorphic data, and is more elementary than the spectral decomposition (7.5). Let us describe it.

For any $P$, we have defined the right $G(\mathbb{A})$-invariant Hilbert space $\mathcal{H}_P$ of functions on $G(\mathbb{A})$, and the dense subspace $\mathcal{H}^0_P$. Let $\mathcal{H}_{P,\mathrm{cusp}}$ be the subspace of vectors $\phi \in \mathcal{H}_P$ such that for almost all $x \in G(\mathbb{A})$, the function $\phi_x(m) = \phi(mx)$ on $M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1$ lies in the space $L^2_{\mathrm{cusp}}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$. Then

$$\mathcal{H}_{P,\mathrm{cusp}} = \bigoplus_{\sigma} \mathcal{H}_{\mathcal{P},\mathrm{cusp},\sigma},$$

where for any irreducible unitary representation $\sigma$ of $M_P(\mathbb{A})^1$, $\mathcal{H}_{P,\mathrm{cusp},\sigma}$ is the subspace of vectors $\phi \in \mathcal{H}_{\mathcal{P},\mathrm{cusp}}$ such that each of the functions $\phi_x$ lies in the space $L^2_{\mathrm{cusp},\sigma}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$. We write $\mathcal{H}^0_{P,\mathrm{cusp}}$ and $\mathcal{H}^0_{\mathcal{P},\mathrm{cusp},\sigma}$ for the respective intersections of $\mathcal{H}_{P,\mathrm{cusp}}$ and $\mathcal{H}_{P,\mathrm{cusp},\sigma}$ with $\mathcal{H}^0_P$.

Suppose that $\Psi(\lambda)$ is an entire function of $\lambda \in \mathfrak{a}^*_{P,\mathbb{C}}$ of Paley-Wiener type, with values in a finite dimensional subspace of functions $x \to \Psi(\lambda, x)$ in $\mathcal{H}^0_{P,\mathrm{cusp},\sigma}$. Then $\Psi(\lambda, x)$ is the Fourier transform in $\lambda$ of a smooth, compactly supported function on $\mathfrak{a}_P$. This means that for any point $\Lambda \in \mathfrak{a}^*_P$, the function

$$\psi(x) = \int_{\Lambda + i\mathfrak{a}^*_P} \mathrm{e}^{(\lambda+\rho_P)(H_P(x))}\Psi(\lambda, x)\mathrm{d}\lambda$$

of $x \in N_P(\mathbb{A})M_P(\mathbb{Q})\backslash G(\mathbb{A})$ is compactly supported in $H_P(x)$.

Lemma 12.2 (Langlands). *The function*

$$(E\psi)(x) = \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} \psi(\delta x), \qquad x \in G(\mathbb{Q})\backslash G(\mathbb{A}),$$

*lies in* $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$.

Lemma 12.3 (Langlands). *Suppose that $\Psi'(\lambda', x)$ is a second such function, attached to a pair $(P', \sigma')$. Then the inner product formula*

$$(E\psi, E\psi') = \int_{\Lambda + i\mathfrak{a}^*_P} \sum_{s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'})} \big(M(s,\lambda)\Psi(\lambda), \Psi'(-s\overline{\lambda})\big)\mathrm{d}\lambda \tag{12.3}$$

*holds if $\Lambda$ is any point in $\mathfrak{a}^*_P$ such that $(\Lambda - \rho_P)(\alpha^\vee) > 0$ for every $\alpha \in \Delta_P$.*

If $\chi$ is the class in $\mathfrak{X}$ represented by a pair $(P, \sigma)$, let $L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ be the closed, $G(\mathbb{A})$-invariant subspace of $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ generated by the functions $E\psi$ attached to $(P, \sigma)$.

Lemma 12.4 (Langlands). *There is an orthogonal decomposition*

$$L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big) = \bigoplus_{\chi \in \mathfrak{X}} L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big). \tag{12.4}$$

Lemmas 12.2–12.4 are discussed in the early part of Langlands's survey article [**Lan1**]. They are the foundations for the rest of the theory, and for Theorem 7.2 in particular. We refer the reader to [**Lan1**] for brief remarks on the proofs, which are relatively elementary. □

The inner product formula (12.3) is especially important. It is used in the proof of both the analytic continuation (a) and the spectral decomposition (b) in Theorem 7.2. Observe that the domain of integration in (12.3) is contained in the region of absolute convergence of the cuspidal operator valued function $M(s,\lambda)$ in the integrand. Once he had proved the meromorphic continuation of this function, Langlands was able to use (12.3) to establish the remaining analytic continuation assertions of Theorem 7.2(a), and the spectral decomposition of the space $L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$. His method was based on a change contour of integration from $\Lambda + i\mathfrak{a}^*_P$ to $i\mathfrak{a}^*_P$, and an elaborate analysis of the resulting residues. It was a tour de force, the details of which comprise the notoriously difficult Chapter 7 of [**Lan5**].

Any class $\chi = \{(\mathcal{P},\sigma)\}$ in $\mathfrak{X}$ determines an associated class $\mathcal{P}_\chi = \{P\}$ of standard parabolic subgroups. We then obtain a decomposition (12.2) from (12.4) by setting

$$L^2_{\mathcal{P}\text{-cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big) = \bigoplus_{\{\chi\in\mathfrak{X}:\mathcal{P}_\chi=\mathcal{P}\}} L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big).$$

However, it is the finer decomposition (12.4) that is more often used. We shall actually apply the obvious variant of (12.4) that holds for $G(\mathbb{A})^1$ in place of $G(\mathbb{A})$, or rather its restriction

$$L^2_{\mathrm{disc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big) = \bigoplus_{\chi\in\mathfrak{X}} L^2_{\mathrm{disc},\chi}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big) \tag{12.5}$$

to the discrete spectrum, in which

$$L^2_{\mathrm{disc},\chi}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big) = L^2_{\mathrm{disc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big) \cap L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big).$$

If $P$ is a standard parabolic subgroup, the correspondence

$$(P_1\cap M_P,\sigma_1) \longrightarrow (P_1,\sigma_1), \qquad\qquad P_1\subset P,\ \{(P_1\cap M_P,\sigma_1)\}\in\mathfrak{X}^{M_P},$$

yields a mapping $\chi_P\to\chi$ from $\mathfrak{X}^{M_P}$ to the set $\mathfrak{X}=\mathfrak{X}^G$. We can then write

$$L^2_{\mathrm{disc}}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big) = \bigoplus_{\chi\in\mathfrak{X}} L^2_{\mathrm{disc},\chi}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big),$$

where $L^2_{\mathrm{disc},\chi}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$ is the sum of those subspaces of $L^2_{\mathrm{disc}}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$ attached to classes $\chi_P\in\mathfrak{X}^{M_P}$ in the fibre of $\chi$. Let $\mathcal{H}_{P,\chi}$ be the subspace of functions $\phi$ in the Hilbert space $\mathcal{H}_P$ such that for almost all $x$, the function $\phi_x(m)=\phi(mx)$ lies in $L^2_{\mathrm{disc},\chi}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$. There is then an orthogonal direct sum

$$\mathcal{H}_P = \bigoplus_\chi \mathcal{H}_{\mathcal{P},\chi}.$$

There is also an algebraic direct sum

$$\mathcal{H}^0_P = \bigoplus_\chi \mathcal{H}^0_{P,\chi}, \tag{12.6}$$

where $\mathcal{H}^0_{P,\chi}$ is the intersection of $\mathcal{H}_{P,\chi}$ with $\mathcal{H}^0_P$. For any $\lambda$ and $f$, we shall write $\mathcal{I}_{P,\chi}(\lambda,f)$ for the restriction of the operator $\mathcal{I}_P(\lambda,f)$ to the invariant subspace $\mathcal{H}_{P,\chi}$ of $\mathcal{H}_P$.

At the end of §7, we described the spectral expansions for both the kernel $K(x,y)$ and the truncated function $k^T(x)$ in terms of Eisenstein series. They were defined by means of an orthonormal basis $\mathcal{B}_P$ of $\mathcal{H}_P$. We can assume that $\mathcal{B}_P$ is compatible with the algebraic direct sum (12.6). In other words,

$$\mathcal{B}_P = \coprod_{\chi\in\mathfrak{X}} \mathcal{B}_{\mathcal{P},\chi},$$

where $\mathcal{B}_{\mathcal{P},\chi}$ is the intersection of $\mathcal{B}_P$ with $\mathcal{H}^0_{\mathcal{P},\chi}$. For any $\chi \in \mathfrak{X}$ we set

$$K_\chi(x,y) = \sum_P n_P^{-1} \int_{i\mathfrak{a}_P^*} \sum_{\phi\in\mathcal{B}_{P,\chi}} E\big(x, \mathcal{I}_{P,\chi}(\lambda, f)\phi, \lambda\big)\overline{E(y,\phi,\lambda)}\mathrm{d}\lambda, \tag{12.7}$$

where $n_P$ is the integer defined in Theorem 7.2(b). It is a consequence of Langlands' construction of the spectral decomposition (7.5) from the more elementary decomposition (12.4) that $K_\chi(x,y)$ is the kernel of the restriction of $R(f)$ to the invariant subspace $L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ of $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$. It follows, either from this or from the definition (12.7), that

$$K(x,y) = \sum_{\chi\in\mathfrak{X}} K_\chi(x,y). \tag{12.8}$$

This is the spectral analogue of the geometric decomposition (10.1).

More generally, suppose that we fix $P$, and use $P_1 \subset P$ in place of $P$ to index the orthonormal bases. Then we have

$$K_P(x,y) = \sum_{\chi\in\mathfrak{X}} K_{P,\chi}(x,y),$$

where $K_{P,\chi}(x,y)$ is equal to

$$\sum_{P_1\subset P} (n^P_{P_1})^{-1} \int_{i\mathfrak{a}_P^*} \sum_{\phi\in\mathcal{B}_{P_1,\chi}} E^P_{P_1}\big(x, \mathcal{I}_{P_1,\chi}(\lambda, f)\phi, \lambda\big)\overline{E^P_{P_1}(y,\phi,\lambda)}\mathrm{d}\lambda.$$

We obtain a decomposition

$$k^T(x) = \sum_{\chi\in\mathfrak{X}} k^T_\chi(x), \tag{12.9}$$

where

$$\begin{aligned} k^T_\chi(x) &= k^T_\chi(x,f) \\ &= \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})} K_{P,\chi}(\delta x, \delta x)\widehat{\tau}_P\big(H_P(\delta x) - T\big). \end{aligned}$$

This is the spectral analogue of the geometric decomposition (10.2) of the truncated kernel.

We have given spectral versions of the constructions at the beginning of §10. However, the spectral analogue of the coarse geometric expansion (10.7) is more difficult. The problem is to obtain an analogue of Corollary 10.1. We know from Theorem 6.1 that

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Big|\sum_\chi k^T_\chi(x)\Big|\mathrm{d}x = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} |k^T(x)|\mathrm{d}x < \infty.$$

To obtain a corresponding expansion for $J^T(f)$, we would need the stronger assertion that the double integral

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_{\chi} |k_\chi^T(x)| \mathrm{d}x$$

is finite. Unlike the geometric case of Corollary 10.1, this is not an immediate consequence of the proof of Theorem 6.1. It requires some new methods.

## 13. A truncation operator

The process that assigns the modified function $k^T(x) = k^T(x,f)$ to the original kernel $K(x,x)$ can be regarded as a construction that is based on the adjoint action of $G$ on itself. It is compatible with the geometry of classes $\mathfrak{o} \in \mathcal{O}$. The process is less compatible with the spectral properties of classes $\chi \in \mathfrak{X}$. However, we still have to deal with the spectral expansion (12.9) of $k^T(x)$. We do so by introducing an operator that systematically truncates functions on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$.

The operator depends on the same parameter $T$ used to define $k^T(x)$. It acts on the space $\mathcal{B}_{\mathrm{loc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ of locally bounded, measurable functions on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$. For any suitably regular point $T \in \mathfrak{a}_0^+$ and any function $\phi \in \mathcal{B}_{\mathrm{loc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$, we define $\Lambda^T\phi$ to be the function in $\mathcal{B}_{\mathrm{loc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ whose value at $x$ equals

$$(13.1) \qquad \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(\mathbb{Q})\backslash G(\mathbb{Q})} \int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})} \phi(n\delta x)\widehat{\tau}_P\big(H_P(\delta x) - T\big)\mathrm{d}n.$$

The inner sum may be taken over a finite set (that depends on $x$), while the integrand is a bounded function of $n$. Notice the formal similarity of the definition with that of $k^T(x)$ in §6. Notice also that if $\phi$ belongs to $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$, then $\Lambda^T\phi = \phi$.

There are three basic properties of the operator $\Lambda^T$ to be discussed in this section. The first is that $\Lambda^T$ is an orthogonal projection.

Proposition 13.1. (a) *For any $P_1$, any $\phi_1 \in \mathcal{B}_{\mathrm{loc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$, and any $x_1 \in G(\mathbb{A})^1$, the integral*

$$\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} (\Lambda^T\phi_1)(n_1x_1)\mathrm{d}n_1$$

*vanishes unless $\varpi\big(H_{P_1}(x_1) - T\big) \leq 0$ for every $\varpi \in \widehat{\Delta}_{P_1}$.*

(b) $\Lambda^T \circ \Lambda^T = \Lambda^T$.

(c) *The operator $\Lambda^T$ is self-adjoint, in the sense that it satisfies the inner product formula*

$$(\Lambda^T\phi_1, \phi_2) = (\phi_1, \Lambda^T\phi_2),$$

*for functions $\phi_1 \in \mathcal{B}_{\mathrm{loc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ and $\phi_2 \in C_c\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$.*

The first assertion of the proposition is Lemma 1.1 of [**A4**]. (The symbol $<$ in the statement of this lemma should in fact be $\leq$.) In the case $G = SL(2)$, it follows directly from classical reduction theory, as illustrated in the earlier Figure 8.3. In general, one has to apply the Bruhat decomposition to elements in the sum over $P(\mathbb{Q})\backslash G(\mathbb{Q})$ that occurs in the definition of $\Lambda^T\phi$. We recall that the Bruhat decomposition is a double coset decomposition

$$G(\mathbb{Q}) = \coprod_{s\in W_0} \bigl(B_0(\mathbb{Q})w_s N_0(\mathbb{Q})\bigr)$$

of $G(\mathbb{Q})$, which in turn leads easily to a characterization

$$P(\mathbb{Q})\backslash G(\mathbb{Q}) \cong \coprod_{s\in W_0^M\backslash W_0} \bigl(w_s^{-1}N_0(\mathbb{Q})w_s \cap N_0(\mathbb{Q})\backslash N_0(\mathbb{Q})\bigr)$$

of $P(\mathbb{Q})\backslash G(\mathbb{Q})$. Various manipulations, which we will not reproduce here, reduce the assertion of (i) to Identity 6.2.

The assertion (ii) follows from (i). Indeed, $\bigl(\Lambda^T(\Lambda^T\phi)\bigr)(x)$ equals the sum over $P_1 \supset P_0$ and $\delta_1 \in P_1(\mathbb{Q})\backslash G(\mathbb{Q})$ of

$$\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} (\Lambda^T\phi)(n_1\delta_1 x)\widehat{\tau}_{P_1}\bigl(H_{P_1}(\delta_1 x) - T\bigr)\mathrm{d}n_1.$$

The term corresponding to $P_1 = G$ equals $(\Lambda^T\phi)(x)$, while if $P_1 \neq G$, the term vanishes by (i) and the definition of $\widehat{\tau}_{P_1}$.

To establish (iii), we observe that

$$\begin{aligned}
&(\Lambda^T\phi_1, \phi_2)\\
&= \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})} \cdot\\
&\qquad \cdot \int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})} \phi_1(n\delta x)\widehat{\tau}_P\bigl(H_P(\delta x) - T\bigr)\overline{\phi_2(x)}\mathrm{d}n\mathrm{d}x\\
&= \sum_P (-1)^{\dim(A_P/A_G)} \int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})} \int_{P(\mathbb{Q})\backslash G(\mathbb{A})^1} \phi_1(nx)\overline{\phi_2(x)}\widehat{\tau}_P\bigl(H_P(x) - T\bigr)\mathrm{d}x\mathrm{d}n\\
&= \sum_P (-1)^{\dim(A_P/A_G)} \int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})} \int_{P(\mathbb{Q})\backslash G(\mathbb{A})^1} \phi_1(x)\overline{\phi_2(nx)}\widehat{\tau}_P\bigl(H_P(x) - T\bigr)\mathrm{d}x\mathrm{d}n\\
&= \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_P (-1)^{\dim(A_P/A_G)} \sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})} \cdot\\
&\qquad \cdot \int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})} \phi_1(x)\overline{\phi_2(n\delta x)}\widehat{\tau}_P\bigl(H_P(\delta x) - T\bigr)\mathrm{d}n\mathrm{d}x\\
&= (\phi_1, \Lambda^T\phi_2).
\end{aligned}$$

□

It is not hard to show from (ii) and (iii) that $\Lambda^T$ extends to an orthogonal projection from the space $L^2\bigl(G(\mathbb{Q})\backslash G(\mathbb{A})\bigr)$ to itself. It is also easy to see that $\Lambda^T$ preserves each of the spaces $L^2_{\mathcal{P}\text{-cusp}}\bigl(G(\mathbb{Q})\backslash G(\mathbb{A})^1\bigr)$ in the cuspidal decomposition (12.2). On the other hand, $\Lambda^T$ is decidedly not compatible with the spectral decomposition (7.5). It is an operator built upon the cuspidal properties of §12, rather than the more sensitive spectral properties of Theorem 7.2.

The second property of the operator $\Lambda^T$ is that it transforms uniformly tempered functions to rapidly decreasing functions. To describe this property quantitatively, we need to choose a height function $\|\cdot\|$ on $G(\mathbb{A})$.

Suppose first that $G$ is a general linear group $GL(m)$, and that $x = (x_{ij})$ is a matrix in $GL(m, \mathbb{A})$. We define

$$\|x_v\|_v = \max_{i,j} |x_{ij,v}|_v$$

if $v$ is a $p$-adic valuation, and

$$\|x_v\|_v = \Big(\sum_{i,j} |x_{ij,v}|_v^2\Big)^{\frac{1}{2}}$$

if $v$ is the archimedean valuation. Then $\|x_v\|_v = 1$ for almost all $v$. The height function

$$\|x\| = \prod_v \|x_v\|_v$$

is therefore defined by a finite product. For arbitrary $G$, we fix a $\mathbb{Q}$-rational injection $r\colon G \to GL(m)$, and define

$$\|x\| = \|r(x)\|.$$

By choosing $r$ appropriately, we can assume that the set of points $x \in G(\mathbb{A})$ with $\|x\| \leq t$ is compact, for any $t > 0$. The chosen height function $\|\cdot\|$ on $G(\mathbb{A})$ then satisfies

$$\|xy\| \leq \|x\|\|y\|, \qquad x, y \in G(\mathbb{A}), \tag{13.2}$$

$$\|x^{-1}\| \leq C_0\|x\|^{N_0}, \qquad x \in G(\mathbb{A}), \tag{13.3}$$

and

$$\big|\big\{x \in G(\mathbb{Q}) : \ \|x\| \leq t\big\}\big| \leq C_0 t^{N_0}, \qquad t \geq 0, \tag{13.4}$$

for positive contants $C_0$ and $N_0$. (See [**Bor2**].)

We shall say that a function $\phi$ on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ is *rapidly decreasing* if for any positive integer $N$ and any Siegel set $\mathcal{S} = \mathcal{S}^G(T_1)$ for $G(\mathbb{A})$, there is a positive constant $C$ such that

$$|\phi(x)| \leq C\|x\|^{-N}$$

for every $x$ in $\mathcal{S}^1 = \mathcal{S} \cap G(\mathbb{A})^1$. The notion of uniformly tempered applies to the space of smooth functions

$$C^\infty\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big) = \varinjlim_{K_0} C^\infty\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big).$$

By definition, $C^\infty(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big)$ is the space of functions on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ that are right invariant under the open compact subgroup $K_0$ of $G(\mathbb{A}_{\mathrm{fin}})$, and are infinitely differentiable as functions on the subgroup $G(\mathbb{R})^1 = G(\mathbb{R}) \cap G(\mathbb{A})^1$ of $G(\mathbb{A})^1$. We can of course also define the larger space $C^r\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ of functions of differentiability class $C^r$ in the same way. If $X$ is a left invariant differential operator on $G(\mathbb{R})^1$ of degree $k \leq r$, and $\phi$ lies in $C^r\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big)$, $X\phi$ is a function in $C^{r-k}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big)$. Let us say that a function $\phi \in C^\infty\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ is *uniformly tempered* if there is an $N_0 \geq 0$ with the property that for every left invariant differentiable operator $X$ on $G(\mathbb{R})^1$, there is a constant $c_X$ such that

$$|(X\phi)(x)| \leq c_X\|x\|^{N_0},$$

for every $x \in G(\mathbb{A})^1$.

PROPOSITION 13.2. (a) *If $\phi \in C^\infty\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ is uniformly tempered, the function $\Lambda^T\phi$ is rapidly decreasing.*

(b) *Given a Siegel set $\mathcal{S}$, positive integers $N$ and $N_0$, and an open compact subgroup $K_0$ of $G(\mathbb{A}_{\mathrm{fin}})$, we can choose a finite set $\{X_i\}$ of left invariant differential operators on $G(\mathbb{R})^1$ and a positive integer $r$ with the property that if $(\Omega, \mathrm{d}\omega)$ is a measure space, and $\phi(\omega)$: $x \to \phi(\omega, x)$ is any measurable function from $\Omega$ to $C^r\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big)$, the supremum*

$$\sup_{x\in\mathcal{S}^1}\Big(\|x\|^N\int_\Omega |\Lambda^T\phi(\omega,x)|\mathrm{d}\omega\Big) \tag{13.5}$$

*is bounded by*

$$\sup_{y\in G(\mathbb{A})^1}\Big(\|y\|^{-N_0}\sum_i\int_\Omega |X_i\phi(\omega,y)|\mathrm{d}\omega\Big). \tag{13.6}$$

It is enough to prove (ii), since it is a refined version of (i). This assertion is Lemma 1.4 of [**A4**], the proof of which is reminiscent of that of Theorem 6.1. The initial stages of the two proofs are in fact identical. We multiply the summand corresponding to $P$ in

$$\begin{aligned}&\Lambda^T\phi(\omega,x)\\&=\sum_P(-1)^{\dim(A_P/A_G)}\sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})}\int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})}\phi(\omega,n\delta x)\mathrm{d}n\cdot\widehat{\tau}_P\big(H_P(\delta x)-T\big)\end{aligned}$$

by the left hand side of (8.1). We then apply the definition (8.2) to the product of functions $\tau_{P_1}^P$ and $\widehat{\tau}_P$ that occurs in the resulting expansion. The function $\Lambda^T\phi(\sigma,x)$ becomes the sum over pairs $P_1\subset P_2$ and elements $\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})$ of the product

$$F^{P_1}(\delta x,T)\sigma_{P_1}^{P_2}\big(H_{P_1}(\delta x)-T\big)\phi_{P_1,P_2}(\omega,\delta x),$$

where

$$\phi_{P_1,P_2}(\omega,y)=\sum_{\{P:P_1\subset P\subset P_2\}}(-1)^{\dim(A_P/A_G)}\int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})}\phi(\omega,ny)\mathrm{d}n. \tag{13.7}$$

Suppose that $y=\delta x$ is such that the first two factors in the last product are both nonzero. Replacing $\delta$ by a left $P_1(\mathbb{Q})$-translate, if necessary, we can assume that

$$y=\delta x=n_*n^*mak,$$

for $k\in K$, elements $n_*$, $n^*$ and $m$ in fixed compact subsets of $N_{P_2}(\mathbb{A})$, $N_{P_1}^{P_2}(\mathbb{A})$ and $M_{P_1}(\mathbb{A})^1$ respectively, and a point $a\in A_{P_1}(\mathbb{R})^0$ with $\sigma_{P_1}^{P_2}\big(H_{P_1}(a)-T\big)\neq 0$. Therefore

$$y=\delta x=n_*a\cdot a^{-1}n^*amk=n_*ab,$$

where $b$ belongs to a fixed compact subset of $G(\mathbb{A})^1$ that depends only on $G$. The next step is to extract an estimate of rapid decrease for the function

$$\phi_{P_1,P_2}(\omega,y)=\phi_{P_1,P_2}(\omega,\delta x)=\phi_{P_1,P_2}(\omega,ab)$$

from the alternating sum over $P$ in (13.7).

At this point the argument diverges slightly from that of Theorem 6.1. The quantitative nature of the assertion (ii) represents only a superficial difference, since similar estimates are implicit in the discussion of §8. However, the integrals in (13.7) are over quotients $N_P(\mathbb{Q})\backslash N_P(\mathbb{A})$ rather than groups $N_P(\mathbb{A})$, a reflection of the left $G(\mathbb{Q})$-invariance of the underlying function $y \to \phi(\sigma, y)$. This alters the way we realize the cancellation in the alternating sum over $P$. It entails having to apply the Fourier inversion formula to a product of groups $\mathbb{Q}\backslash\mathbb{A}$, in place of the Poisson summation formula for a product of groups $\mathbb{A}$. The problem is that the quotient $\mathfrak{n}_{P_1}^{P_2}(\mathbb{Q})\backslash\mathfrak{n}_{P_1}^{P_2}(\mathbb{A})$ does not correspond with $N_{P_1}^{P_2}(\mathbb{Q})\backslash N_{P_1}^{P_2}(\mathbb{A})$ under the exponential mapping. However, the problem may be resolved by a straightforward combinatorial argument that appears in [**Har4**, Lemma 11]. One constructs a finite set of pairs

$$(N_I^-, N_I), \qquad\qquad N_{P_2} \subset N_I^- \subset N_I \subset N_{P_1},$$

of $\mathbb{Q}$-rational groups, where $N_I^-$ is normal in $N_I$ with abelian quotient $N^I$. Each index $I$ parametrizes a subset

$$\{\beta_{I,\alpha} \in \Phi_{P_1}^{P_2} :\ \alpha \in \Delta_{P_1}^{P_2}\}$$

of roots of the parabolic subgroup $M_{P_2} \cap P_1$ of $M_{P_2}$ such that $\beta_{I,\alpha}$ contains $\alpha$ in its decomposition into simple roots. If $X_{I,\alpha} \in \mathfrak{n}_{P_1}(\mathbb{Q})$ stands for a root vector relative to $\beta_{I,\alpha}$, the space

$$\mathfrak{n}^I(\mathbb{Q}) = \bigoplus_{\alpha\in\Delta_{P_1}^{P_2}} \mathbb{Q}X_{I,\alpha}$$

becomes a linear complement for the Lie algebra of $N_I^-(\mathbb{Q})$ in that of $N_I(\mathbb{Q})$. The combinatorial argument yields an expansion of $\phi_{P_1,P_2}(\omega, ab)$ as linear combination over $I$ of functions

$$\sum_{\xi\in\mathfrak{n}^I(\mathbb{Q})'} \int_{\mathfrak{n}^I(\mathbb{Q})\backslash\mathfrak{n}^I(\mathbb{A})} \int_{N_I^-(\mathbb{Q})\backslash N_I^-(\mathbb{A})} \phi\big(\omega, u\exp(X)ab\big)\psi\big(\langle X, \xi\rangle\big)\mathrm{d}u\mathrm{d}X, \tag{13.8}$$

where

$$\mathfrak{n}^I(\mathbb{Q})' = \Big\{\xi = \sum_{\alpha\in\Delta_{P_1}^{P_2}} r_\alpha X_{I,\alpha} :\ r_\alpha \in \mathbb{Q}^*\Big\}.$$

(See [**A4**, p. 94].)

One can estimate (13.8) as in the proof of Theorem 6.1. In fact, it is not hard to show that for any positive integer $n$, the product of $\mathrm{e}^{n\|H_{P_1}(a)\|}$ with the integral of the absolute value of (13.8) over $\omega$ has a bound of the form (13.6). But

$$\mathrm{e}^{n\|H_{P_0}(a)\|} \geq c_1\|a\|^{n\varepsilon} \geq c_2\|n^*ab\|^{n\varepsilon} = c_2\|\delta x\|^{n\varepsilon},$$

for positive constants $c_1$, $c_2$ and $\varepsilon$. Moreover, it is known that there is a positive constant $c$ such that

$$\|\delta x\| \geq c\|x\|,$$

for any $x$ in the Siegel set $\mathcal{S}$, and any $\delta \in G(\mathbb{Q})$. It follows that the supremum

$$\sup_{x\in\mathcal{S}} \sup_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})} \Big(\|x\|^{n\varepsilon}\int_\Omega \big|\phi_{P_1,P_2}(\omega,\delta x)\big|\mathrm{d}\omega\Big)$$

has a bound of the form (13.6). Since this supremum is independent of $\delta$, we have only to estimate the sum

$$\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})} F^{P_1}(\delta x, T)\sigma_{P_1}^{P_2}\big(H_{P_1}(\delta x) - T\big).$$

It follows from the definition (8.3) and the fact that both $F^{P_1}(\cdot, T)$ and $\sigma_{P_1}^{P_2}(\cdot)$ are characteristic functions that the summand corresponding to $\delta$ is bounded by

$\widehat{\tau}_{P_1}\big(H_{P_1}(\delta x)-T\big)$. In §6 we invoked Lemma 5.1 of [**A3**] in order to say that the sum over $\delta$ in (6.1) could be taken over a finite set. The lemma actually asserts that

$$\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})} \widehat{\tau}_{P_1}\big(H_{P_1}(\delta x)-T\big)\le c_T\|x\|^{N_1},$$

for positive constants $c_T$ and $N_1$. We obtain an estimate (13.6) for (13.5) by choosing $n\ge\varepsilon^{-1}(N+N_1)$. □

The proof of Proposition 13.2 we have just sketched is that of [**A4**, Lemma 1.4]. The details in [**A4**] are a little hard to follow, thanks to less than perfect exposition and some typographical errors. Perhaps the discussion above will make them easier to read.

The most immediate application of Proposition 13.2 is to an Eisenstein series $x\to E(x,\phi,\lambda)$. Among the many properties established by Langlands in the course of proving Theorem 7.2 was the fact that Eisenstein series are uniformly slowly increasing. More precisely, there is a positive integer $N_0$ such that for any vector $\phi\in\mathcal{H}_P^0$ and any left invariant differential operator $X$ on $G(\mathbb{R})^1$, there is an inequality

$$|XE(x,\phi,\lambda)|\le c_{X,\phi}(\lambda)\|x\|^{N_0},\qquad x\in G(\mathbb{A}),$$

in which $c_{X,\phi}(\lambda)$ is a locally bounded function on the set of $\lambda\in\mathfrak{a}^*_{P,\mathbb{C}}$ at which $E(x,\phi,\lambda)$ is analytic. It follows from Proposition 13.2 that for any $N$ and any Siegel set $\mathcal{S}$, there is a locally bounded function $c_{N,\phi}(\lambda)$ on the set of $\lambda$ at which $E(x,\phi,\lambda)$ is analytic such that

$$|\Lambda^T E(x,\phi,\lambda)|\le c_{N,\phi}(\lambda)\|x\|^{-N}, \tag{13.9}$$

for every $x\in\mathcal{S}^1$. In particular, the truncated Eisenstein series $\Lambda^T E(x,\phi,\lambda)$ is square integrable on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$. As we shall see, the spectral expansion of the trace formula depends on being able to evaluate the inner product of two truncated Eisenstein series.

The third property of the truncation operator is one of cancellation. It concerns the partial truncation operator $\Lambda^{T,P_1}$ attached to a standard parabolic subgroup $P_1\supset P_0$. If $\phi$ is any function in $\mathcal{B}_{\mathrm{loc}}\big(P_1(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$, we define $\Lambda^{T,P_1}\phi$ to be the function in $\mathcal{B}_{\mathrm{loc}}\big(M_{P_1}(\mathbb{Q})N_{P_1}(\mathbb{A})\backslash G(\mathbb{A})^1\big)$ whose value at $x$ equals

$$\sum_{\{Q:P_0\subset Q\subset P_1\}}(-1)^{\dim(A_Q/A_{P_1})}\sum_{\delta\in Q(\mathbb{Q})\backslash P_1(\mathbb{Q})}\int_{N_Q(\mathbb{Q})\backslash N_Q(\mathbb{A})}\phi(n\delta x)\widehat{\tau}_Q^{P_1}\big(H_Q(\delta x)-T\big).$$

Proposition 13.3. *If $\phi$ belongs to $\mathcal{B}_{\mathrm{loc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$, then*

$$\sum_{P_1\supset P_0}\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})}\Lambda^{T,P_1}\phi(\delta x)\tau_{P_1}\big(H_{P_1}(\delta x)-T\big)=\phi(x).$$

*More generally, if $\phi$ belongs to $\mathcal{B}_{\mathrm{loc}}\big(P(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$ for some $P\supset P_0$, the sum*

$$\sum_{\{P_1:P_0\subset P_1\subset P\}}\sum_{\delta\in P_1(\mathbb{Q})\backslash P(\mathbb{Q})}\Lambda^{T,P_1}\phi(\delta x)\tau^P_{P_1}\big(H_{P_1}(\delta x)-T\big) \tag{13.10}$$

*equals*

$$\int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})}\phi(nx)dn.$$

If we substitute the definition of $\Lambda^{T,P_1}\phi$ into (13.10), we obtain a double sum over $Q$ and $P_1$. Combining the double sum over $Q(\mathbb{Q})\backslash P_1(\mathbb{Q})$ and $P_1(\mathbb{Q})\backslash P(\mathbb{Q})$ into a single sum over $Q(\mathbb{Q})\backslash P(\mathbb{Q})$, we write (13.10) as the sum over parabolic subgroups $Q$, with $P_0 \subset Q \subset P$, and elements $\delta \in Q(\mathbb{Q})\backslash P(\mathbb{Q})$ of the product of

$$\int_{N_Q(\mathbb{Q})\backslash N_Q(\mathbb{A})} \phi(n\delta x)$$

with

$$\sum_{\{P_1 : Q \subset P_1 \subset P\}} (-1)^{\dim(A_Q/A_{P_1})} \widehat{\tau}_Q^{P_1}\big(H_Q(\delta x) - T\big)\tau_{P_1}^P\big(H_{P_1}(\delta x) - T\big).$$

Since $\tau_{P_1}^P\big(H_{P_1}(\delta x) - T\big) = \tau_{P_1}^P\big(H_Q(\delta x) - T\big)$, we can apply (8.11) to the alternating sum over $P_1$. This proves that the alternating sum vanishes unless $Q = P$, in which case it is trivially equal to 1. The formula of the lemma follows. (See [**A4**, Lemma 1.5].) □

## 14. The coarse spectral expansion

The truncation operator $\Lambda^T$ acts on functions on $G(\mathbb{Q})\backslash G(\mathbb{A})^1$. If $h$ is a function of two variables and $\Lambda$ is a linear operator on any space of functions in $G(\mathbb{A})$, we write $\Lambda_1 h$ and $\Lambda_2 h$ for the transforms of $h$ obtained by letting $\Lambda$ act separately on the first and second variables respectively. We want to consider the case that $\Lambda = \Lambda^T$, and $h(x, y)$ equals the $\chi$-component $K_\chi(x, y)$ of the kernel $K(x, y)$ of $R(f)$. We recall that the parameter $T$ in both the operator $\Lambda^T$ and the modified kernel $k^T(x)$ is a suitably regular point in $\mathfrak{a}_0^+$.

Theorem 14.1. (a) *The double integral*

$$\sum_{\chi \in \mathfrak{X}} \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda_2^T K_\chi(x, x)\mathrm{d}x \tag{14.1}$$

*converges absolutely.*

(b) *If $T$ is suitably regular, in a sense that depends only on the support of $f$, the double integral*

$$\sum_{\chi \in \mathfrak{X}} \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k_\chi^T(x)\mathrm{d}x \tag{14.2}$$

*also converges absolutely.*

(c) *If $T$ is as in (ii), we have*

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k_\chi^T(x)\mathrm{d}x = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda_2^T K_\chi(x, x)\mathrm{d}x,$$

*for any $\chi \in \mathfrak{X}$.*

The assertions of Theorem 14.1 are among the main results of [**A4**]. Their proof is given in §2 of that paper. We shall try to give some idea of the argument.

The assertion (i) requires a quantitative estimate for the spectral expansion of the kernel

$$K(x, y) = \sum_{\gamma \in G(\mathbb{Q})} f(x^{-1}\gamma y).$$

The sum here can obviously be taken over elements $\gamma$ in the support of the function $u \to f(x^{-1}uy)$. Since the support equals $x \cdot \mathrm{supp} f \cdot y^{-1}$, we can apply the properties (13.2)–(13.4) of the height function $\|\cdot\|$. We see that

$$|K(x,y)| \leq c(f)\|x\|^{N_1}\|y\|^{N_1},$$

for a positive number $N_1$ that depends only on $G$. For any $\chi \in \mathfrak{X}$, $K_\chi(x,y)$ is the kernel of the restriction of $R(f)$ to the invariant subspace $L^2_\chi\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$ of $L^2\big(G(\mathbb{Q})\backslash G(\mathbb{A})\big)$. It follows from the discussion at the end of §7 that the sum

$$\sum_{\chi \in \mathfrak{X}} |K_\chi(x,y)|$$

of absolute values is bounded by a finite sum of products

$$\Big(\sum_\chi K_{\chi,1}(x,x)\Big)^{\frac{1}{2}}\Big(\sum_\chi K_{\chi,2}(y,y)\Big)^{\frac{1}{2}} = \big(K_1(x,x)\big)^{\frac{1}{2}}\big(K_2(y,y)\big)^{\frac{1}{2}}.$$

of kernels $K_i(\cdot,\cdot)$ attached positive definite functions

$$f_i = h_i * h_i^*, \qquad h_i \in C_c^r\big(G(\mathbb{A})\big).$$

It follows that

$$\sum_\chi |K_\chi(x,y)| = c(f)\|x\|^{N_1}\|y\|^{N_1}, \qquad x,y \in G(\mathbb{A}),$$

for some constant $c(f)$ depending on $f$.

A similar estimate holds for derivatives of the kernel. Suppose that $X$ and $Y$ are left invariant differential operators on $G(\mathbb{R})$ of degrees $d_1$ and $d_2$. Suppose that $f$ belongs to $C_c^r\big(G(\mathbb{A})\big)$, for some large positive integer $r$. The corresponding kernel then satisfies

$$X_1Y_2K_\chi(x,y) = K_\chi^{X,Y}(x,y), \qquad \chi \in \mathfrak{X},$$

where $K^{X,Y}(x,y)$ is the kernel attached to a function $f_{X,Y}$ in $C_c^{r-d_1-d_2}\big(G(\mathbb{A})\big)$. It follows that

$$\sum_{\chi \in \mathfrak{X}} |X_1Y_2K_\chi(x,y)| \leq c(f_{X,Y})\|x\|^{N_1}\|y\|^{N_1},$$

for all $x,y \in G(\mathbb{A})$.

We combine the last estimate with Proposition 13.2(b). Choose the objects $\mathcal{S}$, $N$, $N_0$ and $K_0$ of Proposition 13.2(b) so that $G(\mathbb{A}) = G(\mathbb{Q})\mathcal{S}$, $N$ is large, $N_0 = N_1$, and $f$ is biinvariant under $K_0$. We can then find a finite set $\{Y_i\}$ of left invariant differential operators on $G(\mathbb{R})$ such that

$$\begin{aligned}
&\sup_{y\in\mathcal{S}}\Big(\|y\|^N\sum_\chi|\Lambda_2^T K_\chi(x,y)|\Big)\\
\leq&\sup_{y\in G(\mathbb{A})}\Big(\sum_\chi\|y\|^{-N_0}\Big|\sum_i(Y_i)_2K_\chi(x,y)\Big|\Big)\\
\leq&\sup_{y\in G(\mathbb{A})}\Big(\sum_i\|y\|^{-N_0}\sum_\chi|(Y_i)_2K_\chi(x,y)|\Big)\\
\leq&\sup_{y\in G(\mathbb{A})}\Big(\sum_i\|y\|^{-N_0}c(f_{1,Y_i})\|x\|^{N_1}\|y\|^{N_1}\Big)\\
\leq&\Big(\sum_i c(f_{1,Y_i})\Big)\|x\|^{N_1},
\end{aligned}$$

for any $x\in G(\mathbb{A})$. Setting $x=y$, we see that there is a constant $c_1=c_1(f)$ such that

$$\sum_\chi|\Lambda_2^T K_\chi(x,x)|\leq c_1\|x\|^{N_1-N},$$

for any $x\in\mathcal{S}$. Since any bounded function is integrable over $\mathcal{S}^1=\mathcal{S}\cap G(\mathbb{A})^1$, we conclude that the sum over $\chi$ of the functions $|\Lambda_2^T K_\chi(x,x)|$ is integrable over $G(\mathbb{Q})\backslash G(\mathbb{A})^1$. This is the assertion (a).

The proof of (b) and (c) begins with an expansion of the function $k_\chi^T(x)=k_\chi^T(x,f)$. We are not thinking of the $\chi$-form of the expansion (8.3) of $k^T(x)$, but rather a parallel expansion in terms of partial truncation operators. We shall derive it as in §8, using Proposition 13.3 in place of Lemma 8.2.

The kernel $K_{P,\chi}(x,y)$ defined in §12 is invariant under left translation of either variable by $N_P(\mathbb{A})$. In particular, we can write

$$K_{P,\chi}(x,y)=\int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})}K_{P,\chi}(x,ny)\mathrm{d}n.$$

It follows from the definition in §12 that $k_\chi^T(x)$ equals

$$\sum_P(-1)^{(A_P/A_G)}\sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})}\widehat{\tau}_P\big(H_P(\delta x)-T\big)\int_{N_P(\mathbb{Q})\backslash N_P(\mathbb{A})}K_{P,\chi}(\delta x,n\delta x)\mathrm{d}n.$$

The integral over $n$ can then be expanded according to Proposition 13.3. The resulting sum over $P_1(\mathbb{Q})\backslash P(\mathbb{Q})$ combines with that over $P(\mathbb{Q})\backslash G(\mathbb{Q})$ to give an expression

$$\sum_{P_1\subset P}(-1)^{\dim(A_P/A_G)}\sum_{\delta\in P_1(\mathbb{Q})\backslash G(\mathbb{Q})}\widehat{\tau}_P\big(H_P(\delta x)-T\big)\tau_{P_1}^P\big(H_{P_1}(\delta x)-T\big)\Lambda_2^{T,P_1}K_{P,\chi}(\delta x,\delta x)$$

for $k_\chi^T(x)$. Applying the expansion (8.2), we write

$$\begin{aligned}
&\widehat{\tau}_P\big(H_P(\delta x)-T\big)\tau_{P_1}^P\big(H_{P_1}(\delta x)-T\big)\\
=&\widehat{\tau}_P\big(H_{P_1}(\delta x)-T\big)\tau_{P_1}^P\big(H_{P_1}(\delta x)-T\big)\\
=&\sum_{\{P_2:P_2\supset P\}}\sigma_{P_1}^{P_2}\big(H_{P_1}(\delta x)-T\big).
\end{aligned}$$

It follows that $k_\chi^T(x)$ has an expansion

$$\sum_{P_1 \subset P_2} \sum_{\delta \in P_1(\mathbb{Q})\backslash G(\mathbb{Q})} \sigma_{P_1}^{P_2}\big(H_{P_1}(\delta x) - T\big)\Lambda_2^{T,P_1} K_{P_1,P_2,\chi}(\delta x, \delta x), \tag{14.3}$$

where

$$K_{P_1,P_2,\chi}(x,y) = \sum_{\{P : P_1 \subset P \subset P_2\}} (-1)^{\dim(A_P/A_G)} K_{P,\chi}(x,y).$$

Observe that (14.3) is the same as the expansion (8.3) (or rather its $\chi$-analogue), except that the partial "cut-off" function $F^{P_1}(\cdot, T)$ has been replaced by the partial truncation operator $\Lambda_2^{T,P_1}$.

We recall from Lemma 8.3 that $\sigma_{P_1}^{P_2}$ vanishes if $P_1 = P_2 \neq G$, so the corresponding summand in (14.3) equals 0. If $P_1 = P_2 = G$, $\sigma_{P_1}^{P_2}$ equals 1, and the corresponding summand in (14.3) equals $\Lambda_2^T K_\chi(x,x)$. It follows that the difference

$$k_\chi^T(x) - \Lambda_2^T K_\chi(x,x)$$

equals the modified expression (14.3) obtained by taking the first sum over $P_1 \subsetneq P_2$. Consider the integral over $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ of the absolute value of this difference. The absolute value is of course bounded by the corresponding double sum of absolute values, in which we can combine the integral with the sum over $P_1(\mathbb{Q})\backslash G(\mathbb{Q})$. It follows that the double integral

$$\sum_{\chi \in \mathfrak{X}} \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} |k_\chi^T(x) - \Lambda_2^T K_\chi(x,x)| \mathrm{d}x$$

is bounded by

$$\sum_{\chi \in \mathfrak{X}} \sum_{P_1 \subsetneq P_2} \int_{P_1(\mathbb{Q})\backslash G(\mathbb{A})^1} \sigma_{P_1}^{P_2}\big(H_{P_1}(x) - T\big)|\Lambda_2^{T,P_1} K_{P_1,P_2,\chi}(x,x)| \mathrm{d}x. \tag{14.4}$$

The assertion (ii) would follow from (i) if it could be shown that (14.4) is finite. In fact, one shows that for $T$ highly regular, the integrand in (14.4) actually vanishes. This obviously suffices to establish both (ii) and (iii).

Consider the integrand in (14.4) attached to a fixed pair $P_1 \subsetneq P_2$. In order to treat the factor $\Lambda_2^{T,P_1} K_{P_1,P_2,\chi}$, one studies the function

$$\begin{aligned}
&\int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} K_{P_1,P_2}(x, n_1 y) \mathrm{d}n_1 \\
&= \sum_{\{P : P_1 \subset P \subset P_2\}} (-1)^{\dim(A_P/A_G)} \int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} K_P(x, n_1 y) \mathrm{d}n_1 \\
&= \sum_P (-1)^{\dim(A_P/A_G)} \int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} \int_{N_P(\mathbb{A})} \sum_{\gamma \in M_P(\mathbb{Q})} f(x^{-1}\gamma n n_1 y) \mathrm{d}n \mathrm{d}n_1.
\end{aligned}$$

In the last summand corresponding to $P$, we change the triple integral to a double integral over the product

$$M_P(\mathbb{Q}) N_P(\mathbb{A})/N_{P_1}(\mathbb{Q}) \times N_{P_1}(\mathbb{A}).$$

This in turn can be written as a triple integral over the product

$$\big(M_P(\mathbb{Q})/M_P(\mathbb{Q}) \cap N_{P_1}(\mathbb{Q})\big) \times \big(N_P(\mathbb{A})/N_P(\mathbb{Q})\big) \times N_{P_1}(\mathbb{A}).$$

The integral over $N_P(\mathbb{A})/N_P(\mathbb{Q})$ can then be absorbed in the integral over $N_{P_1}(\mathbb{A})$. Since

$$M_P(\mathbb{Q})/M_P(\mathbb{Q}) \cap N_{P_1}(\mathbb{Q}) \cong P(\mathbb{Q})/P_1(\mathbb{Q}) \times M_{P_1}(\mathbb{Q}),$$

the sum over $P$ takes the form

$$\begin{aligned}
&\sum_P (-1)^{\dim(A_P/A_G)} \sum_{\gamma \in P_1(\mathbb{Q})\backslash P(\mathbb{Q})} \int_{N_{P_1}(\mathbb{A})} \sum_{\gamma_1 \in M_{P_1}(\mathbb{Q})} f(x^{-1}\gamma^{-1}\gamma_1 n_1 y)\mathrm{d}n_1 \\
=&\sum_P (-1)^{\dim(A_P/A_G)} \sum_{\gamma \in P_1(\mathbb{Q})\backslash P(\mathbb{Q})} K_{P_1}(\gamma x, y).
\end{aligned}$$

Let $F(P_1, P_2)$ be the set of elements in $P_1(\mathbb{Q})\backslash P_2(\mathbb{Q})$ which do not lie in $P_1(\mathbb{Q})\backslash P(\mathbb{Q})$ for any $P$ with $P_1 \subset P \subsetneq P_2$. The alternating sum over $P$ and $\gamma$ then reduces to a sum over $\gamma \in F(P_1, P_2)$, by Identity 6.2. We have established that

$$(14.5) \quad \int_{N_{P_1}(\mathbb{Q})\backslash N_{P_1}(\mathbb{A})} K_{P_1,P_2}(x, n_1 y)\mathrm{d}n_1 = (-1)^{\dim(A_{P_2}/A_G)} \sum_{\gamma \in F(P_1,P_2)} K_{P_1}(\gamma x, y).$$

There remain two steps to showing that the integrand in (14.4) vanishes. The first is to show that for any $x$ and $y$, $\Lambda_2^{T,P_1} K_{P_1,P_2,\chi}(x,y)$ depends linearly on the function of $m \in M_{P_1}(\mathbb{Q})\backslash M_{P_1}(\mathbb{A})^1$ obtained from the left hand side of (14.5) by replacing $y$ by $my$. This is related to the decompositions of §12, and is easily established from the estimates we have discussed. The other is to show that if $T$ is highly regular relative to $\mathrm{supp}(f)$, and $\sigma_{P_1}^{P_2}\big(H_{P_1}(x) - T\big) \neq 0$, then $K_{P_1}(\gamma x, mx) = 0$ for all $m$ and any $\gamma \in F(P_1, P_2)$. This is a consequence of the Bruhat decomposition for $G(\mathbb{Q})$. In the interests of simplicity (rather than efficiency), we shall illustrate the ideas in the concrete example of $G = GL(2)$, referring the reader to [**A4**, §2] for the general case.

Assume that $G = GL(2)$, $P_1 = P_0$ and $P_2 = G$. The partial truncation operator $\Lambda^{T,P_1}$ is then given simply by an integral over $N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})$. Therefore

$$\Lambda_2^{T,P_1} K_{P_1,P_2,\chi}(x,y) = \int_{N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})} \big(K_\chi(x, ny) - K_{P_0,\chi}(x, ny)\big)\mathrm{d}n.$$

If $\chi = (G, \pi)$, the integral of $K_\chi(x, ny)$ over $n$ vanishes, since $\pi$ is a cuspidal automorphic representation of $G(\mathbb{A})$, while $K_{P_0,\chi}(x, ny)$ vanishes by definition. The integrand in (14.4) thus vanishes in this case for any $T$.

For $G = GL(2)$, we have reduced the problem to the remaining case that $\chi$ is represented by a pair $(P_0, \sigma_0)$. Since $M_{P_0}$ is the group of diagonal matrices in $GL(2)$, we can identify $\sigma_0$ with a pair of characters on the group $\mathbb{Q}^*\backslash\mathbb{A}^1$. It follows directly from the definitions that

$$\begin{aligned}
\int_{N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})} K_{P_0,\chi}(x, ny)\mathrm{d}n &= K_{P_0,\chi}(x, y) \\
&= \int_{M_{P_0}(\mathbb{Q})\backslash M_{P_0}(\mathbb{A})^1} K_{P_0}(x, my)\sigma_0(m)\mathrm{d}m.
\end{aligned}$$

The spectral decomposition of the kernel $K(x,y)$ also leads to a formula

$$\begin{aligned}&\int_{N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})} K_\chi(x,ny)\mathrm{d}n\\ &= \int_{M_{P_0}(\mathbb{Q})\backslash M_{P_0}(\mathbb{A})^1}\int_{N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})} K(x,nmy)\sigma_0(m)\mathrm{d}n\mathrm{d}m.\end{aligned}$$

Indeed, the required contribution from the terms in $K(x,y)$ corresponding to the Hilbert space $\mathcal{H}_G$ can be inferred from the fact that the representation of $G(\mathbb{A})$ on $\mathcal{H}_G$ is a sum of cuspidal automorphic representations and one-dimensional automorphic representations. To obtain the contribution from the terms in $K(x,y)$ corresponding to $\mathcal{H}_{P_0}$, we use the fact that for any $\phi \in \mathcal{H}^0_{P_0}$, the function

$$y \longrightarrow \int_{N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})} E(ny,\phi,\lambda)\mathrm{d}n$$

also belongs to $\mathcal{H}^0_{P_0}$. Combining the two formulas, we see that

$$\begin{aligned}&\Lambda_2^{T,P_1}K_{P_1,P_2,\chi}(x,y)\\ &= \int_{M_{P_0}(\mathbb{Q})\backslash M_{P_0}(\mathbb{A})^1}\int_{N_{P_0}(\mathbb{Q})\backslash N_{P_0}(\mathbb{A})} K_{P_1,P_2}(x,nmy)\sigma_0(m)\mathrm{d}n\mathrm{d}m\\ &= \int_{M_{P_0}(\mathbb{Q})\backslash M_{P_0}(\mathbb{A})^1}\sum_{\gamma\in F(P_1,P_2)} K_{P_1}(\gamma x,my)\sigma_0(m)\mathrm{d}m,\end{aligned}$$

for any $x$ and $y$. This completes the first step in the case of $G = GL(2)$.

For the second step, we note that

$$\begin{aligned}F(P_1,P_2) = F(P_0,G) &= P_0(\mathbb{Q})\backslash\big(G(\mathbb{Q}) - P_0(\mathbb{A})\big)\\ &= \left\{M_{P_0}(\mathbb{Q})\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}N_{P_0}(\mathbb{Q})\right\},\end{aligned}$$

by the Bruhat decomposition for $GL(2)$. Setting $y = x$, we write

$$\begin{aligned}&\sum_{\gamma\in F(P_1,P_2)} K_{P_1}(\gamma x,mx)\\ &= \sum_{\gamma\in F(P_1,P_2)}\int_{N_{P_0}(\mathbb{A})} f(x^{-1}\gamma^{-1}nmx)\mathrm{d}n\\ &= \int_{N_{P_0}(\mathbb{A})}\sum_{\nu\in N_{P_0}(\mathbb{Q})}\sum_{\mu\in M_{P_0}(\mathbb{Q})} f\left(x^{-1}\nu\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\mu nmx\right)\mathrm{d}n,\end{aligned}$$

for any $m \in M_P(\mathbb{A})^1$. We need to show that if $T$ is highly regular relative to $\mathrm{supp}(f)$, the product of any summand with $\sigma_{P_1}^{P_2}\big(H_{P_1}(x) - T\big)$ vanishes for each $x \in G(\mathbb{A})^1$. Assume the contrary, and write

$$x = n_1\begin{pmatrix}r & 0\\ 0 & r^{-1}\end{pmatrix}m_1k_1, \qquad n_1 \in N_{P_0}(\mathbb{A}),\ r\in(\mathbb{R}^*)^0,\ m_1\in M_{P_0}(\mathbb{A})^1,\ k_1\in K.$$

On the one hand, the number

$$\sigma_{P_1}^{P_2}\big(H_{P_1}(x)-T\big) = \sigma_{P_0}^{G}\big(H_{P_0}(x)-T\big) = \tau_{P_0}\left(H_{P_0}\begin{pmatrix}r & 0\\ 0 & r^{-1}\end{pmatrix} - T\right)$$

is positive, so that $r$ is large relative to $\mathrm{supp}(f)$. On the other hand, it follows from the discussion above that the point

$$x^{-1}\nu\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\mu n m x$$

belongs to $\mathrm{supp}(f)$, for some $\nu \in N_{P_0}(\mathbb{Q})$, $\mu \in M_{P_0}(\mathbb{Q})$, $n \in N_{P_0}(\mathbb{A})$, and $m \in M_{P_0}(\mathbb{A})^1$. Substituting for $x$, we see that there is a point $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$ in $GL(2,\mathbb{A})^1$, with $|c| = r^2$, which lies in the fixed compact set $K \cdot \mathrm{supp}\, f \cdot K$. This is a contradiction. The argument in the case of $G = GL(2)$ is thus complete. □

We have finished our remarks on the proof of Theorem 14.1. We can now treat the double integral (14.2) as we did its geometric analogue (10.3) in §10. By Fubini's theorem, we obtain an absolutely convergent expression

$$J^T(f) = \sum_{\chi\in\mathfrak{X}} J^T_\chi(f)$$

whose terms are given by absolutely convergent integrals

$$J^T_\chi(f) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} k^T_\chi(x,f)\mathrm{d}x, \qquad \chi\in\mathfrak{X}. \tag{14.6}$$

Following the discussion of §10, we analyze $J^T_\chi(f)$ as a function of $T$ by means of the proof of Theorem 9.1. Defined initially for $T \in \mathfrak{a}_0^+$ sufficiently regular, we see that $J^T_\chi(f)$ extends to any $T \in \mathfrak{a}_0$ as a polynomial function whose degree is bounded by the dimension of $\mathfrak{a}_0^G$. We then set

$$J_\chi(f) = J^{T_0}_\chi(f), \qquad \chi\in\mathfrak{X},$$

for the point $T_0 \in \mathfrak{a}_0^G$ given by (9.4). By the proof of Proposition 9.3, each distribution $J_\chi(f)$ is independent of the choice of minimal parabolic subgroup $P_0 \in \mathcal{P}(M_0)$.

The new distributions $J_\chi(f) = J^G_\chi(f)$ are again generally not invariant. Applying the proof of Theorem 9.4 to the absolutely convergent integral (14.6), we obtain the variance property

$$J_\chi(f^y) = \sum_{Q\supset P_0} J^{M_Q}_\chi(f_{Q,y}), \qquad \chi\in\mathfrak{X},\ y\in G(\mathbb{A}). \tag{14.7}$$

As before, $J^{M_Q}_\chi(f_{Q,y})$ is defined as a finite sum of distributions $J^{M_Q}_{\chi_Q}(f_{Q,y})$, in which $\chi_Q$ ranges over the preimage of $\chi$ in $\mathfrak{X}^{M_Q}$ under the mapping of $\mathfrak{X}^{M_Q}$ to $\mathfrak{X}$. Once again, $\chi$ need not lie in the image of the map $\mathfrak{X}^{M_Q} \to \mathfrak{X}$ attached to any proper parabolic subgroup $Q \subsetneq G$. This is the case precisely when $\chi$ is *cuspidal*, in the sense that it is defined by a pair $(G,\sigma)$. When $\chi$ is cuspidal, the distribution $J_\chi(f)$ is in fact invariant.

The expansion of $J^T(f)$ in terms of distributions $J^T_\chi(f)$ extends by polynomial interpolation to all values of $T$. Setting $T = T_0$, we obtain an identity

$$J(f) = \sum_{\chi\in\mathfrak{X}} J_\chi(f), \qquad f\in C^\infty_c\big(G(\mathbb{A})\big). \tag{14.8}$$

This is what we will call the coarse spectral expansion. The distributions $J_\chi(f)$ for which $\chi$ is cuspidal are to be regarded as general analogues of the spectral terms in the trace formula for compact quotient.

## 15. Weighted characters

This section is parallel to §11. It is aimed at the problem of describing the summands $J_\chi(f)$ in the coarse spectral expansion more explicitly. At this point, we can give a partial solution. We shall express $J_\chi(f)$ as a weighted character for "generic" classes $\chi \in \mathfrak{X}$.

For any $\chi \in \mathfrak{X}$, $J^T_\chi(f)$ is defined by the formula (14.6). However, Theorem 14.1(iii) and the definition (12.7) provide another expression

$$\begin{aligned} J^T_\chi(f) &= \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda^T_2 K_\chi(x,x)\mathrm{d}x \\ &= \sum_P n_P^{-1} \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Big( \int_{i\mathfrak{a}^*_P} \sum_{\phi\in\mathcal{B}_{P,\chi}} E\big(x, \mathcal{I}_P(\lambda,f)\phi,\lambda\big)\overline{\Lambda^T E(x,\phi,\lambda)}\mathrm{d}\lambda\Big)\mathrm{d}x \end{aligned}$$

for $J^T_\chi(f)$. This second formula is better suited to computation.

Suppose that $\lambda \in i\mathfrak{a}^*_P$. The function $E(x,\phi',\lambda)$ is slowly increasing for any $\phi' \in \mathcal{H}^0_{P,\chi}$, while the function $\Lambda^T E(x,\phi,\lambda)$ is rapidly decreasing by (13.9). The integral

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} E(x,\phi',\lambda)\overline{\Lambda^T E(x,\phi,\lambda)}\mathrm{d}x$$

therefore converges, and consequently defines a Hermitian bilinear form on $\mathcal{H}^0_{P,\chi}$. By the intertwining property of Eisenstein series, this bilinear form behaves in the natural way under the the actions of $K$ and $\mathcal{Z}_\infty$ on $\mathcal{H}^0_{P,\chi}$. It may therefore be written as

$$\big(M^T_{P,\chi}(\lambda)\phi',\phi\big),$$

for a linear operator $M^T_{P,\chi}(\lambda)$ on $\mathcal{H}^0_{P,\chi}$. Since $\Lambda^T$ is a self-adjoint projection, by Proposition 13.1, we see that

$$\big(M^T_{P,\chi}(\lambda)\phi',\phi\big) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda^T E(x,\phi',\lambda)\overline{\Lambda^T E(x,\phi,\lambda)}\mathrm{d}x, \tag{15.1}$$

for any vectors $\phi'$ and $\phi$ in $\mathcal{H}^0_{P,\chi}$. It follows that the operator $M^T_{P,\chi}(\lambda)$ is self-adjoint and positive definite.

The following result can be regarded as a spectral analogue of Theorem 11.1.

THEOREM 15.1. *If $T \in \mathfrak{a}^+_{P_0}$ is suitably regular, in a sense that depends only on the support of $f$, the double integral*

$$\sum_P n_P^{-1} \int_{i\mathfrak{a}^*_P} \operatorname{tr}\big(M^T_{P,\chi}(\lambda)\mathcal{I}_{P,\chi}(\lambda,f)\big)\mathrm{d}\lambda \tag{15.2}$$

*converges absolutely, and equals $J^T_\chi(f)$.*

This is Theorem 3.2 of [**A4**]. It includes the implicit assertion that the operator in the integrand is of trace class, as well as that of the absolute convergence of the integral. The precise assertion is Theorem 3.1 of [**A4**], which states that the expression

$$\sum_\chi \sum_P \int_{i\mathfrak{a}^*_P} \|M^T_{P,\chi}(\lambda)\mathcal{I}_{P,\chi}(\lambda,f)\|_1 \mathrm{d}\lambda$$

is finite. As usual, $\|\cdot\|_1$ denotes the trace class norm, taken here for operators on the Hilbert space $\mathcal{H}_{P,\chi}$.

Apart from the last convergence assertion, Theorem 15.1 is a formal consequence of the expression above for $J^T_\chi(f)$. It follows from the definition of $M^T_{P,\chi}(\lambda)$, once we know that the integral over $G(\mathbb{Q})\backslash G(\mathbb{A})^1$ in the expression can be taken inside the integral over $\lambda$ and the sum over $\phi$. The convergence assertion is a modest extension of Theorem 14.1(i). Its proof combines the same two techniques, namely the estimates for $K(x,y)$ obtained from Selberg's positivity argument, and the estimates for $\Lambda^T$ given by Proposition 13.2. We refer the reader to §3 of [**A4**]. $\square$

Suppose that $P$ is fixed. Since the inner product (15.1) depends only on the image of $T$ in the intersection $(\mathfrak{a}^G_{P_0})^+$ of $\mathfrak{a}^+_{P_0}$ with $\mathfrak{a}^G_{P_0}$, we shall assume for the rest of this section that $T$ actually lies in $(\mathfrak{a}^G_{P_0})^+$. It turns out that the inner product can be computed explicitly for cuspidal Eisenstein series. The underlying reason for this is that the constant term

$$\int_{N_Q(\mathbb{Q})\backslash N_Q(\mathbb{A})} E(nx,\phi,\lambda)\mathrm{d}n, \qquad\qquad \phi\in\mathcal{H}^0_P,\ \lambda\in\mathfrak{a}^*_{P,\mathbb{C}},$$

defined for any standard $Q\supset P_0$, has a relatively simple formula if $\phi$ is cuspidal.

Suppose that $\phi$ belongs to $\mathcal{H}^0_{P,\mathrm{cusp}}$ and that $\lambda$ lies in $\mathfrak{a}^*_{P,\mathbb{C}}$. If $Q$ is associated to $P$, we have the basic formula

$$\int_{N_Q(\mathbb{Q})\backslash N_Q(\mathbb{A})} E(nx,\phi,\lambda)\mathrm{d}n = \sum_{s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)} \big(M(s,\lambda)\phi\big)(x)\mathrm{e}^{(s\lambda+\rho_Q)(H_Q(x))}$$

This is established in the domain of absolute convergence of Eisenstein series from the integral formula for $M(s,\lambda)\phi$ and the Bruhat decomposition for $G(\mathbb{Q})$ [**Lan1**, Lemma 3]. More generally, suppose that $Q$ is arbitrary. Then

$$\int_{N_Q(\mathbb{Q})\backslash N_Q(\mathbb{A})} E(nx,\phi,\lambda)\mathrm{d}n = \sum_{s\in W(P;Q)} E^Q\big(x,M(s,\lambda)\phi,s\lambda\big), \tag{15.3}$$

where we have written $E^Q(\cdot,\cdot,\cdot)=E^Q_{P_1}(\cdot,\cdot,\cdot)$, for the group $P_1$ such that $s$ belongs to $W(\mathfrak{a}_P,\mathfrak{a}_{P_1})$. This is established inductively from the first formula by showing that for any $Q'\subsetneq Q$, the $Q'$-constant terms of each sides are equal. The formula (15.3) allows us to express the truncated Eisenstein series $\Lambda^T E(x,\phi,\lambda)$, for $\lambda$ in its domain of absolute convergence, in terms of the signs $\varepsilon_Q$ and characteristic functions $\phi_Q$ defined in §11.

LEMMA 15.2. *Suppose that $\phi\in\mathcal{H}^0_{P,\mathrm{cusp}}$ and $\lambda\in\Lambda+i\mathfrak{a}^*_P$, where $\Lambda$ is any point in the affine chamber $\rho_P+(\mathfrak{a}^*_P)^+$. Then*

$$\Lambda^T E(x,\phi,\lambda)=\sum_{Q\supset P_0}\ \sum_{\delta\in Q(\mathbb{Q})\backslash G(\mathbb{Q})}\psi_Q(\delta x), \tag{15.4}$$

*where for any $y\in G(\mathbb{A})$, $\psi_Q(y)$ is the sum over $s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)$ of the expression*

$$\varepsilon_Q(s\Lambda)\phi_Q\big(s\Lambda,H_Q(\delta x)-T_Q\big)\mathrm{e}^{(s\lambda+\rho_Q)(H_Q(y))}\big(M(s,\lambda)\phi\big)(y). \tag{15.5}$$

This is Lemma 4.1 of [**A4**]. To prove it, we note that for any $Q$, $s$, and $\delta$, the expression

$$\varepsilon_Q(s\Lambda)\phi_Q\big(s\Lambda,H_Q(\delta x)-T_Q\big)$$

equals

$$\sum_{\{R\supset Q:s\in W(P;R)\}}(-1)^{\dim(A_R/A_G)}\widehat{\tau}_R\big(H_R(\delta x)-T_R\big),$$

by the identity of (11.7) and (11.6) established in §11. We substitute this into the formula (15.5) for $\psi_Q(\delta x)$. We then take the sum over $\delta$ in (15.4) inside the resulting sums over $s$ and $R$. This allows us to decompose it into a double sum over $\xi \in Q(\mathbb{Q})\backslash R(\mathbb{Q})$ and $\delta \in R(\mathbb{Q})\backslash G(\mathbb{Q})$. The sum

$$\sum_{\xi\in Q(\mathbb{Q})\backslash R(\mathbb{Q})} \mathrm{e}^{(s\lambda+\rho_Q)(H_Q(\xi\delta x))}\big(M(s,\lambda)\phi\big)(\xi\delta x)$$

converges absolutely to $E^R\big(\delta x, M(s,\lambda)\phi, s\lambda\big)$. It follows that the right hand side of (15.4) equals

$$\sum_R (-1)^{\dim(A_R/A_G)}\sum_\delta \Big\{\sum_s E^R\big(\delta x, M(s,\lambda)\phi, s\lambda\big)\Big\}\widehat{\tau}_R\big(H_R(\delta x)-T_R\big),$$

with $\delta$ and $s$ summed over $R(\mathbb{Q})\backslash G(\mathbb{Q})$ and $W(P;R)$ respectively. Moreover, the last expression in the brackets equals

$$\int_{N_R(\mathbb{Q})\backslash N_R(\mathbb{A})} E(n\delta x, \phi, \lambda)\mathrm{d}n,$$

by (15.3). It then follows from the definition (13.1) that the right hand side of (15.4) equals the truncated Eisenstein series on the right hand side of (15.4). (The elementary convergence arguments needed to justify these manipulations are given on p. 114 of [**A4**].) □

For any $Q$, we treat the sum $\psi_Q$ in the last lemma as a function on $N_Q(\mathbb{A})M_Q(\mathbb{Q})\backslash G(\mathbb{A})^1$. It then follows from the definition of the characteristic functions $\phi_Q(s\Lambda,\cdot)$ and our choice of $\Lambda$ that $\psi_Q(x)$ is rapidly decreasing in $H_Q(x)$. This is slightly weaker than the condition of compact support imposed on the function $\psi$ in §12. However, we shall still express the right hand side of (15.4) as the sum over $Q$ of functions $(E\psi_Q)(x)$, following the notation of Lemma 12.2. In fact, the inner product formula (12.3) is easily seen to hold under the slightly weaker conditions here. We shall sketch how to use it to compute the inner product of truncated Eisenstein series.

One has first to compute the Fourier transform

$$\Psi_Q(\mu, x) = \int_{A_Q(\mathbb{R})^0\cap G(\mathbb{A})^1} \mathrm{e}^{-(\mu+\rho_Q)(H_Q(ax))}\psi_Q(ax)\mathrm{d}a,$$

for any $\mu \in i\mathfrak{a}_Q^*$. This entails computing the integral

$$\int_{A_Q(\mathbb{R})^0\cap G(\mathbb{A})^1} \mathrm{e}^{(s\lambda-\mu)(H_Q(ax))}\varepsilon_Q(s\Lambda)\phi_Q\big(s\Lambda, H_Q(ax)-T_Q\big)\mathrm{d}a,$$

which can be written as

$$\int_{\mathfrak{a}_Q^G} \mathrm{e}^{(s\lambda-\mu)(H)}\varepsilon_Q(s\Lambda)\phi_Q(s\Lambda, H-T_Q)\mathrm{d}H,$$

after the obvious change of variables. A second change of variables

$$H = \sum_{\alpha\in\Delta_Q} t_\alpha\alpha^\vee, \qquad t_\alpha\in\mathbb{R},$$

simplifies the integral further. It becomes a product of integrals of rapidly decreasing exponential functions over half lines, each of which contributes a linear form in $s\lambda-\mu$ to the denominator. We have of course to multiply the resulting expression by the relevant Jacobian determinant, which equals the volume of $\mathfrak{a}_Q^G$ modulo the lattice $\mathbb{Z}(\Delta_Q^\vee)$ generated by $\Delta_Q^\vee$. The result is

$$\Psi_Q(\mu,x) = \sum_{s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)} \mathrm{e}^{(s\lambda-\mu)(T)}\big(M(s,\lambda)\phi\big)(x)\theta_Q(s\lambda-\mu)^{-1}, \tag{15.6}$$

where

$$\theta_Q(s\lambda-\mu) = \mathrm{vol}\big(\mathfrak{a}_Q^G/\mathbb{Z}(\Delta_Q^\vee)\big)^{-1}\prod_{\alpha\in\Delta_Q}(s\lambda-\mu)(\alpha^\vee). \tag{15.7}$$

It is worth emphasizing that $\Psi_Q(\mu,x)$ is a rather simple function of $\mu$, namely a linear combination of products of exponentials with quotients of polynomials. We have taken the real part $\Lambda$ of $\lambda$ to be any point in $\rho_P + (\mathfrak{a}_P^*)^+$. Assume from now on that it is also highly regular, in the sense that $\Lambda(\alpha^\vee)$ is large for every $\alpha\in\Delta_P$. Then $\Psi_Q(\mu,x)$ is an analytic function of $\mu$ in the tube in $\mathfrak{a}_{Q,\mathbb{C}}^*$ over a ball $B_Q$ around 0 in $\mathfrak{a}_Q^*$ of large radius. Moreover, for any $\Lambda_Q\in B_Q$,

$$\Psi_Q(\mu):\ x \longrightarrow \Psi_Q(\mu,x), \qquad\qquad \mu\in\Lambda_Q + i(\mathfrak{a}_Q^G)^*,$$

is a square integrable function of $\mu$ with values in a finite dimensional subspace of $\mathcal{H}^0_{Q,\mathrm{cusp}}$.

Consider another set of data $P'$, $\phi'\in\mathcal{H}^0_{P',\mathrm{cusp}}$ and $\lambda'\in\Lambda'+i\mathfrak{a}_{P'}^*$, where $P'$ is associated to $P$ and $\Lambda'$ is a highly regular point in $\rho_{P'}+(\mathfrak{a}_{P'}^*)^+$. These give rise to a corresponding pair of functions $\psi_{Q'}(x)$ and $\Psi_{Q'}(\mu',x)$, for each standard $Q'$ associated to $P'$. Following the notation of Lemma 12.2, we write the inner product

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda^T E(x,\phi,\lambda)\overline{\Lambda^T E(x,\phi',\lambda')}\mathrm{d}x \tag{15.8}$$

as

$$\sum_{Q,Q'}\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1}(E\psi_Q)(x)\overline{(E\psi_{Q'})(x)}\mathrm{d}x.$$

We are taking for granted the extension of Lemma 12.3 to the rapidly decreasing functions $\psi_Q$ and $\psi_{Q'}$. It yields the further expression

$$\sum_{Q,Q'}\int_{\Lambda_Q+i(\mathfrak{a}_Q^G)^*}\sum_{t\in W(\mathfrak{a}_Q,\mathfrak{a}_{Q'})}\big(M(t,\mu)\Psi_Q(\mu),\Psi_{Q'}(-t\overline{\mu})\big)\mathrm{d}\mu$$

for the inner product, where $\Lambda_Q$ is any point in the intersection of $\rho_Q+(\mathfrak{a}_Q^G)^*$ with the ball $B_Q$. It follows from (15.6) (and its analogue for $P'$) that the inner product (15.8) equals the sum over $Q$ and $s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)$, and the integral over $\mu\in\Lambda_Q+i(\mathfrak{a}_Q^G)^*$, of the product of

$$\theta_Q(s\lambda-\mu)^{-1}\mathrm{e}^{(s\lambda-\mu)(T)} \tag{15.9}$$

with

$$\sum_{Q'}\sum_t\sum_{s'}\theta_{Q'}(s'\bar{\lambda}'+t\mu)^{-1}\mathrm{e}^{(s'\bar{\lambda}'+t\mu)(T)}\big(M(t,\mu)M(s,\lambda)\phi, M(s',\lambda')\phi'\big). \tag{15.10}$$

The inner sums in (15.10) are over elements $t\in W(\mathfrak{a}_Q,\mathfrak{a}_{Q'})$ and $s'\in W(\mathfrak{a}_{P'},\mathfrak{a}_{Q'})$.

There are three more steps. The first is to show that (15.10) is an analytic function of $\mu$ if the real part of $\mu$ is any point in $\rho_Q+(\mathfrak{a}_Q^*)^+$. The operator valued functions $M(t,\mu)$ are certainly analytic, since the integral formula (7.2) converges uniformly in the given domain. The remaining functions $\theta_Q(s'\bar{\lambda}'+t\mu)^{-1}$ of $\mu$ have singularities along hyperplanes

$$\{\mu : (s'\bar{\lambda}'+t\mu)(\alpha^\vee)=0\}, \qquad \alpha \in \Delta_{Q'},$$

for fixed $Q'$, $t$, $s'$ and $\lambda'$. However, each such hyperplane occurs twice in the sum (15.10), corresponding to a pair of multi-indices $(Q', t, s')$ and $(Q'_\alpha, s_\alpha t, s_\alpha s')$ that differ by a simple reflection about $\alpha$. (By definition, $Q'_\alpha$ is the standard parabolic subgroup such that $s_\alpha$ belongs to $W(\mathfrak{a}_{Q'}, \mathfrak{a}_{Q'_\alpha})$.) It is a consequence of the functional equations (7.4) that

$$\big(M(s_\alpha t,\mu)M(s,\lambda)\phi, M(s_\alpha s',\lambda')\phi'\big) = \big(M(t,\mu)M(s,\lambda)\phi, M(s',\lambda')\phi'\big),$$

whenever $(s'\bar{\lambda}'+t\mu)(\alpha^\vee)=0$. It then follows that the singularities cancel from the sum (15.10), and therefore that (15.10) is analytic in the given domain. (This argument is a basic part of the theory of $(G,M)$-families, to be discussed in §17.)

The second step is to show that if $s\neq 1$, the integral over $\mu$ of the product of (15.9) and (15.10) vanishes. For any such $s$, there is a root $\alpha\in\Delta_{Q'}$ such that $(s\Lambda_Q)(\alpha^\vee)<0$. As a function of $\mu$, (15.9) is analytic on any of the affine spaces

$$(\Lambda_Q + r\varpi_\alpha) + i(\mathfrak{a}_Q^G)^*, \qquad 0\leq r<\infty.$$

We have just seen that the same property holds for the function (15.10). We can therefore deform the contour of integration from $\Lambda_Q + i(\mathfrak{a}_Q^G)^*$ to the affine space attached to any $r$. The function $M(t,\mu)$ is bounded independently of $r$ on this affine space, as is the product

$$\mathrm{e}^{-\mu(T)}\mathrm{e}^{(t\mu)(T)}.$$

This leaves only the product

$$\theta_Q(s\lambda-\mu)^{-1}\theta_{Q'}(s'\bar{\lambda}'+t\mu)^{-1},$$

which is the inverse of a polynomial in $\mu$ of degree twice the dimension of the affine space. The integral attached to $r$ therefore approaches 0 as $r$ approaches infinity. The original integral therefore vanishes.

The final step is to set $s=1$ in (15.9) and (15.10), and then integrate the product of the resulting two expressions over $\mu$ in $\Lambda_Q + i(\mathfrak{a}_Q^G)^*$. The group $Q$ actually equals $P$ when $s$ equals 1. However, the point $\Lambda_Q$ in $(\mathfrak{a}_Q^G)^* = (\mathfrak{a}_P^G)^*$ does not equal the real part $\Lambda$ of $\lambda$. Indeed, the conditions we have imposed imply that $(\Lambda-\Lambda_Q)(\alpha^\vee)>0$ for each $\alpha\in\Delta_Q$. We change the contour of integration from $\Lambda_Q + i(\mathfrak{a}_Q^G)^*$ to the affine space

$$\Lambda_Q + r\rho_P + i(\mathfrak{a}_Q^G)^*,$$

for a large positive number $r$. As in the second step, the integral approaches 0 as $r$ approaches infinity. In this case, however, the function

$$\theta_Q(s\lambda-\mu)=\theta_P(\lambda-\mu)$$

contributes a multidimensional residue at $\mu=\lambda$. Using a change of variables

$$\mu = \sum_{\alpha\in\Delta_P} z_\alpha\varpi_\alpha, \qquad z_\alpha\in\mathbb{C},$$

one sees without difficulty that the residue equals the value of (15.10) at $s=1$ and $\mu=\lambda$. This value is therefore equal to the original inner product (15.8). Since the original indices of summation $Q$ and $s$ have disappeared, we may as well reintroduce them in place of the indices $Q'$ and $t$ in (15.9). We then have the following inner product formula.

PROPOSITION 15.3 (Langlands). *Suppose that $\phi \in \mathcal{H}^0_{P,\mathrm{cusp}}$ and $\phi' \in \mathcal{H}^0_{P',\mathrm{cusp}}$, for standard parabolic subgroups $P$ and $P'$. The inner product*

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda^T E(x,\phi,\lambda)\overline{\Lambda^T E(x,\phi',\lambda')}\mathrm{d}x$$

*is then equal to the sum*

$$\sum_Q \sum_s \sum_{s'} \theta_Q(s\lambda + s'\bar{\lambda}')^{-1} \mathrm{e}^{(s\lambda+s'\bar{\lambda}')(T)}\big(M(s,\lambda)\phi, M(s',\lambda')\phi'\big), \tag{15.11}$$

*taken over $Q \supset P_0$, $s \in W(\mathfrak{a}_P,\mathfrak{a}_Q)$ and $s' \in W(\mathfrak{a}_{P'},\mathfrak{a}_Q)$, as meromorphic functions of $\lambda \in \mathfrak{a}^*_{P,\mathbb{C}}$ and $\lambda' \in \mathfrak{a}^*_{P',\mathbb{C}}$.*

The discussion above has been rather dense. However, it does yield the required formula if the real parts of $\lambda$ and $\lambda'$ are suitably regular points in $(\mathfrak{a}^*_P)^+$ and $(\mathfrak{a}^*_{P'})^+$ respectively. Since both sides are meromorphic in $\lambda$ and $\bar{\lambda}'$, the formula holds in general. $\square$

The argument we have given was taken from §4 of [**A4**]. The formula stated by Langlands [**Lan1**, §9] actually differs slightly from (15.11). It contains an extra signed sum over the ordered partitions $\mathfrak{p}$ of the set $\Delta_Q$. The reader might find it an interesting combinatorial exercise to prove directly that this formula reduces to (15.11).

We shall say that a class $\chi \in \mathfrak{X}$ is *unramified* if for every pair $(P,\pi)$ in $\chi$, the stabilizer of $\pi$ in $W(\mathfrak{a}_P,\mathfrak{a}_P)$ is $\{1\}$. This is obviously completely parallel to the corresponding geometric definition in §11. Assume that $\chi$ is unramified, and that $(P,\pi)$ is a fixed pair in $\chi$. We shall use Proposition 15.3 to evaluate the distribution $J_\chi(f)$.

Suppose that $\phi$ and $\phi'$ are two vectors in the subspace $\mathcal{H}^0_{P,\mathrm{cusp},\pi}$ of $\mathcal{H}_P$. This represents the special case of Proposition 15.3 with $P' = P$. The factor

$$\big(M(s,\lambda)\phi, M(s',\lambda')\phi'\big)$$

in (15.11) vanishes if $s \neq s'$, since $M(s,\lambda)\phi$ and $M(s',\lambda')\phi'$ lie in the orthogonal subspaces $\mathcal{H}_{Q,\mathrm{cusp},s\pi}$ and $\mathcal{H}_{Q,\mathrm{cusp},s'\pi}$ of $\mathcal{H}_Q$. We use the resulting simplification to compute the inner product (15.1). We have of course to interchange the roles of $(\phi,\lambda)$ and $(\phi',\lambda')$, and then let $\lambda'$ approach a fixed point $\lambda \in i\mathfrak{a}^*_P$. Writing $\lambda' = \lambda + \zeta$, for a small point $\zeta \in i\mathfrak{a}^*_P$ in general position, we obtain

$$\begin{aligned}\big(M^T_{P,\chi}(\lambda)\phi',\phi\big) &= \lim_{\zeta\to 0}\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \Lambda^T E(x,\phi',\lambda+\zeta)\overline{\Lambda^T E(x,\phi,\lambda)}\mathrm{d}x \\ &= \lim_{\zeta\to 0}\sum_Q \sum_{s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)} \theta_Q(s\zeta)^{-1}\mathrm{e}^{(s\zeta)(T)}\big(M(s,\lambda+\zeta)\phi', M(s,\lambda)\phi\big).\end{aligned}$$

In particular, the last limit exists, and takes values in a finite dimensional space of functions of the highly regular point $T \in (\mathfrak{a}^G_0)^+$. (This is also easy to show directly.) We can therefore extend both the limit and the operator $M^T_{P,\chi}(\lambda)$ to all values of $T \in \mathfrak{a}^G_{P_0}$ so that the identity remains valid. Now, let $M(\widetilde{w}_s,\lambda)$ be the operator on $\mathcal{H}_P$ defined by analytic continuation from the analogue of (7.2) in which $w_s$ has been replaced by the representative $\widetilde{w}_s$ of $s$ in $K$. Since $M(s,\lambda)$ is unitary, we see easily from the definition (9.4) that

$$\begin{aligned}&\big(M(s,\lambda+\zeta)\phi', M(s,\lambda)\phi\big) = \big(M(s,\lambda)^{-1}M(s,\lambda+\zeta)\phi',\phi\big)\\ &= \mathrm{e}^{-(s\zeta)(T_0)}\big(M(\widetilde{w}_s,\lambda)^{-1}M(\widetilde{w}_s,\lambda+\zeta)\phi',\phi\big).\end{aligned}$$

It follows that

$$\big(M^{T_0}_{P,\chi}(\lambda)\phi',\phi\big) = \lim_{\zeta\to 0}\sum_Q\sum_s \theta_Q(s\zeta)^{-1}\big(M(\widetilde{w}_s,\lambda)^{-1}M(\widetilde{w}_s,\lambda+\zeta)\phi',\phi\big).$$

This formula does not depend on the choice of $\pi$. To compute the value

$$\mathrm{tr}\big(M^{T_0}_{P,\chi}(\lambda)\mathcal{I}_{P,\chi}(\lambda,f)\big) \tag{15.12}$$

at $T=T_0$ of the integrand in (15.2), we need only replace $\phi'$ by $\mathcal{I}_{P,\chi}(\lambda,f)\phi$, and then sum $\phi$ over a suitable orthonormal basis of $\mathcal{H}_{P,\chi}$.

Recall that $\mathcal{I}_P(\pi_\lambda)$ denotes the representation of $G(\mathbb{A})$ obtained by parabolic induction from the representation

$$\pi_\lambda(m) = \pi(m)\mathrm{e}^{\lambda(H_P(m))}, \qquad m\in M(\mathbb{A}),$$

of $M_P(\mathbb{A})$. We can also write $M(\widetilde{w}_s,\pi_\lambda)$ for the intertwining operator from $\mathcal{I}_P(\pi_\lambda)$ to $\mathcal{I}_Q(s\pi_\lambda)$ associated to an element $s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)$. Finally, let $m_{\mathrm{cusp}}(\pi)$ denote the multiplicity of $\pi$ in the representation $R_{M_P,\mathrm{cusp}}$. Since

$$\mathcal{H}_{P,\chi} = \bigoplus_{s\in W(\mathfrak{a}_P,\mathfrak{a}_P)}\mathcal{H}_{P,\mathrm{cusp},s\pi},$$

the representation $\mathcal{I}_{P,\chi}(\lambda)$ is then isomorphic to a direct sum of

$$|W(\mathfrak{a}_P,\mathfrak{a}_P)|m_{\mathrm{cusp}}(\pi) \tag{15.13}$$

copies of the representation $\mathcal{I}_P(\pi_\lambda)$. The trace (15.12) is therefore equal to the product of (15.13) with

$$\mathrm{tr}\big(\mathcal{M}_P(\pi_\lambda)\mathcal{I}_P(\pi_\lambda,f)\big),$$

where $\mathcal{M}_P(\pi_\lambda)$ is the operator on underlying Hilbert space of $\mathcal{I}_P(\pi_\lambda)$ defined explicitly in terms of intertwining operators by

$$\mathcal{M}_P(\pi_\lambda) = \lim_{\zeta\to 0}\Big(\sum_Q\sum_{s\in W(\mathfrak{a}_P,\mathfrak{a}_Q)}\theta_Q(s\lambda)^{-1}M(\widetilde{w}_s,\pi_\lambda)^{-1}M(\widetilde{w}_s,\pi_{\lambda+\zeta})\Big). \tag{15.14}$$

Since $P$ has been fixed, we shall let $P_1$ index the sum over standard parabolic subgroups in the formula (15.2) for $J^T_\chi(f)$. If $P_1$ does not belong to $\mathcal{P}_\chi$, it turns out that $\mathcal{H}_{P_1,\chi}=\{0\}$. This is a consequence of Langlands's construction [**Lan5**, §7] of the full discrete spectrum in terms of residues of cuspidal Eisenstein series. For the construction includes a description of the inner product on the residual discrete spectrum in terms of residues of cuspidal self-intertwining operators. Since $\chi$ is unramified, there are no such operators, and the residual discrete spectrum associated to $\chi$ is automatically zero. This leaves only groups $P_1$ in the set $\mathcal{P}_\chi$. For any such $P_1$, the value at $T=T_0$ of the corresponding integral in (15.2) equals the integral over $\lambda\in i\mathfrak{a}^*_M$ of (15.11). Since

$$n_P^{-1}|\mathcal{P}_\chi||W(\mathfrak{a}_P,\mathfrak{a}_P)| = 1,$$

we obtain the following theorem.

THEOREM 15.4. *Suppose that $\chi = \{(P, \pi)\}$ is unramified. Then*

$$J_\chi(f) = m_{\mathrm{cusp}}(\pi) \int_{i\mathfrak{a}_P^*} \mathrm{tr}\big(\mathcal{M}_P(\pi_\lambda)\mathcal{I}_P(\pi_\lambda, f)\big) d\lambda. \tag{15.15}$$

□

# Part II. Refinements and Applications

## 16. The first problem of refinement

We have completed the general steps outlined in §6. The coarse geometric expansion of §10 and the coarse spectral expansion of §14 give us an identity

$$\sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}(f) = \sum_{\chi\in\mathcal{X}} J_\chi(f), \qquad f \in C_c^\infty\big(G(\mathbb{A})\big), \tag{16.1}$$

that holds for any reductive group $G$. We have also seen how to evaluate the distributions $J_{\mathfrak{o}}(f)$ and $J_\chi(f)$ explicitly for unramified classes $\mathfrak{o}$ and $\chi$.

From now on, we shall generally work over an arbitrary number field $F$, whose adele ring $\mathbb{A}_F$ we denote simply by $\mathbb{A}$. We write $S_\infty$ for the set of archimedean valuations of $F$, and we let $q_v$ denote the order of the residue class field of the nonarchimedean completion $F_v$ attached to any $v \notin S_\infty$. We are now taking $G$ to be a fixed, connected reductive algebraic group over $F$. We write $S_{\mathrm{ram}} = S_{\mathrm{ram}}(G)$ for the finite set of valuations of $F$ outside of which $G$ is unramified. Thus, for any $v \notin S_{\mathrm{ram}}$, $G$ is quasisplit over $F_v$, and splits over some finite unramified extension of $F_v$.

The notation of Part I carries over with $F$ in place of $\mathbb{Q}$. So do the results, since they are valid for the group $G_1 = R_{F/\mathbb{Q}}G$ over $\mathbb{Q}$ obtained from $G$ by restriction of scalars. For example, the real vector space $\mathfrak{a}_{G_1}$ is canonically isomorphic to its analogue $\mathfrak{a}_G$ for $G$. The kernel $G(\mathbb{A})^1$ of the canonical mapping $H_G\colon G(\mathbb{A}) \to \mathfrak{a}_G$ is isomorphic to $G_1(\mathbb{A})^1$. It is a factor in a direct product decomposition

$$G(\mathbb{A}) = G(\mathbb{A})^1 \times A_\infty^+,$$

whose other factor

$$A_\infty^+ = A_{G_1}(\mathbb{R})^0$$

embeds diagonally in the connected, abelian Lie group

$$\prod_{v\in S_\infty} A_G(F_v)^0.$$

We shall apply the notation and results of Part I without further comment.

The results in Part I that culminate in the identity (16.1) are the content of the papers [**A3**], [**A4**] and [**A5**, §1–3], and a part of [**A1**, §1–3]. We note in passing that there is another possible approach to the problem, which was used more recently in a local context [**A19**]. It exploits the cruder truncation operation of simply multiplying functions by the local analogue of the characteristic function $F^G(\cdot, T)$. Although the methods of [**A19**] have not been applied globally, they could conceivably shorten some of the arguments. On the other hand, such methods are perhaps less natural in the global context. They would lead to functions of $T$ that are asymptotic to the relevant polynomials, rather than being actually equal to them.

The identity (16.1) can be regarded as a first approximation to a general trace formula. Let us write $\mathcal{X}_{\mathrm{cusp}}$ for the set of cuspidal classes in $\mathcal{X}$. A class $\chi \in \mathcal{X}_{\mathrm{cusp}}$ is thus of the form $(G, \pi)$, where $\pi$ is a cuspidal automorphic representation of $G(\mathbb{A})^1$. For any such $\chi$, the explicit formula of §15 specializes to

$$J_\chi(f) = a^G(\pi) f_G(\pi),$$

where

$$f_G(\pi) = \operatorname{tr}\big(\pi(f)\big) = \operatorname{tr}\Big(\int_{G(\mathbb{A})^1} f(x)\pi(x)\mathrm{d}x\Big)$$

and

$$a^G(\pi) = m_{\mathrm{cusp}}(\pi).$$

Recall that $m_{\mathrm{cusp}}(\pi)$ is the multiplicity of $\pi$ in the representation $R_{\mathrm{cusp}}$ of $G(\mathbb{A})^1$ on $L^2_{\mathrm{cusp}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1\big)$. In particular,

$$\operatorname{tr}\big(R_{\mathrm{cusp}}(f)\big) = \sum_{\chi\in\mathcal{X}_{\mathrm{cusp}}} J_\chi(f).$$

The identity (16.1) can thus be written as a trace formula

$$\operatorname{tr}\big(R_{\mathrm{cusp}}(f)\big) = \sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}(f) - \sum_{\chi\in\mathcal{X}-\mathcal{X}_{\mathrm{cusp}}} J_\chi(f). \tag{16.1$'$}$$

The problem is that the explicit formulas we have obtained so far do not apply to all of the terms on the right.

It is also easy to see that (16.1) generalizes the Selberg trace formula (1.3) for compact quotient. Let us write $\mathcal{O}_{\mathrm{anis}}$ for the set of anisotropic classes in $\mathcal{O}$. A class $\mathfrak{o}\in\mathcal{O}_{\mathrm{anis}}$ is thus of form $\{\gamma\}$, where $\gamma$ represents an anisotropic conjugacy class in $G(\mathbb{Q})$. (Recall that an anisotropic class is one that does not intersect $P(\mathbb{Q})$ for any proper $P\subsetneq G$.) For any such $\mathfrak{o}$, the explicit formula of §11 specializes to

$$J_{\mathfrak{o}}(f) = a^G(\gamma)f_G(\gamma),$$

where

$$f_G(\gamma) = \int_{G(\mathbb{A})_\gamma\backslash G(\mathbb{A})} f(x^{-1}\gamma x)\mathrm{d}x$$

and

$$a^G(\gamma) = \operatorname{vol}\big(G(F)_\gamma\backslash G(\mathbb{A})^1_\gamma\big).$$

The identity (16.1) can therefore be written

(16.1)$''$

$$\sum_{\gamma\in\Gamma_{\mathrm{anis}}(G)} a^G(\gamma)f_G(\gamma) + \sum_{\mathfrak{o}\in\mathcal{O}-\mathcal{O}_{\mathrm{anis}}} J_{\mathfrak{o}}(f) = \sum_{\pi\in\Pi_{\mathrm{cusp}}(G)} a^G(\pi)f_G(\pi) + \sum_{\chi\in\mathcal{X}-\mathcal{X}_{\mathrm{cusp}}} J_\chi(f),$$

where $\Gamma_{\mathrm{anis}}(G)$ is the set of conjugacy classes in $G(F)$ that do not intersect any proper group $P(F)$, and $\Pi_{\mathrm{cusp}}(G)$ is the set of equivalence classes of cuspidal automorphic representations of $G(\mathbb{A})^1$. Recall that $G(F)\backslash G(\mathbb{A})^1$ is compact if and only if $G$ has no proper rational parabolic subgroup $P$. In this case $\mathcal{O}=\mathcal{O}_{\mathrm{anis}}$ and $\mathcal{X}=\mathcal{X}_{\mathrm{cusp}}$, and (16.1)$''$ reduces to the trace formula for compact quotient discussed in §1.

For general $G$, the equivalent formulas (16.1), (16.1)$'$, and (16.1)$''$ are of limited use as they stand. Without explicit expressions for all of the distributions $J_{\mathfrak{o}}(f)$ and $J_\chi(f)$, one cannot get much information about the discrete (or cuspidal) spectrum. In the language of [**CLL**], we need to refine the coarse geometric and spectral expansions we have constructed.

What exactly are we looking for? The unramified cases solved in §11 and §15 will serve as guidelines.

The weighted orbital integral on the right hand side of the formula (11.9) is defined explicitly in terms of $f$. It is easier to handle than the original global construction of the distribution $J_{\mathfrak{o}}(f)$ on the left hand side of the formula. We would like to have a similar formula in general. The problem is that the right hand side of (11.9) does not make sense for more general classes $\mathfrak{o} \in \mathcal{O}$. It is in fact not so simple to define weighted orbital integrals for arbitrary elements in $M$. We shall do so in §18. Then in §19, we shall describe a general formula for $J_{\mathfrak{o}}(f)$ as a linear combination of weighted orbital integrals.

The weighted character on the right hand side of (15.15) is also defined explicitly in terms of $f$. It is again easier to handle than the global construction of the distribution $J_\chi(f)$ on the left hand side. Weighted characters are actually rather easy to define in general. However, this advantage is accompanied by a delicate analytic problem that does not occur on the geometric side. It concerns an interchange of two limits that arises when one tries to evaluate $J_\chi(f)$ for general classes $\chi \in \mathcal{X}$. We shall describe the solution of the analytic problem in §20. In §21 we shall give a general formula for $J_\chi(f)$ as a linear combination of weighted characters.

We adjust our focus slightly in Part II, which is to say, for the rest of the paper. We have already agreed to work over a general number field $F$ instead of $\mathbb{Q}$. We shall make three further changes, all minor, in the conventions of Part I.

The first is a small change of notation. If $H$ is a connected algebraic group over a given field $k$, and $\gamma$ belongs to $H(k)$, we shall denote the centralizer of $\gamma$ in $H$ by $H_{\gamma,+}$ instead of $H_\gamma$. We reserve the symbol $H_\gamma$ for the Zariski connected component of 1 in $H_{\gamma,+}$. Then $H_\gamma$ is a connected algebraic group over $k$, which is reductive if $H$ is reductive and $\gamma$ is semisimple. This convention leads to a slightly different way of writing the formula (11.9) for unramified classes $\mathfrak{o} \in \mathcal{O}$. In particular, suppose that $\mathfrak{o}$ is anisotropic. Then

$$J_{\mathfrak{o}}(f) = a^G(\gamma) f_G(\gamma),$$

where we now write

$$a^G(\gamma) = \mathrm{vol}\big(G_\gamma(F)\backslash G_\gamma(\mathbb{A})^1\big)$$

and

$$f_G(\gamma) = \int_{G_\gamma(\mathbb{A})\backslash G(\mathbb{A})} f(x^{-1}\gamma x)\mathrm{d}x.$$

This would seem to be in conflict with the notation of $(16.1)''$, since the group $G_\gamma(\mathbb{A})^1$ here is of finite index in the group denoted $G(\mathbb{A})^1_\gamma$ above. There is in fact no discrepancy, for the reason that the two factors $a^G(\gamma)$ and $f_G(\gamma)$ depend in either case on an implicit and unrestricted choice of Haar measure on the given isotropy group.

The second change is to make the discussion more canonical by allowing the minimal parabolic subgroup $P_0$ to vary. We have, after all, shown that the distributions $J_{\mathfrak{o}}(f)$ and $J_\chi(f)$ are independent of $P_0$. Some new notation is required, which we may as well formulate for an arbitrary field $k$ that contains $F$. We can of course regard $G$ as a reductive algebraic group over $k$. Parabolic subgroups certainly make sense in this context, as do other algebraic objects we have discussed.

By a Levi subgroup of $G$ over $k$, we mean an $k$-rational Levi component of some $k$-rational parabolic subgroup of $G$. Any such group $M$ is reductive, and comes with a maximal $k$-split central torus $A_M$, and a corresponding real vector space $\mathfrak{a}_M$. (A Levi subgroup $M$ of $G$ over $F$ is also a Levi subgroup over $k$, but $A_M$ and $\mathfrak{a}_M$ depend on the choice of base field. Failure to remember this can lead to embarrassing errors!) Given $M$, we write $\mathcal{L}(M) = \mathcal{L}^G(M)$ for the set of Levi subgroups of $G$ over $k$ that contain $M$, and $\mathcal{F}(M) = \mathcal{F}^G(M)$ for the set of parabolic subgroups of $G$ over $k$ that contain $M$. Any element $Q \in \mathcal{F}(M)$ has a unique Levi component $M_Q$ in $\mathcal{L}(M)$, and hence a canonical Levi decomposition $Q = M_Q N_Q$. We write $\mathcal{P}(M)$ for the subset of groups $Q \in \mathcal{F}(M)$ such that $M_Q = M$. For any $P \in \mathcal{P}(M)$, the roots of $(P, A_M)$ determine an open chamber $\mathfrak{a}_P^+$ in the vector space $\mathfrak{a}_M$. Similarly, the corresponding coroots determine a chamber $(\mathfrak{a}_M^*)_P^+$ in the dual space $\mathfrak{a}_M^*$.

The sets $\mathcal{P}(M)$, $\mathcal{L}(M)$ and $\mathcal{F}(M)$ are all finite. They can be described in terms of the geometry on the space $\mathfrak{a}_M$. To see this, we use the singular hyperplanes in $\mathfrak{a}_M$ defined by the roots of $(G, A_M)$. For example, the correspondence $P \to \mathfrak{a}_P^+$ is a bijection from $\mathcal{P}(M)$ onto the set of connected components in the complement in $\mathfrak{a}_M$ of the set of singular hyperplanes. We shall say that two groups $P, P' \in \mathcal{P}(M)$ are adjacent if their chambers share a common wall. The mapping $L \to \mathfrak{a}_L$ is a bijection from $\mathcal{L}(M)$ onto the set of subspaces of $\mathfrak{a}_M$ obtained by intersecting singular hyperplanes. The third set $\mathcal{F}(M)$ is clearly the disjoint union over $L \in \mathcal{L}(M)$ of the sets $\mathcal{P}(L)$. The mapping $Q \to \mathfrak{a}_Q^+$ is therefore a bijection from $\mathcal{F}(M)$ onto the set of "facets" in $\mathfrak{a}_M$, obtained from chambers of subspaces $\mathfrak{a}_L$. Since any element in $\mathfrak{a}_M$ belongs to a unique facet, there is a surjective mapping from $\mathfrak{a}_M$ to $\mathcal{F}(M)$.

Suppose for example that $G$ is the split group $SL(3)$, that $k$ is any field, and that $M = M_0$ is the standard minimal Levi subgroup. The singular hyperplanes in the two dimensional space $\mathfrak{a}_M$ are illustrated in Figure 16.1. The set $\mathcal{P}(M)$ is bijective with the six open chambers in the diagram. The set $\mathcal{L}(M)$ has five elements, consisting of the two-dimensional space $\mathfrak{a}_M$, the three one-dimensional lines, and the zero-dimensional origin. The set $\mathcal{F}(M)$ has thirteen elements, consisting of six open chambers, six half lines, and the origin. The intuition gained from Figure 16.1, simple though it is, is often useful in understanding operations we perform in general.

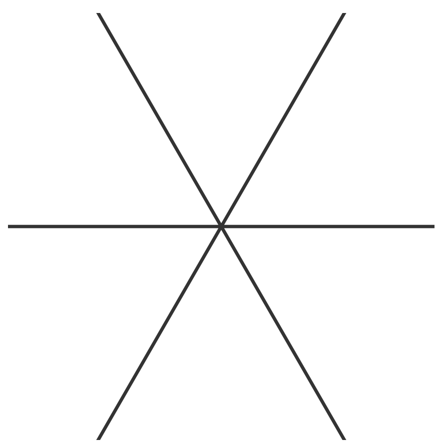


FIGURE 16.1. The three singular hyperplanes in the two dimensional space $\mathfrak{a}_M = \mathfrak{a}_0$ attached to $G = SL(3)$.

Suppose now that $k = F$. Even though we do not fix the minimal parabolic subgroup as in Part I, we shall work with a fixed minimal Levi subgroup $M_0$ of $G$

over $\mathbb{Q}$. We denote the associated sets $\mathcal{L}(M_0)$ and $\mathcal{F}(M_0)$ by $\mathcal{L} = \mathcal{L}^G$ and $\mathcal{F} = \mathcal{F}^G$, respectively.

The variance formulas (10.6) and (14.7) can be written without reference to $P_0$. The reason is that for a given $P_0 \in \mathcal{P}(M_0)$, any group $R \in \mathcal{F}$ is the image under some element in the restricted Weyl group $W_0 = W_0^G$ of a unique group $Q \in \mathcal{F}$ with $Q \supset P_0$. It is an easy consequence of the definitions that $J_{\mathfrak{o}}^{M_R}(f_{R,y})$ equals $J_{\mathfrak{o}}^{M_Q}(f_{Q,y})$ for any $\mathfrak{o}$, and that $J_\chi^{M_R}(f_{R,y})$ equals $J_\chi^{M_Q}(f_{Q,y})$ for any $\chi$. The order of the preimage of $Q$ in $\mathcal{F}$ is equal to the quotient $|W_0^{M_Q}||W_0^G|^{-1}$. Letting $Q$ now stand for an arbitrary group in $\mathcal{F}$, we can write the earlier formulas as

$$J_{\mathfrak{o}}(f^y) = \sum_{Q \in \mathcal{F}} |W_0^{M_Q}||W_0^G|^{-1} J_{\mathfrak{o}}^{M_Q}(f_{Q,y}), \qquad \mathfrak{o} \in \mathcal{O}, \tag{16.2}$$

and

$$J_\chi(f^y) = \sum_{Q \in \mathcal{F}} |W_0^{M_Q}||W_0^G|^{-1} J_\chi^{M_Q}(f_{Q,y}), \qquad \chi \in \mathcal{X}. \tag{16.3}$$

The third point is a slight change of emphasis. The distributions $J_{\mathfrak{o}}(f)$ and $J_\chi(f)$ in (16.1) depend only on the restriction of $f$ to $G(\mathbb{A})^1$. We have in fact identified $f$ implicitly with its restriction to $G(\mathbb{A})^1$, in writing $R_{\mathrm{cusp}}(f)$ above for example. Let us now formalize the convention by setting $C_c^\infty\big(G(\mathbb{A})^1\big)$ equal to the space of functions on $G(\mathbb{A})^1$ obtained by restriction of functions in $C_c^\infty\big(G(\mathbb{A})\big)$. We can then take the test function $f$ to be an element in $C_c^\infty\big(G(\mathbb{A})^1\big)$ rather than $C_c^\infty\big(G(\mathbb{A})\big)$, thereby regarding (16.1) as an identity of distributions on $G(\mathbb{A})^1$. This adjustment is obviously quite trivial. However, as we shall see in §22, it raises an interesting philosophical question that is at the heart of some key operations on the trace formula.

## 17. $(G, M)$-families

The terms in the refined trace formula will have some interesting combinatorial properties. To analyze them, one introduces the notion of a $(G, M)$-family of functions. We shall see that among other things, $(G, M)$-families provide a partial unification of the study of weighted orbital integrals and weighted characters.

We are now working in the setting of the last section. Then $G$ is defined over the fixed number field $F$, and hence over any given extension $k$ of $F$. Let $M$ be a Levi subgroup of $G$ over $k$. Suppose that for each $P \in \mathcal{P}(M)$,

$$c_P(\lambda), \qquad \lambda \in i\mathfrak{a}_M^*,$$

is a smooth function on the real vector space $i\mathfrak{a}_M^*$. The collection

$$\{c_P(\lambda) : \ P \in \mathcal{P}(M)\}$$

is called a $(G, M)$-*family* if $c_P(\lambda) = c_{P'}(\lambda)$, for any pair of adjacent groups $P, P' \in \mathcal{P}(M)$, and any point $\lambda$ in the hyperplane spanned by the common wall of the chambers $i(\mathfrak{a}_M^*)_P^+$ and $i(\mathfrak{a}_M^*)_{P'}^+$. We shall describe a basic operation that assigns a supplementary smooth function $c_M(\lambda)$ on $i\mathfrak{a}_M^*$ to any $(G, M)$-family $\{c_P(\lambda)\}$.

The algebraic definitions of §4 and §5 of course hold with the field $k$ in place of $\mathbb{Q}$. In particular, for any $P \in \mathcal{P}(M)$ we have the simple roots $\Delta_P$ of $(P, A_M)$, and the associated sets $\Delta_P^\vee$, $\widehat{\Delta}_P$ and $(\widehat{\Delta}_P)^\vee$. We are assuming we have fixed a suitable

Haar measure on the subspace $\mathfrak{a}_M^G = \mathfrak{a}_P^G$ of $\mathfrak{a}_M$. We then define a homogeneous polynomial

$$\theta_P(\lambda) = \mathrm{vol}\big(\mathfrak{a}_M^G/\mathbb{Z}(\Delta_P^\vee)\big)^{-1} \cdot \prod_{\alpha\in\Delta_P} \lambda(\alpha^\vee), \qquad \lambda \in i\mathfrak{a}_M^*,$$

on $i\mathfrak{a}_M^*$, where $\mathbb{Z}(\Delta_P^\vee)$ is the lattice spanned by the basis $\Delta_P^\vee$ of $\mathfrak{a}_M^G$.

Lemma 17.1. *For any $(G,M)$-family $\{c_P(\lambda)\}$, the sum*

$$c_M(\lambda) = \sum_{P\in\mathcal{P}(M)} c_P(\lambda)\theta_P(\lambda)^{-1} \tag{17.1}$$

*extends to a smooth function of $\lambda \in i\mathfrak{a}_M^*$.*

The only possible singularities of $c_M(\lambda)$ are simple poles along hyperplanes in $i\mathfrak{a}_M^*$ of the form $\lambda(\alpha^\vee) = 0$. These in turn come from adjacent pairs $P$ and $P'$ for which $\alpha$ and $\alpha' = (-\alpha)$ are respective simple roots. Using the fact that $c_P(\lambda) = c_{P'}(\lambda)$ for any $\lambda$ on the hyperplane, one sees directly that the simple poles cancel, and therefore that $c_M(\lambda)$ does extend to a smooth function. (See [**A5**, Lemma 6.2].) □

We often write $c_M = c_M(0)$ for the value of $c_M(\lambda)$ at $\lambda = 0$. It is in this form that the $(G,M)$-families from harmonic analysis usually appear.

We shall first describe a basic example that provides useful geometric intuition. Suppose that

$$\mathcal{Y} = \{Y_P : \ P \in \mathcal{P}(M)\}$$

is a family of points in $\mathfrak{a}_M$ parametrized by $\mathcal{P}(M)$. We say that $\mathcal{Y}$ is a *positive, $(G,M)$-orthogonal set* if for every pair $P$ and $P'$ of adjacent groups in $\mathcal{P}(M)$, whose chambers share the wall determined by the uniquely determined simple root $\alpha \in \Delta_P$,

$$Y_P - Y_{P'} = r_\alpha \alpha^\vee,$$

for a nonnegative number $r_\alpha$. Assume that this condition holds. The collection

$$c_P(\lambda, \mathcal{Y}) = \mathrm{e}^{\lambda(Y_P)}, \qquad \lambda \in i\mathfrak{a}_M^*,\ P \in \mathcal{P}(M), \tag{17.2}$$

is then a $(G,M)$-family of functions, which extend analytically to all points $\lambda$ in the complex space $\mathfrak{a}_{M,\mathbb{C}}^*$. As with any $(G,M)$-family, the associated smooth function $c_M(\lambda,\mathcal{Y})$ depends on the choice of Haar measure on $\mathfrak{a}_M^G$. In this case, the function has a simple interpretation.

Observe first that

$$Y_P = Y_P^G + Y_G, \qquad Y_P^G \in \mathfrak{a}_P^G,\ Y_G \in \mathfrak{a}_G,$$

where $Y_G$ is independent of the choice of $P \in \mathcal{P}(M)$. Subtracting the fixed point $Y_G \in \mathfrak{a}_G$ from each $Y_P$, we can assume that $Y_P \in \mathfrak{a}_M^G$. Now in §11, we attached a sign $\varepsilon_P(\Lambda)$ and a characteristic function $\phi_P(\Lambda,\cdot)$ on $\mathfrak{a}_M$ to each $P \in \mathcal{P}(M)$ and $\Lambda \in \mathfrak{a}_M$. Suppose that $\Lambda$ is in general position, and that $\lambda$ is any point in $\mathfrak{a}_{M,\mathbb{C}}^*$ whose real part equals $\Lambda$. The function

$$\varepsilon_P(\Lambda)\phi_P(\Lambda, H - Y_P)\mathrm{e}^{\lambda(H)}, \qquad H \in \mathfrak{a}_M^G,$$

is then rapidly decreasing. By writing

$$H = \sum_{\alpha\in\Delta_P} t_\alpha \alpha^\vee, \qquad t_\alpha \in \mathbb{R},$$

we deduce easily that the integral of this function over $H$ equals

$$e^{\lambda(Y_P)}\theta_P(\lambda)^{-1} = c_P(\lambda,\mathcal{Y})\theta_P(\lambda)^{-1}.$$

It then follows that

$$\sum_{P\in\mathcal{P}(M)} e^{\lambda(Y_P)}\theta_P(\lambda)^{-1} = \int_{\mathfrak{a}_M^G} \psi_M(H,\mathcal{Y})e^{\lambda(H)}dH, \tag{17.3}$$

where

$$\psi_M(H,\mathcal{Y}) = \sum_{P\in\mathcal{P}(M)} \varepsilon_P(\Lambda)\phi_P(\Lambda, H-Y_P).$$

LEMMA 17.2. *The function*

$$H \longrightarrow \psi_M(H,\mathcal{Y}), \qquad H\in\mathfrak{a}_M^G,$$

*is the characteristic function of the convex hull in* $\mathfrak{a}_M^G$ *of* $\mathcal{Y}$.

The main step in the proof of Lemma 17.2 is the combinatorial lemma of Langlands mentioned at the end of §8. This result asserts that

$$\sum_{Q\supset P} \varepsilon_P^Q(\Lambda)\phi_P^Q(\Lambda,H)\tau_Q(H) = \begin{cases} 1, & \text{if } \Lambda(\alpha^\vee)>0,\ \alpha\in\Delta_P,\\ 0, & \text{otherwise,}\end{cases} \tag{17.4}$$

for any $P\in\mathcal{P}(M)$ and $H\in\mathfrak{a}_M$, where $\varepsilon_P^Q$ and $\phi_P^Q$ denote objects attached to the parabolic subgroup $P\cap M_Q$ of $M_Q$. Langlands's geometric proof of (17.4) was reproduced in [**A1**, §2]. There is a different combinatorial proof [**A3**, Corollary 6.3], which combines an induction argument with (8.10) and Identity 6.2. Given the formula (17.4), one then observes that $\psi_M(H,\mathcal{Y})$ is independent of the point $\Lambda$. This follows inductively from the expression obtained by summing the left hand side of (17.4) over $P\in\mathcal{P}(M)$ [**A1**, Lemma 3.1]. Finally, by varying $\Lambda$, one shows that

$$\psi_M(H,\mathcal{Y}) = \begin{cases} 1, & \text{if } \varpi(H-Y_P)\le 0,\ \varpi=\widehat{\Delta}_P,\ P\in\mathcal{P}(M),\\ 0, & \text{otherwise.}\end{cases}$$

The inequalities on the right characterize the convex hull of $\mathcal{Y}$, according to the Krein-Millman theorem. (See [**A1**, Lemma 3.2].) □

The convex hull of $\mathcal{Y}$ is of course compact. It follows that the integral on the right hand side of (17.3) converges absolutely, uniformly for $\lambda\in i\mathfrak{a}^*_{M,\mathbb{C}}$. We can therefore identify the smooth function $c_M(\lambda,\mathcal{Y})$ with the Fourier transform of the characteristic function of the convex hull of $\mathcal{Y}$. Its value $c_M(\mathcal{Y})$ at $\lambda=0$ is simply the volume of the convex hull. We have actually been assuming that the point $Y_G\in\mathfrak{a}_G$ attached to $\mathcal{Y}$ equals zero. However, if $Y_G$ is nonzero, the convex hull of $\mathcal{Y}$ represents a compactly supported distribution in the affine subspace $Y_G+\mathfrak{a}_M^G$ of $\mathfrak{a}_M$. The last two assertions therefore remain valid for any $\mathcal{Y}$.

Consider the case that $G=SL(3)$ and $M$ equals the standard minimal Levi subgroup. The convex hull of a typical set $\mathcal{Y}$ is illustrated in Figure 17.1, a diagram on which one could superimpose six convex cones, as in the earlier special case of Figure 11.1. The six points $Y_P$ are the six vertices in the diagram. We have chosen them here to lie in the associated chambers $\mathfrak{a}_P^+$. Notice that with this condition, the intersection of the convex hull with the closure of a chamber $\mathfrak{a}_P^+$ equals a set of the kind illustrated in Figure 9.2. This suggests that the characteristic function $\psi_M(H,\mathcal{Y})$ is closely related to the functions $\Gamma'_P(\cdot,Y_P)$ defined in §9.

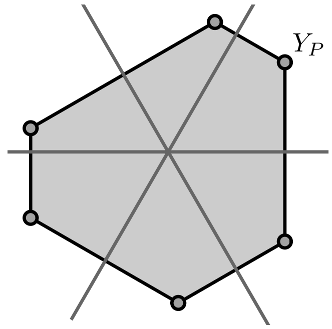



FIGURE 17.1. The convex hull of six points $\{Y_P\}$ in the two dimensional space $\mathfrak{a}_0$ attached to $SL(3)$. Observe that its intersection with any of the six chambers $\mathfrak{a}_P^+$ in the diagram is a region like that in Figure 9.2.

Suppose that $X$ is any point in $\mathfrak{a}_M^G$. According to Lemma 9.2, the function $H \to \Gamma'_P(H, X)$ on $\mathfrak{a}_M^G$ is compactly supported for any $P \in \mathcal{P}(M)$. The integral

$$\int_{\mathfrak{a}_M^G} \Gamma'_P(H, X) e^{\lambda(H)} dH \tag{17.5}$$

therefore converges uniformly to an analytic function of $\lambda \in \mathfrak{a}_{M,\mathbb{C}}^*$. To compute it, we first note that for any $P \in \mathcal{P}(M)$,

$$\begin{aligned} &\sum_{Q \supset P} (-1)^{\dim(A_Q/A_G)} \tau_P^Q(H) \widehat{\tau}_Q(H - X) \\ =&\sum_{Q \supset P} (-1)^{\dim(A_Q/A_G)} \tau_P^Q(H) \sum_{Q' \supset Q} (-1)^{\dim(A_{Q'}/A_G)} \widehat{\tau}_Q^{Q'}(H) \Gamma'_{Q'}(H, X) \\ =&\sum_{Q' \supset P} \Big( \sum_{\{Q : P \subset Q \subset Q'\}} (-1)^{\dim(A_Q/A_{Q'})} \tau_P^Q(H) \widehat{\tau}_Q^{Q'}(H) \Big) \Gamma'_Q(H, X) \\ =&\Gamma'_P(H, X), \end{aligned}$$

by the inductive definition (9.1) and the formula (8.10). Suppose that the real part of $\lambda$ lies in the negative chamber $-(\mathfrak{a}_M^*)_P^+$. Then the integral

$$\int_{\mathfrak{a}_M^G} \tau_P^Q(H) \widehat{\tau}_Q(H - X) e^{\lambda(H)} dH$$

converges. Changing variables by writing

$$H = \sum_{\varpi \in \widehat{\Delta}_P^Q} t_\varpi \varpi^\vee + \sum_{\alpha \in \Delta_Q} t_\alpha \alpha^\vee, \qquad t_\varpi, t_\alpha \in \mathbb{R},$$

one sees without difficulty that the integral equals

$$(-1)^{\dim(A_P/A_G)} e^{\lambda_Q(X)} \widehat{\theta}_P^Q(\lambda)^{-1} \theta_Q(\lambda_Q)^{-1},$$

where $\lambda_Q$ is the projection of $\lambda$ onto $\mathfrak{a}_{Q,\mathbb{C}}^*$, and

$$\widehat{\theta}_P^Q(\lambda) = \widehat{\theta}_{P \cap M_Q}(\lambda) = \mathrm{vol}\big(\mathfrak{a}_P^Q / \mathbb{Z}\big((\widehat{\Delta}_P^Q)^\vee\big)\big)^{-1} \prod_{\varpi \in \widehat{\Delta}_P^Q} \lambda(\varpi^\vee).$$

(See [**A5**, p. 15].) It follows that the original integral (17.5) equals

$$\sum_{Q\supset P}(-1)^{\dim(A_P/A_Q)}\mathrm{e}^{\lambda_Q(X)}\widehat{\theta}_P^Q(\lambda)^{-1}\theta_Q(\lambda_Q)^{-1}. \tag{17.6}$$

In particular, the function (17.6) extends to an analytic function of $\lambda\in\mathfrak{a}^*_{M,\mathbb{C}}$.

Suppose now that for a given $P\in\mathcal{P}(M)$, $c_P(\lambda)$ is an arbitrary smooth function of $\lambda\in i\mathfrak{a}^*_M$. Motivated by the computation above, we set

$$c'_P(\lambda)=\sum_{Q\supset P}(-1)^{\dim(A_P/A_Q)}c_Q(\lambda_Q)\widehat{\theta}_P^Q(\lambda)^{-1}\theta_Q(\lambda_Q)^{-1}, \tag{17.7}$$

where $c_Q$ is the restriction of $c_P$ to $i\mathfrak{a}^*_Q$, and $\lambda_Q$ is again the projection of $\lambda$ onto $i\mathfrak{a}^*_Q$. Then $c'_P$ is defined on the complement of a finite set of hyperplanes in $i\mathfrak{a}^*_M$.

Lemma 17.3. *$c'_Q(\lambda)$ extends to a smooth function of $\lambda\in i\mathfrak{a}^*_M$.*

The lemma is not surprising, given what we have established in the special case that $c_P(\lambda)=\mathrm{e}^{\lambda(X)}$. One can either adapt the discussion above to the more general case, as in [**A3**, Lemma 6.1], or approximate $c_P(\lambda)$ by functions of the form $\mathrm{e}^{\lambda(X)}$, and apply the results above directly. □

Assume now that $\{c_P(\lambda) : P\in\mathcal{P}(M)\}$ is a general $(G,M)$-family. There are two restriction operations that give rise to two new families. Suppose that $Q\in\mathcal{F}(M)$. If $R$ belongs to $\mathcal{P}^{M_Q}(M)$, we set

$$c_R^Q(\lambda)=c_{Q(R)}(\lambda),$$

where $Q(R)$ is the unique group in $\mathcal{P}(M)$ that is contained in $Q$, and whose intersection with $M_Q$ equals $R$. Then $\{c_R^Q(\lambda) : R\in\mathcal{P}^{M_Q}(M)\}$ is an $(M_Q,M)$-family. The other restriction operation applies to a given group $L\in\mathcal{L}(M)$. If $\lambda$ lies in the subspace $i\mathfrak{a}^*_L$ of $i\mathfrak{a}^*_M$, and $Q$ is any group in $\mathcal{P}(L)$, we set

$$c_Q(\lambda)=c_P(\lambda),$$

for any group $P\in\mathcal{P}(M)$ with $P\subset Q$. Since we started with a $(G,M)$-family, this function is independent of the choice of $P$, and the resulting collection $\{c_Q(\lambda) : Q\in\mathcal{P}(L)\}$ is a $(G,L)$-family. Observe that the definition (17.7) can be applied to any $Q$. It yields a smooth function $c'_Q(\lambda)$ on $i\mathfrak{a}^*_L$ that depends only on $c_Q(\lambda)$. Again, we often write $d'_Q=d'_Q(0)$ for the value of $d'_Q(\lambda)$ at $\lambda=0$.

Let $\{d_P(\lambda) : P\in\mathcal{P}(M)\}$ be a second $(G,M)$-family. Then the pointwise product

$$(cd)_P(\lambda)=c_P(\lambda)d_P(\lambda),\qquad P\in\mathcal{P}(M),$$

is also a $(G,M)$-family.

Lemma 17.4. *The product $(G,M)$-family satisfies the splitting formula*

$$(cd)_M(\lambda)=\sum_{Q\in\mathcal{P}(M)}c_M^Q(\lambda)d'_Q(\lambda_Q).$$

*In particular the values at $\lambda=0$ of the functions in the formula satisfy*

$$(cd)_M=\sum_{Q\in\mathcal{F}(M)}c_M^Qd'_Q. \tag{17.8}$$

The lemma is an easy consequence of a formula

$$c_P(\lambda)\theta_P(\lambda)^{-1} = \sum_{Q\supset P} c'_Q(\lambda_Q)\theta_P^Q(\lambda)^{-1}, \qquad P \in \mathcal{P}(M), \tag{17.9}$$

where $\theta_P^Q = \theta_{P\cap M_Q}$, which we obtain by inverting the definition (17.7). To derive (17.9), we write

$$\begin{aligned}
&\sum_{Q\supset P} c'_Q(\lambda_Q)\theta_P^Q(\lambda)^{-1}\\
&= \sum_{Q\supset P}\sum_{Q'\supset Q}(-1)^{\dim(A_Q/A_{Q'})}c_{Q'}(\lambda_{Q'})\widehat{\theta}_Q^{Q'}(\lambda_Q)^{-1}\theta_{Q'}(\lambda_{Q'})^{-1}\theta_P^Q(\lambda)^{-1}\\
&= \sum_{Q'\supset P} c_{Q'}(\lambda_{Q'})\theta_{Q'}(\lambda_{Q'})^{-1}\Big(\sum_{\{Q:P\subset Q\subset Q'\}}(-1)^{\dim(A_Q/A_{Q'})}\theta_P^Q(\lambda)^{-1}\widehat{\theta}_Q^{Q'}(\lambda_Q)^{-1}\Big).
\end{aligned}$$

The expression in the brackets may be written as a Fourier transform

$$\int_{\mathfrak{a}_P^{Q'}}\Big(\sum_{\{Q:P\subset Q\subset Q'\}}(-1)^{\dim(A_P/A_Q)}\widehat{\tau}_P^Q(H)\tau_Q^{Q'}(H)\Big)e^{\lambda(H)}dH,$$

provided that the real part of $\lambda$ lies in $-(\mathfrak{a}_M^*)_P^+$. The identity (8.11) tells us that the expression equals 0 or 1, according to whether $Q'$ properly contains $P$ or not. The formula (17.9) follows. Once we have (17.9), we see that

$$\begin{aligned}
(cd)_M(\lambda) &= \sum_{P\in\mathcal{P}(M)} c_P(\lambda)d_P(\lambda)\theta_P(\lambda)^{-1}\\
&= \sum_P c_P(\lambda)\sum_{Q\supset P} d'_Q(\lambda_Q)\theta_P^Q(\lambda)^{-1}\\
&= \sum_{Q\in\mathcal{F}(M)}\Big(\sum_{\{P\in\mathcal{P}(M):P\subset Q\}} c_P(\lambda)\theta_P^Q(\lambda)^{-1}\Big)d'_Q(\lambda_Q)\\
&= \sum_{Q\in\mathcal{F}(M)} c_M^Q(\lambda)d'_Q(\lambda_Q),
\end{aligned}$$

as required. (See [**A3**, Lemma 6.3].) □

Suppose for example that $c_P(\lambda) = 1$ for each $P$ and $\lambda$. This is the family attached to the trivial positive $(G, M)$-orthogonal set $\mathcal{Y} = 0$. Then $c_M^Q(\lambda)$ equals 0 unless $Q$ lies in the subset $\mathcal{P}(M)$ of $\mathcal{F}(M)$, in which case it equals 1. It follows that

$$d_M(\lambda) = \sum_{P\in\mathcal{P}(M)} d'_P(\lambda). \tag{17.10}$$

In the case that $d_P(\lambda)$ is of the special form (17.2), this formula matches the intuition we obtained from Figure 17.1 and Figure 9.2. For general $\{d_P(\lambda)\}$, and for $\{c_P(\lambda)\}$ subject only to a supplementary condition that the numbers

$$c_M^L = c_M^Q, \qquad L \in \mathcal{L}(M),\ Q \in \mathcal{P}(L), \tag{17.11}$$

be independent of the choice of $Q$, (17.10) can be applied to the splitting formula (17.8). We obtain a simpler splitting formula

$$(cd)_M = \sum_{L\in\mathcal{L}(M)} c_M^L d_L \tag{17.12}$$

Suppose that $\{c_P(\lambda)\}$ and $\{d_P(\lambda)\}$ correspond to positive $(G, M)$-orthogonal sets $\mathcal{Y} = \{Y_P\}$ and $\mathcal{Z} = \{Z_P\}$. Then the product family $\{(cd)_P(\lambda)\}$ corresponds to the sum $\mathcal{Y} + \mathcal{Z} = \{Y_P + Z_P\}$. In this case, (17.8) is similar to a classical formula for mixed volumes. In the case that $G = SL(3)$ and $M$ is minimal, it is illustrated in Figure 17.2.

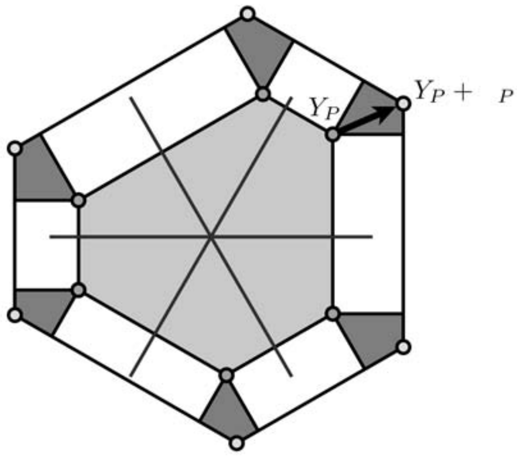



FIGURE 17.2. The entire region is the convex hull of six points $\{Y_P + Z_P\}$ in the two dimensional space $\mathfrak{a}_0$ attached to $SL(3)$. The inner shaded region is the convex hull of the six points $\{Y_P\}$. For any $P$, the area of the darker shaded region with vertex $Y_P$ equals the area of a region in Figure 9.2. The areas of the six rectangular regions represent mixed volumes between the sets $\{Y_P\}$ and $\{Z_P\}$.

In addition to the splitting formula (17.8), there is a descent formula that relates the two restriction operations we have defined. It applies in fact to a generalization of the second operation.

Suppose that $M$ contains a Levi subgroup $M_1$ of $G$ over some extension $k_1$ of $k$. Then $\mathfrak{a}_M$ is contained in the vector space $\mathfrak{a}_{M_1}$ attached to $M_1$. Suppose that $\{c_{P_1}(\lambda_1) : \ P_1 \in \mathcal{P}(M_1)\}$ is a $(G_1, M_1)$-family. If $P$ belongs to $\mathcal{P}(M)$ and $\lambda$ lies in the subspace $i\mathfrak{a}_M^*$ of $i\mathfrak{a}_{M_1}^*$, we set

$$c_P(\lambda) = c_{P_1}(\lambda)$$

for any $P_1 \in \mathcal{P}(M_1)$ with $P_1 \subset P$. This function is independent of the choice of $P_1$, and the resulting collection $\{c_P(\lambda) : \ P \in \mathcal{P}(M)\}$ is a $(G, M)$-family. We would like to express the supplementary function $c_M(\lambda)$ in terms of corresponding functions $c_{M_1}^{Q_1}(\lambda_1)$ attached to groups $Q_1 \in \mathcal{F}(M_1)$. A necessary step is of course to fix Haar measures on each of the spaces $\mathfrak{a}_{M_1}^{L_1}$, as $L_1 = L_{Q_1}$ ranges over $\mathcal{L}(M_1)$. For example, we could fix a suitable Euclidean inner product on the space $\mathfrak{a}_{M_1}$, and then take the Haar measure on $\mathfrak{a}_{M_1}^{L_1}$ attached to the restricted inner product. For each $L_1$, we introduce a nonnegative number $d_{M_1}^G(M, L_1)$ to make the relevant measures compatible. We define $d_{M_1}^G(M, L_1)$ to be 0 unless the natural map

$$\mathfrak{a}_{M_1}^M \oplus \mathfrak{a}_{M_1}^{L_1} \longrightarrow \mathfrak{a}_{M_1}^G$$

is an isomorphism, in which case $d^G_{M_1}(M, L_1)$ is the factor by which the product Haar measure on $\mathfrak{a}^L_{M_1} \oplus \mathfrak{a}^{L_1}_{M_1}$ must be multiplied in order to be equal to the Haar measure on $\mathfrak{a}^G_{M_1}$. (The measure on $\mathfrak{a}^M_{M_1}$ is the quotient of the chosen measures on $\mathfrak{a}^G_{M_1}$ and $\mathfrak{a}^G_M$.)

There is one other choice to be made. Given $M$ and $M_1$, we select a small vector $\xi$ in general position in $\mathfrak{a}^M_{M_1}$. If $L_1$ is any group $\mathcal{L}(M_1)$ with $d^G_{M_1}(M, L_1) \neq 0$, the affine space $\xi + \mathfrak{a}^G_M$ intersects $\mathfrak{a}^G_{L_1}$ at one point. This point is nonsingular, and so belongs to a chamber $\mathfrak{a}^+_{Q_1}$, for a unique group $Q_1 \in \mathcal{P}(L_1)$. The point $\xi$ thus determines a section

$$L_1 \longrightarrow Q_1, \qquad\qquad L_1 \in \mathcal{L}(M_1),\ d^G_{M_1}(M, L_1) \neq 0,$$

from $L_1$ to its fibre $\mathcal{P}(L_1)$.

LEMMA 17.5. *For $F_1 \supset F$, $M_1 \subset M$, and $\{c_{P_1}(\lambda_1)\}$ as above, we have*

$$c_M(\lambda) = \sum_{L_1 \in \mathcal{L}(M_1)} d^G_{M_1}(M, L_1) c^{Q_1}_{M_1}(\lambda), \qquad\qquad \lambda \in i\mathfrak{a}^*_M.$$

*In particular, the values at $\lambda = 0$ of these functions satisfy*

$$c_M = \sum_{L_1 \in \mathcal{L}(M_1)} d^G_{M_1}(M, L_1) c^{Q_1}_{M_1}. \tag{17.13}$$

Lemma 17.5 is proved under slightly more general conditions in [**A13**, Proposition 7.1]. We shall be content to illustrate it geometrically in a very special case. Suppose that $k = k_1$, $G = SL(3)$, $M$ is a maximal Levi subgroup, $M_1$ is a minimal Levi subgroup, and $\{c_{P_1}(\lambda_1)\}$ is of the special form (17.2). The points $\{Y_{P_1}\}$ are the six vertices of the polytope in Figure 17.3. They are of course bijective with the set of minimal parabolic subgroups $P_1 \in \mathcal{P}(M_1)$. The six edges in the polytope are bijective with the six maximal parabolic subgroups $Q_1 \in \mathcal{F}(M_1)$. The two vertical edges are perpendicular to $\mathfrak{a}_M$, so the corresponding coefficients $d^G_{M_1}(M, L_1)$ vanish. The remaining four edges occur in pairs, corresponding to two pairs of groups $Q_1 \in \mathcal{P}(L_1)$ attached to the two maximal Levi subgroups $L_1 \neq M$. However, the upward pointing vector $\xi \in \mathfrak{a}^M_{M_1}$ singles out the upper two edges. The projections of these two edges onto the line $\mathfrak{a}_M$ are disjoint (apart from the interior vertex), with union equal to the line segment obtained by intersecting $\mathfrak{a}_M$ with the polytope. The length of this line segment is the sum of the lengths of the two upper edges, scaled in each case by the associated coefficient $d^G_{M_1}(M, L_1)$.

If this simple example is not persuasive, the reader could perform some slightly more complicated geometric experiments. Suppose that $\dim(\mathfrak{a}_{M_1}) = 3$ and $\{c_{P_1}(\lambda_1)\}$ is of the special form (17.2), but that $k$, $G$, $k_1$, and $M_1$ are otherwise arbitrary. It is interesting to convince oneself geometrically of the validity of the lemma in the two cases $\dim\, \mathfrak{a}_M = 1$ and $\dim\, \mathfrak{a}_M = 2$. The motivation for the general proof is based on these examples. □

We sometimes use a variant of Lemma 17.5, which is included in the general formulation of [**A13**, Propositon 7.1]. It concerns the case that $F = F_1$, but where $M$ is embedded diagonally in the Levi subgroup $\mathcal{M} = M \times M$ of $\mathcal{G} = G \times G$. Then $\mathfrak{a}_M$ is embedded diagonally in the space $\mathfrak{a}_{\mathcal{M}} = \mathfrak{a}_M \oplus \mathfrak{a}_M$. Elements in $\mathcal{L}(\mathcal{M})$ consist of pairs $\mathcal{L} = (L_1, L_2)$, for Levi subgroups $L_1, L_2 \in \mathcal{L}(M)$ of $G$. (We have written

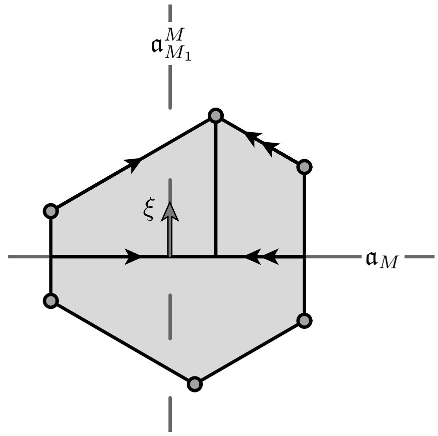



FIGURE 17.3. An illustration of the proof of Lemma 17.5, with $G = SL(3)$, $M$ maximal, and $M_1 = M_0$ minimal. The two upper edges of the polytope project onto the two interior intervals on the horizontal axis. In each case, the projection contracts the length by the appropriate determinant $d^G_{M_1}(M, L_1)$.

$\mathcal{M}$, $\mathcal{G}$, and $\mathcal{L}$ in place of $M_1$, $G_1$, and $L_1$, since we are now using $L_1$ to denote the first component of $\mathcal{L}$.) The corresponding coefficient in (17.13) satisfies

$$d^{\mathcal{G}}_{\mathcal{M}}(M, \mathcal{L}) = 2^{\frac{1}{2}\dim(\mathfrak{a}^G_M)} d^G_M(L_1, L_2),$$

while if $P$ belongs to $\mathcal{P}(M)$, the pair $\mathcal{P} = (P, P)$ in $\mathcal{P}(\mathcal{M})$ satisfies

$$\theta_{\mathcal{P}}(\lambda) = 2^{\frac{1}{2}\dim(\mathfrak{a}^G_M)} \theta_P(\lambda), \qquad \lambda \in i\mathfrak{a}^*_M.$$

We choose a small point $\xi$ in general position in the space

$$\mathfrak{a}^M_{\mathcal{M}} = \{(H, -H) : \ H \in \mathfrak{a}_M\},$$

and let

$$(L_1, L_2) \ \longrightarrow \ (Q_1, Q_2), \qquad L_1, L_2 \in \mathcal{L}(M), \ d^G_M(L_1, L_2) \neq 0,$$

be the corresponding section from $(L_1, L_2)$ to its fibre $\mathcal{P}(L_1) \times \mathcal{P}(L_2)$. If $\xi$ is written in the form $\frac{1}{2}\xi_1 - \frac{1}{2}\xi_2$, $Q_i$ is in fact the group in $\mathcal{P}(L_i)$ such that $\xi_i$ belongs to $\mathfrak{a}^+_{Q_i}$.

LEMMA 17.6. *The product $(G, M)$-family of Lemma 17.4 satisfies the alternate splitting formula*

$$(cd)_M(\lambda) = \sum_{L_1, L_2 \in \mathcal{L}(M)} d^G_M(L_1, L_2) c^{Q_1}_M(\lambda) c^{Q_2}_M(\lambda).$$

*In particular, the values at $\lambda = 0$ of the functions in the formula satisfy*

$$(cd)_M = \sum_{L_1, L_2 \in \mathcal{L}(M)} d^G_M(L_1, L_2) c^{Q_1}_M d^{Q_1}_M. \tag{17.14}$$

(See [**A13**, Corollary 7.4].) □

## 18. Local behaviour of weighted orbital integrals

We now consider the refinement of the coarse geometric expansion (10.7). In this section, we shall construct the general weighted orbital integrals that are to be the local ingredients. In the next section, we shall describe how to expand $J(f)$ as a linear combination of weighted orbital integrals, with certain global coefficients.

Recall that invariant orbital integrals (1.4) arose naturally at the beginning of the article. Weighted orbital integrals are noninvariant analogues of these distributions. We define them by scaling the invariant measure $dx$ with a function $v_M(x)$ obtained from a certain $(G, M)$-family.

The simplest case concerns the setting at the end of §16, in which $k$ is a completion $F_v$ of $F$. Then $M$ is a Levi subgroup of $G$ over $F_v$. We also have to fix a suitable maximal compact subgroup $K_v$ of $G(k) = G(F_v)$. If $x_v$ is an element in $G(F_v)$, and $P$ belongs to $\mathcal{P}(M)$, we form the point $H_P(x_v)$ in $\mathfrak{a}_M$ as in §4. It is a consequence of the definitions that

$$\{Y_P = -H_P(x_v) : \ P \in \mathcal{P}(M)\}.$$

is a positive $(G, M)$-orthogonal set. The functions

$$v_P(\lambda, x_v) = \mathrm{e}^{-\lambda(H_P(x_v))}, \qquad \lambda \in i\mathfrak{a}_M^*, \ P \in \mathcal{P}(M),$$

then form a $(G, M)$-family. The associated smooth function

$$v_M(\lambda, x_v) = \sum_{P\in\mathcal{P}(M)} v_P(\lambda, x_v)\theta_P(\lambda)^{-1}$$

is the Fourier transform of the characteristic function of the convex hull in $\mathfrak{a}_M^G$ of the projection onto $\mathfrak{a}_M^G$ of the points $\{-H_P(x_v) : \ P \in \mathcal{P}(M)\}$. The number

$$v_M(x_v) = v_M(0, x_v) = \lim_{\lambda\to 0} \sum_{P\in\mathcal{P}(M)} v_P(\lambda, x_v)\theta_P(\lambda)^{-1}$$

equals the volume of this convex hull.

For the trace formula, we need to consider the global case that $k = F$. Until further notice, the maximal compact subgroup $K = \prod K_v$ of $G(\mathbb{A})$ will remain fixed. Suppose that $M$ is a Levi subgroup in the finite set $\mathcal{L} = \mathcal{L}(M_0)$, and that $x$ belongs to $G(\mathbb{A})$. The collection

$$v_P(\lambda, x) = \mathrm{e}^{-\lambda(H_P(x))}, \qquad \lambda \in i\mathfrak{a}_M^*, \ P \in \mathcal{P}(M), \tag{18.1}$$

is then a $(G, M)$-family of functions. The limit

$$v_M(x) = \lim_{\lambda\to 0} \sum_{P\in\mathcal{P}(M)} v_P(\lambda, x)\theta_P(\lambda)^{-1} \tag{18.2}$$

exists and equals the volume of the convex hull in $\mathfrak{a}_M^G$ of the projection of the points $\{-H_P(x) : \ P \in \mathcal{P}(M)\}$. To see how this function is related to the discussion of §11, choose a parabolic subgroup $P \in \mathcal{P}(M)$, and a minimal parabolic subgroup $P_0$ of $G$ over $\mathbb{Q}$ that is contained in $P$. The correspondence

$$(P', s) \ \longrightarrow \ Q = w_s^{-1} P' w_s, \qquad P' \supset P_0, \ s \in W(\mathfrak{a}_P, \mathfrak{a}_{P'}),$$

is then a bijection from the disjoint union over $P'$ of the sets $W(\mathfrak{a}_P, \mathfrak{a}_{P'})$ onto the set $\mathcal{P}(M)$, with the property that

$$s^{-1} H_{P'}(\widetilde{w}_s x) = H_Q(x).$$

It follows that $v_M(x)$ equals the weight function $v_P(x)$ of Theorem 11.2.

The local and global cases are of course related. For any $x \in G(\mathbb{A})$, we can write

$$H_P(x) = \sum_v H_P(x_v), \qquad P \in \mathcal{P}(M),$$

where $x_v$ is the component of $x$ in $G(F_v)$. For almost every valuation $v$, $x_v$ lies in $K_v$, and $H_P(x_v) = 0$. We obtain a finite sum

$$H_P(x) = \sum_{v \in S} H_P(x_v),$$

where $S$ is a finite set of valuations that contains the set $S_\infty$ of archimedean valuations. We may therefore fix $S$, and take $x$ to be a point in the product

$$G(F_S) = \prod_{v \in S} G(F_v).$$

The $(G, M)$-family $\{v_P(\lambda, x)\}$ decomposes into a pointwise product

$$v_P(\lambda, x) = \prod_{v \in S} v_P(\lambda, x_v), \qquad \lambda \in i\mathfrak{a}_M^*,\ P \in \mathcal{P}(M),$$

of $(G, M)$-families $\{v_P(\lambda, x_v)\}$. We can therefore use the splitting formula (17.14) and the descent formula (17.13) (with $k = F$ and $k_1 = F_v$) to express the volume $v_M(x)$ in terms of volumes associated to the points $x_v \in G(F_v)$.

We fix the Levi subgroup $M$ of $G$ over $F$. We also fix an arbitrary finite set $S$ of valuations, and write $K_S = \prod_{v \in S} K_v$ for the maximal compact subgroup of $G(F_S)$. Suppose that $\gamma = \prod \gamma_v$ is an element in $M(F_S)$. Our goal is to construct a weighted orbital integral of a function $f \in C_c^\infty\big(G(F_S)\big)$ over the space of $F_S$-valued points in the conjugacy class of $G$ induced from $\gamma$. More precisely, let $\gamma^G$ be the union of those conjugacy classes in $G(F_S)$ that for any $P \in \mathcal{P}(M)$ intersect $\gamma N_P(F_S)$ in a nonempty open set. We shall define the weighted orbital integral attached to $M$ and $\gamma$ by means of a canonical, noninvariant Borel measure on $\gamma^G$.

For any $v$, the connected centralizer $G_{\gamma_v}$ is an algebraic group over $F_v$. We regard the product $G_\gamma = \prod_{v \in S} G_{\gamma_v}$ as a scheme over $F_S$, which is to say simply that

$$G_\gamma(F_S) = \prod_{v \in S} G_{\gamma_v}(F_v).$$

It is known [**R**] that this group is unimodular, and hence that there is a right invariant measure $dx$ on the quotient $G_\gamma(F_S)\backslash G(F_S)$. The correspondence $x \to x^{-1}\gamma x$ is a surjective mapping from $G_\gamma(F_S)\backslash G(F_S)$ onto the conjugacy class of $\gamma$ in $G(F_S)$, with finite fibres (corresponding to the connected components in the full centralizer $G_{\gamma,+}(F_S)$). Now if $\gamma$ is not semisimple, the preimage in $G_\gamma(F_S)\backslash G(F_S)$ of a compact subset of the conjugacy class of $\gamma$ (in the topology induced from $G(F_S)$) need not be compact. Nevertheless, a theorem of Deligne and Rao [**R**] asserts that the measure $dx$ defines a $G(F_S)$-invariant Borel measure on the conjugacy class of $\gamma$. We obtain a continuous $G(F_S)$-invariant linear form

$$f \longrightarrow \int_{G_\gamma(F_S)\backslash G(F_S)} f(x^{-1}\gamma x)\mathrm{d}x, \qquad f \in C_c^\infty(F_S),$$

on $C_c^\infty\big(G(F_S)\big)$.

Suppose first that $G_\gamma$ is contained in $M$. In other words, $G_\gamma = M_\gamma$. This condition holds for example if $\gamma$ is the image in $M(F_S)$ of an element in $M(F)$ that represents an unramified class $\mathfrak{o} \in \mathcal{O}$, as in Theorem 11.2. With this condition, we define the weighted orbital integral

$$J_M(\gamma, f) = J_M^G(\gamma, f)$$

of $f \in C_c^\infty\big(G(F_S)\big)$ at $\gamma$ by

$$J_M(\gamma, f) = |D(\gamma)|^{\frac{1}{2}} \int_{G_\gamma(F_S)\backslash G(F_S)} f(x^{-1}\gamma x) v_M(x) \mathrm{d}x. \tag{18.3}$$

The normalizing factor

$$D(\gamma) = D^G(\gamma) = \prod_{v \in S} D^G(\gamma_v)$$

is the generalized Weyl discriminant

$$\prod_{v \in S} \det\big(1 - \mathrm{Ad}(\sigma_v)\big)_{\mathfrak{g}/\mathfrak{g}_{\sigma_v}},$$

where $\sigma_v$ is the semisimple part of $\gamma_v$, and $\mathfrak{g}_{\sigma_v}$ is the Lie algebra of $G_{\sigma_v}$. Its presence in the definition simplifies some formulas. Since $G_\gamma$ is contained in $M$, and $v_M(mx)$ equals $v_M(x)$ for any $m \in M(F_S)$, the integral is well defined.

Lemma 18.1. *Suppose that $y$ is any point in $G(F_S)$. Then*

$$J_M(\gamma, f^y) = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\gamma, f_{Q,y}), \tag{18.4}$$

*where*

$$f_{Q,y}(m) = \delta_Q(m)^{\frac{1}{2}} \int_{K_S} \int_{N_Q(F_S)} f(k^{-1}mnk) u'_Q(k, y) \mathrm{d}n \mathrm{d}k, \tag{18.5}$$

*for $m \in M_Q(F_S)$, and*

$$u'_Q(k, y) = \int_{\mathfrak{a}_Q^G} \Gamma'_Q\big(H, -H_Q(ky)\big) \mathrm{d}H. \tag{18.6}$$

This formula is Lemma 8.2 of [**A5**]. It probably does not come as a surprise, since the global distributions $J_\mathfrak{o}(f)$ satisfy a similar formula (16.2), and Theorem 11.2 tells us that for many $\mathfrak{o}$, $J_\mathfrak{o}(f)$ is a weighted orbital integral.

To prove the lemma, we first write

$$\begin{aligned} J_M(\gamma, f^y) &= |D(\gamma)|^{\frac{1}{2}} \int_{G_\gamma(F_S)\backslash G(F_S)} f(yx^{-1}\gamma x y^{-1}) v_M(x) \mathrm{d}x \\ &= |D(\gamma)|^{\frac{1}{2}} \int_{G_\gamma(F_S)\backslash G(F_S)} f(x^{-1}\gamma x) v_M(xy) \mathrm{d}x. \end{aligned}$$

We then observe that

$$\begin{aligned} v_P(\lambda, xy) = \mathrm{e}^{-\lambda(H_P(xy))} &= \mathrm{e}^{-\lambda(H_P(x))} \mathrm{e}^{-\lambda(H_P(k_P(x)y))} \\ &= v_P(\lambda, x) u_P(\lambda, x, y), \end{aligned}$$

where

$$u_P(\lambda, x, y) = \mathrm{e}^{-\lambda(H_P(k_P(x)y))},$$

and $k_P(x)$ is the point $K_S$ such that $xk_P(x)^{-1}$ belongs to $P(F_S)$. It is a consequence of Lemma 17.4 that

$$v_M(xy) = \sum_{Q\in\mathcal{F}(M)} v_M^Q(x)u'_Q(x,y).$$

If $k$ belongs to $K_S$, it follows from the definition (18.6) of $u'_Q(k,y)$, and the equality of (17.5) with (17.6) established in §17, that $u'_Q(k,y)$ is indeed of the form (17.7). Making two standard changes of variables in the integral over $x$ in $G_\gamma(F_S)\backslash G(F_S)$, we write

$$\begin{aligned}
&|D(\gamma)|^{\frac{1}{2}}\int f(x^{-1}\gamma x)v_M(xy)\mathrm{d}x\\
&= \sum_{Q\in\mathcal{F}(M)} |D(\gamma)|^{\frac{1}{2}}\int f(x^{-1}\gamma x)v_M^Q(x)u'_Q(x,y)\mathrm{d}x\\
&= \sum_Q |D(\gamma)|^{\frac{1}{2}}\int\int\int f(k^{-1}n^{-1}m^{-1}\gamma mnk)v_M^Q(m)u'_Q(k,y)\mathrm{d}m\mathrm{d}n\mathrm{d}k\\
&= \sum_Q |D^M(\gamma)|^{\frac{1}{2}}\delta_Q(\gamma)^{\frac{1}{2}}\int\int\int f(k^{-1}m^{-1}\gamma mnk)v_M^Q(m)u'_Q(k,y)\mathrm{d}n\mathrm{d}k\mathrm{d}m\\
&= \sum_Q |D^M(\gamma)|^{\frac{1}{2}}\int f_{Q,y}(m^{-1}\gamma m)v_M^Q(m)\mathrm{d}m,
\end{aligned}$$

for integrals over $m$, $n$, and $k$ in $M_{Q,\gamma}(F_S)\backslash M_Q(F)$, $N_Q(F_S)$, and $K_S$ respectively. This equals the right hand side of (18.4), as required. □

The distribution (18.3) is to be regarded as a local object, despite the fact that $M$ is a Levi subgroup of $G$ over $F$. It can be reduced to the more elementary distributions

$$J_{M_v}(\gamma_v, f_v), \qquad \gamma_v\in M_v(F_v),\ f_v\in C_c^\infty\big(G(F_v)\big),$$

defined for Levi subgroups $M_v$ of $G$ over $F_v$ by the obvious analogues of (18.3).

Suppose for example that $S$ is a disjoint union of two sets of valuations $S_1$ and $S_2$. Suppose that

$$f = f_1f_2, \qquad f_i\in C_c^\infty\big(G(F_{S_i})\big)$$

and that

$$\gamma=\gamma_1\gamma_2, \qquad \gamma_i\in M(F_{S_i}).$$

We continue to assume that $G_\gamma = M_\gamma$, so that $G_{\gamma_i}=M_{\gamma_i}$ for $i=1,2$. We apply the general splitting formula (17.14) to the $(G,M)$-family

$$v_P(\lambda,x_1,x_2) = v_P(\lambda,x_1)v_P(\lambda,x_2), \qquad P\in\mathcal{P}(M),\ x_i\in G(F_{S_i}).$$

We then deduce from (18.3) that

$$J_M(\gamma,f) = \sum_{L_1,L_2\in\mathcal{L}(M)} d_M^G(L_1,L_2)J_M^{L_1}(\gamma_1,f_{Q_1})J_M^{L_2}(\gamma_2,f_{Q_2}), \tag{18.7}$$

where $(L_1,L_2)\to(Q_1,Q_2)$ is the section in (17.14), and

$$f_{i,Q_i}(m_i) = \delta_{Q_i}(m_i)^{\frac{1}{2}}\int_{K_{S_i}}\int_{N_{Q_i}(F_{S_i})} f_i(k_i^{-1}m_in_ik_i),\mathrm{d}n_i\mathrm{d}k_i,$$

for $m_i \in M_{Q_i}(F_{S_i})$. If we apply this result inductively, we can reduce the compound distributions (18.3) to the simple case that $S$ contains one element.

Suppose that $S$ does consist of one element $v$. Assume that $M_v$ is a Levi subgroup of $G$ over $F_v$, and that $\gamma_v$ is an element in $M_v(F_v)$ with $G_{\gamma_v} = M_{v,\gamma_v}$. Then $M_{v,\gamma_v} = M_{\gamma_v}$ and $M_{\gamma_v} = G_{\gamma_v}$. The first of these conditions implies that the induced class $\gamma_v^M$ equals the conjugacy class of $\gamma_v$ in $M(F_v)$. The second implies that the distribution

$$J_M(\gamma_v^M, f_v) = J_M(\gamma_v, f_v)$$

is defined by (18.3), for any $f_v \in C_c^\infty\big(G(F_v)\big)$. We apply the general descent formula (17.13) to the $(G, M)$-family

$$v_P(\lambda, x_v), \qquad\qquad P \in \mathcal{P}(M),\ x_v \in G(F_v).$$

We then deduce from (18.3) that

$$J_M(\gamma_v^M, f_v) = \sum_{L_v \in \mathcal{L}(M_v)} d_{M_v}^G(M, L_v) J_{M_v}^{L_v}(\gamma_v, f_{v,Q_v}), \tag{18.8}$$

where $L_v \to Q_v$ is the section in (17.13). The two formulas (18.7) and (18.8) together provide the required reduction of (18.3).

Suppose now that $\gamma \in M(F_S)$ is arbitrary. In the most extreme case, for example, $\gamma$ could be the identity element in $M(F_S)$. The problem of defining a weighted orbital integral is now much harder. We cannot form the integral (18.3), since $v_M(x)$ is no longer a well defined function on $G_\gamma(F_S)\backslash G(F_S)$. Nor can we change the domain of integration to $M_\gamma(F_S)\backslash G(F_S)$, since the integral might then not converge.

What we do instead is to replace $\gamma$ by a point $a\gamma$, for a small variable point $a \in A_M(F_S)$ in general position. Then $G_{a\gamma} = M_{a\gamma}$, so we can define $J_M(a\gamma, f)$ by the integral (18.3). The idea is to construct a distribution $J_M(\gamma, f)$ from the values taken by $J_M(a\gamma, f)$ around $a = 1$. This is somewhat subtle. To get an idea of what happens, let us consider the special case of $GL(2)$.

Assume that $F = \mathbb{Q}$, $G = GL(2)$, $M = M_0$ is minimal, $S$ is the archimedean valuation $v_\infty$, and $\gamma = 1$. Then $a\gamma = a = \begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix}$, for distinct positive real numbers $t_1$ and $t_2$. Since

$$G_{a\gamma}(\mathbb{R})\backslash G(\mathbb{R}) = M(\mathbb{R})\backslash P(\mathbb{R})K_\mathbb{R} \cong N_P(\mathbb{R})K_\mathbb{R},$$

where $P$ is the standard Borel subgroup of upper triangular matrices, the integral (18.3) can be written as

$$J_M(a, f) = |D(a)|^{\frac{1}{2}} \int_{K_\mathbb{R}} \int_{N_P(\mathbb{R})} f(k^{-1}n^{-1}ank) v_M(n) \mathrm{d}n \mathrm{d}k. \tag{18.9}$$

It is easy to compute the function $v_M(n)$. We first write

$$\begin{aligned} v_M(n) &= \lim_{\lambda \to 0} \big(\mathrm{e}^{-\lambda(H_P(n))}\theta_P(\lambda)^{-1} + \mathrm{e}^{-\lambda(H_{\overline{P}}(n))}\theta_{\overline{P}}(\lambda)^{-1}\big) \\ &= \lim_{\lambda \to 0} (1 - \mathrm{e}^{-\lambda(H_{\bar{P}}(n))})\theta_P(\lambda)^{-1} \\ &= \lim_{\lambda \to 0} \lambda\big(H_{\overline{P}}(n)\big)\lambda(\alpha^\vee)^{-1}\mathrm{vol}\big(\mathfrak{a}_M^G/\mathbb{Z}(\alpha^\vee)\big) \\ &= \mathrm{e}_1^*\big(H_{\overline{P}}(n)\big), \end{aligned}$$

where $\overline{P}$ is the Borel subgroup of lower triangular matrices, $\alpha$ is the simple root of $(P, A_P)$, $\mathrm{e}_1^*$ is the linear form on $\mathfrak{a}_M \cong \mathbb{R}^2$ defined by projecting $\mathbb{R}^2$ onto the first component, and the measure on $\mathfrak{a}_M^G = \{(H, -H) : \ H \in \mathbb{R}\}$ is defined by Lebesgue measure on $\mathbb{R}$. We then note that $n$ lies in a set $N_{\overline{P}}(\mathbb{R}) \begin{pmatrix} u & 0 \\ 0 & u^{-1} \end{pmatrix} K_{\mathbb{R}}$, for a positive real number $u$, and hence that

$$\mathrm{e}_1^*\big(H_{\overline{P}}(n)\big) = \log |u| = \log \|(1,0)n\|.$$

It follows that if $n = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$, then

$$v_M(n) = \log \|(1,x)\| = \frac{1}{2} \log(1+x^2). \tag{18.10}$$

We make the standard change of variables

$$n \longrightarrow \nu = a^{-1}n^{-1}an = \begin{pmatrix} 1 & x(1-t_1^{-1}t_2) \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} \tag{18.11}$$

in the last integral over $N_P(\mathbb{R})$. This entails multiplying the factor $|D(a)|^{\frac{1}{2}}$ by the Jacobian determinant

$$|D(a)|^{-\frac{1}{2}} \mathrm{e}^{\rho_P(a)} = |D(a)|^{-\frac{1}{2}} (t_1 t_2^{-1})^{\frac{1}{2}}$$

of the transformation. We conclude that $J_M(a, f)$ equals

$$(t_1 t_2^{-1})^{\frac{1}{2}} \int_{K_{\mathbb{R}}} \int_{\mathbb{R}} f\left(k^{-1} \begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix} \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} k\right) \Big(\frac{1}{2} \log\big(1 + \xi^2 (1 - t_1^{-1}t_2)^{-2}\big)\Big) \mathrm{d}\xi \mathrm{d}k.$$

The logarithmic factor in the last expression for $J_M(a, f)$ blows up at $a = 1$. However, we can modify it by adding a logarithmic factor

$$r_M^G(a) = \log |\alpha(a) - \alpha(a)^{-1}| = \log |t_1 t_2^{-1} - t_1^{-1} t_2|$$

that is independent of $\xi$. This yields a locally integrable function

$$\xi \longrightarrow \frac{1}{2} \log\Big((t_1 t_2^{-1} + 1)^2 \big((1 - t_1^{-1}t_2)^2 + \xi^2\big)\Big), \qquad \xi \in \mathbb{R},$$

whose integral over any compact subset of $\mathbb{R}$ is bounded near $a = 1$. Observe that

$$\begin{aligned} &(t_1 t_2^{-1})^{\frac{1}{2}} \int_{K_{\mathbb{R}}} \int_{\mathbb{R}} f\left(k^{-1} \begin{pmatrix} t_1 & 0 \\ 0 & t_2 \end{pmatrix} \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} k\right) \mathrm{d}\xi \mathrm{d}k \\ &= |D(a)|^{\frac{1}{2}} \int_{K_{\mathbb{R}}} \int_{N_P(\mathbb{R})} f(k^{-1}n^{-1}ank) \mathrm{d}n \mathrm{d}k \\ &= J_G(a, f). \end{aligned}$$

It follows from the dominated convergence theorem that the limit

$$\lim_{a \to 1} \big(J_M(a, f) + r_M^G(a) J_G(a, f)\big)$$

exists, and equals the integral

$$J_M(1, f) = \int_{K_{\mathbb{R}}} \int_{\mathbb{R}} f\left(k^{-1} \begin{pmatrix} 1 & \xi \\ 0 & 1 \end{pmatrix} k\right) \log(2|\xi|) \mathrm{d}\xi \mathrm{d}k.$$

This is how we define the weighted orbital integral in the case $G = GL(2)$. As a distribution on $GL(2, \mathbb{R})$, it is given by a noninvariant Borel measure on the conjugacy class $1^G$ of the matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

For arbitrary $F$, $G$, $M$, $S$, and $\gamma$, the techniques are more elaborate. However, the basic method is similar. One begins with the general analogue of the formula (18.9), valid for a fixed group $P \in \mathcal{P}(M)$. One then computes the function $v_M(n)$ as above, using a variable irreducible right $G$-module over $F$ in place of the standard two-dimensional $GL(2)$-module, and a highest weight vector in place of $(1,0)$. If

$$\nu \;\longrightarrow\; n = n(\nu, \gamma a)$$

is the inverse of the bijection $n \to (\gamma a)^{-1} n^{-1} (\gamma a) n$ of $N_P(\mathbb{R})$, the problem becomes that of understanding the behaviour of the function

$$v_M\big(n(\nu, \gamma a)\big)$$

near $a = 1$. This leads to general analogues of the factor $r_M^G(a)$ defined above for $GL(2)$.

THEOREM 18.2. *For any $F$, $G$, $M$, $S$, and $\gamma \in M(F_S)$, there are canonical functions*

$$r_M^L(\gamma, a), \qquad\qquad L \in \mathcal{L}(M),$$

*defined for small points $a \in A_M(F_S)$ in general position, such that the limit*

$$J_M(\gamma, f) = \lim_{a \to 1} \sum_{L \in \mathcal{L}(M)} r_M^L(\gamma, a) J_L(a\gamma, f) \tag{18.12}$$

*exists and equals the integral of $f$ with respect to a Borel measure on the set $\gamma^G$.*

This is Theorem 5.2 of [**A12**], one of the principal results of [**A12**]. There are two basic steps in its proof. The first is construct the functions $r_M^L(\gamma, a)$. The second is to establish the existence and properties of the limit.

The function $r_M^L(\gamma, a)$ is understood to depend only on $L$, $M$, $\gamma$, and $a$ (and not $G$), so we need only construct it when $L = G$. In this case, the function is defined as the limit

$$r_M^G(\gamma, a) = \lim_{\lambda \to 0} \Big( \sum_{P \in \mathcal{P}(M)} r_P(\lambda, \gamma, a) \theta_P(\lambda)^{-1} \Big)$$

associated to a certain $(G, M)$-family

$$r_P(\lambda, \gamma, a) = \prod_{v \in S} \prod_{\beta_v} r_{\beta_v}\Big(\tfrac{1}{2}\lambda, u_v, a_v\Big), \qquad\qquad \lambda \in i\mathfrak{a}_M^*.$$

The factors in this last product are defined in terms of the Jordan decomposition $\gamma_v = \sigma_v u_v$ of the $v$-component of $\gamma$. Let $P_{\sigma_v}$ be the parabolic subgroup $P \cap G_{\sigma_v}$ of $G_{\sigma_v}$. The indices $\beta_v$ then range over the reduced roots of $(P_{\sigma_v}, A_{M_{\sigma_v}})$. Any such $\beta_v$ determines a Levi subgroup $G_{\sigma_v, \beta_v}$ of $G_{\sigma_v}$, and a maximal parabolic subgroup $P_{\sigma_v, \beta_v} = P_{\sigma_v} \cap G_{\sigma_v, \beta_v}$ of $G_{\sigma_v, \beta_v}$ with Levi component $M_{\sigma_v}$. We will not describe the factors in the product further, except to say that they are of the form

$$r_{\beta_v}(\Lambda, u_v, a_v) = |a_v^{\beta_v} - a_v^{-\beta_v}|^{\rho(\beta_v, u_v)\Lambda(\beta_v^\vee)}, \qquad\qquad \Lambda = \tfrac{1}{2}\lambda,$$

for positive constants $\rho(\beta_v, u_v)$, and that they are defined by subjecting $G_{\sigma_v, \beta_v}$, $M_{\sigma_v}$, and $u_v$ to an analysis similar to that of the special case $GL(2)$, $M_0$, and 1 (with $v = v_\infty$) above.

The existence of the limit (18.12) is more subtle. The functions $r_{\beta_v}(\Lambda, u_v, a_v)$ are defined so as to make the associated limits for $G_{\sigma_v, \beta_v}$, $M_{\sigma_v}$, and $u_v$ exist.

However, these limits are simpler. They concern a variable $a_v$ that is essentially one-dimensional (since $M_{\sigma_v}$ is a maximal Levi subgroup in $G_{\sigma_v,\beta_v}$), while the variable $a$ in (18.12) is multidimensional (since $M$ is an arbitrary Levi subgroup of $G$). The existence of the general limit depends on algebraic geometry, specifically a surprising application by Langlands of Zariski's main theorem [**A12**, §4], and some elementary analysis [**A12**, Lemma 6.1]. The fact that the resulting distribution $f \to J_M(\gamma, f)$ is a measure is a consequence of the proof of the existence of the limit. $\square$

Once we have defined the general distributions $J_M(\gamma, f)$, we can extend the properties established in the special case that $G_\gamma = M_\gamma$. First of all, we note that $J_M(\gamma, f)$ depends only on the conjugacy class of $\gamma$ in $M(F_S)$. It is also easy to see from the definition (18.12) that

$$J_{M_1}(\gamma_1, f) = J_M(\gamma, f),$$

where $\gamma_1 = w_s \gamma w_s^{-1} = \widetilde{w}_s \gamma \widetilde{w}_s^{-1}$ and $M_1 = w_s M w_s^{-1}$, for elements $\gamma \in M(F_S)$ and $s \in W_0$.

Suppose that $y$ lies in $G(F_S)$, and that $\gamma \in M(F_S)$ is arbitrary. It then follows from (18.12) and Lemma 18.1 that

$$\begin{aligned} J_M(\gamma, f^y) &= \lim_{a \to 1} \sum_{L \in \mathcal{L}(M)} r_M^L(\gamma, a) J_L(a\gamma, f^y) \\ &= \lim_{a \to 1} \sum_{L \in \mathcal{L}(M)} \sum_{Q \in \mathcal{F}(L)} r_M^L(\gamma, a) J_L^{M_Q}(a\gamma, f_{Q,y}) \\ &= \lim_{a \to 1} \sum_{Q \in \mathcal{F}(M)} \Big( \sum_{L \in \mathcal{L}^{M_Q}(M)} r_M^L(\gamma, a) J_L^{M_Q}(a\gamma, f_{Q,y}) \Big) \\ &= \sum_{Q \in \mathcal{F}(L)} J_M^{M_Q}(\gamma, f_{Q,y}). \end{aligned}$$

The formula (18.4) therefore holds in general.

The splitting formula (18.7) and the descent formula (18.8) also hold in general. In particular, the general distributions $J_M(\gamma, f)$ can be reduced to the more elementary local distributions $J_{M_v}(\gamma_v, f_v)$. The proof entails application to the general definition (18.12) of the special cases of these formulas already established. One has to also apply Lemmas 17.5 and 17.6 to the coefficients $r_M^L(\gamma, a)$ in (18.12). The argument is not difficult, but is more complicated than the general proof of (18.4) above. We refer the reader to the proofs of Theorem 8.1 and Proposition 9.1 of [**A13**].

## 19. The fine geometric expansion

We now turn to the global side of the problem. It would be enough to express the distribution $J_{\mathfrak{o}}(f)$ in explicit terms, for any $\mathfrak{o} \in \mathcal{O}$. We solved the problem for unramified classes $\mathfrak{o}$ in §11 by writing $J_{\mathfrak{o}}(f)$ as a weighted orbital integral. We would like to have a similar formula that applies to an arbitrary class $\mathfrak{o}$.

The general weighted orbital integrals defined in the last section are linear forms on the space $C_c^\infty\big(G(F_S)\big)$, where $S$ is any finite set of valuations. Assume that $S$ is a large finite set that contains the archimedean valuations $S_\infty$, and write $C_c^\infty\big(G(F_S)^1\big)$ for the space of functions on $G(F_S)^1 = G(F_S) \cap G(\mathbb{A})^1$ obtained by restriction of functions in $C_c^\infty\big(G(F_S)\big)$. If $\gamma$ belongs to the intersection of $M(F_S)$

with $G(F_S)^1$, we can obviously define the corresponding weighted orbital integral as a linear form on $C_c^\infty\big(G(F_S)^1\big)$. Let

$$\chi^S = \prod_{v \notin S} \chi_v$$

be the characteristic function of the maximal compact subgroup

$$K^S = \prod_{v \notin S} K_v$$

of $G(\mathbb{A}^S)$. The mapping $f \to f\chi^S$ is then an injection of $C_c^\infty\big(G(F_S)^1\big)$ into $C_c^\infty\big(G(\mathbb{A})^1\big)$. We shall identify $C_c^\infty\big(G(F_S)^1\big)$ with its image in $C_c^\infty\big(G(\mathbb{A})^1\big)$. We can thus form the distribution $J_{\mathfrak{o}}(f)$ for any $f \in C_c^\infty\big(G(F_S)^1\big)$. Our goal is to write it explicitly in terms of weighted orbital integrals of $f$.

Suppose first that $\mathfrak{o}$ consists entirely of unipotent elements. Then $\mathfrak{o} = \mathfrak{o}_{\mathrm{unip}} = \mathcal{U}_G(F)$, where $\mathcal{U}_G$ is the closed variety of unipotent elements in $G$. It is this class in $\mathcal{O}$ that is furthest from being unramified, and which is consequently the most difficult to handle. In general, there are infinitely many $G(F)$-conjugacy classes in $\mathcal{U}_G(F)$. However, we say that two elements $\gamma_1, \gamma_2 \in \mathcal{U}_G(F)$ are $(G,S)$-*equivalent* if they are $G(F_S)$-conjugate. The associated set $\big(\mathcal{U}_G(F)\big)_{G,S}$ of equivalence classes is then finite. The next theorem gives an expansion of the distribution

$$J_{\mathrm{unip}}(f) = J^G_{\mathrm{unip}}(f) = J^G_{\mathfrak{o}_{\mathrm{unip}}}(f)$$

whose terms are indexed by the finite sets $\big(\mathcal{U}_M(F)\big)_{M,S}$.

THEOREM 19.1. *For any $S$ as above, there are uniquely determined coefficients*

$$a^M(S,u), \qquad M \in \mathcal{L},\ u \in \big(\mathcal{U}_M(F)\big)_{M,S},$$

*with*

$$a^M(S,1) = \mathrm{vol}\big(M(F)\backslash M(\mathbb{A})^1\big), \tag{19.1}$$

*such that*

$$J_{\mathrm{unip}}(f) = \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{u \in (\mathcal{U}_M(F))_{M,S}} a^M(S,u) J_M(u,f), \tag{19.2}$$

*for any $f \in C_c^\infty\big(G(F_S)^1\big)$.*

This is the main result, Theorem 8.1, of the paper [**A10**]. The full proof is too long for the space we have here. However, the basic idea is easy to describe.

Assume inductively that the theorem is valid if $G$ is replaced by any proper Levi subgroup. It is understood that the coefficients $a^M(S,u)$ depend only on $M$ (and not $G$). The induction hypothesis therefore implies that the coefficients have been defined whenever $M$ is proper in $G$. We can therefore set

$$T_{\mathrm{unip}}(f) = J_{\mathrm{unip}}(f) - \sum_{\substack{M \in \mathcal{L} \\ M \neq G}} |W_0^M||W_0^G|^{-1} \sum_{u \in (\mathcal{U}_M(F))_{M,S}} a^M(S,u) J_M(u,f),$$

for any $f \in C_c^\infty\big(G(F_S)^1\big)$. Suppose that $y \in G(F_S)$. By (16.2) and (18.4), we can write the difference

$$T_{\mathrm{unip}}(f^y) - T_{\mathrm{unip}}(f)$$

as the difference between the global expression

$$J_{\mathrm{unip}}(f^y) - J_{\mathrm{unip}}(f) = \sum_{\substack{Q \in \mathcal{F} \\ Q \neq G}} |W_0^{M_Q}||W_0^G|^{-1} J_{\mathrm{unip}}^{M_Q}(f_{Q,y})$$

and the local expression

$$\begin{aligned} &\sum_{M \neq G} |W_0^M||W_0^G|^{-1} \sum_{u \in (\mathcal{U}_M(\mathbb{Q}))_{M,S}} a^M(S,u)\big(J_M(u,f^y) - J_M(u,f)\big) \\ &= \sum_{M \neq G} \sum_{\substack{Q \in \mathcal{F}(M) \\ Q \neq G}} |W_0^{M_Q}||W_0^G|^{-1} \sum_u |W_0^M||W_0^{M_Q}|^{-1} a^M(S,u) J_M^{M_Q}(u, f_{Q,y}). \end{aligned}$$

The difference between $T_{\mathrm{unip}}(f^y)$ and $T_{\mathrm{unip}}(f)$ is therefore equal to the sum over $Q \in \mathcal{F}$ with $Q \neq G$ of the product of $|W_0^{M_Q}||W_0^G|^{-1}$ with the expression

$$J_{\mathrm{unip}}^{M_Q}(f_{Q,y}) - \sum_{M \in \mathcal{L}^{M_Q}} |W_0^M||W_0^{M_Q}|^{-1} \sum_{u \in (\mathcal{U}_M(F))_{M,S}} a^M(S,u) J_M^{M_Q}(u, f_{Q,y}).$$

The last expression vanishes by our induction assumption. It follows that $T_{\mathrm{unip}}(f^y)$ equals $T_{\mathrm{unip}}(f)$, and therefore that the distribution $T_{\mathrm{unip}}$ on $G(F_S)^1$ is invariant.

Recall that $J_{\mathrm{unip}}(f)$ is the value at $T = T_0$ of the polynomial

$$J_{\mathrm{unip}}^T(f) = \int_{G(F)\backslash G(\mathbb{A})^1} k_{\mathrm{unip}}^T(x,f)\mathrm{d}x,$$

where

$$k_{\mathrm{unip}}^T(x,f) = \sum_{P \supset P_0} (-1)^{\dim(A_P/A_G)} \sum_{\delta \in P(F)\backslash G(F)} K_{P,\mathrm{unip}}(\delta x, \delta x)\widehat{\tau}_P\big(H_P(\delta x) - T\big)$$

and

$$K_{P,\mathrm{unip}}(\delta x, \delta x) = \sum_{u \in \mathcal{U}_M(F)} \int_{N_P(\mathbb{A})} f(x^{-1}\delta^{-1}un\delta x)\mathrm{d}n.$$

It follows that $J_{\mathrm{unip}}(f)$ vanishes for any function $f \in C_c^\infty\big(G(\mathbb{A})^1\big)$ that vanishes on the unipotent set in $G(F_S)^1$. For any such function, the distributions $J_M(u,f)$ all vanish as well, according to Theorem 18.2. We conclude that the invariant distribution $T_{\mathrm{unip}}$ annihilates any function in $C_c^\infty\big(G(F_S)^1\big)$ that vanishes on the unipotent set. It follows from this that

$$T_{\mathrm{unip}}(f) = \sum_u a^G(S,u) J_G(u,f),$$

for coefficients $a^G(S,u)$ parametrized by unipotent classes $u$ in $G(F_S)$.

It remains to show that $a^G(S,u)$ vanishes unless $u$ is the image of a unipotent class in $G(F)$, and to evaluate $a^G(S,u)$ explicitly as a Tamagawa number in the case that $u = 1$. This is the hard part. The two assertions are plausible enough. The integrand $k_{\mathrm{unip}}^T(x,f)$ above is supported on the space of $G(\mathbb{A})$-conjugacy classes that come from $F$-rational unipotent classes. Moreover, the contribution to $k_{\mathrm{unip}}^T(x,f)$ from the class 1 equals $f(1)$, which is obviously independent of $x$ and $T$. The integral over $G(F)\backslash G(\mathbb{A})^1$ of this contribution converges, and equals the product

$$\mathrm{vol}\big(G(F)\backslash G(\mathbb{A})^1\big) f(1) = \mathrm{vol}\big(G(F)\backslash G(\mathbb{A})^1\big) J_G(1,f).$$

However, $J_{\mathrm{unip}}(f)$ is defined in terms of the polynomial $J_{\mathrm{unip}}^T(f)$, which depends on a fixed minimal parabolic subgroup $P_0 \in \mathcal{P}(M_0)$, and is equal to an integral whose convergence we can control only for suitably regular points $T \in \mathfrak{a}_{P_0}^+$. Among other difficulties, the dependence of $J_{\mathrm{unip}}^T(f)$ on the local components $f_v$ is not at all transparent. It is therefore not trivial to deduce the remaining two assertions from the intuition we have.

There are two steps. The first is to approximate $J_{\mathrm{unip}}^T(f)$ by the integral of the function

$$K_{\mathrm{unip}}(x,x) = \sum_{u \in \mathcal{U}_G(F)} f(x^{-1}ux)$$

over a compact set. The assertion is that

$$\Big| J_{\mathrm{unip}}^T(f) - \int_{G(F)\backslash G(\mathbb{A})^1} F^G(x,T) K_{\mathrm{unip}}(x,x)\mathrm{d}x \Big| \le \mathrm{e}^{-\frac{1}{2}d_{P_0}(T)}, \tag{19.3}$$

where $F^G(\cdot,T)$ is the compactly supported function on $G(F)\backslash G(\mathbb{A})^1$ defined in §8, and

$$d_{P_0}(T) = \inf_{\alpha \in \Delta_{P_0}} \alpha(T).$$

This inequality is Theorem 3.1 of [**A10**]. Its proof includes an assertion that $F^G(\cdot,T)$ equals the image of the constant function 1 on $G(F)\backslash G(\mathbb{A})^1$ under the truncation operator $\Lambda^T$ [**A10**, Lemma 2.1]. The estimate (19.3), incidentally, is reminiscent of our remarks on the local trace formula at the beginning of §16.

The second step is to solve a kind of lattice point problem. Let $U$ be a unipotent conjugacy class in $G(F)$. If $v$ is a valuation in $S$ and $\varepsilon > 0$, one can define a function $f_{U,v}^{\varepsilon} \in C_c^{\infty}\big(G(\mathbb{A})^1\big)$ that, roughly speaking, truncates the function $f(x)$ whenever the distance from $x_v$ to the $G(F_v)$-conjugacy class of $U$ is greater than $\varepsilon$. (See the beginning of §4 of [**A10**]. The function $f_{U,v}^{\varepsilon}$ equals $f$ at any point in $G(\mathbb{A})^1$ that is conjugate to any point in $\overline{U}(F)$, where $\overline{U}$ is the Zariski closure of $U$.) One then establishes an inequality

$$\int_{G(F)\backslash G(\mathbb{A})^1} F^G(x,T) \sum_{\gamma \in G(F) - \overline{U}(F)} |f_{U,v}^{\varepsilon}(x^{-1}\gamma x)|\mathrm{d}x \le \varepsilon^r \|f\|(1+\|T\|)^{d_0}, \tag{19.4}$$

where $\|\cdot\|$ is a continuous seminorm on $C_c^{\infty}\big(G(\mathbb{A})^1\big)$, and $d_0 = \dim(\mathfrak{a}_0)$. This inequality is the main technical result, Lemma 4.1, of the paper [**A10**]. Its proof in §5-6 of [**A10**] relies on that traditional technique for lattice point problems, the Poisson summation formula.

The inequalities (19.3) and (19.4) are easily combined. By letting $\varepsilon$ approach 0, one deduces the remaining two assertions of Theorem 19.1 from the definition of $J_{\mathrm{unip}}(f) = J_{\mathrm{unip}}^{T_0}(f)$ in terms of $J_{\mathrm{unip}}^T(f)$. (See [**A10**, §4].) $\square$

**Remark.** The explicit formula (19.1) for $a^M(S,1)$ is independent of the set $S$. For nontrivial elements $u \in \mathcal{U}_M(F)$, the coefficients $a^M(S,u)$ do depend on $S$. One sees this in the case $G = GL(2)$ from the term (v) on p. 516 of [**JL**]. As a matter of fact, it is only in the case $G = GL(2)$ that the general coefficients $a^M(S,u)$ have been evaluated. It would be very interesting to understand them better in other examples, although this does not seem to be necessary for presently conceived applications of the trace formula.

The case $\mathfrak{o} = \mathfrak{o}_{\mathrm{unip}}$ we have just discussed is the the most difficult. It is the furthest from the unramified case solved explicitly in §11. For a general class $\mathfrak{o}$, one fashions a descent argument from the techniques of §11. This reduces the problem of computing $J_{\mathfrak{o}}(f)$ to the unipotent case of Theorem 19.1.

We need a couple of definitions before we can state the general result. We say that a semisimple element $\sigma \in G(F)$ is $F$-*elliptic* if $A_{G_\sigma}$ equals $A_G$. In the case $G = GL(n)$, for example, a diagonal element $\sigma$ in $G(F)$ is $F$-elliptic if and only if it is a scalar.

Suppose that $\gamma$ is an element in $G(F)$ with semisimple Jordan component $\sigma$, and that $S$ is a large finite set of valuations of $F$ that contains $S_\infty$. We shall say that a second element $\gamma'$ in $G(F)$ is $(G, S)$-equivalent to $\gamma$ if there is a $\delta \in G(F)$ with the following two properties.

(i) $\sigma$ is also the semisimple Jordan component of $\delta^{-1}\gamma'\delta$.
(ii) The unipotent elements $\sigma^{-1}\gamma$ and $\sigma^{-1}\delta^{-1}\gamma'\delta$ in $G_\sigma(F)$ are $(G_\sigma, S)$-equivalent, in the sense of the earlier definition.

There could be several classes $u \in \big(\mathcal{U}_{G_\sigma}(F)\big)_{G_\sigma,S}$ such that $\sigma u$ is $(G, S)$-equivalent to $\gamma$. The set of such $u$, which we write simply as $\{u : \ \sigma u \sim \gamma\}$, has a transitive action under the finite group

$$\iota^G(\sigma) = G_{\sigma,+}(F)/G_\sigma(F).$$

We define

$$a^G(S, \gamma) = \varepsilon^G(\sigma)|\iota^G(\sigma)|^{-1} \sum_{\{u:\sigma u\sim\gamma\}} a^{G_\sigma}(S, u), \tag{19.5}$$

where

$$\varepsilon^G(\sigma) = \begin{cases} 1, & \text{if } \sigma \text{ is } F\text{-elliptic in } G, \\ 0, & \text{otherwise.} \end{cases}$$

Then $a^G(S, \gamma)$ depends only on the $(G, S)$-equivalence class of $\gamma$. If $\gamma$ is semisimple, we can use (19.1) to express $a^G(S, \gamma)$. In this case, we see that

$$a^G(S, \gamma) = \varepsilon^G(\gamma)|\iota^G(\gamma)|^{-1}\mathrm{vol}\big(G_\gamma(F)\backslash G_\gamma(\mathbb{A})^1\big), \tag{19.6}$$

and in particular, that $a^G(S, \gamma)$ is independent of $S$.

THEOREM 19.2. *Suppose that $\mathfrak{o}$ is any class in $\mathcal{O}$. Then there is a finite set $S_{\mathfrak{o}}$ of valuations of $F$ that contains $S_\infty$ such that for any finite set $S \supset S_{\mathfrak{o}}$ and any function $f \in C_c^\infty\big(G(F_S)^1\big)$,*

$$J_{\mathfrak{o}}(f) = \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in(M(F)\cap\mathfrak{o})_{M,S}} a^M(S, \gamma)J_M(\gamma, f), \tag{19.7}$$

*where $\big(M(F) \cap \mathfrak{o}\big)_{M,S}$ is the finite set of $(M, S)$-equivalence classes in $M(F) \cap \mathfrak{o}$, and $J_M(\gamma, f)$ is the general weighted orbital integral of $f$ defined in §18.*

This is the main result, Theorem 8.1, of the paper [**A11**]. The strategy is to establish formulas of descent that reduce each side of the putative formula (19.7) to the unipotent case (19.2). We are speaking of what might be called "semisimple descent" here. It pertains to the Jordan decomposition, and is therefore different from the property of "parabolic descent" in the formula (18.8). We shall attempt to give a brief idea of the proof.

The reduction is actually a generalization of the unramified case treated in §11. In particular, it begins with the formula

$$J_{\mathfrak{o}}^T(f) = \int_{G(F)\backslash G(\mathbb{A})^1} \widetilde{k}_{\mathfrak{o}}^T(x,f)\mathrm{d}x$$

of Theorem 11.1. We recall that

$$\widetilde{k}_{\mathfrak{o}}^T(x,f) = \sum_{P\supset P_0}(-1)^{\dim(A_P/A_G)}\sum_{\delta\in P(F)\backslash G(F)}\widetilde{K}_{P,\mathfrak{o}}(\delta x,\delta x)\widehat{\tau}_P\big(H_P(\delta x)-T\big),$$

where $P_0\in\mathcal{P}(M_0)$ is a fixed minimal parabolic subgroup. The definition (11.1) expresses $\widetilde{K}_{P,\mathfrak{o}}(\delta x,\delta x)$ in terms of $f$, and the Jordan decomposition of elements $\gamma\in M_P\cap\mathfrak{o}$. The formula contains integrals over unipotent adelic groups $N_R(\mathbb{A}) = N_P(\mathbb{A})_{\gamma_s}$, where $R$ is the parabolic subgroup $P\cap G_{\gamma_s}$ of $G_{\gamma_s}$. It is therefore quite plausible that $J_{\mathfrak{o}}^T(f)$ can be reduced to unipotent distributions $J_{\mathrm{unip}}^{H,T_H}(\Phi)$ attached to reductive subgroups $H$ of $G$, and functions $\Phi\in C_c^\infty\big(H(\mathbb{A})^1\big)$ obtained from $f$ and $T$ by descent. However, the combinatorics of the reduction are somewhat complicated.

One begins as in §11 by fixing a pair $(P_1,\alpha_1)$ that represents the anisotropic rational datum of $\mathfrak{o}$. Then $P_1$ is a parabolic subgroup, which is standard relative to the fixed minimal parabolic subgroup $P_0\in\mathcal{P}(M_0)$ used to construct $J_{\mathfrak{o}}(f)$. One also fixes an element $\sigma=\gamma_1$ in the anisotropic (semisimple) conjugacy class $\alpha_1$ in $M_{P_1}(F)$. Then $P_{1\sigma}=P_1\cap G_\sigma$ is a minimal parabolic subgroup of $G_\sigma$, with Levi component $M_{1\sigma}=M_{P_1}\cap G_\sigma$. The groups $H$ above are Levi subgroups $M_\sigma$ of $G_\sigma$ in the finite set $\mathcal{L}^\sigma=\mathcal{L}^{G_\sigma}(M_{1\sigma})$. The corresponding functions $\Phi=\Phi_y$ of descent in $C_c^\infty\big(M_\sigma(\mathbb{A})^1\big)$ depend on $T$, and among other things, a set of representatives $y$ of $G_\sigma(\mathbb{A})\backslash G(\mathbb{A})$ in $G(\mathbb{A})$. (See [**A11**, p. 199].)

We take $S_{\mathfrak{o}}$ to be any finite set of valuations of $F$ that contains $S_\infty$, and such that any $v\notin S_{\mathfrak{o}}$ satisfies the following four conditions.

(i) $|D^G(\sigma)|_v=1$.
(ii) The intersection $K_{\sigma,v}=K_v\cap G_\sigma(F_v)$ is an admissible maximal compact subgroup of $G_\sigma(F_v)$.
(iii) $\sigma K_v\sigma^{-1}=K_v$.
(iv) If $y_v\in G(F_v)$ is such that $y_v^{-1}\sigma\mathcal{U}_{G_\sigma}(F_v)y_v$ meets $\sigma K_v$, then $y_v$ belongs to $G_\sigma(F_v)K_v$.

(See [**A11**, p. 203].) We choose $S\supset S_{\mathfrak{o}}$ and $f\in C_c^\infty\big(G(F)^1\big)$, as in the statement of the theorem. It then turns out that for any group $M_\sigma\in\mathcal{L}^\sigma$, the corresponding functions of descent $\Phi_y$ all lie in the subspace $C_c^\infty\big(M_\sigma(F_S)^1\big)$ of $C_c^\infty\big(M_\sigma(\mathbb{A})^1\big)$.

Recall that $J_{\mathfrak{o}}(f)$ is the value at $T=T_0$ of the polynomial $J^T(f)$. The unipotent distribution $J_{\mathrm{unip}}^{M_\sigma}(\Phi_y)$ is the value of a polynomial $J_{\mathrm{unip}}^{M_\sigma,T_\sigma}(\Phi_y)$ of $T_\sigma$ in a subspace $\mathfrak{a}_{1\sigma}$ of $\mathfrak{a}_0$ at a fixed point $T_{0\sigma}$. In the descent formula, the groups $M_\sigma$ are of the form $M_R$, where $R$ ranges over the set $\mathcal{F}^\sigma=\mathcal{F}^{G_\sigma}(M_{1\sigma})$. The formula is

$$J_{\mathfrak{o}}(f)=|\iota^G(\sigma)|^{-1}\int_{G_\sigma(\mathbb{A})\backslash G(\mathbb{A})}\Big(\sum_{R\in\mathcal{F}^\sigma}|W_0^{M_R}||W_0^{G_\sigma}|^{-1}J_{\mathrm{unip}}^{M_R}(\Phi_{R,y,T_1})\Big)\mathrm{d}y, \tag{19.8}$$

where $\Phi_{R,y,T_1}$ is obtained from the general descent function $\Phi_y$ by specializing $T$ to the point $T_1=T_0-T_{0\sigma}$ [**A11**, Lemma 6.2]. Since the general functions $\Phi_y$ and their specializations $\Phi_{R,y,T_1}$ are somewhat technical, we have not attempted to define them. However, their construction is formally like that of the functions $f_{Q,y}$ in (18.5). In particular, it relies on the splitting formula of Lemma 17.4.

The formula (19.8) of geometric descent has an analogue for weighted orbital integrals. Suppose that $M$ is a Levi subgroup of $G$ that contains $M_1 = M_{P_1}$. Then $\sigma$ is contained in $M(F)$. Set $\gamma = \sigma u$, where $u$ is a unipotent element in $M_\sigma(F_S)$. The formula is

$$J_M(\gamma, f) = \int_{G_\sigma(F_S)\backslash G(F_S)} \Big( \sum_{R\in\mathcal{F}^\sigma(M_\sigma)} J_{M_\sigma}^{M_R}(u, \Phi_{R,y,T_1}) \Big) \mathrm{d}y, \tag{19.9}$$

where $f$ is any function in $C_c^\infty\big(G(F_S)^1\big)$, and $\mathcal{F}^\sigma(M_\sigma) = \mathcal{F}^{G_\sigma}(M_\sigma)$ ([**A11**, Corollary 8.7]).

The formulas (19.8) and (19.9) of geometric descent must seem rather murky, given the limited extent of our discussion. However, the reader will no doubt agree that the existence of such formulas is plausible. Taking them for granted, one can well imagine that an application of Theorem 19.1 to the distributions in these formulas would lead to an expansion of $J_{\mathfrak{o}}(f)$. The required formula (19.7) for $J_{\mathfrak{o}}(f)$ does indeed follow from Theorem 19.1, used in conjunction with the definition (19.5) of the coefficients $a^M(S,\gamma)$. □

If $\Delta$ is a compact neighbourhood of 1 in $G(\mathbb{A})^1$, we write $C_\Delta^\infty\big(G(\mathbb{A})^1\big)$ for the subspace of functions in $C_c^\infty\big(G(\mathbb{A})^1\big)$ that are supported on $\Delta$. For example, we could take $\Delta$ to be the set

$$\Delta_N = \{x \in G(\mathbb{A}) : \ \log\|x\| \le N\}$$

attached to a positive number $N$. In this case we write $C_N^\infty\big(G(\mathbb{A})^1\big)$ in place of $C_{\Delta_N}^\infty\big(G(\mathbb{A})^1\big)$. For any $\Delta$, we can certainly find a finite set $S$ of valuations of $F$ containing $S_\infty$, such that $\Delta$ is the product of a compact neighbourhood of 1 in $G(F_S)^1$ with $K^S$. We write $S_\Delta^0$ for the minimal such set. We also write

$$C_\Delta^\infty\big(G(F_S)^1\big) = C_\Delta^\infty\big(G(\mathbb{A})^1\big) \cap C_c^\infty\big(G(F_S)^1\big),$$

for any finite set $S \supset S_\Delta^0$. The fine geometric expansion is given by the following corollary of the last theorem.

Corollary 19.3. *Given a compact neighbourhood $\Delta$ of 1 in $G(\mathbb{A})^1$, we can find a finite set $S_\Delta \supset S_\Delta^0$ of valuations of $F$ such that for any finite set $S \supset S_\Delta$, and any $f \in C_\Delta^\infty\big(G(F_S)^1\big)$,*

$$J(f) = \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in(M(F))_{M,S}} a^M(S,\gamma) J_M(\gamma, f), \tag{19.10}$$

*where $\big(M(F)\big)_{M,S}$ is the set of $(M,S)$-equivalence classes in $M(F)$. The summands on the right hand side of (19.10) vanish for all but finite many $\gamma$.*

The corollary is Theorem 9.2 of [**A11**]. It follows immediately from Theorem 19.2 above, once we know that there is a finite subset of $\mathcal{O}$ outside of which $J_{\mathfrak{o}}(f)$ vanishes for any $f \in C_\Delta^\infty\big(G(\mathbb{A})^1\big)$. This property follows immediately from [**A11**, Lemma 9.1], which asserts that there are only finitely many classes $\mathfrak{o} \in \mathcal{O}$ such that the set

$$\{x^{-1}\gamma x : \ x \in G(\mathbb{A}), \ \gamma \in \mathfrak{o}\}$$

meets $\Delta$, and is proved in the appendix of [**A11**]. □

## 20. Application of a Paley-Wiener theorem

The next two sections will be devoted to the refinement of the coarse spectral expansion (14.8). These sections are longer and more intricate than anything so far. For one reason, there are results from a number of different sources that we need to discuss. Moreover, we have included more details than in some of the earlier arguments. The refined spectral expansion is deeper than its geometric counterpart, dependent as it is on Eisenstein series, and we need to get a feeling for the techniques. In particular, it is important to understand how global intertwining operators intervene in the "discrete part" of the spectral expansion.

The spectral side is complicated by the presence of a delicate analytic problem, with origins in the theory of Eisenstein series. It can be described as that of interchanging two limits. We shall see how to resolve the problem in this section. The computations of the fine spectral expansion will then be treated in the next section.

In order to use the results of Part I, we shall work for the time being with a fixed minimal parabolic subgroup $P_0 \in \mathcal{P}(M_0)$. Suppose that $\chi \in \mathfrak{X}$ indexes one of the summands in the coarse spectral expansion. According to Theorem 15.1,

$$J_\chi^T(f) = \sum_{P \supset P_0} n_P^{-1} \int_{i\mathfrak{a}_P^*} \operatorname{tr}\bigl(M_{P,\chi}^T(\lambda)\mathcal{I}_{P,\chi}(\lambda, f)\bigr)d\lambda,$$

where $T \in \mathfrak{a}_{P_0}^+$ is suitably regular, and $M_{P,\chi}^T(\lambda)$ is the operator on $\mathcal{H}_{P,\chi}$ defined by the inner product (15.1) of truncated Eisenstein series. In the next section, we shall see that the explicit inner product formula for truncated Eisenstein series in Proposition 15.3 holds in general, provided it is interpreted as an asymptotic formula in $T$. We might therefore hope to compute $J_\chi^T(f)$ as an explicit polynomial in $T$ by letting the distance

$$d_{P_0}(T) = \inf_{\alpha \in \Delta_{P_0}} \alpha(T)$$

approach infinity. However, any such computation seems to require estimates for the derivatives of $M_{P,\chi}^T(\lambda)$ that are uniform in $\lambda$. This would amount to estimating derivatives in $\lambda$ of Eisenstein series outside the domain of absolute convergence, something that is highly problematical. On the other hand, if we could multiply the integrand in the formula for $J_\chi^T(f)$ above by a smooth, compactly supported cut-off function in $\lambda$, the computations ought to be manageable. The analytic problem is to show that one can indeed insert such a cut-off function.

In the formula for $J_\chi^T(f)$ we have just quoted from Part I, $f$ belongs to $C_c^\infty\bigl(G(\mathbb{A})\bigr)$. We are now taking $f$ to be a function in $C_c^\infty\bigl(G(\mathbb{A})^1\bigr)$. For any such $f$, the integrand in the formula is a well defined function of $\lambda$ in $i\mathfrak{a}_P^*/i\mathfrak{a}_G^*$. The formula remains valid for $f \in C_c^\infty\bigl(G(\mathbb{A})^1\bigr)$, so long as we take the integral over $\lambda \in i\mathfrak{a}_P^*/i\mathfrak{a}_G^*$.

The class $\chi \in \mathfrak{X}$ will be fixed for the rest of this section. We shall first state three preliminary lemmas, all of which are consequences of Theorem 14.1 and its proof. For any $P \supset P_0$, we write

$$\mathcal{H}_{P,\chi} = \bigoplus_\pi \mathcal{H}_{P,\chi,\pi},$$

where $\pi$ ranges over the set $\Pi_{\mathrm{unit}}\bigl(M_P(\mathbb{A})^1\bigr)$ of equivalence classes of irreducible unitary representations of $M_P(\mathbb{A})^1$, and $\mathcal{H}_{P,\chi,\pi}$ is the intersection of $\mathcal{H}_{P,\chi}$ with the subspace $\mathcal{H}_{P,\pi}$ of vectors $\phi \in \mathcal{H}_P$ such that for each $x \in G(\mathbb{A})$, the function $\phi_x(m) = \phi(mx)$ in $L^2_{\mathrm{disc}}\big(M_P(\mathbb{Q})\backslash M_P(\mathbb{A})^1\big)$ is a matrix coefficient of $\pi$. We write $\mathcal{I}_{P,\chi,\pi}(\lambda, f)$ for the restriction of $\mathcal{I}_{P,\chi}(\lambda, f)$ to $\mathcal{H}_{P,\chi,\pi}$. We then set

$$\Psi^T_\pi(\lambda, f) = n_P^{-1}\mathrm{tr}\big(M^T_{P,\chi}(\lambda)\mathcal{I}_{P,\chi,\pi}(\lambda, f)\big),$$

for any $f \in C^\infty_c\big(G(\mathbb{A})^1\big)$ and $\lambda \in i\mathfrak{a}^*_P/i\mathfrak{a}^*_G$.

LEMMA 20.1. *There are positive constants $C_0$ and $d_0$ such that for any $f \in C^\infty_c\big(G(\mathbb{A})^1\big)$, any $n \geq 0$, and any $T \in \mathfrak{a}_0$ with $d_{P_0}(T) > C_0$,*

$$\sum_{P\supset P_0}\int_{i\mathfrak{a}^*_P/i\mathfrak{a}^*_G}\sum_\pi |\Psi^T_\pi(\lambda, f)|(1 + \|\lambda\|)^n \mathrm{d}\lambda \leq c_{n,f}(1 + \|T\|)^{d_0},$$

*for a constant $c_{n,f}$ that is independent of $T$.*

The lemma is a variant of Proposition 14.1(a). One obtains the factor $(1 + \|\lambda\|)^n$ in the estimate by choosing a suitable differentiable operator $\Delta$ on $G(\mathbb{R})$, and applying the arguments of Theorem 14.1(a) to $\Delta f$ in place of $f$. (See [**A7**, Proposition 2.1]. One can in fact take $d_0 = \dim \mathfrak{a}_0$.) □

LEMMA 20.2. *There is a constant $C_0$ such that for any $N > 0$ and any $f \in C^\infty_N\big(G(\mathbb{A})^1\big)$, the expression*

$$\sum_{P\supset P_0}\sum_\pi\int_{i\mathfrak{a}^*_P/i\mathfrak{a}^*_G}\Psi^T_\pi(\lambda, f)\mathrm{d}\lambda \tag{20.1}$$

*equals $J^T_\chi(f)$, and is hence a polynomial in $T$ of degree bounded by $d_0 = \dim \mathfrak{a}_0$, whenever*

$$d_{P_0}(T) > C_0(1 + N).$$

The expression equals

$$\int_{G(F)\backslash G(\mathbb{A})^1}\Lambda^T_1\Lambda^T_2 K_\chi(x, x)\mathrm{d}x = \int_{G(F)\backslash G(\mathbb{A})^1}\Lambda^T_2 K_\chi(x, x)\mathrm{d}x.$$

The lemma follows from Theorem 14.1(c), and an analysis of how the proof of this result depends quantitatively on the support of $f$. (See [**A7**, Proposition 2.2].) □

If $\tau_1, \tau_2 \in \Pi_{\mathrm{unit}}(K_\mathbb{R})$ are irreducible unitary representations of $K_\mathbb{R}$, set

$$f_{\tau_1,\tau_2}(x) = \int_{K_\mathbb{R}}\int_{K_\mathbb{R}}\mathrm{tr}\big(\tau_1(k_1)\big)f(k_1^{-1}xk_2^{-1})\mathrm{tr}\big(\tau_2(k_2)\big)\mathrm{d}k_1\mathrm{d}k_2,$$

for any function $f \in C^\infty_c\big(G(\mathbb{A})^1\big)$. Then

$$f(x) = \sum_{\tau_1,\tau_2} f_{\tau_1,\tau_2}(x).$$

LEMMA 20.3. *There is a decomposition*

$$J^T_\chi(f) = \sum_{\tau_1,\tau_2} J^T_\chi(f_{\tau_1,\tau_2}).$$

The lemma follows easily from an inspection of how the estimates of the proof of Theorem 14.1 depend on left and right translation of $f$ by $K_{\mathbb{R}}$. (See [**A7**, Proposition 2.3].) $\square$

The last three lemmas form the backdrop for our discussion of the analytic problem. The third lemma allows us to assume that $f$ belongs to the Hecke algebra

$$\mathcal{H}(G) = \mathcal{H}\big(G(\mathbb{A})^1\big) = \mathcal{H}\big(G(\mathbb{A})^1, K\big)$$

of $K$-finite functions in $C_c^\infty\big(G(\mathbb{A})^1\big)$. We recall that $f$ is $K$-finite if the space of functions on $G(\mathbb{A})^1$ spanned by left and right $K$-translates of $f$ is finite dimensional. The second lemma describes the qualitative behaviour of $J_\chi^T(f)$ as a function of $T$, quantitatively in terms of the support of $f$. If we could somehow construct a family of new functions in $\mathcal{H}(G)$ in terms of the operators $\mathcal{I}_{P,\chi,\pi}(f)$, with some control over their supports, we might be able to bring this lemma to bear on our analytic difficulties.

Our rescue comes in the form of a Paley-Wiener theorem, or rather a corollary of the theorem that deals with multipliers. Multipliers are defined in terms of infinitesimal characters. To describe them, we have to fix an appropriate Cartan subalgebra.

For each archimedean valuation $v \in S_\infty$ of $F$, we fix a real vector space

$$\mathfrak{h}_v = i\mathfrak{b}_v \oplus \mathfrak{a}_0,$$

where $\mathfrak{b}_v$ is a Cartan subalgebra of the compact Lie group $K_v \cap M_0(F_v)$. We then set

$$\mathfrak{h} = \mathfrak{h}_\infty = \bigoplus_{v \in S_\infty} \mathfrak{h}_v.$$

This space can be identified with a split Cartan subalgebra of the Lie group $G_s^*(F_\infty)$, where

$$F_\infty = F_{S_\infty} = \bigoplus_{v \in S_\infty} F_v,$$

and $G_s^*$ is a split $F$-form of the group $G$. In particular, the complex Weyl group $W = W_\infty$ of the Lie group $G(F_\infty)$ acts on $\mathfrak{h}$. The space $\mathfrak{h}$ comes with a canonical projection $\mathfrak{h} \mapsto \mathfrak{a}_P$, for any standard parabolic subgroup $P \supset P_0$, whose transpose is an injection $\mathfrak{a}_P^* \subset \mathfrak{h}^*$ of dual spaces. It is convenient to fix a positive definite, $W$-invariant inner product $(\cdot,\cdot)$ of $\mathfrak{h}$. The corresponding Euclidean norm $\|\cdot\|$ on $\mathfrak{h}$ restricts to a $W_0$-invariant Euclidean norm on $\mathfrak{a}_0$. We assume that it is dominated by the height function on $G(\mathbb{A})$ fixed earlier, in the sense that

$$\|H\| \leq \log \|\exp H\|, \qquad H \in \mathfrak{a}_0.$$

The infinitesimal character of an irreducible representation $\pi_\infty \in \Pi\big(G(F_\infty)\big)$ is represented by a $W$-orbit $\nu_{\pi_\infty}$ in the complex dual space $\mathfrak{h}_{\mathbb{C}}^*$ of $\mathfrak{h}$. It satisfies

$$\pi_\infty(zf_\infty) = \langle h(z), \nu_{\pi_\infty}\rangle \pi_\infty(f_\infty), \qquad z \in \mathcal{Z}_\infty,\ f_\infty \in C_c^\infty\big(G(F_\infty)\big),$$

where $h\colon \mathcal{Z}_\infty \to S(\mathfrak{h}_{\mathbb{C}})^W$ is the isomorphism of Harish-Chandra, from the algebra $\mathcal{Z}_\infty$ of bi-invariant differential operators on $G(F_\infty)$ onto the algebra of $W$-invariant polynomials on $\mathfrak{h}_{\mathbb{C}}^*$, that plays a central role in his work on representations of real groups. The algebra $\mathcal{Z}_\infty$ acts on the Hecke algebra $\mathcal{H}\big(G(\mathbb{A})\big)$ of $G(\mathbb{A})$ through the $G(F_\infty)$-component of a given function $f$. However, the space of functions $zf$, $z \in \mathcal{Z}_\infty$, is not rich enough for us to exploit Lemma 20.2.

Let $\mathcal{E}(\mathfrak{h})^W$ be the convolution algebra of $W$-invariant, compactly supported distributions on $\mathfrak{h}$. According to the classical Paley-Wiener theorem, the adjoint Fourier transform $\alpha \to \widehat{\alpha}$ is an isomorphism from $\mathcal{E}(\mathfrak{h})^W$ onto the algebra of entire, $W$-invariant functions $\widehat{\alpha}(\nu)$ on $\mathfrak{h}^*_{\mathbb{C}}$ of exponential type that are slowly increasing on cylinders

$$\{\nu \in \mathfrak{h}^*_{\mathbb{C}} : \ \|\mathrm{Re}(\nu)\| \leq r\}, \qquad r \geq 0.$$

The subalgebra $C_c^\infty(\mathfrak{h})^W$ is mapped onto the subalgebra of functions $\widehat{\alpha}$ that are rapidly decreasing on cylinders. (By the adjoint Fourier transform $\widehat{\alpha}$ we mean the transpose-inverse of the standard Fourier transform on functions, rather than simply the transpose. In other words,

$$\widehat{\alpha}(\nu) = \int_{\mathfrak{h}} \alpha(H) e^{\nu(H)} dH,$$

in case $\alpha$ is a function.)

We write $\mathcal{H}\big(G(F_\infty)\big) = \mathcal{H}\big(G(F_\infty), K_\infty\big)$ for the Hecke algebra of $K_\infty = \prod_{v \in S_\infty} K_v$ finite functions in $C_c^\infty\big(G(F_\infty)\big)$, and $\mathcal{H}_N\big(G(F_\infty)\big)$ for the subspace of functions in $\mathcal{H}\big(G(F_\infty)\big)$ supported on the set

$$\{x_\infty \in G(F_\infty) : \ \log \|x_\infty\| \leq N\}.$$

Theorem 20.4. *There is a canonical action*

$$\alpha : \ f_\infty \longrightarrow f_{\infty,\alpha}, \qquad \alpha \in \mathcal{E}(\mathfrak{h})^W, \ f_\infty \in \mathcal{H}\big(G(F_\infty)\big),$$

*of* $\mathcal{E}(\mathfrak{h})^W$ *on* $\mathcal{H}\big(G(F_\infty)\big)$ *with the property that*

$$\pi_\infty(f_{\infty,\alpha}) = \widehat{\alpha}(\nu_{\pi_\infty})\pi_\infty(f_\infty),$$

*for any* $\pi_\infty \in \Pi\big(G(F_\infty)\big)$. *Moreover, if* $f_\infty$ *belongs to* $\mathcal{H}_N\big(G(F_\infty)\big)$ *and* $\alpha$ *is supported on the subset of points* $H \in \mathfrak{h}$ *with* $\|H\| \leq N_\alpha$, *then* $f_{\infty,\alpha}$ *lies in* $\mathcal{H}_{N+N_\alpha}\big(G(F_\infty)\big)$.

(See [**A9**, Theorem 4.2].) □

This is the multiplier theorem we will apply to the expression (20.1). We shall treat (20.1) as a linear functional of $f$ in the Hecke algebra $\mathcal{H}(G) = \mathcal{H}\big(G(\mathbb{A})^1\big)$. If $\mathfrak{h}^1$ is the subspace of points in $\mathfrak{h}$ whose projection onto $\mathfrak{a}_G$ vanishes, we shall take $\alpha$ to be in the subspace $\mathcal{E}(\mathfrak{h}^1)^W$ of distributions in $\mathcal{E}(\mathfrak{h})^W$ supported on $\mathfrak{h}^1$. If $f$ belongs to the Hecke algebra $\mathcal{H}\big(G(\mathbb{A})\big)$ on $G(\mathbb{A})$, we define $f_\alpha$ to be the function in $\mathcal{H}\big(G(\mathbb{A})\big)$ obtained by letting $\alpha$ act on the archimedean component of $f$. The restriction of $f_\alpha$ to $G(\mathbb{A})^1$ will then depend only on the restriction of $f$ to $G(\mathbb{A})^1$. In other words, $f_\alpha \in \mathcal{H}(G)$ is defined for any $f \in \mathcal{H}(G)$. We shall substitute functions of this form into (20.1).

Suppose that $P \supset P_0$ and $\pi \in \Pi_{\mathrm{unit}}\big(M_P(\mathbb{A})^1\big)$ are as in (20.1). Then $\pi$ is the restriction to $M_P(\mathbb{A})^1$ of a unitary representation

$$\pi_\infty \otimes \pi_{\mathrm{fin}}, \qquad \pi_\infty \in \Pi_{\mathrm{unit}}\big(M_P(F_\infty)\big), \ \pi_{\mathrm{fin}} \in \Pi_{\mathrm{unit}}\big(M_P(\mathbb{A}_{\mathrm{fin}})\big),$$

of $M_P(\mathbb{A})$. We obtain a linear form $\nu_\pi = \nu_{\pi_\infty}$ on $\mathfrak{h}_{\mathbb{C}}$, which we decompose

$$\nu_\pi = X_\pi + iY_\pi, \qquad X_\pi, Y_\pi \in \mathfrak{h}^*,$$

into real and imaginary parts. These points actually stand for orbits in $\mathfrak{h}^*$ of the complex Weyl group of $M_P(F_\infty)$, but we can take them to be fixed representatives of the corresponding orbits. Then $X_\pi$ is uniquely determined by $\pi$, while the imaginary part $Y_\pi$ is determined by $\pi$ only modulo $\mathfrak{a}_P^*$. However, we may as well identify $Y_\pi$ with the unique representative in $\mathfrak{h}^*$ of the coset in $\mathfrak{h}^*/\mathfrak{a}_P^*$ of smallest norm $\|Y_\pi\|$. This amounts to taking the representation $\pi_\infty$ of $M_P(F_\infty)$ to be invariant under the subgroup $A_{M_P,\infty}^+$ of $M_P(F_\infty)$, a convention that is already implicit in the notation $\mathcal{H}_{P,\pi}$ above.

If $B$ is any $W$-invariant function on $i\mathfrak{h}^*$, we define a function

$$B_\pi(\lambda) = B(iY_\pi + \lambda), \qquad \lambda \in i\mathfrak{a}_P^*,$$

on $i\mathfrak{a}_P^*$. We also write

$$B^\varepsilon(\nu) = B(\varepsilon\nu), \qquad \nu \in i\mathfrak{h}^*,$$

for any $\varepsilon > 0$. We shall want $B$ to be rapidly decreasing on $i\mathfrak{h}^*/i\mathfrak{a}_G^*$. An obvious candidate would be the Paley-Wiener function $\widehat{\alpha}$ attached to a function $\alpha \in C_c^\infty(\mathfrak{h}^1)^W$. However, the point of this exercise is to allow $B$ to be an arbitrary element in the space $\mathcal{S}(i\mathfrak{h}^*/i\mathfrak{a}_G^*)^W$ of $W$-invariant Schwartz functions on $i\mathfrak{h}^*/i\mathfrak{a}_G^*$.

The next theorem provides the way out of our analytic difficulties.

THEOREM 20.5. (a) *For any $B \in \mathcal{S}(i\mathfrak{h}^*/i\mathfrak{a}_G^*)^W$ and $f \in \mathcal{H}(G)$, there is a unique polynomial $P^T(B,f)$ in $T$ such that the difference*

$$\sum_{P \supset P_0} \int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \sum_\pi \Psi_\pi^T(\lambda, f) B_\pi(\lambda) d\lambda - P^T(B,f) \tag{20.2}$$

*approaches* 0 *as $T$ approaches infinity in any cone*

$$\mathfrak{a}_{P_0}^r = \{T \in \mathfrak{a}_0 : \ d_{P_0}(T) > r\|T\|\}, \qquad r > 0.$$

(b) *If $B(0) = 1$, then*

$$J_\chi^T(f) = \lim_{\varepsilon \to 0} P^T(B^\varepsilon, f).$$

This is the main result, Theorem 6.3, of the paper [**A7**]. We shall sketch the proof.

The idea is to approximate $B$ by Paley-Wiener functions $\widehat{\alpha}$, for $\alpha \in C_c^\infty(\mathfrak{h}^1)^W$. Assume that $f$ belongs to the space

$$\mathcal{H}_N(G) = \mathcal{H}(G) \cap C_N^\infty\big(G(\mathbb{A})^1\big),$$

for some fixed $N > 0$, and that $\alpha$ is a general element in $\mathcal{E}(\mathfrak{h}^1)^W$. Then $f_\alpha$ lies in $\mathcal{H}_{N+N_\alpha}(G)$. For any $P \supset P_0$ and $\lambda \in i\mathfrak{a}_P^*$, $\mathcal{I}_P(\lambda, f_\alpha)$ is an operator on $\mathcal{H}_P$ whose restriction to $\mathcal{H}_{P,\chi,\pi}$ equals

$$\widehat{\alpha}(\nu_\pi + \lambda)\mathcal{I}_{P,\chi,\pi}(\lambda, f).$$

Applying Lemma 20.2 with $f_\alpha$ in place of $f$, we see that the expression

$$\sum_{P \supset P_0} \int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \sum_\pi \widehat{\alpha}(\nu_\pi + \lambda) \Psi_\pi^T(\lambda, f) d\lambda \tag{20.3}$$

equals $J_\chi^T(f_\alpha)$ whenever $d_{P_0}(T) > C_0(1 + N + N_\alpha)$, and is hence a polynomial in $T$ in this range. The sum over $\pi$ in (20.3) can actually be taken over a finite set that depends only on $\chi$ and $f$. This is implicit in Langlands's proof of Theorem 7.2, specifically his construction of the full discrete spectrum from residues of cuspidal Eisenstein series.

Suppose that $\alpha$ belongs to the subspace $C_c^\infty(\mathfrak{h}^1)^W$ of $\mathcal{E}(\mathfrak{h}^1)^W$. Then $J_\chi^T(f_\alpha)$ equals

$$\sum_{P\supset P_0}\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*}\sum_\pi\int_{\mathfrak{h}^1}\Psi_\pi^T(\lambda,f)\mathrm{e}^{(\nu_\pi+\lambda)(H)}\alpha(H)\mathrm{d}H\mathrm{d}\lambda.$$

By Lemma 20.1, integral

$$\psi_\pi^T(H,f)=\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*}\Psi_\pi^T(\lambda,f)\mathrm{e}^{\lambda(H)}d\lambda$$

converges to a bounded, smooth function of $H\in\mathfrak{h}^1$. It follows that

$$J_\chi^T(f_\alpha)=\int_{\mathfrak{h}^1}\Big(\sum_{P\supset P_0}\sum_\pi\psi_\pi^T(H,f)\mathrm{e}^{\nu_\pi(H)}\Big)\alpha(H)\mathrm{d}H,$$

whenever $d_{P_0}(T)>C_0(1+N+N_\alpha)$. Since $C_c^\infty(\mathfrak{h}^1)^W$ is dense in $\mathcal{E}(\mathfrak{h}^1)$ (in the weak topology), the assertion actually holds for any $\alpha\in\mathcal{E}(\mathfrak{h}^1)^W$ (with the integral being interpreted as evaluation of the distribution $\alpha$).

If $H$ is any point in $\mathfrak{h}^1$, let $\delta_H$ be the Dirac measure on $\mathfrak{h}^1$ at $H$. The symmetrization

$$\alpha_H=|W|^{-1}\sum_{s\in W}\delta_{s^{-1}H}$$

belongs to $\mathcal{E}(\mathfrak{h}^1)^W$. The function

$$p^T(H,f)=J_\chi^T(f_{\alpha_H})$$

is therefore a well defined polynomial in $T$, of degree bounded by $d_0$. The support of $\alpha_H$ is contained in the ball about the origin of radius $\|H\|$, so we can take $N_{\alpha_H}=\|H\|$. It follows that

$$p^T(H,f)=\sum_{P\supset P_0}\sum_\pi|W|^{-1}\sum_{s\in W}\psi_\pi^T(s^{-1}H,f)\mathrm{e}^{\nu_\pi(s^{-1}H)},\tag{20.4}$$

for all $H$ and $T$ with $d_{P_0}(T)>C_0(1+N+\|H\|)$. The right hand expression may be regarded as a triple sum over a finite set. It follows that $p^T(H,f)$ is a smooth function of $H\in\mathfrak{h}^1$ for all $T$ in the given domain, and hence for all $T$, by polynomial interpolation. Observe that $\alpha_0=\delta_0$, and therefore that $f_{\alpha_0}=f$. It follows that

$$p^T(0,f)=J_\chi^T(f).$$

To study the right hand side of (20.4), we group the nonzero summands with a given real exponent $X_\pi$. More precisely, we define an equivalence relation on the triple indices of summation in (20.4) by setting $(P',\pi',s')\sim(P,\pi,s)$ if $s'X_{\pi'}=sX_\pi$. If $\Gamma$ is any equivalence class, we set $X_\Gamma=sX_\pi$, for any $(P,\pi,s)\in\Gamma$. We also define

$$\psi_\Gamma^T(H,f)=|W|^{-1}\sum_{(P,\pi,s)\in\Gamma}\mathrm{e}^{iY_\pi(s^{-1}H)}\psi_\pi^T(s^{-1}H,f).$$

Then $\psi_\Gamma^T(H,f)$ is a bounded, smooth function of $H\in\mathfrak{h}^1$ that is defined for all $T$ with $d_{P_0}(T)$ greater than some absolute constant. In fact, Lemma 20.1 implies that for any invariant differential operator $D$ on $\mathfrak{h}^1$, there is a constant $c_{D,f}$ such that

$$|D\psi_\Gamma^T(H,f)|\le c_{D,f}(1+\|T\|)^{d_0},\qquad H\in\mathfrak{h}^1,\ d_{P_0}(T)>C_0,\tag{20.5}$$

for constants $C_0$ and $d_0$ independent of $f$. In particular, we can assume that the constants $C_0$ in (20.4) and (20.5) are the same. Let $\mathcal{E}=\mathcal{E}_f$ be the finite set of equivalence classes $\Gamma$ such that the function $\psi_\Gamma^T(H,f)$ is not identically zero. It then follows from (20.4) that

$$\sum_{\Gamma\in\mathcal{E}} \mathrm{e}^{X_\Gamma(H)}\psi_\Gamma^T(H,f) - p^T(H,f) = 0,$$

whenever $d_{P_0}(T) > C_0(1+N+\|H\|)$. The proof of Theorem 20.5 rests on an argument that combines this last identity with the inequality (20.5). We shall describe it in detail for a special case.

Suppose that there is only one class $\Gamma$, and that $X_\Gamma = 0$. In other words, if $\pi$ indexes a nonzero summand in (20.4), $X_\pi$ vanishes. The identity (20.4) becomes

$$\psi_\Gamma^T(H,f) - p^T(H,f) = 0, \qquad d_{P_0}(T) > C_0(1+N+\|H\|). \tag{20.6}$$

It is easy to deduce in this case that $p^T(H,f)$ is a slowly increasing function of $H$. In fact, we claim that for every invariant differential operator $D$ on $\mathfrak{h}^1$, there is a constant $c_{D,f}$ such that

$$|Dp^T(H,f)| \leq c_{D,f}(1+\|H\|)^{d_0}(1+\|T\|)^{d_0}, \tag{20.7}$$

for all $H\in\mathfrak{h}^1$ and $T\in\mathfrak{a}_0$. Since $p^T(H,f)$ is a polynomial in $T$ whose degree is bounded by $d_0$, it would be enough to establish an estimate for each of the coefficients of $p^T(H,f)$ as functions of $H$. For any $H$, we choose $T$ so that $d_{P_0}(T)$ is greater than $C_0(1+N+\|H\|)$, but so that $\|T\|$ is less than $C_1(1+\|H\|)$, for some large constant $C_1$ (depending on $C_0$ and $N$). It follows from (20.6) and (20.5) that

$$\begin{aligned}|Dp^T(H,f)| = |D\psi_\Gamma^T(H,f)| &\leq c_{D,f}(1+\|T\|)^{d_0}\\ &\leq c_{D,f}\big(1+C_1(1+\|H\|)\big)^{d_0} \leq c'_{D,f}(1+\|H\|)^{d_0},\end{aligned}$$

for some constant $c'_{D,f}$. Letting $T$ vary within the chosen domain, we obtain a similar estimate for each of the coefficients of $p^T(H,f)$ by interpolation. The claimed inequality (20.7) follows.

We shall now prove Theorem 20.5(a), in the special case under consideration. We can write

$$B(\nu) = \int_{\mathfrak{h}_1} \mathrm{e}^{\nu(H)}\beta(H)\mathrm{d}H,$$

where $\beta\in\mathcal{S}(\mathfrak{h}^1)^W$ is the standard Fourier transform $\widehat{B}$ of the given function $B\in\mathcal{S}(i\mathfrak{h}^*/i\mathfrak{a}_G^*)^W$. We then form the integral

$$P^T(B,f) = p^T(\beta,f) = \int_{\mathfrak{h}^1} p^T(H,f)\beta(H)\mathrm{d}H,$$

which converges by (20.7). This is the required polynomial in $T$. We have to show that it is asymptotic to the expression

$$\sum_{P\supset P_0}\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*}\sum_\pi \Psi_\pi^T(\lambda,f)B_\pi(\lambda)\mathrm{d}\lambda. \tag{20.8}$$

We write the expression (20.8) as

$$\begin{aligned}
&\sum_P \int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \sum_\pi \Psi_\pi^T(\lambda, f) \int_{\mathfrak{h}^1} \mathrm{e}^{(iY_\pi+\lambda)(H)}\beta(H)\mathrm{d}H\mathrm{d}\lambda \\
&= \int_{\mathfrak{h}^1} \sum_P \sum_\pi \psi_\pi^T(H, f)\mathrm{e}^{iY_\pi(H)}\beta(H)\mathrm{d}H \\
&= \int_{\mathfrak{h}^1} \sum_P \sum_\pi |W|^{-1} \sum_{s\in W} \psi_\pi^T(s^{-1}H, f)\mathrm{e}^{iY_\pi(s^{-1}H)}\beta(H)\mathrm{d}H \\
&= \int_{\mathfrak{h}^1} \psi_\Gamma^T(H, f)\beta(H)\mathrm{d}H,
\end{aligned}$$

by the definition of $B_\pi(\lambda)$, the definition of $\psi_\pi^T(H, f)$, the fact that $\beta(H)$ is $W$-symmetric, and our assumption that $X_\Gamma = 0$. It follows that the difference (20.2) between $p^T(B, f)$ and (20.8) has absolute value bounded by the integral

$$\int_{\mathfrak{h}^1} |\psi_\Gamma^T(H, f) - p^T(H, f)||\beta(H)|\mathrm{d}H.$$

We can assume that $T$ lies in a fixed cone $\mathfrak{a}_{P_0}^r$, and is large. If $d_{P_0}(T)$ is greater than $C_0(1+N+\|H\|)$, the integrand vanishes by (20.6). We may therefore restrict the domain of integration to the subset of points $H \in \mathfrak{h}^1$ with

$$\|H\| \geq C_0^{-1}d_{P_0}(T) - (1+N) \geq C_0^{-1}r\|T\| - (1+N) \geq r_1\|T\|,$$

for some fixed positive number $r_1$. For any such $H$, we have

$$\begin{aligned}
|\psi_\Gamma^T(H, f) - p^T(H, f)| &\leq |\psi_\Gamma^T(H, f)| + |p^T(H, f)| \\
&\leq c_1(1+\|H\|)^{2d_0},
\end{aligned}$$

for some $c_1 > 0$, by (20.5) and (20.7). We also have

$$|\beta(H)| \leq c_2(1+\|H\|)^{-(1+2d_0+2\dim\,\mathfrak{h}^1)}, \qquad H \in \mathfrak{h}^1,$$

for some $c_2 > 0$. The integral is therefore bounded by

$$c_1c_2 \int_{\|H\|\geq r_1\|T\|} (1+\|H\|)^{2d_0}(1+\|H\|)^{-(1+2d_0+2\dim\,\mathfrak{h}^1)}\mathrm{d}H,$$

a quantity that is in turn bounded by an expression

$$c_1c_2r_1^{-1}\|T\|^{-1} \int_{\mathfrak{h}^1} (1+\|H\|)^{-2\dim\,\mathfrak{h}^1}\mathrm{d}H$$

that approaches 0 as $T$ approaches infinity. It follows that the difference (20.2) approaches 0 as $T$ approaches infinity in $\mathfrak{a}_{P_0}^r$. We have established Theorem 20.5(a), in the special case under consideration, by combining (20.5), (20.6), and (20.7).

Next we prove Theorem 20.5(b), in the given special case. Recall that $J_\chi^T(f)$ is the value of $p^T(H, f)$ at $H = 0$. We have to show that this equals the limit of $P^T(B^\varepsilon, f)$ as $\varepsilon$ approaches 0, under the assumption that $B(0) = 1$. Now

$$(\widehat{B}^\varepsilon)(H) = (\widehat{B})_\varepsilon(H) = \beta_\varepsilon(H),$$

where

$$\beta_\varepsilon(H) = \varepsilon^{-(\dim\,\mathfrak{h}^1)}\beta(\varepsilon^{-1}H).$$

Therefore

$$\begin{aligned} P^T(B^\varepsilon, f) = p^T(\beta_\varepsilon, f) &= \int_{\mathfrak{h}^1} p^T(H,f)\beta_\varepsilon(H)\mathrm{d}H \\ &= J_\chi^T(f) + \int_{\mathfrak{h}^1} \big(p^T(H,f) - p^T(0,f)\big)\beta_\varepsilon(H)\mathrm{d}H, \end{aligned}$$

since $J_\chi^T(f) = p^T(0,f)$, and $\int \beta_\varepsilon = B^\varepsilon(0) = 1$. But if we combine the mean value theorem with (20.7), we see that

$$|p^T(H,f) - p^T(0,f)| \le c\|H\|(1+\|H\|)^{d_0}(1+\|T\|)^{d_0},$$

for some fixed $c > 0$, and all $H$ and $T$. We can assume that $\varepsilon \le 1$. Then

$$\begin{aligned} &\int_{\mathfrak{h}^1} |p^T(H,f) - p^T(0,f)||\beta_\varepsilon(H)|\mathrm{d}H \\ &= \varepsilon^{-\dim(\mathfrak{h}^1)} \int_{\mathfrak{h}^1} |p^T(H,f) - p^T(0,f)||\beta(\varepsilon^{-1}H)|\mathrm{d}H \\ &= \int_{\mathfrak{h}^1} |p^T(\varepsilon H,f) - p^T(0,f)||\beta(H)|\mathrm{d}H \\ &\le c\int_{\mathfrak{h}^1} \varepsilon\|H\|(1+\|\varepsilon H\|)^{d_0}(1+\|T\|)^{d_0}|\beta(H)|\mathrm{d}H \\ &\le c'\varepsilon(1+\|T\|)^{d_0}, \end{aligned}$$

where

$$c' = c\int_{\mathfrak{h}^1} \|H\|(1+\|H\|)^{d_0}|\beta(H)|\mathrm{d}H.$$

It follows that

$$\lim_{\varepsilon\to 0}\big(p^T(B^\varepsilon,f) - J_\chi^T(f)\big) = 0,$$

as required.

We have established Theorem 20.5 in the special case that there is only one class $\Gamma \in \mathcal{E}$, and that $X_\Gamma = 0$. In general, there are several classes, so there can be nonzero points $X_\Gamma$. In place of (20.6), we have the more general identity

$$\sum_\Gamma \mathrm{e}^{X_\Gamma(H)}\psi_\Gamma^T(H,f) - p^T(H,f) = 0, \qquad d_{P_0}(T) > C_0(1+N+\|H\|).$$

In particular, $p^T(H,f)$ can have exponential growth in $H$, and need not be tempered. It cannot be integrated against a Schwartz function $\beta$ of $H$. Now each function $\psi_\Gamma^T(H,f)$ is tempered in $H$, by (20.7). The question is whether it is asymptotic to a polynomial in $T$. In other words, does the polynomial $p^T(H,f)$ have a $\Gamma$-component $\mathrm{e}^{X_\Gamma(H)}p_\Gamma^T(H,f)$?

To answer the question, we take $H_\Gamma$ to be the point in $\mathfrak{h}^1$ such that the inner product $(H_\Gamma, H)$ equals $X_\Gamma(H)$, for each $H \in \mathfrak{h}$. We claim that for fixed $H$, the function

$$t \longrightarrow \psi_\Gamma^T(tH_\Gamma + H, f), \qquad t \in \mathbb{R},$$

is a finite linear combination of unitary exponential functions. To see this, we first note that the function equals

$$|W|^{-1}\sum_{(P,\pi,s)\in\Gamma} \mathrm{e}^{iY_\pi(s^{-1}(tH_\Gamma+H))}\psi_\pi^T\big(s^{-1}(tH_\Gamma+H),f\big).$$

For any $(P, \pi, s) \in \Gamma$, the linear form $X_\pi = s^{-1}X_\Gamma$ is the real part of the infinitestimal character of a unitary representation $\pi$ of $M_P(\mathbb{A})^1$. It follows that the corresponding point $s^{-1}H_\Gamma$ in $\mathfrak{h}^1$ lies in the kernel $\mathfrak{h}^P$ of the projection of $\mathfrak{h}$ onto $\mathfrak{a}_P$. On the other hand, the function

$$\psi_\pi^T(H, f) = \int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \Psi_\pi^T(\lambda, f)\mathrm{e}^{\lambda(H)}d\lambda$$

is invariant under translation by $\mathfrak{h}^P$. Consequently

$$\psi_\pi^T\big(s^{-1}(tH_\Gamma + H)\big) = \psi_\pi^T(s^{-1}H).$$

The claim follows.

It is now pretty clear that we can construct the $\Gamma$-component of the polynomial

$$p^T(H, f) = \sum_{\Gamma\in\mathcal{E}} \mathrm{e}^{X_\Gamma(H)}\psi_\Gamma^T(H, f), \qquad d_{P_0}(T) \geq C_0(1 + N + \|H\|),$$

in terms of its direction of real exponential growth. If one examines the question more closely, taking into consideration the derivation of (20.7) above, one obtains the following lemma.

Lemma 20.6. *There are functions*

$$p_\Gamma^T(H, f) \qquad H \in \mathfrak{h}^1,\ \Gamma \in \mathcal{E},$$

*which are smooth in $H$ and polynomials in $T$ of degree at most $d_0$, such that*

$$p^T(H, f) = \sum_{\Gamma\in\mathcal{E}} \mathrm{e}^{X_\Gamma(H)}p_\Gamma^T(H, f),$$

*and such that if $D$ is any invariant differential operator on $\mathfrak{h}^1$, then*

$$(20.6)' \qquad \big|D\big(\psi_\Gamma^T(H, f) - p_\Gamma^T(H, f)\big)\big| \leq c_{D,f}\mathrm{e}^{-\delta d_{P_0}(T)}(1 + \|T\|)^{d_0},$$

*for all $H$ and $T$ with $d_{P_0}(T) > C_0(1 + N + \|H\|)$, and*

$$(20.7)' \qquad |Dp_\Gamma^T(H, f)| \leq c_{D,f}(1 + \|H\|)^{d_0}(1 + \|T\|)^{d_0},$$

*for all $H$ and $T$, with $C_0$, $\delta$ and $c_{D,f}$ being positive constants.*

See [**A7**, Proposition 5.1]. □

Given Lemma 20.6, we set

$$p_\Gamma^T(\beta, f) = \int_{\mathfrak{h}^1} p_\Gamma^T(H, f)\beta(H)\mathrm{d}H,$$

for any function $\beta \in \mathcal{S}(\mathfrak{h}^1)^W$ and any $\Gamma \in \mathcal{E}$. We then argue as above, using the inequalities (20.5), (20.6)′ and (20.7)′ in place of (20.5), (20.6), and (20.7). We deduce that for any $\Gamma$ and $\beta$,

$$(20.9(\text{a})) \qquad \lim_{T\to\infty}\Big(\int_{\mathfrak{h}^1} \psi_\Gamma^T(H, f)\beta(H)\mathrm{d}H - p_\Gamma^T(\beta, f)\Big) = 0, \qquad T \in \mathfrak{a}_{P_0}^r,$$

and that

$$(20.9(\text{b})) \qquad \lim_{\varepsilon\to 0} p_\Gamma^T(\beta_\varepsilon, f) = p_\Gamma^T(0, f),$$

if $\int \beta = 1$, exactly as in the proofs of (a) and (b) in the special case of Theorem 20.5 above. (See [**A7**, Lemmas 6.2 and 6.1].)

To establish Theorem 20.5 in general, we set

$$P^T(B,f) = \sum_{\Gamma\in\mathcal{E}} p_\Gamma^T(\beta,f), \qquad \beta = \widehat{B}.$$

Then, as in the proof of the special case of Theorem 20.5(a) above, we deduce that

$$\sum_P \int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \sum_\pi \Psi_\pi^T(\lambda,f)B_\pi(\lambda)\mathrm{d}\lambda = \sum_{\Gamma\in\mathcal{E}} \int_{\mathfrak{h}^1} \psi_\Gamma^T(H,f)\beta(H)\mathrm{d}H.$$

It follows from (20.9(a)) that the difference between the expression on the right hand side of this identity and $P^T(B,f)$ approaches 0 as $T$ approaches infinity in $\mathfrak{a}_{P_0}^r$. The same is therefore true of the difference between the expression on the left hand side of the identity and $P^T(B,f)$. This gives Theorem 20.5(a). For Theorem 20.5(b), we use (20.9(b)) to write

$$\begin{aligned}\lim_{\varepsilon\to 0} P^T(B^\varepsilon,f) &= \lim_{\varepsilon\to 0}\sum_{\Gamma\in\mathcal{E}} p_\Gamma^T(\beta_\varepsilon,f)\\ &= \sum_{\Gamma\in\mathcal{E}} p_\Gamma^T(0,f) = p^T(0,f) = J_\chi^T(f),\end{aligned}$$

if $B(0) = \int\beta = 1$. This completes our discussion of the proof of Theorem 20.5. □

## 21. The fine spectral expansion

We have taken care of the primary analytic obstruction to computing the distributions $J_\chi(f)$. Its resolution is contained in Theorem 20.5, which applies to objects $\chi\in\mathcal{X}$, $P_0\in\mathcal{P}(M_0)$, $f\in\mathcal{H}(G)$, and $B\in\mathcal{S}(i\mathfrak{h}^*/i\mathfrak{a}_G^*)^W$, with $B(0)=1$. We take $B$ to be compactly supported. The function

$$B_\pi(\lambda) = B(iY_\pi+\lambda), \qquad \lambda\in i\mathfrak{a}_P^*,$$

attached to any $P\supset P_0$ and $\pi\in\Pi_{\mathrm{unit}}\big(M_P(\mathbb{A})^1\big)$ then belongs to $C_c^\infty(i\mathfrak{a}_P^*/i\mathfrak{a}_G^*)$.

Suppose that $a^T$ and $b^T$ are two functions defined on some cone $d_{P_0}(T)>C_0$ in $\mathfrak{a}_0$. We shall write $a^T\sim b^T$ if $a^T-b^T$ approaches 0 as $T$ approaches infinity in any cone $\mathfrak{a}_{P_0}^r$. Theorem 20.5(a) tells us that

$$\begin{aligned}P^T(B,f) &\sim \sum_{P\supset P_0}\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*}\sum_\pi \Psi_\pi^T(\lambda,f)B_\pi(\lambda)\mathrm{d}\lambda\\ &= \sum_{P\supset P_0} n_P^{-1}\sum_\pi\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*}\mathrm{tr}\big(M_{P,\chi}^T(\lambda)\mathcal{I}_{P,\chi,\pi}(\lambda,f)\big)B_\pi(\lambda)\mathrm{d}\lambda,\end{aligned}$$

where $P^T(B,f)$ is a polynomial in $T$ that depends linearly on $B$. The fact that each $B_\pi(\lambda)$ has compact support is critical. It removes the analytic problem of reconciling an asymptotic limit in $T$ with an integral in $\lambda$ over a noncompact space. Our task is to compute $P^T(B,f)$ explicitly, as a bilinear form in the functions $\{B_\pi(\lambda)\}$ and the operators $\{\mathcal{I}_{P,\chi,\pi}(\lambda,f)\}$. We will then obtain an explicit formula for $J_\chi^T(f)$ from the assertion

$$J_\chi^T(f) = \lim_{\varepsilon\to 0} P^T(B^\varepsilon,f)$$

of Theorem 20.5(b).

The operator $M_{P,\chi}^T(\lambda)$ is defined by (15.1) in terms of an inner product of truncated Eisenstein series attached to $P$. Proposition 15.3 gives the explicit inner product formula of Langlands, which applies to the special case that the Eisenstein series are cuspidal. It turns out that the same formula holds asymptotically in $T$ for arbitrary Eisenstein series.

THEOREM 21.1. *Suppose that $\phi \in \mathcal{H}_P^0$ and $\phi' \in \mathcal{H}_{P'}^0$, for standard parabolic subgroups $P, P' \supset P_0$. Then the difference between the inner product*

$$\int_{G(F)\backslash G(\mathbb{A})^1} \Lambda^T E(x,\phi,\lambda)\overline{\Lambda^T E(x,\phi',\lambda')}\mathrm{d}x$$

*and the sum*

$$\sum_Q \sum_s \sum_{s'} \theta_Q(s\lambda + s'\overline{\lambda}')^{-1}\mathrm{e}^{(s\lambda+s'\overline{\lambda}')(T)}\big(M(s,\lambda)\phi, M(s',\lambda')\phi'\big) \tag{21.1}$$

*over $Q \supset P_0$, $s \in W(\mathfrak{a}_P, \mathfrak{a}_Q)$, and $s' \in W(\mathfrak{a}_{P'}, \mathfrak{a}_Q)$ is bounded by a product*

$$c(\lambda,\lambda',\phi,\phi')\mathrm{e}^{-\varepsilon d_{P_0}(T)},$$

*where $\varepsilon > 0$, and $c(\lambda,\lambda',\phi,\phi')$ is a locally bounded function on the set of points $\lambda \in \mathfrak{a}_{P,\mathbb{C}}^*$ and $\lambda' \in \mathfrak{a}_{P',\mathbb{C}}^*$ at which the Eisenstein series are analytic.*

This is [**A6**, Theorem 9.1], which is the main result of the paper [**A6**]. The proof begins with the special case already established for cuspidal Eisenstein series in Proposition 15.3. One then uses the results of Langlands in [**Lan5**, §7], which express arbitrary Eisenstein series in terms of residues of cuspidal Eisenstein series. This process is not canonical in general. Nevertheless, one can still show that (21.1) is an asymptotic approximation for the expression obtained from the appropriate residues of the corresponding formula for cuspidal Eisenstein series. □

Let us write $\omega^T(\lambda,\lambda',\phi,\phi')$ for the expression (21.1). If $B_\chi$ is any function in $C_c^\infty(i\mathfrak{a}_P^*/i\mathfrak{a}_G^*)$, the theorem tells us that

$$\begin{aligned}&\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \big(M_{P,\chi}^T(\lambda)\mathcal{I}_{P,\chi}(\lambda,f)\phi,\phi\big)B_\chi(\lambda)\mathrm{d}\lambda\\ \sim &\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*} \omega^T\big(\lambda,\lambda,\mathcal{I}_{P,\chi}(\lambda,f)\phi,\phi\big)B_\chi(\lambda)\mathrm{d}\lambda.\end{aligned}$$

We shall apply this asymptotic formula to the functions $B_\chi = B_\pi$. Since $f$ is $K$-finite, $\mathcal{I}_{P,\chi}(\lambda,f)\phi$ vanishes for all but finitely many vectors $\phi$ in the orthonormal basis $\mathcal{B}_{P,\chi}$ of $\mathcal{H}_{P,\chi}$. This is a consequence of Langlands' construction of the discrete spectrum, as we have noted earlier. We assume that $\mathcal{B}_{P,\chi}$ is a disjoint union of orthonormal bases $\mathcal{B}_{P,\chi,\pi}$ of the spaces $\mathcal{H}_{P,\chi,\pi}$. It then follows that

$$\begin{aligned}&P^T(B,f)\\ \sim &\sum_{P\supset P_0} n_P^{-1}\sum_\pi \Big(\int_{i\mathfrak{a}_P^*/i\mathfrak{a}_G^*}\sum_{\phi\in\mathcal{B}_{P,\chi,\pi}} \omega^T\big(\lambda,\lambda,\mathcal{I}_{P,\chi,\pi}(\lambda,f)\phi,\phi\big)B_\pi(\lambda)\mathrm{d}\lambda\Big).\end{aligned}$$

The problem is to find an explicit polynomial function of $T$, for any $P$ and $\pi$, which is asymptotic in $T$ to the expression in the brackets.

Suppose that $P$, $\pi$, and $\phi$ are fixed, and that $\lambda$ lies in $i\mathfrak{a}_P^*$. Changing the indices of summation in the definition (21.1), we write

$$\omega^T\big(\lambda,\lambda,\mathcal{I}_{P,\chi,\pi}(\lambda,f)\phi,\phi\big)$$

as the limit as $\lambda'$ approaches $\lambda$ of the expression

$$\begin{aligned}&\omega^T\big(\lambda',\lambda,\mathcal{I}_P(\lambda,f)\phi,\phi\big)\\ =&\sum_{P_1}\sum_{t'}\sum_{t}\theta_{P_1}(t'\lambda'-t\lambda)^{-1}\mathrm{e}^{(t'\lambda'-t\lambda)(T)}\big(M(t',\lambda')\mathcal{I}_P(\lambda,f)\phi,M(t,\lambda)\phi\big)\\ =&\sum_{s}\sum_{P_1}\sum_{t}\theta_{P_1}\big(t(s\lambda'-\lambda)\big)^{-1}\mathrm{e}^{(t(s\lambda'-\lambda))(T)}\big(M(ts,\lambda')\mathcal{I}_P(\lambda,f)\phi,M(t,\lambda)\phi\big),\end{aligned}$$

for sums over $P_1\supset P_0$ and $t',t\in W(\mathfrak{a}_P,\mathfrak{a}_{P_1})$, and for $s=t^{-1}t'$ ranging over the group $W(M_P)=W(\mathfrak{a}_P,\mathfrak{a}_P)$. Since $\lambda$ is purely imaginary, the adjoint of the operator $M(t,\lambda)$ equals $M(t,\lambda)^{-1}$. The sum

$$\sum_{\phi\in\mathcal{B}_{P,\chi,\pi}}\omega^T\big(\lambda,\lambda,\mathcal{I}_{P,\chi,\pi}(\lambda,f)\phi,\phi\big)\tag{21.2}$$

therefore equals the limit as $\lambda'$ approaches $\lambda$ of

$$\sum_{s}\sum_{(P_1,t)}\theta_{P_1}\big(t(s\lambda'-\lambda)\big)^{-1}\mathrm{e}^{(t(s\lambda'-\lambda))(T)}\mathrm{tr}\big(M(t,\lambda)^{-1}M(ts,\lambda')\mathcal{I}_{P,\chi,\pi}(\lambda,f)\big).$$

Set $M=M_P$. The correspondence

$$(P_1,t)\longrightarrow Q=w_t^{-1}P_1w_t,\qquad\qquad P_1\supset P_0,\ t\in W(\mathfrak{a}_P,\mathfrak{a}_{P_1}),$$

is then a bijection from the set of pairs $(P_1,t)$ in the last sum onto the set $\mathcal{P}(M)$. For any group $Q\in\mathcal{P}(M)$ and any element $s\in W(M)$, there is a unitary intertwining operator

$$M_{Q|P}(s,\lambda):\ \mathcal{H}_P\longrightarrow\mathcal{H}_Q,\qquad\qquad\lambda\in i\mathfrak{a}_M^*.$$

It is defined by analytic continuation from the analogue of the integral formula (7.2), in which $P'$ is replaced by $Q$. If $(P_1,t)$ is the preimage of $Q$, it is easy to see from the definitions that

$$M(ts,\lambda')=tM_{Q|P}(s,\lambda')\mathrm{e}^{(s\lambda'+\rho_Q)(T_0-t^{-1}T_0)},$$

where $t$: $\mathcal{H}_Q\to\mathcal{H}_{P_1}$ is the operator defined by

$$(t\phi)(x)=\phi(w_t^{-1}x),\qquad\qquad\phi\in\mathcal{H}_P.$$

The point $T_0$ is used as in §15 to measure the discrepancy between the two representatives $w_t$ and $\widetilde{w}_t$ of the element $t\in W_0$. (See [**A8**, (1.4)].) It follows that

$$M(t,\lambda)^{-1}M(ts,\lambda')=M_{Q|P}(\lambda)^{-1}M_{Q|P}(s,\lambda')\mathrm{e}^{(s\lambda'-\lambda)(T_0-t^{-1}T_0)},$$

where $M_{Q|P}(\lambda)=M_{Q|P}(1,\lambda)$. Next, we define a point $Y_Q(T)$ to be the projection onto $\mathfrak{a}_M$ of the point

$$t^{-1}(T-T_0)+T_0.$$

Then

$$\mathrm{e}^{(t(s\lambda'-\lambda))(T)}\mathrm{e}^{(s\lambda'-\lambda)(T_0-t^{-1}T_0)}=\mathrm{e}^{(s\lambda'-\lambda)(Y_Q(T))}.$$

Finally, it is clear that

$$\theta_{P_1}\big(t(s\lambda'-\lambda)\big)^{-1}=\theta_Q(s\lambda'-\lambda)^{-1}.$$

It follows that (21.2) equals the limit as $\lambda'$ approaches $\lambda$ of the sum over $s\in W(M)$ of

$$\sum_{Q\in\mathcal{P}(M)}\mathrm{tr}\big(M_{Q|P}(\lambda)^{-1}M_{Q|P}(s,\lambda')\mathcal{I}_{P,\chi,\pi}(\lambda,f)\big)\mathrm{e}^{(s\lambda'-\lambda)(Y_Q(T))}\theta_Q(s\lambda'-\lambda)^{-1}.\tag{21.3}$$

The expression (21.3) looks rather like the basic function (17.1) we have attached to any $(G, M)$-family. We shall therefore study it as a function of the variable

$$\Lambda = s\lambda' - \lambda.$$

The expression becomes

$$\sum_{Q\in\mathcal{P}(M)} c_Q(\Lambda)d_Q(\Lambda)\theta_Q(\Lambda)^{-1},$$

where

$$c_Q(\Lambda) = \mathrm{e}^{\Lambda(Y_Q(T))},$$

and

$$d_Q(\Lambda) = \mathrm{tr}\big(M_{Q|P}(\lambda)^{-1}M_{Q|P}(s,\lambda')\mathcal{I}_{\mathcal{P},\chi,\pi}(\lambda, f)\big).$$

It follows easily from the definition of $Y_Q(T)$ that $\{c_Q(\Lambda)\}$ is a $(G, M)$-family. The operators $M_{Q|P}(s, \lambda')$ in the second factor satisfy a functional equation

$$M_{Q'|P}(s,\lambda') = M_{Q'|Q}(s\lambda')M_{Q|P}(s,\lambda'), \qquad Q' \in \mathcal{P}(M).$$

It follows easily from this that $\{d_Q(\Lambda)\}$ is also a $(G, M)$-family. (See [**A8**, p. 1298]. Of course $d_Q(\Lambda)$ depends on the kernel of the mapping $(\lambda', \lambda) \to \Lambda$ as well as on $\Lambda$, but at the moment we are only interested in the variable $\Lambda$.) The expression (21.3) therefore reduces to something we have studied, namely the function $(cd)_M(\Lambda)$ attached to the $(G, M)$-family $\{(cd)_Q(\Lambda)\}$. By Lemma 17.1, the function has no singularities in $\Lambda$. It follows that the expression (21.3) extends to a smooth function of $(\lambda', \lambda)$ in $i\mathfrak{a}^*_M \times i\mathfrak{a}^*_M$.

Remember that we are supposed to take the limit, as $\lambda'$ approaches $\lambda$, of the sum over $s \in W(M)$ of (21.3). We will then want to integrate the product of $B_\pi(\lambda)$ with the resulting function of $\lambda$ over the space $i\mathfrak{a}^*_M/i\mathfrak{a}^*_G$. From what we have just observed, the integral and limit may be taken inside the sum over $s$. It turns out that the asymptotic limit in $T$ may also be taken inside the sum over $s$. In other words, it is possible to find an explicit polynomial in $T$ that is asymptotic to the integral over $\lambda$ of the product $B_\pi(\lambda)$ with value at $\lambda' = \lambda$ of (21.3). We shall describe how to do this, using the product formula of Lemma 17.4.

Suppose that $s \in W(M)$ is fixed. Let $L$ be the smallest Levi subgroup in $\mathcal{L}(M)$ that contains a representative of $s$. Then $\mathfrak{a}_L$ equals the kernel of $s$ in $\mathfrak{a}_M$. The element $s$ therefore belongs to the subset

$$W^L(M)_{\mathrm{reg}} = \{t \in W^L(M) : \ \ker(t) = \mathfrak{a}_L\},$$

of regular elements in $W^L(M)$. Given $s$, we set $\lambda' = \lambda + \zeta$, where $\zeta$ is restricted to lie in the subspace $i\mathfrak{a}^*_L$ of $i\mathfrak{a}^*_M$ associated to $s$. Then $s\zeta = \zeta$, and

$$\Lambda = (s\lambda - \lambda) + \zeta$$

is the decomposition of $\Lambda$ relative to the direct sum

$$i\mathfrak{a}^*_M = i(\mathfrak{a}^L_M)^* \oplus i\mathfrak{a}^*_L.$$

If $\lambda_L$ is the projection of $\lambda$ onto $i\mathfrak{a}^*_L$, the mapping

$$(\lambda, \zeta) \longrightarrow (\Lambda, \lambda_L), \qquad \lambda \in i\mathfrak{a}^*_M,\ \zeta \in i\mathfrak{a}^*_L,$$

is a linear automorphism of the vector space $i\mathfrak{a}^*_M \oplus i\mathfrak{a}^*_L$. In particular, the points $\lambda$ and $\lambda' = \lambda + \zeta$ are uniquely determined by $\Lambda$ and $\lambda_L$. Let us write

$$c_Q(\Lambda, T) = \mathrm{e}^{\Lambda(Y_Q(T))}$$

and

$$d_Q(\Lambda, \lambda_L) = \mathrm{tr}\big(M_{Q|P}(\lambda)^{-1} M_{Q|P}(s, \lambda + \zeta)\mathcal{I}_{P,\chi,\pi}(\lambda, f)\big),$$

in order to keep track of our two $(G, M)$-families on the supplementary variables. They of course remain $(G, M)$-families in the variable $\Lambda$. For $\lambda' = \lambda + \zeta$ as above, the expression (21.3) equals

$$\sum_{Q\in\mathcal{F}(M)} c_Q(\Lambda, T) d_Q(\Lambda, \lambda_L)\theta_Q(\Lambda)^{-1} = \sum_{S\in\mathcal{F}(M)} c^S_M(\Lambda, T) d'_S(\Lambda_S, \lambda_L),$$

by the product formula of Lemma 17.4. To evaluate (21.3) at $\lambda' = \lambda$, we set $\zeta = 0$. This entails simply replacing $\Lambda$ by $s\lambda - \lambda$. The value of (21.3) at $\lambda' = \lambda$ therefore equals

$$\sum_{S\in\mathcal{F}(M)} c^S_M(s\lambda - \lambda, T) d'_S\big((s\lambda - \lambda)_S, \lambda_L\big).$$

We have therefore to consider the integral

$$\int_{i\mathfrak{a}^*_M/i\mathfrak{a}^*_G} \Big(\sum_{S\in\mathcal{F}(M)} c^S_M(s\lambda - \lambda, T) d'_S\big((s\lambda - \lambda)_S, \lambda_L\big)\Big) B_\pi(\lambda)\mathrm{d}\lambda, \tag{21.4}$$

for $M = M_P$, $\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$, $L \in \mathcal{L}(M)$ and $s \in W^L(M)_{\mathrm{reg}}$, and for $T$ in a fixed domain $\mathfrak{a}^r_{P_0}$. We need to show that the integral is asymptotic to an explicit polynomial in $T$. This will allow us to construct $P^T(B)$ simply by summing the product of this polynomial with $n_P^{-1}$ over $P \supset P_0$, $\pi$, $L$, and $s$.

We first decompose the integral (21.4) into a double integral over $i(\mathfrak{a}^L_M)^*$ and $i\mathfrak{a}^*_L/i\mathfrak{a}^*_G$. If $\lambda$ belongs to $i\mathfrak{a}^*_M$, $s\lambda - \lambda$ depends only on the projection $\mu$ of $\lambda$ onto $i(\mathfrak{a}^L_M)^*$. Since the mapping

$$F_s : \ \mu \longrightarrow s\mu - \mu$$

is a linear isomorphism of $i(\mathfrak{a}^L_M)^*$, (21.4) equals the product of the inverse

$$|\det(s-1)_{\mathfrak{a}^L_M}|^{-1}$$

of the determinant of this mapping with the sum over $S \in \mathcal{F}(M)$ of

$$\int_{i(\mathfrak{a}^L_M)^*}\int_{i\mathfrak{a}^*_L/i\mathfrak{a}^*_G} c^S_M(\mu, T) d'_S(\mu_S, \lambda) B_\pi\big(F_s^{-1}(\mu) + \lambda\big)\mathrm{d}\lambda\mathrm{d}\mu. \tag{21.5}$$

Next, we note that the dependence of the integral on $T$ is through the term $c^S_M(\mu, T)$. For fixed $S$, the set

$$\mathcal{Y}^S_M(T) = \{Y_{S(R)}(T) : \ R \in \mathcal{P}^{M_S}(M)\}$$

is a positive $(M_S, M)$-orthogonal set of points in $\mathfrak{a}_M$, which all project to a common point $Y_S(T)$ in $\mathfrak{a}_S$. It follows from Lemma 17.2 that

$$c^S_M(\mu, T) = \int_{Y_S(T)+\mathfrak{a}^{M_S}_M} \psi^S_M(H, T)\mathrm{e}^{\mu(H)}\mathrm{d}H,$$

where $\psi^S_M(\cdot, T)$ is the characteristic function of the convex hull in $\mathfrak{a}_M$ of $\mathcal{Y}^S_M(T)$. We can therefore write (21.5) as

$$\int_{Y_S(T)+\mathfrak{a}^{M_S}_M} \psi^S_M(H, T)\phi_S(H)\mathrm{d}H, \tag{21.6}$$

where

$$\phi_S(H) = \int_{i(\mathfrak{a}_M^L)^*} \int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} \mathrm{e}^{\mu(H)} d'_S(\mu_S, \lambda) B_\pi\big(F_s^{-1}(\mu) + \lambda\big) \mathrm{d}\lambda \mathrm{d}\mu,$$

for any $H \in \mathfrak{a}_M$. Since $d'_S(\cdot,\cdot)$ is smooth, and $B_\pi(\cdot)$ is both smooth and compactly supported, $\phi_S(H)$ is a Schwartz function on $\mathfrak{a}_M/\mathfrak{a}_L$.

There are two cases to consider. Suppose first that $S$ does not belong to the subset $\mathcal{F}(L)$ of $\mathcal{F}(M)$. Then $\mathfrak{a}_S$ is not contained in $\mathfrak{a}_L$, and $Y_S(T)$ projects to a nonzero point $Y_S(T)_M^L$ in $\mathfrak{a}_M^L$. In fact, it follows easily from the fact that $T$ lies in $\mathfrak{a}_{P_0}^r$ that

$$\|Y_S(T)_M^L\| \geq r_1 \|T\|,$$

for some $r_1 > 0$. The function $\psi_M^S(\cdot, T)$ is supported on a compact subset of the affine space $Y_S(T) + \mathfrak{a}_M^{M_S}$ whose volume is bounded by a polynomial in $T$. One combines this with the fact that $\phi_S(H)$ is a Schwartz function on $\mathfrak{a}_M/\mathfrak{a}_L$ to show that (21.6) approaches 0 as $T$ approaches infinity in $\mathfrak{a}_{P_0}^r$. (See [**A8**, p. 1306].)

We can therefore assume that $S$ belongs to $\mathcal{F}(L)$. Then

$$\mathfrak{a}_M^{M_S} = \mathfrak{a}_M^L \oplus \mathfrak{a}_L^{M_S}.$$

Since $\phi_S$ is $\mathfrak{a}_L$-invariant, we are free to write (21.6) as

$$\int_{Y_S(T)+\mathfrak{a}_L^{M_S}} \Big( \int_{\mathfrak{a}_M^L} \phi_S(U) \psi_M^S(U + H, T) \mathrm{d}U \Big) \mathrm{d}H.$$

As it turns out, we can simplify matters further by replacing

$$\psi_M^S(U + H, T)$$

with $\psi_L^S(H,T)$, where $\psi_L^S(H,T)$ is the characteristic function in $\mathfrak{a}_L$ of the set $\mathcal{Y}_L^S(T)$ obtained in the obvious way from $\mathcal{Y}_M^S(T)$. More precisely, the difference between the last expression and the product

$$\int_{\mathfrak{a}_M^L} \phi_S(U) \mathrm{d}U \cdot \int_{Y_S(T)+\mathfrak{a}_L^{M_S}} \psi_L^S(H,T) \mathrm{d}H \tag{21.7}$$

approaches 0 as $T$ approaches infinity in $\mathfrak{a}_{P_0}^r$. Suppose for example that $G = SL(3)$, $M = M_0$ is minimal, $M_S = G$, and that $L$ is a standard maximal Levi subgroup $M_1$. Then $Y_S(T) = 0$, and the difference

$$\psi_L^S(H,T) - \psi_M^S(U + H, T), \qquad U \in \mathfrak{a}_M^L,\ H \in \mathfrak{a}_L,$$

is the characteristic function of the darker shaded region in Figure 21.1. Since $\phi_S(U)$ is rapidly decreasing on the vertical $\mathfrak{a}_M^L$-axis in the figure, the integral over $(U, H)$ of its product with the difference above does indeed approach 0. In the general case, the lemmas in [**A8**, §3] show that the convex hull of $\mathcal{Y}_M^S(T)$ has the same qualitative behaviour as is Figure 21.1. (See [**A8**, p. 1307–1308].)

The problem thus reduces to the computation of the product (21.7), for any element $S \in \mathcal{F}(L)$. The first factor in the product can be written as

$$\int_{\mathfrak{a}_M^L} \phi_S(U) \mathrm{d}U = \int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} d'_S(0, \lambda) B_\pi(\lambda) \mathrm{d}\lambda,$$

by the Fourier inversion formula in $\mathfrak{a}_M^L$. The second factor equals

$$\int_{Y_S(T)+\mathfrak{a}_L^{M_S}} \psi_L^S(H,T) \mathrm{d}H = c_L^S(0, T)$$

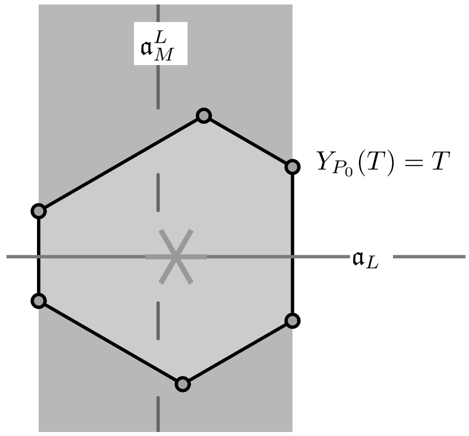



FIGURE 21.1. The vertices represent the six points $Y_P(T)$, as $P$ ranges over $\mathcal{P}(M_0)$. Since $T$ ranges over a set $\mathfrak{a}_{P_0}^r$, the distance from any vertex to the horizontal $\mathfrak{a}_L$-axis is bounded below by a positive multiple of $\|T\|$.

by Lemma 17.2, and is therefore a polynomial in $T$. In particular, (21.7) is already a polynomial in $T$. To express its contribution to the asymptotic value of (21.4), we need only sum $S$ over $\mathcal{F}(L)$. We conclude that (21.4) differs from the polynomial

$$|\det(s-1)_{\mathfrak{a}_M^L}|^{-1} \int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} \Big( \sum_{S \in \mathcal{F}(L)} c_L^S(0,T) d_S'(0,\lambda) \Big) B_\pi(\lambda) \mathrm{d}\lambda \tag{21.8}$$

by an expression that approaches 0 as $T$ approaches infinity in $\mathfrak{a}_{P_0}^r$.

The sum

$$\sum_{S \in \mathcal{F}(L)} c_L^S(0,T) d_S'(0,\lambda) \tag{21.9}$$

in (21.8) comes from a product

$$c_{Q_1}(\Lambda, T) d_{Q_1}(\Lambda, \lambda), \qquad Q_1 \in \mathcal{P}(L),\ \Lambda \in i\mathfrak{a}_L^*,$$

of $(G,L)$-families. By Lemma 17.4, it equals the value at $\Lambda = 0$ of the sum

$$\sum_{Q_1 \in \mathcal{P}(L)} c_{Q_1}(\Lambda, T) d_{Q_1}(\Lambda, \lambda) \theta_{Q_1}(\Lambda)^{-1}.$$

Recall the definition of the $(G,M)$-family $\{d_Q(\Lambda,\lambda)\}$ of which the $(G,L)$-family $\{d_{Q_1}(\Lambda,\lambda)\}$ is the restriction. Since $\lambda$ and $\Lambda$ lie in the subspace $i\mathfrak{a}_L^*$ of $i\mathfrak{a}_M^*$, $\lambda_L$ equals $\lambda$, and

$$\zeta = \Lambda - (s\lambda - \lambda) = \Lambda.$$

It follows from the definitions and the functional equations of the global intertwining operators that

$$\begin{aligned} d_Q(\Lambda,\lambda) &= \mathrm{tr}\big(M_{Q|P}(\lambda)^{-1} M_{Q|P}(s, \lambda+\Lambda) \mathcal{I}_{P,\chi,\pi}(\lambda, f)\big) \\ &= \mathrm{tr}\big(M_{Q|P}(\lambda)^{-1} M_{Q|P}(\lambda+\Lambda) M_{P|P}(s, \lambda+\Lambda) \mathcal{I}_{P,\chi,\pi}(\lambda, f)\big), \end{aligned}$$

for any $Q \in \mathcal{P}(M)$. Since the point $\lambda + \Lambda$ lies in the space $i\mathfrak{a}_L^*$ fixed by $s$, the operator

$$M_P(s, 0) = M_{P|P}(s, \lambda + \Lambda)$$

is independent of $\lambda$ and $\Lambda$. To deal with the other operators, we define

$$\mathcal{M}_Q(\Lambda, \lambda, P) = M_{Q|P}(\lambda)^{-1} M_{Q|P}(\lambda + \Lambda),$$

and

$$\begin{aligned}\mathcal{M}_Q^T(\Lambda, \lambda, P) &= \mathrm{e}^{\Lambda(Y_Q(T))} M_{Q|P}(\lambda)^{-1} M_{Q|P}(\lambda + \Lambda) \\ &= c_Q(\Lambda, T) \mathcal{M}_Q(\Lambda, \lambda, P),\end{aligned}$$

for any $Q \in \mathcal{P}(M)$. As functions of $\Lambda$ in the larger domain $i\mathfrak{a}_M^*$, these objects form two $(G, M)$-families as $Q$ varies over $\mathcal{P}(M)$. With $\Lambda$ restricted to $i\mathfrak{a}_L^*$ as above, the functions

$$\mathcal{M}_{Q_1}^T(\Lambda, \lambda, P) = \mathcal{M}_Q^T(\Lambda, \lambda, P), \qquad\qquad Q_1 \in \mathcal{P}(L),\ Q \subset Q_1,$$

form a $(G, L)$-family as $Q_1$ varies over $\mathcal{P}(L)$. It follows from the definitions that (21.9) equals

$$\begin{aligned}&\lim_{\Lambda \to 0} \sum_{Q_1 \in \mathcal{P}(L)} c_{Q_1}(\Lambda, T) d_{Q_1}(\Lambda, \lambda) \theta_{Q_1}(\Lambda)^{-1} \\ &= \lim_{\Lambda \to 0} \sum_{Q_1 \in \mathcal{P}(L)} \mathrm{tr}\big(\mathcal{M}_{Q_1}^T(\Lambda, \lambda, P) M_P(s, 0) \mathcal{I}_{P,\chi,\pi}(\lambda, f)\big) \theta_{Q_1}(\Lambda)^{-1} \\ &= \lim_{\Lambda \to 0} \mathrm{tr}\big(\mathcal{M}_L^T(\Lambda, \lambda, P) M_P(s, 0) \mathcal{I}_{P,\chi,\pi}(\lambda, f)\big) \\ &= \mathrm{tr}\big(\mathcal{M}_L^T(\lambda, P) M_P(s, 0) \mathcal{I}_{P,\chi,\pi}(\lambda, f)\big).\end{aligned}$$

We substitute this formula into (21.8). The resulting expression is the required polynomial approximation to (21.4).

The following proposition is Theorem 4.1 of [**A8**]. We have completed a reasonably comprehensive sketch of its proof.

PROPOSITION 21.2. *For any $f \in \mathcal{H}(G)$ and $B \in C_c^\infty(i\mathfrak{h}^*/i\mathfrak{a}_G^*)^W$, the polynomial $P^T(B, f)$ equals the sum over $P \supset P_0$, $\pi \in \Pi_{\mathrm{unit}}\big(M_P(\mathbb{A})^1\big)$, $L \in \mathcal{L}(M_P)$, and $s \in W^L(M_P)_{\mathrm{reg}}$ of the product of*

$$n_P^{-1} |\det(s - 1)_{\mathfrak{a}_P^L}|^{-1}$$

*with*

$$\int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} \mathrm{tr}\big(\mathcal{M}_L^T(\lambda, P) M_P(s, 0) \mathcal{I}_{P,\chi,\pi}(\lambda, f)\big) B_\pi(\lambda) \mathrm{d}\lambda. \qquad \square$$

Recall that

$$J_\chi^T(f) = \lim_{\varepsilon \to 0} P^T(B^\varepsilon, f),$$

where $B^\varepsilon(\nu) = B(\varepsilon\nu)$, and $B(0)$ is assumed to be 1. Therefore

$$J_\chi(f) = J_\chi^{T_0}(f) = \lim_{\varepsilon \to 0} P^{T_0}(B^\varepsilon, f).$$

Now

$$\begin{aligned}\mathcal{M}_L^{T_0}(\lambda, P) &= \lim_{\Lambda\to 0} \sum_{Q_1\in\mathcal{P}(L)} c_{Q_1}(\Lambda, T_0)\mathcal{M}_{Q_1}(\Lambda,\lambda,P)\theta_{Q_1}(\Lambda)^{-1}\\ &= \lim_{\Lambda\to 0} \sum_{Q_1\in\mathcal{P}(L)} \mathrm{e}^{\Lambda(Y_{Q_1}(T_0))}\mathcal{M}_{Q_1}(\Lambda,\lambda,P)\theta_{Q_1}(\Lambda)^{-1}\\ &= \lim_{\Lambda\to 0} \mathrm{e}^{\Lambda(T_0)} \sum_{Q_1\in\mathcal{P}(L)} \mathcal{M}_{Q_1}(\Lambda,\lambda,P)\theta_{Q_1}(\Lambda)^{-1}\\ &= \mathcal{M}_L(\lambda,P),\end{aligned}$$

since $Y_{Q_1}(T_0)$ is just the projection of $T_0$ onto $\mathfrak{a}_L$. We substitute this into the formula above. The canonical point $T_0 \in \mathfrak{a}_0$ is independent of the minimal parabolic subgroup $P_0 \in \mathcal{P}(M_0)$ we fixed at the beginning of the section. Moreover, if $M = M_P$, the function

$$\operatorname{tr}\big(\mathcal{M}_L(\lambda,P)M_P(s,0)\mathcal{I}_{P,\chi,\pi}(\lambda,f)\big)$$

is easily seen to be independent of the choice of $P \in \mathcal{P}(M)$. We can therefore rewrite the formula of Proposition 21.2 in terms of Levi subgroups $M \in \mathcal{L}$ rather than standard parabolic subgroups $P \supset P_0$. Making the appropriate adjustments to the coefficients, one obtains the following formula as a corollary of the last one. (See [**A8**, Theorem 5.2].)

COROLLARY 21.3. *For any $f \in \mathcal{H}(G)$, the linear form $J_\chi(f)$ equals the limit as $\varepsilon$ approaches 0 of the expression obtained by taking the sum over $M \in \mathcal{L}$, $L \in \mathcal{L}(M)$, $\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$, and $s \in W^L(M)_{\mathrm{reg}}$ of the product of*

$$|W_0^M||W_0^G|^{-1}|\det(s-1)_{\mathfrak{a}_M^L}|^{-1}$$

*with*

$$\int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} \operatorname{tr}\big(\mathcal{M}_L(\lambda,P)M_P(s,0)\mathcal{I}_{P,\chi,\pi}(\lambda,f)\big)B_\pi^\varepsilon(\lambda)\mathrm{d}\lambda. \qquad \square$$

The final step is to get rid of the function $B_\pi^{\mathcal{E}}$ and the associated limit in $\varepsilon$. Recall that $B$ had the indispensable role of truncating the support of integrals that would otherwise be unmanageable. The function

$$B_\pi^\varepsilon(\lambda) = B\big(\varepsilon(iY_\pi+\lambda)\big)$$

is compactly supported in $\lambda \in i\mathfrak{a}_L^*/i\mathfrak{a}_G^*$, but converges pointwise to 1 as $\varepsilon$ approaches 0. If we can show that the integral

$$\int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} \operatorname{tr}\big(\mathcal{M}_L(\lambda,P)M_P(s,0)\mathcal{I}_{P,\chi,\pi}(\lambda,f)\big)\mathrm{d}\lambda \tag{21.10}$$

converges absolutely, we could remove the limit in $\varepsilon$ by an appeal to the dominated convergence theorem. One establishes absolute convergence by normalizing the intertwining operators from which the operator $\mathcal{M}_L(\lambda,P)$ is constructed.

Suppose that $\pi_v \in \Pi\big(M(F_v)\big)$ is an irreducible representation of $M(F_v)$, for a Levi subgroup $M \in \mathcal{L}$ and a valuation $v$ of $F$. We write

$$\pi_{v,\lambda}(m_v) = \pi_v(m_v)\mathrm{e}^{\lambda(H_M(m_v))}, \qquad m_v \in M(F_v),$$

as usual, for the twist of $\pi_v$ by an element $\lambda \in \mathfrak{a}_{M,\mathbb{C}}^*$. If $P \in \mathcal{P}(M)$, $\mathcal{I}_P(\pi_{v,\lambda})$ denotes the corresponding induced representation of $G(F_v)$, acting on a Hilbert space $\mathcal{H}_P(\pi_v)$ of vector valued functions on $K_v$. If $Q \in \mathcal{P}(M)$ is another parabolic subgroup, and $\phi$ belongs to $\mathcal{H}_P(\pi_v)$, the integral

$$\int_{N_Q(F_v)\cap N_P(F_v)\backslash N_Q(F_v)} \phi(n_v x_v)\mathrm{e}^{(\lambda+\rho_P)(H_P(n_v x_v))}\mathrm{e}^{-(\lambda+\rho_Q)(H_Q(x_v))}\mathrm{d}n_v$$

converges if the real part of $\lambda$ is highly regular in the chamber $(\mathfrak{a}_M^*)_P^+$. It defines an operator

$$J_{Q|P}(\pi_{v,\lambda}):\ \mathcal{H}_P(\pi_v)\ \longrightarrow\ \mathcal{H}_Q(\pi_v)$$

that intertwines the local induced representations $\mathcal{I}_P(\pi_{v,\lambda})$ and $\mathcal{I}_Q(\pi_{v,\lambda})$. One knows that $J_{Q|P}(\pi_{v,\lambda})$ can be analytically continued to a meromorphic function of $\lambda \in \mathfrak{a}_{M,\mathbb{C}}^*$ with values in the corresponding space of intertwining operators. (See [**Har5**], [**KnS**], and [**Sha1**].) This is a local analogue of Langlands' analytic continuation of the global operators $M_{Q|P}(\lambda)$. Unlike the operators $M_{Q|P}(\lambda)$, however, the local operators $J_{Q|P}(\pi_{v,\lambda})$ are not transitive in $Q$ and $P$. For example, if $\overline{P}$ is the group in $\mathcal{P}(M)$ opposite to $P$, Harish-Chandra has proved that

$$J_{P|\bar{P}}(\pi_{v,\lambda})J_{\bar{P}|P}(\pi_{v,\lambda}) = \mu_M(\pi_{v,\lambda})^{-1},$$

where $\mu_M(\pi_{v,\lambda})$ is a meromorphic scalar valued function that is closely related to the Plancherel density. To make the operators $J_{Q|P}(\pi_{v,\lambda})$ have better properties, one must multiply them by suitable scalar normalizing factors.

Theorem 21.4. *For any $M$, $v$, and $\pi_v \in \Pi\big(M(F_v)\big)$, one can choose meromorphic scalar valued functions*

$$r_{Q|P}(\pi_{v,\lambda}), \qquad\qquad \lambda \in \mathfrak{a}_{M,\mathbb{C}}^*,\ \ P, Q \in \mathcal{P}(M),$$

*such that the normalized intertwining operators*

$$R_{Q|P}(\pi_{v,\lambda}) = r_{Q|P}(\pi_{v,\lambda})^{-1}J_{Q|P}(\pi_{v,\lambda}) \tag{21.11}$$

*have the following properties.*

(i) $R_{Q'|P}(\pi_{v,\lambda}) = R_{Q'|Q}(\pi_{v,\lambda})R_{Q|P}(\pi_{v,\lambda}), \qquad Q', Q, P \in \mathcal{P}(M)$.

(ii) *The $K_v$-finite matrix coefficients of $R_{Q|P}(\pi_{v,\lambda})$ are rational functions of the variables $\{\lambda(\alpha^\vee):\ \alpha \in \Delta_P\}$ if $v$ is archimedean, and the variables $\{q_v^{-\lambda(\alpha^\vee)}:\ \alpha \in \Delta_P\}$ if $v$ is nonarchimedean.*

(iii) *If $\pi_v$ is unitary, the operator $R_{Q|P}(\pi_{v,\lambda})$ is unitary for $\lambda \in i\mathfrak{a}_M^*$, and hence analytic.*

(iv) *If $G$ is unramified at $v$, and $\phi \in \mathcal{H}(\pi_v)$ is the characteristic function of $K_v$, $R_{Q|P}(\pi_{v,\lambda})\phi$ equals $\phi$.*

See [**A15**, Theorem 2.1] and [**CLL**, Lecture 15]. The factors $r_{Q|P}(\pi_{v,\lambda})$ are defined as products, over reduced roots $\beta$ of $(Q, A_M)$ that are not roots of $(P, A_M)$, of meromorphic functions $r_\beta(\pi_\lambda)$ that depend only on $\lambda(\beta^\vee)$. The main step is to establish the property

$$r_{P|\bar{P}}(\pi_{v,\lambda})r_{\bar{P}|P}(\pi_{v,\lambda}) = \mu_M(\pi_{v,\lambda})^{-1}, \tag{21.12}$$

in the case that $M$ is maximal. □

**Remarks.** 1. The assertions of the theorem are purely local. They can be formulated for Levi subgroups and parabolic subgroups that are defined over $F_v$.

2. Suppose that $\bigotimes_v \pi_v$ is an irreducible representation of $M(\mathbb{A})$, whose restriction to $M(\mathbb{A})^1$ we denote by $\pi$. The product

$$R_{Q|P}(\pi_\lambda) = \bigotimes_v R_{Q|P}(\pi_{v,\lambda}) \tag{21.13}$$

is then a well defined transformation of the dense subspace $\mathcal{H}_P^0(\pi)$ of $K$-finite vectors in $\mathcal{H}_P(\pi)$. Indeed, for any $\phi \in \mathcal{H}_P^0(\pi)$, $R_{Q|P}(\pi_\lambda)\phi$ can be expressed as a finite product by (iv). If $\pi$ is unitary and $\lambda \in i\mathfrak{a}_M^*$, $R_{Q|P}(\pi_\lambda)$ extends to a unitary transformation of the entire Hilbert space $\mathcal{H}_P(\pi)$.

Suppose that $\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$ is any representation that occurs in the discrete part $R_{M,\mathrm{disc}}$ of $R_M$. In other words, the subspace $\mathcal{H}_{P,\pi}$ of $\mathcal{H}_P$ is nonzero. The restriction of the global intertwining operator $M_{Q|P}(\lambda)$ to $\mathcal{H}_{P,\pi}$ can be expressed in terms of the local intertwining operators above. It is isomorphic to $m_{\mathrm{disc}}(\pi)$-copies of the operator

$$J_{Q|P}(\pi_\lambda) = \bigotimes_v J_{Q|P}(\pi_{v,\lambda}),$$

defined for any unitary extension $\bigotimes_v \pi_v$ of $\pi$ to $M(\mathbb{A})$ by analytic continuation in $\lambda$. If $\{r_{Q|P}(\pi_{v,\lambda})\}$ is any family of local normalizing factors that for each $v$ satisfy the conditions of Theorem 21.4, the scalar-valued product

$$r_{Q|P}(\pi_\lambda) = \prod_v r_{Q|P}(\pi_{v,\lambda}) \tag{21.14}$$

is also defined by analytic continuation in $\lambda$, and is analytic for $\lambda \in i\mathfrak{a}_M^*$. Let $R_{Q|P}(\lambda)$ be the operator on $\mathcal{H}_P$ whose restriction to any subspace $\mathcal{H}_{P,\pi}$ equals the product of $r_{Q|P}(\pi_\lambda)^{-1}$ with the restriction of $M_{Q|P}(\lambda)$. In other words, the restriction of $R_{Q|P}(\lambda)$ to $\mathcal{H}_{P,\pi}$ is isomorphic to $m_{\mathrm{disc}}(\pi)$-copies of the operator (2.13). We define

$$r_Q(\Lambda, \pi_\lambda, P) = r_{Q|P}(\pi_\lambda)^{-1} r_{Q|P}(\pi_{\lambda+\Lambda}), \qquad \mathcal{H}_{P,\pi} \neq \{0\}, \tag{21.15}$$

and

$$\mathcal{R}_Q(\Lambda, \lambda, P) = R_{Q|P}(\lambda)^{-1} R_{Q|P}(\lambda + \Lambda),$$

for points $\Lambda$ and $\lambda$ in $i\mathfrak{a}_M^*$. Then $\{r_Q(\Lambda, \pi_\lambda)\}$ and $\{\mathcal{R}_Q(\Lambda, \lambda, P)\}$ are new $(G, M)$-families of $\Lambda$. They give rise to functions $r_L(\pi_\lambda, P)$ and $\mathcal{R}_L(\lambda, P)$ of $\lambda$ for any $L \in \mathcal{L}(M)$. We write $r_L(\pi_\lambda) = r_L(\pi_\lambda, P)$, since this function is easily seen to be independent of the choice of $P$.

Lemma 21.5. (a) *There is a positive integer $n$ such that*

$$\int_{i\mathfrak{a}_M^*/i\mathfrak{a}_G^*} |r_L(\pi_\lambda)|(1 + \|\lambda\|)^{-n} d\lambda < \infty.$$

(b) *The integral (21.10) converges absolutely.*

The integrand in (21.10) depends only on the restriction $\mathcal{M}_L(\lambda, P)_{\chi,\pi}$ of the operator $\mathcal{M}_L(\lambda, P)$ to $\mathcal{H}_{P,\chi,\pi}$. But $\mathcal{M}_L(\lambda, P)_{\chi,\pi}$ can be defined in terms of the product of the two new $(G, M)$-families above. Moreover, we are free to apply the simpler version (17.12) of the usual splitting formula. This is because for any $S \in \mathcal{F}(L)$ and $Q \in \mathcal{P}(S)$, the number

$$r_L^S(\pi_\lambda) = r_L^Q(\pi_\lambda)$$

is independent of the choice of $Q$ [**A8**, Corollary 7.4]. Therefore

$$\mathcal{M}_L(\lambda, P)_{\chi,\pi} = \sum_{S\in\mathcal{F}(L)} r_L^S(\pi_\lambda)\mathcal{R}_S(\lambda, P)_{\chi,\pi},$$

where $\mathcal{R}_S(\lambda, P)_{\chi,\pi}$ denotes the restriction of $\mathcal{R}_S(\lambda, P)$ to $\mathcal{H}_{P,\chi,\pi}$. The integral (21.10) can therefore be decomposed as a sum

$$\sum_{S\in\mathcal{F}(L)} \int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} r_L^S(\pi_\lambda)\mathrm{tr}\big(\mathcal{R}_S(\lambda, P)M_P(s, 0)\mathcal{I}_{P,\chi,\pi}(\lambda, f)\big)\mathrm{d}\lambda. \tag{21.16}$$

Since $f$ lies in the Hecke algebra $\mathcal{H}(G)$, the operator $\mathcal{I}_{P,\chi,\pi}(\lambda, f)$ is supported on a finite dimensional subspace of $\mathcal{H}_{P,\chi,\pi}$. Moreover, it is an easy consequence of the conditions (ii)–(iv) of Theorem 21.4 that any matrix coefficient of the operator $\mathcal{R}_S(\lambda, P)$ is a rational function in finitely many complex variables $\{\lambda(\alpha^\vee), q_v^{-\lambda(\alpha^\vee)}\}$, which is analytic for $\lambda \in i\mathfrak{a}_M^*$. Since $\mathcal{I}_{P,\chi,\pi}(\lambda, f)$ is rapidly decreasing in $\lambda$, part (b) of the lemma follows inductively from (a). (See [**A8**, §8].)

It is enough to establish part (a) in the case that $M$ is a maximal Levi subgroup. This is because for general $M$ and $L$, $r_M^L(\pi_\lambda)$ can be written as a finite linear combination of products

$$r_M^{M_1}(\pi_\lambda)\dots r_M^{M_p}(\pi_\lambda),$$

for Levi subgroups $M_1, \dots, M_p$ in $\mathcal{L}(M)$, with $\dim(\mathfrak{a}_M/\mathfrak{a}_{M_i}) = 1$, such that the mapping

$$\mathfrak{a}_M/\mathfrak{a}_G \longrightarrow \bigoplus_{i=1}^{p}(\mathfrak{a}_M/\mathfrak{a}_{M_i})$$

is an isomorphism. (See [**A8**, §7].) In case $M$ is maximal, one combines (21.16) with estimates based on Selberg's positivity argument used to prove Theorem 14.1(a). (See [**A8**, §8–9].) □

It is a consequence of Langlands' construction of the discrete spectrum of $M$ in terms of residues of cuspidal Eisenstein series that the sum over $\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$ in Corollary 21.3 can be taken over a finite set. Lemma 21.5(b) asserts that for any $\pi$, the integral (21.10) converges absolutely. Combining the dominated convergence theorem with the formula of Corollary 21.3, we obtain the following theorem.

THEOREM 21.6. *For any $f \in \mathcal{H}(G)$, the linear form $J_\chi(f)$ equals the sum over $M \in \mathcal{L}$, $L \in \mathcal{L}(M)$, $\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$, and $s \in W^L(M)_{\mathrm{reg}}$ of the product of*

$$|W_0^M||W_0^G|^{-1}|\det(s-1)_{\mathfrak{a}_M^G}|^{-1} \tag{21.17}$$

*with*

$$\int_{i\mathfrak{a}_L^*/i\mathfrak{a}_G^*} \mathrm{tr}\big(\mathcal{M}_L(\lambda, P)M_P(s, 0)\mathcal{I}_{P,\chi,\pi}(\lambda, f)\big)\mathrm{d}\lambda.$$

(See [**A8**, Theorem 8.2].) □

**Remarks.** 3. There is an error in [**A8**, §8]. It is the ill-considered inequality stated on p. 1329 of [**A8**], three lines above the expression (8.4), which was taken from [**A5**, (7.6)]. The inequality seems to be false if $f$ lies in the complement of $\mathcal{H}(G)$ in $C_c^\infty\big(G(\mathbb{A})^1\big)$, and $\pi$ is nontempered. Consequently, the last formula for $J_\chi(f)$ does not hold if $f$ lies in the complement of $\mathcal{H}(G)$ in $C_c^\infty\big(G(\mathbb{A})^1\big)$.

The fine spectral expansion of $J(f)$ is the sum over $\chi \in \mathcal{X}$ of the formulas for $J_\chi(f)$ provided by the last theorem. It is convenient to express this expansion in terms of infinitesimal characters.

A representation $\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$ has an archimedean infinitesimal character, consisting of a $W$-orbit of points $\nu_\pi = X_\pi + iY_\pi$ in $\mathfrak{h}^*_{\mathbb{C}}/i\mathfrak{a}^*_G$. The imaginary part $Y_\pi$ is really an $\mathfrak{a}^*_M$-coset in $\mathfrak{h}^*$, but as in §20, we can identify it with the unique point in the coset for which the norm $\|Y_\pi\|$ is minimal. We then define

$$\Pi_{t,\mathrm{unit}}\big(M(\mathbb{A})^1\big) = \big\{\pi \in \Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big) : \ \|\mathrm{Im}(\nu_\pi)\| = \|Y_\pi\| = t\big\},$$

for any nonnegative real number $t$.

Recall that a class $\chi \in \mathcal{X}$ is a $W_0$-orbit of pairs $(M_1, \pi_1)$, with $\pi_1$ being a cuspidal automorphic representation of $M_1(\mathbb{A})^1$. Setting $\nu_\chi = \nu_{\pi_1}$, we define a linear form

$$J_t(f) = \sum_{\{\chi \in \mathcal{X} : \|\mathrm{Im}(\nu_\chi)\| = t\}} J_\chi(f), \qquad t \geq 0,\ f \in \mathcal{H}(G),$$

in which the sum may be taken over a finite set. Then

$$J(f) = \sum_{t \geq 0} J_t(f).$$

We also write $\mathcal{I}_{P,t}(\lambda, f)$ for the restriction of the operator $\mathcal{I}_P(\lambda, f)$ to the invariant subspace

$$\mathcal{H}_{P,t} = \bigoplus_{\{(\chi,\pi) : \|\mathrm{Im}(\nu_\chi)\| = t\}} \mathcal{H}_{P,\chi,\pi},$$

of $\mathcal{H}_P$. It is again a consequence of Langlands' construction of the discrete spectrum that if $\|\mathrm{Im}(\nu_\chi)\| = t$, the space $\mathcal{H}_{P,\chi,\pi}$ vanishes unless $\pi$ belongs to $\Pi_{t,\mathrm{unit}}\big(M(\mathbb{A})^1\big)$. In other words, the representation $\mathcal{I}_{P,t}(\lambda)$ is equivalent to a direct sum of induced representations of the form $\mathcal{I}_P(\pi_\lambda)$, for $\pi \in \Pi_{t,\mathrm{unit}}\big(M(\mathbb{A})^1\big)$. The fine spectral expansion is then given by the following corollary of Theorem 21.6.

Corollary 21.7. *For any $f \in \mathcal{H}(G)$, the linear form $J(f)$ equals the sum over $t \geq 0$, $M \in \mathcal{L}$, $L \in \mathcal{L}(M)$, and $s \in W^L(M)_{\mathrm{reg}}$ of the product of the coefficient (21.17) with the linear form*

$$\int_{i\mathfrak{a}^*_L/i\mathfrak{a}^*_G} \mathrm{tr}\big(\mathcal{M}_L(\lambda, P) M_P(s,0) \mathcal{I}_{P,t}(\lambda, f)\big) d\lambda. \tag{21.18}$$

□

The fine spectral expansion is thus an explicit sum of integrals. Among these integrals, the ones that are discrete have special significance. They correspond to the terms with $L = G$. The discrete part of the fine spectral expansion attached to any $t$ equals the linear form

(21.19)
$$I_{t,\mathrm{disc}}(f) = \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{s \in W(M)_{\mathrm{reg}}} |\det(s-1)_{\mathfrak{a}^G_M}|^{-1} \mathrm{tr}\big(M_P(s,0)\mathcal{I}_{P,t}(0,f)\big).$$

It contains the $t$-part of the discrete spectrum, as well as singular points in the $t$-parts of continuous spectra. Observe that we have not shown that the sum over $t$ of these distributions converges. To do so, one would need to extend Müller's solution of the trace class conjecture [**Mul**], as has been done in the case $G = GL(n)$ by Müller and Speh [**MS**]. It is only after $I_{t,\mathrm{disc}}(f)$ has been enlarged to the linear form $J_t(f)$, by including the corresponding continuous terms, that the spectral arguments we have discussed yield the absolute convergence of the sum over $t$. However, it turns out that this circumstance does not effect our ability to use trace formulas to compare discrete spectra on different groups.

## 22. The problem of invariance

In the last four sections, we have refined both the geometric and spectral sides of the original formula (16.1). Let us now step back for a moment to assess the present state of affairs. The fine geometric expansion of Corollary 19.3 is transparent in its overall structure. It is a simple linear combination of weighted orbital integrals, taken over Levi subgroups $M \in \mathcal{L}$. The fine spectral expansion of Corollary 21.7 is also quite explicit, but it contains a more complicated double sum over Levi subgroups $M \subset L$. In order to focus our discussion on the next stage of development, we need to rewrite the spectral side so that it is parallel to the geometric side.

We shall first revisit the fine geometric expansion. This expansion is a sum of products of local distributions $J_M(\gamma, f)$ with global coefficients $a^M(S,\gamma)$, where $S \supset S_{\mathrm{ram}}$ is a large finite set of valuations depending on the support of $f$, and $\gamma \in \big(M(F)\big)_{M,S}$ is an $(M,S)$-equivalence class. Let us write

$$\Gamma(M)_S = \big(M(F)\big)_{M,S}, \qquad S \supset S_{\mathrm{ram}}, \tag{22.1}$$

in order to emphasize that this set is a quotient of the set $\Gamma(M)$ of conjugacy classes in $M(F)$. The semi-simple component $\gamma_s$ of a class $\gamma \in \Gamma(M)_S$ can be identified with a semisimple conjugacy class in $M(F)$. By choosing $S$ to be large, we guarantee that for any class $\gamma$ with $J_M(\gamma, f) \neq 0$, the set

$$\mathrm{Int}\big(M(\mathbb{A}^S)\big)\gamma_s = \{m^{-1}\gamma_s m : \ m \in M(\mathbb{A}^S)\}$$

intersects the maximal compact subgroup $K_M^S$ of $M(\mathbb{A}^S)$. If $S$ is *any* finite set containing $S_{\mathrm{ram}}$, and $\gamma$ is a class in $\Gamma(M)_S$, we shall write

$$a^M(\gamma) = \begin{cases} a^M(S,\gamma), & \text{if } \mathrm{Int}\big(M(\mathbb{A}^S)\big)\gamma_s \cap K_M^S \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{22.2}$$

If $f$ belongs to $\mathcal{H}(G) = \mathcal{H}\big(G(\mathbb{A})^1\big)$, we also write

$$J_M(\gamma, f) = J_M(\gamma, f_S),$$

where $f_S$ is the restriction of $f$ to the subgroup $G(F_S)^1$ of $G(\mathbb{A})^1$. We can then write the fine geometric expansion slightly more elegantly as the limit over increasing sets $S$ of expressions

$$\sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) I_M(\gamma, f).$$

The limit stabilizes for large finite sets $S$.

To write the spectral expansion in parallel form, we have first to introduce suitable weighted characters $J_M(\pi, f)$. Suppose that $\pi$ belongs to $\Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$. Then $\pi$ can be identified with an orbit $\pi_\lambda$ of $i\mathfrak{a}_M^*$ in $\Pi_{\mathrm{unit}}\big(M(\mathbb{A})\big)$. In the last section, we defined normalized intertwining operators $R_{Q|P}(\pi_\lambda)$ in terms (21.11) and (21.13) of a suitable choice of local normalizing factors $\{r_{Q|P}(\pi_{v,\lambda})\}$. We now introduce the corresponding $(G, M)$-family

$$\mathcal{R}_Q(\Lambda, \pi_\lambda, P) = R_{Q|P}(\pi_\lambda)^{-1} R_{Q|P}(\pi_{\lambda+\Lambda}), \qquad Q \in \mathcal{P}(M),\ \Lambda \in i\mathfrak{a}_M^*,$$

of operators on $\mathcal{H}_P(\pi)$, which we use to define the linear form

$$J_M(\pi_\lambda, \widetilde{f}) = \operatorname{tr}\big(\mathcal{R}_M(\pi_\lambda, P)\mathcal{I}_P(\pi_\lambda, \widetilde{f})\big) \qquad \widetilde{f} \in \mathcal{H}\big(G(\mathbb{A})\big), \tag{22.3}$$

on $\mathcal{H}\big(G(\mathbb{A})\big)$. We then set

$$J_M(\pi, f) = \int_{i\mathfrak{a}_M^*} J_M(\pi_\lambda, \widetilde{f})\mathrm{d}\lambda, \qquad f \in \mathcal{H}(G), \tag{22.4}$$

where $\widetilde{f}$ is any function in $\mathcal{H}\big(G(\mathbb{A})\big)$ whose restriction to $G(\mathbb{A})^1$ equals $f$. The last linear form does indeed depend only on $\pi$ and $f$. It is the required weighted character.

The core of the fine spectral expansion is the $t$-discrete part $I_{t,\mathrm{disc}}(f)$, defined for any $t \geq 0$ and $f \in \mathcal{H}(G)$ by (21.19). The term "discrete" refers obviously to the fact that we can write the distribution as a linear combination

$$I_{t,\mathrm{disc}}(f) = \sum_{\pi \in \Pi_{t,\mathrm{unit}}(G(\mathbb{A})^1)} a_{\mathrm{disc}}^G(\pi) f_G(\pi) \tag{22.5}$$

of irreducible characters, with complex coefficients $a_{\mathrm{disc}}^G(\pi)$. It is a consequence of Langlands' construction of the discrete spectrum that for any $f$, the sum may be taken over a finite set. (See [**A14**, Lemmas 4.1 and 4.2].) Let $\Pi_{t,\mathrm{disc}}(G)$ be the subset of irreducible constituents of induced representations

$$\sigma_\lambda^G, \qquad M \in \mathcal{L},\ \sigma \in \Pi_{t,\mathrm{unit}}\big(M(\mathbb{A})^1\big),\ \lambda \in i\mathfrak{a}_M^*/i\mathfrak{a}_G^*,$$

of $G(\mathbb{A})^1$, where the representation $\sigma_\lambda$ of $M(\mathbb{A}) \cap G(\mathbb{A})^1$ satisfies the two conditions.

(i) $a_{\mathrm{disc}}^M(\sigma) \neq 0$.

(ii) there is an element $s \in W^G(\mathfrak{a}_M)_{\mathrm{reg}}$ such that $s\sigma_\lambda = \sigma_\lambda$.

As a discrete subset of $\Pi_{t,\mathrm{unit}}\big(G(\mathbb{A})^1\big)$, $\Pi_{t,\mathrm{disc}}(G)$ is a convenient domain for the coefficients $a_{\mathrm{disc}}^G(\pi)$.

It is also useful to introduce a manageable domain of induced representations in $\Pi_{t,\mathrm{unit}}\big(G(\mathbb{A})^1\big)$. We define a set

$$\Pi_t(G) = \{\pi_\lambda^G :\ M \in \mathcal{L},\ \pi \in \Pi_{t,\mathrm{disc}}(M),\ \lambda \in i\mathfrak{a}_M^*/i\mathfrak{a}_G^*\}, \tag{22.6}$$

equipped with the measure $d\pi_\lambda^G$ for which

$$\int_{\Pi_t(G)} \phi(\pi_\lambda^G)\mathrm{d}\pi_\lambda^G = \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\pi \in \Pi_{t,\mathrm{disc}(M)}} \int_{i\mathfrak{a}_M^*/i\mathfrak{a}_G^*} \phi(\pi_\lambda^G)\mathrm{d}\lambda, \tag{22.7}$$

for any reasonable function $\phi$ on $\Pi_t(G)$. If $\pi$ belongs to a set $\Pi_{t,\mathrm{disc}}(M)$, the global normalizing factors $r_{Q|P}(\pi_\lambda)$ can be defined by analytic continuation of a product (21.14). We can therefore form the $(G, M)$-family $\{r_Q(\Lambda, \pi_\lambda)\}$ as in (21.15). The associated function $r_M(\pi_\lambda) = r_M^G(\pi_\lambda)$ is analytic in $\lambda$, and satisfies the estimate of Lemma 21.5(a). We define a coefficient function on $\Pi_t(G)$ by setting

$$a^G(\pi_\lambda^G) = a_{\mathrm{disc}}^M(\pi) r_M^G(\pi_\lambda), \qquad M \in \mathcal{L},\ \pi \in \Pi_{t,\mathrm{disc}}(M),\ \lambda \in i\mathfrak{a}_M^*/i\mathfrak{a}_G^*. \tag{22.8}$$

It is not hard to show that the right hand side of this expression depends only on the induced representation $\pi_\lambda^G$, at least on the complement of a set of measure 0 in $\Pi_t(G)$.

For any $M \in \mathcal{L}$, we write $\Pi(M)$ for the union over $t \geq 0$ of the sets $\Pi_t(M)$. The analogues of (22.7) and (22.8) for $M$ provide a measure $d\pi$ and a function $a^M(\pi)$ on $\Pi(M)$. Since we have now terminated our relationship with the earlier parameter of truncation, we allow ourselves henceforth to let $T$ stand for a positive real number. With this notation, we write $\Pi(M)^T$ for the union over $t \leq T$ of the sets $\Pi_t(M)$. The refined spectral expansion then takes the form of a limit, as $T$ approaches infinity, of a sum of integrals over the sets $\Pi(M)^T$.

We can now formulate the refined trace formula as an identity between two parallel expansions. We state it as a corollary of the results at the end of §19 and §21.

Corollary 22.1. *For any $f \in \mathcal{H}(G)$, $J(f)$ has a geometric expansion*

$$J(f) = \lim_S \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) J_M(\gamma, f) \tag{22.9}$$

*and a spectral expansion*

$$J(f) = \lim_T \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Pi(M)^T} a^M(\pi) J_M(\pi, f) \mathrm{d}\pi. \tag{22.10}$$

The geometric expansion (22.9) is essentially that of Corollary 19.3, as we noted above. The spectral expansion (22.10) is a straightforward reformulation of the expansion of Corollary 21.7, which is established in the first part of the proof of Theorem 4.4 of [**A14**]. One applies the appropriate analogue of the splitting formula (21.16) to the integral (21.18). This gives an expansion of $J_t(f)$ as a triple sum over Levi subgroups $M \subset L \subset S$ and a simple sum over $s \in W^L(\mathfrak{a}_M)_{\mathrm{reg}}$. One then observes that the sum over $M$ gives rise to a form of the distribution $I^L_{t,\mathrm{disc}}$, for which one can substitute the analogue of (22.5). Having removed the original sum over $M$, we are free to write $M$ in place of the index $S$. The expression (21.18) becomes a sum over $M \in \mathcal{L}(L)$ and an integral over $\lambda \in i\mathfrak{a}_L^*/i\mathfrak{a}_G^*$. The last step is to rewrite the integral as a double integral over the product of $i\mathfrak{a}_L^*/i\mathfrak{a}_M^*$ with $i\mathfrak{a}_M^*/i\mathfrak{a}_G^*$. The spectral expansion (22.10) then follows from the definitions of the linear forms $J_M(\pi, f)$, the coefficients $a^M(\pi)$, and the measure $d\pi$. □

Although the refined trace formula of Corollary 22.1 is a considerable improvement over its predecessor (16.1), it still has defects. There are of course the questions inherent in the two limits. These difficulties were mentioned briefly in §19 (in the remark following Theorem 19.1) and in §21 (at the end of the section). The spectral problem has been solved for $GL(n)$, while the geometric problem is open for any group other than $GL(2)$. Both problems will be relevant to any attempt to exploit the trace formula of $G$ in isolation. However, they seem to have no bearing on our ability to compare trace formulas on different groups. We shall not discuss them further.

Our concern here is with the failure of the linear forms $J_M(\gamma, f)$ and $J_M(\pi, f)$ to be invariant. There is also the disconcerting fact that they depend on a non-canonical choice of maximal compact subgroup $K$ of $G(\mathbb{A})$. Of course, the domain $\mathcal{H}(G)$ of the linear forms already depends on $K$, through its archimedean component $K_\infty$. However, even when we can extend the linear forms to the larger domain $C_c^\infty\big(G(\mathbb{A})^1\big)$, which we can invariably do in the geometric case, they are still fundamentally dependent on $K$.

To see why the lack of invariance is a concern, we recall the Jacquet-Langlands correspondence described in §3. Their mapping $\pi \to \pi^*$ of automorphic representations was governed by a correspondence $f \to f^*$ from functions $f$ on the multiplicative group $G(\mathbb{A})$ of an adelic quaternion algebra, and functions $f^*$ on the adelic group $G^*(\mathbb{A})$ attached to $G^* = GL(2)$. The correspondence of functions was defined by identifying invariant orbital integrals. It is expected that for any $G$, the set of strongly regular invariant orbital integrals spans a dense subspace of the entire space of invariant distributions. (The same is expected of the set of irreducible tempered characters.) We might therefore be able to transfer invariant distributions between suitably related groups. However, we cannot expect to be able to transfer distributions that are not invariant.

The problem is to transform the identity between the expansions (22.9) and (22.10) into a more canonical formula, whose terms are invariant distributions. How can we do this? The first thing to observe is that the weighted orbital integrals in (22.9) and the weighted characters in (22.10) fail to be invariant in a similar way. By the construction of §18, the weighted orbital integrals satisfy the relation

$$J_M(\gamma, f^y) = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\gamma, f_{Q,y}),$$

for any $f \in C_c^\infty\big(G(\mathbb{A})^1\big)$, $\gamma \in \Gamma(M)_S$, and $y \in G(\mathbb{A})$. A minor technical lacuna arises here when we restrict $f$ to the domain $\mathcal{H}(G)$ of the weighted characters, since the transformation $f \to f^y$ does not send $\mathcal{H}(G)$ to itself. However, the convolutions $L_h f = h * f$ and $R_h f = f * h$ of $f$ by a fixed function $h \in \mathcal{H}(G)$ do preserve $\mathcal{H}(G)$. We define a linear form on $\mathcal{H}(G)$ to be invariant if for any such $h$ it assumes the same values at $L_h f$ and $R_h f$. The relation above is equivalent to a formula

$$J_M(\gamma, L_h f) = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\gamma, R_{Q,h} f), \tag{22.11}$$

where

$$R_{Q,h} f = \int_{G(\mathbb{A})^1} h(y)(R_{y^{-1}} f)_{Q,h} \mathrm{d}y$$

and $(R_{y^{-1}} f)(x) = f(xy)$, which applies equally well to functions $f$ in either $C_c^\infty\big(G(\mathbb{A})^1\big)$ or $\mathcal{H}(G)$. It is no surprise to discover that the weighted characters satisfy a similar formula, since we know that the original distributions $J_{\mathfrak{o}}(f)$ and $J_\chi(f)$ satisfy the parallel variance formulas (16.2) and (16.3). It follows from Lemma 6.2 of [**A15**] that

$$J_M(\pi, L_h f) = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\pi, R_{Q,h} f), \tag{22.12}$$

for any $f \in \mathcal{H}(G)$, $\pi \in \Pi(M)$ and $h \in \mathcal{H}(G)$.

We have just seen that the two families of linear forms in the trace formula satisfy parallel variance formulas. It seems entirely plausible that we could construct an invariant distribution by taking a typical noninvariant distribution from one of the two families, and subtracting from it some combination of noninvariant distributions from the other family. Two questions arise. What would be the precise mechanics of the process? At a more philosophical level, should we subtract some combination of weighted characters from a given weighted orbital integral, or should we start with a weighted character and subtract from it some combination of weighted orbital integrals? We shall discuss the second question in the rest of this section, leaving the first question for the beginning of the next section.

Consider the example that $G = GL(2)$, and $M$ is the minimal Levi subgroup $GL(1) \times GL(1)$ of the diagonal matrices. Suppose that $f \in \mathcal{H}(G)$, and that $S$ is a large finite set of valuations. We can then identify $f$ with a function on $\mathcal{H}\big(G(F_S)^1\big)$. The weighted orbital integral $\gamma \to J_M(\gamma, f)$ is a compactly supported, locally integrable function on the group

$$M(F_S)^1 = \{(a,b) \in F_S \times F_S : \ |a| = |b| = 1\}.$$

The weighted character $\pi \to J_M(\pi, f)$ is a Schwartz function on the group $\Pi_{\mathrm{unit}}\big(M(F_S)^1\big)$ of unitary characters on $M(F_S)^1$. We could form the distribution

$$J_M(\gamma, f) - \int_{\Pi_{\mathrm{unit}}(M(F_S)^1)} \pi(\gamma^{-1}) J_M(\pi, f) \mathrm{d}\pi, \qquad \gamma \in M(F_S)^1, \tag{22.13}$$

by modifying the weighted orbital integral. We could also form the distribution

$$J_M(\pi, f) - \int_{M(F_S)^1} \pi(\gamma) J_M(\gamma, f) \mathrm{d}\gamma, \qquad \pi \in \Pi_{\mathrm{unit}}\big(M(F_S)^1\big), \tag{22.14}$$

by modifying the weighted character. According to the variance formulas above, each of these distributions is invariant. Which one should we take?

The terms in the trace formula for $G = GL(2)$ that are not invariant are the ones attached to our minimal Levi subgroup $M$. They can be written as

$$\frac{1}{2}\mathrm{vol}\big(M(\mathfrak{o}_S)\backslash M(F_S)^1\big) \sum_{\gamma \in M(\mathfrak{o}_S)} J_M(\gamma, f)$$

and

$$\frac{1}{2} \sum_{\pi \in \Pi(M(\mathfrak{o}_S)\backslash M(F_S)^1)} J_M(\pi, f)$$

respectively, for the discrete, cocompact subring

$$\mathfrak{o}_S = \{\gamma \in F : \ |\gamma|_v \leq 1, \ v \notin S\}$$

of $F_S$. Can we apply the Poisson summation formula to either of these expressions? Such an application to the first expression would yield an invariant trace formula for $GL(2)$ with terms of the form (22.14). An application of Poisson summation to the second expression would yield an invariant trace formula with terms of the form (22.13).

We need to be careful. Continuing with the example $G = GL(2)$, suppose that $\widetilde{f}$ lies in the Hecke algebra $\mathcal{H}\big(G(F_S)\big)$ on $G(F_S)$, and consider $J_M(\gamma, \widetilde{f})$ and $J_M(\pi, \widetilde{f})$ as functions on the larger groups $M(F_S)$ and $\Pi_{\mathrm{unit}}\big(M(F_S)\big)$ respectively. The function $J_M(\gamma, \widetilde{f})$ is still compactly supported. However, it has singularities at points $\gamma$ whose eigenvalues at some place $v \in S$ are equal. Indeed, in the example $v = \mathbb{R}$ examined in §18, we saw that the weighted orbital integral had a logarithmic singularity. If the logarithmic term is removed, the resulting function of $\gamma$ is bounded, but it still fails to be smooth. Langlands showed that the function was nevertheless well enough behaved to be able to apply the Poisson summation formula. He made the trace formula for $GL(2)$ invariant in this way, using the distributions (22.14) in his proof of base change for $GL(2)$ [**Lan9**]. A particular advantage of this approach is a formulation of the contribution of weighted orbital integrals in terms of a continuous spectral variable, which can be separated from the discrete spectrum. For groups of higher rank, however, the singularities of weighted orbital integrals seem to be quite unmanageable.

The other function $J_M(\pi, \widetilde{f})$ belongs to the Schwartz space on $\Pi_{\mathrm{unit}}\bigl(M(F_S)\bigr)$, but it need not lie in the Paley-Wiener space. This is because the operator-valued weight factor

$$\mathcal{R}_M(\pi, P), \qquad \pi \in \Pi_{\mathrm{unit}}\bigl(M(F_S)\bigr),$$

is a rational function in the continuous parameters of $\pi$, which acquires poles in the complex domain $\Pi\bigl(M(F_S)\bigr)$. Therefore $J_M(\pi, \widetilde{f})$ is not the Fourier transform of a compactly supported function on $M(F_S)$. This again does not preclude applying Poisson summation in the case under consideration. However, it does not seem to bode well for higher rank.

What is one to do? I would argue that it is more natural in general to work with the geometric invariant distributions (22.13) than with their spectral counterparts (22.14). Weighted characters satisfy splitting formulas analogous to (18.7). In the example under consideration, the formula is

$$J_M(\pi, \widetilde{f}) = \sum_{v \in S} \bigl(J_M(\pi_v, f_v) \cdot \prod_{w \neq v} f_{w,G}(\pi_w)\bigr),$$

where $\pi = \bigotimes\limits_{v \in S} \pi_v$ and $\widetilde{f} = \prod\limits_{v \in S} f_v$, and $J_M(\pi_v, f_v)$ is the local weighted character defined by the obvious analogue of (22.3). It follows from this that the Fourier transform

$$J_M^\wedge(\gamma, \widetilde{f}) = \int_{\Pi_{\mathrm{unit}}(M(F_S))} \pi^\vee(\gamma) J_M(\pi, \widetilde{f}) \mathrm{d}\pi, \qquad \gamma \in M(F_S),$$

of $J_M(\pi, \widetilde{f})$ is equal to a sum of products

$$J_M^\wedge(\gamma, \widetilde{f}) = \sum_{v \in S} \Bigl(J_M^\wedge(\gamma_v, f_v) \cdot \prod_{w \neq v} f_{w,G}(\gamma_w)\Bigr),$$

for $\gamma = \prod\limits_{v \in S} \gamma_v$. The invariant orbital integrals $f_{w,G}(\gamma_w)$ are all compactly supported, even though the functions $J_M^\wedge(\gamma_v, f_v)$ are not. Remember that we are supposed to take the Poisson summation formula for the diagonal subgroup

$$M(F_S)^1 = \Bigl\{\gamma \in M(F_S) : \ H_M(\gamma) = \sum_{v \in S} H_M(\gamma_v) = 0\Bigr\}$$

of $M(F_S)$. The intersection of this subgroup with any set that is a product of a noncompact subset of $M(F_v)$ with compact subsets of each of the complementary groups $M(F_w)$ is compact. It follows that if $f$ belongs to $\mathcal{H}\bigl(G(F_S)^1\bigr)$, the weighted character

$$J_M(\pi, f) = \int_{i\mathfrak{a}_M^*} J_M(\pi_\lambda, \widetilde{f}) \mathrm{d}\lambda, \qquad \pi \in \Pi_{\mathrm{unit}}\bigl(M(F_S)^1\bigr),$$

that actually occurs in the trace formula belongs to the Paley-Wiener space on $\Pi_{\mathrm{unit}}\bigl(M(F_S)^1\bigr)$ after all.

Suppose now that $G$ is arbitrary. It turns out that the phenomenon we have just described for $GL(2)$ holds in general. The underlying reason again is the fact that the weighted characters occur on the spectral side in the form of integrals (22.4), rather than as a discrete sum of linear forms (22.3). Otherwise said, the fine spectral expansion of Corollary 21.7 is composed of continuous integrals (21.18), while the fine geometric expansion of Corollary 19.3 is given by a discrete sum.

What if it had been the other way around? What if the weighted orbital integrals had occurred on the geometric side in the form of integrals

$$\int_{A^+_{M,\infty}} J_M(\gamma a, \widetilde{f})\mathrm{d}a, \qquad\qquad f \in \mathcal{H}(G),\ \gamma \in \Gamma(M)_S,$$

over the subgroup $A^+_{M,\infty}$ of $M(\mathbb{A})$, with $\widetilde{f}$ now being a function in $\mathcal{H}\big(G(\mathbb{A})\big)$ such that

$$f(x) = \int_{A^+_{M,\infty}} \widetilde{f}(xz)\mathrm{d}z,$$

while the weighted characters had occurred as a discrete sum of distributions (22.4)? It would then have been more natural to work with the general analogues of the spectral invariant distributions (22.14), rather than their geometric counterparts (22.13). Were this the case, we might want to identify $f \in \mathcal{H}(G)$ with a function on the quotient $A^+_{G,\infty}\backslash G(\mathbb{A})$. We would then identify $\Pi(M)$ with a family of representations of $A^+_{M,\infty}\backslash M(\mathbb{A})$. In the example $G = GL(2)$ above, this would lead to an application of the Poisson summation formula to the discrete image of $M(\mathfrak{o}_S)$ in $A^+_{M,\infty}\backslash M(\mathbb{A})$, rather than to the discrete subgroup $M(\mathfrak{o}_S)$ of $M(\mathbb{A})^1$.

These questions are not completely hypothetical. In the local trace formula [**A19**], which we do not have space to discuss here, weighted characters and weighted orbital integrals both occur continuously. One could therefore make the local trace formula invariant in one of two natural ways. One could equally well work with the general analogues of either of the two families (22.13) or (22.14) of invariant distributions.

## 23. The invariant trace formula

We have settled on trying to make the trace formula invariant by adding combinations of weighted characters to a given weighted orbital integral. We can now focus on the mechanics of the process.

For flexibility, we take $S$ to be any finite set of valuations of $F$. The trace formula applies to the case that $S$ is large, and contains $S_{\mathrm{ram}}$. In the example of $G = GL(2)$ in §22, the correction term in the invariant distribution (22.13) is a Fourier transform of the function

$$J_M(\pi, f), \qquad\qquad \pi \in \Pi_{\mathrm{unit}}\big(M(F_S)\big).$$

In the general case, $M$ of course need not be abelian. The appropriate analogue of the abelian dual group is not the set $\Pi_{\mathrm{unit}}\big(M(F_S)\big)$ of all unitary representations. It is rather the subset $\Pi_{\mathrm{temp}}\big(M(F_S)\big)$ of representations $\pi \in \Pi_{\mathrm{unit}}\big(M(F_S)\big)$ that are tempered, in the sense that the distributional character $f \to f_G(\pi)$ on $G(F_S)$ extends to a continuous linear form on Harish-Chandra's Schwartz space $\mathcal{C}\big(G(F_S)\big)$. Tempered representations are the spectral ingredients of Harish-Chandra's general theory of local harmonic analysis. They can be characterized as irreducible constituents of representations obtained by unitary induction from discrete series of Levi subgroups.

The tempered characters provide a mapping

$$f \longrightarrow f_G(\pi), \qquad\qquad f \in \mathcal{H}\big(G(F_S)\big),\ \pi \in \Pi_{\mathrm{temp}}\big(G(F_S)\big),$$

from $\mathcal{H}\big(G(F_S)\big)$ onto a space $\mathcal{I}\big(G(F_S)\big)$ of complex-valued functions on $\Pi_{\mathrm{temp}}\big(G(F_S)\big)$. The image of this mapping has been characterized in terms of the internal parameters of $\Pi_{\mathrm{temp}}\big(G(F_S)\big)$ ([**CD**], [**BDK**]). Roughly speaking, $\mathcal{I}\big(G(F_S)\big)$ is the space of all functions in $\Pi_{\mathrm{temp}}\big(G(F_S)\big)$ that have finite support in all discrete parameters, and lie in the relevant Paley-Wiener space in each continuous parameter. Consider a linear form $i$ on $\mathcal{I}\big(G(F_S)\big)$ that is continuous with respect to the natural topology. The corresponding linear form

$$f \longrightarrow i(f_G), \qquad\qquad f \in \mathcal{H}\big(G(F_S)\big),$$

on $\mathcal{H}\big(G(F_S)\big)$ is both continuous and invariant. Conversely, suppose that $I$ is any continuous, invariant linear form on $\mathcal{H}\big(G(F_S)\big)$. We say that $I$ is *supported on characters* if $I(f) = 0$ for any $f \in \mathcal{H}\big(G(F_S)\big)$ with $f_G = 0$. If this is so, there is a continuous linear form $\widehat{I}$ on $\mathcal{I}\big(G(F_S)\big)$ such that

$$\widehat{I}(f_G) = I(f), \qquad\qquad f \in \mathcal{H}\big(G(F_S)\big).$$

We refer to $\widehat{I}$ as the invariant Fourier transform of $I$. It is believed that every continuous, invariant linear form on $\mathcal{I}\big(G(F_S)\big)$ is supported on characters. This property is known to hold in many cases, but I do not have a comprehensive reference. The point is actually not so important here, since in making the trace formula invariant, one can show directly that the relevant invariant forms are supported on characters.

We want to apply these notions to Levi subgroups $M$ of $G$. In particular, we use the associated embedding $\widehat{I} \to I$ of distributions as a substitute for the Fourier transform of functions in (22.13). However, we have first to take care of the problem mentioned in the last section. Stated in the language of this section, the problem is that the function

$$\pi \longrightarrow J_M(\pi, f), \qquad\qquad \pi \in \Pi_{\mathrm{temp}}\big(M(F_S)\big),$$

attached to any $f \in \mathcal{H}\big(G(F_S)\big)$ does not generally lie in $\mathcal{I}\big(M(F_S)\big)$. To deal with it, we introduce a variant of the space $\mathcal{I}\big(M(F_S)\big)$.

We shall say that a set $S$ has the closure property if it either contains an archimedean valuation $v$, or contains only nonarchimedean valuations with a common residual characteristic. We assume until further notice that $S$ has this property. The image

$$\mathfrak{a}_{G,S} = H_G\big(G(F_S)\big)$$

of $G(F_S)$ in $\mathfrak{a}_G$ is then a closed subgroup of $\mathfrak{a}_G$. It equals $\mathfrak{a}_G$ if $S$ contains an archimedean place, and is a lattice in $\mathfrak{a}_G$ otherwise. In spectral terms, the action $\pi \to \pi_\lambda$ of $i\mathfrak{a}_G^*$ on $\Pi_{\mathrm{temp}}\big(G(F_S)\big)$ lifts to the quotient

$$i\mathfrak{a}_{G,S}^* = i(\mathfrak{a}_G^*/\mathfrak{a}_{G,S}^\vee), \qquad\qquad \mathfrak{a}_{G,S}^\vee = \mathrm{Hom}(\mathfrak{a}_{G,S}, 2\pi\mathbb{Z}),$$

of $i\mathfrak{a}_G^*$. If $\phi$ belongs to $\mathcal{I}\big(G(F_S)\big)$, we set

$$\phi(\pi, Z) = \int_{i\mathfrak{a}_{G,S}^*} \phi(\pi_\lambda) e^{-\lambda(Z)} d\lambda \qquad\qquad \pi \in \Pi_{\mathrm{temp}}\big(G(F_S)\big),\ Z \in \mathfrak{a}_{G,S}.$$

This allows us to identify $\mathcal{I}\big(G(F_S)\big)$ with a space of functions $\phi$ on $\Pi_{\mathrm{temp}}\big(G(F_S)\big) \times \mathfrak{a}_{G,S}$ such that

$$\phi(\pi_\lambda, Z) = e^{\lambda(Z)} \phi(\pi, Z).$$

If $f$ belongs to $\mathcal{H}\big(G(F_S)\big)$, we have

$$f_G(\pi, Z) = \mathrm{tr}\big(\pi(f^Z)\big) = \mathrm{tr}\Big(\int_{G(F_S)^Z} f(x)\pi(x)\mathrm{d}x\Big),$$

where $f^Z$ is the restriction of $f$ to the closed subset

$$G(F_S)^Z = \{x \in G(F_S) :\ H_G(x) = Z\}$$

of $G(F_S)$. In particular, $f_G(\pi, 0)$ is the character of the restriction of $\pi$ to the subgroup $G(F_S)^1$ of $G(F_S)$.

We use the interpretation of $\mathcal{I}\big(G(F_S)\big)$ as a space of functions on $G(F_S) \times \mathfrak{a}_{G,S}$ to define a larger space $\mathcal{I}_{\mathrm{ac}}\big(G(F_S)\big)$. It is clear that

$$\mathcal{I}\big(G(F_S)\big) = \varinjlim_{\Gamma} \mathcal{I}\big(G(F_S)\big)_{\Gamma},$$

where $\Gamma$ ranges over finite sets of irreducible representations of the compact group $K_S = \prod_{v \in S} K_v$, and $\mathcal{I}\big(G(F_S)\big)_{\Gamma}$ is the space of functions $\phi \in \mathcal{I}\big(G(F_S)\big)$ such that $\phi(\pi, Z)$ vanishes for any $\pi \in \Pi_{\mathrm{temp}}\big(G(F_S)\big)$ whose restriction to $K_S$ does not contain some representation in $\Gamma$. For any $\Gamma$, we define $\mathcal{I}_{\mathrm{ac}}\big(G(F_S)\big)_{\Gamma}$ to be the space of functions $\phi$ on $G(F_S) \times \mathfrak{a}_{G,S}$ with the property that for any function $b \in C_c^{\infty}(\mathfrak{a}_{G,S})$, the product

$$\phi(\pi, Z)b(Z), \qquad\qquad \pi \in \Pi_{\mathrm{temp}}\big(G(F_S)\big),\ Z \in \mathfrak{a}_{G,S},$$

lies in $\mathcal{I}\big(G(F_S)\big)_{\Gamma}$. We then set

$$\mathcal{I}_{\mathrm{ac}}\big(G(F_S)\big) = \varinjlim_{\Gamma} \mathcal{I}_{\mathrm{ac}}\big(G(F_S)\big)_{\Gamma}.$$

It is also clear that

$$\mathcal{H}\big(G(F_S)\big) = \varinjlim_{\Gamma} \mathcal{H}\big(G(F_S)\big)_{\Gamma},$$

where $\mathcal{H}\big(G(F_S)\big)_{\Gamma}$ is the subspace of functions in $\mathcal{H}\big(G(F_S)\big)$ that transform on each side under $K_S$ according to representations in $\Gamma$. We define $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)_{\Gamma}$ to be the space of functions $f$ on $G(F_S)$ such that each product

$$f(x)b\big(H_G(x)\big), \qquad\qquad x \in G(F_S),\ b \in C_c^{\infty}(\mathfrak{a}_{G,S}),$$

belongs to $\mathcal{H}\big(G(F_S)\big)_{\Gamma}$. We then set

$$\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big) = \varinjlim_{\Gamma} \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)_{\Gamma}.$$

The functions $f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$ thus have "almost compact support", in the sense that $f^Z$ has compact support for any $Z \in \mathfrak{a}_{G,S}$. If $f$ belongs to $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$, we set

$$f_G(\pi, Z) = \mathrm{tr}\big(\pi(f^Z)\big), \qquad\qquad \pi \in \Pi_{\mathrm{temp}}\big(G(F_S)\big),\ Z \in \mathfrak{a}_{G,S}.$$

Then $f \to f_G$ is a continuous linear mapping from $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$ onto $\mathcal{I}_{\mathrm{ac}}\big(G(F_S)\big)$. The mapping $I \to \widehat{I}$ can obviously be extended to an isomorphism from the space of continuous linear forms on $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$ that are supported on characters, and the space of continuous linear forms on $\mathcal{I}_{\mathrm{ac}}\big(G(F_S)\big)$.

Having completed these preliminary remarks, we are now in a position to interpret the set of weighted characters attached to $M$ as a transform of functions. Suppose that $f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$. We first attach a general meromorphic function

$$J_M(\pi_\lambda, f^Z) = \mathrm{tr}\big(\mathcal{R}_M(\pi_\lambda, P)\mathcal{I}_P(\pi_\lambda, f^Z)\big), \qquad \lambda \in \mathfrak{a}^*_{M,\mathbb{C}}, \tag{23.1}$$

to any $M \in \mathcal{L}$, $\pi \in \Pi\big(M(F_S)\big)$ and $Z \in \mathfrak{a}_{G,S}$. We can then attach a natural linear form $J_M(\pi, X, f)$ to any $X \in \mathfrak{a}_{M,S}$. For example, if $J_M(\pi_\lambda, f^Z)$ is analytic for $\lambda \in i\mathfrak{a}^*_M$, we set

$$J_M(\pi, X, f) = \int_{i\mathfrak{a}^*_{M,S}/i\mathfrak{a}^*_{G,S}} J_M(\pi_\lambda, f^Z)\mathrm{e}^{-\lambda(X)}\mathrm{d}\lambda, \tag{23.2}$$

where $Z$ is the image of $X$ in $\mathfrak{a}_{G,S}$. (In general, one must take a linear combination of integrals over contours $\varepsilon_P + i\mathfrak{a}^*_{M,S}/i\mathfrak{a}^*_{G,S}$, for groups $P \in \mathcal{P}(M)$ and small points $\varepsilon_P \in (\mathfrak{a}^*_M)^+_P$. See [**A15**, §7].) The premise underlying (23.2) holds if $\pi$ is unitary. If in addition, $S \supset S_{\mathrm{ram}}$ and $X = 0$, (23.2) reduces to the earlier definition (22.4). Our transform is given by the special case that $\pi$ belongs to the subset $\Pi_{\mathrm{temp}}\big(M(F_S)\big)$ of $\Pi_{\mathrm{unit}}\big(M(F_S)\big)$. We define $\phi_M(f)$ to be the function

$$(\pi, X) \longrightarrow \phi_M(f, \pi, X) = J_M(\pi, X, f), \qquad \pi \in \Pi_{\mathrm{temp}}\big(M(F_S)\big),\ X \in \mathfrak{a}_{M,S},$$

on $\Pi_{\mathrm{temp}}\big(M(F_S)\big) \times \mathfrak{a}_{M,S}$.

Proposition 23.1. *The mapping*

$$f \longrightarrow \phi_M(f), \qquad f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big),$$

*is a continuous linear transformation from $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$ to $\mathcal{I}_{\mathrm{ac}}\big(M(F_S)\big)$.*

This is Theorem 12.1 of [**A15**]. The proof in [**A15**] is based on a study of the residues of the meromorphic functions

$$\lambda \longrightarrow J_M(\pi_\lambda, f^Z), \qquad \lambda \in \mathfrak{a}^*_{M,\mathbb{C}},\ \pi \in \Pi_{\mathrm{temp}}\big(M(F_S)\big).$$

A somewhat simpler proof is implicit in the results of [**A13**]. (See the remark on p. 370 of [**A13**].) It is based on the splitting and descent formulas for the functions (23.1), which are parallel to (18.7) and (18.8), and are consequences of Lemmas 17.5 and 17.6. These formulas in turn yield splitting and descent formulas for the linear forms (23.2), and consequently, for the functions $\phi_M(f, \pi, X)$. They reduce the problem to the special case that $S$ contains one element $v$, $M$ is replaced by a Levi subgroup $M_v$ over $F_v$, and $\pi$ is replaced by a tempered representation $\pi_v$ of $M_v(F_v)$ that is not properly induced. The family of such representations can be parametrized by a set that is discrete modulo the action of the connected group $i\mathfrak{a}^*_{M_v,F_v} = i\mathfrak{a}^*_{M_v,\{v\}}$. The proposition can then be established from the definition of $\mathcal{I}_{\mathrm{ac}}\big(M(F_S)\big)$. □

It is the mappings $\phi_M$ that allow us to transform the various noninvariant linear forms to invariant forms. We state the construction as a pair of parallel theorems, to be followed by an extended series of remarks. The first theorem describes the general analogues of the invariant linear forms (22.13). The second theorem describes associated spectral objects. Both theorems apply to a fixed finite set of valuations $S$ with the closure property, and a Levi subgroup $M \in \mathcal{L}$.

THEOREM 23.2. *There are invariant linear forms*

$$I_M(\gamma, f) = I_M^G(\gamma, f), \qquad\qquad \gamma \in M(F_S),\ f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big),$$

*that are supported on characters, and satisfy*

$$I_M(\gamma, f) = J_M(\gamma, f) - \sum_{\substack{L \in \mathcal{L}(M) \\ L \neq G}} \widehat{I}_M^L\big(\gamma, \phi_L(f)\big). \tag{23.3}$$

THEOREM 23.3. *There are invariant linear forms*

$$I_M(\pi, X, f) = I_M^G(\pi, X, f), \qquad \pi \in \Pi\big(M(F_S)\big),\ X \in \mathfrak{a}_{M,S},\ f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big),$$

*that are supported on characters, and satisfy*

$$I_M(\pi, X, f) = J_M(\pi, X, f) - \sum_{\substack{L \in \mathcal{L}(M) \\ L \neq G}} \widehat{I}_M^L\big(\pi, X, \phi_L(f)\big). \tag{23.4}$$

**Remarks.** 1. In the special case that $G$ equals $GL(2)$, $M$ is minimal, and $S$ contains the set $S_{\mathrm{ram}} = S_\infty$, the right hand side of (23.3) reduces to the original expression (22.13). For in this case, the value of $\phi_M(\pi, X, f)$ at $X = 0$ equals the function $J_M(\pi, f)$ in (22.13). Since the linear form $I_M^M(\gamma)$ in this case is just the evaluation map of a function on $M(F_S)^1$ at $\gamma$, $\widehat{I}_M^M\big(\gamma, \phi_M(f)\big)$ reduces to the integral in (23.13) by the Fourier inversion formula for the abelian group $M(F_S)^1$.

2. The formulas (23.3) and (23.4) amount to inductive definitions of $I_M(\gamma, f)$ and $I_M(\pi, X, f)$. We need to know that these linear forms are supported on characters in order that the summands on the right hand sides of the two formulas be defined.

3. The theorems give nothing new in the case that $M = G$ and $X = Z$. For it follows immediately from the definitions that

$$I_G(\gamma, f) = J_G(\gamma, f) = f_G(\gamma)$$

and

$$I_G(\pi, Z, f) = J_G(\pi, Z, f) = f_G(\pi, Z).$$

4. The linear forms $I_M(\gamma, f)$ of Theorem 23.2 are really the primary objects. We see inductively from (23.3) that $I_M(\gamma, f)$ depends only on $f^Z$, where $Z = H_G(\gamma)$. In particular, $I_M(\gamma, f)$ is determined by its restriction to the subspace $\mathcal{H}\big(G(F_S)\big)$ of $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$. One can in fact show that as a continuous linear form on $\mathcal{H}\big(G(F_S)\big)$, $I_M(\gamma, f)$ extends continuously to the Schwartz space $\mathcal{C}\big(G(F_S)\big)$ [**A21**]. In other words, $I_M(\gamma, f)$ is a tempered distribution. It has an independent role in local harmonic analysis.

5. The linear forms $I_M(\pi, X, f)$ of Theorem 23.3 are secondary objects, but they are still interesting. We see inductively from (23.4) that $I_M(\pi, X, f)$ depends only on $f^Z$, where $Z$ is the image of $X$ in $\mathfrak{a}_{G,S}$, so $I_M(\pi, X, f)$ is also determined by its restriction to $\mathcal{H}\big(G(F_S)\big)$. However, it is not a tempered distribution. If $\pi$ is tempered,

$$J_M(\pi, X, f) = \phi_M(f, \pi, X) = \widehat{I}_M^M\big(\pi, X, \phi_M(f)\big),$$

by definition. It follows inductively from (23.4) that

$$I_M(\pi, X, f) = \begin{cases} f_G(\pi, Z), & \text{if } M = G, \\ 0, & \text{otherwise,} \end{cases} \tag{23.5}$$

in this case. But if $\pi$ is nontempered, $I_M(\pi, X, f)$ is considerably more complicated. Suppose for example that $G$ is semisimple, $M$ is maximal, $F = \mathbb{Q}$, $S = \{v_\infty\}$, and $\pi = \sigma_\mu$, for $\sigma \in \Pi_{\mathrm{temp}}\big(M(\mathbb{R})\big)$ and $\mu \in \mathfrak{a}^*_{M,\mathbb{C}}$. We assume that $\mathrm{Re}(\mu)$ is in general position. Then

$$J_M(\pi, X, f) = \int_{i\mathfrak{a}^*_M} J_M(\sigma_{\mu+\lambda}, f)\mathrm{e}^{-\lambda(X)}\mathrm{d}\lambda,$$

while

$$\phi_M(f, \pi, X) = \mathrm{e}^{\mu(X)} \int_{i\mathfrak{a}^*_M} J_M(\sigma_\lambda, f)\mathrm{e}^{-\lambda(X)}\mathrm{d}\lambda.$$

It follows that $I_M(\pi, X, f)$ is the finite sum of residues

$$\sum_\eta \operatorname*{Res}_{\Lambda=\eta}\big(J_M(\sigma_\Lambda, f)\mathrm{e}^{(\mu-\Lambda)(X)}\big),$$

obtained in deforming one contour of integration to the other. In general, $I_M(\pi, X, f)$ is a more elaborate combination of residues of general functions $J^{L_2}_{L_1}(\sigma_\Lambda, f)$.

6. The linear forms $J_M(\gamma, f)$ and $J_M(\pi, X, f)$ are strongly dependent on the choice of maximal compact subgroup $K_S = \prod_{v\in S} K_v$ of $G(F_S)$. However, it turns out that the invariant forms $I_M(\gamma, f)$ and $I_M(\pi, X, f)$ are independent of $K_S$. The proof of this fact is closely related to that of invariance, which we will discuss presently. (See [**A24**, Lemma 3.4].) The invariant linear forms are thus canonical objects, even though their construction is quite indirect.

7. The trace formula concerns the case that $S \supset S_{\mathrm{ram}}$, $\gamma \in M(F_S)^1$, and $X = 0$. In this case, the summands corresponding to $L$ in (23.3) and (23.4) depend only on the image of $\phi_L(f)$ in the invariant Hecke algebra $\mathcal{I}\big(L(F_S)^1\big)$ on $L(F_S)^1$. We can therefore take $f$ to be a function in $\mathcal{H}\big(G(F_S)^1\big)$, and treat $\phi_M$ as the mapping from $\mathcal{H}\big(G(F_S)^1\big)$ to $\mathcal{I}\big(L(F_S)^1\big)$ implicit in Proposition 23.1. In fact, since these spaces both embed in the corresponding adelic spaces, we can take $f$ to be a function in the space $\mathcal{H}(G) = \mathcal{H}\big(G(\mathbb{A})^1\big)$, and $\phi_M$ to be a mapping from $\mathcal{H}(G)$ to the adelic space $\mathcal{I}(M) = \mathcal{I}\big(M(\mathbb{A})^1\big)$. This is of course the setting of the invariant trace formula. Recall that on the geometric side, $\gamma$ represents a class in the subset $\Gamma(M)_S$ of conjugacy classes in $M(F_S)$. We write

$$I_M(\gamma, f) = I_M(\gamma, f_S) \tag{23.6}$$

as before, where $f_S$ is the restriction of $f$ to the subgroup $G(F_S)^1$ of $G(\mathbb{A})^1$. On the spectral side, $\pi$ is a representation in the subset $\Pi(M)$ of $\Pi_{\mathrm{unit}}\big(M(\mathbb{A})^1\big)$. In this case, we write

$$I_M(\pi, f) = I_M(\pi_S, 0, f_S), \tag{23.7}$$

where $S \supset S_{\mathrm{ram}}$ is any finite set outside of which both $f$ and $\pi$ are unramified, and $\pi_S \in \Pi_{\mathrm{unit}}\big(M(F_S)^1\big)$ is the $M(F_S)^1$-component of $\pi$, or rather a representative in $\Pi_{\mathrm{unit}}\big(M(\mathbb{A})\big)$ of that component.

8. The distributions $I_M(\gamma, f)$ satisfy splitting and descent formulas. We have

$$I_M(\gamma, f) = \sum_{L_1, L_2 \in \mathcal{L}(M)} d^G_M(L_1, L_2)\widehat{I}^{L_1}_M(\gamma_1, f_{1,L_1})\widehat{I}^{L_2}_M(\gamma_2, f_{2,L_2}), \tag{23.8}$$

and

$$I_M(\gamma^M_v, f_v) = \sum_{L_v \in \mathcal{L}(M_v)} d^G_{M_v}(M, L_v)\widehat{I}^{L_v}_{M_v}(\gamma_v, f_{v,L_v}), \tag{23.9}$$

under the respective conditions of (18.7) and (18.8). (In (23.8), we of course have also to ask each of the two subsets $S_1$ and $S_2$ of $S$ satisfy the closure property.) The formulas are established from the inductive definition (23.3), the formulas (18.7) and (18.8), and corresponding formulas for the functions $J_M(\pi_\lambda, f)$. (See [**A13**, Proposition 9.1 and Corollary 8.2]. If $f$ belongs to $\mathcal{H}\big(G(F_S)\big)$ and $L \in \mathcal{L}(M)$, $f_L$ is the function

$$\pi \longrightarrow f_L(\pi) = f_G(\pi^G), \qquad\qquad \pi \in \Pi_{\mathrm{temp}}\big(L(F_S)\big),$$

in $\mathcal{I}\big(L(F_S)\big)$. It is the image in $\mathcal{I}\big(L(F_S)\big)$ of any of the functions $f_Q \in \mathcal{H}\big(G(F_S)\big)$, but is independent of the choice of $Q \in \mathcal{P}(L)$.) The linear forms $J_M(\pi, X, f)$ satisfy their own splitting and descent formulas. Since these are slightly more complicated to state, we simply refer the reader to [**A13**, Proposition 9.4 and Corollary 8.5]. One often needs to apply the splitting and descent formulas to the linear forms (23.6) and (23.7) that are relevant to the trace formula. This is why one has to formulate the definitions in terms of spaces $\mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$ and $\mathcal{I}_{\mathrm{ac}}\big(L(F_S)\big)$, for general sets $S$, even though the objects (23.6) and (23.7) can be constructed in terms of the simpler spaces $\mathcal{H}(G)$ and $\mathcal{I}(L)$.

The two theorems are really just definitions, apart from the assertions that the linear forms are supported on characters. These assertions can be established globally, by exploiting the invariant trace formula of which they are the terms. In so doing, one discovers relations between the linear forms (23.3) and (23.4) that are essential for comparing traces on different groups. We shall therefore state the invariant trace formula as a third theorem, which is proved at the same time as the other two.

The invariant trace formula is completely parallel to the refined noninvariant formula of Corollary 22.1. It consists of two different expansions of a linear form $I(f) = I^G(f)$ on $\mathcal{H}(G)$ that is the invariant analogue of the original form $J(f)$. We assume inductively that for any $L \in \mathcal{L}$ with $L \neq G$, $I^L$ has been defined, and is supported on characters. We can then define $I(f)$ inductively in terms of $J(f)$ by setting

$$I(f) = J(f) - \sum_{\substack{L\in\mathcal{L}\\ L\neq G}} |W_0^L||W_0^G|^{-1}\widehat{I}^L\big(\phi_L(f)\big), \qquad f \in \mathcal{H}(G). \tag{23.10}$$

The (refined) invariant trace formula is then stated as follows.

THEOREM 23.4. *For any $f \in \mathcal{H}(G)$, $I(f)$ has a geometric expansion*

$$I(f) = \lim_{S} \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma(M)_S} a^M(\gamma) I_M(\gamma, f), \tag{23.11}$$

*and a spectral expansion*

$$I(f) = \lim_{T} \sum_{M\in S} |W_0^M||W_0^G|^{-1} \int_{\Pi(M)^T} a^M(\pi) I_M(\pi, f) d\pi. \tag{23.12}$$

**Remarks.** 9. The limit in (23.11) stabilizes for large $S$. Moreover, for any such $S$, the corresponding sums over $\gamma$ can be taken over finite sets. One can in fact be more precise. Suppose that $f$ belongs to the subspace $\mathcal{H}\big(G(F_V)^1\big)$ of $\mathcal{H}(G)$, for some finite set $V \supset S_{\mathrm{ram}}$, and is supported on a compact subset $\Delta$ of $G(\mathbb{A})^1$. Then the double sum in (23.11) is independent of $S$, so long as $S$ is large in a sense that depends only on $V$ and $\Delta$. Moreover, for any such $S$, each sum over

$\gamma$ can be taken over a finite set that depends only on $V$ and $\Delta$. These facts can be established by induction from the corresponding properties of the noninvariant geometric expansion (22.9). Alternatively, they can be established directly from [**A14**, Lemma 3.2], as on p. 513 of [**A14**].

10. For any $T$, the integral in (23.12) converges absolutely. This follows by induction from the corresponding property of the noninvariant spectral expansion (22.10). There is a weak quantitative estimate for the convergence of the limit, which is to say the convergence of the sum

$$I(f) = \sum_{t \geq 0} I_t(f)$$

of the linear forms

$$I_t(f) = I_t^G(f) = \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Pi_t(M)} a^M(\pi) I_M(\pi, f) \mathrm{d}\pi,$$

in terms of the multipliers of §20. For any $r \geq 0$, set

$$\mathfrak{h}_u^*(r, T) = \{\nu \in \mathfrak{h}_u^* : \ \|\mathrm{Re}(\nu)\| \leq r, \ \|\mathrm{Im}\,\nu\| \geq T\},$$

where $\mathfrak{h}_u^*$ is a subset of $\mathfrak{h}_{\mathbb{C}}^*/i\mathfrak{a}_G^*$ that is defined as on p. 536 of [**A14**], and contains the infinitesimal characters of all unitary representations of $G(F_\infty)^1$. Then for any $f \in \mathcal{H}(G)$, there are positive constants $C$, $k$ and $r$ with the following property. For any positive numbers $T$ and $N$, and any $\alpha$ in the subspace

$$C_N^\infty(\mathfrak{h}^1)^W = \{\alpha \in C_c^\infty(\mathfrak{h}^1)^W : \ \|\mathrm{supp}\,\alpha\| \leq N\}$$

of $\mathcal{E}(\mathfrak{h}^1)^W$, the estimate

$$\sum_{t > T} |I_t(f_\alpha)| \leq C\mathrm{e}^{kT} \sup_{\nu \in \mathfrak{h}_u^*(r,T)} \big(|\widehat{\alpha}(\nu)|\big) \tag{23.13}$$

holds. (See [**A14**, Lemma 6.3].) This "weak multiplier estimate" serves as a substitute for the absolute convergence of the spectral expansion. It is critical for applications.

As we noted above, the three theorems are proved together. We assume inductively that they all hold if $G$ is replaced by a proper Levi subgroup $L$.

It is easy to establish that the various linear forms are invariant. Fix $S$ and $M$ as in the first two theorems, and let $h$ be any function in $\mathcal{H}\big(G(F_S)\big)$. It follows easily from (22.12) that

$$\phi_L(L_h f) = \sum_{Q \in \mathcal{F}(L)} \phi_L^{M_Q}(R_{Q,h} f), \qquad f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big),$$

for any $L \in \mathcal{L}(M)$. It then follows from (22.11) and the definition (23.3) that

$$\begin{aligned} & I_M(\gamma, L_h f) \\ & = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\gamma, R_{Q,h} f) - \sum_{\substack{L \in \mathcal{L}(M) \\ L \neq G}} \sum_{Q \in \mathcal{F}(L)} \widehat{I}_M^L\big(\gamma, \phi_L^{M_Q}(R_{Q,h} f)\big) \\ & = \sum_{Q \in \mathcal{F}(M)} \Big( J_M^{M_Q}(\gamma, R_{Q,h} f) - \sum_{\substack{L \in \mathcal{L}^{M_Q}(M) \\ L \neq G}} \widehat{I}_M^L\big(\gamma, \phi_L^{M_Q}(R_{Q,h} f)\big)\Big), \end{aligned}$$

for any element $\gamma \in M(F_S)$. If $Q \neq G$, the associated summand can be written

$$\begin{aligned} & J_M^{M_Q}(\gamma, R_{Q,h}f) - \sum_{L \in \mathcal{L}^{M_Q}(M)} \widehat{I}_M^L\big(\gamma, \phi_L^{M_Q}(R_{Q,h}f)\big) \\ & = \Big(J_M^{M_Q}(\gamma, R_{Q,h}f) - \sum_{\substack{L \in \mathcal{L}^{M_Q}(M) \\ L \neq M_Q}} \widehat{I}_M^L\big(\gamma, \phi_L^{M_Q}(R_{Q,h}f)\big)\Big) - I_M^{M_Q}(\gamma, R_{Q,h}f). \end{aligned}$$

It therefore vanishes by (23.3). If $Q = G$, the corresponding summand equals

$$I_M^G(\gamma, R_{G,h}f) = I_M(\gamma, R_h f),$$

again by (23.3). Therefore $I_M(\gamma, L_h f)$ equals $I_M(\gamma, R_h f)$. It follows that $I_M(\gamma, \cdot)$ is an invariant distribution. Similarly, $I_M(\pi, X, \cdot)$ is an invariant linear form for any $\pi \in \Pi\big(M(F_S)\big)$ and $X \in \mathfrak{a}_{M,S}$. A minor variant of the argument establishes that the linear form $I$ in (23.10) is invariant as well.

It is also easy to establish the required expansions of Theorem 23.4. To derive the geometric expansion (23.11), we apply what we already know to the terms on the right hand side of the definition (23.10). That is, we substitute the geometric expansion (22.9) for $J(f)$, and we apply (23.11) inductively to the summand $\widehat{I}^L\big(\phi_L(f)\big)$ attached to any $L \neq G$. We see that $I(f)$ equals the difference between the expressions

$$\lim_S \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) J_M(\gamma, f)$$

and

$$\lim_S \sum_{L \neq G} |W_0^L||W_0^G|^{-1} \sum_{M \in \mathcal{L}^L} |W_0^M||W_0^L|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) \widehat{I}_M^L\big(\gamma, \phi_L(f)\big).$$

The second expression can be written as

$$\lim_S \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) \sum_{\substack{L \in \mathcal{L}(M) \\ L \neq G}} \widehat{I}_M^L\big(\gamma, \phi_L(f)\big).$$

Therefore $I(f)$ equals

$$\begin{aligned} & \lim_S \sum_M |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) \Big(J_M(\gamma, f) - \sum_{\substack{L \in \mathcal{L}(M) \\ L \neq G}} \widehat{I}_M^L\big(\gamma, \phi_L(f)\big)\Big) \\ & = \lim_S \sum_M |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) I_M(\gamma, f), \end{aligned}$$

by (23.3). This is the required geometric expansion (23.11). An identical argument yields the spectral expansion (23.12).

We have established the required expansions of Theorem 23.4. We have also shown that the terms in the expansions are invariant linear forms. The identity between the two expansions can thus be regarded as an invariant trace formula. If we knew that any invariant linear form was supported on characters, the inductive definitions of Theorem 23.2 and Theorem 23.3 would be complete, and we would be finished. Lacking such knowledge, we use the invariant trace formula to establish the property directly for the specific invariant linear forms in question.

PROPOSITION 23.5. *The linear forms of Theorem 23.3 can be expressed in terms of those of Theorem 23.2. In particular, if the linear forms $\{I_M(\gamma)\}$ are all supported on characters, so are the linear forms $\{I_M(\pi, X)\}$.*

The first assertion of the proposition might be more informative if it contained the phrase "in principle", since the algorithm is quite complicated. It is based on the fact that the various residues that determine the linear forms $\{I_M(\pi, X)\}$ are themselves determined by the asymptotic values in $\gamma$ of the linear forms $\{I_M(\gamma)\}$. We shall be content to illustrate the idea in a very special case.

Suppose that $G = SL(2)$, and that $M$ is minimal. Since $M$ is also maximal, the observations of Remark 5 above are relevant. Assume then that $F = \mathbb{Q}$, $S = \{v_\infty\}$, and $\pi = \sigma_\mu$, as earlier. For simplicity, we assume also that $f \in \mathcal{H}\big(G(\mathbb{R})\big)$ is invariant under the central element $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, and that $\sigma$ is the trivial representation of $M(\mathbb{R})$. It then follows from Remark 5 that for any $X \in \mathfrak{a}_M$, $I_M(\pi, X, f)$ equals the sum of residues of the function

$$\Lambda \longrightarrow \big(J_M(\sigma_\Lambda, f)\mathrm{e}^{(\mu-\Lambda)(X)}\big) \tag{23.14}$$

obtained in deforming a contour of integration from $(\mu + i\mathfrak{a}_M^*)$ to $i\mathfrak{a}_M^*$.

On the other hand,

$$\begin{aligned} I_M(\gamma, f) &= J_M(\gamma, f) - \widehat{I}_M^M\big(\gamma, \phi_M(f)\big) \\ &= J_M(\gamma, f) - \int_{i\mathfrak{a}_M^*} J_M(\sigma_\lambda, f)\mathrm{e}^{-\lambda(H_M(\gamma))}\mathrm{d}\lambda, \end{aligned}$$

for any $\gamma \in M(\mathbb{R})$. Given $X$, we choose $\gamma$ so that $H_M(\gamma) = X$. Since $f$ is compactly supported, $J_M(\gamma, f)$ is compactly supported in $X$. However, the integral over $i\mathfrak{a}_M^*$ is not generally compactly supported in $X$, since its inverse transform $\lambda \to J_M(\sigma_\lambda, f)$ can have poles in the complex domain. Therefore $I_M(\gamma, f)$ need not be compactly supported in $X$. In fact, it is the failure of $I_M(\gamma, f)$ to have compact support that determines the residues of the function (23.14). For if we apply the proof of the classical Paley-Wiener theorem to the integral over $i\mathfrak{a}_M^*$, we see that the family of functions

$$\gamma \longrightarrow I_M(\gamma_1\gamma, f), \qquad \gamma_1 \in C,$$

in which $C$ is a suitable compact subset of $M(\mathbb{R})$ and $\gamma$ is large relative to $C$ and $f$, spans a finite dimensional vector space. Moreover, it is easy to see that this space is canonically isomorphic to the space of functions of $X$ spanned by the space of residues of (23.14). It follows that the distributions $I_M(\gamma, f)$ determine the residues (23.14), and hence the linear forms $I_M(\pi, X, f)$. In particular, if $I_M(\gamma, f)$ vanishes for all such $\gamma$, then $I_M(\pi, X, f)$ vanishes for all $X$. Applied to the case that $f_G = 0$, this gives the second assertion of the proposition in the special case under consideration.

For general $G$ and $M$, the ideas are similar, but the details are considerably more elaborate. When the dimension of $\mathfrak{a}_M/\mathfrak{a}_G$ is greater than 1, we have to be concerned with partial residues and with functions whose support is compact in various directions. These are best handled with the supplementary mappings and linear forms of [**A13**, §4]. The first assertion of the proposition is implicit in the results of [**A13**, §4–5]. The second assertion is part of Theorem 6.1 of [**A13**]. $\square$

It remains to show that the distributions of Theorem 23.2 are supported on characters. From the splitting formula (23.8), one sees easily that it is enough to treat the case that $S$ contains one valuation $v$. We therefore fix a function $f_v \in \mathcal{H}\big(G(F_v)\big)$ with $f_{v,G} = 0$. The problem is to show that $I_M(\gamma_v, f_v) = 0$, for any $M \in \mathcal{L}$ and $\gamma_v \in M(F_v)$. How can we use the invariant trace formula to do this? We begin by choosing an arbitrary function $f^v \in \mathcal{H}\big(G(\mathbb{A}^v)\big)$ and letting $f$ be the restriction of $f_v f^v$ to $G(\mathbb{A})^1$. We have then to isolate the corresponding geometric expansion (23.11) in the invariant trace formula. But how is this possible, when our control of the spectral side provided by Proposition 23.5 requires an a priori knowledge of the terms on the geometric side?

The point is that the terms on the spectral side are not arbitrary members of the family defined by Theorem 23.3. They are of the form $I_M(\pi, 0, f)$, where $S \supset S_{\mathrm{ram}}$ is large enough that $f$ belongs to $\mathcal{H}\big(G(F_S)^1\big)$, and $\pi \in \Pi_{\mathrm{unit}}\big(M(F_S)\big)$. We need to show only that these terms vanish. Combining an induction argument with the splitting formula [**A13**, Proposition 9.4], one reduces the problem to showing that $I_M(\pi_v, X_v, f_v)$ vanishes for any $\pi_v \in \Pi_{\mathrm{unit}}\big(M(F_v)\big)$ and $X_v \in \mathfrak{a}_{M_v,F_v}$. The fact that $\pi_v$ is unitary is critical. The representation need not be tempered, but within the Grothendieck group it can be expressed as an integral linear combination of induced (standard) representations

$$\sigma^M_{v,\Lambda}, \qquad\qquad \sigma_v \in \Pi_{\mathrm{temp}}(M_v),\ \Lambda \in (\mathfrak{a}^M_{M_v})^*,$$

for Levi subgroups $M_v$ of $M$ over $F_v$. If $M_v = M$, $\Lambda$ equals 0, and

$$I_M(\sigma^M_{v,\Lambda}, X_v, f_v) = I_M(\sigma_v, X_v, f_v) = 0,$$

by (23.5). If $M_v \neq M$, we use the descent formula [**A13**, Corollary 8.5] to write $I_M(\sigma^M_{v,\Lambda}, X_v, f_v)$ in terms of linear forms

$$\widehat{I}^{L_v}_{M_v}(\sigma_{v,\Lambda}, Y_v, f_{v,L_v}), \qquad\qquad Y_v \in \mathfrak{a}_{M_v,F_v},$$

for Levi subgroups $L_v \in \mathcal{L}(M_v)$ with $L_v \neq G$. It follows from Proposition 23.5 and our induction hypotheses that $I_M(\sigma^M_{v,\Lambda}, X_v, f_v)$ again equals 0. Therefore $I_M(\pi_v, X_v, f_v)$ vanishes, and so therefore do the integrands on the spectral side.

We conclude that for the given function $f$, the spectral expansion (23.12) of $I(f)$ vanishes. Therefore the geometric expansion (23.11) of $I(f)$ also vanishes. In dealing with the distributions $I_M(\gamma, f)$ in this expansion, we are free to apply the splitting formula (23.8) recursively to the valuations $v \in S$. If $L \in \mathcal{L}(M)$ is a proper Levi subgroup of $G$, the induction hypotheses imply that $\widehat{I}^L_M(\gamma_v, f_v)$ vanishes for any element $\gamma_v \in M(F_v)$. It follows that

$$I_M(\gamma, f) = I_M(\gamma_v, f_v) f^v_M(\gamma^v), \qquad\qquad \gamma \in \Gamma(M)_S,$$

where $\gamma = \gamma_v \gamma^v$ is the decomposition of $\gamma$ relative to the product

$$M(F_S) = M(F_v) \times M(F^v_S).$$

Therefore

$$\lim_S \sum_{M \in \mathcal{L}} |W^M_0||W^G_0|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma) I_M(\gamma_v, f_v) f^v_M(\gamma^v) = 0. \tag{23.15}$$

We are attempting to show that $I_M(\gamma_v, f_v) = 0$, for any $M \in \mathcal{L}$ and $\gamma_v \in M(F_v)$. The definition (18.12) reduces the problem to the case that $M_{\gamma_v} = G_{\gamma_v}$. A further reduction based on invariant orbital integrals on $M(F_v)$ allows us to assume that $\gamma_v$ is strongly $G$-regular, in the sense that its centralizer in $G$ is a maximal torus $T_v$. Finally, in view of the descent formula (23.9), we can assume that $T_v$ is elliptic in $M$ over $F_v$, which is to say that $T_v$ lies in no proper Levi subgroup of $M$ over $F_v$. The problem is of course local. To solve it, one should really start with objects $G_1$, $M_1$, and $T_1$ over a local field $F_1$, together with a function $f_1 \in \mathcal{H}\big(G_1(F_1)\big)$ such that $f_{1,G_1} = 0$. One then chooses global objects $F$, $G$, $M$, and $T$ such that $F_1 = F_v$, $G_1 = G_v$, $M_1 = M_v$, and $T_1 = T_v$ for some valuation $v$ of $F$, as for example on p. 526 of [**A14**]. Among the general constraints on the choice of $G$, $M$, and $T$ is a condition that $T(F)$ be dense in $T(F_v)$. This reduces the problem to showing that $I_M(\delta, f_v)$ vanishes for any $G$-regular element $\delta \in T(F)$.

We can now sketch the proof of the remaining global argument. To exploit the identity (23.15), we have to allow the complementary function $f^v \in \mathcal{H}\big(G(\mathbb{A}^v)\big)$ to vary. We first fix a large finite set $V$ of valuations containing $v$, outside of which $G$ and $T$ are unramified. We then restrict $f^v$ to functions of the form $f_V^v f^V$, with $f^V$ being the product over $w \notin V$ of characteristic functions of $K_w$, whose support is contained in a fixed compact neighbourhood $\Delta^v$ of $\delta^v$ in $G(\mathbb{A}^v)$. According to Remark 9, the sums over $\gamma \in \Gamma(M)_S$ can be then taken over finite sets that are independent of $f^v$, for a fixed finite set of valuations $S \supset V$ that is also independent of $f^v$. Since the factors $f_M^v(\gamma^v)$ in (23.15) are actually distributions, we can allow $f_V^v$ to be a function in $C_c^\infty\big(G(F_V^v)\big)$. We choose this function so that it is supported on a small neighborhood of the image $\delta_V^v$ of $\delta^v$ in $G(F_V^v)$, and so that $f_M^v(\delta^v) = 1$. It is then easy to see that (23.15) reduces to an identity

$$\sum_\gamma c(\gamma) I_M(\gamma_v, f_v) = 0,$$

where $\gamma$ is summed over the conjugacy classes in $M(F)$ that are $G(F_w)$-conjugate to $\delta$ for any $w \in V - \{v\}$ and are $G(F_w)$-conjugate to a point in $K_w$ for every $w \notin V$, and where each coefficient $c(\gamma)$ is positive. A final argument, based on the Galois cohomology of $T$, establishes that any such $\gamma$ is actually $G(F)$-conjugate to $\delta$. This means that $\gamma = w_s^{-1} \delta w_s$ for some element $w_s \in W(M)$, and hence that

$$I_M(\gamma, f_v) = I_M(\delta, f_v).$$

(See [**A14**, pp. 527–529].) It follows that

$$I_M(\delta, f_v) = 0,$$

as required.

We have completed our sketch of the proof that the linear forms of Theorems 23.2 and 23.3 are supported on characters. The proof is a generalization of an argument introduced by Kazhdan to study invariant orbital integrals. (See [**Ka1**], [**Ka2**].) With its completion, we have also finished the collective proof of the three theorems. □

We have just devoted what might seem to be a disproportionate amount of space to a fairly arcane point. We have done so deliberately. Our proof that the linear forms $I_M(\gamma, f)$ and $I_M(\pi, X, f)$ are supported on characters can serve as a model for a family of more sophisticated arguments that are part of the general comparison of trace formulas. Instead of showing that $I_M(\gamma, f)$ and $I_M(\pi, X, f)$ vanish for certain functions $f$, as we have done here, one has to establish identities among corresponding linear forms for suitably related functions on different groups.

The invariant trace formula of Theorem 23.4 simplifies if we impose local vanishing conditions on the function $f$. We say that a function $f \in \mathcal{H}(G)$ is *cuspidal* at a place $w$ if it is the restriction to $G(\mathbb{A})^1$ of a finite sum of functions $\prod_v f_v$ whose $w$-component $f_w$ is cuspidal. This means that for any proper Levi subgroup $M_w$ of $G$ over $F_w$, the function

$$f_{w,M_w}(\pi_w) = f_{w,G}(\pi_w^G), \qquad \pi_w \in \Pi_{\mathrm{temp}}\big(M(F_w)\big),$$

in $\mathcal{I}\big(M_w(F_w)\big)$ vanishes.

COROLLARY 23.6. (a) *If $f$ is cuspidal at one place $w$, then*

$$I(f) = \lim_T \int_{\Pi_{\mathrm{disc}}(G)^T} a_{\mathrm{disc}}^G(\pi) f_G(\pi),$$

*where $\Pi_{\mathrm{disc}}(G)^T$ is the intersection of $\Pi_{\mathrm{disc}}(G)$ with $\Pi(G)^T$.*

(b) *If $f$ is cuspidal at two places $w_1$ and $w_2$, then*

$$I(f) = \lim_S \sum_{\gamma \in \Gamma(G)_S} a^G(\gamma) f_G(\gamma).$$

To establish the simple form of the spectral expansion in (a), one applies the splitting formula [**A13**, Proposition 9.4] to the linear forms $I_M(\pi, f)$ in (23.12). Combined with an argument similar to that following Proposition 23.5 above, this establishes that $I_M(\pi, f) = 0$, for any $M \neq G$, and for $f$ as in (a). Since the distribution

$$f_G(\pi) = I_G(\pi, f)$$

vanishes for any $\pi$ in the complement of $\Pi_{\mathrm{disc}}(G)^T$ in $\Pi(G)^T$, the expansion (a) follows. To establish the simple form of the spectral expansion in (b), one applies the splitting formula (23.8) to the terms $I_M(\gamma, f)$ in (23.11). This establishes that $I_M(\gamma, f) = 0$, for any $M \neq G$, and for $f$ as in (b). The expansion in (b) follows. (See the proof of Theorem 7.1 of [**A14**].) □

## 24. A closed formula for the traces of Hecke operators

In the next three sections, we shall give three applications of the invariant trace formula. The application in this section might be called the "finite case" of the trace formula. It is a finite closed formula for the traces of Hecke operators on general spaces of automorphic forms. The result can be regarded as an analogue for higher rank of Selberg's explicit formula for the traces of Hecke operators on classical spaces of modular forms.

In this section, we revert to the setting that $F = \mathbb{Q}$, in order to match standard notation for Shimura varieties. We also assume for simplicity that $A_G$ is the split component of $G$ over $\mathbb{R}$ as well as over $\mathbb{Q}$. The group

$$G(\mathbb{R})^1 = G(\mathbb{R}) \cap G(\mathbb{A})^1$$

then has compact center. The finite case of the trace formula is obtained by specializing the archimedean component of the function $f \in \mathcal{H}(G)$ in the general invariant trace formula. Before we do so, we shall formulate the problem in terms somewhat more elementary than those of recent sections.

Suppose that $\pi_{\mathbb{R}} \in \Pi_{\mathrm{unit}}\big(G(\mathbb{R})\big)$ is an irreducible unitary representation of $G(\mathbb{R})$, and that $K_0$ is an open compact subgroup of $G(\mathbb{A}_{\mathrm{fin}})$. We can write

$$L^2_{\mathrm{disc}}\big(\pi_{\mathbb{R}}, G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big) \tag{24.1}$$

for the $\pi_{\mathbb{R}}$-isotypical component of $L^2_{\mathrm{disc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big)$, which is to say, the largest subspace of $L^2_{\mathrm{disc}}\big(G(\mathbb{Q})\backslash G(\mathbb{A})^1/K_0\big)$ that decomposes under the action of $G(\mathbb{R})^1$ into a sum of copies of the restriction of $\pi_{\mathbb{R}}$ to $G(\mathbb{R})^1$. We can also write

$$L^2_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0) = L^2_{\mathrm{disc}}\big(\pi_{\mathbb{R}}, G(\mathbb{Q})\backslash G(\mathbb{A})/K_0, \zeta_{\mathbb{R}}\big)$$

for the space of functions $\phi$ on $G(\mathbb{Q})\backslash G(\mathbb{A})/K_0$ such that

$$\phi(zx) = \xi_{\mathbb{R}}(z)\phi(x), \qquad z \in A_G(\mathbb{R})^0,$$

where $\zeta_{\mathbb{R}}$ is the central character of $\pi_{\mathbb{R}}$ on $A_G(\mathbb{R})^0$, and such that the restriction of $\phi$ to $G(\mathbb{A})^1$ lies in the space (24.1). The restriction mapping from $G(\mathbb{A})$ to $G(\mathbb{A})^1$ is then a $G(\mathbb{R})^1$-isomorphism from $L^2_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$ onto the space (24.1). The action of $G(\mathbb{R})$ by right translation on $L^2_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$ is isomorphic to a direct sum of copies of $\pi_{\mathbb{R}}$, with finite multiplicity $m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$. One would like to compute the nonnegative integer $m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$.

More generally, suppose that $h$ belongs to the nonarchimedean Hecke algebra $\mathcal{H}\big(G(\mathbb{A}_{\mathrm{fin}}), K_0\big)$ attached to $K_0$. Let $R_{\mathrm{disc}}(\pi_{\mathbb{R}}, h)$ be the operator on $L^2_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$ obtained by right convolution of $h$. As an endomorphism of the $G(\mathbb{R})$-module $L^2_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$, $R_{\mathrm{disc}}(\pi_{\mathbb{R}}, h)$ can be regarded as a square matrix of rank equal to $m_{\mathrm{disc}}(\pi_{\mathbb{R}}, K_0)$. One would like a finite closed formula for its trace.

The problem just posed is too broad. However, it is reasonable to consider the question when $\pi_{\mathbb{R}}$ belongs to a restricted class of representations. We shall assume that $\pi_{\mathbb{R}}$ belongs to the subset $\Pi_{\mathrm{temp},2}\big(G(\mathbb{R})\big)$ of representations in $\Pi_{\mathrm{unit}}\big(G(\mathbb{R})\big)$ that are square integrable modulo the center of $G(\mathbb{R})$. Selberg's formula [**Sel1**] describes the solution to this problem in the case that $G = SL(2)$, $K_0 = K_{\mathrm{fin}}$ is maximal, and $\pi_{\mathbb{R}}$ is any representation in the set $\Pi_2\big(G(\mathbb{R})\big) = \Pi_{\mathrm{temp},2}\big(G(\mathbb{R})\big)$ that is also integrable.

The set $\Pi_{\mathrm{temp},2}\big(G(\mathbb{R})\big)$ is known as the discrete series, since it consists of those unitary representations of $G(\mathbb{R})$ whose restrictions to $G(\mathbb{R})^1$ occur discretely in the local spectral decomposition of $L^2\big(G(\mathbb{R})^1\big)$. The set is nonempty if and only if $G$ has a maximal torus $T_G$ that is elliptic over $\mathbb{R}$, which is to say that $T_G(\mathbb{R})/A_G(\mathbb{R})$ is compact. Assume for the rest of this section that $T_G$ exists, and that $T_G(\mathbb{R})$ is contained in the subgroup $K_{\mathbb{R}}A_G(\mathbb{R})$ of $G(\mathbb{R})$. Then $\Pi_{\mathrm{temp},2}\big(G(\mathbb{R})\big)$ is a disjoint union of finite sets $\Pi_2(\mu)$, parametrized by the irreducible finite dimensional representations $\mu$ of $G(\mathbb{R})$ with unitary central character. For any such $\mu$, the set $\Pi_2(\mu)$ consists of those representations in $\Pi_{\mathrm{temp},2}\big(G(\mathbb{R})\big)$ with the same infinitesimal character and central character as $\mu$. It is noncanonically bijective with the set of right cosets of the Weyl group $W(K_{\mathbb{R}}, T_G)$ of $K_{\mathbb{R}}$ in the Weyl group $W(G, T_G)$ of $G$. In particular, the number of elements in any packet $\Pi_2(\mu)$ equals the quotient

$$w(G) = |W(K_{\mathbb{R}}, T_G)|^{-1}|W(G, T_G)|.$$

The facts we have just stated are part of Harish-Chandra's classification of discrete series. The classification depends on a deep theory of characters that Harish-Chandra developed expressly for the purpose. We recall that the character of an arbitrary irreducible representation $\pi_{\mathbb{R}}$ of $G(\mathbb{R})$ is defined initially as the distribution

$$f_{\mathbb{R}} \longrightarrow f_{\mathbb{R},G}(\pi_{\mathbb{R}}) = \mathrm{tr}\big(\pi_{\mathbb{R}}(f_{\mathbb{R}})\big), \qquad f_{\mathbb{R}} \in C_c^{\infty}\big(G(\mathbb{R})\big),$$

on $G(\mathbb{R})$. Harish-Chandra proved the fundamental theorem that a character equals a locally integrable function $\Theta(\pi_{\mathbb{R}}, \cdot)$ on $G(\mathbb{R})$, whose restriction to the open dense set $G_{\mathrm{reg}}(\mathbb{R})$ of strongly regular elements in $G(\mathbb{R})$ is analytic [**Har1**], [**Har2**]. That is,

$$f_{\mathbb{R},G}(\pi_{\mathbb{R}}) = \int_{G_{\mathrm{reg}}(\mathbb{R})} f_{\mathbb{R}}(x)\Theta(\pi_{\mathbb{R}}, x)\mathrm{d}x, \qquad f_{\mathbb{R}} \in C_c^\infty\big(G(\mathbb{R})\big).$$

After he established his character theorem, Harish-Chandra was able to prove a simple formula for the character values of any representation $\pi_{\mathbb{R}} \in \Pi_{\mathrm{temp},2}\big(G(\mathbb{R})\big)$ in the discrete series on the regular elliptic set

$$T_{G,\mathrm{reg}}(\mathbb{R}) = T_G(\mathbb{R}) \cap G_{\mathrm{reg}}.$$

The formula is a signed sum of exponential functions that is remarkably similar to the formula of Weyl for the character of a finite dimensional representation $\mu$. However, there are two essential differences. The first is that the sum over the full Weyl group $W(G, T_G)$ in Weyl's formula is replaced by a sum over the Weyl group $W(K_{\mathbb{R}}, T_G)$ of $K_{\mathbb{R}}$. This is the reason that there are $w(G)$ representations $\pi_{\mathbb{R}}$ associated to $\mu$. The second difference is that the real group $G(\mathbb{R})$ generally has several conjugacy classes of maximal tori $T(\mathbb{R})$ over $\mathbb{R}$. This means that the character of $\pi_{\mathbb{R}}$ has also to be specified on tori other than $T_G$. Harish-Chandra gave an algorithm for computing the values of $\Theta(\pi_{\mathbb{R}}, \cdot)$ on any set $T_{\mathrm{reg}}(\mathbb{R})$ in terms of its values on $T_{G,\mathrm{reg}}(\mathbb{R})$. The resulting expression is again a linear combination of exponential functions, but now with more general integral coefficients, which can be computed explicitly from Harish-Chandra's algorithm. (For a different way of looking at the algorithm, see [**GKM**].)

We return to the problem we have been discussing. We are going to impose another restriction. Rather than evaluating the trace of a single matrix $R_{\mathrm{disc}}(\pi_{\mathbb{R}}, h)$, we have to be content at this point with a formula for the sum of such traces, taken over $\pi_{\mathbb{R}}$ in a packet $\Pi_2(\mu)$. (Given $\mu$, we shall actually sum over the packet $\Pi_2(\mu^\vee)$, where

$$\mu^\vee(x) = {}^t\mu(x)^{-1}, \qquad x \in G(\mathbb{R}),$$

is the contragredient of $\mu$.) This restriction is dictated by the present state of the invariant trace formula. There is a further refinement of the trace formula, the stable trace formula, which we shall discuss in §29. It is expected that if the stable trace formula is combined with the results we are about to describe, explicit formulas for the individual traces can be established.

We fix the irreducible finite dimensional representation $\mu$ of $G(\mathbb{R})$. The formula for the corresponding traces of Hecke operators is obtained by specializing the general invariant trace formula. In particular, it will retain the general structure of a sum over groups $M \in \mathcal{L}$. Each summand contains a product of three new factors, which we now describe.

The most interesting factor is a local function

$$\Phi'_M(\mu, \gamma_{\mathbb{R}}), \qquad \gamma_{\mathbb{R}} \in M(\mathbb{R}),$$

on $M(\mathbb{R})$ attached to the archimedean valuation $v_\infty$. Assume first that $\gamma_{\mathbb{R}}$ lies in $T_M(\mathbb{R}) \cap G_{\mathrm{reg}}$, where $T_M$ is a maximal torus in $M$ over $\mathbb{R}$ such that $T_M(\mathbb{R})/A_M(\mathbb{R})$ is compact. In this case, we set

$$\Phi'_M(\mu, \gamma_{\mathbb{R}}) = |D_M^G(\gamma_{\mathbb{R}})|^{\frac{1}{2}} \sum_{\pi_{\mathbb{R}} \in \Pi_2(\mu)} \Theta(\pi_{\mathbb{R}}, \gamma_{\mathbb{R}}), \tag{24.2}$$

where
$$D_M^G(\gamma_{\mathbb{R}}) = D^G(\gamma_{\mathbb{R}})D^M(\gamma_{\mathbb{R}})^{-1}$$
is the relative Weyl discriminant. It is a straightforward consequence of the character formulas for discrete series that $\Phi'_M(\mu,\gamma_{\mathbb{R}})$ extends to a continuous function on the torus $T_M(\mathbb{R})$. (See [**A16**, Lemma 4.2].) If $\gamma_{\mathbb{R}} \in M(\mathbb{R})$ does not belong to any such torus, we set $\Phi'_M(\mu,\gamma_{\mathbb{R}}) = 0$. The function $\Phi'_M(\mu,\gamma_{\mathbb{R}})$ on $M(\mathbb{R})$ is complicated enough to be interesting (because it involves characters of discrete series on nonelliptic tori in $G(\mathbb{R})$), but simple enough to be given explicitly (because there are concrete formulas for such characters). It is supported on the semisimple elements in $M(\mathbb{R})$, and is invariant under conjugation by $M(\mathbb{R})$.

The second factor is a local term attached to the nonarchimedean valuations. If $\gamma$ is any semisimple element in $M(\mathbb{Q})$, we write

(24.3)
$$h'_M(\gamma) = \delta_P(\gamma_{\mathrm{fin}})^{\frac{1}{2}} \int_{K_{\mathrm{fin}}} \int_{N_P(\mathbb{A}_{\mathrm{fin}})} \int_{M_\gamma(\mathbb{A}_{\mathrm{fin}})\backslash M(\mathbb{A}_{\mathrm{fin}})} h(k^{-1}m^{-1}\gamma mnk)dndndk,$$
where $P$ is any group in $\mathcal{P}(M)$, $\delta_P(\gamma_{\mathrm{fin}})$ is the modular function on $P(\mathbb{A}_{\mathrm{fin}})$, and $K_{\mathrm{fin}}$ is our maximal compact subgroup of $G(\mathbb{A}_{\mathrm{fin}})$. ( In [**A16**], this function was denoted $h_M(\gamma)$ rather than $h'_M(\gamma)$. However, the symbol $h_M(\gamma)$ has since been used to denote the normalized orbital integral
$$h_M(\gamma) = |D^M(\gamma_{\mathrm{fin}})|^{\frac{1}{2}} h'_M(\gamma).)$$
Since the integrals in (24.3) reduce to finite linear combinations of values assumed by the locally constant function $h$, $h'_M(\gamma)$ can in principle by computed explicitly.

The third factor is a global term. It is defined only for semisimple elements $\gamma \in M(\mathbb{Q})$ that lie in $T_M(\mathbb{R})$, for a maximal torus $T_M$ in $M$ over $\mathbb{R}$ such that $T_M(\mathbb{R})/A_M(\mathbb{R})$ is compact. For any such $\gamma$, we set

(24.4)
$$\chi(M_\gamma) = (-1)^{q(M_\gamma)}\mathrm{vol}\big(\overline{M}_\gamma(\mathbb{Q})\backslash\overline{M}_\gamma(\mathbb{A}_{\mathrm{fin}})\big)w(M_\gamma),$$
where
$$q(M_\gamma) = \frac{1}{2}\dim\big(M_\gamma(\mathbb{R})/K_{\gamma,\mathbb{R}}A_M(\mathbb{R})^0\big)$$
is one-half the dimension of the symmetric space attached to $M_\gamma$, while $\overline{M}_\gamma$ is an inner twist of $M_\gamma$ over $\mathbb{Q}$ such that $\overline{M}_\gamma(\mathbb{R})/A_M(\mathbb{R})^0$ is compact, and $w(M_\gamma)$ is the analogue for $M_\gamma$ of the positive integer $w(G)$ defined for $G$ above. The volume in the product $\chi(M_\gamma)$ is taken with respect to the inner twist of a chosen Haar measure on $M_\gamma(\mathbb{A}_{\mathrm{fin}})$. We note that the product of the Haar measure on $M_\gamma(\mathbb{A}_{\mathrm{fin}})$ with the invariant measures in the definition of $h'_M(\gamma)$ determines a Haar measure on $G(\mathbb{A}_{\mathrm{fin}})$. This measure is supposed to coincide with the Haar measure used to define the original operator $R_{\mathrm{disc}}(\pi_{\mathbb{R}},h)$ by right convolution of $h$ on $G(\mathbb{A})$.

THEOREM 24.1. *Suppose that the highest weight of the finite dimensional representation $\mu$ of $G(\mathbb{R})$ is nonsingular. Then for any element $h \in \mathcal{H}\big(G(\mathbb{A}_{\mathrm{fin}}),K_0\big)$, the sum*

(24.5)
$$\sum_{\pi_{\mathbb{R}}\in\Pi_2(\mu^\vee)} \mathrm{tr}\big(R_{\mathrm{disc}}(\pi_{\mathbb{R}},h)\big)$$

*equals the geometric expansion*

(24.6)
$$\sum_{M\in\mathcal{L}}(-1)^{\dim(A_M/A_G)}|W_0^M||W_0^G|^{-1}\sum_{\gamma\in\Gamma(M)}|\iota^M(\gamma)|^{-1}\chi(M_\gamma)\Phi'_M(\mu,\gamma)h'_M(\gamma).$$

To establish the formula, one has to specialize the function $f$ in the general invariant trace formula. The finite dimensional representation $\mu$ satisfies

$$\mu(zx) = \zeta_{\mathbb{R}}(z)^{-1}\mu(x), \qquad z \in A_G(\mathbb{R})^0,\ x \in G(\mathbb{R}),$$

for a unitary character $\zeta_{\mathbb{R}}$ on $A_G(\mathbb{R})^0$. Its contragredient $\mu^\vee$ has central character $\zeta_{\mathbb{R}}$ on $A_G(\mathbb{R})^0$. The associated packet $\Pi_2(\mu^\vee)$ is contained in the set $\Pi_{\mathrm{temp}}\big(G(\mathbb{R}),\zeta_{\mathbb{R}}\big)$ of tempered representations of $G(\mathbb{R})$ whose central character on $A_G(\mathbb{R})^0$ equals $\zeta_{\mathbb{R}}$. Now the characterization [**CD**] of the invariant image $\mathcal{I}\big(G(\mathbb{R})\big)$ of $\mathcal{H}\big(G(\mathbb{R})\big)$ applies equally well to the $\zeta_{\mathbb{R}}^{-1}$-equivariant analogue $\mathcal{H}\big(G(\mathbb{R}),\zeta_{\mathbb{R}}\big)$ of the Hecke algebra. It implies that there is a function $f_{\mathbb{R}}$ in $\mathcal{H}\big(G(\mathbb{R}),\zeta_{\mathbb{R}}\big)$ such that

$$f_{\mathbb{R},G}(\pi_{\mathbb{R}}) = \begin{cases} 1, & \text{if } \pi_{\mathbb{R}} \in \Pi_2(\mu^\vee), \\ 0, & \text{otherwise,} \end{cases} \tag{24.7}$$

for any representation $\pi_{\mathbb{R}} \in \Pi_{\mathrm{temp}}\big(G(\mathbb{R}),\zeta_{\mathbb{R}}\big)$. The restriction $f$ of the product $f_{\mathbb{R}}h$ to $G(\mathbb{A})^1$ is then a function in $\mathcal{H}(G)$. We shall substitute it into the invariant trace formula.

Since $f_{\mathbb{R},G}$ vanishes on the complement of the discrete series in $\Pi_{\mathrm{temp}}\big(G(\mathbb{R}),\zeta_{\mathbb{R}}\big)$, $f_{\mathbb{R}}$ is cuspidal. By Corollary 23.6(a), the spectral expansion of $I(f)$ simplifies. We obtain

$$\begin{aligned} I(f) &= \lim_T \sum_{\pi \in \Pi_{\mathrm{disc}}(G)^T} a^G_{\mathrm{disc}}(\pi) f_G(\pi) \\ &= \sum_t I_{t,\mathrm{disc}}(f) \\ &= \sum_t \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{s \in W(M)_{\mathrm{reg}}} |\det(s-1)_{\mathfrak{a}_M^G}|^{-1} \mathrm{tr}\big(M_P(s,0)\mathcal{I}_{P,t}(0,f)\big). \end{aligned}$$

The irreducible constituents of the representation $\mathcal{I}_{P,t}(0,f)$ could well be nontempered. However, given that $s \in W(M)$ is regular, and that the tempered support of $f_{\mathbb{R},G}$ contains no representation with singular infinitesimal character, one deduces that

$$\mathrm{tr}\big(M_P(s,0)\mathcal{I}_{P,t}(0,f)\big) = 0,$$

as long as $M \neq G$. (See [**A16**, p. 268].) The terms with $M \neq G$ therefore vanish. The expansion reduces simply to

$$I(f) = \sum_t \sum_{\pi \in \Pi_{t,\mathrm{disc}}(G)} m_{\mathrm{disc}}(\pi)\mathrm{tr}\big(\pi(f_{\mathbb{R}}h)\big), \tag{24.8}$$

the contribution from the discrete spectrum. There can of course be nontempered representations $\pi$ with $m_{\mathrm{disc}}(\pi) \neq 0$. But the condition that the highest weight of $\mu$ be nonsingular is stronger than the conditions on $f_{\mathbb{R},G}$ used to derive (24.8). It can be seen to imply that the summands in (24.8) corresponding to nontempered archimedean components $\pi_{\mathbb{R}}$ vanish. (The proof on p. 283 on [**A16**], which uses the classification of unitary representations $\pi_{\mathbb{R}}$ with cohomology, anticipates Corollary 24.2 below.) It follows that

$$I(f) = \sum_t \sum_{\{\pi : \pi_{\mathbb{R}} \in \Pi_2(\mu^\vee)\}} m_{\mathrm{disc}}(\pi) f_{\mathbb{R},G}(\pi_{\mathbb{R}}) h_M(\pi_{\mathrm{fin}}).$$

This in turn implies that $I(f)$ equals the sum (24.5).

The problem is then to compute the geometric expansion (23.11) of $I(f)$, for the chosen function $f$ defined by $f_{\mathbb{R}}h$. Consider the terms

$$I_M(\gamma, f), \qquad\qquad M \in \mathcal{L},\ \gamma \in \Gamma(M)_S,$$

in (23.11). We apply the splitting formula (23.8) successively to the valuations in $S$. If $L \in \mathcal{L}(M)$ is proper in $G$, the contribution $\widehat{I}^L_M(\gamma_{\mathbb{R}}, f_{\mathbb{R},L})$ to the formula vanishes. It follows that

$$I_M(\gamma, f) = I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})h_M(\gamma_{\mathrm{fin}}).$$

The sum of traces (24.5) therefore equals

$$\lim_S \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma(M)_S} a^M(\gamma) I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})h_M(\gamma_{\mathrm{fin}}). \tag{24.9}$$

The problem reduces to that of computing the archimedean component $I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})$, for elements $\gamma_{\mathbb{R}} \in M(\mathbb{R})$.

Suppose that $\gamma_{\mathbb{R}} = t_{\mathbb{R}}$ is strongly $G$-regular. In this case, the main theorem of [**A1**] provides a formula for $I_M(t_{\mathbb{R}}, f_{\mathbb{R}})$ in terms of character values of discrete series at $t_{\mathbb{R}}$. The proof uses differential equations and boundary conditions satisfied by $I_M(t_{\mathbb{R}}, f_{\mathbb{R}})$ to reduce the problem to the case $M = G$, which had been solved earlier by Harish-Chandra [**Har3**]. A more conceptual proof of the same formula came later, as a consequence of the local trace formula [**A20**, Theorem 5.1]. (A $p$-adic analogue for Lie algebras of this result is contained in the lectures of Kottwitz [**Ko8**].) If $t_{\mathbb{R}}$ is elliptic in $M(\mathbb{R})$, the formula asserts that $I_M(t_{\mathbb{R}}, f_{\mathbb{R}})$ equals the product of

$$(-1)^{\dim(A_M/A_G)}\mathrm{vol}\big(T_M(\mathbb{R})/A_M(\mathbb{R})^0\big)^{-1}$$

with

$$\sum_{\pi_{\mathbb{R}}\in\Pi_2(G(\mathbb{R}),\zeta_{\mathbb{R}})} |D^G(t_{\mathbb{R}})|^{\frac{1}{2}}\Theta(\pi_{\mathbb{R}}, t_{\mathbb{R}}) f_{\mathbb{R},G}(\pi_{\mathbb{R}}),$$

where $T_M$ is the centralizer of $t_{\mathbb{R}}$. It follows that

$$I_M(t_{\mathbb{R}}, f_{\mathbb{R}}) = (-1)^{\dim(A_M/A_G)}\mathrm{vol}\big(T_M(\mathbb{R})/A_M(\mathbb{R})^0\big)^{-1}|D^M(t_{\mathbb{R}})|^{\frac{1}{2}}\Phi'_M(\mu, t_{\mathbb{R}}). \tag{24.10}$$

If $t_{\mathbb{R}}$ is not elliptic in $M(\mathbb{R})$, the formula of [**A1**] (or just the descent formula (23.9)) tells us that $I_M(t_{\mathbb{R}}, f_{\mathbb{R}})$ vanishes. Since $\Phi'_M(\mu, t_{\mathbb{R}})$ vanishes by definition in this case, (24.10) holds for any strongly $G$-regular element $t_{\mathbb{R}}$.

It remains to sketch a generalization of (24.10) to arbitrary elements $\gamma_{\mathbb{R}} \in M(\mathbb{R})$. From the definitions (18.12) and (23.3), we deduce that

$$I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}}) = \lim_{a_{\mathbb{R}}\to 1} \sum_{L\in\mathcal{L}(M)} r^L_M(\gamma_{\mathbb{R}}, a_{\mathbb{R}}) I_L(a_{\mathbb{R}}\gamma_{\mathbb{R}}, f_{\mathbb{R}}),$$

for small points $a_{\mathbb{R}} \in A_M(\mathbb{R})$ in general position. Since $f_{\mathbb{R}}$ is cuspidal, the descent formula (23.9) implies that the summands on the right with $L \neq M$ vanish. Replacing $\gamma_{\mathbb{R}}$ by $a_{\mathbb{R}}\gamma_{\mathbb{R}}$, if necessary, we can therefore assume that the centralizer of $\gamma_{\mathbb{R}}$ in $G$ is contained in $M$. In this case, $I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})$ can be approximated by functions $I_M(t_{\mathbb{R}}, f_{\mathbb{R}})$, for $G$-regular elements $t_{\mathbb{R}}$ in $M(\mathbb{R})$ that are close to the semisimple part $\sigma_{\mathbb{R}}$ of $\gamma_{\mathbb{R}}$. We can actually assume that $\sigma_{\mathbb{R}}$ lies in an elliptic torus $T_M$, again by the descent formula (23.9). The approximation of $I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})$ then takes the form of a limit formula

$$I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}}) = \lim_{t_{\mathbb{R}}\to\sigma_{\mathbb{R}}} \big(\partial(h_{u_{\mathbb{R}}}) I_M(t_{\mathbb{R}}, f_{\mathbb{R}})\big),$$

where $\partial(h_{u_{\mathbb{R}}})$ is a harmonic differential operator on $T_M(\mathbb{R})$ attached to the unipotent part $u_{\mathbb{R}}$ of $\gamma_{\mathbb{R}}$ [**A16**, Lemma 5.2]. One can compute the limit from the properties of the function $\Phi'_M(\mu, t_{\mathbb{R}})$ on the right hand side of (24.10). The fact that this function is constructed from a sum of characters of discrete series in the packet $\Pi_2(\mu)$ is critical. One uses it to show that the limit vanishes unless $u_{\mathbb{R}} = 1$. The conclusion [**A16**, Theorem 5.1] is that

$$I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}}) = (-1)^{\dim(A_M/A_G)} v(M_{\gamma_{\mathbb{R}}})^{-1} |D^M(\gamma_{\mathbb{R}})|^{\frac{1}{2}} \Phi'_M(\mu, \gamma_{\mathbb{R}}),$$

where

$$v(M_{\gamma_{\mathbb{R}}}) = (-1)^{q(M_{\gamma_{\mathbb{R}}})} \mathrm{vol}\big(\overline{M}_{\gamma_{\mathbb{R}}}(\mathbb{R})/A_M(\mathbb{R})^0\big) w(M_{\gamma_{\mathbb{R}}})^{-1}.$$

In particular, $I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})$ vanishes unless $\gamma_{\mathbb{R}}$ is semisimple and lies in an elliptic maximal torus $T_M$.

We substitute the general formula for $I_M(\gamma_{\mathbb{R}}, f_{\mathbb{R}})$ into the expression (24.9) for $I(f)$. We see that the summand in (24.9) corresponding to $\gamma \in \Gamma(M)_S$ vanishes unless $\gamma$ is semisimple. Since $(M,S)$-equivalence of semisimple elements in $\Gamma(M)_S$ is the same as $M(\mathbb{Q})$-conjugacy, we can sum $\gamma$ over the set $\Gamma(M)$ instead of $\Gamma(M)_S$, removing the limit over $S$ at the same time. We can also write

$$\begin{aligned} &|D^M(\gamma_{\mathbb{R}})|^{\frac{1}{2}} \Phi'_M(\mu, \gamma_{\mathbb{R}}) h_M(\gamma_{\mathrm{fin}}) \\ &= |D^M(\gamma_{\mathbb{R}}) D^M(\gamma_{\mathrm{fin}})|^{\frac{1}{2}} \Phi'_M(\mu, \gamma_{\mathbb{R}}) h'_M(\gamma_{\mathrm{fin}}) \\ &= \Phi'_M(\mu, \gamma) h'_M(\gamma), \end{aligned}$$

for any semisimple element $\gamma \in M(\mathbb{Q})$, by the product formula for $\mathbb{Q}$. Finally, it follows from the definitions (19.5) and (22.2) of $a^M(\gamma)$, together with the main theorem of [**Ko6**], that

$$a^M(\gamma) v(M_{\gamma_{\mathbb{R}}})^{-1} = \chi(M_\gamma) |\iota^M(\gamma)|^{-1},$$

again for any semisimple element $\gamma \in M(\mathbb{Q})$. We conclude that $I(f)$ is equal to the required expression (24.6). Since it is also equal to the original sum (24.5), the theorem follows. □

**Remarks.** 1. The theorem from [**Ko6**] we have just appealed to is that the coefficient

$$a^G(1) = \mathrm{vol}\big(G(F)\backslash G(\mathbb{A})^1\big)$$

is invariant under inner twisting of $G$. Kottwitz was able to match the terms with $M = G$ and $\gamma = 1$ in the fine geometric expansion (22.9) for any two groups related by inner twisting. This completed the proof of the Weil conjecture on Tamagawa numbers, following a suggestion from [**JL**, §16]. It represents a different and quite striking application of the general trace formula, which clearly illustrates the need for a fine geometric expansion. Unfortunately, we do not have space to discuss it further.

2. The condition that the highest weight of $\mu$ be nonsingular was studied by F. Williams [**Wi**], in connection with multiplicity formulas for compact quotient. It is weaker than the condition that the relevant discrete series representations be integrable, which was used in the original multiplicity formulas of Langlands [**Lan2**].

If our condition on the highest weight of $\mu$ is removed, the expression (24.6) still makes sense. To what does it correspond?

Assume that

$$\mu: \ G \longrightarrow GL(V)$$

is an irreducible finite dimensional representation of $G$ that is defined over $\mathbb{Q}$. This represents a slight change of perspective. On the one hand, we are asking that the restriction of $\mu$ to the center of $G$ be algebraic, and that the representation itself be defined over $\mathbb{Q}$. On the other, we are relaxing the condition that the central character $\zeta_{\mathbb{R}}^{-1}$ of $\mu$ on $A_G(\mathbb{R})^0$ be unitary. The corresponding packet $\Pi_2(\mu)$ still exists, but it is now contained only in the set $\Pi_2\big(G(\mathbb{R})\big)$ of general (not necessarily tempered) representations of $G(\mathbb{R})$ that are square integrable modulo the center. We define the function $\Phi'_M(\gamma_{\mathbb{R}}, \mu)$ exactly as before.

If $K'_{\mathbb{R}} = K_{\mathbb{R}} A_G(\mathbb{R})^0$, the quotient

$$X = G(\mathbb{R})/K'_{\mathbb{R}}$$

is a globally symmetric space with respect to a fixed left $G(\mathbb{R})$-invariant metric. Let us assume that none of the simple factors of $G$ is anisotropic over $\mathbb{R}$. We assume also that the open compact subgroup $K_0 \subset G(\mathbb{A}_{\mathrm{fin}})$ is small enough that the action of $G(\mathbb{Q})$ on the product of $X$ with $G(\mathbb{A}_{\mathrm{fin}})/K_0$ has no fixed points. The quotient

$$\mathcal{M}(K_0) = G(\mathbb{Q})\backslash\big(X \times G(\mathbb{A}_{\mathrm{fin}})/K_0\big)$$

is then a finite union of locally symmetric spaces. Moreover, the restriction of the representation $\mu$ to $G(\mathbb{Q})$ determines a locally constant sheaf

$$\mathcal{F}_\mu(K_0) = V(\mathbb{C}) \underset{G(\mathbb{Q})}{\times} \big(X \times G(\mathbb{A}_{\mathrm{fin}})/K_0\big)$$

on $\mathcal{M}(K_0)$.

One can form the $L^2$-cohomology

$$H^*_{(2)}\big(\mathcal{M}(K_0), \mathcal{F}_\mu(K_0)\big) = \bigoplus_{q \geq 0} H^q_{(2)}\big(\mathcal{M}(K_0), \mathcal{F}_\mu(K_0)\big)$$

of $\mathcal{M}(K_0)$ with values in $\mathcal{F}_\mu$. It is a finite dimensional graded vector space, which reduces to ordinary de Rham cohomology in the case that $\mathcal{M}(K_0)$ is compact. The element $h$ in the Hecke algebra $\mathcal{H}\big(G(\mathbb{A}_{\mathrm{fin}}), K_0\big)$ acts by right convolution on any reasonable space of functions or differential forms on $\mathcal{M}(K_0)$. It yields an operator

$$H^*_{(2)}\big(h, \mathcal{F}_\mu(K_0)\big) = \bigoplus_{q} H^q_{(2)}\big(h, \mathcal{F}_\mu(K_0)\big)$$

on the $L^2$-cohomology space. Let

$$\mathcal{L}_\mu(h) = \sum_q (-1)^q \mathrm{tr}\Big(H^q_{(2)}\big(h, \mathcal{F}_\mu(K_0)\big)\Big) \tag{24.11}$$

be its Lefschetz number.

Corollary 24.2. *The Lefschetz number $\mathcal{L}_\mu(h)$ equals the product of $(-1)^{q(G)}$ with the geometric expression (24.6).*

The reduction of the corollary to the formula of the theorem depends on the spectral decomposition of $L^2$-cohomology [**BC**], and the Vogan-Zuckermann classification [**VZ**] of unitary representations of $G(\mathbb{R})$ with $\big(\mathfrak{g}(\mathbb{R}), K'_{\mathbb{R}}\big)$-cohomology. These matters are discussed in §2 of [**A16**]. We shall include only a few words here.

The space $H^q_{(2)}\big(\mathcal{M}(K_0), \mathcal{F}_\mu(K_0)\big)$ is defined by square-integrable differential $q$-forms on $\mathcal{M}(K_0)$. Consider the case that $\mathcal{M}(K_0)$ is compact. Elements in the space are then defined by smooth, differential $q$-forms on $\mathcal{M}(K_0)$ with values in $\mathcal{F}_\mu(K_0)$. By thinking carefully about the nature of such objects, one is led to a canonical isomorphism

$$H^q_{(2)}\big(\mathcal{M}(K_0), \mathcal{F}_\mu(K_0)\big) \cong \bigoplus_{\pi \in \Pi_{\mathrm{unit}}(G(\mathbb{A}),\zeta_{\mathbb{R}})} m_{\mathrm{disc}}(\pi)\big(H^q\big(\mathfrak{g}(\mathbb{R}), K'_{\mathbb{R}}; \pi_{\mathbb{R}} \otimes \mu\big) \otimes \pi_{\mathrm{fin}}^{K_0}\big),$$

in which $\Pi_{\mathrm{unit}}\big(G(\mathbb{A}), \zeta_{\mathbb{R}}\big)$ denotes the set of representations in $\Pi\big(G(\mathbb{A}), \zeta_{\mathbb{R}}\big)$ that are unitary modulo $A_G(\mathbb{R})^0$, $H^q\big(\mathfrak{g}(\mathbb{R}), K'_{\mathbb{R}}; \cdot\big)$ represents the $\big(\mathfrak{g}(\mathbb{R}), K'_{\mathbb{R}}\big)$-cohomology groups defined in [**BW**, Chapter II], for example, and $\pi_{\mathrm{fin}}^{K_0}$ stands for the space of $K_0$-invariant vectors for the finite component $\pi_{\mathrm{fin}}$ of $\pi$. (See [**BW**, Chapter VII].) This isomorphism is compatible with the canonical action of the Hecke algebra $\mathcal{H}\big(G(\mathbb{A}_{\mathrm{fin}}), K_0\big)$ on each side. It follows that there is a canonical isomorphism of operators

$$H^q_{(2)}\big(h; \mathcal{F}_\mu(K_0)\big) \cong \bigoplus_{\pi \in \Pi_{\mathrm{unit}}(G(\mathbb{A}),\zeta_{\mathbb{R}})} m_{\mathrm{disc}}(\pi) \cdot \dim\big(H^q\big(\mathfrak{g}(\mathbb{R}), K'_{\mathbb{R}}; \pi_{\mathbb{R}} \otimes \mu\big)\big) \cdot \pi_{\mathrm{fin}}(h).$$

In the paper [**BC**], Borel and Casselman show that this isomorphism carries over to the case of noncompact quotient (with our assumption that $G(\mathbb{R})$ has discrete series). Define

$$\chi_\mu(\pi_{\mathbb{R}}) = \sum_q (-1)^q \dim\big(H^q\big(\mathfrak{g}(\mathbb{R}), K'_{\mathbb{R}}; \pi_{\mathbb{R}} \otimes \mu\big)\big),$$

for any unitary representation $\pi_{\mathbb{R}}$ of $G(\mathbb{R})$. It then follows that

$$\mathcal{L}_\mu(h) = \sum_{\pi \in \Pi_{\mathrm{unit}}(G(\mathbb{A}),\zeta_{\mathbb{R}})} m_{\mathrm{disc}}(\pi)\chi_\mu(\pi_{\mathbb{R}})\mathrm{tr}\big(\pi_{\mathrm{fin}}(h)\big). \tag{24.12}$$

The second step is to describe the integers $\chi_\mu(\pi_{\mathbb{R}})$. This is done in [**CD**]. The result can be expressed as an identity

$$\chi_\mu(\pi_{\mathbb{R}}) = (-1)^{q(G)} f_{\mathbb{R},G}(\pi_{\mathbb{R}}), \qquad \pi_{\mathbb{R}} \in \Pi_{\mathrm{unit}}\big(G(\mathbb{R}), \zeta_{\mathbb{R}}\big),$$

where $f_{\mathbb{R}} \in \mathcal{H}\big(G(\mathbb{R}), \zeta_{\mathbb{R}}\big)$ is a function that satisfies (24.7). It follows that if $\pi$ is as in (24.8), then

$$\begin{aligned}\chi_\mu(\pi_{\mathbb{R}})\mathrm{tr}\big(\pi_{\mathrm{fin}}(h)\big) &= (-1)^{q(G)}\mathrm{tr}\big(\pi_{\mathbb{R}}(f_{\mathbb{R}})\big)\mathrm{tr}\big(\pi_{\mathrm{fin}}(h)\big) \\ &= (-1)^{q(G)}\mathrm{tr}\big(\pi(f_{\mathbb{R}}h)\big),\end{aligned}$$

where $\pi_{\mathbb{R}} \otimes \pi_{\mathrm{fin}}$ is the representation in $\Pi_{\mathrm{unit}}\big(G(\mathbb{A}), \zeta_{\mathbb{R}}\big)$ whose restriction to $G(\mathbb{A})^1$ equals $\pi$. It follows from (24.8) and (24.12) that

$$\mathcal{L}_\mu(h) = (-1)^{q(G)} I(f).$$

Since we have already seen that $I(f)$ equals the geometric expression (24.6), the corollary follows. □

The formula of Corollary 24.2 is relevant to Shimura varieties. The reader will recall from the lectures of Milne [**Mi**] that with further conditions on $G$, the space $\mathcal{M}(K_0)$ becomes the set of complex points of a Shimura variety. It is a fundamental problem for Shimura varieties to establish reciprocity laws between the analytic data contained in Hecke operators on $L^2$-cohomology, and the arithmetic data contained in $\ell$-adic representations of Galois groups on étale cohomology. Following the strategy that was successful for $GL(2)$ [**Lan4**], one would try to compare geometric sides of two Lefschetz formulas. Much progress has been made in the case that $\mathcal{M}(K_0)$ is compact [**Ko7**]. In the general case, the formula of Corollary 24.2 could serve as the basic analytic Lefschetz formula. (One still has to "stabilize" this formula, a problem closely related to that of computing the individual summands in (24.5), as opposed to their sum.) The other ingredient would be a Lefschetz trace formula for Frobenius-Hecke correspondences on the $\ell$-adic intersection cohomology of the Bailey-Borel compactification $\overline{\mathcal{M}(K_0)}$, and a comparison of its gemetric terms with those of the analytic formula.

The general problem is still far from being solved. However, Goresky, Kottwitz, and Macpherson have taken an important step. They have established a formula for the Lefschetz numbers of Hecke correspondences in the complex intersection cohomology of $\overline{\mathcal{M}(K_0)}$, whose geometric terms match those of the analytic formula [**GKM**]. Since one knows that the spectral sides of the two formulas match, by Zucker's conjecture [**Lo**], [**SS**], the results of Goresky, Kottwitz, and Macpherson can be regarded as a topological proof of the formula of Corollary 24.2. It is hoped that their methods can be applied to $\ell$-adic intersection cohomology.

## 25. Inner forms of $GL(n)$

The other two applications each entail a comparison of trace formulas. They concern higher rank analogues of the Jacquet-Langlands correspondence, and the theorem of Saito-Shintani and Langlands on base change for $GL(2)$. These two applications are the essential content of the monograph [**AC**]. Since we are devoting only limited space to them here, our discussion will have to be somewhat selective.

The two comparisons were treated together in [**AC**]. However, it is more instructive to discuss them separately. In this section we will discuss a partial generalization of the Jacquet-Langlands correspondence from $GL(2)$ to $GL(n)$. We shall describe a term by term comparison of the invariant trace formula of the multiplicative group of a central simple algebra with that of $GL(n)$.

We return to the general setting of Part II, in which $G$ is defined over a number field $F$. In this section, $G^*$ will stand for the general linear group $GL(n)$ over $F$. We take $G$ to be an inner twist of $G^*$ over $F$. This means that $G$ is equipped with an isomorphism $\psi$: $G \to G^*$ such that for every element $\tau$ in $\Gamma_F = \mathrm{Gal}(\overline{F}/F)$, the relation

$$\psi \circ \tau(\psi)^{-1} = \mathrm{Int}\big(a(\tau)\big)$$

holds for some element $a(\tau)$ in $G^*$.

The general classification of reductive groups over local and global fields assigns a family of invariants

$$\big\{\mathrm{inv}_v = \mathrm{inv}_v(G, \psi)\big\}$$

to $(G, \psi)$, parametrized by the valuations $v$ of $F$. The local invariant $\mathrm{inv}_v$ is attached to the localization of $(G, \psi)$ at $F_v$, and takes values in the cyclic group $(\mathbb{Z}/n\mathbb{Z})$. It can assume any value if $v$ is nonarchimedean, but satisfies the constraints $2\mathrm{inv}_v = 0$ if $F_v \cong \mathbb{R}$, and $\mathrm{inv}_v = 0$ if $F_v \cong \mathbb{C}$. The elements in the family $\{\mathrm{inv}_v\}$ vanish for almost all $v$, and satisfy the global constraint

$$\sum_v \mathrm{inv}_v = 0.$$

Conversely, given $G^*$ and any set of invariants $\{\mathrm{inv}_v\}$ in $\mathbb{Z}/n\mathbb{Z}$ with these constraints, there is an essentially unique inner twist $(G, \psi)$ of $G^*$ with the given invariants. These assertions are special cases of Theorems 1.2 and 2.2 of [**Ko5**].

(To see this, one has to identify $(\mathbb{Z}/n\mathbb{Z})$ with the group of characters on the center $\widehat{Z}_{\mathrm{sc}}$ of the complex dual group $SL(n,\mathbb{C})$ of $G^*_{ad} = PGL(n)$.) We write $n = d_v m_v$, where $d_v$ is the order of the element $\mathrm{inv}_v$ in $(\mathbb{Z}/n\mathbb{Z})$, and $n = dm$, where $d$ is the least common multiple of the integers $d_v$.

The notation $\{\mathrm{inv}_v\}$ is taken from the older theory of central simple algebras. Since the inner automorphisms $\mathrm{Int}\big(a(\tau)\big)$ of $G^*$ extend to the matrix algebra $M_n(\overline{F})$, one sees easily that $\psi$ extends to an isomorphism from $A \bigotimes \overline{F}$ to $M_n(\overline{F})$, where $A$ is a central simple algebra over $F$ such that

$$G(k) = (A \bigotimes k)^*,$$

for any $k \supset F$, and the tensor products are taken over $F$. It is a consequence of the theory of such algebras [**We**] that $A$ is isomorphic to $M_m(D)$, where $D$ is a division algebra over $F$ of degree $d$. Similarly, for any $v$, the local algebra $A_v = A \bigotimes F_v$ is isomorphic to $M_{m_v}(D_v)$. It follows that $G(F) \cong GL(m, D)$ and $G(F_v) \cong GL(m_v, D_v)$. (These facts can also be deduced from the two theorems quoted from [**Ko5**].) In particular, the minimal Levi subgroup $M_0$ of $G$ we suppose to be fixed is isomorphic to a product of $m$ copies of multiplicative groups of $D$.

It is easy to see that by replacing $\psi$ with some conjugate

$$\mathrm{Int}(g)^{-1} \circ \psi, \qquad g \in G^*(\overline{F}),$$

if necessary, we can assume that the image $M_0^* = \psi(M_0)$ is defined over $F$. The mapping

$$M \longrightarrow M^* = \psi(M), \qquad M \in \mathcal{L},$$

is then a bijection from $\mathcal{L}$ onto the set of Levi subgroups $\mathcal{L}(M_0^*)$ in $G^*$. For any $M \in \mathcal{L}$, there is a bijection $P \to P^*$ from $\mathcal{P}(M)$ to $\mathcal{P}(M^*)$. Similar remarks apply to any completion $F_v$ of $F$. We can choose a point $g_v \in M_0^*(F_v)$ such that the conjugate

$$\psi_v = \mathrm{Int}(g_v)^{-1} \circ \psi$$

maps a fixed minimal Levi subgroup $M_{v0} \subset M_0$ over $F_v$ to a Levi subgroup $M_{v0}^* \subset M_0^*$ over $F_v$. The mapping $M_v \to M_v^*$ is then a bijection from $\mathcal{L}_v = \mathcal{L}(M_{v0})$ to $\mathcal{L}(M_{v0}^*)$. In the special case that $\mathrm{inv}_v = 0$, the isomorphism $\psi_v$ from $G$ to $G^*$ is defined over $F_v$.

In order to transfer functions from $G$ to $G^*$, one has first to be able to transfer conjugacy classes. Working with a general field $k \supset F$, we start with a semisimple conjugacy class $\sigma \in \Gamma_{\mathrm{ss}}\big(G(k)\big)$ in $G(k)$. The image $\psi(\sigma)$ of $\sigma$ in $G^*$ generates a semisimple conjugacy class in $G(\overline{k})$. Since

$$\tau\big(\psi(\sigma)\big) = \tau(\psi)\big(\tau(\sigma)\big) = \mathrm{Int}\big(a(\tau)\big)^{-1}\psi(\sigma),$$

for any element $\tau \in \mathrm{Gal}(\overline{k}/k)$, the characteristic polynomial of this conjugacy class has coefficients in $k$. It follows from rational canonical form that the conjugacy class of $\psi(\sigma)$ intersects $G(k)$. It therefore determines a canonical semisimple conjugacy class $\sigma^* \in \Gamma_{\mathrm{ss}}\big(G^*(k)\big)$. We thus obtain a canonical injection $\sigma \to \sigma^*$ from $\Gamma_{\mathrm{ss}}\big(G(k)\big)$ into $\Gamma_{\mathrm{ss}}\big(G^*(k)\big)$. Now if $\sigma$ is a semisimple element in $G(k)$, it is easy to see that $G_\sigma(k)$ is isomorphic to $GL(m_\sigma, D_\sigma)$, where $D_\sigma$ is a division algebra of rank $d_\sigma$ over an extension field $k_\sigma$ of degree $\mathrm{e}_\sigma$ over $k$, with $n = d_\sigma \mathrm{e}_\sigma m_\sigma$, while $G^*_{\sigma^*}(k)$ is isomorphic to $GL(d_\sigma m_\sigma, k_\sigma)$. The unipotent classes $u$ in $G_\sigma(k)$ correspond to partitions of $m_\sigma$. For any such $u$, let $u^*$ be the unipotent class in $G^*_{\sigma^*}(k)$ that corresponds to the partition of $d_\sigma m_\sigma$ obtained by multiplying the components of the first partition by $d_\sigma$. Then

$$\gamma = \sigma u \ \longrightarrow\ \gamma^* = \sigma^* u^*$$

is a canonical injection from the set $\Gamma\big(G(k)\big)$ of all conjugacy classes in $G(k)$ into the corresponding set $\Gamma\big(G^*(k)\big)$ in $G^*(k)$.

Suppose that $k = F$. If $M \in \mathcal{L}$, $\psi$ restricts to an inner twist from $M$ to the Levi subgroup $M^*$ of $G^*$. It therefore defines an injection $\gamma_M \to \gamma_M^*$ from $\Gamma(M)$ to $\Gamma(M^*)$, by the prescription above. If $\gamma \in \Gamma(G)$ is the induced class $\gamma_M^G$, it follows immediately from the definitions that $\gamma^*$ is the induced class $(\gamma_M^*)^{G^*}$ in $\Gamma(G^*)$.

For the transfer of functions, we need to take $k$ to be a completion $F_v$ of $F$, and $\gamma$ to be a strongly regular class $\gamma_v \in \Gamma_{\mathrm{reg}}\big(G(F_v)\big)$ in $G(F_v)$. Suppose that $f_v$ is a function in $\mathcal{H}\big(G(F_v)\big)$. We define a function $f_v^*$ on $\Gamma_{\mathrm{reg}}\big(G^*(F_v)\big)$ by setting

$$f_v^*(\gamma_v^*) = \begin{cases} f_{v,G}(\gamma_v), & \text{if } \gamma_v \text{ maps to } \gamma_v^*, \\ 0, & \text{if } \gamma_v^* \text{ is not in the image of } \Gamma\big(G(F_v)\big), \end{cases}$$

for any class $\gamma_v^* \in \Gamma_{\mathrm{reg}}\big(G^*(F_v)\big)$. In the case that $\mathrm{inv}_v = 0$, $f_v^*$ is the image in $\mathcal{I}\big(G^*(F_v)\big)$ of the function $f_v \circ \psi_v^{-1}$ in $\mathcal{H}\big(G^*(F_v)\big)$. In particular, if $v$ is also nonarchimedean (so that $G$ is unramified at $v$), and $f_v$ is the characteristic function of the maximal compact subgroup $K_v$, $f_v^*$ is the image in $\mathcal{I}\big(G^*(F_v)\big)$ of the characteristic function of the maximal compact subgroup $K_v^* = \psi(K_v)$ of $G^*(F_v)$.

The next theorem applies to any valuation $v$ of $F$.

Theorem 25.1. (Deligne, Kazhdan, Vigneras)
(a) *For any $f_v \in \mathcal{H}\big(G(F_v)\big)$, the function $f_v^*$ belongs to $\mathcal{I}\big(G^*(F_v)\big)$. In other words, $f_v^*$ represents the set of strongly regular orbital integrals of some function in $\mathcal{H}\big(G^*(F_v)\big)$.*

(b) *There is a canonical injection $\pi_v \to \pi_v^*$ from $\Pi_{\mathrm{temp}}\big(G(F_v)\big)$ into $\Pi_{\mathrm{temp}}\big(G^*(F_v)\big)$ such that*

$$f_v^*(\pi_v^*) = e(G_v) f_{v,G}(\pi_v), \qquad f_v \in \mathcal{H}\big(G(F_v)\big),$$

*where $e(G_v)$ is the sign attached to the reductive group $G$ over $F_v$ by Kottwitz* [**Ko2**].

These results were established in [**DKV**]. The largely global argument makes use of a simple version of the trace formula, such as the formula provided by Corollary 23.6 for functions $f \in \mathcal{H}(G)$ that are cuspidal at two places. Part (a) is Theorem B.2.c of [**DKV**]. Part (b) follows from Theorems B.2.a, B.2.c, and B.2.d of [**DKV**]. □

The assertions of the theorem remain valid if $G$ is replaced by a Levi subgroup. This is because a Levi subgroup is itself a product of groups attached to central semisimple algebras.

Recall that the invariant trace formula depends on a choice of normalizing factors for local intertwining operators. In the case of the group $G^* = GL(n)$, Shahidi [**Sha2**] has shown that Langlands' conjectural definition of normalizing factors in terms of $L$-functions satisfies the required properties. Now if we are to be able to compare terms in the general trace formulas of $G$ and $G^*$, we will need a set of local normalizing factors for $G$ that are compatible with those of $G^*$. Suppose then that $v$ is a valuation, and that $M_v \in \mathcal{L}_v$. It is enough to define normalizing factors $r_{P|Q}(\pi_{v,\lambda})$ for tempered representations $\pi_v \in \Pi_{\mathrm{temp}}\big(M(F_v)\big)$, by [**A15**, Theorem 2.1]. We set

$$r_{P|Q}(\pi_{v,\lambda}) = r_{P^*|Q^*}(\pi^*_{v,\lambda}), \qquad \pi_v \in \Pi_{\mathrm{temp}}\big(M(F_v)\big),\ P,Q \in \mathcal{P}(M), \tag{25.1}$$

where the right hand side is Langlands' canonical normalizing factor for $GL(n)$.

LEMMA 25.2. *The functions (25.1) give valid normalizing factors for $G$.*

This is [**AC**, Lemma 2.2.1]. One has to show that the functions (25.1) satisfy the conditions of Theorem 21.4. The main point is to establish the basic identity (21.12) that relates the normalizing factors to Harish-Chandra's $\mu$-function $\mu_M(\pi_{v,\lambda})$. To establish this identity, one first deduces that

$$\mu_M(\pi_{v,\lambda}) = \mu_{M^*}(\pi^*_{v,\lambda})$$

from the formula

$$f_v(1) = e(G_v)f^*_v(1), \qquad f_v \in \mathcal{H}\big(G(F_v)\big),$$

the Plancherel formulas for $G(F_v)$ and $G^*(F_v)$, and the relationship between $\mu$-functions and corresponding Plancherel densities. The required identity for $G$ then follows from its analogue for $G^*$ established by Shahidi. $\square$

Suppose that $f$ is the restriction to $G(\mathbb{A})^1$ of a function in $\mathcal{H}\big(G(\mathbb{A})\big)$ of the form $\prod f_v$. Let $f^*$ be the corresponding restriction of the function $\prod f^*_v$. Then $f \to f^*$ extends to a linear mapping from $\mathcal{H}(G)$ to $\mathcal{I}(G^*)$. It takes any subspace $\mathcal{H}\big(G(F_S)^1\big)$ of $\mathcal{H}(G)$ to the corresponding subspace $\mathcal{I}\big(G^*(F_S)^1\big)$ of $\mathcal{I}(G)$.

We define

$$I^{\mathcal{E}}(f) = \widehat{I}^*(f^*), \qquad f \in \mathcal{H}(G^*),$$

where $I^* = I^{G^*}$ is the distribution given by either side of the invariant trace formula for $G^*$. We of course also have the corresponding distribution $I = I^G$ from the trace formula for $G$. One of the main problems is to show that $I^{\mathcal{E}}(f) = I(f)$. There seems to be no direct way to do this. One employs instead an indirect strategy of comparing terms, both geometric and spectral, in the two trace formulas.

If $S$ is a finite set of valuations of $F$ that contains $S_{\mathrm{ram}}$, and $\gamma$ belongs to $\Gamma(M)_S$, we define

$$a^{M,\mathcal{E}}(\gamma) = a^{M^*}(\gamma^*), \qquad M \in \mathcal{L}, \tag{25.2}$$

and

$$I^{\mathcal{E}}_M(\gamma, f) = \widehat{I}_{M^*}(\gamma^*, f^*), \qquad f \in \mathcal{H}(G). \tag{25.3}$$

More generally, the definition (25.3) applies to any finite set of valuations $S$ with the closure property, any conjugacy class $\gamma$ in $M(F_S)$, and any function $f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$.

LEMMA 25.3. *There is an expansion*

$$I^{\mathcal{E}}(f) = \lim_S \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma(M)_S} a^{M,\mathcal{E}}(\gamma) I^{\mathcal{E}}_M(\gamma, f). \tag{25.4}$$

This is Proposition 2.5.1 of [**AC**]. By definition,

$$\begin{aligned} I^{\mathcal{E}}(f) &= \widehat{I}^*(f^*) \\ &= \lim_S \sum_{L\in\mathcal{L}^*} |W_0^L||W_0^G|^{-1} \sum_{\beta\in\Gamma(L)_S} a^L(\beta)\widehat{I}_L(\beta, f^*), \end{aligned}$$

where $\mathcal{L}^*$ is the finite set of Levi subgroups of $G^*$ that contain the standard minimal Levi subgroup. This can in turn be written

$$\lim_S \sum_{\{L\}} |W^{G^*}(L)|^{-1} \sum_{\beta\in\Gamma(L)_S} a^L(\beta)\widehat{I}_L(\beta, f^*),$$

where $\{L\}$ is a fixed set of representations of conjugacy classes in $\mathcal{L}^*$. A global vanishing property [**A5**, Proposition 8.1] asserts that $\widehat{I}_L(\beta, f^*)$ vanishes unless the pair $(L, \beta)$ comes from $G$, in the sense that it is conjugate to the image $(M^*, \gamma^*)$ of a pair $(M, \gamma)$. We can assume in this case that our representative $L$ actually equals $M^*$. Moreover, $M^*$ is $G^*$-conjugate to another group $M_1^*$ if and only if $M$ is $G$-conjugate to $M_1$. Since $W^{G^*}(M^*) = W^G(M)$, we see that

$$\begin{aligned} I^{\mathcal{E}}(f) &= \lim_S \sum_{\{M\}} |W^G(M)|^{-1} \sum_{\gamma\in\Gamma(M)_S} a^{M^*}(\gamma^*)\widehat{I}_{M^*}(\gamma^*, f^*) \\ &= \lim_S \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma(M)_S} a^{M^*}(\gamma^*)\widehat{I}_{M^*}(\gamma^*, f^*), \end{aligned}$$

where $\{M\}$ is a fixed set of representatives of conjugacy classes in $\mathcal{L}$. This in turn equals the right hand side of (25.4). □

If we could somehow establish identities between the terms in (25.4) and their analogues in the geometric expansion of $I(f)$, we would know that $I^{\mathcal{E}}(f)$ equals $I(f)$. We could then try to compare the spectral expansions. In practice, one has to consider the two kinds of expansions simultaneously. Before we try to do this, however, we must first establish a spectral expansion of $I^{\mathcal{E}}(f)$ in terms of objects associated with $G$. The process is slightly more subtle than the geometric case just treated. This is because the local correspondence $\pi_v \to \pi_v^*$ works only for tempered representations, while nontempered representations occur on the two spectral sides.

We have been writing $\Pi\big(G(\mathbb{A})^1\big)$ for the set of irreducible representations of $G(\mathbb{A})^1$. If $\tau$ belongs to the corresponding set for $G^*$, we can write

$$f_G^*(\tau) = \sum_{\pi\in\Pi(G(\mathbb{A})^1)} \delta_G(\tau, \pi) f_G(\pi), \qquad f \in \mathcal{H}(G),$$

for uniquely determined complex numbers $\delta_G(\tau, \pi)$. This definition would be superfluous if we were concerned only with the tempered case. For if $\tau$ and $\pi$ are tempered,

$$\delta_G(\tau, \pi) = \begin{cases} 1, & \text{if } \tau = \pi^*, \\ 0, & \text{otherwise,} \end{cases}$$

since the product $\prod\limits_v e(G_v)$ of signs equals 1. If $\tau$ and $\pi$ are nontempered, however, $\delta(\tau, \pi)$ could be more complicated. This is because the decompositions of irreducible representations into standard representations for $G$ and $G^*$ might not be compatible.

If $\pi \in \Pi\big(G(\mathbb{A})^1\big)$, we define

$$a^{G,\mathcal{E}}_{\mathrm{disc}}(\pi) = \sum_{\tau \in \Pi_{\mathrm{disc}}(G^*)} a^{G^*}_{\mathrm{disc}}(\tau)\delta_G(\tau,\pi). \tag{25.5}$$

It is not hard to show that the sum may be taken over a finite set [**AC**, Lemma 2.9.1]. Using the coefficients $a^{G,\mathcal{E}}_{\mathrm{disc}}(\pi)$ in place of $a^G_{\mathrm{disc}}(\pi)$, we modify the definition of the set $\Pi_{t,\mathrm{disc}}(G)$ in §22. This gives us a discrete subset $\Pi^{\mathcal{E}}_{t,\mathrm{disc}}(G)$ of $\Pi\big(G(\mathbb{A})^1\big)$ for every $t \geq 0$. We then form the larger subset

$$\Pi^{\mathcal{E}}_t(G) = \big\{\pi^G_\lambda :\ M \in \mathcal{L},\ \pi \in \Pi^{\mathcal{E}}_{t,\mathrm{disc}}(M),\ \lambda \in i\mathfrak{a}^*_M/i\mathfrak{a}^*_G\big\}$$

of $\Pi\big(G(\mathbb{A})^1\big)$, equipped with a measure $d\pi^G_\lambda$ defined as in (22.7). Finally, we define a function $a^{G,\mathcal{E}}$ on $\Pi^{\mathcal{E}}_t(G)$ by setting

$$a^{G,\mathcal{E}}(\pi^G_\lambda) = a^{M,\mathcal{E}}_{\mathrm{disc}}(\pi) r^G_M(\pi_\lambda), \tag{25.6}$$

as in (22.8). The ultimate aim, in some sense, is to show that the discrete coefficients $a^{G,\mathcal{E}}_{\mathrm{disc}}(\pi)$ and $a^G_{\mathrm{disc}}(\pi)$ match. We now assume inductively that this is true if $G$ is replaced by any proper Levi subgroup $M$. Then $\Pi^{\mathcal{E}}_{t,\mathrm{disc}}(M)$ equals $\Pi_{t,\mathrm{disc}}(M)$, and in particular, consists of unitary representations of $M(\mathbb{A})^1$. It follows that the function $a^{G,\mathcal{E}}(\pi^G_\lambda)$ is analytic, and slowly increasing in the sense of Lemma 21.5.

The extra complication arises when we try to describe the function $a^{G,\mathcal{E}}$ as a pullback of the corresponding function for $G^*$. Suppose that $\pi \in \Pi\big(M(\mathbb{A})^1\big)$ and $\tau \in \Pi\big(M^*(\mathbb{A})^1\big)$ are representations with $\delta_M(\tau,\pi) \neq 0$. Given a point $\lambda \in i\mathfrak{a}^*_M/i\mathfrak{a}^*_G$ in general position, and groups $P, Q \in \mathcal{P}(M)$, we set

$$r_{Q|P}(\tau_\lambda,\pi_\lambda) = r_{Q^*|P^*}(\tau_{S,\lambda})^{-1} r_{Q|P}(\pi_{S,\lambda}),$$

where $S \supset S_{\mathrm{ram}}$ is a large finite set of valuations, and $\tau_S$ and $\pi_S$ are the $S$-components of $\tau$ and $\pi$. The condition that $\delta_M(\tau,\pi) \neq 0$ implies that $\tau_v \cong \pi_v$ for almost all $v$ [**AC**, Corollary 2.8.3], so that $r_{Q|P}(\tau_\lambda,\pi_\lambda)$ is independent of the choice of $S$. Moreover, $r_{Q|P}(\tau_\lambda,\pi_\lambda)$ is a rational function in the relevant variables $\lambda(\alpha^\vee)$ or $q_v^{-\lambda(\alpha^\vee)}$ attached to valuations $v$ in $S$ [**A15**, Proposition 5.2]. As $Q$ varies, we obtain a $(G,M)$-family of functions

$$r_Q(\Lambda,\tau_\lambda,\pi_\lambda,P) = \delta_M(\tau,\pi) r_{Q|P}(\tau_{\lambda+\Lambda},\pi_{\lambda+\Lambda}) r_{Q|P}(\tau_\lambda,\pi_\lambda)^{-1}$$

of $\Lambda \in i\mathfrak{a}^*_M$, which we define for *any* $\tau$ and $\pi$.

Assume now that $\pi$ belongs to $\Pi^{\mathcal{E}}_{t,\mathrm{disc}}(M)$. For any representation $\tau \in \Pi\big(M^*(\mathbb{A})^1\big)$, the $(G,M)$-family of global normalizing factors

$$\delta_M(\tau,\pi) r_Q(\Lambda,\pi_\lambda,P), \qquad Q \in \mathcal{P}(M),$$

is defined, and equals the product of $(G,M)$-families

$$r_{Q^*}(\Lambda,\tau_\lambda,P^*) r_Q(\Lambda,\tau_\lambda,\pi_\lambda,P), \qquad Q \in \mathcal{P}(M).$$

It follows from the product formula (17.12) that

$$\delta_M(\tau,\pi) r^G_M(\pi_\lambda) = \sum_{L \in \mathcal{L}(M)} r^{L^*}_{M^*}(\tau_\lambda) r^G_L(\tau_\lambda,\pi_\lambda).$$

Multiplying each side of this last identity by $a^{M^*}_{\mathrm{disc}}(\tau)$, and then summing over $\tau$, we obtain an identity

$$a^{G,\mathcal{E}}_M(\pi^G_\lambda) = \sum_{\tau \in \Pi_{t,\mathrm{disc}}(M^*)} \sum_{L \in \mathcal{L}(M)} a^{L^*}_{M^*}(\tau^{L^*}_\lambda) r^G_L(\tau_\lambda,\pi_\lambda). \tag{25.7}$$

The description of the coefficient $a_M^{G,\mathcal{E}}(\pi_\lambda^G)$ as a pullback of coefficients from $G^*$ is thus more elaborate than its geometric counterpart. This has to be reflected in the construction of the corresponding linear forms that occur in the spectral expansion of $I^{\mathcal{E}}(f)$. Suppose that $S$ is any finite set of valuations with the closure property. The function $r_M^L(\tau_\lambda, \pi_\lambda)$ can obviously be defined for representations $\tau \in \Pi\big(M^*(F_S)\big)$ and $\pi \in \Pi\big(M(F_S)\big)$. If either $\tau$ or $\pi$ is in general position, $r_M^L(\tau_\lambda, \pi_\lambda)$ is an analytic function of $\lambda$ in $i\mathfrak{a}_{M,S}^*/i\mathfrak{a}_{L,S}^*$. In this case, we define linear forms

$$I_M^{\mathcal{E}}(\pi, X, f), \qquad\qquad X \in \mathfrak{a}_{M,S},\ f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big),$$

inductively by setting

(25.8)

$$\widehat{I}_{M^*}(\tau, X, f^*) = \sum_{L \in \mathcal{L}(M)} \sum_{\pi \in \Pi(M(F_S))} \int_{i\mathfrak{a}_{M,S}^*/i\mathfrak{a}_{L,S}^*} r_M^L(\tau_\lambda, \pi_\lambda) I_L^{\mathcal{E}}(\pi_\lambda^L, X_L, f) \mathrm{e}^{-\lambda(X)} \mathrm{d}\lambda,$$

for any $\tau$. (For arbitrary $\tau$ and $\pi$, the functions $r_M^L(\tau_\lambda, \pi_\lambda)$ can acquire poles in the domain of integration, and one has to take a linear combination of integrals over contours $\varepsilon_P + i\mathfrak{a}_{M,S}^*/i\mathfrak{a}_{L,S}^*$. See [**AC**, pp. 124–126]. The general definition in [**AC**] avoids induction, but is a three stage process that is based on standard representations.) It is of course the summands with $L \neq M$ in (25.8) that we assume inductively to be defined. The summand of $L = M$ equals

$$\sum_{\pi \in \Pi(M(F_S))} \delta_M(\tau, \pi) I_M^{\mathcal{E}}(\pi, X, f).$$

By applying the local vanishing property [**AC**, Proposition 2.10.3] to the left hand side of the relation (25.8), one shows without difficulty that $I_M^{\mathcal{E}}(\pi, X, f)$ is well defined by this relation. We extend the definition to adelic representations $\pi \in \Pi\big(M(\mathbb{A})^1\big)$ and functions $f \in \mathcal{H}(G)$ by taking $S \supset S_{\mathrm{ram}}$ to be large. If in addition, $\pi$ is unitary, we write

$$I_M^{\mathcal{E}}(\pi, f) = I_M^{\mathcal{E}}(\pi, 0, f),$$

as before.

Lemma 25.4. *There is an expansion*

$$I^{\mathcal{E}}(f) = \lim_T \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Pi^{\mathcal{E}}(M)^T} a^{M,\mathcal{E}}(\pi) I_M^{\mathcal{E}}(\pi, f) \mathrm{d}\pi. \tag{25.9}$$

This is Proposition 2.12.2 of [**AC**]. The inductive definition (25.8) we have given here leads to a two step proof. The first step is a duplication of the proof of Lemma 25.3, while the second is an application of the formulas (25.7) and (25.8).

We begin by writing

$$\begin{aligned} I^{\mathcal{E}}(f) &= \widehat{I}^*(f^*) \\ &= \lim_T \sum_{L \in \mathcal{L}^*} |W_0^L||W_0^{G^*}|^{-1} \int_{\Pi(L)^T} a^L(\tau) \widehat{I}_L(\tau, f^*) \mathrm{d}\tau, \end{aligned}$$

by the spectral expansion (23.12) for $G^*$. The global vanishing property [**A14**, Proposition 8.2] asserts that $\widehat{I}_L(\tau, f^*)$ vanishes unless $L$ is conjugate to the image of a group $M$ in $\mathcal{L}$. Using the elementary counting argument from the proof of Lemma 25.3, we see that

$$I^{\mathcal{E}}(f) = \lim_{T} \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Pi(M^*)^T} a^{M^*}(\tau)\widehat{I}_{M^*}(\tau, f^*)\mathrm{d}\tau.$$

For the second step, we have to substitute the formula (25.8), with $S \supset S_{\mathrm{ram}}$ large and $X = 0$, for $\widehat{I}_{M^*}(\tau, f^*)$. More correctly, we substitute the version of (25.8) that is valid if any of the functions $r_M^L(\tau_\lambda, \pi_\lambda)$ have poles, since we do not know a priori that the representations $\pi$ over which we sum are unitary. We then substitute the explicit form (22.7) of the measure $d\tau$ on $\Pi(M^*)^T$. In the resulting multiple (seven-fold, as a matter of fact) sum-integral, it is not difficult to recognize the expansion (25.7). There is some minor effort involved in keeping track of the various constants and domains of the integration. This accounts for the length of some of the arguments in [**AC**]. In the end, however, the expression collapses to the required expansion (25.9). $\square$

THEOREM 25.5. *If $\gamma$ belongs to $\Gamma(M)_S$ for some $S \supset S_{\mathrm{ram}}$, then*

$$I_M^{\mathcal{E}}(\gamma, f) = I_M(\gamma, f) \tag{25.10}$$

*and*

$$a^{M,\mathcal{E}}(\gamma) = a^M(\gamma). \tag{25.11}$$

THEOREM 25.6. *If $\pi$ belongs to the union of $\Pi(M)^T$ and $\Pi^{\mathcal{E}}(M)^T$, for some $T > 0$, then*

$$I_M^{\mathcal{E}}(\pi, f) = I_M(\pi, f) \tag{25.12}$$

*and*

$$a^{M,\mathcal{E}}(\pi) = a^M(\pi). \tag{25.13}$$

Theorems 25.5 and 25.6 correspond to Theorems A and B in Sections 2.5 and 2.9 of [**AC**], which are the main results of Chapter 2 of [**AC**]. They are proved together, by an argument that despite its length sometimes seems to move forward of its own momentum. In following our sketch of the proof, the reader might keep in mind the earlier argument used in §21 to establish that the terms in the invariant trace formula are supported on characters.

The combined proof of the two theorems is by double induction on $n$ and $\dim(A_M)$. The first induction hypothesis immediately implies that the global formulas (25.11) and (25.13) are valid for proper Levi subgroups $M \neq G$. If $M = G$, on the other hand, the local formulas (25.10) and (25.12) hold by definition, the two sides in each case being equal to $f_G(\gamma)$ and $f_G(\pi)$ respectively. We apply these observations to the identity obtained from the right hand sides of (25.4) and (25.9). Combining the resulting formula with the invariant trace formula for $G$, we see that the limit over $S$ of the sum of

$$\sum_{M \neq G} |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma)\big(I_M^{\mathcal{E}}(\gamma, f) - I_M(\gamma, f)\big) \tag{25.14}$$

and

$$\sum_{\gamma \in \Gamma(G)_S} \big(a^{G,\mathcal{E}}(\gamma) - a^G(\gamma)\big) f_G(\gamma) \tag{25.15}$$

equals the limit over $T$ of the sum of

$$\sum_{M\neq G} |W_0^M||W_0^G|^{-1} \int_{\Pi(M)^T} a^M(\pi)\big(I_M^{\mathcal{E}}(\pi,f) - I_M(\pi,f)\big)\mathrm{d}\pi \tag{25.16}$$

and

$$\int_{\Pi^*(G)^T} \big(a^{G,\mathcal{E}}(\pi) - a^G(\pi)\big) f_G(\pi)\mathrm{d}\pi, \tag{25.17}$$

where $\Pi^*(G)^T$ is the union of $\Pi^{\mathcal{E}}(G)^T$ with $\Pi(G)^T$.

The linear forms $I_M(\gamma,f)$ and $I_M(\pi,X,f)$ were defined for any finite set $S$ with the closure property, and any $f \in \mathcal{H}_{\mathrm{ac}}\big(G(F_S)\big)$. They each satisfy splitting and descent formulas. The linear forms $I_M^{\mathcal{E}}(\gamma,f)$ and $I_M^{\mathcal{E}}(\pi,X,f)$ have been defined in the same context, and satisfy parallel splitting and descent formulas. The required local identities (25.10) and (25.12) can be broadened to formulas

$$I_M^{\mathcal{E}}(\gamma,f) = I_M(\gamma,f), \qquad \gamma \in \Gamma\big(M(F_S)\big), \tag{25.18}$$

and

$$I_M^{\mathcal{E}}(\pi,X,f) = I_M(\pi,X,f), \qquad \pi \in \Pi\big(M(F_S)\big),\ X \in \mathfrak{a}_{M,S}, \tag{25.19}$$

which we postulate for any $f \in \mathcal{H}\big(G(F_S)\big)$. These general identities were originally established only up to some undetermined constants [**AC**, Theorem 2.6.1], but they were later resolved by the local trace formula [**A18**, Theorem 3.C]. We assume inductively that (25.18) and (25.19) hold if $n$ is replaced by a smaller integer. This allows us to simplify the local terms in (25.14) and (25.16). In so doing, we can assume that the function $f \in \mathcal{H}(G)$ is the restriction to $G(\mathbb{A})^1$ of a product of $\prod f_v$.

Consider first the expression (25.16). We recall that Proposition 23.5 applies to the linear forms $I_M(\gamma,f)$ and $I_M(\pi,X,f)$. This proposition can also be adapted to the linear forms $I_M^{\mathcal{E}}(\gamma,f)$ and $I_M^{\mathcal{E}}(\pi,X,f)$ [**AC**, §2.8]. Its first assertion implies that either of the two spectral linear forms can be expressed in terms of its geometric counterpart. The analogue of the more specific second assertion of Proposition 23.5 can be formulated to say that if (25.18) holds for all $M$, $S$, $\gamma$ and $f$, then so does (25.19) [**AC**, Theorem 2.10.2]. We combine this with the splitting and descent formulas satisfied by the terms in the brackets in (25.16). As in §23, the fact that the representations $\pi \in \Pi(M)$ are unitary is critical to the success of the argument. Following the corresponding discussion after Proposition 23.5, one deduces that the required local identity (25.12) is valid. The expression (25.14) therefore vanishes.

Now consider the expression (25.14). It follows from the splitting formulas (23.8) and [**AC**, $(2.3.4)^{\mathcal{E}}$], together with our induction hypotheses, that

$$I_M^{\mathcal{E}}(\gamma,f) - I_M(\gamma,f) = \sum_v \varepsilon_M(f_v,\gamma_v) f^v(\gamma^v),$$

where

$$\varepsilon_M(f_v,\gamma_v) = I_M^{\mathcal{E}}(\gamma_v,f_v) - I_M(\gamma_v,f_v),$$

and

$$f^v(\gamma^v) = \prod_{w\neq v} f_w(\gamma_w).$$

If $v$ does not belong to the set $S_{\mathrm{ram}}$, $\varepsilon_M(f_v,\gamma_v)=0$, since $G$ and $G^*$ are isomorphic over $F_v$. The expression (25.14) therefore reduces to

$$\sum_{M\neq G}|W_0^M||W_0^G|^{-1}\sum_{\gamma\in\Gamma(M)_S}a^M(\gamma)\Big(\sum_{v\in S_{\mathrm{ram}}}\varepsilon_M(f_v,\gamma_v)f_G^v(\gamma^v)\Big). \tag{25.20}$$

The remaining global coefficients can also be simplified. Consider a class $\gamma\in\Gamma(G)_S$ in (25.15) whose semisimple part is represented by a noncentral element $\sigma\in G(F)$. Then $G_\sigma$ is a proper subgroup of $G$. It follows from the definitions (19.6), (22.2), and (25.2), together with our induction hypothesis, that $a^{G,\mathcal{E}}(\gamma)$ equals $a^G(\gamma)$. The expression (25.15) therefore reduces to

$$\sum_{z\in A_G(F)}\sum_{u\in\Gamma_{\mathrm{unip}}(G)_S}\big(a^{G,\mathcal{E}}(zu)-a^G(zu)\big)f_G(zu), \tag{25.21}$$

where $\Gamma_{\mathrm{unip}}(G)_S=\big(\mathcal{U}_G(F)\big)_{G,S}$ is the set of unipotent classes in $\Gamma(G)_S$.

Consider a representation $\pi\in\Pi^*(G)^T$ in (25.17) that does not lie in the union $\Pi^*_{t,\mathrm{disc}}(G)$ of $\Pi^{\mathcal{E}}_{t,\mathrm{disc}}(G)$ and $\Pi_{t,\mathrm{disc}}(G)$, for any $t$. The induction hypothesis we have taken on includes the earlier assumption that the coefficients $a^{M,\mathcal{E}}_{\mathrm{disc}}$ and $a^M_{\mathrm{disc}}$ are equal, for any $M\neq G$. It follows from the definitions (22.8) and (25.6) that $a^{G,\mathcal{E}}(\pi)$ equals $a^G(\pi)$. The expression (25.17) therefore reduces to

$$\sum_{t\leq T}\sum_{\pi\in\Pi^*_{t,\mathrm{disc}}(G)}\big(a^{G,\mathcal{E}}_{\mathrm{disc}}(\pi)-a^G_{\mathrm{disc}}(\pi)\big)f_G(\pi).$$

We conclude that the limit in $T$ of the sum of (25.16) and (25.17) equals

$$\sum_{t}\sum_{\pi\in\Pi^*_{t,\mathrm{disc}}(G)}\big(a^{G,\mathcal{E}}_{\mathrm{disc}}(\pi)-a^G_{\mathrm{disc}}(\pi)\big)f_G(\pi). \tag{25.22}$$

This expression is conditionally convergent, in the sense that the iterated sums converge absolutely.

Using the induction hypothesis, we have reduced the original four expressions to (25.20), (25.21), and (25.22). It follows that if $S\supset S_{\mathrm{ram}}$ is large, in a sense that depends only on the support of $f$, the sum of (25.20) and (25.21) equals (25.22). The rest of the proof is harder. It consists of several quite substantial steps, each of which we shall attempt to sketch in a few words.

The first step concerns the summands in (25.20). The problem at this stage is to establish something weaker than the required vanishing of these summands. It is to show that for any $M\in\mathcal{L}$ and $v\in S_{\mathrm{ram}}$, and for certain $f_v\in\mathcal{H}\big(G(F_v)\big)$, the function

$$\varepsilon_M(f_v):\ \gamma_v\ \longrightarrow\ \varepsilon_M(f_v,\gamma_v)=I^{\mathcal{E}}_M(\gamma_v,f_v)-I_M(\gamma_v,f_v),\qquad \gamma_v\in\Gamma_{\mathrm{reg}}\big(M(F_v)\big),$$

belongs to $\mathcal{I}_{\mathrm{ac}}\big(M(F_v)\big)$. The functions $I^{\mathcal{E}}_M(\gamma_v,f_v)$ and $I_M(\gamma_v,f_v)$ are smooth on the strongly $G$-regular set in $M(F_v)$, but as $\gamma_v$ approaches the boundary, they acquire singularities over and above those attached to invariant orbital integrals on $M(F_v)$. The problem is to show that these supplementary singularities cancel.

If $v$ is nonarchimedean, let $\mathcal{H}\big(G(F_v)\big)^0$ be the subspace of functions $f_v\in\mathcal{H}\big(G(F_v)\big)$ such that for every central element $z_v\in A_G(F_v)$ and every nontrivial unipotent element $u_v\neq 1$ in $G(F_v)$, $f_G(z_vu_v)$ vanishes. If $v$ is archimedean, we set $\mathcal{H}\big(G(F_v)\big)^0$ equal to $\mathcal{H}\big(G(F_v)\big)$. The result is that the correspondence

$$f_v \longrightarrow \varepsilon_M(f_v), \qquad\qquad f_v \in \mathcal{H}\big(G(F_v)\big)^0,$$

is a continuous linear mapping from $\mathcal{H}\big(G(F_v)\big)^0$ to $\mathcal{I}_{\mathrm{ac}}\big(M(F_v)\big)$. If $v$ is archimedean, one establishes the result by combining the induction hypothesis with the differential equations and boundary conditions [**A12**, §11–13] satisfied by weighted orbital integrals . If $v$ is nonarchimedean, one combines the induction hypotheses with the germ expansion [**A12**, §9] of weighted orbital integrals about a singular point. In this case, one has also to make use of the explicit formulas for weighted orbital integrals of supercuspidal matrix coefficients, in order to match the germs corresponding to $u_v = 1$. (See [**AC**, Proposition 2.13.2].) Let $\mathcal{H}\big(G(\mathbb{A})^0\big)$ be the subspace of $\mathcal{H}\big(G(\mathbb{A})\big)$ spanned by products $f = \prod f_v$ such that for every $v \in S_{\mathrm{ram}}$, $f_v$ belongs to $\mathcal{H}\big(G(F_v)\big)^0$. The result above then implies that the correspondence

$$f \longrightarrow \varepsilon_M(f) = \sum_{v \in S_{\mathrm{ram}}} \varepsilon_M(f_v) f_G^v$$

is a continuous linear mapping from $\mathcal{H}\big(G(\mathbb{A})\big)^0$ to $\mathcal{I}_{\mathrm{ac}}\big(M(\mathbb{A})\big)$.

Suppose now that $M \in \mathcal{L}$ is fixed. We formally introduce the second induction hypothesis that the analogue of (25.18), for any $L \in \mathcal{L}$ with $\dim(A_L) < \dim(A_M)$, holds for any $S$. We define $\mathcal{H}\big(G(\mathbb{A}), M\big)$ to be the space of functions $f$ in $\mathcal{H}\big(G(\mathbb{A})\big)$ that are $M$-*cuspidal* at two nonarchimedean places $v$, in the sense that the local functions $f_{v,L}$ vanish unless $L$ contains a conjugate of $M$. We also define $\mathcal{H}\big(G(\mathbb{A}), M\big)^0$ to be the space of functions $f$ in the intersection

$$\mathcal{H}\big(G(\mathbb{A}), M\big) \cap \mathcal{H}\big(G(\mathbb{A})\big)^0$$

that satisfy one additional condition. We ask that $f$ vanish at any element in $G(\mathbb{A})$ whose component at each finite place $v$ belongs to $A_G(F_v)$. In combination with the definition of $\mathcal{H}\big(G(\mathbb{A})\big)^0$, this last condition is designed to insure that the terms $f_G(zu)$ in (25.21) all vanish. Notice that $f$ may be modified at any archimedean place without affecting the condition that it lie in $\mathcal{H}\big(G(\mathbb{A}), M\big)^0$.

Suppose that $f$ belongs to $\mathcal{H}\big(G(\mathbb{A}), M\big)^0$. The last induction hypothesis then implies that the summand in (25.20) corresponding to any Levi subgroup that is not conjugate to our fixed group $M$ vanishes. The expression (25.20) reduces to

$$|W(M)|^{-1} \sum_{\gamma \in \Gamma(M)_S} a^M(\gamma)\varepsilon_M(f,\gamma). \tag{25.23}$$

It is an easy consequence of the original induction hypothesis and the splitting formulas that the function $\varepsilon_M(f)$ in $\mathcal{I}_{\mathrm{ac}}\big(M(\mathbb{A})\big)$ is cuspidal at two places. It then follows from the simple form of the geometric expansion for $M$ in Corollary 23.6 that the original expansion (25.20) equals the product of $|W(M)|^{-1}$ with $\widehat{I}^M\big(\varepsilon_M(f)\big)$. The conditions on $f$ imply that the second expression (25.21) vanishes. Recall that the third expression (25.22) was the ultimate reduction of the spectral expansion of $I^{\mathcal{E}}(f) - I(f)$. Since the third expression equals the sum of the first two, we can write

$$\sum_t \big(I_t^{\mathcal{E}}(f) - I_t(f)\big) - |W(M)|^{-1} \sum_t \widehat{I}_t^M\big(\varepsilon_M(f)\big) = 0, \tag{25.24}$$

in the notation of Remark 10 in §23. (See [**AC**, (2.15.1)].)

The second step is to apply the weak multiplier estimate (23.13) to the sums in (25.24). Suppose that $f \in \mathcal{H}\big(G(\mathbb{A}), M\big)^0$ is fixed. If $\alpha \in \mathcal{E}(\mathfrak{h}^1)^W$ is any multiplier, $f_\alpha$ also belongs to $\mathcal{H}\big(G(\mathbb{A}), M\big)^0$, and the identity (25.24) remains valid with $f_\alpha$ in place of $\alpha$. It is a consequence of the definitions that $I_t^{\mathcal{E}}(f_\alpha) = \widehat{I}_t(f_\alpha^*)$. One shows also that $\varepsilon_M(f_\alpha) = \varepsilon_M(f)_\alpha$ [**AC**, Corollary 2.14.4]. It then follows from (23.13) that there are positive constants $C$, $k$ and $r$ such that for any $T > 0$, any $N \geq 0$, and any $\alpha$ in the subspace $C_N^\infty(\mathfrak{h}^1)^W$ of $\mathcal{E}(\mathfrak{h}^1)^W$, the sum

$$\Big| \sum_{t \leq T} \big(I_t^{\mathcal{E}}(f_\alpha) - I_t(f_\alpha)\big) - |W(M)|^{-1} \sum_{t \leq T} \widehat{I}_t^M\big(\varepsilon_M(f_\alpha)\big)\Big| \tag{25.25}$$

is bounded by

$$C\mathrm{e}^{kN} \sup_{\nu \in \mathfrak{h}_u^*(r,T)} \big(|\widehat{\alpha}(\nu)|\big). \tag{25.26}$$

To exploit the last inequality, one fixes a point $\nu_1$ in $\mathfrak{h}_u^*$. Enlarging $r$ if necessary, we can assume that $\nu_1$ lies in the space $\mathfrak{h}_u^*(r) = \mathfrak{h}_u^*(r, 0)$. It is then possible to choose a function $\alpha_1 \in C_c^\infty(\mathfrak{h}^1)^W$ such that $\widehat{\alpha}_1$ maps $\mathfrak{h}_u^*(r)$ to the unit interval, and such that the inverse image of 1 under $\widehat{\alpha}_1$ is the $W$-orbit $W(\nu_1)$ of $\nu_1$ [**AC**, Lemma 2.15.2]. If $\alpha_1$ belongs to $C_{N_1}^\infty(\mathfrak{h}^1)^W$, and $r$ and $k$ are as in (25.26), we chose $T > 0$ so that

$$|\widehat{\alpha}_1(\nu)| \leq \mathrm{e}^{-2kN_1}$$

for all $\nu \in \mathfrak{h}_u^*(r, T)$. We then apply the inequality, with $\alpha$ equal to the function $\alpha_m$ obtained by convolving $\alpha_1$ with itself $m$ times. Since $\widehat{\alpha}_m(\nu)$ equals $\widehat{\alpha}_1(\nu)^m$, the expression (25.26) approaches 0 as $m$ approaches infinity. One shows independently that the second sum in (25.25) also approaches 0 as $m$ approaches infinity [**AC**, p. 183–188]. Therefore, the first sum in (25.25) approaches 0 as $m$ approaches infinity. But this first sum equals the double sum

$$\sum_{t \leq T} \sum_{\pi \in \Pi_{t,\mathrm{disc}}^*(G)} \big(a_{\mathrm{disc}}^{G,\mathcal{E}}(\pi) - a_{\mathrm{disc}}^G(\pi)\big) f_G(\pi) \alpha_1(\nu_\pi)^m,$$

which can be taken over a finite set that is independent of $m$. We can assume that $T \geq \|\mathrm{Im}(\nu_1)\|$. It follows that the double sum approaches

$$\sum_{\pi \in \Pi_{\nu_1,\mathrm{disc}}^*(G)} \big(a_{\mathrm{disc}}^{G,\mathcal{E}}(\pi) - a_{\mathrm{disc}}^G(\pi)\big) f_G(\pi)$$

as $m$ approaches infinity, where $\Pi_{\nu_1,\mathrm{disc}}^*(G)$ is the set of representations $\pi$ in the set $\Pi_{\mathrm{disc}}^*(G)$ with $\nu_\pi = \nu_1$. Summing over the infinitesimal characters $\nu_1$ with $\|\mathrm{Im}(\nu_1)\| = t$, we conclude that

$$\sum_{\pi \in \Pi_{t,\mathrm{disc}}^*(G)} \big(a_{\mathrm{disc}}^{G,\mathcal{E}}(\pi) - a_{\mathrm{disc}}^G(\pi)\big) f_G(\pi) = 0, \tag{25.27}$$

for any $t \geq 0$.

The identity (25.27) holds for any function $f$ in $\mathcal{H}\big(G(\mathbb{A}), M\big)^0$. The third step is to show that it extends to any $f$ in the larger space $\mathcal{H}\big(G(\mathbb{A}), M\big)$. This is a fairly standard argument. On the one hand, the left hand side of (25.27) is a linear combination of point measures in the spectral variables of $f_G$. On the other hand, the linear forms whose kernels define the subspace $\mathcal{H}\big(G(\mathbb{A}), M\big)^0$ of $\mathcal{H}\big(G(\mathbb{A}), M\big)$ are easily seen to be continuous in the spectral variables. Playing one against the other, one sees that (25.27) does indeed remain valid for any $f$ in $\mathcal{H}\big(G(\mathbb{A}), M\big)$. (See [**AC**, §2.16].) In particular, (25.22) vanishes for any such $f$. Since (25.20) equals (25.23), we deduce that the sum of (25.23) and (25.21) vanishes for any function $f$ in $\mathcal{H}\big(G(\mathbb{A}), M\big)$.

The fourth step is to apply what we have just established to the expression (25.23). Suppose that for each $v \in S_{\mathrm{ram}}$, $f_v$ is a given function in $\mathcal{H}\big(G(F_v)\big)$. Suppose also that $\gamma_1$ is a fixed $G$-regular element in $M(F)$ that is $M$-elliptic at two unramified places $w_1$ and $w_2$. At the places $w \notin S_{\mathrm{ram}}$, we choose functions $f_w \in \mathcal{H}\big(G(F_w)\big)$ so that $f_{w,G}(\gamma_1) = 1$, and so that the product $f = \prod f_v$ lies in $\mathcal{H}\big(G(\mathbb{A})\big)$. We fix $f_w$ for $w$ distinct from $w_1$ and $w_2$, but for $w$ equal to $w_1$ or $w_2$, we allow the support of $f_w$ to shrink around a small neighbourhood of $\gamma_1$ in $G(F_w)$. Then $f$ belongs to $\mathcal{H}\big(G(\mathbb{A}), M\big)$. Since the support of $f$ remains within a fixed compact set, we can take $S$ to be some fixed finite set containing $S_{\mathrm{ram}}$, $w_1$, and $w_2$. We can also restrict the sum in (25.23) to a finite set that is independent of $f$. (See Remark 9 in §23.)

Since we are shrinking $f_{w_1}$ and $f_{w_2}$ around $\gamma_1$, the terms $f_G(zu)$ in (25.21) all vanish. In addition, the function

$$\varepsilon_M(f,\gamma) = \sum_{v \in S_{\mathrm{ram}}} \varepsilon_M(f_v,\gamma) f_M^v(\gamma)$$

in (25.23) is supported on the subset $\Gamma_{G\text{-reg}}(M)$ of $G$-regular classes in $\Gamma(M)_S$. It is in fact supported on classes $\gamma$ that are $G(F_{w_i})$-conjugate to $\gamma_1$. For the group $G$ at hand, any such class is actually $G(F)$-conjugate to $\gamma_1$, and hence equal to $w_s^{-1}\gamma_1 w_s$, for some $s \in W(M)$. But

$$\varepsilon_M(f_v, w_s^{-1}\gamma_1 w_s) f_M^v(w_s^{-1}\gamma_1 w_s) = \varepsilon_M(f_v,\gamma_1) f_M^v(\gamma_1).$$

Moreover, since $\gamma_1$ is $F$-elliptic in $M$, the coefficients

$$a^M(w_s^{-1}\gamma_1 w_s) = a^M(\gamma_1) = \mathrm{vol}\big(M_{\gamma_1}(F)\backslash M_{\gamma_1}(\mathbb{A})^1\big)$$

are all positive. The vanishing of the sum of (25.23) and (25.21) thus reduces to the identity

$$\varepsilon_M(f,\gamma_1) = \sum_{v \in S_{\mathrm{ram}}} \varepsilon_M(f_v,\gamma_1) f_M^v(\gamma_1) = 0.$$

This holds for any choice of functions $f_v \in \mathcal{H}\big(G(F_v)\big)$ at the places $v \in S_{\mathrm{ram}}$.

Consider a fixed valuation $v \in S_{\mathrm{ram}}$. It follows from what we have just established that if $f_{v,G}(\gamma_1) = 0$, then $\varepsilon_M(f_v,\gamma_1) = 0$. This in turn implies that if $f_v$ is arbitrary, then

$$\varepsilon_M(f_v,\gamma_1) = \varepsilon_v(\gamma_1) f_{v,M}(\gamma_1),$$

for a complex number $\varepsilon_v(\gamma_1)$ depending on the chosen element $\gamma_1 \in M(F)$. Now, it is known that $G(F)$ is dense in $G(F_S)$, for any finite set $S \supset S_{\mathrm{ram}}$. Letting the $G$-regular point $\gamma_1 \in M(F)$ vary, we see that

$$\varepsilon_M(f_v,\gamma_v) = \varepsilon_v(\gamma_v) f_{v,M}(\gamma_v), \qquad \gamma_v \in \Gamma_{G\text{-reg}}\big(M(F_v)\big),\ f_v \in \mathcal{H}\big(G(F_v)\big),$$

for a function $\varepsilon_M$ on $\Gamma_{G\text{-reg}}\big(M(F_v)\big)$ that is smooth.

The last identity is a watershed. It represents a critical global contribution to a local problem. It is also the input for one of the elementary applications of the local trace formula in the article [**A18**]. The result in question is Theorem 3C of

[**A18**], which asserts that the function $\varepsilon_M(\gamma_v)$ actually vanishes. We can therefore conclude that the linear form

$$\varepsilon_M(f_v,\gamma_v) = I_M^{\mathcal{E}}(\gamma_v, f_v) - I_M(\gamma_v, f_v), \qquad f \in \mathcal{H}\big(G(F_v)\big),$$

vanishes for any $G$-regular class $\gamma_v$ in $M(F_v)$. It is then not hard to see from the definitions (18.3), (18.12), (23.3) and (25.3) that the linear form vanishes for any element $\gamma_v \in M(F_v)$ at all.

The fourth step we have just sketched completes the induction argument on $M$. Indeed, the general identity (25.18) follows for any $S$ from the splitting formula (23.8), and the case $S = \{v\}$ just established. In particular, the required identity (25.10) is valid for any $M$. We have already noted that (25.18) implies the companion identity (25.19). In particular, both required local identities (25.10) and (25.12) of the two theorems are valid for any $M$.

The last step is to extract what remains of the required global identities (25.11) and (25.13) from the properties of the expressions (25.21) and (25.23) we have found. Since we have completed the induction argument on $M$, and since $\mathcal{H}\big(G(\mathbb{A}), M_0\big)$ equals $\mathcal{H}\big(G(\mathbb{A})\big)$ by definition, the identity (25.27) holds for any function $f \in \mathcal{H}\big(G(\mathbb{A})\big)$. The sum in (25.27) can be taken over a finite set that depends only on a choice of open compact subgroup $K_0 \subset G(\mathbb{A}_{\mathrm{fin}})$ under which $f$ is bi-invariant. It is then not hard to show that the coefficients

$$a_{\mathrm{disc}}^{G,\mathcal{E}}(\pi) - a_{\mathrm{disc}}^{G}(\pi), \qquad \pi \in \Pi_{t,\mathrm{disc}}^*(G),$$

in (25.27) vanish. This completes the proof of (25.13). Since (25.27) vanishes for any $f$, so does the expression (25.22). We have already established that (25.20) vanishes. It follows that the remaining expression (25.21) vanishes for any $f \in \mathcal{H}(G)$. By varying $f$, one deduces that the coefficients

$$a^{G,\mathcal{E}}(zu) - a^G(zu), \qquad z \in A_G(F),\ u \in \Gamma_{\mathrm{unip}}(G)_S,$$

in (25.21) vanish. This completes the proof of (25.11). It also finishes the original induction argument on $n$. (See [**AC**, §2.16] and [**A18**, §2–3].) □

For global applications, the most important assertion of the two theorems is the identity (25.13) of global coefficients. It implies that

$$I_{t,\mathrm{disc}}(f) = I_{t,\mathrm{disc}}^*(f^*), \tag{25.28}$$

for any $t \geq 0$ and $f \in \mathcal{H}(G)$. Given the explicit definition (21.19) of $I_{t,\mathrm{disc}}(f)$, one could try to use (25.28) to establish an explicit global correspondence $\pi \to \pi^*$ from automorphic representations in the discrete spectrum of $G$ to automorphic representations in the discrete spectrum of $G^*$. However, this has not been done. So far as I know, the best results are due to Vigneras [**Vi**], who establishes the correspondence in the special case that for any $v$, $G(F_v)$ is either the multiplicative group of a division algebra, or is equal to $GL(n, F_v)$. (See also [**HT**].) Since the local condition implies that $G(F)\backslash G(\mathbb{A})^1$ is compact, this special case relies only on the trace formula for compact quotient, and a simple version of the trace formula (such as that of Corollary 23.6) for $GL(n)$. The general problem seems to be accesssible, at least in part, and would certainly be interesting.

## 26. Functoriality and base change for $GL(n)$

The third application of the invariant trace formula is to cyclic base change for $GL(n)$. This again entails a comparison of trace formulas. The base change comparison is very similar to that for inner twistings of $GL(n)$. We recall that the two were actually treated together in [**AC**]. Having just discussed the inner twisting comparison in some detail, we shall devote most of this section to some broader questions related to base change.

Base change is a special case of Langlands' general principle of functoriality. It is also closely related to a separate case of functoriality, Langlands' conjectural formulation of nonabelian class field theory. We have alluded to functoriality earlier, without actually stating it. Let us make up for this omission now.

For the time being, $G$ is to be a general group over the number field $F$. In fact, we regard $G$ as a group over some given extension $k$ of $F$. The theory of algebraic groups assigns to $G$ a canonical based root datum

$$\Psi(G) = (X, \Delta, X^\vee, \Delta^\vee),$$

equipped with an action of the Galois group

$$\Gamma_k = \mathrm{Gal}(\overline{k}/k).$$

Recall that there are many based root data attached to $G$. They are in bijection with pairs $(B, T)$, where $T$ is a maximal torus in $G$, and $B$ is a Borel subgroup of $G$ containing $T$. However, there is a *canonical* isomorphism between any two of them, given by any inner automorphism of $G$ between the corresponding two pairs. It is this property that gives rise to the canonical based root domain $\Psi(G)$. By construction, the group $\mathrm{Aut}\big(\Psi(G)\big)$ of automorphisms of $\Psi(G)$ is canonically isomorphic to the group

$$\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Int}(G)$$

of outer automorphisms of $G$. The $\Gamma_k$-action on $\Psi(G)$ comes from a choice of isomorphism $\psi_s$ from $G$ to a split group $G_s^*$. It is given by the homomorphism from $\Gamma_k$ to $\mathrm{Out}(G)$ defined by

$$\sigma \longrightarrow \psi_s \circ \sigma(\psi_s)^{-1}, \qquad \sigma \in \Gamma_k.$$

(See [**Spr2**, §1], [**Ko3**, (1.1)–(1.2)].)

Recall that a *splitting* of $G$ is a pair $(B, T)$, together with a set $\{X_\alpha : \alpha \in \Delta\}$ of nonzero vectors in the associated root spaces $\{\mathfrak{g}_\alpha : \alpha \in \Delta\}$. There is a canonical isomorphism from the group $\mathrm{Out}(G)$, and hence also the group $\mathrm{Aut}\big(\Psi(G)\big)$, onto the group of automorphisms of $G$ that preserve a given splitting [**Spr2**, Proposition 2.13]. Recall also that an action of any finite group by automorphisms on $G$ is called an *L-action* if it preserves some splitting of $G$. We define a *dual group* of $G$ to be a complex reductive group $\widehat{G}$, equipped with an $L$-action of $\Gamma_k$, and a $\Gamma_k$-isomorphism from $\Psi(\widehat{G})$ to the dual

$$\Psi(G)^\vee = (X^\vee, \Delta^\vee, X, \Delta)$$

of $\Psi(G)$. Suppose for example that $G$ is a torus $T$. Then

$$\Psi(T) = \big(X(T), \emptyset, X(T)^\vee, \emptyset\big),$$

where $X(T)^\vee = \mathrm{Hom}\big(X(T), \mathbb{Z}\big)$ is the dual of the additive character group $X(T)$. The dual group of $T$ is the complex dual torus

$$\widehat{T} = X(T) \otimes \mathbb{C}^*,$$

defined as a tensor product over $\mathbb{Z}$ of two abelian groups. In general, $\widehat{G}$ comes with the structure that assigns to any pair $(B,T)$ for $G$, and any pair $(\widehat{B},\widehat{T})$ for $\widehat{G}$, a $\Gamma_k$-isomorphism from $\widehat{T}$ to a dual torus for $T$.

An *L-group* for $G$ can take one of several forms. The Galois form is a semidirect product

$$^LG = \widehat{G} \rtimes \Gamma_k,$$

with respect to the $L$-action of $\Gamma_k$ on $\widehat{G}$. For many purposes, one can replace the profinite group $\Gamma_k$ with a finite group $\Gamma_{k'/k} = \mathrm{Gal}(k'/k)$, for a Galois extension $k'/k$ over which $G$ splits. For example, if $G$ is a group such as $GL(n)$ that splits over $k$, one can often work with $\widehat{G}$ instead of the full $L$-group. If $k$ is a local or global field, one sometimes replaces $\Gamma_k$ with the corresponding Weil group $W_k$, which we recall is a locally compact group equipped with a continuous homomorphism into $\Gamma_k$ [**Tat2**]. The Weil form of the $L$-group is a semidirect product

$$^LG = \widehat{G} \rtimes W_k$$

obtained by pulling back the $L$-action from $\Gamma_k$ to $W_k$. The symbol $^LG$ is generally used in this way to denote any of the forms of the $L$-group. Suppose that $k$ is the completion $F_v$ of $F$ with respect to a valuation $v$. The local Galois group $\Gamma_{F_v}$ or Weil group $W_{F_v}$ comes with a conjugacy class of embeddings into its global counterpart $\Gamma_F$ or $W_F$. There is consequently a conjugacy class of embeddings of the local $L$-group $^LG_v$ into $^LG$, which is trivial on $\widehat{G}$.

Suppose that as a group over $F$, $G$ is unramified at a given place $v$. As we recall, this means that $v$ is nonarchimedean, that $G$ is quasisplit over $F_v$, and that $G$ splits over a finite unramified extension $F_v'$ of $F_v$. We recall also that $\Gamma_{F_v'/F_v}$ is a finite cyclic group, with a canonical generator the Frobenius automorphism $\mathrm{Frob}_v$. We take the finite form

$$^LG_v = \widehat{G} \rtimes \Gamma_{F_v'/F_v}$$

of the $L$-group of $G$ over $F_v$ determined by the outer automorphism $\mathrm{Frob}_v$ of $\widehat{G}$. We can choose a pair $(B_v,T_v)$ defined over $F_v$ such that the torus $T_v$ splits over $F_v'$, and a hyperspecial maximal compact subgroup $K_v$ of $G(F_v)$ that lies in the apartment of $T_v$ [**Ti**]. The *unramified* representations of $G(F_v)$ (relative to $K_v$) are the irreducible representations whose restrictions to $K_v$ contain the trivial representation.

If $\lambda$ belongs to the space $\mathfrak{a}^*_{T_v,\mathbb{C}}$, and $1_{v,\lambda}$ is the unramified quasicharacter

$$t_v \longrightarrow q_v^{-\lambda(H_{T_v}(t_v))}, \qquad t_v \in T_v(F_v),$$

the induced representation $\mathcal{I}_{B_v}(1_{v,\lambda})$ contains the trivial representation of $K_v$ with multiplicity 1. This representation need not be irreducible. However, it does have a unique irreducible constituent $\pi_{v,\lambda}$ that contains the trivial representation of $K_v$, and is hence unramified. Obviously $\pi_{v,\lambda}$ depends only on the image of $\lambda$ in the quotient of $\mathfrak{a}^*_{T_v,\mathbb{C}}$ by the discrete subgroup

$$\Lambda_v = \Big(\frac{2\pi i}{\log q_v}\Big)\mathrm{Hom}(\mathfrak{a}_{T_v,F_v},\mathbb{Z}) = \frac{i}{\log q_v}\mathfrak{a}^\vee_{T_v,F_v}.$$

It also depends only on the orbit of $\lambda$ under the restricted Weyl group $W_{v0}$ of $(G, A_{T_v})$. The correspondence $\lambda \to \pi_{v,\lambda}$ is thus a mapping from the quotient

$$W_{v0}\backslash\mathfrak{a}^*_{T_v,\mathbb{C}}/\Lambda_v \tag{26.1}$$

to the set of unramified representations of $G(F_v)$. One shows that the mapping is a bijection. (See [**Ca**], for example.) On the other hand, there is a canonical homomorphism $\lambda \to q_v^{-\lambda}$ from $\mathfrak{a}^*_{T_v,\mathbb{C}}/\Lambda_v$ to the complex torus $\widehat{T}_v$, which takes a point in (26.1) to a $W_{v0}$-orbit of points in $\widehat{T}_v$. One shows that the correspondence

$$\lambda \longrightarrow q_v^{-\lambda} \rtimes \mathrm{Frob}_v$$

is a bijection from (26.1) onto the set of semisimple conjugacy classes in ${}^LG_v$ whose image in $\Gamma_{F'_v/F_v}$ equals $\mathrm{Frob}_v$. (See [**Bor3**, (6.4), (6.5)], for example.) It follows that there is a canonical bijection

$$\pi_v \longrightarrow c(\pi_v)$$

from the set of unramified representations of $G(F_v)$ onto the set of semisimple conjugacy classes in ${}^LG_v$ that project to $\mathrm{Frob}_v$. This mapping is due to Langlands [**Lan3**], and in itself justifies the introduction of the $L$-group.

The reader may recall that the symbol $c(\pi_v)$ also appeared earlier. It was introduced in §2 (in the special case $F = \mathbb{Q}$) to denote the homomorphism from the unramified Hecke algebra $\mathcal{H}_v = \mathcal{H}(G_v, K_v)$ to $\mathbb{C}$ attached to $\pi_v$. The two uses of the symbol are consistent. They are related by the Satake isomorphism from $\mathcal{H}_v$ to the complex co-ordinate algebra on the space (26.1). (See [**Ca**, (4.2)], for example. By the co-ordinate algebra on (26.1), we mean the subalgebra of $W_{v0}$-invariant functions in the co-ordinate algebra of $\mathfrak{a}^*_{T_v,\mathbb{C}}/\Lambda_v$, regarded as a subtorus of $\widehat{T}_v$.) The complex valued homomorphisms of $\mathcal{H}_v$ are therefore bijective with the points in (26.1), and hence with the set of semisimple conjugacy classes in ${}^LG_v$ that project to $\mathrm{Frob}_v$.

Suppose now that $\pi$ is an automorphic representation of $G$. Then $\pi = \bigotimes \pi_v$, where $\pi_v$ is unramified for almost all $v$. We choose a finite form ${}^LG = \widehat{G} \rtimes \Gamma_{F'/F}$ of the global $L$-group, for some finite Galois extension $F'$ of $F$ over which $G$ splits, and a finite set of valuations $S$ outside of which $\pi$ and $F'$ are unramified. For any $v \notin S$, we then write $c_v(\pi)$ for the image of $c(\pi_v)$ under the canonical conjugacy class of embeddings of ${}^LG_v = \widehat{G} \rtimes \Gamma_{F'_v/F_v}$ into ${}^LG$. This gives a correspondence

$$\pi \longrightarrow c(\pi) = \{c_v(\pi) : \ v \notin S\}$$

from automorphic representations of $G$ to families of semisimple conjugacy classes in ${}^LG$. The construction becomes independent of the choice of $F'$ and $S$ if we agree to identify to families of conjugacy classes that are equal almost everywhere.

An automorphic representation thus carries some very concrete data, namely the complex parameters that determine the conjugacy classes in the associated family. The interest stems not so much from the values assumed by individual classes $c_v(\pi)$, but rather in the relationships among the different classes implicit in the requirement that $\pi$ be automorphic. Following traditions from number theory and algebraic geometry, Langlands wrapped the data in analytic garb by introducing an unramified $L$-function

$$L^S(s, \pi, r) = \prod_{v \notin S} \det\big(1 - r\big(c_v(\pi)\big)q_v^{-s}\big)^{-1}, \tag{26.2}$$

for any automorphic representation $\pi$, any reasonable finite dimensional representation

$$r : \ {}^LG \longrightarrow GL(N, \mathbb{C}),$$

and any finite set $S$ of valuations outside of which $\pi$ and $r$ are unramified. He observed that the product converged for $\mathrm{Re}(s)$ large, and conjectured that it had analytic continuation with functional equation.

Langlands' principle of functoriality [**Lan3**] postulates deep and quite unexpected reciprocity laws among the families $c(\pi)$ attached to different groups. Assume that $G$ is quasisplit over $F$, and that $G'$ is a second connected reductive group over $F$. Suppose that

$$\rho :\ {}^LG' \longrightarrow {}^LG$$

is an $L$-homomorphism of $L$-groups. (Besides satisfying the obvious conditions, an $L$-homomorphism between two groups that each project onto a common Galois or Weil group is required to be compatible with the two projections.) The principle of functoriality asserts that for any automorphic representation $\pi'$ of $G'$, there is an automorphic representation $\pi$ of $G$ such that

$$c(\pi) = \rho\big(c(\pi')\big). \tag{26.3}$$

In other words, $c_v(\pi) = \rho\big(c_v(\pi')\big)$ for every valuation $v$ outside some finite set $S$. Functoriality thus postulates a correspondence $\pi' \to \pi$ of automorphic representations, which depends only on the $\widehat{G}$-orbit of $\rho$. We shall recall three basic examples.

Suppose that $G$ is an inner form of a quasiplit group $G^*$, equipped with an inner twist

$$\psi :\ G \to G^*.$$

In other words, $\psi$ is an isomorphism such that $\psi \circ \sigma(\psi)^{-1}$ is an inner automorphism of $G^*$ for every $\sigma \in \Gamma_F$. It determines an $L$-isomorphism

$${}^L\psi :\ {}^LG \longrightarrow {}^LG^*,$$

which allows us to identify the two $L$-groups. Functoriality asserts that the set of automorphic families $\{c(\pi)\}$ of conjugacy classes for $G$ is contained in the set of such families $\{c(\pi^*)\}$ for $G^*$. Our last section was devoted to the study of this question in the case $G^* = GL(n)$. It is pretty clear from the conclusion (25.28), together with the explicit formula for $I_{t,\mathrm{disc}}(f)$ and the fact that $f_v = f_v^*$ for almost all $v$, that something pretty close to the assertion of functoriality holds in this case. However, the precise nature of the correspondence remains open.

Langlands introduced the second example in his original article [**Lan3**], as a particularly vivid illustration of the depth of functoriality. It concerns the case that $G$ is an arbitrary quasisplit group, and $G'$ is the trivial group $\{1\}$. The $L$-group ${}^LG'$ need not be trivial, since it can take the form of the Galois group $\Gamma_F$. Functoriality applies to a continuous homomorphism

$$\rho :\ \Gamma_F \longrightarrow {}^LG$$

whose composition with the projection of ${}^LG$ on $\Gamma_F$ equals the identity. Since $\Gamma_F$ is totally disconnected, $\rho$ can be identified with an $L$-homomorphism from $\Gamma_{F'/F}$ to the restricted form ${}^LG = \widehat{G} \rtimes \Gamma_{F'/F}$ of the $L$-group of $G$ given by some finite Galois extension $F'$ of $F$. Let $S$ be any finite set of valuations $v$ of $F$ outside of which $F'$ is unramified. Then for any $v \notin S$, $F'_{v'}/F_v$ is an unramified extension of local fields, for any (normalized) valuation $v'$ of $F'$ over $v$. Its Galois group is cyclic, with a canonical generator $\mathrm{Frob}_v = \mathrm{Frob}_{v,F'}$, whose conjugacy class in $\Gamma_{F'/F}$ is independent of the choice of $v'$. Thus, $\rho$ gives rise to a family

$$\{\rho(\mathrm{Frob}_v) : \ v \notin S\}$$

of conjugacy classes of finite order in ${}^LG$. If $\pi'$ is the trivial automorphic representation of $G' = \{1\}$ and $v \notin S$, the image of $c_v(\pi')$ in the group ${}^LG' = \Gamma_{F'/F}$ equals $\mathrm{Frob}_v$, by construction. Functoriality asserts that there is an automorphic representation $\pi$ of $G$ such that for any $v \notin S$, the class $c_v(\pi)$ in ${}^LG$ equals $\rho(\mathrm{Frob}_v)$. A more general assertion applies to the Weil form of ${}^LG'$. In this form, functoriality attaches an automorphic representation $\pi$ to any $L$-homomorphism

$$\phi : \ W_F \longrightarrow {}^LG$$

of the global Weil group into ${}^LG$.

The third example is general base change. It applies to an arbitrary group $G'$ over $F$, and a finite extension $E$ of $F$ over which $G'$ is quasisplit. Given these objects, we take $G$ to be the group $R_{E/F}(G'_E)$ over $F$ obtained from the quasisplit group $G'$ over $E$ by restriction of scalars. Following [**Bor3**, §4–5], we identify $\widehat{G}$ with the group of functions $g$ from $\Gamma_F$ to $\widehat{G}'$ such that

$$g(\sigma\tau) = \sigma g(\tau), \qquad \sigma \in \Gamma_E, \ \tau \in \Gamma_F,$$

with pointwise multiplication, and $\Gamma_F$-action

$$(\tau_1 g)(\tau) = g(\tau\tau_1), \qquad \tau, \tau_1 \in \Gamma_F.$$

We then obtain an $L$-homomorphism

$$\rho : \ {}^LG' \longrightarrow {}^LG$$

by mapping any $g' \in {}^LG'$ to the function

$$g(\tau) = \tau g', \qquad \tau \in \Gamma_F,$$

on $\Gamma_F$. This case of functoriality can be formulated in slightly more concrete terms. The restriction of scalars functor provides a canonical isomorphism from $G(\mathbb{A})$ onto $G'(\mathbb{A}_E)$, which takes $G(F)$ to $G'(E)$. The automorphic representations of $G$ are therefore in bijection with those of $G'_E$. This means that we can work with the $L$-group ${}^LG'_E = \widehat{G}' \rtimes \Gamma_E$ of $G'_E$ instead of ${}^LG$. Base change becomes a conjectural correspondence $\pi' \to \pi$ of automorphic representations of $G'$ and $G'_E$ such that for any valuation $v$ of $F$ for which $\pi'$ and $E$ are unramified, and any valuation $w$ of $E$ over $v$, the associated conjugacy classes are related by

$$c_v(\pi') = c_w(\pi)^{f_w}, \qquad f_w = \deg(E_w/F_v).$$

We should bear in mind that Langlands also postulated a local principle of functoriality. This takes the form of a conjectural correspondence $\pi'_v \to \pi_v$ of irreducible representations of $G'(F_v)$ and $G(F_v)$, for any $v$ and any local $L$-homomorphism $\rho_v$ of local $L$-groups, which is compatible with the global functoriality correspondence $\pi' \to \pi$. Representations $\pi_v$ of the local groups $G(F_v)$ are important for the functional equations of $L$-functions (among many other things). Langlands conjectured the existence of local $L$-functions $L(s, \pi_v, r_v)$, which reduce to the relevant factors of (26.2) in the unramified case, and local $\varepsilon$-factors

$$\varepsilon(s, \pi_v, r_v, \psi_v) = a q_v^{-bs}, \qquad a \in \mathbb{C}, \ b \in \mathbb{Z},$$

which equal 1 in the unramified case, such that the finite product

$$\varepsilon(s,\pi,r)=\prod_v \varepsilon(s,\pi_v,r_v,\psi_v)$$

is independent of the nontrivial additive character $\psi$ of $\mathbb{A}/F$ of which $\psi_v$ is the restriction, and such that the product

$$L(s,\pi,r)=\Big(\prod_{v\notin S} L(s,\pi_v,r_v)\Big)L^S(s,\pi,r)$$

satisfies the functional equation

$$L(s,\pi,r)=\varepsilon(s,\pi,r)L(1-s,\pi,r^\vee). \tag{26.4}$$

We have written $r^\vee$ here for the contragredient of the representation $r$. The local $L$-functions and $\varepsilon$-factors should be compatible with the local version of functoriality, in the sense that

$$L(s,\pi'_v,r_v\circ\rho_v)=L(s,\pi_v,r_v)$$

and

$$\varepsilon(s,\pi_v,r_v\circ\rho_v,\psi_v)=\varepsilon(s,\pi_v,r_v,\psi_v).$$

These relations are obvious in the unramified case. In general, they imply corresponding relations for global $L$-functions and $\varepsilon$-factors.

Suppose now that $G=GL(n)$. The constructions above are, not surprisingly, more explicit in this case. There is no harm in reviewing them in concrete terms.

Let $v$ be a nonarchimedean valuation, and take $(B_v,T_v)$ to be the standard pair $(B,M_0)$. If $\lambda$ belongs to $\mathfrak{a}^*_{T_v,\mathbb{C}}\cong\mathbb{C}^n$, the induced representation $\mathcal{I}_{B_v}(1_{v,\lambda})$ acts by right translation on the space of functions $\phi$ on $G(F_v)$ such that

$$\phi(bx)=|b_{11}|^{\lambda_1+\frac{n-1}{2}}|b_{22}|^{\lambda_2+\frac{n-3}{2}}\dots|b_{nn}|^{\lambda_n-(\frac{n-1}{2})}\phi(x),\quad b\in B_v(F_v),\ x\in G(F_v).$$

It has a unique irreducible constituent $\pi_{v,\lambda}$ that contains the trivial representation of $K_v=GL(n,\mathfrak{o}_v)$. Two such representations $\pi_{v,\lambda'}$ and $\pi_{v,\lambda}$ are equivalent if and only if the corresponding vectors $\lambda',\lambda\in\mathbb{C}^n$ are related by

$$(\lambda'_1,\dots,\lambda'_n)\equiv(\lambda_{\sigma(1)},\dots,\lambda_{\sigma(n)})\Big(\mathrm{mod}\Big(\frac{2\pi i}{\log\, q_v}\Big)\mathbb{Z}^n\Big),$$

for some permutation $\sigma\in S_n$. The dual group $\widehat{G}$ equals $GL(n,\mathbb{C})$. We give it the canonical structure, which assigns to the standard pairs $(B,T)$ and $(\widehat{B},\widehat{T})$ in $G$ and $\widehat{G}$ the obvious isomorphism of $\widehat{T}$ with the complex dual torus of $T$. Since the action of $\Gamma_F$ on $\widehat{G}$ is trivial, we can take the restricted form ${}^LG=\widehat{G}$ of the $L$-group. The semisimple conjugacy class of the representation $\pi_{v,\lambda}$ is then given by

$$c(\pi_{v,\lambda})=\left\{\begin{pmatrix} q_v^{-\lambda_1} & & 0\\ & \ddots & \\ 0 & & q_v^{-\lambda_n}\end{pmatrix}\right\}.$$

Given an automorphic representation $\pi$ of $GL(n)$, let $S$ be any finite set of valuations outside of which $\pi$ is unramified. Then $\pi$ gives rise to a family

$$c(\pi)=\big\{c_v(\pi)=c(\pi_v):\ v\notin S\big\}$$

of semisimple conjugacy classes in $\widehat{G}=GL(n,\mathbb{C})$. It is known that if $\pi$ occurs in the spectral decomposition of $L^2\big(G(F)\backslash G(\mathbb{A})\big)$, it is uniquely determined by the family $c(\pi)$ [**JaS**]. This remarkable property is particular to $G=GL(n)$.

Consider a continuous $n$-dimensional representation $r$ of $\Gamma_F$. Then $r$ lifts to a representation of a finite group $\Gamma_{F'/F}$, for a finite Galois extension $F'$ of $F$. We may as well take $F'$ to be the minimal such extension, for which $\Gamma_{F'}$ is the kernel of $r$. Let $S$ be any finite set of valuations outside of which $F'$ is unramified. The representation $r$ then gives rise to a family

$$c(r) = \bigl\{c_v(r) = r(\mathrm{Frob}_v):\ v \notin S\bigr\}$$

of semisimple conjugacy classes in $\widehat{G} = GL(n, \mathbb{C})$. As in the automorphic setting, the equivalence class of $r$ is uniquely determined by $c(r)$. For the Tchebotarev density theorem characterizes $F'$ as the Galois extension of $F$ for which

$$\mathrm{Spl}_{F'/F} = \{v \notin S:\ c_v(r) = 1\}$$

is the set of valuations outside of $S$ that split completely in $F'$. Since the Tchebotarev theorem deals in densities of subsets, the characterization is independent of the choice of $S$. The theorem also implies that every conjugacy class in the group $\Gamma_{F'/F}$ is of the form $\mathrm{Frob}_v$, for some $v \notin S$. The character of $r$ is therefore determined by the family $c(r)$.

According to the second example of functoriality above, specialized to the case that $G = GL(n)$, there should be an automorphic representation $\pi$ attached to any $r$ such that $c(\pi) = c(r)$. Consider the further specialization to the case that $n = 1$. The one dimensional characters of the group $\Gamma_F$ are the characters of its abelianization $\Gamma_F^{\mathrm{ab}}$. The case $n = 1$ of Langlands' Galois representation conjecture could thus be interpreted as the existence of a surjective dual homomorphism

$$GL(1, F)\backslash GL(1, \mathbb{A}) = F^*\backslash\mathbb{A}^* \ \longrightarrow\ \Gamma_F^{\mathrm{ab}}. \tag{26.5}$$

The condition $c(\pi) = c(r)$ specializes to the requirement that the composition of (26.5) with the projection of $\Gamma_F^{\mathrm{ab}}$ onto the Galois group of any finite abelian extension $F'$ of $F$ satisfy

$$x_v \ \longrightarrow\ (\mathrm{Frob}_v)^{\mathrm{ord}(x_v)}, \qquad\qquad x_v \in F_v^*,$$

where $v$ is any valuation that is unramified in $F'$, $\mathrm{Frob}_v$ is the corresponding Frobenius element in the abelian group $\Gamma_{F'/F}$, and

$$\mathrm{ord}(x_v) = -\log_{q_v}(|x_v|).$$

The mapping (26.5) has been known for many years. It is the Artin reciprocity law, which is at the heart of class field theory. (See [**Has**], [**Tat1**].) Langlands' Galois representation conjecture thus represents a nonabelian analogue of class field theory. If $n = 2$ and $\Gamma_{F'/F}$ is solvable, it was established as a consequence of cyclic base change for $GL(2)$ [**Lan9**], [**Tu**]. If $n$ is arbitrary and $\Gamma_{F'/F}$ is nilpotent, it is a consequence [**AC**, Theorem 3.7.3] of cyclic base change for $GL(n)$, the ostensible topic of this section. Other cases for $n = 2$ have been established [**BDST**], as have a few other cases in higher rank.

Besides extending class field theory, Langlands' Galois representation conjecture has important implications for Artin $L$-functions

$$L^S(s, r) = \prod_{v \notin S} \det\bigl(1 - r(\mathrm{Frob}_v)q_v^{-s}\bigr)^{-1}.$$

If $r$ corresponds to $\pi$, it is clear that

$$L^S(s, r) = L^S(s, \pi), \tag{26.6}$$

where $L^S(s,\pi)$ is the automorphic $L$-function for $GL(n)$ relative to the standard, $n$-dimensional representation of ${}^LG = GL(n,\mathbb{C})$. It has been known for some time how to construct the local $L$-functions and $\varepsilon$-factors in this case so that the functional equation (26.4) holds [**GoJ**]. These results are now part of the larger theory of Rankin-Selberg $L$-functions $L(s,\pi_1\times\pi_2)$, attached to representations $\pi_1\otimes\pi_2$ of $GL(n_1)\times GL(n_2)$, and the representation

$$(g_1,g_2):\ X \longrightarrow g_1Xg_2^{-1}, \qquad\qquad X\in M_{n_1\times n_2}(\mathbb{C}),$$

of $GL(n_1,\mathbb{C})\times GL(n_2,\mathbb{C})$ [**JPS**]. In fact, there is a broader theory still, known as the Langlands-Shahidi method, which exploits the functional equations from the theory of Eisenstein series. It pertains to automorphic $L$-functions of a maximal Levi subgroup $M$ of a given group, and the representation of ${}^LM$ on the Lie algebra of a unipotent radical [**GS**]. Be that as it may, our refined knowledge of the automorphic $L$-function in the special case encompassed by the right hand side of (26.6) would establish critical analytic properties of the Artin $L$-function on the left hand side of (26.6).

The Langlands conjecture for Galois representations (which we re-iterate is but a special case of functoriality) is still far from being solved in general. However, it plays an important role purely as a conjecture in motivating independent operations on automorphic representations. Nowhere is this more evident than in the question of cyclic base change of prime order for the group $GL(n)$.

Suppose that $E$ is a Galois extension of $F$, with cyclic Galois group $\{1,\sigma,\dots,\sigma^{\ell-1}\}$ of prime order $\ell$. To be consistent with the description of base change above, we change notation slightly. We write $r'$ instead of $r$ for a continuous $n$-dimensional representation of $\Gamma_F$, leaving $r$ to stand for a continuous $n$-dimensional representation of $\Gamma_E$. Regarded as equivalence classes of representations, these two families come with two bijections $r\to r^\sigma$ and $r'\to r'\otimes\eta$ of order $\ell$, where $r^\sigma(\tau)=r(\sigma\tau\sigma^{-1})$, and $\eta$ is the pullback to $\Gamma_F$ of the character on $\Gamma_{E/F}$ that maps the generator $\sigma$ to $e^{\frac{2\pi i}{\ell}}$. The main operation is the mapping $r'\to r$ obtained by restricting $r'$ to the subgroup $\Gamma_E$ of $\Gamma_F$. This mapping is characterized in terms of conjugacy classes by the relation

$$c_w(r)=\begin{cases} c_v(r'), & \text{if } v \text{ splits in } E,\\ c_v(r')^\ell, & \text{otherwise,}\end{cases}$$

for any valuation $v$ of $F$ at which $r'$ and $E$ are unramified and any valuation $w$ over $v$, and satisfies the following further conditions.

(i) The image of the mapping is the set of $r$ with $r^\sigma=r$.

(ii) If $r'$ is irreducible, the fibre of its image equals

$$\{r',r'\otimes\eta,\dots,r'\otimes\eta^{\ell-1}\}.$$

(iii) If $r'$ is irreducible, its image $r$ is irreducible if and only if $r'\neq r'\otimes\eta$, which is to say that the fibre in (ii) contains $\ell$ elements.

(iv) If $r'$ is irreducible and $r'=r'\otimes\eta$, its image equals a direct sum

$$r=r_1\oplus r_1^\sigma\oplus\cdots\oplus r_1^{\sigma^{\ell-1}},$$

for an irreducible representation $r_1$ of degree $n_1=n\ell^{-1}$ such that $r_1^\sigma\neq r_1$. Conversely, the preimage of any such direct sum consists of a representation $r'$ that is irreducible and satisfies $r'=r'\otimes\eta$.

These conditions are all elementary consequences of the fact that $\Gamma_E$ is a normal subgroup of prime index in $\Gamma_F$. For example, the representation $r'$ in (iv) is obtained by induction of the representation $r_1$ from $\Gamma_E$ to $\Gamma_F$.

Base change is a mapping of automorphic representations with completely parallel properties. We write $\pi'$ and $\pi$ for (equivalence classes of) automorphic representations of $GL(n)_F$ and $GL(n)_E$ respectively. The two families come with bijections $\pi \to \pi^\sigma$ and $\pi' \to \pi' \otimes \eta$ of order $\ell$, where $\pi^\sigma(x) = \pi\bigl(\sigma(x)\bigr)$, and $\eta$ has been identified with the 1-dimensional automorphic representation of $GL(n)_F$ obtained by composing the determinant on $GL(n, \mathbb{A})$ with the pullback of $\eta$ to $GL(1, \mathbb{A})$ by (26.5). The results of [**AC**] were established for cuspidal automorphic representations, and the larger class of "induced cuspidal" representations. For $GL(n)_E$, this larger class consists of induced representations

$$\pi = \pi_1 \boxtimes \cdots \boxtimes \pi_p = \mathrm{Ind}_P^G(\pi_1 \otimes \cdots \otimes \pi_p),$$

where $P$ is the standard parabolic subgroup of $GL(n)$ corresponding to a partition $(n_1, \dots, n_p)$, and $\pi_i$ is a *unitary* cuspidal automorphic representation of $GL(n_i)_E$. Any such representation is automorphic, by virtue of the theory of Eisenstein series.

THEOREM 26.1. (Base change for $GL(n)$). *There is a mapping $\pi' \to \pi$ from induced cuspidal automorphic representations of $GL(n)_F$ to induced cuspidal automorphic representations of $GL(n)_E$, which is characterized by the relation*

$$c_w(\pi) = \begin{cases} c_v(\pi'), & \text{if } v \text{ splits in } E, \\ c_v(\pi')^\ell, & \text{otherwise}, \end{cases} \tag{26.7}$$

*for any valuation $v$ of $F$ at which $\pi'$ and $E$ are unramified and any valuation $w$ of $F$ over $v$, and which satisfies the following further conditions.*

(i) *The image of the mapping is the set of $\pi$ with $\pi^\sigma = \pi$.*

(ii) *If $\pi'$ is cuspidal, the fibre of its image equals*

$$\{\pi', \pi' \otimes \eta, \dots, \pi' \otimes \eta^{\ell-1}\}.$$

(iii) *If $\pi'$ is cuspidal, its image $\pi$ is cuspidal if and only if $\pi' \neq \pi' \otimes \eta$, which is to say that the fibre in (ii) contains $\ell$ elements.*

(iv) *If $\pi'$ is cuspidal and $\pi' = \pi' \otimes \eta$, its image equals a sum*

$$\pi = \pi_1 \boxtimes \pi_1^\sigma \boxtimes \cdots \boxtimes \pi_1^{\sigma^{\ell-1}},$$

*for a cuspidal automorphic representation $\pi_1$ of $GL(n_1)_E$ such that $\pi_1^\sigma \neq \pi_1$. Conversely, the preimage of any such sum consists of a representation $\pi'$ that is cuspidal and satisfies $\pi' = \pi' \otimes \eta$.*

**Remark.** The theorem provides two mappings of cuspidal automorphic representations. Base change gives an $\ell$ to 1 mapping $\pi' \to \pi$, from the set of cuspidal representations of $\pi'$ of $GL(n)_F$ with $\pi' \neq \pi' \otimes \eta$ onto the set of cuspidal representations $\pi$ of $GL(n)_E$ with $\pi = \pi^\sigma$. The second mapping is given by (iv), and is known as *automorphic induction.* It is an $\ell$ to 1 mapping $\pi_1 \to \pi'$, from the set of cuspidal autmorphic representations $\pi_1$ of $GL(n_1)_E$ with $\pi_1 \neq \pi_1^\sigma$ onto the set of cuspidal automorphic representations $\pi'$ of $GL(n)_E$ with $\pi' = \pi' \otimes \eta$.

Theorem 26.1 contains the main results of [**AC**]. It is proved by a comparison of two trace formulas. One is the invariant trace formula for the group $GL(n)_F$.

The other is the invariant twisted trace formula, applied to the automorphism of the group $R_{E/F}\big(GL(n)_E\big)$ determined by $\sigma$.

The twisted trace formula is a generalization of the ordinary trace formula. It applies to an $F$-rational automorphism $\theta$ of finite order of a connected reductive group $G$ over $F$. The twisted trace formula was introduced by Saito for classical modular forms [**Sai**], by Shintani for the associated automorphic representations of $GL(2)$ (see [**Shin**]), and by Langlands for general automorphic representations of $GL(2)$ [**Lan9**]. The idea, in the special case of compact quotient, for example, is to express the trace of an operator

$$R(f) \circ \theta, \qquad f \in \mathcal{H}\big(G(\mathbb{A})\big),$$

in terms of twisted orbital integrals

$$\int_{G_{\theta\gamma}(A)\backslash G(\mathbb{A})} f\big(x^{-1}\gamma\theta(x)\big)\mathrm{d}x, \qquad \gamma \in G(F).$$

This gives a geometric expression for a sum of twisted characters

$$\sum_{\pi} m(\pi)\mathrm{tr}\big(\pi(f) \circ \theta\big),$$

taken over irreducible representations $\pi$ of $G$ such that $\pi^\theta = \pi$. It is exactly the sort of formula needed to quantify the proposed image of the base change map.

In general, our discussion that led to the invariant trace formula applies also to the twisted case. (See [**CLL**], [**A14**].) Most of the results in fact remain valid as stated, if we introduce a minor change in notation. We take $G$ to be a connected component of a (not necessarily connected) reductive group over $F$ such that $G(F)$ is not empty. We write $G^+$ for the reductive group generated by $G$, and $G^0$ for the connnected component of 1 in $G^+$. We then consider distributions on $G(\mathbb{A})$ that are invariant with respect to the action of $G^0(\mathbb{A})$ on $G(\mathbb{A})$ by conjugation. The analogue of the Hecke algebra becomes a space $\mathcal{H}(G)$ of functions on a certain closed subset $G(\mathbb{A})^1$ of $G(\mathbb{A})$. The objects of Theorems 23.2, 23.3, and 23.4 can all be formulated in this context, and the invariant twisted trace formula becomes the identity of Theorem 23.4. (See [**A14**].) It holds for any $G$, under one condition. We require that the twisted form of the archimedean trace Paley-Wiener theorem of Clozel-Delorme [**CD**] hold for $G$. This condition, which was established by Rogawski in the $p$-adic case [**Ro2**], is needed to characterize the invariant image $\mathcal{I}(G)$ of the twisted Hecke algebra $\mathcal{H}(G)$. (See also [**KR**].)

For base change, we take

$$G = G^0 \rtimes \theta, \qquad G^0 = R_{E/F}\big(GL(n)_E\big),$$

where $\theta$ is the automorphism of $G^0$ defined by the generator $\sigma$ of $\Gamma_{E/F}$. We also set $G' = GL(n)_F$. Our task is to compare the invariant twisted trace formula of $G$ with the invariant trace formula of $G'$. The problem is very similar to the comparison for inner twistings of $GL(n)$, treated at some length in §25. In fact, we recall that the two comparisons were actually treated together in [**AC**]. We shall add only a few words here, concentrating on aspects of the problem that are different from those of §25.

The first step is to define a mapping $\gamma \to \gamma'$, which for any $k \supset F$ takes the set $\Gamma\big(G(k)\big)$ of $G^0(k)$-orbits in $G(k)$ to the set $\Gamma\big(G'(k)\big)$ of conjugacy classes in $G'(k)$. The mapping is analogous to the injection $\gamma \to \gamma^*$ of §25. In place of the inner twist, one uses the norm mapping from number theory, which in the present context becomes the mapping $\gamma \to \gamma^\ell$ from $G$ to $G^0$. Setting $k = F_v$, one uses the mapping $\gamma_v \to \gamma'_v$ to transfer twisted orbital integrals on $\Gamma_{\text{reg}}\big(G(F_v)\big)$. This gives a transformation $f_v \to f'_v$, from functions $f_v \in \mathcal{H}\big(G(F_v)\big)$ to functions $f'$ on $\Gamma_{\text{reg}}\big(G'(F_v)\big)$. One then combines Theorem 25.1 with methods of descent to show that $f'_v$ lies in the invariant Hecke algebra $\mathcal{I}\big(G'(F_v)\big)$.

One has then to combine the mappings $f_v \to f'_v$ into a global correspondence of adelic functions. This is more complicated than it was in §25. The problem is that $G'(F_v)$ is distinct from $G(F_v)$ at all places, not just the unramified ones. At almost all places $v$, we want $f_v$ to be the characteristic function of the compact subset $K_v = K_v^0 \rtimes \theta = G(\mathfrak{o}_v)$ of $G(F_v)$, and $f'_v$ to be the image in $\mathcal{I}\big(G'(F_v)\big)$ of the characteristic function of the maximal compact subgroup $K'_v = G'(\mathfrak{o}_v)$ of $G'(F_v)$. However, we do not know a priori that this is compatible with the transfer of orbital integrals. The assertion that the two mappings are in fact compatible is a special case of the twisted fundamental lemma. It was established in the case at hand by Kottwitz [**Ko4**]. The result of Kottwitz allows us to put the local mappings together. We obtain a mapping $f \to f'$ from $\mathcal{H}(G)$ to $\mathcal{I}(G')$, which takes any subspace $\mathcal{H}\big(G(F_S)^1\big)$ of $\mathcal{H}(G)$ to the corresponding subspace $\mathcal{I}\big(G'(F_S)^1\big)$ of $\mathcal{I}(G')$.

The next step is to extend the fundamental lemma to more general functions in an unramified Hecke algebra $\mathcal{H}\big(G(F_v), K_v^0\big)$. More precisely, one needs to show that at an unramified place $v$, the canonical mapping from $\mathcal{H}\big(G(F_v), K_v^0\big)$ to $\mathcal{H}\big(G'(F_v), K'_v\big)$ defined by Satake isomorphisms is compatible with the transfer of orbital integrals. This was established in [**AC**, §1.4], using the special case established by Kottwitz, and the simple forms of Corollary 23.6 of the two trace formulas. Further analysis of the two simple trace formulas allows one to establish local base change [**AC**, §1.6–1.7]. The result is a mapping $\pi'_v \to \pi_v$ of tempered representations, which satisfies local forms of the conditions of the theorem, and is the analogue of Theorem 25.1(b).

The expansions (23.11) and (23.12) represent the two sides of the invariant twisted trace formula for $G$. We define "endoscopic" forms $I_M^{\mathcal{E}}(\gamma, f)$, $a^{M,\mathcal{E}}(\gamma)$, $I_M^{\mathcal{E}}(\pi, f)$ and $a^{M,\mathcal{E}}(\pi)$ of the terms in the two expansions by using the mapping $f \to f'$ to pull back the corresponding terms from $G'$. The constructions are similar to those of §25, but with one essential difference. In the present situation, we have to average spectral objects $\widehat{I}_{M'}(\cdot, f')$ and $a^{M'}(\cdot)$ over representations $\tau \otimes \xi$, for characters $\xi$ on $M'(\mathbb{A})$ obtained from the original character $\eta$ on $\Gamma_{E/F}$. The reason for this is related to condition (ii) of the theorem, which in turn is a consequence of the fact that the norm mapping is not surjective. However, the averaging operation is not hard to handle. It is an essential part of the discussion in [**AC**, §2.10–2.12]. The identities of Theorems 25.5 and 25.6 can therefore be formulated in the present context. Their proof is more or less the same as in §25.

The analogue of the global spectral identity (25.13) (with $M = G$) is again what is most relevant for global applications. It leads directly to an identity

$$I_{t,\text{disc}}(f) = \widehat{I}'_{t,\text{disc}}(f'), \qquad f \in \mathcal{H}(G), \tag{26.8}$$

of $t$-discrete parts of the two trace formulas. One extracts global information from the last identity by allowing local components $f_v$ of $f$ to vary over unramified Hecke algebras $\mathcal{H}\big(G(F_v), K_v^0\big)$. By combining general properties of the distributions in (26.8) with operations on Rankin-Selberg $L$-functions $L(s, \pi_1 \times \pi_2)$, one establishes all the assertions of the theorem. (See [**AC**, Chapter 3].) □

We remark that the proof of base change in [**AC**] works only for cyclic extensions of prime degree (despite assertions in [**AC**] to the contrary). The mistake, which occurred in Lemma 6.1 of [**AC**], was pointed out by Lapid and Rogawski. In the case of $G = GL(2)$, they characterized the image of base change for a general cyclic extension by combining the special case of Theorem 26.1 established by Langlands [**Lan9**] with a second comparison of trace formulas [**LR**]. There is also a gap in the density argument at the top of p. 196 of [**AC**], which was filled in [**A29**, Lemma 8.2].

We know that the spectral decomposition for $GL(n)$ contains more than just induced cuspidal representations. In particular, the discete spectrum contains more than the cuspidal automorphic representations. The classification of the discrete spectrum for $GL(n)$ came after [**AC**]. It was established through a deep study by Moeglin and Waldspurger of residues of cuspidal Eisenstein series [**MW2**], following earlier work of Jacquet [**J**].

THEOREM 26.2. (Moeglin-Waldspurger). *The irreducible representations $\pi$ of $GL(n, \mathbb{A})$ that occur in $L^2_{\mathrm{disc}}\big(GL(n, F)\backslash GL(n, \mathbb{A})^1\big)$ have multiplicity one, and are parametrized by pairs $(k, \sigma)$, where $n = kp$ is divisible by $k$, and $\sigma$ in an irreducible unitary cuspidal automorphic representation of $GL(k, \mathbb{A})$. If $P$ is the standard parabolic subgroup of $GL(n)$ of type $(k, \dots, k)$, and $\rho_\sigma$ is the nontempered representation*

$$(\sigma \otimes \cdots \otimes \sigma) \cdot \delta_P^{\frac{1}{2}} : \ m \rightarrow \big(\sigma(m_1)|\det m_1|^{\frac{p-1}{2}}\big) \otimes \cdots \otimes \big(\sigma(m_p)|\det m_p|^{-\frac{p-1}{2}}\big)$$

*of $M_P(\mathbb{A}) \cong GL(k, \mathbb{A})^p$, then $\pi$ is the unique irreducible quotient of the induced representation $\mathcal{I}_P(\rho_\sigma)$.* $\square$

If we combine Theorem 26.1 with Theorem 26.2 (and the theory of Eisenstein series), we obtain a base change mapping $\pi' \rightarrow \pi$ for any representation $\pi'$ of $G'(\mathbb{A})$ that occurs in the spectral decomposition of $L^2\big(G'(F)\backslash G'(\mathbb{A})\big)$. It would be interesting, and presumably not difficult, to describe the general properties of this mapping. It would also be interesting to try to establish the last step of the proof of Theorem 26.1 without recourse to the argument based on $L$-functions in [**AC**, Chapter 3]. This might be possible with a careful study of the fine structure of the distributions on each side of (26.8).

As a postscript to this section on base change, we note that there is a suggestive way to look at the theorem of Moeglin and Waldspurger. It applies to those representations $\pi$ in the discrete spectrum for which the underlying cuspidal automorphic representation $\sigma$ is attached to an irreducible representation

$$\mu : \ W_F \ \longrightarrow \ GL(k, \mathbb{C})$$

of the global Weil group, according to the special case of functoriality we discussed earlier. One expects $\sigma$ to be tempered. This means that $\mu$ is (conjugate to) a unitary representation, or equivalently, that its image in $GL(k, \mathbb{C})$ is bounded. We are assuming that $n = kp$, for some positive integer $p$. Let $\nu$ be the irreducible representation of the group $SL(2, \mathbb{C})$ of degree $p$. We then represent the automorphic representation $\pi$ by the irreducible $n$-dimensional representation

$$\psi = \mu \otimes \nu : \ W_F \times SL(2, \mathbb{C}) \ \longrightarrow \ GL(n, \mathbb{C})$$

of the product of $W_F$ with $SL(2, \mathbb{C})$. Set

$$\phi_\psi(w) = \psi\left(w, \begin{pmatrix} |w|^{\frac{1}{2}} & 0 \\ 0 & |w|^{-\frac{1}{2}} \end{pmatrix}\right), \qquad w \in W_F,$$

where $|w|$ is the canonical absolute value on $W_F$. By comparing the unramified constituents of $\pi$ with the unramified images of Frobenius classes in $W_F$, we see that $\pi$ is the automorphic representation corresponding to the $n$-dimensional representation $\phi_\psi$ of $W_F$. Thus, according to functoriality, there is a mapping $\psi \to \pi$, from the set of irreducible $n$-dimensional representations of $W_F \times SL(2,\mathbb{C})$ whose restriction to $W_F$ is bounded, into the set of automorphic representations $\pi$ of $GL(n)$ that occur in the discrete spectrum. Removing the condition that $\psi$ be irreducible gives rise to representations $\pi$ that occur in the general spectrum.

## 27. The problem of stability

We return to the general trace formula. The invariant trace formula of Theorem 23.4 still has one serious deficiency. The invariant distributions on each side are not usually stable. We shall discuss the notion of stability, and why it is an essential consideration in any general attempt to compare trace formulas on different groups.

Stability was discovered by Langlands in attempting to understand how to generalize the Jacquet-Langlands correspondence. We discussed the extension of this correspondence from $GL(2)$ to $GL(n)$ in §25, but it is for groups other than $GL(n)$ that the problems arise. Suppose then that $G$ is an arbitrary connected reductive group over our number field $F$. We fix an inner twist

$$\psi:\ G \longrightarrow G^*,$$

where $G^*$ is a quasisplit reductive group over $F$. One would like to establish the reciprocity laws between automorphic representations of $G$ and $G^*$ predicted by functoriality.

To use the trace formula, we would start with a test function $f$ for $G$. For the time being, we take $f$ to be a function in $C_c^\infty\big(G(\mathbb{A})^1\big)$, which we assume is the restriction of a product of functions

$$\prod_v f_v, \qquad\qquad f_v \in C_c^\infty\big(G(F_v)\big).$$

If we were to follow the prescription of Jacquet-Langlands, we would map $f$ to a function $f^*$ on $G(\mathbb{A})^1$ obtained by restriction of a product $\prod f_v^*$ of functions on the local groups $G^*(F_v)$. Each function $f_v^*$ would be attached to the associated function $f_v$ on $G(F_v)$ by imposing a matching condition for the local invariant orbital integrals of $f_v$ and $f_v^*$. This would in turn require a correspondence $\gamma_v \to \gamma_v^*$ between strongly regular conjugacy classes. How might such a correspondence be defined in general?

In the special case discussed in §25, the correspondence of strongly regular elements can be formulated explicitly in terms of characteristic polynomials. For any $k \supset F$, one matches a characteristic polynomial on the matrix algebra $M_n(k)$ with its variant for the central simple algebra that defines $G$. Now the coefficients of characteristic polynomials have analogues for the general group $G$. For example, one can take any set of generators of the algebra of $G$-invariant polynomials on $G$. These objects can certainly be used to transfer semisimple conjugacy classes from $G$ to $G^*$. However, invariant polynomials measure only *geometric* conjugacy classes, that is, conjugacy classes in the group of points over an algebraically closed field. In general, if $k$ is not algebraically closed, and $G$ is just about any group other than $GL(n)$ (or one of its inner twists), there can be nonconjugate elements in $G(k)$ that are conjugate over an algebraic closure $G(\overline{k})$. For example, in the case that $G = SL(2)$ and $k = \mathbb{R}$, the relation

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$$

represents conjugacy over $G(\mathbb{C})$ of nonconjugate elements in $G(\mathbb{R})$. This phenomenon obviously complicates the problem of transferring conjugacy classes.

Langlands defined two strongly regular elements in $G(k)$ to be *stably conjugate* if they were conjugate over an algebraic closure $G(\overline{k})$. Stable conjugacy is thus an equivalence relation that is weaker than conjugacy. Suppose that $\delta$ belongs to the set $\Delta_{\mathrm{reg}}\big(G(k)\big)$ of (strongly regular) stable conjugacy classes in $G(k)$. The image $\psi(\delta)$ of $\delta$ in $G^*$ yields a well defined conjugacy class in $G(\overline{k})$. If $\sigma$ belongs to $\mathrm{Gal}(\overline{k}/k)$,

$$\sigma\big(\psi(\delta)\big) = \sigma(\psi)\sigma(\delta) = \alpha(\sigma)^{-1}\psi(\delta),$$

for the inner automorphism $\alpha(\sigma) = \psi \circ \sigma(\psi)^{-1}$ of $G^*$. The geometric conjugacy class of $\psi(\delta)$ is therefore defined over $k$. Because $G^*$ is quasisplit and $\psi(\delta)$ is semisimple, an important theorem of Steinberg [**Ste**] implies that the geometric conjugacy class has a representative in $G(k)$. This representative is of course not unique, but it does map to a well defined stable conjugacy class $\delta^* \in \Delta_{\mathrm{reg}}\big(G^*(k)\big)$. We therefore have an injection $\delta \to \delta^*$ from $\Delta_{\mathrm{reg}}\big(G(k)\big)$ to $\Delta_{\mathrm{reg}}\big(G^*(k)\big)$, determined canonically by $\psi$. (The fact that Steinberg's theorem holds only for quasisplit groups is responsible for the mapping not being surjective.) We cannot however expect to be able to transfer ordinary conjugacy classes $\gamma \in \Gamma_{\mathrm{reg}}\big(G(k)\big)$ from $G$ to $G^*$.

Besides the subtle global questions it raises for the trace formula, stable conjugacy also has very interesting implications for local harmonic analysis. Suppose that $k$ is one of the local fields $F_v$. In this case, there are only finitely many conjugacy classes in any stable class. One defines the stable orbital integral of a function $f_v \in C_c^\infty\big(G(F_v)\big)$ over a (strongly regular) stable conjugacy class $\delta_v$ as a finite sum

$$f_v^G(\delta_v) = \sum_{\gamma_v} f_{v,G}(\gamma_v)$$

of invariant orbital integrals, taken over the conjugacy classes $\gamma_v$ in the stable class $\delta_v$. (It is not hard to see how to choose compatible invariant measures on the various domains $G_{\gamma_v}(F_v)\backslash G(F_v)$.) An invariant distribution $S_v$ on $G(F_v)$ is said to be *stable* if its value at $f_v$ depends only on the set of stable orbital integrals $\{f_v^G(\delta_v)\}$ of $f_v$. Under this condition, there is a continuous linear form $\widehat{S}_v$ on the space of functions

$$S\mathcal{I}\big(G(F_v)\big) = \big\{f_v^G : \ f_v \in \mathcal{H}\big(G(F_v)\big)\big\}$$

on $\Delta_{\mathrm{reg}}\big(G(F_v)\big)$ such that

$$S_v(f_v) = \widehat{S}_v(f_v^G), \qquad\qquad f_v \in \mathcal{H}\big(G(F_v)\big).$$

We thus have a whole new class of distributions on $G(F_v)$, which is more restrictive than the family of invariant distributions. Is there some other way to characterize it?

In general terms, one becomes accustomed to thinking of conjugacy classes as being dual to irreducible characters. From the perspective of local harmonic analysis, the semisimple conjugacy classes in $G(F_v)$ could well be regarded as dual analogues of irreducible tempered characters on $G(F_v)$. The relation of stable conjugacy ought then to determine a parallel relationship on the set of tempered characters. (In [**Ko1**], Kottwitz extended the notion of stable conjugacy to arbitrary semisimple elements.) Langlands called this hypothetical relationship *$L$-equivalence*, and referred to the corresponding equivalence classes as *$L$-packets*, since they seemed to preserve the local $L$-functions and $\varepsilon$-factors attached to irreducible representations of $G(F_v)$. He also realized that in the case $F_v = \mathbb{R}$, there was already a good candidate for this relationship in the work of Harish-Chandra.

Recall from §24 that $G(\mathbb{R})$ has a discrete series if and only if $G$ has an elliptic maximal torus $T_G$ over $\mathbb{R}$. In this case, the discrete series occur in finite packets $\Pi_2(\mu)$. On the other hand, any strongly regular elliptic conjugacy class for $G(\mathbb{R})$ intersects $T_{G,\mathrm{reg}}(\mathbb{R})$. Moreover, two elements in $T_{G,\mathrm{reg}}(\mathbb{R})$ are $G(\mathbb{R})$-conjugate if and only if they lie in the same $W(K_{\mathbb{R}}, T_G)$-orbit, and are stably conjugate if and only if they are in the same orbit under the full Weyl group $W(G, T_G)$. This is because $W(K_{\mathbb{R}}, T_G)$ is the subgroup of elements in $W(G, T_G)$ that are actually induced by conjugation from points in $G(\mathbb{R})$. It can then be shown from Harish-Chandra's algorithm for the characters of discrete series that the sum of characters

$$\Theta(\mu, \gamma) = \sum_{\pi_{\mathbb{R}} \in \Pi_2(\mu)} \Theta(\pi_{\mathbb{R}}, \gamma), \qquad \gamma \in G_{\mathrm{reg}}(\mathbb{R}),$$

attached to representations in a packet $\Pi_2(\mu)$, depends only on the stable conjugacy class of $\gamma$, rather than its actual conjugacy class. In other words, the distribution

$$f_{\mathbb{R}} \longrightarrow \sum_{\pi_{\mathbb{R}} \in \Pi_2(\mu)} f_{\mathbb{R},G}(\pi_{\mathbb{R}}), \qquad f_{\mathbb{R}} \in C_c^{\infty}\big(G(\mathbb{R})\big),$$

on $G(\mathbb{R})$ is stable. This fact justifies calling $\Theta(\mu, \gamma)$ a "stable character", and designating the sets $\Pi_2(\mu)$ the $L$-packets of discrete series. It also helps to explain why the sum over $\pi_{\mathbb{R}} \in \Pi_2(\mu)$, which occurs on each side of the "finite case" of the trace formula in Theorem 24.1, is a natural operation. Langlands used the $L$-packet structure of discrete series as a starting point for a classification of the irreducible representations of $G(\mathbb{R})$, and a partition of the representations into $L$-packets governed by their local $L$-functions [**Lan11**]. (Knapp and Zuckerman [**KZ2**] later determined the precise structure of the $L$-packets outside the discrete series.) The Langlands classification for real groups applies to all irreducible representations, but it is only for the tempered representations that the sum of the characters in an $L$-packet is stable.

Let us return to the invariant trace formula. The basic questions raised by the problem of stability can be posed for the simplest terms on the geometric side. Let $\Gamma_{\mathrm{reg,ell}}(G)$ be the set of conjugacy classes $\gamma$ in $G(F)$ that are both strongly regular and elliptic. An element $\gamma \in G(F)$ represents a class in $\Gamma_{\mathrm{reg,ell}}(G)$ if and only if the centralizer $G_\gamma$ is a maximal torus in $G$ that is elliptic, in the usual sense that $A_{G_\gamma} = A_G$. It follows from the definitions that

$$\Gamma_{\mathrm{reg,ell}}(G) \subset \Gamma_{\mathrm{anis}}(G) \subset \Gamma(G)_S.$$

The elements in $\Gamma_{\mathrm{reg,ell}}(G)$ are in some sense the generic elements in the set $\Gamma(G)_S$, which we recall indexes the terms in the sum with $M = G$ on the geometric side. The *regular elliptic* part

$$I_{\mathrm{reg,ell}}(f) = \sum_{\gamma \in \Gamma_{\mathrm{reg,ell}}(G)} a^G(\gamma) f_G(\gamma) \tag{27.1}$$

of the trace formula therefore represents the generic part of this sum.

The first question that comes to mind is the following. Is the distribution

$$f \longrightarrow I_{\mathrm{reg,ell}}(f), \qquad\qquad f \in C_c^\infty\big(G(\mathbb{A})^1\big),$$

stable? In other words, does $I_{\mathrm{reg,ell}}(f)$ depend only on the family of stable orbital integrals $\{f_v^G(\delta_v)\}$? An affirmative answer could solve many of the global problems created by stability. For in order to compare $I_{\mathrm{reg,ell}}$ with its analogue on $G^*$, it would then only be necessary to transfer $f$ to a function $f^*$ on $\Delta_{G\text{-reg}}\big(G^*(\mathbb{A})\big)$, something we could do by the local correspondence of stable conjugacy classes.

A cursory glance seems to suggest that the answer is indeed affirmative. The volume

$$a^G(\gamma) = \mathrm{vol}\big(G_\gamma(\mathbb{Q})\backslash G_\gamma(\mathbb{A})^1\big), \qquad\qquad \gamma \in \Gamma_{\mathrm{reg,ell}}(G),$$

depends only on the stable conjugacy class $\delta$ of $\gamma$ in $G(F)$, since it depends only on the $F$-isomorphism class of the maximal torus $G_\gamma$. We can therefore write

$$I_{\mathrm{reg,ell}}(f) = \sum_\delta a^G(\delta)\Big(\sum_{\gamma\to\delta} f_G(\gamma)\Big), \tag{27.2}$$

where $\delta$ is summed over the set $\Delta_{\mathrm{reg,ell}}(G)$ of elliptic stable classes in $\Delta_{\mathrm{reg}}\big(G(F)\big)$, $\gamma$ is summed over the preimage of $\delta$ in $\Gamma_{\mathrm{reg,ell}}(G)$, and $a^G(\delta) = a^G(\gamma)$. The sum over $\gamma$ looks as if it might be stable in $f$. However, a closer inspection reveals that it is not. For we are demanding that the distribution be stable in each component $f_v$ of $f$. If

$$\delta_{\mathbb{A}} = \prod_{v\in S}\delta_v, \qquad\qquad \delta_v \in \Delta_{\mathrm{reg}}\big(G(F_v)\big),$$

is a product of local stable classes with a rational representative $\delta$, each ordinary conjugacy class $\gamma_{\mathbb{A}} = \prod \gamma_v$ in $\delta_{\mathbb{A}}$ would also have to have a rational representative $\gamma$. It turns out that there are not enough rational conjugacy classes $\gamma$ for this to happen. Contrary to our initial impression then, the distribution $I_{\mathrm{reg,ell}}(f)$ is not generally stable in $f$.

Since $I_{\mathrm{reg,ell}}(f)$ need not be stable, the question has to be reformulated in terms of stabilizing this distribution. The problem may be stated in general terms as follows.

*Express $I_{\mathrm{reg,ell}}(f)$ as the sum of a canonical stable distribution $S^G_{\mathrm{reg,ell}}(f)$ with an explicit error term.*

The first group to be investigated was $SL(2)$. Labesse and Langlands stabilized the full trace formula for this group, as well as for its inner forms, and showed that the solution had remarkable implications for the corresponding spectral decompositions [**Lab1**], [**She1**], [**LL**]. Langlands also stabilized $I_{\mathrm{reg,ell}}$ in the general case, under the assumption of two conjectures in local harmonic analysis [**Lan10**].

In his general stabilization of $I_{\mathrm{reg,ell}}(f)$, Langlands constructed the stable component $S^G_{\mathrm{reg,ell}}$ explicitly. He expressed the error term in terms of corresponding stable components attached to groups $G'$ of dimension smaller than $G$. The groups $\{G'\}$ are all quasisplit. Together with the group $G' = G^*$ of dimension equal to $G$, they are known as *elliptic endoscopic groups* for $G$. For each $G'$, Langlands formulated a conjectural correspondence $f \to f'$ between test functions for $G$ and $G'$. His stabilization then took the form

$$I_{\mathrm{reg,ell}}(f) = \sum_{G'} \iota(G, G') \widehat{S}'_{G\text{-reg,ell}}(f'), \tag{27.3}$$

for explicitly determined coefficients $\iota(G, G')$, and stable linear forms $S'_{G\text{-reg,ell}}$ attached to $G'$. In case $G' = G^*$, the corresponding terms satisfy $\iota(G, G') = 1$ and $S^*_{G\text{-reg,ell}} = S^*_{\mathrm{reg,ell}}$. The stable component of $I_{\mathrm{reg,ell}}(f)$ is the associated summand

$$S^G_{\mathrm{reg,ell}}(f) = \widehat{S}^*_{\mathrm{reg,ell}}(f^*).$$

For arbitrary $G'$, $S'_{G\text{-reg,ell}}$ is the strongly $G$-regular part of $S'_{\mathrm{reg,ell}}$, obtained from classes in $\Delta_{\mathrm{reg,ell}}(G')$ whose image in $G$ remains strongly regular.

Langlands' stabilization is founded on class field theory. Specifically, it depends on the application of Tate-Nakayama duality to the Galois cohomology of algebraic groups. The basic relationship is easy to describe. Suppose that $\delta$ is a strongly regular element in $G(k)$, for some $k \supset F$. The centralizer of $\delta$ in $G$ is a maximal torus $T$ over $k$. Suppose that $\gamma \in G(k)$ is stably conjugate to $\delta$. Then $\gamma$ equals $g^{-1}\delta g$, for some element $g \in G(\overline{k})$. If $\sigma$ belongs to $\mathrm{Gal}(\overline{k}/k)$, we have

$$\delta = \sigma(\delta) = \sigma(g\gamma g^{-1}) = \sigma(g)\gamma\sigma(g)^{-1} = t(\sigma)^{-1}\delta t(\sigma),$$

where $t(\sigma)$ is the 1-cocycle $g\sigma(g)^{-1}$ from $\mathrm{Gal}(\overline{k}/k)$ to $T(\overline{k})$. One checks that a second element $\gamma_1 \in G(k)$ in the stable class of $\delta$ is $G(k)$-conjugate at $\gamma$ if and only if the corresponding 1-cocycle $t_1(\sigma)$ has the same image as $t(\sigma)$ in the Galois cohomology group

$$H^1(k, T) = H^1\big(\Gamma_k, T(\overline{k})\big), \qquad \Gamma_k = \mathrm{Gal}(\overline{k}/k).$$

Conversely, an arbitrary class in $H^1(k, T)$ comes from an element $\gamma$ if and only if it is represented by a 1-cocycle of the form $g\sigma(g)^{-1}$. The mapping $\gamma \to t$ therefore defines a bijection from the set of $G(k)$-conjugacy classes in the stable conjugacy class of $\delta$ to the kernel

$$\mathcal{D}(T) = \mathcal{D}(T/k) = \ker\big(H^1(k, T) \to H^1(k, G)\big). \tag{27.4}$$

Keep in mind that $H^1(k, G)$ is only a set with distinguished element 1, since $G$ is generally nonabelian. The preimage $\mathcal{D}(T)$ of this element in $H^1(k, T)$ therefore need not be a subgroup. However, $\mathcal{D}(T)$ is contained in the subgroup

$$\mathcal{E}(T) = \mathcal{E}(T/k) = \mathrm{im}\big(H^1(k, T_{\mathrm{sc}}) \to H^1(k, T)\big)$$

of $H^1(k, T)$, where $T_{\mathrm{sc}}$ is the preimage of $T$ in the simply connected cover $G_{\mathrm{sc}}$ of the derived group of $G$. This is because the canonical map $\mathcal{D}(T_{\mathrm{sc}}) \to \mathcal{D}(T)$ is surjective. If $H^1(k, G_{\mathrm{sc}}) = \{1\}$, which is the case whenever $k$ is a nonarchimedean local field [**Spr1**, §3.2], $\mathcal{D}(T)$ actually equals the subgroup $\mathcal{E}(T)$. This is one of the reasons why one works with the groups $\mathcal{E}(T)$ in place of $H^1(T, G)$, and why the simply connected group $G_{\mathrm{sc}}$ plays a significant role in the theory.

In the case that $k$ is a local or global field, Tate-Nakayama duality applies class field theory to the groups $H^1(k, T)$. If $k$ is a completion $F_v$ of $F$, it provides a canonical isomorphism

$$H^1(F_v, T) \xrightarrow{\sim} \pi_0(\widehat{T}^{\Gamma_v})^*$$

of $H^1(F_v,T)$ with the group of characters on the finite abelian group $\pi_0(\widehat{T}^{\Gamma_v})$. We have written $\Gamma_v$ here for the Galois group $\Gamma_{F_v} = \mathrm{Gal}(\overline{F}_v/F_v)$, which acts on the complex dual torus

$$\widehat{T} = X(T) \otimes \mathbb{C}^*$$

through its action on the group of rational characters $X(T)$. As usual, $\pi_0(\cdot)$ denotes the set of connected components of a topological space. If $k = F$, Tate-Nakayama duality characterizes the group

$$H^1\big(F, T(\overline{\mathbb{A}})/T(\overline{F})\big) = H^1\big(\Gamma_F, T(\mathbb{A}_{\bar{F}})/T(\overline{F})\big) = H^1\big(\Gamma_{F'/F}, T(\mathbb{A}_{F'})/T(F')\big),$$

where $F'$ is some finite Galois extension of $F$ over which $T$ splits. It provides a canonical isomorphism

$$H^1\big(F, T(\overline{\mathbb{A}})/T(\overline{F})\big) \stackrel{\sim}{\longrightarrow} \pi_0(\widehat{T}^{\Gamma})^*,$$

where the Galois group $\Gamma = \Gamma_F = \mathrm{Gal}(\overline{F}/F)$ again acts on the complex torus $\widehat{T}$ through its action on $X(T)$. If we combine this with the long exact sequence of cohomology attached to the exact sequence of $\Gamma$-modules

$$1 \longrightarrow T(\overline{F}) \longrightarrow T(\overline{\mathbb{A}}) \longrightarrow T(\overline{\mathbb{A}})/T(\overline{F}) \longrightarrow 1,$$

and the isomorphism

$$H^1\big(F, T(\overline{\mathbb{A}})\big) \cong \bigoplus_v H^1(F_v, T)$$

provided by Shapiro's lemma, we obtain a characterization of the diagonal image of $H^1(F,T)$ in the direct sum over $v$ of the groups $H^1(F_v,T)$. It is given by a canonical isomorphism from the cokernel

$$\mathrm{coker}^1(F,T) = \mathrm{coker}\Big(H^1(F,T) \longrightarrow \bigoplus_v H^1(F_v,T)\Big) \tag{27.5}$$

onto the image

$$\mathrm{im}\Big(\bigoplus_v \pi_0(\widehat{T}^{\Gamma_v})^* \longrightarrow \pi_0(\widehat{T}^{\Gamma})^*\Big).$$

If these results are combined with their analogues for $T_{\mathrm{sc}}$, they provide similar assertions for the subgroups $\mathcal{E}(T/k)$ of $H^1(k,T)$. In the local case, one has only to replace $\pi_0(\widehat{T}^{\Gamma_v})$ by the group $\mathcal{K}(T/F_v)$ of elements in $\pi_0\big(\big(\widehat{T}/Z(\widehat{G})\big)^{\Gamma_v}\big)$ whose image in $H^1\big(F_v, Z(\widehat{G})\big)$ is trivial. In the global case, one replaces $\pi_0(\widehat{T}^{\Gamma})$ by the group $\mathcal{K}(T/F)$ of elements in $\pi_0\big(\big(\widehat{T}/Z(\widehat{G})\big)^{\Gamma}\big)$ whose image in $H^1\big(F, Z(\widehat{G})\big)$ is locally trivial, in the sense that their image in $H^1\big(F_v, Z(\widehat{G})\big)$ is trivial for each $v$. (See [**Lan10**], [**Ko5**].)

To simplify the discussion, assume for the present that $G = G_{\mathrm{sc}}$. Then $\mathcal{E}(T/k) = H^1(k,T)$, for any $k$. Moreover, $\mathcal{K}(T/F_v) = \pi_0(\widehat{T}^{\Gamma_v})$ and $\mathcal{K}(T/F) = \pi_0(\widehat{T}^{\Gamma})$, since $Z(\widehat{G}) = 1$. In fact, $\pi_0(\widehat{T}^{\Gamma})$ equals $\widehat{T}^{\Gamma}$ if $T$ is elliptic in $G$ over $F$.

We recall that Langlands' stabilization (27.3) of $I_{\mathrm{reg,ell}}(f)$ was necessitated by the failure of each $G(\mathbb{A})$-conjugacy class in the $G(\mathbb{A})$-stable class of $\delta \in \Delta_{\mathrm{reg,ell}}(G)$ to have a representative in $G(F)$. The cokernel (27.5) gives a measure of this failure. Langlands' construction treats the quantity in brackets on the right hand side of (27.2) as the value at 1 of a function on the finite abelian group $\mathrm{coker}^1(F,T)$. The critical step is to expand this function according to Fourier inversion on

$\mathrm{coker}^1(F,T)$. One has to keep track of the $G(F)$-conjugacy classes in the $G(\mathbb{A})$-conjugacy class of $\delta$, which by the Hasse principle for $G = G_{\mathrm{sc}}$ are in bijection with the finite abelian group

$$\ker^1(F,T) = \ker\Big(H^1(F,T) \ \longrightarrow\ \bigoplus_v H^1(F_v,T)\Big).$$

The formula (27.2) becomes an expansion

$$I_{\mathrm{reg,ell}}(f) = \sum_{\delta\in\Delta_{\mathrm{reg,ell}}(G)} a^G(\delta)\iota(T)\sum_{\kappa\in\widehat{T}^\Gamma} f_G^\kappa(\delta), \tag{27.6}$$

where $T = G_\delta$ denotes the centralizer of (some fixed representative of) $\delta$, $\iota(\tau)$ equals the product of $|\widehat{T}^\Gamma|^{-1}$ with $|\ker^1(F,T)|$, and

$$f_G^\kappa(\delta) = \sum_{\{\gamma_{\mathbb{A}}\in\Gamma(G(\mathbb{A})):\gamma_{\mathbb{A}}\sim\delta\}} f_G(\gamma_{\mathbb{A}})\kappa(\gamma_{\mathbb{A}}).$$

The last sum is of course over the $G(\mathbb{A})$-conjugacy classes $\gamma_{\mathbb{A}} = \prod\gamma_v$ in the stable class of $\delta$ in $G(\mathbb{A})$. For any such $\gamma_{\mathbb{A}}$, it can be shown that $\gamma_v$ is $G(F_v)$-conjugate to $\delta$ for almost all $v$. It follows that $\gamma_{\mathbb{A}}$ maps to an element $t_{\mathbb{A}} = \bigoplus t_v$ in the direct sum of the groups $H^1(F_v,T)$. This in turn maps to a point in the cokernel (27.5), and hence to a character in $(\widehat{T}^\Gamma)^*$. The coefficient $\kappa(\gamma_{\mathbb{A}})$ is the value of this character at $\kappa$.

Suppose for example that $G = SL(2)$. The eigenvalues of $\delta$ then lie in a quadratic extension $E$ of $F$, and $T = G_\delta$ is the one-dimensional torus over $F$ such that

$$T(F) \cong \{t\in E^*:\ t\sigma(t) = 1\}, \qquad\qquad \Gamma_{E/F} = \{1,\sigma\}.$$

The nontrivial element $\sigma\in\Gamma_{E/F}$ acts on $X(T)\cong\mathbb{Z}$ by $m\to(-m)$, and therefore acts on $\widehat{T} = \mathbb{Z}\otimes\mathbb{C}^*\cong\mathbb{C}^*$ by $z\to z^{-1}$. It follows that $\pi_0(\widehat{T}^\Gamma) = \widehat{T}^\Gamma$ is isomorphic to the subgroup $\{\pm1\}$ of $\mathbb{C}^*$. Similarly, $\pi_0(\widehat{T}^{\Gamma_v}) = \widehat{T}^{\Gamma_v}\cong\{\pm1\}$ if $v$ does not split in $E$, while $\pi_0(\widehat{T}^{F_v}) = \pi_0(\widehat{T}) = \{1\}$ if $v$ does split. In particular, if $\kappa$ is the nontrivial element in $\pi_0(\widehat{T}^\Gamma)$, the local $\kappa$-orbital integral $f_{v,G}^\kappa(\delta) = f_{v,G}^\kappa(\delta_v)$ equals a difference of two orbital integrals if $v$ does not split, and is a simple orbital integral otherwise. The characterization we have described here for the various groups $H^1(F_v,T)$, and for the diagonal image of $H^1(F,T)$ in their direct sum, is typical of what happens in general. In the present situation $\ker^1(F,T) = \{1\}$, so that $H^1(F,T)$ can in fact be identified with its diagonal image.

The expression (27.6) is part of the stabilization (27.3) of $I_{\mathrm{reg,ell}}(f)$. We need to see how it gives rise to the quasisplit groups $G'$ of (27.3).

Suppose that $T$ and $\kappa$ are as in (27.6). We choose an embedding $\widehat{T}\subset\widehat{G}$ of the dual torus of $T$ into $\widehat{G}$ that is admissible, in the sense that it is the mapping assigned to a choice of some pair $(\widehat{B},\widehat{T})$ in $\widehat{G}$, and some Borel subgroup $B$ of $G$ containing $T$. Let $s'$ be the image of $\kappa$ in $\widehat{G}$, and let $\widehat{G}' = \widehat{G}_{s'}$ be its connected centralizer in $\widehat{G}$. Then $\widehat{G}'$ is a reductive subgroup of $\widehat{G}$. It is known that there is an $L$-embedding

$${}^LT = \widehat{T}\rtimes W_F \ \hookrightarrow\ {}^LG = \widehat{G}\rtimes W_F,$$

for the Weil forms of the $L$-groups of $T$ and $G$, which restricts to the given embedding of $\widehat{T}$ into $\widehat{G}$ [**LS1**, (2.6)]. Fix such an embedding, and set

$$\mathcal{G}' \ = \ {}^L T \widehat{G}'.$$

Then $\mathcal{G}'$ is an $L$-subgroup of ${}^L G$, which commutes with $s'$. It provides a split extension

$$1 \longrightarrow \widehat{G}' \longrightarrow \mathcal{G}' \longrightarrow W_F \longrightarrow 1 \tag{27.7}$$

of $W_F$ by $\widehat{G}'$. In particular, it determines an action of $W_F$ on $\widehat{G}'$ by outer automorphisms, which factors through a finite quotient of $\Gamma_F$. Let $G'$ be any quasisplit group over $F$ for which $\widehat{G}'$, with the given action of $\Gamma_F$, is a dual group. We have obtained a correspondence

$$(T, \kappa) \ \longrightarrow \ (G', \mathcal{G}', s').$$

We can choose a maximal torus $T' \subset G'$ in $G'$ over $F$, together with an isomorphism from $T'$ to $T$ over $F$ that is admissible, in the sense that the associated isomorphism $\widehat{T}' \to \widehat{T}$ of dual groups is the composition of an admissible embedding $\widehat{T}' \subset \widehat{G}'$ with an inner automorphism of $\widehat{G}$ that takes $\widehat{T}'$ to $\widehat{T}$. Let $\delta' \in T'(F)$ be the associated preimage of the original point $\delta \in T(F)$. The tori $T$ and $T'$ are the centralizers in $G$ and $G'$ of $\delta$ and $\delta'$. The two points $\delta$ and $\delta'$ are therefore the primary objects. They become part of a larger correspondence

$$(\delta, \kappa) \ \longrightarrow \ \big((G', \mathcal{G}', s'), \delta'\big). \tag{27.8}$$

Elements $\delta' \in G'(F)$ obtained in this way are said to be *images* from $G$ [**LS1**, (1.3)].

Suppose now that $G$ is arbitrary. Motivated by the last construction, one defines an *endoscopic datum* for $G$ to be a triplet $(G', \mathcal{G}', s', \xi')$, where $G'$ is a quasisplit group over $F$, $\mathcal{G}'$ is a split extension of $W_F$ by a dual group $\widehat{G}'$ of $G'$, $s'$ is a semisimple element in $\widehat{G}$, and $\xi'$ is an $L$-embedding of $\mathcal{G}'$ into ${}^L G$. It is required that $\xi'(\widehat{G}')$ be equal to the connected centralizer of $s'$ in $\widehat{G}$, and that

$$\xi'(u')s' = s'\xi'(u')a(u'), \qquad u' \in \mathcal{G}', \tag{27.9}$$

where $a$ is a 1-cocycle from $W_F$ to $Z(\widehat{G})$ that is locally trivial, in the sense that its image in $H^1\big(W_{F_v}, Z(\widehat{G})\big)$ is trivial for every $v$. The quasisplit group $G'$ is called an *endoscopic group* for $G$. An *isomorphism* of endoscopic data $(G', \mathcal{G}', s', \xi')$ and $(G'_1, \mathcal{G}'_1, s'_1, \xi'_1)$ is an isomorphism $\alpha$: $G' \to G'_1$ over $F$ for which, roughly speaking, there is dual isomorphism induced by some element in $\widehat{G}$. More precisely, it is required that there be an $L$-isomorphism $\beta$: $\mathcal{G}'_1 \to \mathcal{G}'$ such that the corresponding mappings $\Psi(G') \xrightarrow{\alpha} \Psi(G'_1)$ and $\Psi(\widehat{G}'_1) \xrightarrow{\beta} \Psi(\widehat{G}')$ of based root data are dual, and an element $g \in \widehat{G}$ such that

$$\xi'\big(\beta(u'_1)\big) = g^{-1}\xi'_1(u'_1)g, \qquad u'_1 \in \mathcal{G}'_1,$$

and

$$s' = g^{-1}s'_1 g z, \qquad z \in Z(\widehat{G})Z(\xi'_1)^0,$$

where $Z(\xi'_1)^0$ is the connected component of 1 in the centralizer in $\widehat{G}$ of $\xi'_1(\mathcal{G}'_1)$. (See [**LS1**, (1.2)].) We write $\mathrm{Aut}_G(G')$ for the group of isomorphisms $\alpha$: $G' \to G'$ of $G'$ as a endoscopic datum for $G$.

We say that $(G', \mathcal{G}', s', \xi')$ is *elliptic* if $Z(\xi')^0 = 1$. This means that the image of $\xi'$ is ${}^L G$ is not contained in ${}^L M$, for any proper Levi subgroup of $G$ over $F$. We write $\mathcal{E}_{\text{ell}}(G)$ for the set of isomorphism classes of elliptic endoscopic data for $G$. It is customary to denote an element in $\mathcal{E}_{\text{ell}}(G)$ by $G'$, even though $G'$ is really only the first component of a representative $(G', \mathcal{G}', s', \xi')$ of an isomorphism class. Any $G' \in \mathcal{E}_{\text{ell}}(G)$ then comes with a finite group

$$\text{Out}_G(G') = \text{Aut}_G(G')/\text{Int}(G'),$$

of outer automorphisms of $G'$ as an endoscopic datum.

Suppose for example that $G = GL(n)$. The centralizer of any semisimple element $s'$ in $\widehat{G} = GL(n, \mathbb{C})$ is a product of general linear groups. It follows that any endoscopic datum for $G$ is represented by a Levi subgroup $M$. In particular, there is only one element in $\mathcal{E}_{\text{ell}}(G)$, namely the endoscopic datum represented by $G$ itself. This is why the problem of stability is trivial for $GL(n)$.

The general definitions tend to obscure the essential nature of the construction. Suppose again that $G = G_{\text{sc}}$. The dual group $\widehat{G}$ is then adjoint, and $Z(\widehat{G}) = 1$. In general, any $G' \in \mathcal{E}_{\text{ell}}(G)$ can be represented by an endoscopic datum for which $\mathcal{G}'$ is a subgroup of ${}^L G$, and $\xi'$ is the identity embedding $\iota'$. The condition (27.9) reduces in the case at hand to the requirement that $\mathcal{G}'$ commute with $s'$. To construct a general element in $\mathcal{E}_{\text{ell}}(G)$, we start with the semisimple element $s' \in \widehat{G}$. The centralizer ${}^L G_{s',+}$ of $s'$ in ${}^L G$ is easily seen to project onto $W_F$. Its quotient by the connected centralizer $\widehat{G}' = \widehat{G}_{s'}$ is an extension of $W_F$ by a finite group. To obtain an endoscopic datum, we need only choose a section

$$\omega' : \; W_F \;\longrightarrow\; {}^L G_{s',+}/\widehat{G}'$$

that can be inflated to a homomorphism $W_F \to {}^L G_{s',+}$. For the product

$$\mathcal{G}' = \widehat{G}'\omega'(W_F)$$

is then a split extension of $W_F$ by $\widehat{G}'$. It determines an $L$-action of $W_F$ on $\widehat{G}'$, and hence a quasisplit group $G'$ over $F$ of which $\widehat{G}'$ is a dual group. The endoscopic datum $(G', \mathcal{G}', s', \iota')$ thus obtained is elliptic if and only if the centralizer of $\mathcal{G}'$ in $\widehat{G}$ is finite, a condition that reduces considerably the possibilities for the pairs $(s', \omega')$. The mapping

$$(s', \omega') \;\longrightarrow\; (G', \mathcal{G}', s', \iota')$$

becomes a bijection from the set of $\widehat{G}$-orbits of such pairs and $\mathcal{E}_{\text{ell}}(G)$. We note that a point $g \in \widehat{G}$ represents an element in $\text{Out}_G(G')$ if and only if it stabilizes $\mathcal{G}'$ and commutes with $s'$.

For purposes of illustration, suppose that $G$ is split as well as being simply connected. We have then to consider semisimple elements $s' \in \widehat{G}$ whose centralizer $\widehat{G}_{s',+}$ has finite center. It is an interesting exercise (which I confess not to have completed) to classify the $\widehat{G}$-orbits of such elements in terms of the extended Coxeter-Dynkin diagram of $\widehat{G}$. For example, elements $s'$ that satisfy the stronger condition that $\widehat{G}' = \widehat{G}_{s'}$ has finite center are represented by vertices in the affine diagram (although in the adjoint group $\widehat{G}$, some of these elements are conjugate). Once we have chosen $s'$, we then select a homomorphism $\omega'$ from $\Gamma_F$ to the finite abelian group $\pi_0(\widehat{G}_{s',+}) = \widehat{G}_{s',+}/\widehat{G}'$ whose image pulls back to a subgroup of $\widehat{G}_{s',+}$ that still has finite center. Suppose for example that $G = SL(2)$, and that $s'$ is the image of $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ in $\widehat{G} = PGL(2, \mathbb{C})$. Then $\widehat{G}_{s',+}$ consists of the group of diagonal matrices, together with a second component generated by the element $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Since the center of $\widehat{G}_{s',+}$ equals $\{1, s'\}$, we obtain elliptic endoscopic data for $G$ by choosing nontrivial homomorphisms from $\Gamma_F$ to the group $\pi_0(\widehat{G}_{s',+}) \cong \mathbb{Z}/2\mathbb{Z}$. The classes in $\mathcal{E}_{\text{ell}}(G)$ other than $G$ itself, are thus parametrized by quadratic extensions $E$ of $F$.

We return to the case of a general group $G$. We have spent most of this section trying to motivate some of the new ideas that arose with the problem of stability. This leaves only limited space for a brief description of the details of Langlands' stabilization (27.3) of $I_{\text{reg,ell}}(f)$.

The general form of the expansion (27.6) is

$$I_{\text{reg,ell}}(f) = \sum_{\delta \in \Delta_{\text{reg,ell}}(G)} a^G(\delta) \iota(T, G) \sum_{\kappa \in \mathcal{K}(T/F)} f_G^\kappa(\delta),$$

where

$$\iota(T, G) = \Big| \ker \Big( \mathcal{E}(T/F) \longrightarrow \bigoplus_v \mathcal{E}(T/F_v) \Big) \Big| |\mathcal{K}(T/F)|^{-1},$$

and $f_G^\kappa(\delta)$ is defined as in (27.6). The correspondence (27.8) is easily seen to have an inverse, which in general extends to a bijection

$$\{(G', \delta')\} \xrightarrow{\sim} \{(\delta, \kappa)\}.$$

The domain of this bijection is the set of equivalence classes of pairs $(G', \delta')$, where $G'$ is an elliptic endoscopic datum for $G$, $\delta'$ is a strongly $G$-regular, elliptic element in $G'(F)$ that is an image from $G$, and equivalence is defined by isomorphisms of endoscopic data. The range is the set of equivalence classes of pairs $(\delta, \kappa)$, where $\delta$ belongs to $\Delta_{\text{reg,ell}}(G)$, $\kappa$ lies in $\mathcal{K}(G_\delta/F)$, and equivalence is defined by conjugating by $G(\overline{F})$. (See [**Lan10**], [**Ko5**, Lemma 9.7].) Given $(G', \delta')$, we set

$$f'(\delta') = f_G^\kappa(\delta) = \sum_{\{\gamma_{\mathbb{A}} \in \Gamma(G(\mathbb{A})) : \gamma_{\mathbb{A}} \sim \delta\}} f_G(\gamma_{\mathbb{A}}) \kappa(\gamma_{\mathbb{A}}). \tag{27.10}$$

We can then write

$$I_{\text{reg,ell}}(f) = \sum_{G' \in \mathcal{E}_{\text{ell}}(G)} |\text{Out}_G(G')|^{-1} \sum_{\delta' \in \Delta_{G\text{-reg,ell}}(G')} a^G(\delta) \iota(G_\delta, G) f'(\delta'),$$

with the understanding that $f'(\delta') = 0$ if $\delta'$ is not an image from $G$. Langlands showed that for any pair $(G', \delta')$, the number

$$\iota(G, G') = \iota(G_\delta, G) \iota(G'_{\delta'}, G')^{-1} |\text{Out}_G(G')|^{-1}$$

was independent of $\delta'$ and $\delta$. (Kottwitz later expressed the product of the first two factors on the right as a quotient $\tau(G)\tau(G')^{-1}$ of Tamagawa numbers [**Ko3**, Theorem 8.3.1].) Set

$$\widehat{S}'_{G\text{-reg,ell}}(f') = \sum_{\delta' \in \Delta_{G\text{-reg,ell}}(G')} b'(\delta') f'(\delta'), \tag{27.11}$$

where

$$b'(\delta') = a^G(\delta) \iota(G'_{\delta'}, G') = \text{vol}\big(G'_{\delta'}(F) \backslash G'_{\delta'}(\mathbb{A})^1\big) \iota(G'_{\delta'}, G').$$

Then

$$I_{\text{reg,ell}}(f) = \sum_{G' \in \mathcal{E}_{\text{ell}}(G)} \iota(G, G') \widehat{S}'_{G\text{-reg,ell}}(f').$$

We have now sketched how to derive the formula (27.3). However, the term $f'(\delta')$ in (27.11) is defined in (27.10) only as a function on $\Delta_{G\text{-reg,ell}}(G')$. One would hope that it is the stable orbital integral at $\delta'$ of a function in $C_c^\infty\big(G'(\mathbb{A})\big)$. The sum in (27.10) can be taken over adelic products $\gamma_{\mathbb{A}} = \prod \gamma_v$, where $\gamma_v$ is a conjugacy class in $G(F_v)$ that lies in the stable class of the image $\delta_v$ of $\delta$ in $G(F_v)$. It follows that

$$f'(\delta') = f_G^\kappa(\delta) = \prod_v f_{v,G}^\kappa(\delta_v),$$

where

$$f_{v,G}^\kappa(\delta_v) = \sum_{\gamma_v \sim \delta_v} f_{v,G}(\gamma_v)\kappa(\gamma_v).$$

Are the local components $\delta'_v \to f_{v,G}^\kappa(\delta_v)$ stable orbital integrals of functions in $C_c^\infty\big(G'(F_v)\big)$? The question concerns the singularities that arise, as the strongly regular points approach 1, for example. Do enough of the singularities of the orbital integrals $f_{v,G}(\gamma_v)$ disappear from the sum so that only singularities of stable orbital integrals on the smaller group $G'(F_v)$ remain?

The question is very subtle. We have been treating $\delta$ as both a stable class in $\Delta_{\text{reg,ell}}(G)$ and a representative in $G(F)$ of that class. The distinction has not mattered so far, since $f'(\delta') = f_G^\kappa(\delta)$ depends only on the class of $\delta$. However, the coefficients $\kappa(\gamma_v)$ in the local functions $f_{v,G}^\kappa(\delta_v)$ are defined in terms of the relative position of $\gamma_v$ and $\delta_v$. The local functions do therefore depend on the choice of $\delta_v$ within its stable class in $G(F_v)$. The solution of Langlands and Shelstad was to replace $\kappa(\gamma_v)$ with a function $\Delta_G(\delta'_v, \gamma_v)$ that they called a *transfer factor.* This function is defined as a product of $\kappa(\gamma_v)$ with an explicit but complicated factor that depends on $\delta'_v$ and $\delta_v$, but not $\gamma_v$. The product $\Delta_G(\delta'_v, \gamma_v)$ then turns out to be independent of the choice of $\delta_v$, and depends only on the local stable class of $\delta'_v$ and local conjugacy class of $\gamma_v$. Moreover, if $\delta'_v$ is the local image of $\delta' \in \Delta_{G\text{-reg,ell}}(G')$, for every $v$, the product over $v$ of the corresponding local transfer factors is equal to the coefficient $\kappa(\gamma_{\mathbb{A}})$ in (27.10). (See [**LS1**, §3,§6].)

There is one further technical complication we should mention. The Langlands-Shelstad transfer factor depends on a choice of $L$-embedding of ${}^LG'$ into ${}^LG$. If $G'$ represents an endoscopic datum $(G', \mathcal{G}', s', \xi')$ with $\mathcal{G}' \subset {}^LG$ and $\xi' = \iota'$, this amounts to a choice of $L$-isomorphism $\widetilde{\xi}'\colon \mathcal{G}' \to {}^LG'$. In the case that $G_{\text{der}}$ is simply connected, such an $L$-isomorphism exists [**Lan8**]. However, it is not canonical, and one does have to choose $\widetilde{\xi}'$ in order to specify the transfer factors. The general situation is more complicated. The problem is that there might not be any such $L$-isomorphism. In this case, one has to modify the construction slightly. One replaces the group $G'$ by a central extension

$$1 \longrightarrow \widetilde{C}' \longrightarrow \widetilde{G}' \longrightarrow G' \longrightarrow 1$$

of $G'$, where $\widetilde{C}'$ is a suitable torus over $F$, and $\widetilde{G}'_{\text{der}}$ is simply connected. One can then take $\widetilde{\xi}'$ to be an $L$-embedding

$$\widetilde{\xi}' : \mathcal{G}' \hookrightarrow {}^L\widetilde{G}',$$

whose existence is again implied by [**Lan8**]. This determines a character $\widetilde{\eta}'$ on $\widetilde{C}'(\mathbb{A})/\widetilde{C}'(F)$, which is dual to the global Langlands parameter defined by the composition

$$W_F \longrightarrow \mathcal{G}' \xrightarrow{\widetilde{\xi}'} {}^L\widetilde{G}' \longrightarrow {}^L\widetilde{C}',$$

for any section $W_F \to \mathcal{G}'$. The transfer factor at $v$ becomes a function $\Delta_G(\delta'_v, \gamma_v)$ of $\delta'_v \in \Delta_{G\text{-reg}}\big(\widetilde{G}'(F_v)\big)$ and $\gamma_v \in \Gamma_{\mathrm{reg}}\big(G(F_v)\big)$, such that

$$\Delta_G(c'_v\delta'_v, \gamma_v) = \widetilde{\eta}'_v(c'_v)^{-1}\Delta_G(\delta'_v, \gamma_v), \qquad\qquad c'_v \in \widetilde{C}'(F_v).$$

It vanishes unless $\delta'_v$ is an image of a stable conjugacy class $\delta_v \in \Delta_{\mathrm{reg}}\big(G(F_v)\big)$ in $G(F_v)$, in which case it is supported on those conjugacy classes $\gamma_v$ in $G(F_v)$ that lie in $\delta_v$. In particular, $\Delta_G(\delta'_v, \gamma_v)$ has finite support in $\gamma_v$, for any $\delta'_v$.

Transfer factors play the role of a kernel in a transform of functions. Consider a function $f_v$ in $G(F_v)$, which we now take to be in the Hecke algebra $\mathcal{H}\big(G(F_v)\big)$. For any such $f_v$, we define an $(\widetilde{\eta}'_v)^{-1}$-equivariant function

$$f'_v(\delta'_v) = f_v^{\widetilde{G}'}(\delta'_v) = \sum_{\gamma_v \in \Gamma_{\mathrm{reg}}(G(F_v))} \Delta_G(\delta'_v, \gamma_v) f_{v,G}(\gamma_v) \tag{27.12}$$

of $\delta'_v \in \Delta_{G\text{-reg}}\big(\widetilde{G}'(F_v)\big)$. Langlands and Shelstad conjecture that $f'_v$ lies in the space $S\mathcal{I}\big(\widetilde{G}'(F_v), \widetilde{\eta}'_v\big)$ [**LS1**]. In other words, $f'_v(\delta'_v)$ can be identified with the stable orbital integral at $\delta'_v$ of some fixed function $h'_v$ in the $(\widetilde{\eta}'_v)^{-1}$-equivariant Hecke algebra $\mathcal{H}\big(\widetilde{G}'(F_v), \widetilde{\eta}'_v\big)$ on $\widetilde{G}'(F_v)$. (Langlands' earlier formulation of the conjecture [**Lan10**] was less precise, in that it postulated the existence of suitable transfer factors.) For archimedean $v$, the conjecture was established by Shelstad [**She3**]. In fact, it was Shelstad's results for real groups that motivated the construction of the general transfer factors $\Delta_G(\delta'_v, \gamma_v)$. (Shelstad actually worked with the Schwartz space $\mathcal{C}\big(G(F_v)\big)$. However, she also characterized the functions $f'_v$ in spectral terms, and in combination with the main theorem of [**CD**], this establishes the conjecture for the space $\mathcal{H}\big(G(F_v)\big)$.)

If $v$ is nonarchimedean, the Langlands-Shelstad conjecture remains open. Consider the special case that $G$, $G'$ and $\widetilde{\eta}'$ are unramified at $v$, and that $f_v$ is the characteristic function of a (hyperspecial) maximal compact subgroup $K_v$ of $G(F_v)$. Then one would like to know not only that $h'_v$ exists, but also that it can be taken to be the characteristic function of a (hyperspecial) maximal compact subgroup $\widetilde{K}'_v$ of $\widetilde{G}'(F_v)$ (or rather, the image of such a function in $\mathcal{H}(\widetilde{G}'_v, \widetilde{\eta}'_v)$.) This variant of the Langlands-Shelstad conjecture is what is known as the *fundamental lemma*. It is discussed in the lectures [**Hal1**] of Hales. Waldspurger has shown that the fundamental lemma actually implies the general transfer conjecture [**Wa2**]. To be precise, if the fundamental lemma holds for sufficiently many unramified pairs $(G_v, G'_v)$, the Langlands-Shelstad transfer conjecture holds for an arbitrary given pair $(G_v, G'_v)$.

The two conjectures together imply the existence of a global mapping

$$f = \prod_v f_v \longrightarrow f' = \prod_v f'_v$$

from $\mathcal{H}(G)$ to the space

$$S\mathcal{I}\big(\widetilde{G}'(\mathbb{A}), \widetilde{\eta}'\big) = \varinjlim_S S\mathcal{I}(\widetilde{G}'_S, \widetilde{\eta}'_S).$$

Such a mapping would complete Langlands' stabilization (27.3) of the regular elliptic term. It would express $I_{\mathrm{reg,ell}}(f)$ as the sum of a stable component, and pullbacks of corresponding stable components for proper endoscopic groups. We shall henceforth assume the existence of the mapping $f \to f'$. The remaining problem of stabilization is then to establish similar relations for the other terms in the invariant trace formula. We would like to show that any such term $I_*$ has a stable component

$$S_* = S_*^G = S_*^{G^*},$$

now regarded as a stable linear form on the Hecke algebra, such that

$$I_*(f) = \sum_{G' \in \mathcal{E}_{\mathrm{ell}}(G)} \iota(G, G') \widehat{S}'_*(f'), \tag{27.13}$$

for any $f \in \mathcal{H}\big(G(\mathbb{A})\big)$. The identity obtained by replacing the terms in the invariant trace formula by their corresponding stable components would then be a stable trace formula. We shall describe the solution to this problem in §29.

In recognition of the inductive nature of the putative identity (27.13), we ought to modify some of the definitions slightly. In the case of $I_* = I_{\mathrm{reg,ell}}$, for example, the term $\widehat{S}'_{G\text{-reg,ell}}(f')$ in (27.3) is not the full stable component of $I'_{\mathrm{reg,ell}}$. We could rectify this minor inconsistency by replacing $I_{\mathrm{reg,ell}}(f)$ with its $H$-regular part $I_{H\text{-reg,ell}}(f)$, for some reductive group $H$ that shares a maximal torus with $G$, and whose roots contain those of $G$. The resulting version

$$I_{H\text{-reg,ell}}(f) = \sum_{G'} \iota(G, G') \widehat{S}'_{H\text{-reg,ell}}(f')$$

of (27.3) is then a true inductive formula.

Another point concerns the function $f'$. We are assuming that $f'$ is the stable image of a function in the $(\widetilde{\eta}')^{-1}$-equivariant algebra $\mathcal{H}\big(\widetilde{G}'(\mathbb{A}), \widetilde{\eta}'\big)$. However, the original function $f$ belongs to the ordinary Hecke algebra $\mathcal{H}(G)$. To put the two functions on an even footing, we fix a central torus $Z$ in $G$ over $F$, and a character $\zeta$ on $Z(\mathbb{A})/Z(F)$. We then write $\widetilde{Z}'$ for the preimage in $\widetilde{G}'$ of the canonical image of $Z$ in $G'$. Global analogues of the local constructions in [**LS1**, (4.4)] provide a canonical extension of $\widetilde{\eta}'$ to a character on $\widetilde{Z}'(\mathbb{A})/\widetilde{Z}'(F)$. We write $\widetilde{\zeta}'$ for the character on $\widetilde{Z}'(\mathbb{A})/\widetilde{Z}'(F)$ obtained from the product of $\widetilde{\eta}'$ with the pullback of $\zeta$. The presumed correspondence $f \to f'$ then takes the form of a mapping from $\mathcal{H}\big(G(\mathbb{A}), \zeta\big)$ to $S\mathcal{I}\big(\widetilde{G}'(\mathbb{A}), \widetilde{\zeta}'\big)$. At the beginning of §29, we shall describe a version of the invariant trace formula that applies to equivariant test functions $f$.

## 28. Local spectral transfer and normalization

We have now set the stage for the final refinement of the trace formula. We shall describe it over the course of the next two sections. This discussion, as well as that of the applications in §30, contains much that is only implicit. However, it also contains remarks that are intended to provide general orientation. A reader who is not an expert should ignore the more puzzling points at first pass, and aim instead at acquiring a sense of the underlying structure.

The problem is to stabilize the invariant trace formula for a general connected group $G$ over $F$. In the case that $G$ is an inner form of $GL(n)$, Theorems 25.5 and 25.6 represent a solution of the problem. They provide a term by term identification of the trace formula for $G$ with the relevant part of the trace formula for the group

$G^* = GL(n)$. In this case, all invariant distributions are stable, and $G^*$ is the only elliptic endoscopic group. The stabilization problem therefore reduces to the comparison of $G$ with its quasisplit inner form.

The case of an inner form of $GL(n)$ is also simpler for the existence of a local correspondence $\pi_v \to \pi_v^*$ of tempered representations. Among other things, this allows us to define normalizing factors for intertwining operators for $G$ in terms of those for $G^*$. We recall that the invariant distributions in the trace formula depend on a choice of normalizing factors. So therefore do any identities among these distributions. For general $G$, the theory of endoscopy predicts a refined local correspondence, which would yield compatible normalizations as a biproduct. However, the full form of this correspondence is presently out of reach. We nevertheless do require some analogue of it in any attempt to stabilize the trace formula.

We shall first describe a makeshift substitute for the local correspondence, which notwithstanding its provisional nature, still depends on the fundamental lemma. We will then review how the actual correspondence is supposed to work. After seeing the two side by side, the reader will probably agree that it is not reasonable at this point to try to construct compatible normalizing factors. Fortunately, there is a second way to normalize weighted orbital integrals, which does not depend on a normalization of intertwining operators. We shall discuss the construction at the end of the section. At the beginning of the next section, we shall describe how the construction leads to another form of the invariant trace formula. It will be this second form of the trace formula that we actually stabilize.

The global stabilization of the next section will be based on two spaces of invariant distributions, which reflect the general duality between conjugacy classes and characters. We may as well introduce them here. We are assuming that $G$ is arbitrary. If $V$ is a finite set of valuations of $F$, we shall write $G_V = G(F_V)$ for simplicity. Suppose that $Z$ is a torus in $G$ over $F$ that is contained in the center, and that $\zeta_V$ is a character on $Z_V$. Let $\mathcal{D}(G_V, \zeta_V)$ be the space of invariant distributions that are $\zeta_V$-equivariant under translation by $Z_V$, and are supported on the preimage in $G_V$ of a finite union of conjugacy classes in $\overline{G}_V = G_V/Z_V$. Let $\mathcal{F}(G_V, \zeta_V)$ be the space of invariant distributions that are $\zeta_V$-equivariant under translation by $Z_V$, and are spanned by irreducible characters on $G_V$. This second space is obviously spanned by the characters attached to the set $\Pi(G_V, \zeta_V)$ of irreducible representations of $G_V$ whose central character on $Z_V$ equals $\zeta_V$. Now, the Hecke algebra on $G_V$ has $\zeta_V^{-1}$-equivariant analogue $\mathcal{H}(G_V, \zeta_V)$, composed of functions $f$ such that

$$f(zx) = \zeta_V(z)^{-1} f(x), \qquad z \in Z_V.$$

Likewise, the invariant Hecke algebra has a $\zeta_V^{-1}$-analogue $\mathcal{I}(G_V, \zeta_V)$. A distribution $D$ in either of the spaces $\mathcal{D}(G_V, \zeta_V)$ or $\mathcal{F}(G_V, \delta_V)$ can be regarded as a linear form

$$D(f) = f_G(D), \qquad f \in \mathcal{H}(G_V, \zeta_V),$$

on either $\mathcal{H}(G_V, \zeta_V)$ or $\mathcal{I}(G_V, \zeta_V)$.

The notation $f_G(D)$ requires further comment. On the one hand, it generalizes the way we have been denoting both invariant orbital integrals $f_G(\gamma)$ and irreducible characters $f_G(\pi)$. But it also has the more subtle interpretation as the value of a linear form on the function $f_G$ in $\mathcal{I}(G_V, \zeta_V)$. Since we have defined $\mathcal{I}(G_V, \zeta_V)$ as a space of functions on $\Pi_{\mathrm{temp}}(G_V, \zeta_V)$, we need to know that $D$ is supported on characters. This is clear if $D$ belongs to $\mathcal{F}(G_V, \zeta_V)$. If $D$ belongs to the other space $\mathcal{D}(G_V, \zeta_V)$, results of Harish-Chandra and Bouaziz [**Bou**] imply that it can be expressed in terms of strongly regular invariant orbital integrals. (See [**A30**, Lemma 1.1].) Since invariant orbital integrals are supported on characters, by the special case of Theorem 23.2 with $M = G$, $D$ is indeed supported on characters. Incidentally, by the special case of Theorem 23.2 and the fact that characters are locally integrable functions, we can identify $\mathcal{I}(G_V, \zeta_V)$ with the space

$$\big\{f_G(\gamma) : \ \gamma \in \Gamma_{G\text{-reg}}(G_V), \ f \in \mathcal{H}(G_V, \zeta_V)\big\}.$$

We have in fact already implicitly done so in our discussion of inner twists and base change for $GL_n$ in §25 and §26. However, we do not have a geometric analogue of [**CD**] that would allow us to characterize $\mathcal{I}(G_V, \zeta_V)$ explicitly as a space of functions on $\Gamma_{G\text{-reg}}(G_V)$.

We write $S\mathcal{D}(G_V, \zeta_V)$ and $S\mathcal{F}(G_V, \zeta_V)$ for the subspaces of stable distributions in $\mathcal{D}(G_V, \zeta_V)$ and $\mathcal{F}(G_V, \zeta_V)$ respectively. We also write $S\mathcal{I}(G_V, \zeta_V)$ for the $\zeta_V^{-1}$-analogue of the stably invariant Hecke algebra. Any distribution $S$ in either $S\mathcal{D}(G_V, \zeta_V)$ or $S\mathcal{F}(G_V, \zeta_V)$ can then be identified with a linear form

$$f^G \longrightarrow f^G(S), \qquad f \in \mathcal{H}(G_V, \zeta_V),$$

on $S\mathcal{I}(G_V, \zeta_V)$. We recall $S\mathcal{I}(G_V, \zeta_V)$ is presently just a space of functions on $\Delta_{G\text{-reg}}(G_V)$. One consequence of the results we are about to describe is a spectral characterization of $S\mathcal{I}(G_V, \zeta_V)$.

Our focus for the rest of this section will be entirely local. We shall consider the second space $\mathcal{F}(G_V, \zeta_V)$, under the condition that $V$ consist of one valuation $v$. We shall regard $G$ and $Z$ as groups over the local field $k = F_v$, and we shall write $\zeta = \zeta_v$, $G = G_v = G(F_v)$, $\mathcal{F}(G, \zeta) = \mathcal{F}(G_v, \zeta_v)$, $\Pi(G, \zeta) = \Pi(G_v, \zeta_v)$, $\mathcal{H}(G, \zeta) = \mathcal{H}(G_v, \zeta_v)$, and so on, for simplicity. With this notation, we write $\mathcal{I}_{\text{cusp}}(G, \zeta)$ for the subspace of functions in $\mathcal{I}(G, \zeta)$ that are supported on the $k$-elliptic subset $\Gamma_{\text{reg,ell}}(G)$ of $\Gamma_{\text{reg}}(G) = \Gamma_{\text{reg}}(G_v)$. We also write $S\mathcal{I}_{\text{cusp}}(G, \zeta)$ for the image of $\mathcal{I}_{\text{cusp}}(G, \zeta)$ in $S\mathcal{I}(G, \zeta)$, and $\mathcal{H}_{\text{cusp}}(G, \zeta)$ for the preimage of $\mathcal{I}_{\text{cusp}}(G, \zeta)$ in $\mathcal{H}(G, \zeta)$. Keep in mind that any element $D \in \mathcal{F}(G, \zeta)$ has a (virtual) character. It is a locally integrable, invariant function $\Theta(D, \cdot)$ on $G_v$ such that

$$f_G(D) = \int_{G_v} f(x)\Theta(D, x)\mathrm{d}x, \qquad f \in \mathcal{H}(G, \zeta).$$

Assume for a moment that $Z$ contains the split component $A_G$ (over $k$) of the center of $G$. The space $\mathcal{I}_{\text{cusp}}(G, \zeta)$ then has the noteworthy property that it is a canonical linear image of $\mathcal{F}(G, \zeta)$. To be precise, there is a surjective linear mapping

$$\mathcal{F}(G, \zeta) \longrightarrow \mathcal{I}_{\text{cusp}}(G, \zeta)$$

that assigns to any element $D \in \mathcal{F}(G, \zeta)$ the elliptic part

$$I_{\text{ell}}(D, \gamma) = \begin{cases} I(D, \gamma), & \text{if } \gamma \in \Gamma_{\text{reg,ell}}(G) \\ 0, & \text{otherwise,} \end{cases}$$

of its normalized character

$$I(D, \gamma) = |D^G(\gamma)|^{\frac{1}{2}}\Theta(D, \gamma).$$

This assertion follows from the case $M = G$ of the general result [**A20**, Theorem 5.1]. What is more, the mapping has a canonical linear section

$$\mathcal{I}_{\mathrm{cusp}}(G,\zeta) \longrightarrow \mathcal{F}(G,\zeta).$$

This is defined by a natural subset $T_{\mathrm{ell}}(G,\zeta)$ [**A22**, §4] of $\mathcal{F}(G,\zeta)$ whose image in $\mathcal{I}_{\mathrm{cusp}}(G,\zeta)$ forms a basis.

The set $T_{\mathrm{ell}}(G,\zeta)$ contains the family $\Pi_2(G,\zeta)$ of square integrable representations of $G_v$ with central character $\zeta$. However, it also contains certain linear combinations of irreducible constituents of induced representations. We can define $T_{\mathrm{ell}}(G,\zeta)$ as the set of $G_v$-orbits of triplets $(L,\sigma,r)$, where $L$ is a Levi subgroup of $G$ over $k = F_v$, $\sigma$ belongs to $\Pi_2(L,\zeta)$, and $r$ is an element in the $R$-group $R_\sigma$ of $\sigma$ whose null space in $\mathfrak{a}_M$ equals $\mathfrak{a}_G$. The $R$-group is an important object in local harmonic analysis that was discovered by Knapp. In general terms, it can be represented as a subgroup of the stabilizer of $\sigma$ in $W(L)$, for which corresponding normalized intertwining operators

$$R_Q(r,\sigma) = A(\sigma_r)R_{r^{-1}Qr|Q}(\sigma), \qquad\qquad Q \in \mathcal{P}(L),\ r \in R_\sigma,$$

form a basis of the space of all operators that intertwine the induced representation $\mathcal{I}_Q(\sigma)$. We write $\sigma_r$ for an extension of the representation $\sigma$ to the group generated by $L_v$ and a representative $\widetilde{w}_r$ of $r$ in $K_v$. Then

$$A(\sigma_r):\ \mathcal{H}_{r^{-1}Qr}(\sigma) \longrightarrow \mathcal{H}_Q(\sigma)$$

is the operator defined by

$$\big(A(\sigma_r)\phi'\big)(x) = \sigma_r(\widetilde{w}_r)\phi'(\widetilde{w}_r^{-1}x), \qquad\qquad \phi' \in \mathcal{H}_{r^{-1}Qr}(\sigma).$$

(See [**A20**, §2].)

We identify elements $\tau \in T_{\mathrm{ell}}(G,\zeta)$ with the distributions

$$f_G(\tau) = \mathrm{tr}\big(R_Q(r,\sigma)\mathcal{I}_Q(\sigma,f)\big), \qquad\qquad f \in C_c^\infty(G),$$

in $\mathcal{F}(G,\zeta)$. It is the associated set of functions

$$I_{\mathrm{ell}}(\tau,\cdot), \qquad\qquad \tau \in T_{\mathrm{ell}}(G,\zeta),$$

that provides a basis of $\mathcal{I}_{\mathrm{cusp}}(G,\zeta)$. In fact, by Theorem 6.1 of [**A20**], these functions form an orthogonal basis of $\mathcal{I}_{\mathrm{cusp}}(G,\zeta)$ with respect to a canonical measure $d\gamma$ on $\Gamma_{\mathrm{reg,ell}}(G/Z)$, whose square norms

$$\|I_{\mathrm{ell}}(\tau)\|^2 = \int_{\Gamma_{\mathrm{reg,ell}}(G/Z)} I_{\mathrm{ell}}(\tau,\gamma)\overline{I_{\mathrm{ell}}(\tau,\gamma)}\mathrm{d}\gamma = n(\tau), \qquad\qquad \tau \in T_{\mathrm{ell}}(G,\zeta),$$

satisfy

$$n(\tau) = |R_{\sigma,r}||\det(1-r)_{\mathfrak{a}_L/\mathfrak{a}_G}|.$$

(As usual $R_{\sigma,r}$ denotes the centralizer of $r$ in $R_\sigma$. See [**A21**, §4].)

The set $T_{\mathrm{ell}}(G,\zeta)$ is part of a natural basis $T(G,\zeta)$ of $\mathcal{F}(G,\zeta)$. This can either be defined directly [**A20**, §3], or built up from elliptic sets attached to Levi subgroups. To remove the dependence on $Z$, we should really let $\zeta$ vary. The union

$$T_{\mathrm{temp,ell}}(G) = \coprod_\zeta T_{\mathrm{ell}}(G,\zeta)$$

is a set of tempered distributions, which embeds in the set

$$T_{\mathrm{ell}}(G) = \big\{\tau_\lambda :\ \tau \in T_{\mathrm{temp,ell}}(G),\ \lambda \in \mathfrak{a}^*_{G,\mathbb{C}}\big\}$$

that parametrizes nontempered elliptic characters

$$\Theta(\tau_\lambda, \gamma) = \Theta(\tau, \gamma)e^{\lambda(H_G(\gamma))}.$$

These two elliptic sets are in turn contained in respective larger sets

$$T_{\mathrm{temp}}(G) = \coprod_{\{M\}} T_{\mathrm{temp,ell}}(M)/W(M)$$

and

$$T(G) = \coprod_{\{M\}} T_{\mathrm{ell}}(M)/W(M),$$

where $\{M\}$ represents the set of conjugacy classes of Levi subgroups of $G$ over $k = F_v$. If $T_*(G)$ is any of the four sets above, we obviously have an associated subset $T_*(G, \zeta)$ attached to any $Z$ and $\zeta$. The distributions $f \to f_G(\tau)$ parametrized by the largest set $T(G, \zeta)$ form a basis of $\mathcal{F}(G, \zeta)$, while the distributions parametrized by $T_{\mathrm{temp}}(G, \zeta)$ give a basis of the subset of tempered distributions in $\mathcal{F}(G, \zeta)$. We thus have bases that are parallel to the more familiar bases $\Pi(G, \zeta)$ and $\Pi_{\mathrm{temp}}(G, \zeta)$ of these spaces given by irreducible characters.

Assume now that $k = F_v$ is nonarchimedean. In this case, one does not have a stable analogue for the set $T_{\mathrm{ell}}(G, \zeta)$. As a substitute, in case $G$ is quasisplit and $Z$ contains $A_G$, we write $\Phi_2(G, \zeta)$ for an indexing set $\{\phi\}$ that parametrizes a fixed family of functions $\{S_{\mathrm{ell}}(\phi, \cdot)\} \subset S\mathcal{I}_{\mathrm{cusp}}(G, \zeta)$ for which the products

$$n(\delta)S_{\mathrm{ell}}(\phi, \delta), \qquad\qquad \delta \in \Delta_{G\text{-reg,ell}}(G),\ \phi \in \Phi_2(G, \zeta),$$

form an orthogonal basis of $S\mathcal{I}_{\mathrm{cusp}}(G, \zeta)$. (The number $n(\delta)$ stands for the number of conjugacy classes in the stable class $\delta$, and is used to form the measure $d\delta$ on $\Delta_{G\text{-reg,ell}}(G/Z)$. The subscript *2* is used in place of *ell* because the complement of $\Pi_2(G, \zeta)$ in $T_{\mathrm{ell}}(G, \zeta)$ is believed to be purely unstable.) We fix the family $\{S_{\mathrm{ell}}(\phi, \cdot)\}$, subject to certain natural conditions [**A22**, Proposition 5.1]. We then form larger sets

$$\Phi_{\mathrm{temp},2}(G) = \coprod_{\zeta} \Phi_2(G, \zeta),$$

$$\Phi_2(G) = \{\phi_\lambda :\ \phi \in \Phi_{\mathrm{temp},2}(G),\ \lambda \in \mathfrak{a}^*_{G,\mathbb{C}}\},$$

$$\Phi_{\mathrm{temp}}(G) = \coprod_{\{M\}} \Phi_{\mathrm{temp},2}(M)/W(M),$$

and

$$\Phi(G) = \coprod_{\{M\}} \Phi_2(M)/W(M),$$

where $S_{\mathrm{ell}}(\phi_\lambda, \delta) = S_{\mathrm{ell}}(\phi, \delta)e^{\lambda(H_G(S))}$, as well as corresponding subsets $\Phi_*(G, \zeta)$ of $\Phi_*(G)$ attached to any $Z$ and $\zeta$. The analogy with the sets $T_*(G, \zeta)$ is clear. What is not obvious, however, is that the elements in $\Phi_*(G, \zeta)$ give stable distributions. The first step in this direction is to define

$$f^G(\phi) = \int_{\Delta_{\mathrm{reg,ell}}(G/Z)} f^G(\delta)S_{\mathrm{ell}}(\phi, \delta)\mathrm{d}\delta, \tag{28.1}$$

for any $f \in \mathcal{H}_{\mathrm{cusp}}(G, \zeta)$ and $\phi \in \Phi_2(G, \zeta)$.

We shall now apply the Langlands-Shelstad transfer of functions. One introduces endoscopic data $G'$ for $G$ over the local field $k = F_v$ by copying the definitions of §27 for the global field $F$. (The global requirement that a certain class in

$H^1\big(F, Z(\widehat{G})\big)$ be locally trivial is replaced by the simpler condition that the corresponding class in $H^1\big(F_v, Z(\widehat{G})\big)$ be trivial, but this is the only difference.) We follow the same notation as in the global constructions of §27. In particular, we write $\mathcal{E}_{\rm ell}(G)$ for the set of isomorphism classes of elliptic endoscopic data for $G$ over $k$.

We are assuming that the fundamental lemma holds, for units of Hecke algebras at unramified places of any group over $F$ that is isomorphic to $G$ over $k = F_v$. The theorem of Waldspurger mentioned at the end of the last section asserts that this global hypothesis (augmented to allow for induction arguments) implies the Langlands-Shelstad transfer conjecture for any endoscopic datum $G'$ for $G$ over $k$. We suppose that for each elliptic endoscopic datum $G' \in \mathcal{E}_{\rm ell}(G)$ of $G$ over $k$, we have chosen sets $\Phi(\widetilde{G}', \widetilde{\zeta}')$, as above. If $f$ belongs to $\mathcal{H}_{\rm cusp}(G,\zeta)$, $f'$ belongs to $S\mathcal{I}(\widetilde{G}', \widetilde{\zeta}')$, by our assumption. Since the orbital integrals of $f$ are supported on the elliptic set, $f'$ in fact belongs to the subspace $S\mathcal{I}_{\rm cusp}(\widetilde{G}', \widetilde{\zeta}')$ of $S\mathcal{I}(\widetilde{G}', \widetilde{\zeta}')$. We can therefore define $f'(\phi')$ by (28.1), for any element $\phi' \in \Phi_2(\widetilde{G}', \widetilde{\zeta}')$. As a linear form on $\mathcal{H}_{\rm cusp}(G,\zeta)$, $f'(\phi')$ is easily seen to be the restriction of some distribution in $\mathcal{F}(G,\zeta)$. It therefore has an expression

$$f'(\phi') = \sum_{\tau \in T_{\rm ell}(G,\zeta)} \Delta_G(\phi', \tau) f_G(\tau), \qquad f \in \mathcal{H}_{\rm cusp}(G,\zeta), \tag{28.2}$$

in terms of the basis $T_{\rm ell}(G,\zeta)$.

The coefficients $\Delta_G(\phi', \tau)$ in (28.2) are to be regarded as spectral transfer factors. They are defined a priori for elements $\phi' \in \Phi_2(\widetilde{G}', \widetilde{\zeta}')$ and $\tau \in T_{\rm ell}(G,\zeta)$. However, it is easy to extend the construction to general elements $\phi' \in \Phi(\widetilde{G}', \widetilde{\zeta}')$ and $\tau \in T(G,\zeta)$. To do so, we represent $\phi'$ and $\tau$ respectively as Weyl orbits $\{\phi'_{M'}\}$ and $\{\tau_M\}$ of elliptic elements $\phi'_M \in \Phi_2(\widetilde{M}', \widetilde{\zeta}')$ and $\tau_M \in T_{\rm ell}(M,\zeta)$ attached to Levi subgroups $\widetilde{M}' \subset \widetilde{G}'$ and $M \subset G$. We then define $\Delta_G(\phi', \sigma) = 0$ unless $M'$ can be identified with an elliptic endoscopic group for $M$, in which case we set

$$\Delta_G(\phi', \tau) = \sum_{w \in W(M)} \Delta_M(\phi'_{M'}, w\tau_M).$$

It is not hard to deduce that for a fixed value of one of the arguments, $\Delta_G(\phi', \tau)$ has finite support in the other.

Suppose now that $f$ is any function in $\mathcal{H}(G,\zeta)$. For any $G' \in \mathcal{E}_{\rm ell}(G)$, we define the *spectral transfer* of $f$ to be the function

$$f'_{\rm gr}(\phi') = \sum_{\tau \in T(G,\zeta)} \Delta_G(\phi', \tau) f_G(\tau), \qquad \phi' \in \Phi'(\widetilde{G}', \widetilde{\zeta}').$$

(The subscript *gr* stands for the grading on the space $\mathcal{I}(G,\zeta)$ provided by the basis $T(G,\zeta)$ of $\mathcal{F}(G,\zeta)$.) It is by no means clear, a priori, that $f'_{\rm gr}$ coincides with the Langlands-Shelstad transfer $f'$. The problem is this. We defined the coefficients $\Delta_G(\phi', \tau)$ by stabilizing elliptic (virtual) characters $T_{\rm ell}(G,\zeta)$ on the elliptic set. However, these characters also take values at nonelliptic elements. Why should their stabilization on the elliptic set, where they are uniquely determined, induce a corresponding stabilization on the nonelliptic set? The answer is provided by the following theorem.

Theorem 28.1. (a) *Suppose that $G$ is quasisplit and that $\phi \in \Phi(G,\zeta)$. Then the distribution*

$$f \longrightarrow f^G_{\mathrm{gr}}(\phi), \qquad f \in \mathcal{H}(G,\zeta),$$

*is stable, and therefore lifts to a linear form*

$$f^G \longrightarrow f^G(\phi), \qquad f \in \mathcal{H}(G,\zeta),$$

*on $S\mathcal{I}(G,\zeta)$.*

(b) *Suppose that $G$ is arbitrary, that $G' \in \mathcal{E}_{\mathrm{ell}}(G)$, and that $\phi' \in \Phi(\widetilde{G}',\widetilde{\zeta}')$. Then*

$$f'(\phi') = f'_{\mathrm{gr}}(\phi'), \qquad f \in \mathcal{H}(G,\zeta).$$

**Remark.** The theorem asserts that for any $\phi' \in \Phi(\widetilde{G}',\widetilde{\zeta}')$, the mapping $f \to f'(\phi')$ is a well defined element in $\mathcal{F}(G,\zeta)$, with an expansion

$$f'(\phi') = \sum_{\tau \in T(G,\zeta)} \Delta_G(\phi',\tau) f_G(\tau), \qquad f \in \mathcal{H}(G,\zeta). \tag{28.3}$$

Since $\Pi(G,\zeta)$ is another basis of $\mathcal{F}(G,\zeta)$, we could also write

$$f'(\phi') = \sum_{\pi \in \Pi(G,\zeta)} \Delta_G(\phi',\pi) f_G(\pi), \tag{28.4}$$

for complex numbers $\Delta_G(\phi',\pi)$.

The two assertions (a) and (b) of the theorem coincide with Theorems 6.1 and 6.2 of [**A22**], the main results of that paper. The proof is global. One chooses a suitable group over $F$ that is isomorphic to $G$ over $k = F_v$. By taking a global test function that is cuspidal at two places distinct from $v$, one can apply the simple trace formula of Corollary 23.6. The fundamental lemma and the Langlands-Shelstad transfer mapping provide a transfer of global test functions to endoscopic groups. One deduces the assertions of the theorem by a variant of the arguments used to establish Theorem 25.1(b) [**DKV**] and local base change [**AC**, §1]. □

We have taken some time to describe a weak form of spectral transfer. This is of course needed to stabilize the general trace formula. However, we would also like to contrast it with the stronger version expected from the theory of endoscopy, which among many other things, ought to give rise to compatible normalizing factors. For we are trying to see why we need another form of the invariant trace formula.

One expects to be able to identify $\Phi(G)$ with the set of Langlands parameters. A Langlands parameter for $G$ is a $\widehat{G}$-conjugacy class of relevant $L$-homomorphisms

$$\phi:\ L_k \longrightarrow {}^LG,$$

from the local Langlands group

$$L_k = W_k \times SU(2)$$

to the Weil form ${}^LG = \widehat{G} \rtimes W_k$ of $G$ over $k = F_v$. (In this context, an $L$-homomorphism is a continuous homomorphism for which the image in $\widehat{G}$ of any element is semisimple, and which commutes with the projections of $L_k$ and ${}^LG$ onto $W_k$. *Relevant* means that if the image of $\phi$ is contained in a Levi subgroup ${}^LM$ of ${}^LG$, then ${}^LM$ must be the $L$-group of a Levi subgroup $M$ of $G$ over $k$.) Let us temporarily let $\Phi(G)$ denote the set of such parameters, rather than the abstract indexing set above. Then $\Phi_{\mathrm{temp}}(G)$ corresponds to those homomorphisms whose image projects to a relatively compact subset of $\widehat{G}$. The subset $\Phi_2(G)$ corresponds to mappings whose images are contained in no proper Levi subgroup ${}^LM$ of ${}^LG$, while $\Phi_{\mathrm{temp},2}(G)$ is of course the intersection of $\Phi_{\mathrm{temp}}(G)$ with $\Phi_2(G)$. For any $\phi$, one writes $S_\phi$ for the centralizer in $\widehat{G}$ of the image of $\phi$, and $\mathcal{S}_\phi$ for the group of connected components of the quotient $\overline{S}_\phi = S_\phi/Z(\widehat{G})^{\Gamma_k}$. The *R-group* $R_\phi$ of $\phi$ is defined as the quotient of $\mathcal{S}_\phi$ by the subgroup of components that act by inner automorphism on $\overline{S}{}^0_\phi$. A choice of Borel subgroup in the connected reductive group $\overline{S}_\phi$ induces an embedding of $R_\phi$ into $\mathcal{S}_\phi$.

In the case that $G$ is abelian, Langlands constructed a natural bijection $\phi \to \pi$ from the set of parameters $\Phi(G)$ onto the set $\Pi(G)$ of quasicharacters on $G$ [**Lan12**]. We can therefore set

$$S_{\mathrm{ell}}(\phi,\gamma) = \Theta(\pi,\gamma) = \pi(\gamma), \qquad \gamma \in G(k),$$

in this case. For example, if $G = GL(1)$, a parameter in $\Phi(G)$ is tantamount to a continuous homomorphism

$$L_k = W_k \times SU(2) \longrightarrow \widehat{G} = \mathbb{C}^*.$$

Since $SU(2)$ is its own derived group, and the abelianization of $W_k$ is isomorphic to $k^* \cong G(k)$, a parameter does indeed correspond to a quasicharacter. If $G$ is a general group, with central torus $Z$, there is a canonical homomorphism from ${}^LG$ to ${}^LZ$. A parameter in $\Phi(G)$ then yields a quasicharacter $\zeta$ on $Z$, whose corresponding parameter is the composition

$$L_k \xrightarrow{\phi} {}^LG \longrightarrow {}^LZ.$$

The entire set of parameters $\Phi(G)$ thus decomposes into a disjoint union over $\zeta$ of the subsets $\Phi(G,\zeta)$ with central quasicharacter $\zeta$ on $Z$.

Suppose that $\zeta$ is a character on $Z$. For each parameter $\phi \in \Phi_{\mathrm{temp}}(G,\zeta)$, it is expected that there is a canonical nonnegative integer valued function $d_\phi(\pi)$ on $\Pi_{\mathrm{temp}}(G,\zeta)$ with finite support, such that the distribution

$$f \longrightarrow f^G(\phi) = \sum_\pi d_\phi(\pi) f_G(\pi), \qquad f \in \mathcal{H}(G,\zeta),$$

is stable. The sum

$$S(\phi,\delta) = \sum_\pi d_\phi(\pi) I(\pi,\gamma), \qquad \gamma \in \Gamma_{\mathrm{reg}}(G),$$

would then depend only on the stable conjugacy class $\delta$ of $\gamma$. Moreover, the finite packets

$$\Pi_\phi = \big\{\pi \in \Pi_{\mathrm{temp}}(G,\zeta) :\ d_\phi(\pi) > 0\big\}, \qquad \phi \in \Phi_{\mathrm{temp}}(G,\zeta),$$

are supposed to be disjoint, and have union equal to $\Pi_{\mathrm{temp}}(G,\zeta)$. The subset $\Pi_{\mathrm{temp},2}(G,\zeta)$ of $\Pi_{\mathrm{temp}}(G,\zeta)$ should be the disjoint union of packets $\Pi_\phi$, in which $\phi$ ranges over the subset $\Phi_{\mathrm{temp},2}(G,\zeta)$ of $\Phi_{\mathrm{temp}}(G,\zeta)$.

Suppose that these properties hold in general, and that $G' \in \mathcal{E}_{\mathrm{ell}}(G)$ and $\phi' \in \Phi_{\mathrm{temp}}(\widetilde{G}',\widetilde{\zeta}')$. Then $f'(\phi')$ is a well defined linear form in $f \in \mathcal{H}(G,\zeta)$. The pair $(\widetilde{G}',\widetilde{\zeta}')$ is constructed in such a way that $\phi'$ maps to a parameter $\phi \in \Phi_{\mathrm{temp}}(G,\zeta)$. For example, if $\widetilde{G}'$ happens to equal $G'$, $\phi$ is just the composition of

$\phi'$ with the underlying embedding of ${}^LG'$ into ${}^LG$. It is believed that the expansion of $f'(\phi')$ into irreducible characters on $G$ takes the form

$$f'(\phi') = \sum_{\pi \in \Pi_\phi} \Delta_G(\phi', \pi) f_G(\pi), \tag{28.5}$$

for complex coefficients $\Delta_G(\phi', \pi)$ that are supported on the packet $\Pi_\phi$.

The other basis $T_{\mathrm{temp}}(G, \zeta)$ would also have a packet structure. For the elements of $\Pi_\phi$ ought to be irreducible constituents of induced representations

$$\mathcal{I}_P(\sigma), \qquad P \in \mathcal{P}(M),\ \sigma \in \Pi_{\phi_M}, \tag{28.6}$$

where $M \subset G$ is a minimal Levi subgroup whose $L$-group ${}^LM \subset {}^LG$ contains the image of $\phi$, and $\phi_M$ is the parameter in $\Phi_{\mathrm{temp},2}(M, \zeta)$ determined by $\phi$. Recall that as a representation in $\Pi_2(M, \zeta)$, $\sigma$ has its own $R$-group $R_\sigma$. In terms of the $R$-group of $\phi$, $R_\sigma$ ought to be the stabilizer of $\sigma$ under the dual action of $R_\phi$ on $M$. Let $T_\phi$ be the subset of $T_{\mathrm{temp}}(G, \zeta)$ represented by triplets

$$(M, \sigma, r), \qquad \sigma \in \Pi_{\phi_M},\ r \in R_\sigma.$$

If the packet $\Pi_\phi$ is defined as above, the packet $T_\phi$ gives rise to a second basis of the subspace of $\mathcal{F}(G, \zeta)$ spanned by $\Pi_\phi$. It provides a second expansion

$$f'(\phi') = \sum_{\tau \in T_\phi} \Delta_G(\phi', \tau) f_G(\tau), \tag{28.7}$$

for complex coefficients $\Delta_G(\phi', \tau)$. As $\phi$ varies over $\Phi_{\mathrm{temp}}(G, \zeta)$, $T_{\mathrm{temp}}(G, \zeta)$ is a disjoint union of the corresponding packets $T_\phi$.

Given their expected properties, Langlands' parameters become canonical indexing sets. If $G$ is quasisplit and $Z$ contains $A_G$, we can set

$$S_{\mathrm{ell}}(\phi, \delta) = \begin{cases} S(\phi, \delta), & \text{if } \delta \in \Delta_{\mathrm{reg,ell}}(G), \\ 0, & \text{otherwise,} \end{cases}$$

for any $\phi \in \Phi_2(G, \zeta)$. The family $\{S_{\mathrm{ell}}(\phi, \cdot)\}$ then serves as the basis of $S\mathcal{I}_{\mathrm{cusp}}(G, \zeta)$ chosen earlier. The improvement of the conjectural transfer (28.5) or (28.7) over the weaker version (28.4) or (28.3) that one can actually prove (modulo the fundamental lemma) is obvious. For example, the hypothetical coefficients in (28.5) are supported on disjoint sets parametrized by $\Phi_{\mathrm{temp}}(G, \zeta)$. However, the actual coefficents in (28.4) could have overlapping supports, for which we have no control.

The hypothetical coefficients in (28.5) are expected to have further striking properties. Suppose for example that $G$ is quasisplit. In this case, it seems to be generally believed that coefficients will give a bijection from $\Pi_\phi$ onto the set $\widehat{\mathcal{S}}_\phi$ of irreducible characters on $\mathcal{S}_\phi$. This bijection would depend on a noncanonical choice of any base point $\pi_1$ in $\Pi_\phi$ at which the integer $d_\phi(\pi_1) = \Delta_G(\phi, \pi_1)$ equals 1. The irreducible character attached to any $\pi \in \Pi_\phi$ ought then to be the function

$$s \to \langle s, \pi | \pi_1 \rangle = \Delta(\phi', \pi)\Delta(\phi', \pi_1)^{-1}, \qquad s \in \mathcal{S}_\phi, \tag{28.8}$$

where $s$ is the projection onto $\mathcal{S}_\phi$ of the semisimple element $s' \in S_\phi$ attached to the elliptic endoscopic datum $G'$. There is also a parallel interpretation that relates the hypothetical coefficients (28.7) and the packets $T_\phi$ with the representation theory of the finite groups $\mathcal{S}_\phi$. In the case that $G$ is not quasisplit, similar properties are expected, but they are weaker and not completely understood. (See [**LL**], [**Lan10**, §IV.2].)

We have been assuming that the parameter $\phi$ is tempered. Suppose now that $\phi$ is a general parameter. Then $\phi$ is the image in $\Phi(G)$ of a twist $\phi_{M,\lambda}$, for a Levi subgroup $M \subset G$, a tempered parameter $\phi_M \in \Phi_{\mathrm{temp}}(M)$, and a point $\lambda$ in the chamber $(\mathfrak{a}_M^*)_P^+$ in $\mathfrak{a}_M^*$ attached to a parabolic subgroup $P \in \mathcal{P}(M)$. The packet $\Pi_\phi$ can be defined to be the set of irreducible representations obtained by taking the unique irreducible quotient (the Langlands quotient) of each representation

$$\mathcal{I}_P(\pi_{M,\lambda}), \qquad \pi_M \in \Pi_{\phi_M}. \tag{28.9}$$

Similar constructions allow one to define the packet $T_\phi$ in terms of the tempered packet $T_{\phi_M}$. One can thus attach conjectural packets to nontempered parameters. The Langlands classification for real groups [**Lan11**] extends to $p$-adic groups, to the extent that it reduces the general classification to the tempered case [**BW**, §XI.2]. Combined with the expected packet structure of tempered representations, it then gives a conjectural classification of $\Pi(G)$ into a disjoint union of finite packets $\Pi_\phi$, indexed by parameters $\phi \in \Phi(G)$. Moreover, for $\phi$, $\phi_M$, and $\lambda$ as above, the finite group $\mathcal{S}_\phi$ equals the corresponding group $\mathcal{S}_{\phi_M}$ attached to the tempered parameter $\phi_M$ for $M$. We can therefore relate the representations in $\Pi_\phi$ to characters on $\mathcal{S}_\phi$, if we are able to relate the representations in the tempered packet $\Pi_{\phi_M}$ with characters in $\mathcal{S}_{\phi_M}$. However, the nontempered analogues of the character relations (28.5) and (28.7) will generally be false.

Suppose that $G = GL(n)$. In this case, the centralizer $S_\phi$ of the image of any parameter $\phi \in \Phi(G)$ is connected. The group $\mathcal{S}_\phi$ is therefore trivial, and the corresponding packet $\Pi_\phi$ should consequently contain exactly one element. The Langlands classification for $G = GL(n)$ thus takes the form of a bijection between parameters $\phi \in \Phi(G)$ and irreducible representations $\pi \in \Pi(G)$. It has recently been established by Harris and Taylor [**HT**] and Henniart [**He**].

We have assumed that the local field $k$ was nonarchimedean. The analogues for archimedean fields $k = F_v$ of the conjectural properties described above have all been established. They are valid as stated, except that $L_k$ is just the Weil group $W_k$, and the correspondence $\Pi_\phi \to \widehat{\mathcal{S}}_\phi$ is an injection rather than a bijection. As we mentioned earlier, the classification of irreducible representations $\Pi(G)$ in terms of parameters $\phi \in \Phi(G)$ was established by Langlands and Knapp-Zuckermann. (See [**KZ1**].) The transfer identities (28.5) and (28.7) for tempered parameters $\phi$, together with the description of packets in terms of characters $\widehat{\mathcal{S}}_\phi$, were established by Shelstad [**She2**], [**She3**]. In particular, there is a classification of irreducible representations of $G(k)$ in terms of simple invariants attached to the dual group ${}^LG$. One would obviously like to have a similar classification for nonarchimedean fields.

One reason for wanting such a classification is to give a systematic construction of $L$-functions for irreducible representations. Suppose that $k = F_v$ is any completion of $F$. One can attach a local $L$-function $L(s, r)$ and $\varepsilon$-factor $\varepsilon(s, r, \psi)$ of the complex variable $s$ to any (continuous, semisimple) representation $r$ of the local Weil group $W_k$, and any nontrivial additive character $\psi\colon k \to \mathbb{C}$. The $\varepsilon$-factors are needed for the functional equations of $L$-functions attached to representations of the global Weil group $W_F$. Deligne's proof [**D1**] that they exist and have the appropriate properties in fact uses global arguments. Suppose that the local Langlands conjecture holds for $G = G_v$. That is, any irreducible representation $\pi \in \Pi(G)$ lies in the packet $\Pi_\phi$ attached to a unique parameter $\phi$. We write $\phi_W$ for the restriction of $\phi$ to the subgroup $W_k$ of $L_k$. Suppose that $\rho$ is a finite dimensional representation of the $L$-group ${}^LG$. We can then define a local $L$-function

$$L(s, \pi, \rho) = L(s, \rho \circ \phi_W) \tag{28.10}$$

and $\varepsilon$-factor

$$\varepsilon(s, \pi, \rho, \psi) = L(s, \rho \circ \phi_W, \psi) \tag{28.11}$$

in terms of corresponding objects for $W_k$. For example, suppose that $k = F_v$ is nonarchimedean, and that $\pi$, $\rho$, and $\psi$ are unramified. Then $\pi$ is parametrized by a semisimple conjugacy class $c = c(\pi)$ in ${}^LG$. The associated parameter $\phi$: $L_k \to {}^LG$ is trivial on both $SU(2)$ and the inertia subgroup $I_k$ of $W_k$. It maps the element $\mathrm{Frob}_k$ that generates the cyclic quotient $W_k/I_k$ to $c$. In this case, $\varepsilon(s, \pi, \rho, \psi) = 1$, and

$$L(s, \pi, \rho) = \det\big(1 - \rho(c)q^{-s}\big),$$

where $q = q_v$.

Langlands has conjectured that local $L$-functions give canonical normalizing factors for induced representations. Suppose that $\pi \in \Pi(M)$ is an irreducible representation of a Levi subgroup $M$ of $G$ over $k = F_v$. Recall that the unnormalized intertwining operators

$$J_{Q|P}(\pi_\lambda) : \mathcal{I}_P(\pi_\lambda) \longrightarrow \mathcal{I}_Q(\pi_\lambda), \qquad P, Q \in \mathcal{P}(M),$$

between induced representations are meromorphic functions of a complex variable $\lambda \in \mathfrak{a}^*_{M,\mathbb{C}}$. Let $\rho_{Q|P}$ be the adjoint representation of ${}^LM$ on the Lie algebra of the intersection of the unipotent radicals of the parabolic subgroups $\widehat{P}$ and $\widehat{\overline{Q}}$ of $\widehat{G}$. We can then set

$$r_{Q|P}(\pi_\lambda) = L(0, \pi_\lambda, \rho_{Q|P})\big(\varepsilon(0, \pi_\lambda, \rho^\vee_{Q|P}, \psi)L(1, \pi_\lambda, \rho_{Q|P})\big)^{-1}, \tag{28.12}$$

assuming of course that the functions on the right have been defined. Langlands conjectured [**Lan5**, Appendix II] that for a suitable normalization of Haar measures on the groups $N_Q \cap N_{\bar{P}}$, these meromorphic functions of $\lambda$ are an admissible set of normalizing factors, in the sense that they satisfy the conditions of Theorem 21.4. It is this conjecture that Shahidi established in case $G = GL(n)$, and that was used in the applications described in §25 and §26. (We recall that for $GL(n)$, the relevant local $L$ and $\varepsilon$-factors were defined independently of Weil groups. Part of the recent proof of the local Langlands classification for $GL(n)$ by Harris-Taylor and Henniart was to show that these $L$ and $\varepsilon$-factors were the same as the ones attached to representations of $W_k$.)

However, we do not have a general classification of representations in the packets $\Pi_\phi$. We therefore cannot use (28.10) and (28.11) to define the factors on the right hand side of (28.12). The canonical normalization factors are thus not available. This is our pretext for normalizing the weighted characters in a different way.

Instead of normalizing factors $r = \{r_{Q|P}(\pi_\lambda)\}$, we use Harish-Chandra's canonical family $\mu = \{\mu_{Q|P}(\pi_\lambda)\}$ of $\mu$-functions. We recall that

$$\mu_{Q|P}(\pi_\lambda) = \big(J_{Q|P}(\pi_\lambda)J_{P|Q}(\pi_\lambda)\big)^{-1} = \big(r_{Q|P}(\pi_\lambda)r_{P|Q}(\pi_\lambda)\big)^{-1},$$

for any $Q, P \in \mathcal{P}(M)$, $\pi \in \Pi(M)$ and $\lambda \in \mathfrak{a}^*_{M,\mathbb{C}}$. Suppose that $\pi$ is in general position, in the sense that the unnormalized intertwining operators $J_{Q|P}(\pi_\lambda)$ are analytic for $\lambda \in i\mathfrak{a}_M^*$. For fixed $P$, the operator valued family

$$\mathcal{J}_Q(\Lambda, \pi, P) = J_{Q|P}(\pi)^{-1} J_{Q|P}(\pi_\Lambda), \qquad Q \in \mathcal{P}(M),$$

is a $(G, M)$-family of functions of $\Lambda \in i\mathfrak{a}_M^*$. The normalized weighted characters used in the original invariant trace formula were constructed from the product $(G, M)$-family

$$\mathcal{R}_Q(\Lambda, \pi, P) = r_Q(\Lambda, \pi, P)^{-1} \mathcal{J}_Q(\Lambda, \pi, P),$$

where

$$r_Q(\Lambda, \pi, P) = r_{Q|P}(\pi)^{-1} r_{Q|P}(\pi_\Lambda).$$

The normalized weighted characters for the second version are to be constructed from the product $(G, M)$-family

$$\mathcal{M}_Q(\Lambda, \pi, P) = \mu_Q(\Lambda, \pi, P) \mathcal{J}_Q(\Lambda, \pi, P), \tag{28.13}$$

where

$$\mu_Q(\Lambda, \pi, P) = \mu_{Q|P}(\pi)^{-1} \mu_{Q|P}(\pi_{\frac{1}{2}\Lambda}).$$

They are defined by setting

$$J_M(\pi, f) = \operatorname{tr}\big(\mathcal{M}_M(\pi, P)\mathcal{I}_P(\pi, f)\big), \qquad f \in \mathcal{H}(G), \tag{28.14}$$

where

$$\mathcal{M}_M(\pi, P) = \lim_{\Lambda \to 0} \sum_{Q \in \mathcal{P}(M)} \mathcal{M}_Q(\Lambda, \pi, P)\theta_Q(\Lambda)^{-1}, \tag{28.15}$$

as usual. Notice that we are using the same notation for the two sets of normalized weighted characters. It there is any danger of confusion, we can always denote the original objects by $J_M^r(\pi, f)$, and the ones we have just constructed by $J_M^\mu(\pi, f)$.

PROPOSITION 28.2. *The linear form $J_M(\pi, f) = J_M^\mu(\pi, f)$, defined for $\pi \in \Pi(M)$ in general position, is independent of the fixed group $P \in \mathcal{P}(M)$. Moreover, if $\pi \in \Pi_{\mathrm{unit}}(M)$ is any unitary representation, $J_M(\pi_\lambda, f)$ is an analytic function of $\lambda \in i\mathfrak{a}_M^*$.*

The two assertions are among the main results of [**A24**]. We know that for the original weighted characters $J_M^r(\pi, \lambda)$, the assertions are simple consequences of the properties of the normalizing factors $r$. We form a second $(G, M)$-family

$$r_Q(\Lambda, \pi) = r_{Q|\overline{Q}}(\pi)^{-1} r_{Q|\overline{Q}}(\pi_{\frac{1}{2}\Lambda}), \qquad Q \in \mathcal{P}(M),$$

from the normalizing factors. The new weighted characters are then related to the original ones by an expansion

$$J_M^\mu(\pi, f) = \sum_{L \in \mathcal{L}(M)} r_M^L(\pi) J_L^r(\pi^L, f),$$

which one derives easily from the relations between the functions $\{r_{Q|P}(\pi_\lambda)\}$ and $\{\mu_{Q|P}(\pi_\lambda)\}$ [**A24**, Lemma 2.1]. The first assertion follows immediately [**A24**, Corollary 2.2]. To establish the second assertion, one shows that for $\pi \in \Pi_{\mathrm{unit}}(M)$, the functions $r_M^L(\pi_\lambda)$ are analytic on $i\mathfrak{a}_M^*$ [**A24**, Proposition 2.3]. □

## 29. The stable trace formula

In this section, we shall discuss the solution to the problem posed at the end of §27. We shall describe how to stabilize all of the terms in the invariant trace formula. The stabilization is conditional upon the fundamental lemma. It is also contingent upon a generalization of the fundamental lemma, which applies to unramified weighted orbital integrals.

The results are contained in the three papers [**A27**], [**A26**], and [**A29**]. They depend on other papers as well, including some still in preparation. Our discussion will therefore have to be quite limited. However, we can at least try to give a coherent statement of the results. The techniques follow the model of inner twistings of $GL(n)$, outlined in some detail in §25. However, the details here are considerably more elaborate. The results discussed in this section are in fact the most technical of the paper.

We have of course to return to the global setting, with which we were preoccupied before the local interlude of the last section. Then $G$ is a fixed reductive group over the number field $F$. There are two preliminary matters to deal with before we can consider the main problem.

The first is to reformulate the invariant trace formula for $G$. Since it is based on the construction at the end of the last section, this second version does not depend on the normalization of intertwining operators. In some ways, it is slightly less elegant than the original version, but the two are essentially equivalent. In particular, our stabilization of the second version would no doubt give a stabilization of the first, if we had the compatible normalizing factors provided by a refined local correspondence of representations.

Our reformulation of the invariant trace formula entails a couple of other minor changes. It applies to test functions $f$ on the group $G_V = G(F_V)$, where $V$ is a finite set of valuations of $F$ that contains the set $S_{\mathrm{ram}} = S_{\mathrm{ram}}(G)$ of places at which $G$ is ramified. We can take $V$ to be large. However, we want to distinguish it from the large finite set $S$ that occurs on the geometric side of the original formula. In relating the two versions of the formula, $S$ would be a finite set of places that is large relative to both $V$ and the support of some chosen test function on $G_V$. The terms in our second version will be indexed by conjugacy classes in $M_V$ (rather than $M(\mathbb{Q})$-conjugacy classes or $(M,S)$-classes) and irreducible representations of $M_V$ (rather than automorphic representations of $M(\mathbb{A})$). In order to allow for induction arguments, we also need to work with equivariant test functions on $G_V$. We fix a suitable central torus $Z \subset G$ over $F$, and a character $\zeta$ on $Z(\mathbb{A})/Z(F)$. We then assume that $V$ contains the larger finite set $S_{\mathrm{ram}}(G,\zeta)$ of valuations at which any of $G$, $Z$ or $\zeta$ ramifies. We write $G_V^Z$ for the subgroup of elements $x \in G_V$ such that $H_G(x)$ lies in the image of $\mathfrak{a}_Z$ in $\mathfrak{a}_G$, and $\zeta_V$ for the restriction of $\zeta$ to $Z_V$. Our test functions are to be taken from the Hecke algebra

$$\mathcal{H}(G,V,\zeta) = \mathcal{H}(G_V^Z,\zeta_V),$$

and its invariant analogue

$$\mathcal{I}(G,V,\zeta) = \mathcal{I}(G_V^Z,\zeta_V).$$

Observe that if $Z$ equals 1, $G_V^Z$ equals the group $G(F_V)^1$. In this case, $\mathcal{H}(G,V,\zeta)$ embeds in the original space $\mathcal{H}(G) = \mathcal{H}\big(G(\mathbb{A})^1\big)$ of test functions.

There is a natural projection from the subspace $\mathcal{H}(G,V)=\mathcal{H}\big(G(F_V)^1\big)$ of $\mathcal{H}(G)$ onto $\mathcal{H}(G,V,\zeta)$. Let $J$ be the basic linear form on $\mathcal{H}(G)$ whose two expansions give the noninvariant trace formula. If $f$ lies in $\mathcal{H}(G)$, and $f_z$ denotes the translate of $f$ by a point $z\in Z(\mathbb{A})^1$, the integral

$$\int_{Z(F)\backslash Z(\mathbb{A})^1} J(f_z)\zeta(z)\mathrm{d}z$$

is well defined. If $f$ belongs to the subspace $\mathcal{H}(G,V)$, the integral depends only on the image of $f$ in $\mathcal{H}(G,V,\zeta)$. It therefore determines a linear form on $\mathcal{H}(G,V,\zeta)$, which we continue to denote by $J$. To make this linear form invariant, we define mappings

$$\phi_M:\ \mathcal{H}(G,V,\zeta)\ \longrightarrow\ \mathcal{I}(M,V,\zeta),\qquad\qquad M\in\mathcal{L},\tag{29.1}$$

in terms of the weighted characters at the end of last section. In other words, the operator valued weight factor is to be attached to a product over $v\in V$ of $(G,M)$-families (28.13), rather than the $(G,M)$-family defined in §23 in terms of normalized intertwining operators. The mapping itself is defined by an integral analogous to (23.2) (with $X=0$), but over a domain $i\mathfrak{a}^*_{M,Z}/i\mathfrak{a}^*_{G,Z}$ (where $i\mathfrak{a}^*_{M,Z}$ is the subspace of elements in $i\mathfrak{a}^*_M$ that vanish on the image of $i\mathfrak{a}^*_Z$ on $i\mathfrak{a}^*_M$). It follows from the proof of Propositions 23.1 and 28.2 that $\phi_M$ does indeed map $\mathcal{H}(G,V,\zeta)$ to $\mathcal{I}(M,V,\zeta)$. We can therefore define an invariant linear form $I=I^G$ on $\mathcal{H}(G,V,\zeta)$ by the analogue of (23.10). The problem is to transform the two expansions of Theorem 23.4 into two expansions of this new linear form.

We define weighted orbital integrals $J_M(\gamma,f)$ for functions $f\in\mathcal{H}(G,V,\zeta)$ exactly as in §18. The element $\gamma$ is initially a conjugacy class in $M_V^Z$. However, $J_M(\gamma,f)$ depends only on the image of $\gamma$ in the space $\mathcal{D}(M_V^Z,\zeta_V)$ of invariant distributions on $M_V^Z$, defined as at the beginning of §28. We can therefore regard $J_M(\cdot,f)$ as a linear form on the subspace $\mathcal{D}_{\mathrm{orb}}(M_V^Z,\zeta_V)$ of $\mathcal{D}(M_V^Z,\zeta_V)$ generated by conjugacy classes. There is actually a more subtle point, which we may as well raise here. As it turns out, stabilization requires that $J_M(\gamma,f)$ be defined for *all* elements in the space $\mathcal{D}(M_V^Z,\zeta)$. If $v$ is nonarchimedean, $\mathcal{D}_{\mathrm{orb}}(M_v,\zeta_v)$ equals $\mathcal{D}(M_v,\zeta_v)$. In this case, there is nothing further to do. However, if $v$ is archimedean, $\mathcal{D}(M_v,\zeta_v)$ is typically much larger than $\mathcal{D}_{\mathrm{orb}}(M_v,\zeta_v)$, thanks to the presence of normal derivatives along conjugacy classes. The construction of weighted orbital integrals at distributions in this larger space demands a careful study of the underlying differential equations. Nevertheless, one can in the end extend $J_M(\gamma,f)$ to a canonical linear form on the space $\mathcal{D}(M_V^Z,\zeta_V)$. (See [**A31**].) One then uses the mappings (29.1) as in (23.3), to define invariant distributions

$$I_M(\gamma,f),\qquad\qquad \gamma\in\mathcal{D}(M_V^Z,\zeta_V),\ \ f\in\mathcal{H}(G,V,\zeta).$$

These distributions, with $\gamma$ restricted to the subspace $\mathcal{D}_{\mathrm{orb}}(M_V^Z,\zeta_V)$ of $\mathcal{D}(M_V^Z,\zeta_V)$, will be the terms in the geometric expansion.

The coefficients in the geometric expansion should really be regarded as elements in $\mathcal{D}(M_V^Z,\zeta_V)$, or rather, the appropriate completion $\widehat{\mathcal{D}}(M_V^Z,\zeta_V)$ of $\mathcal{D}(M_V^Z,\zeta_V)$. As such, they have a natural pairing with the linear forms $I_M(\cdot,f)$ on $\mathcal{D}(M_V^Z,\zeta_V)$. However, we would like to work with an expansion like that of (23.11). We therefore identify $\widehat{\mathcal{D}}(M_V^Z,\zeta_V)$ with the dual space of $\mathcal{D}(M_V^Z,\zeta_V)$ by fixing a suitable basis of $\Gamma(M_V^Z,\zeta_V)$ of $\mathcal{D}(M_V^Z,\zeta_V)$. Since we can arrange that the elements in

$$\Gamma_{\mathrm{orb}}(M_V^Z, \zeta_V) = \Gamma(M_V^Z, \zeta_V) \cap \mathcal{D}_{\mathrm{orb}}(M_V^Z, \zeta_V)$$

be parametrized by conjugacy classes in $\overline{M}_V = M_V/Z_V$, we will still be dealing essentially with conjugacy classes. We define coefficient functions on $\Gamma(M_V^Z, \zeta_V)$ by compressing the corresponding coefficients in (23.11). It is done in two stages. For a given $\gamma_M \in \Gamma(M_V, \zeta_V)$, we choose a large finite set $S \supset V$, and take $k$ to be a conjugacy class in $\overline{M}(F_V^S)$ that meets $K_V^S$. We then define a function $a_{\mathrm{ell}}^M(\gamma_M \times k)$ as a certain finite linear combination of coefficients $a^M(\gamma)$ in (23.11), taken over those $(M, S)$-equivalence classes $\gamma \in \Gamma(M)_S$ that map to $\gamma_M \times k$ [**A27**, (2.6)]. For any given $k$, we can form the unramified weighted orbital integral

$$r_M^G(k) = J_M(k, u_S^V), \tag{29.2}$$

where $u_S^V = u_S^{V,\zeta}$ is the projection onto $\mathcal{H}(G_S^V, \zeta_S^V)$ of the characteristic function of $K_S^V$. If $\gamma$ is now an element in $\Gamma(G_V^Z, \zeta_V)$, we set

$$a^G(\gamma) = \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_k a_{\mathrm{ell}}^M(\gamma_M \times k) r_M^G(k), \tag{29.3}$$

where $\gamma \to \gamma_M$ is the restriction operator that is adjoint to induction of conjugacy classes (and invariant distributions). (See [**A27**, (2.8), (1.9)].)

PROPOSITION 29.1. *Suppose that $f \in \mathcal{H}(G, V, \zeta)$. Then*

$$I(f) = \sum_{M \in \mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma \in \Gamma(M,V,\zeta)} a^M(\gamma) I_M(\gamma, f),$$

*where $\Gamma(M, V, \zeta)$ is a discrete subset of $\Gamma(M_V^Z, \zeta_V)$ that contains the support of $a^M(\gamma)$, and on which $I_M(\gamma, f)$ has finite support.*

See [**A27**, Proposition 2.2]. □

The spectral expansion of $I(f)$ begins with the decomposition

$$I(f) = \sum_{t \geq 0} I_t(f), \qquad f \in \mathcal{H}(G, V, \zeta), \tag{29.4}$$

relative to the norms $t$ of archimedean infinitesimal characters. The summand $I_t(f)$ is as in Remark 10 of §23, the invariant version of a linear form $J_t(f)$ on $\mathcal{H}(G, V, \zeta)$ defined as at the end of §21. The sum itself satisfies the weak multiplier estimate (23.13), and hence converges absolutely. We shall describe the spectral expansion of $I_t(f)$.

We define weighted characters $J_M(\pi, f)$, for functions $f \in \mathcal{H}(G, V, \zeta)$, by a minor modification of the construction of §22. The element $\pi$ lies in $\Pi_{\mathrm{unit}}(M_V, \zeta_V)$, and can therefore be regarded as a distribution in the space $\mathcal{F}(G_V^Z, \zeta_V)$. As with the mappings (29.1), $J_M(\pi, f)$ is defined in terms of the product over $v \in V$ of $(G, M)$-families in (28.13), and an integral analogous to (22.4), but over a domain $i\mathfrak{a}_{M,Z}^*/i\mathfrak{a}_{G,Z}^*$. We then form corresponding invariant distributions $I_M(\pi, f)$ from the mappings (29.1) as in (23.4) (or rather the special case of (23.4) with $X = 0$).

The coefficients in the spectral expansion are parallel to those in the geometric expansion. The analogues of the classes $k$ in (29.3) are families

$$c = \{c_v : \ v \notin V\}$$

of semisimple conjugacy classes in ${}^L M$. We allow only those classes of the form $c = c(\pi^V)$, where $\pi^V = \pi^V(c)$ is an unramified representation of $M^V = M(\mathbb{A}^V)$

whose $Z^V$-central character is equal to the corresponding component $\zeta^V$ of $\zeta$. There is an obvious action

$$c \longrightarrow c_\lambda = \{c_{v,\lambda} : \ v \notin S\}, \qquad \lambda \in i\mathfrak{a}^*_{M,Z},$$

such that $\pi^V(c_\lambda) = \pi^V(c)_\lambda$. If $\pi^V(c)$ is unitary, we write

$$\pi \times c = \pi \otimes \pi^V(c)$$

for the representation in $\Pi_{\mathrm{unit}}\big(M(\mathbb{A}), \zeta\big)$ attached to any representation $\pi$ in $\Pi_{\mathrm{unit}}(M_V, \zeta_V)$. Similar notation holds if $\pi$ belongs to the quotient $\Pi_{\mathrm{unit}}(M_V^Z, \zeta_V)$ of $\Pi_{\mathrm{unit}}(M_V, \zeta_V)$, with the understanding that $\pi$ is identified with a representative in $\Pi_{\mathrm{unit}}(M_V, \zeta_V)$. We define $\Pi_{t,\mathrm{disc}}(M, V, \zeta)$ to be the set of representations $\pi \in \Pi_{\mathrm{unit}}(M_V^Z, \zeta_V)$ such that for some $c$, $\pi \times c$ belongs to the subset $\Pi_{t,\mathrm{disc}}(M, \zeta)$ of $\Pi_{t,\mathrm{disc}}(M)$ attached to $\zeta$. We also define $\mathcal{C}^V_{\mathrm{disc}}(M, \zeta)$ to be the set of $c$ such that $\pi \times c$ belongs to $\Pi_{t,\mathrm{disc}}(M, \zeta)$, for some $t$ and some $\pi \in \Pi_{t,\mathrm{disc}}(M, \zeta)$.

If $c$ belongs to $\mathcal{C}^V_{\mathrm{disc}}(M, \zeta)$ and $\lambda \in \mathfrak{a}^*_{M,Z,\mathbb{C}}$, the unramified $L$-function

$$L(s, c_\lambda, \rho) = \prod_{v \notin V} \det\big(1 - \rho(c_{v,\lambda}) q_v^{-s}\big)^{-1}$$

converges absolutely for $\mathrm{Re}(s)$ large. In case $\rho$ is the representation $\rho_{Q|P}$ of ${}^L M$, it is known that $L(s, c_\lambda, \rho)$ has analytic continuation as a meromorphic function of $s$, and that for any fixed $s$, $L(s, c_\lambda, \rho)$ is a meromorphic function of $\lambda \in \mathfrak{a}^*_{M,Z,\mathbb{C}}$. Following (28.12), we define the unramified normalizing factor

$$r_{Q|P}(c_\lambda) = L(0, c_\lambda, \rho_{Q|P}) L(1, c_\lambda, \rho^\vee_{Q|P})^{-1}, \qquad P, Q \in \mathcal{P}(M).$$

We then define a $(G, M)$-family

$$r_Q(\Lambda, c_\lambda) = r_{Q|\bar{Q}}(c_\lambda)^{-1} r_{Q|\bar{Q}}(c_{\lambda + \frac{1}{2}\Lambda}), \qquad Q \in \mathcal{P}(M),$$

and a corresponding meromorphic function

$$r^G_M(c_\lambda) = \lim_{\Lambda \to 0} \sum_{Q \in \mathcal{P}(M)} r_Q(\Lambda, c_\lambda) \theta_Q(\Lambda)^{-1} \tag{29.5}$$

of $\lambda$. One shows that $r^G_M(c_\lambda)$ is an analytic function of $\lambda \in i\mathfrak{a}^*_{M,Z}$, whose integral against any rapidly decreasing function of $\lambda$ converges [**A27**, Lemma 3.2]. If $\pi$ is now a representation in $\Pi_{t,\mathrm{unit}}(G_V, \zeta_V)$, we define

$$a^G(\pi) = \sum_{M \in \mathcal{L}} |W_0^M| |W_0^G|^{-1} \sum_c a^M_{\mathrm{disc}}(\pi_M \times c) r^G_M(c), \tag{29.6}$$

where $\pi \to \pi_M$ is the restriction operation that is adjoint to induction of characters. We define a subset $\Pi_t(G, V, \zeta)$ of $\Pi_{t,\mathrm{unit}}(G_V, \zeta_V)$, which contains the support of $a^G(\pi)$, and a measure $d\pi$ on $\Pi_t(G, V, \zeta)$ by following the appropriate analogues of (22.6) and (22.7). (See [**A27**, p. 205].)

Proposition 29.2. *Suppose that $f \in \mathcal{H}(G, V, \zeta)$. Then*

$$I_t(f) = \sum_{M \in \mathcal{L}} |W_0^M| |W_0^G|^{-1} \int_{\Pi_t(M, V, \zeta)} a^M(\pi) I_M(\pi, f) \mathrm{d}\pi.$$

(See [**A27**, Proposition 3.3].) □

The identity obtained from (29.4) and Propositions 29.1 and 29.2 is the required reformulation of the invariant trace formula. Observe that the spectral factors $r_M^G(c)$ in the coefficients (29.6) are constructed from canonical unramified normalizing factors, while their counterparts $r_M^G(\pi_\lambda)$ in the earlier coefficients (22.8) were constructed from noncanonical global normalizing factors. This is a consequence of the modified definition of the mappings (29.1). The geometric factors $r_M^G(k)$ in the coefficients (29.3) have no counterparts in the earlier coefficients (19.6). They occur in the original geometric expansion (23.11) instead as implicit factors of the distributions $I_M(\gamma, f)$. This is because the set $V$ is fixed, whereas $S$ is large, in a sense that depends on the support of $f \in \mathcal{H}(G, V, \zeta)$.

The second preliminary matter pertains directly to the notion of stability. If $T$ is a maximal torus in $G$ over $F$, and $v$ is archimedean, the subset $\mathcal{D}(T/F_v)$ of $\mathcal{E}(T/F_v)$ in (27.4) can be proper. On the other hand, the $v$-components of the summands $f_G^\kappa(\delta)$ in Langlands' stabilization (27.6) are parametrized by points $\kappa_v$ in the dual group $\mathcal{K}(T/F_v)$ of $\mathcal{E}(T/F_v)$. If $\mathcal{D}(T/F_v)$ is proper in $\mathcal{E}(T/F_v)$, the mapping

$$f_v \longrightarrow f_{v,G}^{\kappa_v}(\delta_v), \qquad \kappa_v \in \mathcal{K}(T/F_v),\ f_v \in \mathcal{H}(G_v),$$

from functions $f_{v,G} \in \mathcal{I}\big(G(F_v)\big)$ to functions on $\mathcal{K}(T/F_v)$, is not surjective. This makes it difficult to characterize the image of the collective transfer mappings

$$\mathcal{I}(G, V, \zeta) \longrightarrow \bigoplus_{G'} S\mathcal{I}(\widetilde{G}', V, \widetilde{\zeta}').$$

It was pointed out by Vogan that the missing elements in $\mathcal{D}(T/F_v)$ could be attached to other groups. He observed that $\mathcal{E}(T/F_v)$ could be expressed as a disjoint union

$$\mathcal{E}(T/F_v) = \coprod \mathcal{D}_{\alpha_v}(T/F_v),$$

over sets $\mathcal{D}_{\alpha_v}(T/F_v)$ attached to a finite collection of groups $G_{\alpha_v}$ over $F_v$ related by inner twisting. (See [**AV**] and [**ABV**] for extensions and applications of this idea.) Kottwitz then formulated the observations of Vogan in terms of the transfer factors. His formulation gives rise to a notion that was called a $K$-group in [**A25**]. Over the global field $F$, a $K$-group is an algebraic variety

$$G = \coprod_\alpha G_\alpha, \qquad \alpha \in \pi_0(G),$$

whose connected components are reductive algebraic groups $G_\alpha$ over $F$, and which is equipped with two kinds of supplementary structure. One consists of cohomological data, which include inner twists $\psi_{\alpha\beta}\colon G_\beta \to G_\alpha$ between any two components. The other is a local product structure, which for any finite set $V \supset V_{\mathrm{ram}}(G)$ allows us to identify the set

$$G_V = \coprod_\alpha G_{\alpha,V} = \coprod_\alpha G_\alpha(F_V)$$

with a product

$$\prod_{v\in V} G_v = \prod_{v \in V} G_v(F_v)$$

of $F_v$-points in local $K$-groups $G_v$ over $F_v$. The local $K$-group $G_v$ is a finite disjoint union

$$G_v = \coprod_{\alpha_v} G_{\alpha_v}$$

of connected groups if $v$ is archimedean, but it is just a connected group if $v$ is nonarchimedean. In particular, the set $V_{\mathrm{ram}}(G) = V_{\mathrm{ram}}(G_\alpha)$ is independent of $\alpha$. (See [**A27**, p. 209–211].)

We assume for the rest of this section that $G$ is a $K$-group over $F$. Many concepts for connected groups carry over to this new setting without change in notation. For example, we define $\Gamma_{\mathrm{reg}}(G_V)$ to be the disjoint union over $\alpha$ of the corresponding sets $\Gamma_{\mathrm{reg}}(G_{\alpha,V})$ for the connected groups $G_\alpha$. Similar conventions apply to the sets $\Pi(G_V)$, $\Pi_{\mathrm{unit}}(G_V)$, and $\Pi_{\mathrm{temp}}(G_V)$ of irreducible representations. We define compatible central character data $Z = \{Z_\alpha\}$ and $\zeta = \{\zeta_\alpha\}$ for $G$ by choosing data $Z_\alpha$ and $\zeta_\alpha$ for any one component $G_\alpha$. This allows us to form the sets $\Pi(G_V, \zeta_V)$, $\Pi_{\mathrm{unit}}(G_V, \zeta_V)$, and $\Pi_{\mathrm{temp}}(G_V, \zeta_V)$ as disjoint unions of corresponding sets attached to components $G_\alpha$. We can also define the vector spaces $\mathcal{H}(G_V, \zeta_V)$, $\mathcal{H}(G, V, \zeta)$, $\mathcal{I}(G_V, \zeta_V)$, $\mathcal{I}(G, V, \zeta)$, $\mathcal{D}(G_V^Z, \zeta_V)$, $\mathcal{F}(G_V^Z, \zeta_V)$, etc., by taking direct sums of the corresponding spaces attached to components $G_\alpha$. Finally, we define sets $\Gamma(G_V^Z, \zeta_V)$, $\Gamma(G, V, \zeta)$, $\Pi_t(G_V^Z, \zeta_V)$, $\Pi_t(G, V, \zeta)$, and $\Pi_{t,\mathrm{disc}}(G, V, \zeta)$, again as disjoint unions of corresponding sets attached to components $G_\alpha$.

There is also a notion of Levi subgroup (or more correctly, Levi $K$-subgroup) $M$ of $G$. For any such $M$, the objects $\mathfrak{a}_M$, $A_M$, $W(M)$, $\mathcal{P}(M)$, $\mathcal{L}(M)$, and $\mathcal{F}(M)$ all have meaning, and play a role similar to that of the connected case. (See [**A25**, §1].) We again write $\mathcal{L}$ for the set $\mathcal{L}(M_0)$ attached to a fixed minimal Levi subgroup $M_0$ of $G$. With these conventions, the objects in the expansions of Proposition 29.1 and 29.2 now all have meaning for the $K$-group $G$ over $F$. The invariant trace formula for $G$ is an identity

$$\begin{aligned} &\sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma(M,V,\zeta)} a^M(\gamma) I_M(\gamma, f) \\ &= \sum_t \sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Pi_t(M,V,\zeta)} a^M(\pi) I_M(\pi, f) d\pi, \end{aligned} \tag{29.7}$$

which holds for any $f \in \mathcal{H}(G, V, \zeta)$. It is obtained by applying (29.4) and Propositions 29.1 and 29.2 to the components $f_\alpha \in \mathcal{H}(G_{\alpha,V}, \zeta_{\alpha,V})$ of $f$, and then summing the resulting expansions over $\alpha$.

Stable conjugacy in $G_V$ has to be formulated slightly differently. We define two strongly regular elements $\gamma \in G_{\alpha,V}$ and $\delta \in G_{\beta,V}$ to be *stably conjugate* if $\psi_{\alpha\beta}(\delta)$ is stably conjugate in $G_{\alpha,V}$ to $\gamma$. We then define $S\mathcal{I}(G_V, \zeta_V)$ as a space of functions on the set $\Delta_{\mathrm{reg}}(G_V)$ of strongly regular stable conjugacy classes in $G_V$. This leads to the notion of a stable distribution on $G_V$, and allows us to define the subspaces $S\mathcal{D}(G_V, \zeta_V)$ and $S\mathcal{F}(G_V, \zeta_V)$ of stable distributions in $\mathcal{D}(G_V, \zeta_V)$ and $\mathcal{F}(G_V, \zeta_V)$ respectively. The conventions here are just minor variations of what we used for connected groups. We define a quasisplit inner twist of $G$ to be a *connected*, quasisplit group $G^*$ over $F$, together with a family of inner twists $\psi_\alpha\colon G_\alpha \to G^*$ of connected groups such that $\psi_\beta = \psi_\alpha \circ \psi_{\alpha\beta}$. For any such $G^*$, there is a canonical injection $\delta \to \delta^*$ from $\Delta_{\mathrm{reg}}(G_V)$ to $\Delta_{\mathrm{reg}}(G_V^*)$. There is also a surjective mapping $S^* \to S$ from the space of stable distributions on $G_V^*$ to the space of stable distributions on $G_V$. We say that $G$ is *quasisplit* if one of the components $G_\alpha$ is quasisplit. In this case, the mapping $\delta \to \delta^*$ is a bijection, and the mapping $S^* \to S$ is an isomorphism.

Because the components $G_\alpha$ of $G$ are related by inner twists, they can all be assigned a common dual group $\widehat{G}$, and a common $L$-group ${}^LG$. We recall that endoscopic data were defined entirely in terms of $\widehat{G}$. We can therefore regard them as objects $G'$ attached to the $K$-group $G$. The same holds for auxiliary data $\widetilde{G}'$ and $\widetilde{\xi}'$ attached to $G'$. Similarly, local endoscopic data $G'_v$, with auxiliary data $\widetilde{G}'_v$ and $\widetilde{\xi}'_v$, are objects attached to the local $K$-group $G_v$.

The main new property is a natural extension of the Langlands-Shelstad construction of local transfer factors to $K$-groups. For any $G_v$, $\widetilde{G}'_v$ and $\widetilde{\xi}'_v$, it provides a function $\Delta_{G_v}(\delta'_v, \gamma_v)$ of $\delta'_v \in \Delta_{G\text{-reg}}(\widetilde{G}'_v)$ and $\gamma_v \in \Gamma_{\mathrm{reg}}(G_v)$. (See [**A25**, §2].) This is the essence of the observations of Kottwitz and Vogan. It has two implications. One is that the transfer factors are now built around sets $\mathcal{D}(T/F_v)$, which are attached to the local $K$-group $G_v$, and are equal to the subgroups $\mathcal{E}(T/F_v)$ of $H^1(F_v, T)$. This places the theory of real and $p$-adic groups on an even footing. The other concerns a related point, which we did not raise earlier. The original Langlands-Shelstad transfer factor attached to $G'_v$ (and $(\widetilde{G}'_v, \widetilde{\xi}'_v)$) depends on an arbitrary multiplicative constant. If $G'_v$ is the localization of a global endoscopic datum, the product over $v$ of these constants equals 1. However, if $G'_v$ is taken in isolation, the constant reflects an intrinsic lack of uniqueness in the correspondence $f_v \to f'_v$. The extension of the transfer factors to $G_v$ still depends on an arbitrary multiplicative constant. However, the constants for the components $G_{\alpha_v}$ of $G_v$ can all be specified in terms of the one constant for $G_v$.

Thus, despite their ungainly appearance, $K$-groups streamline some aspects of the study of connected groups. This is the reason for introducing them. If we are given a connected reductive group $G_1$ over $F$, we can find a $K$-group $G$ over $F$ such that $G_{\alpha_1} = G_1$ for some $\alpha_1 \in \pi_1(G)$. Moreover, $G$ is uniquely determined by $G_1$, up to a natural notion of isomorphism. In particular, for any connected quasisplit group $G^*$, there is a quasisplit $K$-group $G$ such that $G_{\alpha^*} = G^*$, for some $\alpha^* \in \pi_0(G)$.

Let $V$ be a fixed finite set of valuations that contains $S_{\mathrm{ram}}(G, Z, \zeta)$. Suppose that for each $v \in V$, $G'_v$ represents an endoscopic datum $(G'_v, \mathcal{G}'_v, s'_v, \xi'_v)$ for $G$ over $F_v$, equipped with auxiliary data $\widetilde{G}'_v \to G'_v$ and $\widetilde{\xi}'_v \colon \mathcal{G}'_v \to {}^L\widetilde{G}'_v$, and a corresponding choice of local transfer factor $\Delta_v = \Delta_{G_v}$. We are assuming the Langlands-Shelstad transfer conjecture. Applied to each of the components $G_{\alpha_v}$ of $G_v$, it gives a mapping $f_v \to f'_v = f_v^{\widetilde{G}'}$ from $\mathcal{H}(G_v, \zeta_v)$ to $S\mathcal{I}(\widetilde{G}'_v, \widetilde{\zeta}'_v)$, which can be identified with a mapping $a_v \to a'_v$ from $\mathcal{I}(G_v, \zeta_v)$ to $S\mathcal{I}(\widetilde{G}'_v, \widetilde{\zeta}'_v)$. We write $\widetilde{G}'_V$, $\widetilde{\zeta}'_V$, and $\widetilde{\xi}'_V$ for the product over $v \in V$ of $\widetilde{G}'_v$, $\widetilde{\zeta}'_v$, and $\widetilde{\xi}'_v$ respectively. The product

$$\prod_v a_v \longrightarrow \prod_v a'_v, \qquad a_v \in \mathcal{I}(G_v, \zeta_v),$$

then gives a linear transformation $a \to a'$ from $\mathcal{I}(G_V, \zeta_V)$ to $S\mathcal{I}(\widetilde{G}'_V, \widetilde{\zeta}'_V)$. This mapping is attached to the product $G'_V$ of data $G'_v$, which we can think of as an endoscopic datum for $G$ over $F_V$, equipped with auxiliary data $\widetilde{G}'_V$ and $\widetilde{\xi}'_V$, and a corresponding product $\Delta_V$ of local transfer factors. We can think of the transfer factor $\Delta_V$ over $F_V$ as the primary object, since it presupposes a choice of the other objects $G'_V$, $\widetilde{G}'_V$, $\widetilde{\zeta}'_V$ and $\widetilde{\xi}'_V$.

Letting $G'_V$ vary, we obtain a mapping

$$\mathcal{I}(G_V, \zeta_V) \longrightarrow \prod_{G'_V} S\mathcal{I}(\widetilde{G}'_V, \widetilde{\zeta}'_V) \tag{29.8}$$

by putting together all of the individual images $a'$. Notice that we have taken a direct product rather than a direct sum. This is because $G'_V$ ranges over the infinite set of endoscopic data, equipped with auxiliary data $\widetilde{G}'_V$ and $\widetilde{\xi}'_V$, rather than the finite set of isomorphism classes. However, the fact that $G$ is a $K$-group makes it possible to characterize the image of $\mathcal{I}(G_V, \zeta_V)$ in this product. The image fits into a sequence of inclusions

$$\mathcal{I}^{\mathcal{E}}(G_V, \zeta_V) \subset \bigoplus_{\{G'_V\}} \mathcal{I}^{\mathcal{E}}(G'_V, G_V, \zeta_V) \subset \prod_{\Delta_V} S\mathcal{I}(\widetilde{G}'_V, \widetilde{\zeta}'_V)$$

in which the summand $\mathcal{I}^{\mathcal{E}}(G'_V, G_V, \zeta_V)$ depends only on the $F_V$-isomorphism class of $G'_V$. Roughly speaking, $\mathcal{I}^{\mathcal{E}}(G'_V, G_V, \zeta_V)$ is the subspace of products $\prod a'_v$ of functions attached to choices of transfer factors $\Delta_V$ for $\{G'_V\}$ that have the appropriate equivariance properties relative to variations in these choices. The space $\mathcal{I}^{\mathcal{E}}(G_V, \zeta_V)$ is defined as the subspace of functions in the direct sum whose various components are compatible under restriction to common Levi subgroups. One shows that the transfer mapping gives an isomorphism

$$a \longrightarrow a^{\mathcal{E}} = \prod_{\Delta_V} a', \qquad a \in \mathcal{I}(G_V, \zeta_V),$$

from $\mathcal{I}(G_V, \zeta_V)$ onto $\mathcal{I}^{\mathcal{E}}(G_V, \zeta_V)$. This in turn determines an isomorphism from the quotient

$$\mathcal{I}(G, V, \zeta) = \mathcal{I}(G_V^Z, \zeta_V)$$

of $\mathcal{I}(G_V, \zeta_V)$ onto the corresponding quotient

$$\mathcal{I}^{\mathcal{E}}(G, V, \zeta) = \mathcal{I}^{\mathcal{E}}(G_V^Z, \zeta_V)$$

of $\mathcal{I}^{\mathcal{E}}(G_V, \zeta_V)$. (See [**A31**].) The image fits into a sequence of inclusions

$$\mathcal{I}^{\mathcal{E}}(G_V^Z, \zeta_V) \subset \bigoplus_{\{G'_V\}} \mathcal{I}^{\mathcal{E}}(G'_V, G_V^Z, \zeta_V) \subset \prod_{G'_V} S\mathcal{I}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big). \tag{29.9}$$

The mappings of functions we have described have dual analogues for distributions. Given $G'_V$ (with auxiliary data $\widetilde{G}'_V$ and $\widetilde{\xi}'_V$), assume that $\delta'$ belongs to the space of stable distributions $S\mathcal{D}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$. If $f$ belongs to $\mathcal{H}(G, V, \zeta)$, the transfer $f'$ of $f$ can be evaluated at $\delta'$. Since $f \to f'(\delta')$ belongs to $\mathcal{D}(G_V^Z, \zeta_V)$, we can write

$$f'(\delta') = \sum_{\gamma \in \Gamma(G_V^Z, \zeta_V)} \Delta_G(\delta', \gamma) f_G(\gamma), \tag{29.10}$$

for complex numbers $\Delta_G(\delta', \gamma)$ that depend linearly on $\delta'$. Now (29.9) is dual to a sequence of surjective linear mappings

$$\prod_{G'_V} S\mathcal{D}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big) \;\mapsto\; \bigoplus_{\{G'_V\}} \mathcal{D}^{\mathcal{E}}(G'_V, G_V^Z, \zeta_V) \;\mapsto\; \mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$$

between spaces of distributions. Since $f'$ is the image of the function $f_G \in \mathcal{I}(G, V, \zeta)$, $f'(\delta')$ depends only on the image $\delta$ of $\delta'$ in $\mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$. In other words, $f'(\delta')$ equals $f_G^{\mathcal{E}}(\delta)$, where $f_G^{\mathcal{E}}$ is the image of $f_G$ in $\mathcal{I}^{\mathcal{E}}(G, V, \zeta)$. The same is therefore true of the coefficients $\Delta_G(\delta', \gamma)$. We can write

$$\Delta_G(\delta, \gamma) = \Delta_G(\delta', \gamma), \qquad \gamma \in \Gamma(G_V^Z, \zeta_V),$$

for complex numbers $\Delta_G(\delta, \gamma)$ that depend linearly on $\delta \in \mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$. We note that the image in $\mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$ of the subspace

$$S\mathcal{D}\big((G_V^*)^{Z^*}, \zeta_V^*\big) \xrightarrow{\sim} S\mathcal{D}(G_V^*, G_V^Z, \zeta_V)$$

can be identified with the space $S\mathcal{D}(G_V^Z, \zeta_V)$ of stable distributions in $\mathcal{D}(G_V^Z, \zeta_V)$.

The constructions above were given in terms of products $G'_V$ of local endoscopic data for $G$. The stabilization of the trace formula is based primarily on global endoscopic data, particularly the subset $\mathcal{E}_{\mathrm{ell}}(G, V)$ of global isomorphism classes in $\mathcal{E}_{\mathrm{ell}}(G)$ that are unramified outside of $V$. If $G'$ is any endoscopic datum for $G$ over $F$, we can form the product $G'_V$ of its completions. We can also attach auxiliary data $\widetilde{G}'_V$ and $\widetilde{\xi}'_V$ for $G'_V$ to global auxiliary data $\widetilde{G}'$ and $\widetilde{\xi}'$ for $G'$. The datum $G'_V$, together with $\widetilde{G}'_V$ and $\widetilde{\xi}'_V$, indexes a component on the right hand side of (29.9). There are of course other components in (29.9) that do not come from global endoscopic data.

We are trying to formulate stable and endoscopic analogues of the terms in the invariant trace formula (29.7). We start with the local terms $I_M(\gamma, f)$ on the geometric side. Specializing the distributional transfer coefficients above to Levi subgroups $M \in \mathcal{L}$, we can define a linear form

$$I_M(\delta, f) = \sum_{\gamma \in \Gamma(M_V^Z, \zeta_V)} \Delta_M(\delta, \gamma) I_M(\gamma, f), \tag{29.11}$$

for any $\delta \in \mathcal{D}^{\mathcal{E}}(M_V^Z, \zeta_V)$. However, the true endoscopic analogue of $I_M(\gamma, f)$ is a more interesting object. It is defined inductively in terms of an important family $\mathcal{E}_{M'}(G)$ of global endoscopic data for $G$.

Suppose that $M'$ represents a global endoscopic datum $(M', \mathcal{M}', s'_M, \xi'_M)$ for $M$, which is elliptic and unramified outside of $V$. We assume that $\mathcal{M}'$ is an $L$-subgroup of ${}^LM$ and that $\xi'_M$ is the identity embedding. We define $\mathcal{E}_{M'}(G)$ to be the set of endoscopic data $(G', \mathcal{G}', s', \xi')$ for $G$, taken up to translation of $s'$ by $Z(\widehat{G})^\Gamma$, in which $s'$ lies in $s'_M Z(\widehat{M})^\Gamma$, $\widehat{G}'$ is the connected centralizer of $s'$ in $\widehat{G}$, $\mathcal{G}'$ equals $\mathcal{M}'\widehat{G}'$, and $\xi'$ is the identity embedding of $\mathcal{G}'$ and ${}^LG$. For each $G' \in \mathcal{E}_{M'}(G)$, we fix an embedding $M' \subset G'$ for which $\widehat{M}' \subset \widehat{G}'$ is a dual Levi subgroup. We also fix auxiliary data $\widetilde{G}' \to G'$ and $\widetilde{\xi}'\colon \mathcal{G}' \to {}^L\widetilde{G}'$ for $G'$. These objects restrict to auxiliary data $\widetilde{M}' \to M'$ and $\widetilde{\xi}'_M\colon \mathcal{M}' \to {}^L\widetilde{M}'$ for $M'$, whose central character data $\widetilde{Z}'$ and $\widetilde{\zeta}'$ are the same as those for $G'$. Observe that $G^*$ belongs to $\mathcal{E}_{M'}(G)$ if and only if $M'$ equals $M^*$. We write

$$\mathcal{E}^0_{M'}(G) = \begin{cases} \mathcal{E}_{M'}(G) - \{G^*\}, & \text{if } G \text{ is quasisplit,} \\ \mathcal{E}_{M'}(G), & \text{otherwise.} \end{cases}$$

For any $G' \in \mathcal{E}_{M'}(G)$, we also define a coefficient

$$\iota_{M'}(G, G') = |Z(\widehat{M}')^\Gamma / Z(\widehat{M})^\Gamma| |Z(\widehat{G}')^\Gamma / Z(\widehat{G})^\Gamma|^{-1}.$$

Suppose that $\delta'$ belongs to $S\mathcal{D}\big((\widetilde{M}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$. We assume inductively that for every $G' \in \mathcal{E}^0_{M'}(G)$, we have defined a stable linear form $S^{\tilde{G}'}_{\tilde{M}'}(\delta', \cdot)$ on $\mathcal{H}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$. We impose natural conditions of equivariance on $S^{\tilde{G}'}_{\tilde{M}'}(\delta', \cdot)$, which imply that the linear form

$$f \longrightarrow \widehat{S}^{\tilde{G}'}_{\tilde{M}'}(\delta', f'), \qquad f \in \mathcal{H}(G, V, \zeta),$$

on $\mathcal{H}(G,V,\zeta)$ depends only on the image of $\delta'$ in the space $\mathcal{D}^{\mathcal{E}}(M'_V, M^Z_V, \zeta_V)$. In particular, the last linear form is independent of the choice of auxiliary data $\widetilde{G}'$ and $\widetilde{\xi}'$. If $G$ is not quasisplit, we define an "endoscopic" linear form

$$I^{\mathcal{E}}_M(\delta', f) = \sum_{G' \in \mathcal{E}_{M'}(G)} \iota_{M'}(G,G') \widehat{S}^{\tilde{G}'}_{\tilde{M}'}(\delta', f'). \tag{29.12}$$

In the case that $G$ is quasisplit, we define a linear form

$$S^G_M(M', \delta', f) = I_M(\delta, f) - \sum_{G' \in \mathcal{E}^0_{M'}(G)} \iota_{M'}(G,G') \widehat{S}^{\tilde{G}'}_{\tilde{M}'}(\delta', f'), \tag{29.13}$$

where $\delta$ is the image of $\delta'$ in $\mathcal{D}^{\mathcal{E}}(M^Z_V, \zeta_V)$. In this case, we also define the endoscopic linear form by the trivial relation

$$I^{\mathcal{E}}_M(\delta', f) = I_M(\delta, f). \tag{29.14}$$

These definitions represent the first stage of an extensive generalization of the constructions of §25. To see this more clearly, we need to replace the argument $\delta'$ in $I^{\mathcal{E}}_M(\delta', f)$ by an element $\gamma \in \mathcal{D}(M^Z_V, \zeta_V)$. It turns out that there is a canonical bilinear form $I^{\mathcal{E}}_M(\gamma, f)$ in $\gamma$ and $f$ such that

$$I^{\mathcal{E}}_M(\delta', f) = \sum_{\gamma \in \Gamma(M^Z_V, \zeta)} \Delta_M(\delta', \gamma) I^{\mathcal{E}}_M(\gamma, f), \tag{29.15}$$

for any $(M', \delta')$. Since $M'$ was chosen to be an endoscopic datum over $F$, $I^{\mathcal{E}}_M(\gamma, f)$ is not uniquely determined by (29.15). However, the definitions (29.13) and (29.14) apply more generally if $M'$ is replaced by an endoscopic datum $M'_V$ over $F_V$. (See [**A25**, §5].) One shows directly that the resulting linear form

$$I^{\mathcal{E}}_M(\delta, f) = I^{\mathcal{E}}_M(\delta', f)$$

depends only on the image $\delta$ of $\delta'$ in $\mathcal{D}^{\mathcal{E}}(M^Z_V, \zeta_V)$. The distribution $I^{\mathcal{E}}_M(\gamma, f)$ is then defined by inversion from the corresponding extension of (29.15). (See [**A31**].)

To complete the inductive definition, one still has to prove something in the special case that $G$ is quasisplit and $M' = M^*$. Then $\delta' = \delta^*$ belongs to $S\mathcal{D}\big((M^*_V)^{Z^*}, \zeta^*_V\big)$, and the image $\delta$ of $\delta'$ in $\mathcal{D}^{\mathcal{E}}(M^Z_V, \zeta_V)$ lies in the subspace $S\mathcal{D}(M^Z_V, \zeta_V)$ of stable distributions. The problem in this case is to show that the linear form

$$S^G_M(\delta, f) = S^G_M(M^*, \delta^*, f) \tag{29.16}$$

is stable. Only then would we have a linear form

$$\widehat{S}^{G^*}_{M^*}(\delta^*, f^*) = S^G_M(\delta, f)$$

on $S\mathcal{I}\big((G^*_V)^{Z^*}, \zeta^*_V\big)$ that is the analogue for $(G^*, M^*)$ of the terms $\widehat{S}^{\tilde{G}'}_{\tilde{M}'}(\delta', f')$ in (29.12) and (29.13). This property is deep, and is a critical part of the stabilization of the general trace formula. In the case that $G$ is quasi-split but $M' \neq M^*$, there is a second question which is as deep as the first. The problem in this case is to show that $S^G_M(M', \delta', f)$ vanishes for any $\delta'$ and $f$.

The analogue for unramified valuations $v \notin V_{\mathrm{ram}}(G)$ of this second problem is of special interest. It represents the generalization of the fundamental lemma to weighted oribital integrals. To state it, we write

$$r^{G_v}_{M_v}(k_v) = J_{M_v}(k_v, u_v), \qquad k_v \in \Gamma_{G\text{-reg}}(M_v),$$

where $u_k$ is the characteristic function of $K_v$ in $G_v(F_v)$, and $M_v$ is a Levi subgroup of $G_v$. Since $v$ is nonarchimedean, the associated component $G_v$ of $G$ is a connected reductive group over $F_v$. In this context, we may as well take $Z_v = 1$, since for any endoscopic datum $G'_v$ over $F_v$, there is a canonical class of $L$-embeddings of ${}^L G'_v$ in ${}^L G_v$ [**Hal1**, §6]. If $M'_v$ is an unramified elliptic endoscopic datum for $M_v$, and $\ell'_v \in \Delta_{G\text{-reg}}(M'_v)$, we write

$$r^{G_v}_{M_v}(\ell'_v) = \sum_{k_v} \Delta_{M_v}(\ell'_v, k_v) r^{G_v}_{M_v}(k_v).$$

We can also obviously write $\mathcal{E}_{M'_v}(G_v)$ and $\iota_{M_v}(G_v, G'_v)$ for the local analogues of the global objects defined earlier.

**Conjecture.** (Generalized fundamental lemma). *For any $M'_v$ and $\ell'_v$, there is an identity*

$$r^{G_v}_{M_v}(\ell'_v) = \sum_{G'_v \in \mathcal{E}_{M'_v}(G_v)} \iota_{M'_v}(G_v, G'_v) s^{G'_v}_{M'_v}(\ell'_v), \tag{29.17}$$

*for functions $s^{G'_v}_{M'_v}(\ell'_v)$ that depend only on $G'_v$, $M'_v$ and $\ell'_v$.*

If $M'_v = M^*_v$ and $\ell'_v = \ell^*_v$, $G^*_v$ belongs to $\mathcal{E}_{M'_v}(G_v)$, and (29.17) represents an inductive definition of $s^{G_v}_{M_v}(\ell^*_v)$. If $M'_v \neq M^*_v$, $G^*_v$ does not belong to $\mathcal{E}_{M'_v}(G_v)$, and (29.17) becomes an identity to be proved. The reader can check that when $M_v = G_v$, the identity reduces to the standard fundamental lemma, which we described near the end of §27. We assume from now on that this conjecture holds for $G$, at least at almost all valuations $v \notin S_{\mathrm{ram}}(G)$, as well as for any other groups that might be required for induction arguments. Since this includes the usual fundamental lemma, it also encompasses our assumption that the Langlands-Shelstad transfer conjecture is valid [**Wa2**].

We can now state the first of four theorems, which together comprise the stabilization of the invariant trace formula. They are all dependent on our assumption that the generalized fundamental lemma holds.

THEOREM 29.3. (a) *If $G$ is arbitrary,*

$$I^{\mathcal{E}}_M(\gamma, f) = I_M(\gamma, f), \qquad \gamma \in \mathcal{D}(M^Z_V, \zeta_V),\ f \in \mathcal{H}(G, V, \zeta).$$

(b) *Suppose that $G$ is quasisplit, and that $\delta'$ belongs to $S\mathcal{D}\big((\widetilde{M}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, for some $M' \in \mathcal{E}_{\mathrm{ell}}(M, V)$. Then the linear form*

$$f \longrightarrow S^G_M(M', \delta', f), \qquad f \in \mathcal{H}(G, V, \zeta),$$

*vanishes unless $M' = M^*$, in which case it is stable.*

The linear forms $I^{\mathcal{E}}_M(\gamma, f)$ and $S^G_M(\delta, f)$ ultimately become terms in endoscopic and stable analogues of the geometric side of (29.7). These objects are to be regarded as the local components of the expansions. The global components are endoscopic and stable analogues of the coefficients $a^G(\gamma)$ in (29.7). As before, the new coefficients really belong to a completion of the appropriate space of distributions. However, we again identify them with elements in a dual space by choosing bases of the relevant spaces of distributions. We fix a basis $\Delta\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$ of $S\mathcal{D}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$ for any $F_V$-endoscopic datum $G'_V$, with auxiliary data $\widetilde{G}'_V$ and

$\widetilde{\xi}'_V$. We also fix a basis $\Delta^{\mathcal{E}}(G_V^Z, \zeta_V)$ of the space $\mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$. Among various conditions, we require that the subset

$$\Delta(G_V^Z, \zeta_V) = \Delta^{\mathcal{E}}(G_V^Z, \zeta_V) \cap S\mathcal{D}(G_V^Z, \zeta_V)$$

of $\Delta^{\mathcal{E}}(G_V^Z, \zeta_V)$ be a basis of $S\mathcal{D}(G_V^Z, \zeta_V)$, and in the case that $G$ is quasisplit, that $\Delta(G_V^Z, \zeta_V)$ be the isomorphic image of the basis $\Delta\big((G_V^*)^{Z^*}, \zeta_V^*\big)$.

We assume inductively that for every $G'$ in the set

$$\mathcal{E}^0_{\mathrm{ell}}(G, V) = \begin{cases} \mathcal{E}_{\mathrm{ell}}(G, V) - \{G^*\}, & \text{if } G \text{ is quasisplit,} \\ \mathcal{E}_{\mathrm{ell}}(G, V), & \text{otherwise,} \end{cases}$$

we have defined a function $b^{\tilde{G}'}(\delta')$ on $\Delta\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$. If $G$ is not quasisplit, we can then define the "endoscopic" coefficient

$$a^{G,\mathcal{E}}(\gamma) = \sum_{G' \in \mathcal{E}_{\mathrm{ell}}(G,V)} \sum_{\delta'} \iota(G, G') b^{\tilde{G}'}(\delta') \Delta_G(\delta', \gamma), \tag{29.18}$$

as a function of $\gamma \in \Gamma(G_V^Z, \zeta_V)$. In the case that $G$ is quasisplit, we define a "stable" coefficient function $b^G(\delta)$ of $\delta \in \Delta^{\mathcal{E}}(G_V^Z, \zeta_V)$ by requiring that

$$\sum_{\delta} b^G(\delta) \Delta_G(\delta, \gamma) = a^G(\gamma) - \sum_{G' \in \mathcal{E}^0_{\mathrm{ell}}(G,V)} \sum_{\delta'} \iota(G, G') b^{\tilde{G}'}(\delta') \Delta_G(\delta', \gamma), \tag{29.19}$$

for any $\gamma \in \Gamma(G_V^Z, \zeta_V)$. In this case, we also define the endoscopic coefficient by the trivial relation

$$a^{G,\mathcal{E}}(\gamma) = a^G(\gamma).$$

In both (29.18) and (29.19), the numbers $\iota(G, G')$ are Langlands' original global coefficients from (27.3), while $\delta'$ and $\delta$ are summed over $\Delta\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$ and $\Delta^{\mathcal{E}}(G_V^Z, \zeta_V)$ respectively. To complete the inductive definition, we set

$$b^{G^*}(\delta^*) = b^G(\delta), \qquad \delta^* \in \Delta\big((G_V^*)^{Z^*}, \zeta_V^*\big),$$

when $G$ is quasisplit and $\delta$ is the preimage of $\delta^*$ in the subset $\Delta(G_V^Z, \zeta_V)$ of $\Delta^{\mathcal{E}}(G_V^Z, \zeta_V)$.

THEOREM 29.4. (a) *If $G$ is arbitrary,*

$$a^{G,\mathcal{E}}(\gamma) = a^G(\gamma), \qquad \gamma \in \Gamma(G_V^Z, \zeta_V).$$

(b) *If $G$ is quasisplit, $b^G(\delta)$ vanishes for any $\delta$ in the complement of $\Delta(G_V^Z, \zeta_V)$ in $\Delta^{\mathcal{E}}(G_V^Z, \zeta_V)$.*

We have completed our description of the geometric ingredients that go into the stabilization of the trace formula. The spectral ingredients are entirely parallel. In place of the spaces of distributions $\mathcal{D}(G_V^Z, \zeta_V)$, $S\mathcal{D}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, $\mathcal{D}^{\mathcal{E}}(G'_V, G_V^Z, \zeta_V)$, and $\mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$, we have spectral analogues $\mathcal{F}(G_V^Z, \zeta_V)$, $S\mathcal{F}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, $\mathcal{F}^{\mathcal{E}}(G'_V, G_V^Z, \zeta_V)$, and $\mathcal{F}^{\mathcal{E}}(G_V^Z, \zeta_V)$. The subspace $S\mathcal{D}(G_V^Z, \zeta_V)$ of $\mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$ is replaced by a corresponding subspace $S\mathcal{F}(G_V^Z, \zeta_V)$ of $\mathcal{F}^{\mathcal{E}}(G_V^Z, \zeta_V)$. In place of the prescribed basis $\Gamma(G_V^Z, \zeta_V)$ of $\mathcal{D}(G_V^Z, \zeta_V)$, we have the basis

$$\Pi(G_V^Z, \zeta_V) = \coprod_{t \geq 0} \Pi_t(G_V^Z, \zeta_V)$$

of $\mathcal{F}(G_V^Z, \zeta_V)$ consisting of irreducible characters. If $\phi'$ belongs to $S\mathcal{F}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, the distribution $f \to f'(\phi')$ belongs to $\mathcal{F}(G_V^Z, \zeta_V)$. It therefore has an expansion

$$f'(\phi') = \sum_{\pi \in \Pi(G_V^Z, \zeta_V)} \Delta(\phi', \pi) f_G(\pi)$$

that is parallel to (29.15). The coefficients

$$\Delta(\phi, \pi) = \Delta(\phi', \pi), \qquad \pi \in \Pi(G_V^Z, \zeta_V),$$

are products over $v$ of local coefficients in (28.4) (or rather, linear extensions in $\phi'_v$ of such coefficients), and depend only on the image $\phi$ of $\phi'$ in $\mathcal{F}^{\mathcal{E}}(G_V^Z, \zeta_V)$.

The definitions (29.13)–(29.16) have obvious spectral variants. They provide linear forms $I_M(\phi, f)$, $I_M^{\mathcal{E}}(\phi', f)$, $S_M^G(M', \phi', f)$, $I_M^{\mathcal{E}}(\pi, f)$, and $S_M(\phi, f)$ in $f \in \mathcal{H}(G, V, \zeta)$, which also depend linearly on the distributions $\phi$, $\phi'$ and $\pi$.

THEOREM 29.5. (a) *If $G$ is arbitrary,*

$$I_M^{\mathcal{E}}(\pi, f) = I_M(\pi, f), \qquad \pi \in \mathcal{F}(M_V^Z, \zeta_V),\ f \in \mathcal{H}(G, V, \zeta).$$

(b) *Suppose that $G$ is quasisplit, and that $\phi'$ belongs to $S\mathcal{F}\big((\widetilde{M}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, for some $M' \in \mathcal{E}_{\mathrm{ell}}(M, V)$. Then the linear form*

$$f \longrightarrow S_M^G(M', \phi', f), \qquad f \in \mathcal{H}(G, V, \zeta),$$

*vanishes unless $M' = M^*$, in which case it is stable.*

The linear forms $I_M^{\mathcal{E}}(\pi, f)$ and $S_M^G(\phi, f)$ ultimately become local terms in endoscopic and stable analogues of the spectral side of (29.7). The global terms are endoscopic and stable analogues of the coefficients $a^G(\pi)$ in (29.7). We fix a basis $\Phi\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$ of the space $S\mathcal{F}\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, for each $G'_V$, $\widetilde{G}'_V$ and $\widetilde{\xi}'_V$, which we can form from local bases $\Phi(\widetilde{G}'_v, \widetilde{\zeta}'_v)$. If $v$ is nonarchimedean, we take $\Phi(\widetilde{G}'_v, \widetilde{\zeta}'_v)$ to be the abstract basis discussed in §28. If $v$ is archimedean, we can identify $\Phi(\widetilde{G}'_v, \widetilde{\zeta}'_v)$ with the relevant set of archimedean Langlands parameters $\phi_v$, thanks to the work of Shelstad. Since any such $\phi_v$ has an archimedean infinitesimal character, there is a decomposition

$$\Phi\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big) = \coprod_{t \geq 0} \Phi_t\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big).$$

We also fix a basis $\Phi^{\mathcal{E}}(G_V^Z, \zeta_V)$ of the space $\mathcal{D}^{\mathcal{E}}(G_V^Z, \zeta_V)$, which can in fact be taken to be a set of equivalence classes in the union of the various bases $\Phi\big((\widetilde{G}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$. Among other things, this implies that the subset

$$\Phi(G_V^Z, \zeta_V) = \Phi^{\mathcal{E}}(G_V^Z, \zeta_V) \cap S\mathcal{F}(G_V^Z, \zeta_V)$$

of $\Phi^{\mathcal{E}}(G_V^Z, \zeta_V)$ is a basis of $S\mathcal{F}(G_V^Z, \zeta_V)$, and in the case that $G$ is quasisplit, is the isomorphic image of the basis $\Phi\big((G_V^*)^{Z^*}, \zeta_V^*\big)$.

Having fixed bases, we can apply the obvious spectral variants of the definitions (29.18) and (29.19). We thereby obtain functions $a^{G,\mathcal{E}}(\pi)$ and $b^G(\phi)$ of $\pi \in \Pi(G_V^Z, \zeta_V)$ and $\phi \in \Phi^{\mathcal{E}}(G_V^Z, \zeta_V)$ respectively.

THEOREM 29.6. (a) *If $G$ is arbitrary,*

$$a^{G,\mathcal{E}}(\pi) = a^G(\pi), \qquad \pi \in \Pi(G_V^Z, \zeta_V).$$

(b) *If $G$ is quasisplit, $b^G(\phi)$ vanishes for any $\phi$ in the complement of $\Phi(G_V^Z, \zeta_V)$ in $\Phi^{\mathcal{E}}(G_V^Z, \zeta_V)$.*

Theorems 29.3 and 29.4 are general analogues of Theorem 25.5 for inner twistings of $GL(n)$. The extra assertions (b) of theorems were not required earlier, since the question of stability is trivial for $GL(n)$. Similarly, Theorems 29.5 and 29.6 are general analogues of Theorem 25.6. Taken together, Theorems 29.3–29.6 amount to a stabilization of the general trace formula. This will become clearer after we have stated the general analogues of Lemmas 25.3 and 25.4.

The four theorems are proved together. As in the special case in §25, the argument is by double induction on $\dim(G/Z)$ and $\dim(A_M)$. The first stage of the proof is to obtain endoscopic and stable analogues of the expansions on each side of (29.7). For this, one needs only the induction assumption that the global assertions (b) of Theorems 29.4 and 29.6 be valid if $(G, \zeta)$ is replaced by $(\widetilde{G}', \widetilde{\zeta}')$, for any $G' \in \mathcal{E}^0_{\mathrm{ell}}(G, V)$.

Let $I$ be the invariant linear form on $\mathcal{H}(G, V, \zeta)$ defined by either of the two sides of (29.7). If $G$ is not quasisplit, we define an "endoscopic" linear form inductively by setting

$$I^{\mathcal{E}}(f) = \sum_{G' \in \mathcal{E}_{\mathrm{ell}}(G,V)} \iota(G, G')\widehat{S}'(f'), \tag{29.20}$$

for stable linear forms $\widehat{S}' = \widehat{S}^{\tilde{G}'}$ on $S\mathcal{I}(\widetilde{G}', V, \widetilde{\zeta}')$. In the case that $G$ is quasisplit, we define a linear form

$$S^G(f) = I(f) - \sum_{G' \in \mathcal{E}^0_{\mathrm{ell}}(G,V)} \iota(G, G')\widehat{S}'(f'). \tag{29.21}$$

We also define the endoscopic linear form by the trivial relation

$$I^{\mathcal{E}}(f) = I(f). \tag{29.22}$$

In the case $G$ is quasisplit, we need to show that the linear form $S^G$ on $\mathcal{I}(G, V, \zeta)$ is stable. Only then will we have a linear form

$$\widehat{S}^{G^*}(f^*) = S^G(f)$$

on $S\mathcal{I}(G^*, V, \zeta^*)$ that is the analogue for $G^*$ of the summands in (29.20) and (29.21) needed to complete the inductive definition. We would also like to show that $I^{\mathcal{E}}(f) = I(f)$. These properties are obviously related to the assertions of the four theorems.

The reader will recognize in the definitions (29.20)–(29.22), taken with the assertions that $S^G(f)$ is stable and $I^{\mathcal{E}}(f) = I(f)$, an analogue of Langlands' stabilization (27.3) of the regular elliptic terms. This construction is in fact a model for the stabilization of any part of the trace formula. For example, let

$$I_{\mathrm{orb}}(f) = \sum_{\gamma \in \Gamma(G,V,\zeta)} a^G(\gamma) f_G(\gamma) \tag{29.23}$$

be the component with $M = G$ in the geometric expansion in (29.7). This sum includes the regular elliptic terms, as well as orbital integrals over more general conjugacy classes. Its complement $I(f) - I_{\mathrm{orb}}(f)$ in $I(f)$, being a sum over $M$ in the complement $\mathcal{L}^0$ of $\{G\}$ in $\mathcal{L}$, can be regarded as the "parabolic" part of the geometric expansion. We define linear forms $I^{\mathcal{E}}_{\mathrm{orb}}(f)$ and $S^G_{\mathrm{orb}}(f)$ on $\mathcal{H}(G, V, \zeta)$ by the obvious analogues of (29.20)–(29.22).

Proposition 29.7. (a) *If $G$ is arbitrary,*

$$I^{\mathcal{E}}(f) - I^{\mathcal{E}}_{\mathrm{orb}}(f) = \sum_{M\in\mathcal{L}^0} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma^{\mathcal{E}}(M,V,\zeta)} a^{M,\mathcal{E}}(\gamma) I^{\mathcal{E}}_M(\gamma,f),$$

*where $\Gamma^{\mathcal{E}}(M,V,\zeta)$ is a natural discrete subset of $\Gamma(M_V^Z,\zeta_V)$ that contains the support of $a^{M,\mathcal{E}}(\gamma)$.*

(b) *If $G$ is quasisplit,*

$$\begin{aligned} &S^G(f) - S^G_{\mathrm{orb}}(f) \\ &= \sum_{M\in\mathcal{L}^0} |W_0^M||W_0^G|^{-1} \sum_{M'\in\mathcal{E}_{\mathrm{ell}}(M,V)} \iota(M,M') \sum_{\delta'\in\Delta(\tilde{M}',V,\tilde{\zeta}')} b^{\tilde{M}'}(\delta') S^G_M(M',\delta',f), \end{aligned}$$

*where $\Delta(\widetilde{M}',V,\widetilde{\zeta}')$ is a natural discrete subset of $\Delta\big((\widetilde{M}'_V)^{\tilde{Z}'},\widetilde{\zeta}'_V\big)$ that contains the support of $b^{\tilde{M}'}(\delta')$.*

See [**A27**, Theorem 10.1]. □

Let $I_t(f)$ be the summand of $t$ on the spectral side of (29.7). We attach linear forms $I^{\mathcal{E}}_t(f)$ and $S^G_t(f)$ to $I_t(f)$ by the analogues of (29.20)–(29.22). The decomposition in (29.7) of $I(f)$ as a sum over $t \geq 0$ of $I_t(f)$ leads to corresponding decompositions

$$I^{\mathcal{E}}(f) = \sum_{t\geq 0} I^{\mathcal{E}}_t(f) \tag{29.24(a)}$$

and

$$S^G(f) = \sum_{t\geq 0} S^G_t(f) \tag{29.24(b)}$$

of $I^{\mathcal{E}}(f)$ and $S^G(f)$. Each of these sums satisfies the analogue of the weak multiplier estimate (23.13), and hence converges absolutely. (See [**A27**, Proposition 10.5].) For any $t$, we write

$$I_{t,\mathrm{unit}}(f) = \int_{\Pi_t(G,V,\zeta)} a^G(\pi) f_G(\pi) d\pi \tag{29.25}$$

for the component with $M = G$ for the spectral expansion of $I_t(f)$ in (29.7). We then define corresponding linear forms $I^{\mathcal{E}}_{t,\mathrm{unit}}(f)$ and $S^G_{t,\mathrm{unit}}(f)$ on $\mathcal{H}(G,V,\zeta)$, again by the obvious analogues of (29.20)–(29.22).

Proposition 29.8. (a) *If $G$ is arbitrary,*

$$I^{\mathcal{E}}_t(f) - I^{\mathcal{E}}_{t,\mathrm{unit}}(f) = \sum_{M\in\mathcal{L}^0} |W_0^M||W_0^G|^{-1} \int_{\Pi^{\mathcal{E}}_t(M,V,\zeta)} a^{M,\mathcal{E}}(\pi) I^{\mathcal{E}}_M(\pi,f) d\pi,$$

*where $\Pi^{\mathcal{E}}_t(M,V,\zeta)$ is a subset of $\Pi_t(M_V^Z,\zeta_V)$, equipped with a natural measure $d\pi$, that contains the support of $a^{M,\mathcal{E}}(\pi)$.*

(b) *If $G$ is quasisplit,*

$$\begin{aligned} &S^G_t(f) - S^G_{t,\mathrm{unit}}(f) \\ &= \sum_{M\in\mathcal{L}^0} |W_0^M||W_0^G|^{-1} \sum_{M'\in\mathcal{E}_{\mathrm{ell}}(M,V)} \iota(M,M') \int_{\Phi_{t'}(\tilde{M}',V,\tilde{\zeta}')} b^{\tilde{M}'}(\phi') S^G_M(M',\phi',f) d\phi', \end{aligned}$$

*where* $t'$ *is a translate of* $t$, *and* $\Phi_{t'}(\widetilde{M}', V, \widetilde{\zeta}')$ *is a subset of* $\Phi_{t'}\big((\widetilde{M}'_V)^{\tilde{Z}'}, \widetilde{\zeta}'_V\big)$, *equipped with a natural measure* $\mathrm{d}\phi'$, *which contains the support of* $b^{\tilde{M}'}(\phi')$.

See [**A27**, Theorem 10.6]. □

In contrast to the special cases of Lemmas 25.3 and 25.4, we have excluded the terms with $M = G$ from the expansions of Propositions 29.7 and 29.8. This was only to keep the notation slightly simpler in the assertions (b). It is a consequence of the definitions that

$$I^{\mathcal{E}}_{\mathrm{orb}}(f) = \sum_{\gamma \in \Gamma^{\mathcal{E}}(G,V,\zeta)} a^{G,\mathcal{E}}(\gamma) f_G(\gamma) \tag{29.26(a)}$$

and

$$S^{G}_{\mathrm{orb}}(f) = \sum_{\delta \in \Delta^{\mathcal{E}}(G,V,\zeta)} b^{G}(\delta) f^{\mathcal{E}}_G(\delta), \tag{29.26(b)}$$

where $\Delta^{\mathcal{E}}(G, V, \zeta)$ is a certain discrete subset of $\Delta^{\mathcal{E}}(G^Z_V, \zeta_V)$ that contains the support of $b^G$. Similarly, we have

$$I^{\mathcal{E}}_{t,\mathrm{unit}}(f) = \int_{\Pi^{\mathcal{E}}_t(G,V,\zeta)} a^{G,\mathcal{E}}(\pi) f_G(\pi) \mathrm{d}\pi \tag{29.27(a)}$$

and

$$S^{G}_{t,\mathrm{unit}}(f) = \int_{\Phi^{\mathcal{E}}_t(G,V,\zeta)} b^{G}(\phi) f^{\mathcal{E}}_G(\phi) \mathrm{d}\phi, \tag{29.27(b)}$$

where $\Phi^{\mathcal{E}}_t(G, V, \zeta)$ is a subset of $\Phi^{\mathcal{E}}_t(G^Z_V, \zeta_V)$, equipped with a natural measure $\mathrm{d}\phi$, that contains the support of $b^G(\phi)$. (See [**A27**, Lemmas 7.2 and 7.3].) We can obviously combine (29.26(a)) and (29.27(a)) with the expansions (a) of Propositions 29.7 and 29.8. This provides expressions for $I^{\mathcal{E}}(f)$ and $I^{\mathcal{E}}_t(f)$ that are more clearly generalizations of those of Lemmas 25.3 and 25.4. On the other hand, the sums in (29.26(b)) and (29.27(b)) are not of the same form as those in the expansions (b) of Propositions 29.7 and 29.8. Their substitution into these expansions leads to expressions for $S^G(f)$ and $S^G_t(f)$ that, without the general assertions (b) of the four theorems, are more ungainly.

We shall say only a few words about the proof of the four theorems. If $G$ is not quasisplit, one works with the identity obtained from (29.24(a)), (29.26(a)), (29.27(a)), and Propositions 29.7(a) and 29.8(a). The problem is to compare the terms in this identity with those of the invariant trace formula (29.7). If $G$ is quasisplit, one works with the identity obtained from (29.24(b)), (29.26(b)), (29.27(b)), and Propositions 29.7(b) and 29.8(b). The problem here is to show that if $f^G = 0$, the appropriate terms in the identity vanish. The arguments are long and complicated, but they do follow the basic model established in §25. In particular, they frequently move forward under their own momentum.

There is one point we should mention explicitly. The geometric coefficients $a^G(\gamma)$ are compound objects, defined (29.3) in terms of the original coefficients $a^M_{\mathrm{ell}}(\gamma_M \times k)$. The identities stated in Theorem 29.4 have analogues that apply to endoscopic and stable forms of the coefficients $a^G_{\mathrm{ell}}(\gamma \times k)$. The role of the generalized fundamental lemma is to reduce Theorem 29.4 to these basic identities [**A27**, Proposition 10.3]. (The case $M = G$ of the generalized fundamental lemma, namely the ordinary fundamental lemma, carries the more obvious burden of establishing the existence of the mappings $f \to f'$.) One has then to reduce these basic identities further to the special case of classes in $G(F_S)$ that are purely unipotent. This turns out to be a major undertaking [**A26**], which depends heavily on Langlands-Shelstad descent for transfer factors [**LS2**]. The reduction to unipotent classes can be regarded as an extension of the stabilization of the semisimple elliptic terms by Langlands [**Lan10**] and Kottwitz [**Ko5**].

The spectral coefficients $a^G(\pi)$ are also compound objects. They are defined (29.6) in terms of the original spectral coefficients $a^M_{\mathrm{disc}}(\pi_M \times c)$. The identities stated in Theorem 29.6 have analogues for endoscopic and stable forms of the coefficients $a^G_{\mathrm{disc}}(\pi \times c)$. It is interesting to note that the generalized fundamental lemma has a spectral variant [**A27**, Proposition 8.3], albeit one which is much less deep, and which has a straightforward proof. (For example, the case $M = G$ of this spectral result is entirely vacuous. The cases with $M \neq G$ reflect relatively superficial aspects of the deeper geometric conjecture.) The role of the spectral result is to reduce Theorem 29.6 to the identities for endoscopic and stable forms of the coefficients $a^G_{\mathrm{disc}}(\gamma \times k)$ [**A27**, Proposition 10.7].

We have touched on a couple of aspects of the first half of the argument. The second half of the proof is contained in [**A29**]. It is based on a comparison of the expansions in Propositions 29.7 and 29.8 with those in (29.7). Among the many reductions on the geometric sides, one establishes the required cancellation of almost all of the terms in $I_{\mathrm{orb}}(f)$, $I^{\mathcal{E}}_{\mathrm{orb}}(f)$, and $S^G_{\mathrm{orb}}(f)$ by appealing to the reductions of Theorem 29.4 described above. Those that remain correspond to unipotent elements. They can be separated from the complementary terms in the expansions by an approximation argument. Among the spectral reductions, one sees that many of the terms in $I_{t,\mathrm{unit}}(f)$, $I^{\mathcal{E}}_{t,\mathrm{unit}}(f)$, and $S^G_{t,\mathrm{unit}}(f)$ also cancel, thanks to the reduction of Theorem 29.6 we have mentioned. Those that remain occur discretely. They can be separated from the complementary terms in the expansions by the appropriate forms of the weak multiplier estimate (23.13).

These sparse comments convey very little sense of the scope of the argument. It will suffice for us to reiterate that much of the collective proof of the four theorems is in attempting to generalize arguments described in the special case of §25. □

COROLLARY 29.9. (a) (Endoscopic trace formula). *The identity*

$$\begin{aligned}&\sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\gamma\in\Gamma^{\mathcal{E}}(M,V,\zeta)} a^{M,\mathcal{E}}(\gamma) I^{\mathcal{E}}_M(\gamma, f)\\ &= \sum_{t\geq 0}\sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Pi^{\mathcal{E}}_t(M,V,\zeta)} a^{M,\mathcal{E}}(\pi) I^{\mathcal{E}}_M(\pi, f)\end{aligned} \tag{29.28(a)}$$

*holds for any $f \in \mathcal{H}(G, V, \zeta)$. Each term in the identity is equal to its corresponding analogue in the invariant trace formula (29.7).*

(b) (Stable trace formula). *If $G$ is quasisplit, the identity*

$$\begin{aligned}&\sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \sum_{\delta\in\Delta(M,V,\zeta)} b^M(\delta) S_M(\delta, f)\\ &= \sum_{t\geq 0}\sum_{M\in\mathcal{L}} |W_0^M||W_0^G|^{-1} \int_{\Phi_t(M,V,\zeta)} b^M(\phi) S_M(\phi, f) \mathrm{d}\phi\end{aligned} \tag{29.28(b)}$$

*holds for any $f \in \mathcal{H}(G, V, \zeta)$. The terms in the identity are all stable in $f$.*

The identity (29.28(a)) follows immediately from Propositions 29.7(a) and 29.8(a) and expansions (29.26(a)) and (29.27(a)), as we have already noted. Assertions (a) of the four theorems give the term by term identification of this identity with the invariant trace formula.

To establish (29.28(b)), we combine the expansions of Propositions 29.7(b) and 29.8(b) with (29.26(b)) and (29.27(b)). This yields a rather complicated formula. However, assertions (b) of the four theorems imply immediately that the formula collapses to the required identity (29.28(b)). Supplementary assertions in Theorems 29.3(b) and 29.5(b) tell us that the linear forms $S_M(\delta, f)$ and $S_M(\phi, f)$ in (29.28(b)) are stable in $f$. $\square$

The endoscopic trace formula (29.28(a)) is a priori quite different from the original formula (29.7). In case $G$ is not quasisplit, it is defined as a linear combination of stable trace formulas for endoscopic groups $G'$. Our conclusion that it is in fact equal to the original formula amounts to a stabilization of the trace formula.

We recall that $G = \coprod G_\alpha$ is a $K$-group over $F$. However, if $f$ is supported on a component $G_\alpha(F_V)$, the sums in (29.28(a)) can be taken over geometric and stable objects attached to $G_\alpha$. Moreover, if $G$ is quasisplit, the stable distributions on $G_V$ are in bijective correspondence with those on $G_V^*$. It follows that the assertions of Corollary 29.9 hold as stated if $G$ is an ordinary connected group over $F$.

There is one final corollary. To state it, we return to the setting of earlier sections. We take $G$ to be a connected reductive group over $F$, and $f$ to be a function in the adelic Hecke algebra $\mathcal{H}(G,\zeta) = \mathcal{H}\big(G(\mathbb{A})^Z, \zeta\big)$. The $t$-discrete part $I_{t,\mathrm{disc}}(f)$ of the trace formula (21.19) represents its spectral core. It is the part that is actually used for applications.

Corollary 29.10. *There are stable linear forms*

$$S^G_{t,\mathrm{disc}}(f), \qquad\qquad f \in \mathcal{H}(G,V),\ t \geq 0,$$

*defined whenever $G$ is quasisplit, such that*

$$I_{t,\mathrm{disc}}(f) = \sum_{G' \in \mathcal{E}_{\mathrm{ell}}(G)} \iota(G,G') \widehat{S}^{\tilde{G}'}_{t,\mathrm{disc}}(f'), \tag{29.29}$$

*for any $G$, $t$ and $f$.*

We define linear forms $I^{\mathcal{E}}_{t,\mathrm{disc}}$ and $S^G_{t,\mathrm{disc}}$ inductively by analogues of (29.20)–(29.22). Recall that there is an expansion

$$I_{t,\mathrm{disc}}(f) = \sum_{\pi \in \Pi_{t,\mathrm{disc}}(G)} a^G_{\mathrm{disc}}(\pi) f_G(\pi),$$

which serves as the definition of the coefficients $a^G_{\mathrm{disc}}(\pi)$, and is parallel to the definition (29.25) of $I_{t,\mathrm{unit}}(f)$. This leads to corresponding expansions of $I^{\mathcal{E}}_{t,\mathrm{disc}}(f)$ and $S^G_{t,\mathrm{disc}}(f)$, which are parallel to (29.27(a)) and (29.27(b)). We have already noted that the assertions of Theorem 29.6 reduce to corresponding assertions for the coefficients of these latter expansions. Theorem 29.6 therefore implies that $I^{\mathcal{E}}_{t,\mathrm{disc}}(f) = I_{t,\mathrm{disc}}(f)$, and that $S^G_{t,\mathrm{disc}}(f)$ is stable in case $G$ is quasisplit. The identity (29.29) then follows from the definition of $I^{\mathcal{E}}_{t,\mathrm{disc}}(f)$. $\square$

## 30. Representations of classical groups

To give some sense of the power of the stable trace formula, we shall describe a broad application. It concerns the representations of classical groups. We shall describe a classification of automorphic representations of classical groups $G$ in terms of those of general linear groups $GL(N)$. Since it depends on the stable trace formula for $G$, the classification is conditional on the fundamental lemma (both the standard version and its generalization (29.17)) for each of the classical groups in question. It also depends on the stabilization of a twisted trace formula for $GL(N)$. The classification is therefore conditional also on the corresponding twisted fundamental lemma (both standard and generalized) for $GL(N)$, as well as twisted analogues (yet to be established) of the results of §29.

It is possible to work in a more general context. One could take a product of general linear groups, equipped with a pair $\alpha = (\theta, \omega)$, where $\theta$ is an outer automorphism, and $\omega$ is an automorphic character of $GL(1)$. This is the setting adopted by Kottwitz and Shelstad in their construction of twisted transfer factors [**KoS**]. There is much to be learned by working in such generality. However, we shall adopt the more restricted setting in which $\alpha = \theta$ is the standard outer automorphism of $GL(N)$. For reasons on induction, it is important to allow $N$ to vary. The groups $G$ will then range the quasisplit classical groups in the three infinite families $SO(2n+1)$, $Sp(2n)$, and $SO(2n)$. The results have yet to be published. My notes apply only to the special case under discussion, but I will try to write them up in greater generality.

The groups $G$ arise as twisted endoscopic groups. For computational purposes, we represent $\theta$ as the automorphism

$$\theta(x) \longrightarrow {}_t x^{-1} = J\, {}^t x^{-1} J^{-1}, \qquad x \in GL(N),$$

of $GL(N)$, where

$${}_t x = J\, {}^t x J = J\, {}^t x J^{-1}, \qquad J = \begin{pmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & 0 \end{pmatrix},$$

is the "second transpose" of $x$, about the second diagonal. Then $\theta$ stabilizes the standard Borel subgroup of $GL(N)$. (For theoretical purposes [**KoS**], it is sometimes better to work with the automorphism

$$\theta'(x) = J'\, {}^t x^{-1} (J')^{-1}, \qquad J' = \begin{pmatrix} 0 & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ (-1)^{N+1} & & 0 \end{pmatrix},$$

that stabilizes the standard splitting in $GL(N)$ as well.) We form the connected component

$$\widetilde{G} = \widetilde{G}_N = GL(N) \rtimes \theta$$

in the nonconnected semidirect product

$$\widetilde{G}^+ = \widetilde{G}_N^+ = GL(N) \rtimes (\mathbb{Z}/2\mathbb{Z}),$$

whose identity component we denote by $\widetilde{G}^0$. Twisted endoscopic data are like ordinary endoscopic data, except that their dual groups are connected centralizers of semisimple automorphisms within the inner class defined by $\widetilde{G}$, rather than the earlier identity class of inner automorphisms. We have then to consider semisimple elements $s$ in the component $\widehat{\widetilde{G}} = \widehat{\widetilde{G}}{}^0 \rtimes \theta$, acting by conjugation on $\widetilde{G}^0$. It suffices to work here with the Galois form of L-groups. In the present context, a twisted endoscopic datum for $\widetilde{G}$ can be taken to be a quasisplit group $G$, together with an admissible $L$-embedding $\xi$ of $\mathcal{G} = {}^LG$ into the centralizer in ${}^L\widetilde{G}^0 = GL(N,\mathbb{C}) \times \Gamma_F$ of some element $s$. We define $G$ to be *elliptic* if $A_G = \{1\}$, which is to say that the group $Z(\widehat{G})^{\Gamma_F}$ is finite. We then write $\mathcal{E}_{\text{ell}}(\widehat{G})$ for the set of isomorphism classes of elliptic (twisted) endoscopic data for $\widetilde{G}$.

Suppose for example that $N$ is odd, and that $s = \theta$. Then the centralizer of $s$ in $\widehat{\widetilde{G}}^0$ is a group we will denote by $O(N,\mathbb{C})$, even though it is really the orthogonal group with respect to the symmetric bilinear attached to $J$. The element $s$ therefore yields a twisted endoscopic group $G$ for which $\widehat{G}$ is the special orthogonal group $SO(N,\mathbb{C})$. Since $N$ is odd, $G$ is isomorphic to the split group $Sp(N-1)$. The group $O(N,\mathbb{C})$ has a second connected component, represented by the central element $(-I)$ in $GL(N,\mathbb{C})$. This means that there are many admissible ways to embed ${}^LG$ into ${}^L\widetilde{G}^0$. They are parametrized by isomorphisms from $\Gamma_F$ to $\mathbb{Z}/2\mathbb{Z}$, which by class field theory correspond to characters $\eta$ on $F^*\backslash\mathbb{A}^*$ with $\eta^2 = 1$. The set of such $\eta$ parametrizes the subset of $\mathcal{E}_{\text{ell}}(\widetilde{G})$ attached to $s$. This phenomenon illustrates a second point of departure in the twisted case. The different embeddings represent distinct isomorphism classes of twisted endoscopic data, even though the underlying twisted endoscopic groups and associated elements $s$ are all the same.

To describe the full set $\mathcal{E}_{\text{ell}}(\widetilde{G})$, we consider decompositions of $N$ into a sum $N_s + N_o$ of nonnegative integers, with $N_s$ even. We then take the diagonal matrix

$$s = \begin{pmatrix} -I_s & & 0 \\ & I_o & \\ 0 & & I_s \end{pmatrix},$$

where $I_s$ is the identity matrix of rank $(N_{s/2})$, and $I_o$ is the identity matrix of rank $N_o$. The centralizer of $s$ in $\widehat{\widetilde{G}}^0$ is a product

$$Sp(N_s,\mathbb{C}) \times O(N_o,\mathbb{C})$$

of complex classical groups, defined again by bilinear forms supported on the second diagonal. It corresponds to a twisted endoscopic group $G$ with dual group

$$\widehat{G} = Sp(N_s,\mathbb{C}) \times SO(N_o,\mathbb{C}).$$

The group $O(N_o,\mathbb{C})$ has two connected components if $N_o > 0$. We have then also to specifiy an idèle class character $\eta$ with $\eta^2 = 1$. If $N_o$ is odd, the twisted endoscopic group is the split group

$$G = SO(N_s + 1) \times Sp(N_o - 1)$$

over $F$. In this case, $\eta$ serves to specify the embedding of ${}^LG$ into ${}^L\widetilde{G}^0$, as in the special case above. We emphasize again that $\eta$ is an essential part of the associated endoscopic datum. If $N_o$ is even, the nonidentity component of $O(N_o,\mathbb{C})$ acts on the identity component $SO(N_o,\mathbb{C})$ as an outer automorphism. In this case, the twisted endoscopic group is the quasisplit group

$$G = SO(N_s + 1) \times SO(N_o, \eta),$$

where $SO(N_o,\eta)$ is the outer twist of the split group $SO(N_o)$ determined by $\eta$. The character $\eta$ also determines an $L$-embedding of ${}^LG$ into ${}^L\widetilde{G}^0$ in this case. If $N_o = 2$, the group $SO(N_o)$ is abelian. In this case, $\eta$ must be nontrivial in order that corresponding twisted endoscopic datum be elliptic. In all other cases, $\eta$ can be arbitrary. It is a straightforward exercise to check that the twisted endoscopic data obtained from triplets $(N_s, N_o, \eta)$ in this way give a complete set of representatives of $\mathcal{E}_{\mathrm{ell}}(\widetilde{G})$.

It is possible to motivate the discussion above in more elementary terms. One does so by analyzing continuous representations

$$r: \Gamma_F \longrightarrow GL(N, \mathbb{C})$$

that are self-contragredient, in the sense that the representation

$$_tr^{-1}: \sigma \longrightarrow {}_tr(\sigma)^{-1}, \qquad \sigma \in \Gamma_F,$$

is equivalent to $r$. Since $r$ is continuous, it factors through a finite quotient of $\Gamma_F$. The analysis is therefore essentially that of the self-contragredient representations of an abstract finite group. One sees that twisted endoscopic data arise naturally in terms of decompositions of $r$ into symplectic and orthogonal components. (See [**A23**, §3].)

The general results are proved by induction on $N$. We therefore have a particular interest in elements $G \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$ that are *primitive*, in the sense either $N_s$ or $N_o$ equals zero. There are three cases. They correspond to $N = N_s$ even, $N = N_o$ odd, and $N = N_o$ even. The associated twisted endoscopic groups are the split group $G = SO(N+1)$ with dual group $\widehat{G} = Sp(N, \mathbb{C})$, the split group $G = Sp(N-1)$ with dual group $\widehat{G} = SO(N, \mathbb{C})$, and the quasisplit group $G = SO(N, \eta)$ with dual group $\widehat{G} = SO(N, \mathbb{C})$. We write $\mathcal{E}_{\mathrm{prim}}(\widetilde{G})$ for the subset of primitive elements in $\mathcal{E}_{\mathrm{ell}}(\widetilde{G})$.

Suppose that $G \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G})$. Regarding $G$ simply as a reductive group over $F$, we can calculate its (standard) elliptic endoscopic data $G' \in \mathcal{E}_{\mathrm{ell}}(G)$. It suffices to consider diagonal matrices $s' \in \widehat{G}$ with entries $\pm 1$. For example, in the first case that $G = SO(N+1)$ and $\widehat{G} = Sp(N, \mathbb{C})$ (for $N$ even), it is enough to take diagonal elements

$$s' = \begin{pmatrix} -I' & & 0 \\ & I'' & \\ 0 & & -I' \end{pmatrix},$$

where $I'$ is the identity matrix of rank $(N'/2)$, and $I''$ is the identity matrix of rank $N''$. The set $\mathcal{E}_{\mathrm{ell}}(G)$ is parametrized by pairs $(N', N'')$ of nonnegative even integers, with $0 \leq N' \leq N''$ and $N = N' + N''$. The corresponding endoscopic groups are the split groups

$$G' = SO(N'+1) \times SO(N''+1),$$

with dual groups

$$\widehat{G}' = Sp(N', \mathbb{C}) \times Sp(N'', \mathbb{C}) \subset Sp(N, \mathbb{C}) = \widehat{G}.$$

In the second case that $G = Sp(N-1)$ and $\widehat{G} = SO(N, \mathbb{C})$, $\mathcal{E}_{\mathrm{ell}}(G)$ is parametrized by pairs of $(N', N'')$ of nonnegative even integers with $N = N' + (N''+1)$, and idèle class characters $\eta'$ with $(\eta')^2 = 1$. The corresponding endoscopic groups are the quasisplit groups

$$G' = SO(N', \eta') \times Sp(N''),$$

with dual groups

$$\widehat{G}' = SO(N', \mathbb{C}) \times SO(N''+1, \mathbb{C}) \subset \widehat{G} = SO(N, \mathbb{C}).$$

In the third case that $G = SO(N, \eta)$ and $\widehat{G} = SO(N, \mathbb{C})$, $\mathcal{E}_{\mathrm{ell}}(G)$ is parametrized pairs of nonnegative even integers $(N', N'')$ with $0 \leq N' \leq N''$ and $N = N' + N''$, and pairs $(\eta', \eta'')$ of idèle class characters with $(\eta')^2 = (\eta'')^2 = 1$ and $\eta = \eta'\eta''$. The corresponding endoscopic groups are the quasisplit groups

$$G' = SO(N', \eta') \times SO(N'', \eta''),$$

with dual groups

$$\widehat{G}' = SO(N', \mathbb{C}) \times SO(N'', \mathbb{C}) \subset SO(N, \mathbb{C}) = \widehat{G}.$$

In the second and third cases, the character $\eta^*$ has to be nontrivial if the corresponding integer $N^*$ equals 2, and in the case $N' = 0$, $\eta'$ must of course be trivial.

Our goal is to try to classify automorphic representations of a group $G \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G})$ by means of the trace formula. The core of the trace formula for $G$ is the $t$-discrete part

(30.1)
$$I_{t,\mathrm{disc}}(f) = \sum_{\{M\}} |W(M)|^{-1} \sum_{s \in W(M)_{\mathrm{reg}}} |\det(s-1)_{\mathfrak{a}_M}|^{-1} \mathrm{tr}\big(M_P(s,0)\mathcal{I}_{P,t}(0,f)\big),$$

of its spectral side. We recall that $f$ is a test function in $\mathcal{H}(G) = \mathcal{H}\big(G(\mathbb{A})\big)$, while $t$ is a nonnegative number that restricts the automorphic constituents of $\mathcal{I}_{P,t}(0,f)$ by specifying the norm of their archimedean infinitesimal characters. The stabilization described in §29 yields the decomposition

$$I_{t,\mathrm{disc}}(f) = \sum_{G' \in \mathcal{E}_{\mathrm{ell}}(G)} \iota(G, G') \widehat{S}^{G'}_{t,\mathrm{disc}}(f') \tag{30.2}$$

stated in Corollary 29.10. We recall that $S^{G'}_{t,\mathrm{disc}}$ is a stable distribution on $G'(\mathbb{A})$, while $f' = f^{G'}$ is the Langlands-Shelstad transfer of $f$. This is the payoff. It is our remuneration for the work done in stabilizing the other terms in the trace formula. But really, how valuable is it? Since $G$ is quasisplit, $G = G^*$ is an element in $\mathcal{E}(G)$. The stabilization does not provide an independent characterization of the distribution $S^G_{t,\mathrm{disc}}$. In fact, (30.2) can be regarded as an inductive definition of $S^G_{t,\mathrm{disc}}$ in terms of $I_{\mathrm{disc},t}$ and corresponding distributions for groups $G'$ of dimension smaller than $G$. Thus, (30.2) amounts to the assertion that one can modify $I_{t,\mathrm{disc}}(f)$ by adding some correction terms, defined inductively in terms of Langlands-Shelstad transfer, so that it becomes stable. A useful property, no doubt, but not something that in itself could classify the automorphic representations of $G$.

What saves the day is the twisted trace formula for $\widetilde{G}$. Let $\widetilde{f}$ be a test function in the Hecke space $\mathcal{H}(\widetilde{G}) = \mathcal{H}\big(\widetilde{G}(\mathbb{A})\big)$ attached to the component $\widetilde{G} = GL(N) \rtimes \theta$. The twisted trace formula is an identity of linear forms whose spectral side also has a discrete part

(30.3)
$$I_{t,\mathrm{disc}}(\widetilde{f}) = \sum_{\{\widetilde{M}^0\}} |W(\widetilde{M}^0)|^{-1} \sum_{s \in W(\widetilde{M}^0)_{\mathrm{reg}}} |\det(s-1)_{\mathfrak{a}^{\widetilde{G}}_{\widetilde{M}^0}}|^{-1} \mathrm{tr}\big(M_{\widetilde{P}^0}(s,0)\mathcal{I}_{\widetilde{P}^0,t}(0)\big)(\widetilde{f})$$

with the same general structure as (30.1). (The first sum is over the set of $\widetilde{G}^0$-orbits of Levi subgroups $\widetilde{M}^0$, while the second sum is over the regular elements in the relevant twisted Weyl set. The other terms are also twisted forms of their analogues in (30.1), for which the reader can consult [**CLL**] and [**A14**, §4].) We assume that the twisted fundamental lemma (both ordinary and weighted) holds for

$\widetilde{G}$, as well as twisted analogues of the other results described in §29. These include the twisted analogue of Waldspurger's theorem that the fundamental lemma implies transfer. We can therefore suppose that the transfer mapping $\widetilde{f} \to \widetilde{f}^G$, defined for any $G \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$ by the twisted transfer factors of Kottwitz-Shelstad [**KoS**], sends $\mathcal{H}(\widetilde{G})$ to the space $S\mathcal{I}(G)$. The stabilization of the twisted trace formula for $\widetilde{G}$ then yields a decomposition

$$I_{t,\mathrm{disc}}(\widetilde{f}) = \sum_{G \in \mathcal{E}_{\mathrm{ell}}(\tilde{G})} \iota(\widetilde{G}, G)\widehat{S}^G_{t,\mathrm{disc}}(\widetilde{f}^G), \tag{30.4}$$

where $S^G_{t,\mathrm{disc}}$ is the (untwisted) stable distribution on $G(\mathbb{A})$ that appears in (30.2), and $\iota(\widetilde{G}, G)$ is an explicit constant. This gives an a priori relationship among the terms $S^G_{t,\mathrm{disc}}$ defined in the formulas (30.2).

By combining the global identities (30.2) and (30.4), one obtains both local and global results. In the end, the interplay between the two formulas yields a classification of representations of odd orthogonal and symplectic groups, and something close to a classification in the even orthogonal case. We shall say little more about the proofs. We shall instead use the rest of the section to try to give a precise statement of the results.

Since everything ultimately depends on the automorphic spectrum of $GL(N)$, we begin with this group. We need to formulate the results of Moeglin and Waldspurger in a way that can be extended to the classical groups in question.

We shall represent the discrete spectrum of $GL(N)$ by a set of formal objects that are parallel to the global parameters at the end of §26. Let $\Psi_2\big(GL(N)\big)$ be the set of formal tensor products

$$\psi = \mu \boxtimes \nu,$$

where $\mu$ is an irreducible, unitary, cuspidal automorphic representation of $GL(m)$, and $\nu$ is the unique irreducible $n$-dimensional representation of the group $SL(2,\mathbb{C})$, for positive integers $m$ and $n$ such that $N = mn$. For any such $\psi$, we form the induced representation

$$\mathcal{I}^G_P\big((\underbrace{\mu \otimes \cdots \otimes \mu}_{n})\delta_P^{\frac{1}{2}}\big), \tag{30.5}$$

of $GL(N,\mathbb{A})$, where $P$ is the standard parabolic subgroup of type $(m,\dots,m)$. We then write $\pi_\psi$ for the unique irreducible quotient of this representation. The theorem of Moeglin and Waldspurger asserts that the mapping $\psi \to \pi_\psi$ is a bijection from $\Psi_2\big(GL(N)\big)$ onto the set of automorphic representations of $GL(N)$ that occur in the discrete spectrum. Set

$$c(\psi) = \big\{c_v(\psi) : \ v \notin S\big\},$$

for any finite set $S \supset S_\infty$ of valuations outside of which $\mu$ is unramified, and semisimple conjugacy classes

$$c_v(\psi) = c_v(\mu) \otimes c_v(\phi_\nu) = c_v(\mu)q_v^{(\frac{n-1}{2})} \oplus \cdots \oplus c_v(\mu)q_v^{-(\frac{n-1}{2})}$$

in $GL(N,\mathbb{C})$. The family $c(\psi)$ then equals the family $c(\pi_\psi)$ attached to $\pi_\psi$ in §26.

We also represent the entire automorphic spectrum of $GL(N)$ by a larger set of formal objects. Let $\Psi\big(GL(N)\big)$ be the set of formal (unordered) direct sums

$$\psi = \ell_1\psi_1 \boxplus \cdots \boxplus \ell_r\psi_r \tag{30.6}$$

for positive integers $\ell_i$, and distinct elements $\psi_i = \mu_i \boxtimes \nu_i$ in $\Psi_2\big(GL(N_i)\big)$. The ranks $N_i$ are positive integers of the form $N_i = m_i n_i$ such that

$$N = \ell_1 N_1 + \cdots + \ell_r N_r = \ell_1 m_1 n_1 + \cdots + \ell_r m_r n_r.$$

For any $\psi$ as in (30.6), take $P$ to be the standard parabolic subgroup with Levi component

$$M = \big(\underbrace{GL(N_1) \times \cdots \times GL(N_1)}_{\ell_1}\big) \times \cdots \times \big(\underbrace{GL(N_r) \times \cdots \times GL(N_r)}_{\ell_r}\big),$$

and form the corresponding induced representation

$$\pi_\psi = \mathcal{I}_P^G\big((\underbrace{\pi_{\psi_1} \otimes \cdots \otimes \pi_{\psi_1}}_{\ell_1}) \otimes \cdots \otimes (\underbrace{\pi_{\psi_r} \otimes \cdots \otimes \pi_{\psi_r}}_{\ell_r})\big). \tag{30.7}$$

As a representation of $GL(N, \mathbb{A})$ induced from a unitary representation, $\pi_\psi$ is known to be irreducible [**Be**]. It follows from the theory of Eisenstein series, and Theorem 7.2 in particular, that $\psi \to \pi_\psi$ is a bijection from $\Psi\big(GL(N)\big)$ onto the set of irreducible representations of $GL(N, \mathbb{A})$ that occur in the spectral decomposition of $L^2\big(GL(N, F)\backslash GL(N, \mathbb{A})\big)$. We set

$$c(\psi) = \{c_v(\psi) : \ v \notin S\},$$

for any finite set $S \supset S_\infty$ outside of which each $\mu_i$ is unramified, and semisimple conjugacy classes

$$c_v(\psi) = \big(\underbrace{c_v(\psi_1) \oplus \cdots \oplus c_v(\psi_1)}_{\ell_1}\big) \oplus \cdots \oplus \big(\underbrace{c_v(\psi_r) \oplus \cdots \oplus c_v(\psi_r)}_{\ell_r}\big),$$

in $GL(N, \mathbb{C})$. Then $c(\psi)$ is again equal to $c(\pi_\psi)$. The theorem of Jacquet and Shalika mentioned in §26 [**JaS**] tells us that the mapping

$$\psi \longrightarrow c(\psi), \qquad \psi \in \Psi\big(GL(N)\big),$$

from $\Psi\big(GL(N)\big)$ to the set of (equivalence classes of) semisimple conjugacy classes in $GL(N, \mathbb{C})$, is injective.

There is an action $\pi_\psi \to \pi_\psi^\theta$ of the outer automorphism $\theta$ on the set of representations $\pi_\psi$. If $\psi$ is an element (30.6) in $\Psi\big(GL(N)\big)$, set

$$\begin{aligned} \psi^\theta &= \ell_1(\mu_1^\theta \boxtimes \nu_1^\theta) \boxplus \cdots \boxplus \ell_r(\mu_r^\theta \boxtimes \nu_r^\theta) \\ &= \ell(\mu_1^\theta \boxtimes \nu_1) \boxplus \cdots \boxplus \ell_r(\mu_r^\theta \boxtimes \nu_r), \end{aligned}$$

where $\mu_i^\theta$ is the contragredient of the cuspidal automorphic representation $\mu_i$ of $GL(m_i)$. (We can write $\nu_i^\theta = \nu_i$, since any irreducible representation of $SL(2, \mathbb{C})$ is self dual.) Then $\pi_\psi^\theta = \pi_{\psi^\theta}$. We introduce a subset

$$\widetilde{\Psi} = \Psi(\widetilde{G}) = \big\{\psi \in \Psi\big(GL(N)\big) : \ \psi^\theta = \psi\big\}$$

of elements in $\Psi\big(GL(N)\big)$ associated to the component

$$\widetilde{G} = \widetilde{G}_N = \ GL(N) \rtimes \theta.$$

It corresponds to those representations $\pi_\psi$ of $GL(N, \mathbb{A})$ that extend to group $\widetilde{G}(\mathbb{A})^+$ generated by $\widetilde{G}(\mathbb{A})$. We shall say that $\psi$ is *primitive* if $r = \ell_1 = n_1 = 1$. In other words, $\psi = \mu_1$ is a self-dual cuspidal automorphic representation of $GL(N)$. In this case $\psi$ has a central character $\eta_\psi$ of order 1 or 2.

We would like to think of the elements in $\widetilde{\Psi}$ as parameters. They ought to correspond to self dual, $N$-dimensional representations of a group $L_F \times SL(2,\mathbb{C})$, where $L_F$ is a global analogue of the local Langlands group $L_{F_v}$. The global Langlands group $L_F$ is purely hypothetical. It should be an extension of the global Weil group $W_F$, equipped with a conjugacy class of embeddings

$$\begin{array}{ccc} L_{F_v} & \longrightarrow & W_{F_v} \\ \downarrow & & \downarrow \\ L_F & \longrightarrow & W_F \end{array}$$

of each local group. The hypothetical group $L_F$ should ultimately play a fundamental role in the automorphic representation theory of any $G$. In the meantime, we attach an ad hoc substitute for $L_F$ to any $\psi$.

The proofs of the results we are going to describe include an extended induction argument. There are in fact both local and global induction hypotheses. We introduce the global hypothesis first, in order to define our substitutes for $L_F$.

**Global induction hypothesis.** *Suppose that $\psi \in \widetilde{\Psi}$ is primitive. Then there is a unique class $G_\psi = (G_\psi, {}^L G_\psi, s_\psi, \xi_\psi)$ of (twisted) elliptic endoscopic data in $\mathcal{E}_{\mathrm{ell}}(\widetilde{G})$ such that*

$$c(\psi) = \xi_\psi\big(c(\pi)\big),$$

*for some irreducible representation $\pi$ of $G(\mathbb{A})$ that occurs in $L^2_{\mathrm{disc}}\big(G(F)\backslash G(\mathbb{A})\big)$. Moreover $G_\psi$ is primitive.*

The assertion is quite transparent. Among all the (twisted) elliptic endoscopic data $G$ for $\widetilde{G}$, there should be exactly one source for the conjugacy class data of $\psi$. If $\psi$ happens to be attached to an irreducible, self-dual representation of a group $L_F$, it is an elementary exercise in linear algebra to show that the assertion is valid. That is, $\psi$ factors through the $L$-group of a unique $G_\psi \in \mathcal{E}_{\mathrm{ell}}(G)$, with $G_\psi$ being primitive. Of course, we do not know that $\psi$ is of this form. We do know that if $G_\psi$ is primitive, the dual group $\widehat{G}_\psi \subset GL(N,\mathbb{C})$ is purely orthogonal or symplectic. If $\eta_\psi \neq 1$ or $N$ is odd, $\widehat{G}_\psi$ is orthogonal, and $\eta_\psi$ determines $G_\psi$ uniquely. However, if $\eta_\psi = 1$ and $N$ is even, $\widehat{G}_\psi$ could be either symplectic or orthogonal. In this case, we will require a deeper property of $\psi$ to characterize $G_\psi$.

In proving the results, one fixes $N$, and assumes inductively that the hypothesis holds if $N$ is replaced by a positive integer $m < N$. The completion of the induction argument is of course part of what needs to be proved. Our purpose here is simply to state the results. Therefore, in order to save space, we shall treat the hypothesis as a separate theorem. In other words, we shall assume that it holds for $m = N$ as well.

Suppose that $\psi$ is an arbitrary element in $\widetilde{\Psi}$. Then $\theta$ acts by permutation on the indices $1 \leq i \leq \ell$ in (30.6). Let $I$ be the set of $i$ with $\psi_i^\theta = \psi_i$. The complement of $I$ is a disjoint union of two sets $J$ and $J'$, with a bijection $j \to j'$ from $J$ to $J'$, such that $\psi_j^\theta = \psi_{j'}$ for every $j \in J$. We can then write

$$\psi = \Big(\boxplus_{i\in I} \ell_i\psi_i\Big) \boxplus \Big(\boxplus_{j\in J} \ell_j(\psi_j \boxplus \psi_{j'})\Big).$$

If $i$ belongs to $I$, we apply the global induction hypothesis to the self-dual, cuspidal automorphic representation $\mu_i$ of $GL(m_i)$. This gives us a canonical datum $G_i = G_{\mu_i}$ in $\mathcal{E}_{\mathrm{prim}}(\widetilde{G}_{m_i})$. If $j$ belongs to $J$, we simply set $G_j = GL(m_j)$. We thus obtain a group $G_\alpha$ over $F$ for any index $\alpha$ in $I$ or $J$. Let ${}^L G_\alpha$ be the Galois form of its $L$-group. We can then form the fibre product

$$\mathcal{L}_\psi = \prod_{\alpha \in I \cup J} \left( {}^L G_\alpha \longrightarrow \Gamma_F \right) \tag{30.8}$$

of these groups over $\Gamma_F$. If $i$ belongs to $I$, the endoscopic datum $G_i$ comes with the standard embedding

$$\widetilde{\mu}_i : \ {}^L G_i \longrightarrow {}^L\big(GL(m_i)\big) = GL(m_i, \mathbb{C}) \times \Gamma_F.$$

If $j$ belongs to $J$, we define a standard embedding

$$\widetilde{\mu}_j : \ {}^L G_j \longrightarrow {}^L\big(GL(2m_j)\big) = GL(2m_j, \mathbb{C}) \times \Gamma_F$$

by setting

$$\widetilde{\mu}_j(g_j \times \sigma) = (g_j \oplus {}_t g_j^{-1}) \times \sigma \qquad\qquad g_j \in \widehat{G}_j = GL(m_j, \mathbb{C}),\ \sigma \in \Gamma_F.$$

We then define the $L$-embedding

$$\widetilde{\psi} : \ \mathcal{L}_\psi \times SL(2, \mathbb{C}) \longrightarrow {}^L\big(GL(N)\big) = GL(N, \mathbb{C}) \times \Gamma_F \tag{30.9}$$

by taking the appropriate direct sum

$$\widetilde{\psi} = \Big( \bigoplus_{i \in I} \ell_i (\widetilde{\mu}_i \otimes \nu_i) \Big) \oplus \Big( \bigoplus_{j \in J} \ell_j (\widetilde{\mu}_j \otimes \nu_j) \Big).$$

We can of course interpret the embedding $\widetilde{\psi} = \psi_{\widetilde{G}}$ also as an $N$-dimensional representation of $\mathcal{L}_\psi \times SL(2, \mathbb{C})$. With either interpretation, we are primarily interested in the equivalence class of $\widetilde{\psi}$, which is a $GL(N, \mathbb{C})$-conjugacy class of homomorphisms from $\mathcal{L}_\psi \times SL(2, \mathbb{C})$ to either $GL(N, \mathbb{C})$ or ${}^L\big(GL(N)\big)$.

Suppose that $G$ belongs to $\mathcal{E}_{\mathrm{ell}}(\widetilde{G})$. We write $\widetilde{\Psi}(G)$ for the set of $\psi \in \widetilde{\Psi}$ such that $\widetilde{\psi}$ factors through ${}^L G$. By this, we mean that there exists an $L$-homomorphism

$$\psi_G : \ \mathcal{L}_\psi \times SL(2, \mathbb{C}) \longrightarrow {}^L G \tag{30.10}$$

such that

$$\xi \circ \psi_G = \widetilde{\psi},$$

where $\xi$ is the embedding of ${}^L G$ into ${}^L\big(GL(N)\big)$ that is part of the twisted endoscopic datum represented by $G$. Since $\widetilde{\psi}$ and $\xi$ are to be regarded as $GL(N, \mathbb{C})$-conjugacy classes of homomorphisms, $\psi_G$ is determined only up to conjugacy by a subgroup of $GL(N, \mathbb{C})$. We define $\mathrm{Aut}_{\widetilde{G}}(G)$ to be the group of automorphisms of ${}^L G$ induced by conjugation of elements in $GL(N, \mathbb{C})$ that normalize the image of ${}^L G$. Then $\psi_G$ is to be regarded as an $\mathrm{Aut}_{\widetilde{G}}(G)$-orbit of $L$-homomorphisms (30.10). One sees easily that the quotient

$$\mathrm{Out}_{\widetilde{G}}(G) = \mathrm{Aut}_{\widetilde{G}}(G) / \mathrm{Int}(\widehat{G})$$

is trivial unless the integer $N_o$ attached to $G$ is even and positive, in which case it equals $\mathbb{Z}/2\mathbb{Z}$. In particular, if $G$ is primitive and equals an even orthogonal group, there can be two $\widehat{G}$-orbits of homomorphisms in the class of $\psi_G$. It is for this reason that we write $\widetilde{\Psi}(G)$ in place of the more natural symbol $\Psi(G)$.

If $\psi$ belongs to $\widetilde{\Psi}(G)$, we form the subgroup

$$S_\psi = S_\psi(G) = \mathrm{Cent}\Big(\widehat{G}, \psi_G\big(\mathcal{L}_\psi \times SL(2, \mathbb{C})\big)\Big) \tag{30.11}$$

of elements in $\widehat{G}$ that centralize the image of $\psi_G$. The quotient

$$\mathcal{S}_\psi = \mathcal{S}_\psi(G) = S_\psi / S_\psi^0 Z(\widehat{G})^\Gamma \tag{30.12}$$

is a finite abelian group, which plays a central role in the theory. Notice that there is a canonical element

$$s_\psi = \psi_G \left(1, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right) \tag{30.13}$$

in $S_\psi$. Its image in $\mathcal{S}_\psi$ (which we denote also by $s_\psi$) will be part of the description of nontempered automorphic representations.

Let $\widetilde{\Psi}_2$ be the subset of elements $\psi \in \widetilde{\Psi}$ such that the indexing set $J$ is empty, and such that $\ell_i = 1$ for each $i \in I$. A general element $\psi \in \widetilde{\Psi}$ always belongs to a set $\widetilde{\Psi}(G)$, for some datum $G \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$. It belongs to a unique such set if and only if it lies in $\widetilde{\Psi}_2$. If $G$ belongs to $\mathcal{E}_{\mathrm{ell}}(\widetilde{G})$, the intersection

$$\widetilde{\Psi}_2(G) = \widetilde{\Psi}(G) \cap \widetilde{\Psi}_2$$

is clearly the set of elements $\psi \in \widetilde{\Psi}(G)$ such that the group $S_\psi$ is finite. We shall write $\widetilde{\Psi}_{\mathrm{prim}}$ for the set of primitive elements in $\widetilde{\Psi}$. Then

$$\widetilde{\Psi}_{\mathrm{prim}} \subset \widetilde{\Psi}_2 \subset \widetilde{\Psi},$$

and

$$\widetilde{\Psi}_{\mathrm{prim}}(G) \subset \widetilde{\Psi}_2(G) \subset \widetilde{\Psi}(G),$$

where

$$\widetilde{\Psi}_{\mathrm{prim}}(G) = \widetilde{\Psi}(G) \cap \widetilde{\Psi}_{\mathrm{prim}}.$$

Suppose that

$$\psi = \psi_1 \boxplus \cdots \boxplus \psi_r$$

belongs to $\widetilde{\Psi}_2$. How do we determine the group $G \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$ such that $\psi$ lies in $\widetilde{\Psi}_2(G)$? To answer the question, we have to be able to write $N = N_s + N_o$ and $\psi = \psi_s \boxplus \psi_o$, where $\psi_s \in \Psi_2(\widetilde{G}_{N_s})$ is the sum of those components $\psi_i$ of symplectic type, and $\psi_o \in \Psi_2(\widetilde{G}_{N_o})$ is the sum of the components $\psi_i$ of orthogonal type. Consider a general component

$$\psi_i = \mu_i \boxtimes \nu_i.$$

The representation $\mu_i \in \Psi_{\mathrm{prim}}(\widetilde{G}_{m_i})$ has a central character $\eta_i = \eta_{\mu_i}$ of order 1 or 2. It gives rise to a datum $G_i \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G}_{m_i})$, according to the global inductive hypothesis, and hence a complex, connected classical group $\widehat{G}_i \subset GL(m, \mathbb{C})$. The $n_i$-dimensional representation $\nu_i$ of $SL(2, \mathbb{C})$ gives rise to a complex, connected classical group $\widehat{H}_i \subset GL(n_i, \mathbb{C})$, which contains its image. By considering principal unipotent elements, for example, the reader can check that $\widehat{H}_i$ is symplectic when $n_i$ is even, and is orthogonal when $n_i$ is odd. The tensor product of the bilinear forms that define $\widehat{G}_i$ and $\widehat{H}_i$ is a bilinear form on $\mathbb{C}^{N_i} = \mathbb{C}^{m_i n_i}$. This yields a complex, connected classical group $\widehat{G}_{\psi_i} \subset GL(N_i, \mathbb{C})$, which contains the image of $\widehat{G}_i \times \widehat{H}_i$ under the tensor product of the two standard representations. In concrete terms, $\widehat{G}_{\psi_i}$ is symplectic if one of $\widehat{G}_i$ and $\widehat{H}_i$ is symplectic and the other is orthogonal, and is orthogonal if both $\widehat{G}_i$ and $\widehat{H}_i$ are of the same type. This allows us to designate $\psi_i$ as either symplectic or orthogonal. It therefore gives us our decomposition $\psi = \psi_s \oplus \psi_o$. The component $\psi_s$ lies in the subset $\widetilde{\Psi}_2(G_s)$ of $\Psi_2(\widetilde{G}_{N_s})$, for the datum $G_s \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G}_{N_s})$ with dual group $\widehat{G}_s = Sp(N_s, \mathbb{C})$. The component $\psi_o$ lies in the subset $\widetilde{\Psi}_2(G_o)$ of $\Psi_2(\widetilde{G}_{N_o})$, for the datum $G_o \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G}_{N_o})$ with dual group $\widehat{G}_o = SO(N_o, \mathbb{C})$, and character

$$\eta_o = \prod_{i=1}^{r} (\eta_i)^{n_i}.$$

The original element $\psi$ therefore lies in $\widetilde{\Psi}_2(G)$, where $G$ is the product datum $G_s \times G_o$ in $\mathcal{E}_{\mathrm{ell}}(\widetilde{G})$. We note that

$$S_\psi(G) = \begin{cases} (Z/2\mathbb{Z})^r, & \text{if each } N_i \text{ is even,} \\ (\mathbb{Z}/2\mathbb{Z})^{r-1}, & \text{otherwise.} \end{cases}$$

Suppose now that $F$ is replaced by a completion $k = F_v$ of $F$. With this condition, we treat $\widetilde{G}^0 = GL(N)$ and $\widetilde{G} = GL(N) \rtimes \theta$ as objects over $k$, to which we add a subscript $v$ if there is any chance of confusion. As we noted in §28, one can introduce endoscopic data over $k$ by copying the definitions for the global field $F$. Similarly, one can introduce twisted endoscopic data for $\widetilde{G}$ over $k$. This gives local forms of the sets $\mathcal{E}_{\mathrm{prim}}(\widetilde{G}) \subset \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$.

We can also construct the sets $\Psi_2\big(GL(N)\big)$, $\Psi\big(GL(N)\big)$, and $\widetilde{\Psi} = \Psi(\widetilde{G})$ as objects over $k$. We define $\Psi_2\big(GL(N)\big)$ to be the set of formal tensor products $\psi = \mu \boxtimes \nu$, where $\mu$ is now an element in the set $\Pi_{\mathrm{temp},2}\big(GL(m,k)\big)$ of tempered irreducible representations of $GL(m,k)$ that are square integrable modulo the center. The other component $\nu$ remains an irreducible, $n$-dimensional representation of $SL(2,\mathbb{C})$, for a positive integer $n$ with $N = mn$. For any such $\psi$, we form the induced representation

$$\mathcal{I}_P^G\big((\underbrace{\mu \otimes \cdots \otimes \mu}_{n})\delta_P^{\frac{1}{2}}\big)$$

of $GL(N,k)$, as in (30.5). It has a unique irreducible quotient $\pi_\psi$, which is known to be unitary. The larger set $\Psi\big(GL(N)\big)$ is again the set of formal direct sums

$$\psi = \ell_1\psi_1 \boxplus \cdots \boxplus \ell_r\psi_r,$$

for positive integers $\ell_i$, and distinct elements $\psi_i = \mu_i \boxtimes \nu_i$ in $\Psi_2\big(GL(N_i)\big)$. For any such $\psi$, we form the induced representation

$$\pi_\psi = \mathcal{I}_P^G\big(\underbrace{(\pi_{\psi_1} \otimes \cdots \otimes \pi_{\psi_1})}_{\ell_1} \otimes \cdots \otimes \underbrace{(\pi_{\psi_r} \otimes \cdots \otimes \pi_{\psi_r})}_{\ell_r}\big)$$

of $GL(N,k)$, as in (30.7). It is irreducible and unitary. Finally, the local set $\widetilde{\Psi}$ is again the subset of elements $\psi$ in the local set $\Psi\big(GL(N)\big)$ such that $\psi^\theta = \psi$. It has subsets

$$\widetilde{\Psi}_{\mathrm{prim}} \subset \widetilde{\Psi}_2 \subset \widetilde{\Psi},$$

defined as in the global case.

We require a local form of our ad hoc substitute for the global Langlands group. Given the results of Harris-Taylor and Henniart, it is likely that one could work with the actual local Langlands group

$$L_{F_v} = \begin{cases} W_{F_v} \times SU(2), & \text{if } v \text{ is nonarchimedean,} \\ W_{F_v}, & \text{if } v \text{ is archimedean.} \end{cases}$$

However, the proof of the results described in this section still requires a local companion to the global induction hypothesis above. We may as well therefore use the local induction hypothesis to define analogues of the groups $\mathcal{L}_\psi$.

The local induction hypothesis depends on being able to attach a twisted character $\widetilde{f} \to \widetilde{f}_{\tilde{G}}(\psi)$ to any $\widetilde{f}$ and $\psi$ in the local sets $\mathcal{H}(\widetilde{G})$ and $\widetilde{\Psi}$. Suppose first that $\psi = \mu \otimes \nu$. By applying the theory of local Whittaker models to the local form of the induced representation (30.5), one can define a canonical extension of the quotient $\pi_\psi$ to $\widetilde{G}^+(k)$. This in turn provides a canonical extension of $\pi_\psi$ to $\widetilde{G}^+(k)$ for a general parameter $\psi \in \widetilde{\Psi}$. We define

$$\widetilde{f}_{\tilde{G}}(\psi) = \operatorname{tr}\big(\pi_\psi(\widetilde{f})\big), \qquad \widetilde{f} \in \mathcal{H}(\widetilde{G}).$$

On the other hand, we are assuming that the twisted form of the Langlands-Shelstad transfer conjecture holds for $k = F_v$. This gives a mapping

$$\widetilde{f} \longrightarrow \widetilde{f}^G$$

from $\mathcal{H}(\widetilde{G})$ to $S\mathcal{I}(G)$, for any twisted endoscopic datum $G$ for $\widetilde{G}$ over $k$.

**Local Induction Hypothesis.** *Suppose that $\psi \in \widetilde{\Psi}$ is primitive. Then there is a unique class $G_\psi \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$ such that $\widetilde{f}_{\tilde{G}}(\psi)$ is the pullback of some stable distribution*

$$h \longrightarrow h^{G_\psi}(\psi), \qquad h \in \mathcal{H}(G_\psi),$$

*on $G_\psi(k)$. In other words,*

$$\widetilde{f}_{\tilde{G}}(\psi) = \widetilde{f}^{G_\psi}(\psi), \qquad \widetilde{f} \in \mathcal{H}(\widetilde{G}).$$

*Moreover, $G_\psi$ is primitive.*

The assertion is less transparent than its global counterpart, for it is tailored to the fine structure of the terms in the spectral identities (30.1) and (30.3). It nonetheless serves the same purpose. Among all the local endoscopic data $G \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$ for $\widetilde{G}$, it singles out one that we can attach to $\psi$. As with the global hypothesis, we shall treat the local induction hypothesis as a separate theorem. In particular, we assume that it holds for $\Psi_{\mathrm{prim}}(\widetilde{G}_m)$, for any $m \leq N$.

We can now duplicate the constructions from the global case. If $\psi \in \widetilde{\Psi}$ is a general local parameter for the component $\widetilde{G} = \widetilde{G}_N$ over $k$, we obtain groups $G_i = G_{\mu_i}$ in $\mathcal{E}(\widetilde{G}_{m_i})$ for each $i$. We can then define the local form of the group $\mathcal{L}_\psi$. It is an extension of the local Galois group $\Gamma_k$, and comes with an $L$-embedding

$$\widetilde{\psi}:\ \mathcal{L}_\psi \times SL(2,\mathbb{C}) \longrightarrow {}^L GL(N) = GL(N,\mathbb{C}) \times \Gamma_k.$$

We again attach a subset $\widetilde{\Psi}(G)$ of $\widetilde{\Psi}$ to any $G \in \mathcal{E}_{\mathrm{ell}}(\widetilde{G})$. Any $\psi \in \widetilde{\Psi}(G)$ comes with an $\mathrm{Aut}_{\tilde{G}}(G)$-orbit of local $L$-embeddings (30.10), with $\xi \circ \psi_G = \widetilde{\psi}$. It also comes with the reductive group $S_\psi = S_\psi(G)$, the finite abelian group $\mathcal{S}_\psi = \mathcal{S}_\psi(G)$, and the element $s_\psi$ in either $S_\psi$ or $\mathcal{S}_\psi$, defined by (30.11), (30.12), and (30.13) respectively.

There are a few more observations to be made in the case $k = F_v$, before we can state the theorems. We first note that the definitions above make sense if $G$ is a general endoscopic datum for $\widetilde{G}$, rather than one that is just elliptic. The more general setting is required in the local context under discussion, since the localization of an elliptic global endoscopic datum need not remain elliptic.

Suppose that $G \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G})$. The putative Langlands-Shelstad mapping $\widetilde{f} \to \widetilde{f}^G$ takes $\mathcal{H}(\widetilde{G}) = \mathcal{H}\big(\widetilde{G}(k)\big)$ to the subspace $S\widetilde{\mathcal{I}}(G) = S\widetilde{\mathcal{I}}\big(G(k)\big)$ of functions in $S\mathcal{I}(G)$ that are invariant under the group $\mathrm{Out}_{\widetilde{G}}(G)$. We recall that this group is trivial unless $G$ is an even special orthogonal group $SO(N)$, in which case it is of order 2. In the latter case, the nontrivial element in $\mathrm{Out}_{\widetilde{G}}(G)$ is induced by conjugation of the nontrivial connected component in $O(N)$. By choosing a $k$-rational element in this component, we obtain an outer automorphism of $G(k)$ (regarded as an abstract group). We can therefore identify $\mathrm{Out}_{\widetilde{G}}(G)$ as a group of outer automorphisms of $G(k)$ of order 1 or 2. We write $\widetilde{\mathcal{I}}(G) = \widetilde{\mathcal{I}}\big(G(k)\big)$ for the space of functions in $\mathcal{I}(G)$ that are symmetric under $\mathrm{Out}_{\widetilde{G}}(G)$, and $\widetilde{\mathcal{H}}(G) = \widetilde{\mathcal{H}}\big(G(k)\big)$ for the space of functions in $\mathcal{H}(G)$ that are symmetric under the image of $\mathrm{Out}_{\widetilde{G}}(G)$ in $\mathrm{Aut}\big(G(k)\big)$ (relative to a suitable section). The mapping $f \to f_G$ then takes $\widetilde{\mathcal{H}}(G)$ onto $\widetilde{\mathcal{I}}(G)$, while the stable orbital integral mapping $f \to f^G$ takes $\widetilde{\mathcal{H}}(G)$ onto $S\widetilde{\mathcal{I}}(G)$. Let $\widetilde{\Pi}(G)$ denote the set of $\mathrm{Out}_{\widetilde{G}}(G)$-orbits in the set $\Pi(G) = \Pi\big(G(k)\big)$ of irreducible representations. We also write $\widetilde{\Pi}_{\mathrm{fin}}(G)$ for the set of formal, finite, nonnegative integral combinations of elements in $\widetilde{\Pi}(G)$. Any element $\pi \in \widetilde{\Pi}_{\mathrm{fin}}$ then determines a linear form

$$f \longrightarrow f_G(\pi), \qquad f \in \widetilde{\mathcal{H}}(G),$$

on $\widetilde{\mathcal{H}}(G)$. We write $\widetilde{\Pi}_{\mathrm{unit}}(G)$ and $\widetilde{\Pi}_{\mathrm{fin,unit}}(G)$ for the subsets of $\widetilde{\Pi}(G)$ and $\widetilde{\Pi}_{\mathrm{fin}}(G)$ built out of unitary representations. By taking the appropriate product, we can extend these definitions to any endoscopic datum $G$ for $\widetilde{G}$.

Suppose again that $G \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G})$, and that $\psi$ belongs to $\widetilde{\Psi}(G)$. Suppose also that $s'$ is a semisimple element in $S_\psi(G)$. Let $\widehat{G}'$ be the connected centralizer $\widehat{G}_{s'}$ of $s'$ in $\widehat{G}$, and set

$$\mathcal{G}' = \widehat{G}'\psi_G(\mathcal{L}_\psi).$$

Then $\mathcal{G}'$ is an $L$-subgroup of ${}^LG$, for which the identity embedding $\xi'$ is an $L$-homomorphism. We take $G' = G_{s'}$ to be a quasisplit group for which $\widehat{G}'$, with the $L$-action of $\Gamma_F$ induced by $\mathcal{G}'$, is a dual group. We thus obtain an endoscopic datum $(G', \mathcal{G}', s', \xi')$ for $G$. Now the set $\widetilde{\Psi}(G')$ can be defined as an obvious Cartesian product of sets we have already constructed. Since $s'$ lies in the centralizer of the image of $\mathcal{L}_\psi$ in $\widehat{G}$, $\psi_G$ factors through ${}^LG'$. We obtain an $L$-embedding

$$\psi_{G'}: \ \mathcal{L}_\psi \times SL(2, \mathbb{C}) \longrightarrow {}^LG'$$

such that

$$\xi' \circ \psi_{G'} = \psi_G,$$

and a corresponding element $\psi' = \psi_{s'}$ in $\widetilde{\Psi}(G')$. Once again, this construction extends to the case that $G$ is a general twisted endoscopic datum for $\widetilde{G}$.

There is one final technical complication. We want the local objects $\psi$ over $k = F_v$ to represent local components at $v$ of global parameters associated to automorphic representations of $GL(N)$. Because we do not know that the extension to $GL(N)$ of Ramanujan's conjecture is valid, we do not know that the local components are tempered. This requires a minor generalization of the local set $\widetilde{\Psi}$ attached to $k = F_v$. We define a larger set $\widetilde{\Psi}^+ = \Psi^+(\widetilde{G})$ of formal direct sums

$$\psi = \ell_1\psi_1 \boxplus \cdots \boxplus \ell_r\psi_r,$$

by relaxing the condition on the representations $\mu_i$ in components $\psi_i = \mu_i \boxtimes \nu_i$. We require only that $\mu_i$ belong to the set $\Pi_2\big(GL(m_i, k)\big)$. In other words, $\mu_i$ is an irreducible representation of $GL(m_i, k)$ that is square integrable modulo the center, but whose central character need not be unitary. This condition applies only to the components $\mu_i$ such that $\mu_i^\theta \neq \mu_i$, since the central character of $\mu_i$ would otherwise have order 2. If $\psi$ belongs to $\widetilde{\Psi}^+$, the twisted character $\widetilde{f} \to \widetilde{f}_{\tilde{G}}(\psi)$ is defined from the tempered case by analytic continuation in the central characters of the components $\mu_i$. The various other objects we have associated to the set $\widetilde{\Psi}$ are also easily formulated for the larger set $\widetilde{\Psi}^+$.

We shall now state the results as three theorems. They are conditional on the fundamental lemma, and the further requirements discussed at the beginning of the section.

THEOREM 30.1. *Assume that $k = F_v$ is local, and that $G \in \mathcal{E}_{\mathrm{prim}}(\widetilde{G})$.*

(a) *For each $\psi \in \widetilde{\Psi}(G)$, there is a stable linear form $h \to h^G(\psi)$ on $\widetilde{\mathcal{H}}(G)$ such that*

$$\widetilde{f}_{\tilde{G}}(\psi) = \widetilde{f}^G(\psi), \qquad \widetilde{f} \in \mathcal{H}(\widetilde{G}).$$

(b) *For each $\psi \in \widetilde{\Psi}(G)$, there is a finite subset $\widetilde{\Pi}_\psi$ of $\widetilde{\Pi}_{\mathrm{fin,unit}}(G)$, together with an injective mapping*

$$\pi \longrightarrow \langle \cdot, \pi \rangle, \qquad \pi \in \widetilde{\Pi}_\psi,$$

*from $\widetilde{\Pi}_\psi$ to the group of characters $\widehat{\mathcal{S}}_\psi(G)$ on $\mathcal{S}_\psi(G)$ that satisfies the following condition. For any $s' \in S_\psi(G)$,*

$$f^{G'}(\psi') = \sum_{\pi \in \widetilde{\Pi}_\psi} \langle s_\psi s, \pi \rangle f_G(\pi), \qquad f \in \widetilde{\mathcal{H}}(G), \tag{30.14}$$

*where $G' = G'_{s'}$, $\psi' = \psi'_{s'}$, and $s$ is the image of $s'$ in $\mathcal{S}_\psi(G)$.*

(c) *Let $\widetilde{\Phi}_{\mathrm{temp}}(G)$ denote the subset of elements in $\widetilde{\Psi}(G)$ for which each of the $SL(2,\mathbb{C})$ components $\nu_i$ is trivial. Then if $\phi \in \widetilde{\Phi}_{\mathrm{temp}}(G)$, the elements in $\widetilde{\Pi}_\phi$ are tempered and irreducible, in the sense that they belong to the set $\widetilde{\Pi}_{\mathrm{temp}}(G)$ of $\mathrm{Out}_{\tilde{G}}(G)$-orbits in $\Pi_{\mathrm{temp}}(G)$. Moreover, every element in $\widetilde{\Pi}_{\mathrm{temp}}(G)$ belongs to exactly one packet $\widetilde{\Pi}_\phi$. Finally, if $k$ is nonarchimedean, the mapping $\widetilde{\Pi}_\phi \to \widehat{\mathcal{S}}_\phi$ is bijective.* □

**Remarks:** 1. The assertions (b) and (c) of the theorem are new only in the nonarchimedean case. (For archimedean $v$, they are special cases of results of Shelstad [**She3**] and Adams, Barbasch, and Vogan [**ABV**].) If $v$ is nonarchimedean, assertion (c) can be combined with the local Langlands conjecture for $GL(N)$ [**HT**], [**He**]. This ought to yield the local Langlands conjecture for $G$, at least in the case that $\mathrm{Out}_{\tilde{G}}(G) = 1$.

2. The transfer mapping $f \to f^{G'}$ in (b) depends on a normalization for the transfer factors $\Delta_G(\delta', \gamma)$ for the quasisplit group $G'$. We assume implicitly that $\Delta_G(\delta', \gamma)$ equals the function denoted $\Delta_0(\delta', \gamma)$ on p. 248 of [**LS1**]. This is the reason that the characters $\langle \cdot, \pi \rangle$ on $\mathcal{S}_\varphi$ attached to an element $\phi \in \widetilde{\Phi}_{\mathrm{temp}}(G)$ are slightly simpler than in the general formulation (28.8).

3. Suppose that $\psi$ lies in the larger set $\widetilde{\Psi}^+(G)$. We can then combine the theorem with a discussion similar to that of (28.9). In particular, we can identify $\psi$ with the image in $\widetilde{\Psi}^+(G)$ induced from a nontempered twist $\psi_{M,\lambda}$, where $M$ is a Levi subgroup of $G$, $\psi_M$ is an element in $\widetilde{\Psi}(M)$, and $\lambda$ is a point in $(\mathfrak{a}_M^*)_P^+$. We can then form the corresponding induced packet

$$\widetilde{\Pi}_\psi = \{\mathcal{I}_P^G(\pi_{M,\lambda}) : \ \pi_M \in \widetilde{\Pi}_{\psi_M}\}$$

for $G(k)$. Since we are dealing with full induced representations, rather than Langlands quotients, the assertions of the theorem extend to $\widetilde{\Pi}_\psi$.

THEOREM 30.2. *Assume that $k = F$ is global, and that $G \in \mathcal{E}_{\text{prim}}(\widetilde{G})$.*

(a) *Suppose that $\psi \in \widetilde{\Psi}(G)$. If $v$ is any valuation of $F$, the localization $\psi_v$ of $\psi$, defined in the obvious way as an element in the set $\widetilde{\Psi}_v^+ = \Psi^+(\widetilde{G}_v)$, has the property that $\mathcal{L}_{\psi_v}$ is contained in $\mathcal{L}_\psi$. In particular, $\psi_v$ belongs to $\widetilde{\Psi}^+(G_v)$, $S_\psi(G)$ is contained in $S_{\psi_v}(G_v)$, and there is a canonical homomorphism $s \to s_v$ from $\mathcal{S}_\psi(G)$ to $\mathcal{S}_{\psi_v}(G_v)$. We can therefore define a global packet*

$$\widetilde{\Pi}_\psi = \Big\{ \bigotimes_v \pi_v : \ \pi_v \in \widetilde{\Pi}_{\psi_v}, \ \langle \cdot, \pi_v \rangle = 1 \textit{ for almost all } v \Big\},$$

*and for each element $\pi = \bigotimes_v \pi_v$ in $\widetilde{\Pi}_\psi$, a character*

$$\langle s, \pi \rangle = \prod_v \langle s_v, \pi_v \rangle, \qquad s \in \mathcal{S}_v,$$

*on $\mathcal{S}_\psi = \mathcal{S}_\psi(G)$.*

(b) *Define a subalgebra of $\mathcal{H}(G)$ by taking the restricted tensor product*

$$\widetilde{\mathcal{H}}(G) = \bigotimes_v^{\text{rest}} \widetilde{\mathcal{H}}(G_v).$$

*Then there is an $\widetilde{\mathcal{H}}(G)$-module isomorphism*

$$L^2_{\text{disc}}\big(G(F)\backslash G(\mathbb{A})\big) \cong \bigoplus_{\psi \in \tilde{\Psi}_2(G)} m_\psi \Big( \bigoplus_{\{\pi \in \tilde{\Pi}_\psi : \langle \cdot, \pi \rangle = \varepsilon_\psi\}} \pi \Big), \tag{30.15}$$

*where $m_\psi$ equals 1 or 2, and*

$$\varepsilon_\psi : \ \mathcal{S}_\psi \ \longrightarrow \ \{\pm 1\}$$

*is a linear character defined explicitly in terms of symplectic root numbers.* □

**Remarks.** 4. The multiplicity $m_\psi$ is defined to be the number of $\widehat{G}$-orbits of embeddings

$$\mathcal{L}_\psi \times SL(2, \mathbb{C}) \ \longrightarrow \ {}^L G$$

in the $\mathrm{Aut}_{\tilde{G}}(G)$-orbit of $\psi_G$. We leave the reader to check that $m_\psi$ equals 1 unless $N$ is even, $\widehat{G} = SO(N, \mathbb{C})$, and the rank $N_i$ of each of the components $\psi_i = \mu_i \otimes \nu_i$ of $\psi$ is also even, in which case $m_\psi = 2$.

5. The sign character $\varepsilon_\psi$ is defined as follows. We first define an orthogonal representation

$$\tau_\psi : \ S_\psi \times \mathcal{L}_\psi \times SL(2, \mathbb{C}) \ \longrightarrow \ GL(\widehat{\mathfrak{g}})$$

on the Lie algebra $\widehat{\mathfrak{g}}$ of $\widehat{G}$ by setting

$$\tau_\psi(s,g,h) = \mathrm{Ad}\big(s\psi_G(g\times h)\big), \qquad s\in S_\psi,\ g\in\mathcal{L}_\psi,\ h\in SL(2,\mathbb{C}).$$

We then write

$$\tau_\psi = \bigoplus_\alpha \tau_\alpha = \bigoplus_\alpha \big(\lambda_\alpha\otimes\mu_\alpha\otimes\nu_\alpha\big),$$

for irreducible representations $\lambda_\alpha$, $\mu_\alpha$ and $\nu_\alpha$ of $S_\psi$, $\mathcal{L}_\psi$ and $SL(2,\mathbb{C})$ respectively. Given the definition of the global group $\mathcal{L}_\psi$, we can regard $L(s,\mu_\alpha)$ as an automorphic $L$-function for a product of general linear groups. One checks that it is among those $L$-functions for which one has analytic continuation, and a functional equation

$$L(s,\mu_\alpha) = \varepsilon(s,\mu_\alpha)L(1-s,\mu_\alpha^\vee).$$

In particular, if $\mu_\alpha^\vee=\mu_\alpha$, $\varepsilon(\frac12,\mu_\alpha)=\pm1$. Let $\mathcal{A}$ be the set of indices $\alpha$ such that

(i) $\tau_\alpha^\vee=\tau_\alpha$ (and hence $\mu_\alpha^\vee=\mu_\alpha$),
(ii) $\dim(\nu_\alpha)$ is even (and hence $\nu_\alpha$ is symplectic),
(iii) $\varepsilon(\frac12,\mu_\alpha)=-1$.

Then

$$\varepsilon_\psi(s) = \prod_{\alpha\in\mathcal{A}} \det\big(\lambda_\alpha(s)\big), \qquad s\in S_\psi. \tag{30.16}$$

Theorem 30.3. *Assume that $F$ is global.*

(a) *Suppose that $G\in\mathcal{E}_{\mathrm{prim}}(\widetilde{G})$, and that $\psi=\mu$ belongs to $\widetilde{\Psi}_{\mathrm{prim}}(G)$. Then $\widehat{G}$ is orthogonal if and only if the symmetric square $L$-function $L(s,\mu,S^2)$ has a pole at $s=1$, while $\widehat{G}$ is symplectic if and only if the skew-symmetric $L$-function $L(s,\mu,\Lambda^2)$ has a pole at $s=1$.*

(b) *Suppose that for $i=1,2$, $G_i\in\mathcal{E}_{\mathrm{prim}}(\widetilde{G}_{N_i})$ and that $\psi_i=\mu_i$ belongs to $\widetilde{\Psi}_{\mathrm{prim}}(G_i)$. Then the corresponding Rankin-Selberg $\varepsilon$-factor satisfies*

$$\varepsilon(\tfrac12,\mu_1\times\mu_2)=1,$$

*provided that $\widehat{G}_1$ and $\widehat{G}_2$ are either both orthogonal or both symplectic.* □

**Remarks:** 6. Suppose that $\mu$ is as in (i). It follows from the fact $\mu^\theta=\mu$ that

$$L(s,\mu\times\mu) = L(s,\mu,S^2)L(2,\mu,\Lambda^2).$$

The Rankin-Selberg $L$-function on the left is known to have a pole of order 1 at $s=1$. One also knows that neither of the two $L$-functions on the right can have a zero at $s=1$. The assertion of (a) is therefore compatible with our a priori knowledge of the relevent $L$-functions. It is also compatible with properties of the corresponding Artin $L$-functions, in case $\mu$ is attached to an irreducible $N$-dimensional representation of $\Gamma_F$ or $W_F$. The assertion is an essential part of both the resolution of the global induction hypothesis and the proof of the multiplicity formula (30.15).

7. Consider the assertion of (b). If $\mu_1$ and $\mu_2$ are both attached to irreducible representations of $W_F$, the conditions of (b) reduce to the requirement that the tensor product of the two representations be orthogonal. The assertion of (b) is known in this case [**D2**]. The general assertion (b) is again intimately related to the global induction hypothesis and the multiplicity formula (30.15).

We shall add a few observations on the "tempered" case of the multiplicity formula (30.15). Assume that $G \in \mathcal{E}_{\text{prim}}(\widetilde{G})$, as in Theorem 30.2. Let us write $L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ for the subspace of $L^2_{\text{disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ whose irreducible constituents transfer to *cuspidal* Eisenstein series for $\widetilde{G}^0 = GL(N)$. (The notation anticipates a successful resolution of the Ramanujan conjecture for $GL(N)$, which given our theorems, would imply that $L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ is indeed the subspace of $L^2_{\text{disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ whose irreducible constituents are tempered.) Let $\widetilde{\Phi}_2(G) = \widetilde{\Phi}_{\text{temp},2}(G)$ be the subset of elements in the global set $\widetilde{\Psi}_2(G)$ for which the $SL(2,\mathbb{C})$-components $\nu_i$ are all trivial. Then $\varepsilon_\phi = 1$ for every $\phi \in \widetilde{\Phi}_2(G)$. The formula (30.15) therefore provides an $\widetilde{\mathcal{H}}(G)$-module isomorphism

$$L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big) \cong \bigoplus_{\phi\in\widetilde{\Phi}_2(G)} m_\phi\Big(\bigoplus_{\{\pi\in\widetilde{\Pi}_\phi:\langle\cdot,\pi\rangle=1\}} \pi\Big). \tag{30.17}$$

Suppose that $N$ is odd, or that $\widehat{G} = Sp(N,\mathbb{C})$. Then $m_\phi = 1$. It is also easy to see that $\widetilde{\mathcal{H}}(G_v) = \mathcal{H}(G_v)$ for any $v$, so that $\widetilde{\mathcal{H}}(G) = \mathcal{H}(G)$ in this case. Moreover, the local packets $\widetilde{\Pi}_{\phi_v} = \Pi_{\phi_v}$ attached to elements $\phi_v$ in the set $\widetilde{\Phi}_{\text{temp}}(G_v) = \Phi_{\text{temp}}(G_v)$ contain only irreducible representations of $G(F_v)$. Now the local component $\phi_v$ of an element $\phi$ in the global set $\widetilde{\Phi}_2(G) = \Phi_2(G)$ could lie in a set $\Phi^+_{\text{temp}}(G_v) \subset \Phi^+(G_v)$ that properly contains $\Phi_{\text{temp}}(G_v)$. However, it is likely that the induced representations that comprise the corresponding packet $\Pi_{\phi_v}$ are still irreducible. (I have not checked this point in general, but it should be a straightforward consequence of the well known structure of generic, irreducible, unitary representations of $GL(N,F_v)$.) Taking the last point for granted, we see that the global packet

$$\Pi_\phi = \Big\{\bigotimes_v \pi_v :\ \pi_v \in \Pi_{\phi_v},\ \langle\cdot,\pi_v\rangle = 1 \text{ for almost all } v\Big\}$$

attached to any $\phi \in \Phi_2(G)$ contains only irreducible representations of $G(\mathbb{A})$. The injectivity of the mapping $\pi \to c(\pi)$ implies that the global packets are disjoint. It then follows from (30.17) that $L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ decomposes with multiplicity 1 in this case.

In the remaining case, $N$ is even and $\widehat{G} = SO(N,\mathbb{C})$. If one of the integers $N_i$ attached to a given global element $\phi \in \widetilde{\Phi}_2(G)$ is odd, $m_\phi$ equals 1. An argument like that above then implies that the irreducible constituents of $L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ attached to $\phi$ have multiplicity 1. However, if the integers $N_i$ attached to $\phi$ are all even, $m_\phi$ equals 2. The multiplicity formula (30.17) then becomes more interesting. It depends in fact on the integers

$$N_{v,i}, \qquad\qquad 1 \leq i \leq \ell_v,$$

attached to the local components $\phi_v$ of $\phi$. If for some $v$, all of these integers are even, (30.17) can be used to show that the irreducible constituents of $L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ attached to $\phi$ again have multiplicity 1. However, it could also happen that for every $v$, one of the integers $N_{v,i}$ is odd. A slightly more elaborate analysis of (30.17) then leads to the conclusion that the irreducible constituents of $L^2_{\text{temp,disc}}\big(G(F)\backslash G(\mathbb{A})\big)$ attached to $\phi$ all have multiplicity 2. This represents a quantitative description of a phenomenon investigated by M. Larsen in terms of representations of Galois groups [**Lar**, p. 253].

The discussion of this section has been restricted to quasisplit orthogonal and symplectic groups. It is of course important to treat other classical groups as well. For example, there ought to be a parallel theory for quasisplit unitary groups over $F$. The case of unitary groups is in fact somewhat simpler. Moreover, a proof of the fundamental lemma for unitary groups has been announced recently by Laumon and Ngo [**LN**]. It is quite possible that their methods could be extended to weighted orbital integrals and their twisted analogues. The goal would be to extend the results of Rogawski for $U(3)$ [**Ro2**], [**Ro3**] to general rank.

Finally, we note that there has been considerable progress recently in applying other methods to classical groups. These methods center around the theory of $L$-functions, and a generalization [**CP**] of Hecke's converse theorem for $GL(2)$. They apply primarily to generic representations (both local and global) of classical groups, but they do not depend on the fundamental lemma. We refer the reader to [**Co**] for a general introduction, and to selected papers [**CKPS1**], [**CKPS2**], [**JiS**] and [**GRS**].

## Afterword: beyond endoscopy

The principle of functoriality is one of the pillars of the Langlands program. It is among the deepest problems in mathematics, and has untold relations to other questions. For example, the work of Wiles suggests that functoriality is inextricably intertwined with that second pillar of the Langlands program, the general analogue of the Shimura-Taniyama-Weil conjecture [**Lan7**].

The theory of endoscopy, which is still largely conjectural, analyzes representations of $G$ in terms of representations of its endoscopic groups $G'$. In its global form, endoscopy amounts to a comparison of trace formulas, namely the invariant (or twisted) trace formula for $G$ with stable trace formulas for $G'$. It includes the applications we discussed in §25, §26, and §30 as special cases. The primary aim of endoscopy is to organize the representations of $G$ into packets. It can be regarded as a first attempt to describe the fibres of the mapping

$$\pi \longrightarrow c(\pi)$$

from automorphic representations to families of conjugacy classes. However, it also includes functorial correspondences for the $L$-homomorphisms

$$\xi' : \ {}^{L}G' \longrightarrow {}^{L}G$$

attached to endoscopic groups $G'$ for $G$ (in cases where $\mathcal{G}'$ can be identified with an $L$-group ${}^{L}G'$).

The general principle of functoriality applies to an $L$-homomorphism

$$\rho : \ {}^{L}G' \longrightarrow {}^{L}G \tag{A.1}$$

attached to any pair $G'$ and $G$ of quasisplit groups. As a strategy for attacking this problem, the theory of endoscopy has obvious theoretical limitations. It pertains, roughly speaking, to the case that ${}^{L}G'$ is the group of fixed points of a semisimple $L$-automorphism of ${}^{L}G$. Most mappings $\rho$ do not fall into this category.

Suppose for example that $G' = GL(2)$ and $G = GL(m+1)$, and that $\rho$ is the $(m+1)$-dimensional representation of $\widehat{G}' = GL(2,\mathbb{C})$ defined by the $m^{\text{th}}$ symmetric power of the standard two-dimensional representation. If $m = 2$, the image of $GL(2,\mathbb{C})$ in ${}^{L}G = GL(3,\mathbb{C})$ is essentially an orthogonal group. In this case, the problem is endoscopic, and is included in the theory of classical groups discussed in §30. (In fact, functoriality was established in this case by other means some years ago [**GeJ**].) In the case $m = 3$ and $m = 4$, functoriality was established recently by Kim and Shahidi [**KiS**] and Kim [**Ki**]. These results came as a considerable surprise. They were proved by an ingenious combination of the converse theorems of Cogdell and Piatetskii-Shapiro with the Langlands-Shahidi method. If $m \geq 5$, however, these methods do not seem to work. Since the problem is clearly not endoscopic in this case, none of the known techniques appear to hold any hope of success.

We are going to conclude with a word about some recent ideas of Langlands[1] [**Lan13**], [**Lan15**]. The ideas are quite speculative. They have yet to be shown to apply even heuristically to new cases of functoriality. However, they have the distinct advantage that everything else appears to fail in principle. The ideas are in any case intriguing. They are based on applications of the trace formula that have never before been considered.

[1] I thank Langlands for enlightening conversations on the topic.

The difficulty in attacking the general case (A.1) of functoriality is that it is hard to characterize the image of ${}^LG'$ in ${}^LG$. If $\rho({}^LG')$ is the group of fixed points of some outer automorphism, $G'$ will be related to a twisted endoscopic group for $G$. The corresponding twisted trace formula isolates automorphic representations of $G$ that are fixed by the outer automorphism. A comparison of this formula with stable trace formulas for the associated collection of twisted endoscopic groups is aimed, roughly speaking, at those $L$-subgroups of ${}^LG$ fixed by automorphisms in the given inner class. If the image of ${}^LG'$ in ${}^LG$ is more general, however, the problem becomes much more subtle. Is it possible to use the trace formula in a way that counts only the automorphic representations of $G(\mathbb{A})$ that are functorial images of automorphic representations of $G'$?

Suppose that $r$ is some finite dimensional representation of ${}^LG$. We write $V_{\mathrm{ram}}(G,r)$ for the finite set of valuations $v$ of $F$ at which either $G$ or $r$ is ramified. For $\rho$ as in (A.1), the composition $r\circ\rho$ is a finite dimensional representation of ${}^LG'$. If this representation contains the trivial representation of ${}^LG'$, and the $L$-function $L(s,\pi',r\circ\rho)$ attached to a given automorphic representation $\pi'$ of $G'$ has the expected analytic continuation, the $L$-function will have a pole at $s=1$. The same would therefore be true of the $L$-function $L(s,\pi,r)$ attached to an automorphic representation $\pi$ of $G$ that is a functorial image of $\pi'$ under $\rho$. On the other hand, so long as $r$ does not contain the trivial representation of ${}^LG$, there will be many automorphic representations $\pi$ of $G$ for which $L(s,\pi,r)$ does not have a pole at $s=1$. One would like to have a trace formula that includes only the automorphic representations $\pi$ of $G$ for which $L(s,\pi,\rho)$ has a pole at $s=1$.

The objects of interest are of course automorphic representations $\pi$ of $G$ that occur in the discrete spectrum. The case that $\pi$ is nontempered is believed to be more elementary, in the sense that it should reduce to the study of tempered automorphic representations of groups $G_\psi$ of dimension smaller than $G$ [**A17**]. The primary objects are therefore the representations $\pi$ that are tempered, and hence cuspidal. If $\pi$ is a tempered, cuspidal automorphic representation of $G$, $L(s,\pi,r)$ should have a pole at $s=1$ of order equal to that of the unramified $L$-function

$$L^V(s,\pi,r)=\prod_{v\notin V}\det\big(1-r(c(\pi_v))q_v^{-s}\big)^{-1},$$

attached to any finite set $V\supset S_{\mathrm{ram}}(G,r)$ outside of which $\pi$ is unramified. The partial $L$-function $L^V(s,\pi,r)$ is not expected to have a zero at $s=1$. The order of its pole will thus equal

$$n(\pi,r)=\operatorname*{Res}_{s=1}\Big(-\frac{d}{ds}\log L^V(s,\pi,r)\Big),$$

a nonnegative integer that is independent of $V$.

We can write

$$\begin{aligned}
&-\frac{d}{ds}\log L^V(s,\pi,r)\\
&=\sum_{v\notin V}\frac{d}{ds}\log\big(\det(1-r(c(\pi_v))q_v^{-s})\big)\\
&=\sum_{v\notin V}\sum_{k=1}^{\infty}\log(q_v)\mathrm{tr}\big(r(c(\pi_v))^k\big)q_v^{-ks},
\end{aligned}$$

for $\mathrm{Re}(s)$ large. Since $\pi$ is assumed to be tempered, the projection of any conjugacy class $c(\pi_v)$ onto $\widehat{G}$ is bounded, in the sense that it intersects any maximal compact subgroup of $\widehat{G}$. It follows that the set of coefficients

$$\bigl\{\mathrm{tr}\bigl(r(c(\pi_v))^k\bigr) :\ v \notin V,\ k \geq 1\bigr\}$$

is bounded, and hence that the last Dirichlet series actually converges for $\mathrm{Re}(s) > 1$. Since $\pi$ is also assumed to be cuspidal automorphic, $L^V(s, \pi, r)$ is expected to have analytic continuation to a meromorphic function on the complex plane. The last Dirichlet series will then have at most a simple pole at $s = 1$, whose residue can be described in terms of the coefficients. Namely, by a familiar application of the Wiener-Ikehara tauberian theorem, there would be an identity

$$n(\pi, r) = \lim_{N\to\infty} \Bigl(V_N^{-1} \sum_{\{v\notin V:\, q_v\leq N\}} \mathrm{tr}\bigl(r(c(\pi_v))\bigr)\Bigr), \tag{A.2}$$

where

$$V_N = |\{v \notin V :\ q_v \leq N\}|.$$

(See [**Ser1**, p. I-29]. Observe that the contribution of the coefficients with $k > 1$ to the Dirchlet series is analytic at $s = 1$, and can therefore be ignored.)

Langlands proposes to apply the trace formula to a family of functions $f_N$ that depend on the representation $r$. We begin with an arbitrary function $f \in \mathcal{H}\bigl(G(\mathbb{A})\bigr)$. If $V \supset S_{\mathrm{ram}}(G)$ is a finite set of valuations such that $f$ belongs to the subspace $\mathcal{H}\bigl(G(F_V)\bigr)$ of $\mathcal{H}\bigl(G(\mathbb{A})\bigr)$, and $\phi$ belongs to the unramified Hecke algebra $\mathcal{H}\bigl(G(\mathbb{A}^V), K^V\bigr)$, the product

$$f_\phi :\ x \ \rightarrow\ f(x)\phi(x^V), \qquad x \in G(\mathbb{A}),$$

also belongs to $\mathcal{H}\bigl(G(\mathbb{A})\bigr)$. We choose the function $\phi = \phi_N$ so that it depends on $r$, as well as a positive integer $N$. Motivated by (A.2), and assuming that $V$ contains the larger finite set $S_{\mathrm{ram}}(G, r)$, we define $\phi_N$ by the requirement that

$$(\phi_N)_G(\pi^V) = \sum_{\{v\notin V:\, q_v\leq N\}} r\bigl(c(\pi_v)\bigr),$$

for any unramified representation $\pi^V$ of $G(\mathbb{A}^V)$. Then

$$n(\pi, r) = \lim_{N\to\infty} \bigl((\phi_N)_G(\pi^V)V_N^{-1}\bigr),$$

for any $\pi$ as in (A.2). The products

$$f_N = f_N^r = f_{\phi_N}, \qquad N \geq 1,$$

or rather their images in $\mathcal{H}(G)$, are the relevant test functions.

Set

$$I_{\mathrm{temp,cusp}}(f) = \mathrm{tr}\bigl(R_{\mathrm{temp,cusp}}(f)\bigr),$$

where $R_{\mathrm{temp,cusp}}$ is the representation of $G(\mathbb{A})^1$ on the subspace of $L^2_{\mathrm{cusp}}\bigl(G(F)\backslash G(\mathbb{A})^1\bigr)$ that decomposes into tempered representations $\pi$ of $G(\mathbb{A})^1$. Suppose that we happen to know that $L^V(s, \pi, r)$ has analytic continuation for each such $\pi$. Then the sum

$$\sum_{\pi} n(\pi, r) m_{\mathrm{cusp}}(\pi) f_G(\pi), \tag{A.3}$$

taken over irreducible tempered representations $\pi$ of $G(\mathbb{A})^1$, equals the limit

$$I^r_{\mathrm{temp,cusp}}(f) = \lim_{N\to\infty}\big(I_{\mathrm{temp,cusp}}(f_N)V_N^{-1}\big).$$

However, it is conceivable that one could investigate the limit $I^r_{\mathrm{temp,cusp}}(f)$ without knowing the analytic continuation of the $L$-functions. The term $I_{\mathrm{temp,cusp}}(f_N)$ in this limit is part of the invariant trace formula for $G$. It is the sum over $t \geq 0$ of the tempered, cuspidal part of the term with $M = G$ in the $t$-discrete part $I_{t,\mathrm{disc}}(f_N)$. (We recall that the linear form $I_{t,\mathrm{disc}}$ is defined by a sum (21.19) over Levi subgroups $M$ of $G$.) For each $N$, one can replace $I_{\mathrm{temp,cusp}}(f_N)$ by the complementary terms of the trace formula. Langlands' hope (referred to as a pipe dream in [**Lan15**]) is that the resulting limit might ultimately be shown to exist, through an analysis of these complementary terms. The expression for the limit so obtained would then provide a formula for the putative sum (A.3).

In general, it will probably be necessary to work with the stable trace formula, rather than the invariant trace formula. This is quite appropriate, since we are assuming that $G$ is quasisplit. The $t$-discrete part

$$S_{t,\mathrm{disc}}(f) = S^G_{t,\mathrm{disc}}(f)$$

of the stable trace formula, defined in Corollary 29.10, has a decomposition

$$S_{t,\mathrm{disc}}(f) = \sum_{c\in\mathcal{C}_{t,\mathrm{disc}}(G)} S_c(f), \qquad f \in \mathcal{H}(G),$$

into Hecke eigenspaces. The indices $c$ here range over "$t$-discrete" equivalence classes of families

$$c^V = \{c_v:\ v \notin V\}, \qquad V \supset S_{\mathrm{ram}}(G),$$

of semisimple conjugacy classes in ${}^LG$ attached to unramified representations $\pi^V = \pi(c^V)$ of $G(\mathbb{A}^V)$. We recall that two such families are equivalent if they are equal for almost all $v$. The eigendistribution $S_c(f)$ is characterized by the property that

$$S_c(f_\phi) = S_c(f)\phi_G(c^V), \qquad \phi \in \mathcal{H}\big(G(\mathbb{A}^V), K^V\big),$$

where $V$ is a large finite set of valuations depending on $f$, $c^V$ is some representative of the equivalence class $c$, and

$$\phi_G(c^V) = \phi_G\big(\pi(c^V)\big).$$

For any $f$ and $t$, the sum in $c$ can be taken over a finite set. Let $\mathcal{C}_{\mathrm{temp,cusp}}(G)$ be the subset of classes in the union

$$\mathcal{C}_{\mathrm{disc}}(G) = \bigcup_{t\geq 0}\mathcal{C}_{t,\mathrm{disc}}(G)$$

that do not lie in the image of $\mathcal{C}_{\mathrm{disc}}(M)$ in $\mathcal{C}_{\mathrm{disc}}(G)$ for any $M \neq G$, and whose components $c_v$ are bounded in $\widehat{G}$. The sum

$$S_{\mathrm{temp,cusp}}(f) = S^G_{\mathrm{temp,cusp}}(f) = \sum_{c\in\mathcal{C}_{\mathrm{temp,cusp}}(G)} S_c(f), \qquad f\in\mathcal{H}(G),$$

is then easily seen to be absolutely convergent.

The sum (A.3) and the limit $I^r_{\mathrm{cusp,temp}}(f)$ have obvious stable analogues. If the partial $L$-function

$$L(s, c^V, r) = L^V(s,\pi,r), \qquad \pi^V = \pi(c^V),$$

attached to a class $c \in \mathcal{C}_{\mathrm{temp,cusp}}(G)$ has analytic continuation, set $n(c, r)$ equal to $n(\pi, r)$. Then

$$n(c,r) = \lim_{N\to\infty} \Big( V_N^{-1} \sum_{\{v \notin V : q_v \le N\}} \mathrm{tr}\big(r(c_v)\big)\Big) = \lim_{N\to\infty} \big( (\phi_N)_G(c^V) V_N^{-1} \big).$$

The notation here reflects the fact that the limit is independent of both $V$ and the representative $c^V$ of $c$. If $L(s, c^V, r)$ has analytic continuation for every $c$, the sum

$$\sum_{c \in \mathcal{C}_{\mathrm{temp,cusp}}(G)} n(c,r) S_c(f) \tag{A.4}$$

equals the limit

$$S^r_{\mathrm{temp,cusp}}(f) = \lim_{N\to\infty} \big( S_{\mathrm{temp,cusp}}(f_N) V_N^{-1} \big).$$

The remarks for $I^r_{\mathrm{temp,cusp}}(f)$ above apply again to the limit $S^r_{\mathrm{temp,cusp}}(f)$ here. Namely, it might be possible to investigate this limit without knowing the analytic continuation of the $L$-functions. Since $S_{\mathrm{temp,cusp}}(f_N)$ is part of the stable trace formula for $G$, we could replace it by the complementary terms in the formula. The ultimate goal would be to show that the limit exists, and that it has an explicit expression given by these complementary terms.

An important step along the way would be to deal with the complementary terms attached to nontempered classes $c$. These terms represent contributions to $S_{t,\mathrm{disc}}(f_N)$ from nontempered representations of $G(\mathbb{A})$ that occur in the discrete spectrum. The conjectural classification in [**A17**] suggests that they can be expressed in terms of groups $G_\psi$ of dimension smaller than $G$. One can imagine that the total contribution of a group $H = G_\psi$ might take the form of a sum

$$\Big( \sum_{\{\psi : G_\psi = H\}} \widehat{S}^H_\psi(f_N^\psi) \Big) V_N^{-1}, \tag{A.5}$$

where $S^H_\psi$ is a component of the linear form $S^H_{\mathrm{temp,cusp}}$ on $\mathcal{H}(H)$, and $f_N \to f_N^\psi$ is a transform from $\mathcal{H}(G)$ to $S\mathcal{I}(H)$. For example, the one-dimensional automorphic representations $\chi$ of $G(\mathbb{A})$ are represented by parameters

$$\psi : \ \Gamma_F \times SL(2,\mathbb{C}) \longrightarrow {}^L G,$$

in which $\psi\left(1, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right)$ is a principal unipotent element in $\widehat{G}$. In this case, $H = G_\psi$ is the co-center of $G$, and

$$\widehat{S}^H_\psi(f_N^\psi) = \int_{G(\mathbb{A})} f_N(x)\chi(x)\mathrm{d}x.$$

In general, the transform $f_N^\psi$ would be defined by nontempered stable characters, and the contribution (A.5) of $G_\psi$ will not have a limit in $N$. One would have to combine the sum over $H$ of these contributions with the sum obtained from the remaining terms in the stable trace formula. More precisely, one would need to show that the difference of the two sums does have a limit in $N$, for which there is an explicit expression. In the process, one could try to establish the global conjectures in [**A17**], in the more exotic cases where endoscopy gives only partial information.

This is a tall order indeed. The most optimistic prediction might be that the program can be carried out with a great deal of work by many mathematicians over a long period of time! However, the potential rewards seem to justify any amount of effort. A successful resolution to the questions raised so far would be spectacular. It would give a complicated, but presumably quite explicit, formula for the linear form $S^r_{\mathrm{temp,cusp}}(f)$ in terms that are primarily geometric. The result would be a stable trace formula for the tempered, cuspidal automorphic representations $\pi$ of $G$ such that $L(s,\pi,r)$ has a pole at $s=1$.

The lesson we have learned from earlier applications is that a complicated trace formula is more useful when it can be compared with something else. The case at hand should be no different. One could imagine that for any $L$-embedding $\rho$ as in (A.1), there might be a mapping $f \to f^{r,\rho}$ from $\mathcal{H}(G)$ to $S\mathcal{I}(G')$ by which one could detect functorial contributions of $\rho$ to $S^r_{\mathrm{temp,cusp}}(f)$. The mapping might perhaps be defined locally. It should certainly vanish unless $\rho$ is unramified outside of $V$, for any finite set $V$ such that $f$ lies in the subspace $\mathcal{H}(G_V)$ of $\mathcal{H}(G)$. We would include only those $\rho$ that are *elliptic*, in the sense that their image is contained in no proper parabolic subgroup of ${}^LG$.

From the theory of endoscopy, we know that we have to treat a somewhat larger class of embeddings $\rho$. We consider the set of $\widehat{G}$-orbits of elliptic $L$-embeddings

$$\rho:\ \mathcal{G}' \longrightarrow {}^LG, \tag{A.1$^*$}$$

where $\mathcal{G}'$ is an extension

$$1 \longrightarrow \widehat{G}' \longrightarrow \mathcal{G}' \longrightarrow W_F \longrightarrow 1$$

for which there is an $L$-embedding $\mathcal{G}' \hookrightarrow {}^L\widetilde{G}'$. It is assumed that $\widehat{G}'$ is the $L$-group of a quasisplit group $G'$, and that $\widetilde{G}' \to G'$ is a $z$-extension of quasisplit groups. For each such $\rho$, we suppose that there is a mapping $f \to f^{r,\rho}$ from $\mathcal{H}(G)$ to $S\mathcal{I}(\widetilde{G}',\widetilde{\eta}')$, for the appropriate character $\widetilde{\eta}'$ on the kernel of the projection $\widetilde{G}' \to G'$, which vanishes unless $\rho$ is unramified outside of $V$. One might hope ultimately to establish an identity

$$S^r_{\mathrm{temp,cusp}}(f) = \sum_{\rho} \sigma(r,\rho)\widehat{S}'_{\mathrm{prim}}(f^{r,\rho}), \tag{A.6}$$

where $\rho$ ranges over classes of elliptic $L$-embeddings $(\mathrm{A.1})^*$, $\sigma(r,\rho)$ are global coefficients determined by $r$ and $\rho$, and $S'_{\mathrm{prim}}$ is a stable linear form on $\mathcal{H}(\widetilde{G}',\widetilde{\eta}')$ that depends only on $\widetilde{G}'$ and $\widetilde{\eta}'$. In fact, $S'_{\mathrm{prim}}$ should be defined by a stable sum of the tempered, cuspidal, automorphic representations $\pi' \in \Pi_{\mathrm{temp}}(\widetilde{G}',\widetilde{\eta}')$ such that for any finite dimensional representation $r'$ of ${}^L\widetilde{G}'$, the order of the pole of $L(s,\pi',r')$ at $s=1$ equals the multiplicity of the trivial representation of ${}^L\widetilde{G}'$ in $r'$. For each $G'$, one would try to construct a trace formula for $S'_{\mathrm{prim}}$ inductively from the formulas for the analogues for $\widetilde{G}'$ of the linear forms $S^r_{\mathrm{temp,cusp}}$. The goal would be to compare the contribution of these formulas to the right hand side of (A.6) with the formula one hopes to obtain for the left hand side. If one could show that the two primarily geometric expressions cancel, one would obtain an identity (A.6).

A formula (A.6) for any $G$ would presumably lead to the general principle of functoriality. Functoriality in turn implies the analytic continuation of the $L$-functions $L(s,\pi,r)$ (for cuspidal automorphic representations $\pi$) and Ramanujan's conjecture (for those cuspidal automorphic representations $\pi$ not attached to the

$SL(2,\mathbb{C})$-parameters of [**A17**]). Both of these implications were drawn in Langlands' original paper [**Lan3**]. It is interesting to note that Langlands' ideas are based on the intuition gained from the analytic continuation and the Ramanujan conjecture. However, his strategy is to bypass these two conjectures, leaving them to be deduced from the principle of functoriality one hopes eventually to establish.

The existence of a formula (A.6) would actually imply something beyond functoriality. Let $\Pi_{\mathrm{prim}}(G)$ be the set of tempered, cuspidal, automorphic representations of $G$ that are *primitive*, in the sense that they are not functorial images of representations $\pi' \in \Pi(\widetilde{G}', \widetilde{\eta}')$, for any $L$-embedding $(\mathrm{A}.1)^*$ with proper image in ${}^LG$. An identity of the form (A.6) implies that if $\pi \in \Pi_{\mathrm{prim}}(G)$, and $r$ is any finite dimensional representation of ${}^LG$, the order of the pole of $L(s,\pi,r)$ at $s=1$ equals the multiplicity of the trivial representation of ${}^LG$ in $r$. This condition represents a kind of converse to functoriality. It implies that any tempered, cuspidal, automorphic representation $\pi$ of $G$ is a functional image under some $\rho$ of a representation $\pi'$ in the associated set $\Pi_{\mathrm{prim}}(\widetilde{G}', \widetilde{\eta}')$. The condition is closely related to the existence of the automorphic Langlands group $L_F$. If it fails, the strategy for attacking the functoriality we have described would seem also to fail.

All of this is implicit in Langlands' paper [**Lan13**], if I have understood it correctly. Langlands is particularly concerned with the case that $G = PGL(2)$, a group for which the stable trace formula is the same as the invariant trace formula, and $r$ is the irreducible representation of $\widehat{G} = SL(2,\mathbb{C})$ of dimension $(m+1)$. In this case, an elliptic homomorphism $\rho$ will be of dihedral, tetrahedral, octahedral, or icosahedral type. For each of the last three types, the image of $\rho$ is actually finite. The poles that any of these three types would contribute to $L$-functions $L(s,\pi,r)$ are quite sparse. (See [**Lan13**, p. 24].) For example, to detect the contribution of an icosahedral homomorphism $\rho$, one would have to take a 12-dimensional representation $r$. For a representation of $\widehat{G}$ of this size, there will be many terms in the putative limit $I^r_{\mathrm{temp,cusp}}(f) = S^r_{\mathrm{temp,cusp}}(f)$ that overwhelm the expected contribution of $\rho$. The analytic techniques required to rule out such terms are well beyond anything that is presently understood. Techniques that can be applied to smaller representations $r$ are discussed in [**Lan13**] and [**Lan15**], and also in the letter [**Sar**] of Sarnak.

## References


[ABV] J. Adams, D. Barbasch, and D. Vogan, *The Langlands Classification and Irreducible Characters for Real Reductive Groups*, Progr. Math. **104**, Birkhauser, Boston, 1992.

[AV] J. Adams and D. Vogan, *L-groups, projective representations, and the Langlands classification*, Amer. J. Math. **114** (1992), 45–138.

[A1] J. Arthur, *The characters of discrete series as orbital integrals*, Inv. Math. **32** (1976), 205–261.

[A2] ———, *Eisenstein series and the trace formula*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. **33** (1979), Part 1, Amer. Math. Soc., 253–274.

[A3] ———, *A trace formula for reductive groups I: Terms associated to classes in $G(\mathbb{Q})$*, Duke Math. J. **245** (1978), 911–952.

[A4] ———, *A trace formula for reductive groups II: Applications of a truncation operator*, Compositio Math. **40** (1980), 87–121.

[A5] ———, *The trace formula in invariant form*, Ann. of Math. **114** (1981), 1–74.

[A6] ———, *On the inner product of truncated Eisenstein series*, Duke Math. J. **49** (1982), 35–70.

[A7] ———, *On a family of distributions obtained from Eisenstein series I: Application of the Paley-Wiener theorem*, Amer. J. Math. **104** (1982), 1243–1288.

[A8] ———, *On a family of distributions obtained from Eisenstein series II: Explicit formulas*, Amer. J. Math. **104** (1982), 1289–1336.

[A9] ———, *A Paley-Wiener theorem for real reductive groups*, Acta Math. **150** (1983), 1–89.

[A10] ———, *A measure on the unipotent variety*, Canad. J. Math. **37** (1985), 1237–1274.

[A11] ———, *On a family of distributions obtained from orbits*, Canad. J. Math. **38** (1986), 179–214.

[A12] ———, *The local behaviour of weighted orbital integrals*, Duke Math. J. **56** (1988), 223–293.

[A13] ———, *The invariant trace formula I. Local theory*, J. Amer. Math. Soc. **1** (1988), 323–383.

[A14] ———, *The invariant trace formula II. Global theory*, J. Amer. Math. Soc. **1** (1988), 501–554.

[A15] ———, *Intertwining operators and residues I. Weighted characters*, J. Funct. Anal. **84** (1989), 19–84.

[A16] ———, *The $L^2$-Lefschetz numbers of Hecke operators*, Invent. Math. **97** (1989), 257–290.

[A17] ———, *Unipotent automorphic representations: conjectures*, Astérisque **171–172** (1989), 13–71.

[A18] ———, *Some problems in local harmonic analysis*, in *Harmonic Analysis on Reductive Groups*, Birkhauser, 1991, 57–78.

[A19] ———, *A local trace formula*, Publ. Math. Inst. Hautes Études Sci. **73** (1991), 5–96.

[A20] ———, *On elliptic tempered characters*, Acta. Math. **171** (1993), 73–138.

[A21] ———, *On the Fourier transforms of weighted orbital integrals*, J. reine. angew. Math. **452** (1994), 163–217.

[A22] ———, *On local character relations*, Selecta Math. **2**, No. 4 (1996), 501–579.

[A23] ———, *The problem of classifying automorphic representations of classical groups*, CRM Proceedings and Lecture Notes, vol. 11 (1997), 1–12.

[A24] ———, *Canonical normalization of weighted characters and a transfer conjecture*, C.R. Math. Rep. Acad. Sci., Canada **20** (2) (1998), 35–52.

[A25] ———, *On the transfer of distributions: weighted orbital integrals*, Duke Math. J. **99** (1999), 209–283.

[A26] ———, *A stable trace formula II. Global descent*, Invent. Math. **143** (2001), 157–220.

[A27] ———, *A stable trace formula I. General expansions*, Journal of the Inst. of Math. Jussieu **1** (2002), 175–277.

[A28] ———, *The principle of functoriality*, Bull. Amer. Math. Soc. **40** (2002), 39–53.

[A29] ———, *A stable trace formula III. Proof of the main theorems*, Annals of Math. **158** (2003), 769–873.

[A30] ———, *Germ expansions for real groups,* preprint.

[A31] ———, *On the transfer of distributions: singular orbital integrals*, in preparation.

[AC] J. Arthur and L. Clozel, *Simple Algebras, Base Change, and the Advanced Theory of the Trace Formula*, Ann. of Math. Studies **120**, Princeton Univ. Press, Princeton, N.J., 1989.

[Be] J. Bernstein, *$P$-invariant distributions on $GL(n)$*, in *Lie Group Representations* II, Lecture Notes in Math., vol. 1041, Springer, New York, 1984, 50–102.

[BDK] J. Bernstein, P. Deligne and D. Kazhdan, *Trace Paley-Wiener theorem for reductive $p$-adic groups*, J. d'Analyse Math. **47** (1986), 180–192.

[Bor1] A. Borel, *Some finiteness properties of adele groups over number fields*, Publ. Math. Inst. Hautes Études Sci. **16** (1963), 5–30.

[Bor2] ———, *Introduction aux groupes arithmétiques*, Actualitées Sci. Ind. no. 1341, Hermann, Paris, 1969.

[Bor3] ———, *Automorphic $L$-functions*, in *Automorphic Forms, Representations and $L$-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 27–62.

[BC] A. Borel and W. Casselman, *$L^2$-cohomology of locally symmetric manifolds of finite volume*, Duke Math. J. **50** (1983), 625–647.

[BH] A. Borel and Harish-Chandra, *Arithmetic subgroups of algebraic groups*, Ann. of Math. **75** (1962), 485–535.

[BW] A. Borel and N. Wallach, *Continuous Cohomology, Discrete Subgroups, and Representations of Reductive Groups*, Ann. of Math. Studies **94**, Princeton Univ. Press, Princeton, N.J., 1980.

[Bou] A. Bouaziz, *Intégrales orbitals sur les groupes de Lie réductives,* Ann. Scient. Éc. Norm. Sup., 4[e] série, **27** (1994), 573–609.

[BDST] K. Buzzard, M. Dickinson, N. Shepard-Barron, and R. Taylor, *On icosahedral Artin representations*, Duke Math. J. **109** (2001), 283–318.

[Ca] P. Cartier, *Representations of $p$-adic groups: A survey*, in *Automorphic Forms, Representations and $L$-functions*, Proc. Sympos. Pure Math. vol. 33, Part 1, Amer. Math. Soc., 1979, 111–156.

[CD] L. Clozel and P. Delorme, *Le théorème de Paley-Wiener invariant pour les groupes de Lie réductif* II, Ann. Scient. Éc. Norm. Sup., 4[e] série, 23 (1990), 193–228.

[CLL] L. Clozel, J.-P. Labesse, and R.P. Langlands, *Morning Seminar on the Trace Formula*, Lecture Notes, Institute for Advanced Study, Princeton, N.J., 1984.

[Co] J. Cogdell, *Lectures on $L$-functions, Converse Theorems, and Functoriality for $GL(n)$*, in *Lectures in Automorphic $L$-functions*, Fields Institute Monographs, Amer. Math. Soc., 2004, 3–96.

[CKPS1] J. Cogdell, H. Kim, I. Piatetski-Shapiro, and F. Shahidi, *On lifting from classical groups to $GL_N$*, Publ. Math. Inst. Hautes Études Sci. **93** (2001), 5–30.

[CKPS2] ———, *Functoriality for the classical groups*, Publ. Math. Inst. Hautes Études Sci. **99** (2004, 163–233.

[CP] J. Cogdell and I. Piatetski-Shapiro, *Converse theorems for $GL_n$*, Publ. Math. Inst. Hautes Études Sci. **79** (1994), 157–204.

[D1] P. Deligne, *Les constantes des équations fonctionnelles des fonctions $L$*, in *Modular Forms of One Variable* II, Lecture Notes in Math. 349, Springer, New York, 1973, 501–597.

[D2] ———, *Les constantes locales de l'équation fonctionnelle de la fonction $L$ d'Artin d'une représentation orthogonale*, Invent. Math. **35** (1976), 299–316.

[DKV] P. Deligne, D. Kazhdan, and M.-F. Vigneras, *Représentations des algèbres centrales simples $p$-adiques*, in *Représentations des groupes réductifs sur un corps local*, Hermann, Paris, 1984, 33–117.

[DL] M. Duflo and J.-P. Labesse, *Sur la formule des traces de Selberg*, Ann. Scient. Éc. Norm. Sup. **4** (1971), 193–284.

[F] D. Flath, *Decomposition of representations into tensor products*, in *Automorphic Forms, Representations and $L$-functions*, Proc. Sympos. Pure Math. vol. 33, Part 1, Amer. Math. Soc., 1979, 179–184.

[Ge] S. Gelbart, *Lectures on the Arthur-Selberg Trace Formula*, University Lecture Series, Amer. Math. Soc., 1995.

[GeJ] S. Gelbart and H. Jacquet, *A relation between automorphic representations of $GL(2)$ and $GL(3)$*, Ann. Scient. Éc. Normale Sup. **11** (1978), 471–552.

[GS] S. Gelbart and F. Shahidi, *Analytic Properties of Automorphic L-Functions*, Perspectives in Mathematics, vol. 6, Academic Press, 1988.

[GRS] D. Ginzburg, S. Rallis, and D. Soudry, *Generic automorphic forms on $SO(2n+1)$: functorial lift to $GL(2n)$, endoscopy and base change*, Internat. Math. Res. Notices **14** (2001), 729–764.

[GoJ] R. Godement and H. Jacquet, *Zeta-functions of Simple Algebras*, Lecture Notes in Math., vol. 260, Springer, New York, 1972.

[GKM] M. Goresky, R. Kottwitz, and R. MacPherson, *Discrete series characters and the Lefschetz formula for Hecke operators*, Duke Math. J. **89** (1997), 477–554.

[Hal1] T. Hales, *A simple definition of transfer factors for unramified groups*, Contemp. Math. **145** (1993), 109–134.

[Hal2] ———, *A statement of the fundamental lemma*, in this volume.

[Har1] Harish-Chandra, *A formula for semisimple Lie groups*, Amer. J. Math. **79**, 733–760.

[Har2] ———, *Invariant eigendistributions on a semisimple Lie group*, Trans. Amer. Math. Soc. **119**, 457–508.

[Har3] ———, *Discrete series for semisimple Lie groups.* II. *Explicit determination of the characters*, Acta Math. **116**, 1–111.

[Har4] ———, *Automorphic Forms on Semisimple Lie Groups*, Lecture Notes in Math., vol. 62, Springer, New York, 1968.

[Har5] ———, *Harmonic analysis on real reductive groups.* I. *The theory of the constant term*, J. Funct. Anal. **19**, 104–204.

[HT] M. Harris and R. Taylor, *On the Geometry and Cohomology of Some Simple Shimura Varieties*, Ann. of Math. Studies **151**, Princeton Univ. Press, Princeton, N.J., 2001.

[Has] H. Hasse, *History of class field theory*, in *Algebraic Number Theory*, Thompson, Washington, D.C., 1967, 266–279.

[He] G. Henniart, *Une preuve simple des conjectures de Langlands de $GL(n)$ sur un corps $p$-adique*, Invent. Math. **139** (2000), 439–455.

[J] H. Jacquet, *Residual spectrum, especially for $GL(n)$*, in *Lie Group Representations* II, Lecture Notes in Math., vol. 1041, Springer, New York, 1984, 185–208.

[JL] H. Jacquet and R. Langlands, *Automorphic Forms on $GL(2)$*, Lecture Notes in Math., vol. 114, Springer, New York, 1970.

[JPS] H. Jacquet, I. Piatetski-Shapiro, and J. Shalika, *Rankin-Selberg convolutions*, Amer. J. Math. **105** (1983), 367–464.

[JaS] H. Jacquet and J. Shalika, *On Euler products and the classification of automorphic representations* II, Amer. J. Math. **103** (1981), 777–815.

[JiS] D. Jiang and D. Soudry, *The local converse theorem for $SO(2n+1)$ and applications*, Ann. of Math. **157** (2003), 743–806.

[Ka1] D. Kazhdan, *On lifting*, in *Lie Group Representations* II, Lecture Notes in Math., vol. 1041, Springer, New York, 1984, 209–249.

[Ka2] ———, *Cuspidal geometry on $p$-adic groups* (Appendix), J. Analyse Math. **47** (1980), 1–36.

[Ki] H. Kim, *Functoriality for the exterior square of $GL_4$ and symmetric fourth power of $GL_2$*, J. of Amer. Math. Soc. **16**, 139–183.

[KiS] H. Kim and F. Shahidi, *Functorial products for $GL_2 \times GL_3$ and symmetric cube for $GL_2$*, Ann. of Math. **155** (2002), 837–893.

[Kna] A. Knapp, *Commutativity of intertwining operators.* II, Bull. Amer. Math. Soc. **82** (1976), 271–273.

[KnS] A. Knapp and E. Stein, *Intertwining operators for semisimple groups* II, Invent. Math. **60** (1980), 9–84.

[KZ1] A. Knapp and G. Zuckerman, *Normalizing factors and L-groups*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 1, Amer. Math. Soc., 1979, 93–106.

[KZ2] ———, *Classification of irreducible tempered representations of semisimple groups*, Ann. of Math. **116** (1982), 389–455.

[Kne] M. Kneser, *Strong approximation*, in *Algebraic Groups and Discontinuous Subgroups*, Proc. Sympos. Pure Math., vol. 9, Amer. Math. Soc., 1966, 187–196.

[Ko1] R. Kottwitz, *Rational conjugacy classes in reductive groups*, Duke Math. J. **49** (1982), 785–806.

[Ko2] ———, *Sign changes in harmonic analysis on reductive groups*, Trans. Amer. Math. Soc. **278** (1983), 289–297.

[Ko3] ———, *Stable trace formula: cuspidal tempered terms*, Duke Math. J. **51** (1984), 611–650.

[Ko4] ———, *Base change for unit elements of Hecke algebras*, Compositio Math. **60** (1986), 237–250.

[Ko5] ———, *Stable trace formula: elliptic singular terms*, Math. Ann. **275** (1986), 365–399.

[Ko6] ———, *Tamagawa numbers,* Ann. of Math. **127** (1988), 629–646.

[Ko7] ———, *Points on some Shimura varieties over finite fields*, J. Amer. Math. Soc. **5** (1992), 373–444.

[Ko8] ———, *Harmonic analysis on reductive $p$-adic groups and Lie algebras*, in this volume.

[KR] R. Kottwitz and J. Rogawski, *The distributions in the invariant trace formula are supported on characters*, Canad. J. Math. **52** (2000), 804–814.

[KoS] R. Kottwitz and D. Shelstad, *Foundations of Twisted Endoscopy*, Astérisque, **255** (1999).

[Lab1] J.-P. Labesse, *$L$-indistinguishable representations and the trace formula for $SL(2)$*, in *Lie Groups and their Representations*, John Wiley, 1975.

[Lab2] ———, *La formule des traces d'Arthur-Selberg*, Seminaire Bourbaki, exposé 636, Astérisque **133–134** (1986), 73–88.

[LL] J.-P. Labesse and R. Langlands, *$L$-indistinguishability for $SL(2)$*, Canad. J. Math. **31** (1979), 726–785.

[Lan1] R. Langlands, *Eisenstein series*, in *Algebraic Groups and Discontinuous Subgroups*, Proc. Sympos. Pure Math., vol. 9, Amer. Math. Soc., 1966, 235–252.

[Lan2] ———, *Dimension of spaces of automorphic forms*, in *Algebraic Groups and Discontinuous Subgroups*, Proc. Sympos. Pure Math., vol. 9, Amer. Math. Soc., 1966, 253–257.

[Lan3] ———, *Problems in the theory of automorphic forms*, in *Lectures in Modern Analysis and Applications,* Lecture Notes in Math. 170, Springer, New York, 1970, 18–61.

[Lan4] ———, *Modular forms and $\ell$-adic representations*, in *Modular Forms of One Variable* II, Lecture Notes in Math. 349, Springer, New York, 1973, 361–500.

[Lan5] ———, *On the Functional Equations Satisfied by Eisenstein Series*, Lecture Notes in Math. 544, Springer, New York, 1976.

[Lan6] ———, *On the notion of an automorphic representation. A supplement to the preceding paper*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 1, Amer. Math. Soc., 1979, 203–208.

[Lan7] ———, *Automorphic representations, Shimura varieties, and motives. Ein Märchen*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 205–246.

[Lan8] ———, *Stable conjugacy: definitions and lemmas*, Canad. J. Math. **31** (1979), 700–725.

[Lan9] ———, *Base Change for $GL(2)$*, Ann. of Math. Studies 96, Princeton Univ. Press, Princeton, N.J., 1980.

[Lan10] ———, *Les débuts d'une formule des traces stables*, Publ. Math. Univ. Paris VII **13**, 1983.

[Lan11] ———, *On the classification of irreducible representations of real algebraic groups*, in *Representation Theory and Harmonic Analysis on Semisimple Lie Groups*, AMS Mathematical Surveys and Monographs, vol. 31, 1989, 101–170.

[Lan12] ———, *Representations of abelian algebraic groups*, Pacific J. Math. (1997), 231–250.

[Lan13] ———, *Endoscopy and beyond*, notes from a lecture, Institute for Advanced Study, Princeton, N.J., 2000.

[Lan14] ———, *The trace formula and its applications: An introduction to the work of James Arthur*, Canad. Math. Bull. **44** (2001), 160–209.

[Lan15] ———, *Beyond endoscopy*, in *Contributions to Automorphic Forms, Geometry, and Number Theory*, Johns Hopkins University Press, 2004, 611–698.

[LS1] R. Langlands and D. Shelstad, *On the definition of transfer factors*, Math. Ann. **278** (1987), 219–271.

[LS2] ———, *Descent for transfer factors, The Grothendieck Festschift*, Vol. II, Birkhauser, Boston, 1990, 485–563.

[LR] E. Lapid and J. Rogawski, *On twists of cuspidal representations of $GL(2)$*, Formum Mathematicum **10** (1998), 175–197.

[Lar] M. Larsen, *On the conjugacy of element-conjugate homomorphisms*, Israel J. Math. **88** (1994), 253–277.

[LN] G. Laumon and B.C. Ngo, *Le lemme fondamental pour les groupes unitaires*, preprint, Université Paris 13, 2004.

[Lo] E. Looijenga, *$L^2$-cohomology of locally symmetric varieties*, Compositio Math. **67** (1988), 3–20.

[Ma] G. Mackey, *Group representations and applications*, lecture notes from lectures at Oxford University, 1966–1967.

[Mi] J. Milne, *Introduction to Shimura varieties*, in this volume.

[MW1] C. Moeglin and J.-L. Waldspurger, *Le spectre résiduel de $GL(n)$*, Ann. Scient. Éc. Norm. Sup. 4[e] série **22** (1989), 605–674.

[MW2] ———, *Spectral Decomposition and Eisenstein Series*, Cambridge Tracts in Mathematics, vol. 113, Cambridge University Press, 1995.

[MT] G.D. Mostow and T. Tamagawa, *On the compactness of arithmetically defined homogeneous spaces*, Ann. of Math. **76** (1962), 446–463.

[Mul] W. Müller, *The trace class conjecture in the theory of automorphic forms*, Ann. of Math. **130** (1989), 473–529.

[MS] W. Müller and B. Speh, *On the absolute convergence of the spectral side of the Arthur trace formula for $GL_n$*, preprint.

[Mur] F. Murnaghan, *Linear algebraic groups*, in this volume.

[P] G. Prasad, *Strong approximation for semi-simple groups over function fields*, Ann. of Math. **105** (1977), 553–572.

[R] R. Rao, *Orbital integrals on reductive groups*, Ann. of Math. **96** (1972), 505–510.

[Ro1] J. Rogawski, *The trace Paley-Wiener theorem in the twisted case*, Trans. Amer. Math. Soc. **309** (1988), 215–229.

[Ro2] ———, *Automorphic Representations of Unitary Groups in Three Variables*, Ann. of Math. Studies **123**, Princeton Univ. Press, Princeton, N.J., 1990.

[Ro3] ———, *The multiplicity formula for A-packets*, in *The Zeta Functions of Picard Modular Surfaces*, Les Publications CRM, Montreal, 1992, 395–420.

[Sai] H. Saito, *Automorphic forms and algebraic extensions of number fields*, Lectures in Math., Kyoto Univ., 1975.

[Sar] P. Sarnak, *Comments on Langland's lecture: "Endoscopy and Beyond"*, http://www.math.princeton.edu//menua/index4.html.

[SS] L. Saper and M. Stern, *$L^2$-cohomology of arithmetic varieties*, Ann. of Math. **132** (1990), 1–69.

[Sel1] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. **20** (1956), 47–87.

[Sel2] ———, *Discontinuous groups and harmonic analysis*, Proc. Int. Cong. Math. 1962, 177–189.

[Ser1] J.-P. Serre, *Abelian $\ell$-adic Representations and Elliptic Curves,* Benjamin, New York, 1968.

[Ser2] ———, *Complex Semisimple Lie Algebras*, Springer, New York, 1987.

[Sha1] F. Shahidi, *On certain L-functions*, Amer. J. Math. **103** (1981), 297–355.

[Sha2] ———, *Fourier transforms of intertwining operators and Plancherel measures for $GL(n)$*, Amer. J. Math. **106** (1984), 67–111.

[She1] D. Shelstad, *Notes on L-indistinguishability (based on a lecture by R.P. Langlands)*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 193–204.

[She2] ———, *Embeddings of L-groups*, Canad. J. Math. **33** (1981), 513–558.

[She3] ———, *L-indistinguishability for real groups*, Math. Ann. **259** (1982), 385–430.

[Shim] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions*, Publ. Math. Soc. Japan **11**, 1971.

[Shin] T. Shintani, *On liftings of holomorphic cusp forms*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 97–110.

[Sho] S. Shokranian, *The Selberg-Arthur Trace Formula*, Lecture Notes in Math. 1503, Springer, New York, 1992.

[Spr1] T. Springer, *Galois cohomology of linear algebraic groups*, in *Algebraic Groups and Discontinuous Subgroups*, Proc. Sympos. Pure Math., vol. 9, Amer. Math. Soc., 1966, 149–158.

[Spr2] ———, *Reductive groups*, in Automorphic Forms, Representations and $L$-functions, Proc. Sympos. Pure Math. vol. 33, Part 1, Amer. Math. Soc., 1979, 3–28.

[Ste] R. Steinberg, *Regular elements of semisimple algebraic groups*, Publ. Math. Inst. Hautes Études Sci. **25** (1965), 49–80.

[Tam1] T. Tamagawa, *On Selberg's trace formula*, J. Fac. Sci. Univ. Tokyo, Sect I.**8** (1960), 363–386.

[Tam2] ———, *Adèles*, in *Algebraic Groups and Discontinuous Subgroups*, Proc. Sympos. Pure Math., vol. 9, Amer. Math. Soc., 1966, 113–121.

[Tat1] J. Tate, *Global class field theory*, in *Algebraic Number Theory*, Thompson, Washington, D.C., 1967, 163–203.

[Tat2] ———, *Number theoretic background*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 2, Amer. Math. Soc., 1979, 3–26.

[Ti] J. Tits, *Reductive groups over local fields*, in *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math. vol. 33, Part 1, Amer. Math. Soc., 1979, 29–70.

[Tu] J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. **5** (1981), 173–175.

[Vi] M.-F. Vigneras, *On the global correspondence between $GL(n)$ and division algebras*, Lecture Notes, Institute for Advanced Study, 1984.

[VZ] D. Vogan and G. Zuckerman, *Unitary representations with nonzero cohomology*, Compositio Math. **53** (1984), 51–90.

[Wa1] J.-L. Waldspurger, *Une formule des traces locale pour les algèbres de Lie $p$-adique*, J. reine angew. Math. **465** (1995), 41–99.

[Wa2] ———, *Le lemme fondamental implique le transfer*, Compositio Math. **105** (1997), 153–236.

[We] A. Weil, *Basic Number Theory*, Springer-Verlag, New York, 1967.

[Wi] F. Williams, *Discrete series multiplicities in $L^2(\Gamma\backslash G)$*, Amer. J. Math. **106** (1984), 137–148.



Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3

*E-mail address*: `arthur@math.toronto.edu`

Clay Mathematics Proceedings
Volume 4, 2005

# Introduction to Shimura Varieties

J.S. Milne

ABSTRACT. This is an introduction to the theory of Shimura varieties, or, in other words, to the arithmetic theory of automorphic functions and holomorphic automorphic forms.

CONTENTS



## Introduction

The arithmetic properties of elliptic modular functions and forms were extensively studied in the 1800s, culminating in the beautiful Kronecker Jugendtraum. Hilbert emphasized the importance of extending this theory to functions of several variables in the twelfth of his famous problems at the International Congress in 1900. The first tentative steps in this direction were taken by Hilbert himself and his students Blumenthal and Hecke in their study of what are now called Hilbert (or Hilbert-Blumenthal) modular varieties. As the theory of complex functions of several variables matured, other quotients of bounded symmetric domains by arithmetic groups were studied (Siegel, Braun, and others). However, the modern theory of Shimura varieties[1] only really began with the development of the theory of abelian varieties with complex multiplication by Shimura, Taniyama, and Weil in the mid-1950s, and with the subsequent proof by Shimura of the existence of canonical models for certain families of Shimura varieties. In two fundamental articles, Deligne recast the theory in the language of abstract reductive groups and extended Shimura's results on canonical models. Langlands made Shimura varieties a central part of his program, both as a source of representations of galois groups and as tests for the conjecture that all motivic $L$-functions are automorphic. These notes are an introduction to the theory of Shimura varieties from the point of view of Deligne and Langlands. Because of their brevity, many proofs have been omitted or only sketched.

*Notations and conventions.* Unless indicated otherwise, vector spaces are assumed to be finite dimensional and free $\mathbb{Z}$-modules are assumed to be of finite rank. The linear dual $\mathrm{Hom}(V, k)$ of a vector space (or module) $V$ is denoted $V^\vee$. For a $k$-vector space $V$ and a $k$-algebra $R$, $V(R)$ denotes $R \otimes_k V$ (and similarly for $\mathbb{Z}$-modules). By a lattice in an $\mathbb{R}$-vector space $V$, I mean a full lattice, i.e., a $\mathbb{Z}$-submodule generated by a basis for $V$. The algebraic closure of a field $k$ is denoted $k^{\mathrm{al}}$.

A superscript $^+$ (resp. $^\circ$) denotes a connected component relative to a real topology (resp. a zariski topology). For an algebraic group, we take the identity connected component. For example, $(O_n)^\circ = \mathrm{SO}_n$, $(\mathrm{GL}_n)^\circ = \mathrm{GL}_n$, and $\mathrm{GL}_n(\mathbb{R})^+$ consists of the $n \times n$ matrices with $\det > 0$. For an algebraic group $G$ over $\mathbb{Q}$, $G(\mathbb{Q})^+ = G(\mathbb{Q}) \cap G(\mathbb{R})^+$. Following Bourbaki, I require compact topological spaces to be separated.

Semisimple and reductive groups, whether algebraic or Lie, are required to be connected. A simple algebraic or Lie group is a semisimple group with no connected proper normal subgroups other than 1 (some authors say almost-simple). For a torus $T$, $X^*(T)$ denotes the character group of $T$. The inner automorphism defined by an element $g$ is denoted $\mathrm{ad}(g)$. The derived group of a reductive group $G$ is denoted $G^{\mathrm{der}}$ (it is a semisimple group). For more notations concerning reductive groups, see p303. For a finite extension of fields $L \supset F$ of characteristic zero, the torus over $F$ obtained by restriction of scalars from $\mathbb{G}_m$ over $L$ is denoted $(\mathbb{G}_m)_{L/F}$.

Throughout, I use the notations standard in algebraic geometry, which sometimes conflict with those used in other areas. For example, if $G$ and $G'$ are algebraic groups over a field $k$, then a homomorphism $G \to G'$ means a homomorphism defined over $k$; if $K$ is a field containing $k$, then $G_K$ is the algebraic group over $K$ obtained by extension of the base field and $G(K)$ is the group of points of $G$ with coordinates in $K$. If $\sigma\colon k \hookrightarrow K$ is a homomorphism of fields and $V$ is an algebraic variety (or other algebro-geometric object) over $k$, then $\sigma V$ has its only possible meaning: apply $\sigma$ to the coefficients of the equations defining $V$.

Let $A$ and $B$ be sets and let $\sim$ be an equivalence relation on $A$. If there exists a canonical surjection $A \to B$ whose fibres are the equivalence classes, then I say

[1] The term "Shimura variety" was introduced by Langlands (1976, 1977), although earlier "Shimura curve" had been used for the varieties of dimension one (Ihara 1968).

that $B$ classifies the elements of $A$ modulo $\sim$ or that it classifies the $\sim$-classes of elements of $A$.

A functor $F\colon \mathsf{A} \to \mathsf{B}$ is fully faithful if the maps $\mathrm{Hom}_{\mathsf{A}}(a, a') \to \mathrm{Hom}_{\mathsf{B}}(Fa, Fa')$ are bijective. The essential image of such a functor is the full subcategory of $\mathsf{B}$ whose objects are isomorphic to an object of the form $Fa$. Thus, a fully faithful functor $F\colon \mathsf{A} \to \mathsf{B}$ is an equivalence if and only if its essential image is $\mathsf{B}$ (Mac Lane 1998, p93).

*References.* In addition to those listed at the end, I refer to the following of my course notes (available at www.jmilne.org/math/).

**AG:** Algebraic Geometry, v5.0, February 20, 2005.
**ANT:** Algebraic Number Theory, v2.1, August 31, 1998.
**CFT:** Class Field Theory, v3.1, May 6, 1997.
**FT:** Fields and galois Theory, v4.0, February 19, 2005.
**MF:** Modular Functions and Modular Forms, v1.1, May 22, 1997.

*Prerequisites.* Beyond the mathematics that students usually acquire by the end of their first year of graduate work (a little complex analysis, topology, algebra, differential geometry,...), I assume familiarity with some algebraic number theory, algebraic geometry, algebraic groups, and elliptic modular curves.

*Acknowledgements.* I thank the Clay Mathematical Institute and the organizers for giving me the opportunity to lecture on Shimura varieties, the Fields Institute for providing a excellent setting for the Summer School, and the audience for its lively participation. Also, I thank Lizhen Ji and Gopal Prasad for their help with references, and F. Hörmann and others for alerting me to errors in earlier versions.

## 1. Hermitian symmetric domains

In this section, I describe the complex manifolds that play the role in higher dimensions of the complex upper half plane, or, equivalently, the open unit disk:

$$\{z \in \mathbb{C} \mid \Im(z) > 0\} = \mathcal{H}_1 \underset{-i\frac{z+1}{z-1} \leftarrow z}{\overset{z \to \frac{z-i}{z+i}}{\rightleftarrows}} \mathcal{D}_1 = \{z \in \mathbb{C} \mid |z| < 1\}.$$

This is a large topic, and I can do little more than list the definitions and results that we shall need.

**Brief review of real manifolds.** A ***manifold*** $M$ ***of dimension*** $n$ is a separated topological space that is locally isomorphic to an open subset of $\mathbb{R}^n$ and admits a countable basis of open subsets. A homeomorphism from an open subset of $M$ onto an open subset of $\mathbb{R}^n$ is called a ***chart*** of $M$.

*Smooth manifolds.* I use smooth to mean $C^\infty$. A ***smooth manifold*** is a manifold $M$ endowed with a ***smooth structure***, i.e., a sheaf $\mathcal{O}_M$ of $\mathbb{R}$-valued functions such that $(M, \mathcal{O}_M)$ is locally isomorphic to $\mathbb{R}^n$ endowed with its sheaf of smooth functions. For an open $U \subset M$, the $f \in \mathcal{O}_M(U)$ are called the ***smooth functions*** on $U$. A smooth structure on a manifold $M$ can be defined by a family $u_\alpha\colon U_\alpha \to \mathbb{R}^n$ of charts such that $M = \bigcup U_\alpha$ and the maps

$$u_\alpha \circ u_\beta^{-1}\colon u_\beta(U_\alpha \cap U_\beta) \to u_\alpha(U_\alpha \cap U_\beta)$$

are smooth for all $\alpha, \beta$. A continuous map $\alpha\colon M \to N$ of smooth manifolds is ***smooth*** if it is a map of ringed spaces, i.e., $f$ smooth on an open $V \subset N$ implies $f \circ \alpha$ smooth on $\alpha^{-1}(V)$.

Let $(M, \mathcal{O}_M)$ be a smooth manifold, and let $\mathcal{O}_{M,p}$ be the ring of germs of smooth functions at $p$. The ***tangent space*** $T_pM$ to $M$ at $p$ is the $\mathbb{R}$-vector space of $\mathbb{R}$-derivations

$X_p\colon \mathcal{O}_{M,p}\to\mathbb{R}$. If $x^1,\ldots,x^n$ are local coordinates at $p$, then $\frac{\partial}{\partial x^1},\ldots,\frac{\partial}{\partial x^n}$ is a basis for $T_pM$ and $dx^1,\ldots,dx^n$ is the dual basis.

Let $U$ be an open subset of a smooth manifold $M$. A ***smooth vector field*** $X$ on $U$ is a family of tangent vectors $X_p\in T_p(M)$ indexed by $p\in U$, such that, for any smooth function $f$ on an open subset of $U$, $p\mapsto X_pf$ is smooth. A ***smooth $r$-tensor field*** on $U$ is a family $t=(t_p)_{p\in M}$ of multilinear mappings $t_p\colon T_pM\times\cdots\times T_pM\to\mathbb{R}$ ($r$ copies of $T_pM$) such that, for any smooth vector fields $X_1,\ldots,X_r$ on an open subset of $U$, $p\mapsto t_p(X_1,\ldots,X_r)$ is a smooth function. A ***smooth $(r,s)$-tensor field*** is a family $t_p\colon (T_pM)^r\times(T_pM)^{\vee s}\to\mathbb{R}$ satisfying a similar condition. Note that to give a smooth $(1,1)$-field amounts to giving a family of endomorphisms $t_p\colon T_pM\to T_pM$ with the property that $p\mapsto t_p(X_p)$ is a smooth vector field for any smooth vector field $X$.

A ***riemannian manifold*** is a smooth manifold endowed with a ***riemannian metric***, i.e., a smooth 2-tensor field $g$ such that, for all $p\in M$, $g_p$ is symmetric and positive definite. In terms of local coordinates $x^1,\ldots,x^n$ at $p$,

$$g_p=\textstyle\sum g_{i,j}(p)dx^i\otimes dx^j,\text{ i.e., } g_p\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right)=g_{ij}(p).$$

A morphism of riemannian manifolds is called an ***isometry***.

A ***real Lie group***[2] $G$ is a smooth manifold endowed with a group structure defined by smooth maps $g_1,g_2\mapsto g_1g_2$, $g\mapsto g^{-1}$.

**Brief review of hermitian forms.** To give a complex vector space amounts to giving a real vector space $V$ together with an endomorphism $J\colon V\to V$ such that $J^2=-1$. A ***hermitian form*** on $(V,J)$ is an $\mathbb{R}$-bilinear mapping $(\,|\,)\colon V\times V\to\mathbb{C}$ such that $(Ju|v)=i(u|v)$ and $(v|u)=\overline{(u|v)}$. When we write

$$(u|v)=\varphi(u,v)-i\psi(u,v),\quad \varphi(u,v),\ \psi(u,v)\in\mathbb{R},\tag{1}$$

then $\varphi$ and $\psi$ are $\mathbb{R}$-bilinear, and

$$\varphi\text{ is symmetric}\qquad \varphi(Ju,Jv)=\varphi(u,v),\tag{2}$$

$$\psi\text{ is alternating}\qquad \psi(Ju,Jv)=\psi(u,v),\tag{3}$$

$$\psi(u,v)=-\varphi(u,Jv),\qquad \varphi(u,v)=\psi(u,Jv).\tag{4}$$

As $(u|u)=\varphi(u,u)$, $(\,|\,)$ is positive definite if and only if $\varphi$ is positive definite. Conversely, if $\varphi$ satisfies (2) (resp. $\psi$ satisfies (3)), then the formulas (4) and (1) define a hermitian form:

$$(u|v)=\varphi(u,v)+i\varphi(u,Jv)\quad(\text{resp. }(u|v)=\psi(u,Jv)-i\psi(u,v))\tag{5}$$

**Complex manifolds.** A $\mathbb{C}$-valued function on an open subset $U$ of $\mathbb{C}^n$ is ***analytic*** if it admits a power series expansion in a neighbourhod of each point of $U$. A ***complex manifold*** is a manifold $M$ endowed with a ***complex structure***, i.e., a sheaf $\mathcal{O}_M$ of $\mathbb{C}$-valued functions such that $(M,\mathcal{O}_M)$ is locally isomorphic to $\mathbb{C}^n$ with its sheaf of analytic functions. A complex structure on a manifold $M$ can be defined by a family $u_\alpha\colon U_\alpha\to\mathbb{C}^n$ of charts such that $M=\bigcup U_\alpha$ and the maps $u_\alpha\circ u_\beta^{-1}$ are analytic for all $\alpha,\beta$. Such a family also makes $M$ into a smooth manifold denoted $M^\infty$. A continuous map $\alpha\colon M\to N$ of complex manifolds is ***analytic*** if it is a map of ringed spaces. A ***riemann surface*** is a one-dimensional complex manifold.

A ***tangent vector*** at a point $p$ of a complex manifold is a $\mathbb{C}$-derivation $\mathcal{O}_{M,p}\to\mathbb{C}$. The tangent spaces $T_pM$ ($M$ as a complex manifold) and $T_pM^\infty$ ($M$ as a smooth manifold) can be identified. Explicitly, complex local coordinates $z^1,\ldots,z^n$ at a point $p$ of $M$ define real local coordinates $x^1,\ldots,x^n,y^1,\ldots,y^n$ with $z^r=x^r+iy^r$.

[2]According to a theorem of Lie, this is equivalent to the usual definition in which "smooth" is replaced by "real-analytic".

The real and complex tangent spaces have bases $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}, \frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^n}$ and $\frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^n}$ respectively. Under the natural identification of the two spaces, $\frac{\partial}{\partial z^r} = \frac{1}{2}\left(\frac{\partial}{\partial x^r} - i\frac{\partial}{\partial y^r}\right)$.

A $\mathbb{C}$-valued function $f$ on an open subset $U$ of $\mathbb{C}^n$ is ***holomorphic*** if it is holomorphic (i.e., differentiable) separately in each variable. As in the one-variable case, $f$ is holomorphic if and only if it is analytic (Hartog's theorem, Taylor 2002, 2.2.3), and so we can use the terms interchangeably.

Recall that a $\mathbb{C}$-valued function $f$ on $U \subset \mathbb{C}$ is holomorphic if and only if it is smooth (as a function of two real variables) and satisfies the Cauchy-Riemann condition. This condition has a geometric interpretation: it requires that $df_p \colon T_pU \to T_{f(p)}\mathbb{C}$ be $\mathbb{C}$-linear for all $p \in U$. It follows that a smooth $\mathbb{C}$-valued function $f$ on $U \subset \mathbb{C}^n$ is holomorphic if and only if the maps $df_p \colon T_pU \to T_{f(p)}\mathbb{C}$ are $\mathbb{C}$-linear for all $p \in U$.

An ***almost-complex structure*** on a smooth manifold $M$ is a smooth tensor field $(J_p)_{p\in M}$, $J_p \colon T_pM \to T_pM$, such that $J_p^2 = -1$ for all $p$, i.e., it is a smoothly varying family of complex structures on the tangent spaces. A complex structure on a smooth manifold endows it with an almost-complex structure. In terms of complex local coordinates $z^1, \ldots, z^n$ in a neighbourhood of a point $p$ on a complex manifold and the corresponding real local coordinates $x^1, \ldots, y^n$, $J_p$ acts by

$$\frac{\partial}{\partial x^r} \mapsto \frac{\partial}{\partial y^r}, \quad \frac{\partial}{\partial y^r} \mapsto -\frac{\partial}{\partial x^r}. \tag{6}$$

It follows from the last paragraph that the functor from complex manifolds to almost-complex manifolds is fully faithful: a smooth map $\alpha \colon M \to N$ of complex manifolds is holomorphic (analytic) if the maps $d\alpha_p \colon T_pM \to T_{\alpha(p)}N$ are $\mathbb{C}$-linear for all $p \in M$. Not every almost-complex structure on a smooth manifold arises from a complex structure — those that do are said to be ***integrable***. An almost-complex structure $J$ on a smooth manifold is integrable if $M$ can be covered by charts on which $J$ takes the form (6) (because this condition forces the transition maps to be holomorphic).

A ***hermitian metric*** on a complex (or almost-complex) manifold $M$ is a riemannian metric $g$ such that

$$g(JX, JY) = g(X, Y) \text{ for all vector fields } X, Y. \tag{7}$$

According to (5), for each $p \in M$, $g_p$ is the real part of a unique hermitian form $h_p$ on $T_pM$, which explains the name. A ***hermitian manifold*** $(M, g)$ is a complex manifold with a hermitian metric, or, in other words, it is a riemannian manifold with a complex structure such that $J$ acts by isometries.

**Hermitian symmetric spaces.** A manifold (riemannian, hermitian, ...) is said to be ***homogeneous*** if its automorphism group acts transitively. It is ***symmetric*** if, in addition, at some point $p$ there is an involution $s_p$ (the ***symmetry at*** $p$) having $p$ as an isolated fixed point. This means that $s_p$ is an automorphism such that $s_p^2 = 1$ and that $p$ is the only fixed point of $s_p$ in some neighbourhood of $p$.

For a riemannian manifold $(M, g)$, the automorphism group is the group $\mathrm{Is}(M, g)$ of isometries. A connected symmetric riemannian manifold is called a ***symmetric space***. For example, $\mathbb{R}^n$ with the standard metric $g_p = \sum dx^i dx^i$ is a symmetric space — the translations are isometries, and $\mathbf{x} \mapsto -\mathbf{x}$ is a symmetry at 0.

For a hermitian manifold $(M, g)$, the automorphism group is the group $\mathrm{Is}(M, g)$ of holomorphic isometries:

$$\mathrm{Is}(M, g) = \mathrm{Is}(M^{\infty}, g) \cap \mathrm{Hol}(M) \tag{8}$$

(intersection inside $\mathrm{Aut}(M^{\infty})$; $\mathrm{Hol}(M)$ is the group of automorphisms of $M$ as a complex manifold). A connected symmetric hermitian manifold is called a ***hermitian symmetric space***.

EXAMPLE 1.1. (a) The complex upper half plane $\mathcal{H}_1$ becomes a hermitian symmetric space when endowed with the metric $\frac{dxdy}{y^2}$. The action

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}), \quad z \in \mathcal{H}_1,$$

identifies $\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ with the group of holomorphic automorphisms of $\mathcal{H}_1$. For any $x+iy \in \mathcal{H}_1$, $x+iy = \left(\begin{smallmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{smallmatrix}\right) i$, and so $\mathcal{H}_1$ is homogeneous. The isomorphism $z \mapsto -1/z$ is a symmetry at $i \in \mathcal{H}_1$, and the riemannian metric $\frac{dxdy}{y^2}$ is invariant under the action of $\mathrm{SL}_2(\mathbb{R})$ and has the hermitian property (7).

(b) The projective line $\mathbb{P}^1(\mathbb{C})$ (= riemann sphere) becomes a hermitian symmetric space when endowed with the restriction (to the sphere) of the standard metric on $\mathbb{R}^3$. The group of rotations is transitive, and reflection along a geodesic (great circle) through a point is a symmetry. Both of these transformations leave the metric invariant.

(c) Any quotient $\mathbb{C}/\Lambda$ of $\mathbb{C}$ by a discrete additive subgroup $\Lambda$ becomes a hermitian symmetric space when endowed with the standard metric. The group of translations is transitive, and $z \mapsto -z$ is a symmetry at 0.

*Curvature.* Recall that, for a plane curve, the curvature at a point $p$ is $1/r$ where $r$ is the radius of the circle that best approximates the curve at $p$. For a surface in 3-space, the principal curvatures at a point $p$ are the maximum and minimum of the signed curvatures of the curves obtained by cutting the surface with planes through a normal at $p$ (the sign is positive or negative according as the curve bends towards the normal or away). Although the principal curvatures depend on the embedding of the surface into $\mathbb{R}^3$, their product, the ***sectional curvature*** at $p$, does not (Gauss's Theorema Egregium) and so it is well-defined for any two-dimensional riemannian manifold. More generally, for a point $p$ on any riemannian manifold $M$, one can define the ***sectional curvature*** $K(p, E)$ of the submanifold cut out by the geodesics tangent to a two-dimensional subspace $E$ of $T_pM$. Intuitively, positive curvature means that the geodesics through a point converge, and negative curvature means that they diverge. The geodesics in the upper half plane are the half-lines and semicircles orthogonal to the real axis. Clearly, they diverge — in fact, this is Poincaré's famous model of noneuclidean geometry in which there are infinitely many "lines" through a point parallel to any fixed "line" not containing it. More prosaically, one can compute that the sectional curvature is $-1$. The Gauss curvature of $\mathbb{P}^1(\mathbb{C})$ is obviously positive, and that of $\mathbb{C}/\Lambda$ is zero.

*The three types of hermitian symmetric spaces.* The group of isometries of a symmetric space $(M, g)$ has a natural structure of a Lie group (Helgason 1978, IV 3.2). For a hermitian symmetric space $(M, g)$, the group $\mathrm{Is}(M, g)$ of holomorphic isometries is closed in the group of isometries of $(M^{\infty}, g)$ and so is also a Lie group.

There are three families of hermitian symmetric spaces (ibid, VIII; Wolf 1984, 8.7):

| Name | example | simply connected? | curvature | $\mathrm{Is}(M,g)^+$ |
|---|---|---|---|---|
| noncompact type | $\mathcal{H}_1$ | yes | negative | adjoint, noncompact |
| compact type | $\mathbb{P}^1(\mathbb{C})$ | yes | positive | adjoint, compact |
| euclidean | $\mathbb{C}/\Lambda$ | not necessarily | zero | |

A Lie group is ***adjoint*** if it is semisimple with trivial centre.

Every hermitian symmetric space, when viewed as hermitian manifold, decomposes into a product $M^0 \times M^- \times M^+$ with $M^0$ euclidean, $M^-$ of noncompact type, and $M^+$ of compact type. The euclidean spaces are quotients of a complex space $\mathbb{C}^g$ by a discrete subgroup of translations. A hermitian symmetric space is ***irreducible*** if it is not the product of two hermitian symmetric spaces of lower dimension. Each of $M^-$ and $M^+$ is a product of irreducible hermitian symmetric spaces, each of which has a simple isometry group.

We shall be especially interested in the hermitian symmetric spaces of noncompact type — they are called ***hermitian symmetric domains.***

EXAMPLE 1.2 (Siegel upper half space). The ***Siegel upper half space*** $\mathcal{H}_g$ ***of degree*** $g$ consists of the symmetric complex $g \times g$ matrices with positive definite imaginary part, i.e.,

$$\mathcal{H}_g = \{Z = X + iY \in M_g(\mathbb{C}) \mid X = X^t, \quad Y > 0\}.$$

Note that the map $Z = (z_{ij}) \mapsto (z_{ij})_{j \geq i}$ identifies $\mathcal{H}_g$ with an open subset of $\mathbb{C}^{g(g+1)/2}$. The symplectic group $\mathrm{Sp}_{2g}(\mathbb{R})$ is the group fixing the alternating form $\sum_{i=1}^{g} x_i y_{-i} - \sum_{i=1}^{g} x_{-i} y_i$:

$$\mathrm{Sp}_{2g}(\mathbb{R}) = \left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \middle| \begin{array}{ll} A^tC = C^tA & A^tD - C^tB = I_g \\ D^tA - B^tC = I_g & B^tD = D^tB \end{array} \right\}.$$

The group $\mathrm{Sp}_{2g}(\mathbb{R})$ acts transitively on $\mathcal{H}_g$ by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} Z = (AZ + B)(CZ + D)^{-1}.$$

The matrix $\left(\begin{smallmatrix} 0 & -I_g \\ I_g & 0 \end{smallmatrix}\right)$ acts as an involution on $\mathcal{H}_g$, and has $iI_g$ as its only fixed point. Thus, $\mathcal{H}_g$ is homogeneous and symmetric as a complex manifold, and we shall see in (1.4) below that $\mathcal{H}_g$ is in fact a hermitian symmetric domain.

**Example: Bounded symmetric domains.** A ***domain*** $D$ in $\mathbb{C}^n$ is a nonempty open connected subset. It is ***symmetric*** if the group $\mathrm{Hol}(D)$ of holomorphic automorphisms of $D$ (as a complex manifold) acts transitively and for some point there exists a holomorphic symmetry. For example, $\mathcal{H}_1$ is a symmetric domain and $\mathcal{D}_1$ is a bounded symmetric domain.

THEOREM 1.3. *Every bounded domain has a canonical hermitian metric (called the Bergman(n) metric). Moreover, this metric has negative curvature.*

PROOF (SKETCH): Initially, let $D$ be any domain in $\mathbb{C}^n$. The holomorphic square-integrable functions $f\colon D \to \mathbb{C}$ form a Hilbert space $H(D)$ with inner product $(f|g) = \int_D f\overline{g}dv$. There is a unique (Bergman kernel) function $K\colon D \times D \to \mathbb{C}$ such that

(a) the function $z \mapsto K(z,\zeta)$ lies in $H(D)$ for each $\zeta$,
(b) $K(z,\zeta) = \overline{K(\zeta,z)}$, and
(c) $f(z) = \int K(z,\zeta)f(\zeta)dv(\zeta)$ for all $f \in H(D)$.

For example, for any complete orthonormal set $(e_m)_{m\in\mathbb{N}}$ in $H(D)$, $K(z,\zeta) = \sum_m e_m(z) \cdot \overline{e_m(\zeta)}$ is such a function. If $D$ is bounded, then all polynomial functions on $D$ are square-integrable, and so certainly $K(z,z) > 0$ for all $z$. Moreover, $\log(K(z,z))$ is smooth and the equations

$$h = \sum h_{ij} dz^i d\bar{z}^j, \quad h_{ij}(z) = \frac{\partial^2}{\partial z^i \partial \bar{z}^j} \log K(z,z),$$

define a hermitian metric on $D$, which can be shown to have negative curvature (Helgason 1978, VIII 3.3, 7.1; Krantz 1982, 1.4). □

The Bergman metric, being truly canonical, is invariant under the action $\mathrm{Hol}(D)$. Hence, a bounded symmetric domain becomes a hermitian symmetric domain for the Bergman metric. Conversely, it is known that every hermitian symmetric domain can be embedded into some $\mathbb{C}^n$ as a bounded symmetric domain. Therefore, a hermitian symmetric domain $D$ has a unique hermitian metric that maps to the Bergman metric under every isomorphism of $D$ with a bounded symmetric domain. On each irreducible factor, it is a multiple of the original metric.

Example 1.4. Let $\mathcal{D}_g$ be the set of symmetric complex matrices such that $I_g - \overline{Z}^t Z$ is positive definite. Note that $(z_{ij}) \mapsto (z_{ij})_{j\geq i}$ identifies $\mathcal{D}_g$ as a bounded domain in $\mathbb{C}^{g(g+1)/2}$. The map $Z \mapsto (Z - iI_g)(Z + iI_g)^{-1}$ is an isomorphism of $\mathcal{H}_g$ onto $\mathcal{D}_g$. Therefore, $\mathcal{D}_g$ is symmetric and $\mathcal{H}_g$ has an invariant hermitian metric: they are both hermitian symmetric domains.

**Automorphisms of a hermitian symmetric domain.**

Lemma 1.5. *Let $(M,g)$ be a symmetric space, and let $p \in M$. Then the subgroup $K_p$ of $\mathrm{Is}(M,g)^+$ fixing $p$ is compact, and*

$$a \cdot K_p \mapsto a \cdot p \colon \mathrm{Is}(M,g)^+/K_p \to M$$

*is an isomorphism of smooth manifolds. In particular, $\mathrm{Is}(M,g)^+$ acts transitively on $M$.*

Proof. For any riemannian manifold $(M,g)$, the compact-open topology makes $\mathrm{Is}(M,g)$ into a locally compact group for which the stabilizer $K'_p$ of a point $p$ is compact (Helgason 1978, IV 2.5). The Lie group structure on $\mathrm{Is}(M,g)$ noted above is the unique such structure compatible with the compact-open topology (ibid. II 2.6). An elementary argument (e.g., MF 1.2) now shows that $\mathrm{Is}(M,g)/K'_p \to M$ is a homeomorphism, and it follows that the map $a \mapsto ap \colon \mathrm{Is}(M,g) \to M$ is open. Write $\mathrm{Is}(M,g)$ as a finite disjoint union $\mathrm{Is}(M,g) = \bigsqcup_i \mathrm{Is}(M,g)^+ a_i$ of cosets of $\mathrm{Is}(M,g)^+$. For any two cosets the open sets $\mathrm{Is}(M,g)^+ a_i p$ and $\mathrm{Is}(M,g)^+ a_j p$ are either disjoint or equal, but, as $M$ is connected, they must all be equal, which shows that $\mathrm{Is}(M,g)^+$ acts transitively. Now $\mathrm{Is}(M,g)^+/K_p \to M$ is a homeomorphism, and it follows that it is a diffeomorphism (Helgason 1978, II 4.3a). □

Proposition 1.6. *Let $(M,g)$ be a hermitian symmetric domain. The inclusions*

$$\mathrm{Is}(M^\infty, g) \supset \mathrm{Is}(M,g) \subset \mathrm{Hol}(M)$$

*give equalities:*

$$\mathrm{Is}(M^{\infty}, g)^{+} = \mathrm{Is}(M, g)^{+} = \mathrm{Hol}(M)^{+}.$$

*Therefore,* $\mathrm{Hol}(M)^{+}$ *acts transitively on* $M$*, and* $\mathrm{Hol}(M)^{+}/K_p \cong M^{\infty}$*.*

PROOF. The first equality is proved in Helgason 1978, VIII 4.3, and the second can be proved similarly. The rest of the statement follows from (1.5). □

Let $H$ be a connected real Lie group. There need not be an algebraic group $G$ over $\mathbb{R}$ such that[3] $G(\mathbb{R})^{+} = H$. However, if $H$ has a faithful finite-dimensional representation $H \hookrightarrow \mathrm{GL}(V)$, then there exists an algebraic group $G \subset \mathrm{GL}(V)$ such that $\mathrm{Lie}(G) = [\mathfrak{h}, \mathfrak{h}]$ (inside $\mathfrak{gl}(V)$) where $\mathfrak{h} = \mathrm{Lie}(H)$ (Borel 1991, 7.9). If $H$, in addition, is semisimple, then $[\mathfrak{h}, \mathfrak{h}] = \mathfrak{h}$ and so $\mathrm{Lie}(G) = \mathfrak{h}$ and $G(\mathbb{R})^{+} = H$ (inside $\mathrm{GL}(V)$). This observation applies to any connected adjoint Lie group and, in particular, to $\mathrm{Hol}(M)^{+}$, because the adjoint representation on the Lie algebra is faithful.

PROPOSITION 1.7. *Let* $(M, g)$ *be a hermitian symmetric domain, and let* $\mathfrak{h} = \mathrm{Lie}(\mathrm{Hol}(M)^{+})$*. There is a unique connected algebraic subgroup* $G$ *of* $\mathrm{GL}(\mathfrak{h})$ *such that*

$$G(\mathbb{R})^{+} = \mathrm{Hol}(M)^{+} \qquad (\textit{inside } \mathrm{GL}(\mathfrak{h})).$$

*For such a* $G$*,*

$$G(\mathbb{R})^{+} = G(\mathbb{R}) \cap \mathrm{Hol}(M) \qquad (\textit{inside } \mathrm{GL}(\mathfrak{h}));$$

*therefore* $G(\mathbb{R})^{+}$ *is the stablizer in* $G(\mathbb{R})$ *of* $M$*.*

PROOF. The first statement was proved above, and the second follows from Satake 1980, 8.5. □

EXAMPLE 1.8. The map $z \mapsto \overline{z}^{-1}$ is an antiholomorphic isometry of $\mathcal{H}_1$, and every isometry of $\mathcal{H}_1$ is either holomorphic or differs from $z \mapsto \overline{z}^{-1}$ by a holomorphic isometry. In this case, $G = \mathrm{PGL}_2$, and $\mathrm{PGL}_2(\mathbb{R})$ acts holomorphically on $\mathbb{C} \setminus \mathbb{R}$ with $\mathrm{PGL}_2(\mathbb{R})^{+}$ as the stabilizer of $\mathcal{H}_1$.

**The homomorphism** $u_p \colon U_1 \to \mathrm{Hol}(D)$**.** Let $U_1 = \{z \in \mathbb{C} \mid |z| = 1\}$ (the circle group).

THEOREM 1.9. *Let* $D$ *be a hermitian symmetric domain. For each* $p \in D$*, there exists a unique homomorphism* $u_p \colon U_1 \to \mathrm{Hol}(D)$ *such that* $u_p(z)$ *fixes* $p$ *and acts on* $T_pD$ *as multiplication by* $z$*.*

EXAMPLE 1.10. Let $p = i \in \mathcal{H}_1$, and let $h \colon \mathbb{C}^{\times} \to \mathrm{SL}_2(\mathbb{R})$ be the homomorphism $z = a + ib \mapsto \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. Then $h(z)$ acts on the tangent space $T_i\mathcal{H}_1$ as multiplication by $z/\overline{z}$, because $\frac{d}{dz}\left(\frac{az+b}{-bz+a}\right)|_i = \frac{a^2+b^2}{(a-bi)^2}$. For $z \in U_1$, choose a square root $\sqrt{z} \in U_1$, and set $u(z) = h(\sqrt{z}) \mod \pm I$. Then $u(z)$ is independent of the choice of $\sqrt{z}$ because $h(-1) = -I$. Therefore, $u$ is a well-defined homomorphism $U_1 \to \mathrm{PSL}_2(\mathbb{R})$ such that $u(z)$ acts on the tangent space $T_i\mathcal{H}_1$ as multiplication by $z$.

Because of the importance of the theorem, I sketch a proof.

[3]For example, the (topological) fundamental group of $\mathrm{SL}_2(\mathbb{R})$ is $\mathbb{Z}$, and so $\mathrm{SL}_2(\mathbb{R})$ has many proper covering groups (even of finite degree). None of them is algebraic.

PROPOSITION 1.11. *Let $(M, g)$ be symmetric space. The symmetry $s_p$ at $p$ acts as $-1$ on $T_pM$, and, for any geodesic $\gamma$ with $\gamma(0) = p$, $s_p(\gamma(t)) = \gamma(-t)$. Moreover, $(M, g)$ is (geodesically) complete.*

PROOF. Because $s_p^2 = 1$, $(ds_p)^2 = 1$, and so $ds_p$ acts semisimply on $T_pM$ with eigenvalues $\pm 1$. Recall that for any tangent vector $X$ at $p$, there is a unique geodesic $\gamma\colon I \to M$ with $\gamma(0) = p$, $\dot{\gamma}(0) = X$. If $(ds_p)(X) = X$, then $s_p \circ \gamma$ is a geodesic sharing these properties, and so $p$ is not an isolated fixed point of $s_p$. This proves that only $-1$ occurs as an eigenvalue. If $(ds_p)(X) = -X$, then $s_p \circ \gamma$ and $t \mapsto \gamma(-t)$ are geodesics through $p$ with velocity $-X$, and so are equal. For the final statement, see Boothby 1975, VII 8.4. □

By a ***canonical tensor*** on a symmetric space $(M, g)$, I mean any tensor canonically derived from $g$, and hence fixed by any isometry of $(M, g)$.

PROPOSITION 1.12. *On a symmetric space $(M, g)$ every canonical $r$-tensor with $r$ odd is zero. In particular, parallel translation of two-dimensional subspaces does not change the sectional curvature.*

PROOF. Let $t$ be a canonical $r$-tensor. Then

$$t_p = t_p \circ (ds_p)^r \overset{1.11}{=} (-1)^r t_p,$$

and so $t = 0$ if $r$ is odd. For the second statement, let $\nabla$ be the riemannian connection, and let $R$ be the corresponding curvature tensor (Boothby 1975, VII 3.2, 4.4). Then $\nabla R$ is an odd tensor, and so is zero. This implies that parallel translation of 2-dimensional subspaces does not change the sectional curvature. □

PROPOSITION 1.13. *Let $(M, g)$ and $(M', g')$ be riemannian manifolds in which parallel translation of 2-dimensional subspaces does not change the sectional curvature. Let $a\colon T_pM \to T_{p'}M'$ be a linear isometry such that $K(p, E) = K(p', aE)$ for every 2-dimensional subspace $E \subset T_pM$. Then $\exp_p(X) \mapsto \exp_{p'}(aX)$ is an isometry of a neighbourhood of $p$ onto a neighbourhood of $p'$.*

PROOF. This follows from comparing the expansions of the riemann metrics in terms of normal geodesic coordinates. See Wolf 1984, 2.3.7. □

PROPOSITION 1.14. *If in (1.13) $M$ and $M'$ are complete, connected, and simply connected, then there is a unique isometry $\alpha\colon M \to M'$ such that $\alpha(p) = p'$ and $d\alpha_p = a$.*

PROOF. See Wolf 1984, 2.3.12. □

I now complete the sketch of the proof of Theorem 1.9. Each $z$ with $|z| = 1$ defines an automorphism of $(T_pD, g_p)$, and one checks that it preserves sectional curvatures. According to (1.11, 1.12, 1.14), there exists a unique isometry $u_p(z)\colon D \to D$ such that $du_p(z)_p$ is multiplication by $z$. It is holomorphic because it is $\mathbb{C}$-linear on the tangent spaces. The isometry $u_p(z) \circ u_p(z')$ fixes $p$ and acts as multiplication by $zz'$ on $T_pD$, and so equals $u_p(zz')$.

**Cartan involutions.** Let $G$ be a connected algebraic group over $\mathbb{R}$, and let $g \mapsto \overline{g}$ denote complex conjugation on $G(\mathbb{C})$. An involution $\theta$ of $G$ (as an algebraic group over $\mathbb{R}$) is said to be ***Cartan*** if the group

$$G^{(\theta)}(\mathbb{R}) \overset{\text{df}}{=} \{g \in G(\mathbb{C}) \mid g = \theta(\overline{g})\} \tag{9}$$

is compact.

EXAMPLE 1.15. Let $G = \mathrm{SL}_2$, and let $\theta = \mathrm{ad}\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. For $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C})$, we have

$$\theta\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \cdot \overline{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)} \cdot \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)^{-1} = \left(\begin{smallmatrix} \overline{d} & -\overline{c} \\ -\overline{b} & \overline{a} \end{smallmatrix}\right).$$

Thus,

$$\begin{aligned} \mathrm{SL}_2^{(\theta)}(\mathbb{R}) &= \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{C}) \mid d = \overline{a},\, c = -\overline{b}\right\} \\ &= \left\{\left(\begin{smallmatrix} a & b \\ -\overline{b} & \overline{a} \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{C}) \mid |a|^2 + |b|^2 = 1\right\} = \mathrm{SU}_2, \end{aligned}$$

which is compact, being a closed bounded set in $\mathbb{C}^2$. Thus $\theta$ is a Cartan involution for $\mathrm{SL}_2$.

THEOREM 1.16. *There exists a Cartan involution if and only if $G$ is reductive, in which case any two are conjugate by an element of $G(\mathbb{R})$.*

PROOF. See Satake 1980, I 4.3. $\square$

EXAMPLE 1.17. Let $G$ be a connected algebraic group over $\mathbb{R}$.

(a) The identity map on $G$ is a Cartan involution if and only if $G(\mathbb{R})$ is compact.

(b) Let $G = \mathrm{GL}(V)$ with $V$ a real vector space. The choice of a basis for $V$ determines a transpose operator $M \mapsto M^t$, and $M \mapsto (M^t)^{-1}$ is obviously a Cartan involution. The theorem says that all Cartan involutions of $G$ arise in this way.

(c) Let $G \hookrightarrow \mathrm{GL}(V)$ be a faithful representation of $G$. Then $G$ is reductive if and only if $G$ is stable under $g \mapsto g^t$ for a suitable choice of a basis for $V$, in which case the restriction of $g \mapsto (g^t)^{-1}$ to $G$ is a Cartan involution; all Cartan involutions of $G$ arise in this way from the choice of a basis for $V$ (Satake 1980, I 4.4).

(d) Let $\theta$ be an involution of $G$. There is a unique real form $G^{(\theta)}$ of $G_{\mathbb{C}}$ such that complex conjugation on $G^{(\theta)}(\mathbb{C})$ is $g \mapsto \theta(\overline{g})$. Then, $G^{(\theta)}(\mathbb{R})$ satisfies (9), and we see that the Cartan involutions of $G$ correspond to the compact forms of $G_{\mathbb{C}}$.

PROPOSITION 1.18. *Let $G$ be a connected algebraic group over $\mathbb{R}$. If $G(\mathbb{R})$ is compact, then every finite-dimensional real representation of $G \to \mathrm{GL}(V)$ carries a $G$-invariant positive definite symmetric bilinear form; conversely, if one faithful finite-dimensional real representation of $G$ carries such a form, then $G(\mathbb{R})$ is compact.*

PROOF. Let $\rho\colon G \to \mathrm{GL}(V)$ be a real representation of $G$. If $G(\mathbb{R})$ is compact, then its image $H$ in $\mathrm{GL}(V)$ is compact. Let $dh$ be the Haar measure on $H$, and choose a positive definite symmetric bilinear form $\langle\,|\,\rangle$ on $V$. Then the form

$$\langle u|v\rangle' = \int_H \langle hu|hv\rangle dh$$

is $G$-invariant, and it is still symmetric, positive definite, and bilinear. For the converse, choose an orthonormal basis for the form. Then $G(\mathbb{R})$ becomes identified with a closed set of real matrices $A$ such that $A^t \cdot A = I$, which is bounded. $\square$

REMARK 1.19. The proposition can be restated for complex representations: if $G(\mathbb{R})$ is compact then every finite-dimensional complex representation of $G$ carries a $G$-invariant positive definite Hermitian form; conversely, if some faithful finite-dimensional complex representation of $G$ carries a $G$-invariant positive definite Hermitian form, then $G$ is compact. (In this case, $G(\mathbb{R})$ is a subgroup of a unitary group instead of an orthogonal group. For a sesquilinear form $\varphi$ to be $G$-invariant means that $\varphi(gu, \overline{g}v) = \varphi(u, v)$, $g \in G(\mathbb{C})$, $u, v \in V$.)

Let $G$ be a real algebraic group, and let $C$ be an element of $G(\mathbb{R})$ whose square is central (so that ad$C$ is an involution). A ***C-polarization*** on a real representation $V$ of $G$ is a $G$-invariant bilinear form $\varphi$ such that the form $\varphi_C$,

$$(u, v) \mapsto \varphi(u, Cv),$$

is symmetric and positive definite.

PROPOSITION 1.20. *If* ad$C$ *is a Cartan involution of $G$, then every finite-dimensional real representation of $G$ carries a $C$-polarization; conversely, if one faithful finite-dimensional real representation of $G$ carries a $C$-polarization, then* ad$C$ *is a Cartan involution.*

PROOF. An $\mathbb{R}$-bilinear form $\varphi$ on a real vector space $V$ defines a sesquilinear form $\varphi'$ on $V(\mathbb{C})$,

$$\varphi' \colon V(\mathbb{C}) \times V(\mathbb{C}) \to \mathbb{C}, \quad \varphi'(u, v) = \varphi_{\mathbb{C}}(u, \overline{v}).$$

Moreover, $\varphi'$ is hermitian (and positive definite) if and only if $\varphi$ is symmetric (and positive definite).

Let $\rho \colon G \to \mathrm{GL}(V)$ be a real representation of $G$. For any $G$-invariant bilinear form $\varphi$ on $V$, $\varphi_{\mathbb{C}}$ is $G(\mathbb{C})$-invariant, and so

$$\varphi'(gu, \overline{g}v) = \varphi'(u, v), \quad \text{all } g \in G(\mathbb{C}), \quad u, v \in V(\mathbb{C}). \tag{10}$$

On replacing $v$ with $Cv$ in this equality, we find that

$$\varphi'(gu, C(C^{-1}\overline{g}C)v) = \varphi'(u, Cv), \quad \text{all } g \in G(\mathbb{C}), \quad u, v \in V(\mathbb{C}), \tag{11}$$

which says that $\varphi'_C$ is invariant under $G^{(\mathrm{ad}C)}$.

If $\rho$ is faithful and $\varphi$ is a $C$-polarization, then $\varphi'_C$ is a positive definite hermitian form, and so $G^{(\mathrm{ad}C)}(\mathbb{R})$ is compact (1.19): ad$C$ is a Cartan involution.

Conversely, if $G^{(\mathrm{ad}C)}(\mathbb{R})$ is compact, then every real representation $G \to \mathrm{GL}(V)$ carries a $G^{(\mathrm{ad}C)}(\mathbb{R})$-invariant positive definite symmetric bilinear form $\varphi$ (1.18). Similar calculations to the above show that $\varphi_{C^{-1}}$ is a $C$-polarization on $V$. □

**Representations of $U_1$.** Let $T$ be a torus over a field $k$, and let $K$ be a galois extension of $k$ splitting $T$. To give a representation $\rho$ of $T$ on a $k$-vector space $V$ amounts to giving an $X^*(T)$-grading $V(K) = \bigoplus_{\chi \in X^*(T)} V_\chi$ on $V(K) =_{\mathrm{df}} K \otimes_k V$ with the property that

$$\sigma(V_\chi) = V_{\sigma\chi}, \quad \text{all } \sigma \in \mathrm{Gal}(K/k), \quad \chi \in X^*(T).$$

Here $V_\chi$ is the subspace of $K \otimes_k V$ on which $T$ acts through $\chi$:

$$\rho(t)v = \chi(t) \cdot v, \quad \text{for } v \in V_\chi, \quad t \in T(K).$$

If $V_\chi \neq 0$, we say that $\chi$ ***occurs*** in $V$.

When we regard $U_1$ as a real algebraic torus, its characters are $z \mapsto z^n$, $n \in \mathbb{Z}$. Thus, $X^*(U_1) \cong \mathbb{Z}$, and complex conjugation acts on $X^*(U_1)$ as multiplication by $-1$. Therefore a representation of $U_1$ on a real vector space $V$ corresponds to a grading $V(\mathbb{C}) = \oplus_{n \in \mathbb{Z}} V^n$ with the property that $V(\mathbb{C})^{-n} = \overline{V(\mathbb{C})^n}$ (complex conjugate). Here $V^n$ is the subspace of $V(\mathbb{C})$ on which $z$ acts as $z^n$. Note that

$V(\mathbb{C})^0 = \overline{V(\mathbb{C})^0}$ and so it is defined over $\mathbb{R}$, i.e., $V(\mathbb{C})^0 = V^0(\mathbb{C})$ for $V^0$ the subspace $V \cap V(\mathbb{C})^0$ of $V$ (see AG 16.7). The natural map

$$V/V^0 \to V(\mathbb{C})/\bigoplus\nolimits_{n\leq 0} V(\mathbb{C})^n \cong \bigoplus\nolimits_{n>0} V(\mathbb{C})^n \tag{12}$$

is an isomorphism. From this discussion, we see that every real representation of $U_1$ is a direct sum of representations of the following types:

(a) $V = \mathbb{R}$ with $U_1$ acting trivially (so $V(\mathbb{C}) = V^0$);
(b) $V = \mathbb{R}^2$ with $z = x + iy \in U_1(\mathbb{R})$ acting as $\left(\begin{smallmatrix} x & y \\ -y & x \end{smallmatrix}\right)^n$, $n > 0$ (so $V(\mathbb{C}) = V^n \oplus V^{-n}$).

**Classification of hermitian symmetric domains in terms of real groups.** The representations of $U_1$ have the same description whether we regard it as a Lie group or an algebraic group, and so every homomorphism $U_1 \to \mathrm{GL}(V)$ of Lie groups is algebraic. It follows that the homomorphism $u_p\colon U_1 \to \mathrm{Hol}(D)^+ \cong G(\mathbb{R})^+$ (see 1.9, 1.7) is algebraic.

THEOREM 1.21. *Let $D$ be a hermitian symmetric domain, and let $G$ be the associated real adjoint algebraic group (1.7). The homomorphism $u_p\colon U_1 \to G$ attached to a point $p$ of $D$ has the following properties:*

(a) *only the characters $z, 1, z^{-1}$ occur in the representation of $U_1$ on $\mathrm{Lie}(G)_{\mathbb{C}}$ defined by $u_p$;*
(b) *$\mathrm{ad}(u_p(-1))$ is a Cartan involution;*
(c) *$u_p(-1)$ does not project to $1$ in any simple factor of $G$.*

*Conversely, let $G$ be a real adjoint algebraic group, and let $u\colon U_1 \to G$ satisfy (a), (b), and (c). Then the set $D$ of conjugates of $u$ by elements of $G(\mathbb{R})^+$ has a natural structure of a hermitian symmetric domain for which $G(\mathbb{R})^+ = \mathrm{Hol}(D)^+$ and $u(-1)$ is the symmetry at $u$ (regarded as a point of $D$).*

PROOF (SKETCH): Let $D$ be a hermitian symmetric domain, and let $G$ be the associated group (1.7). Then $G(\mathbb{R})^+/K_p \cong D$ where $K_p$ is the group fixing $p$ (see 1.6). For $z \in U_1$, $u_p(z)$ acts on the $\mathbb{R}$-vector space

$$\mathrm{Lie}(G)/\,\mathrm{Lie}(K_p) \cong T_pD$$

as multiplication by $z$, and it acts on $\mathrm{Lie}(K_p)$ trivially. From this, (a) follows.

The symmetry $s_p$ at $p$ and $u_p(-1)$ both fix $p$ and act as $-1$ on $T_pD$ (see 1.11); they are therefore equal (1.14). It is known that the symmetry at a point of a symmetric space gives a Cartan involution of $G$ if and only if the space has negative curvature (see Helgason 1978, V 2; the real form of $G$ defined by $\mathrm{ad}s_p$ is that attached to the compact dual of the symmetric space). Thus (b) holds.

Finally, if the projection of $u(-1)$ into a simple factor of $G$ were trivial, then that factor would be compact (by (b); see 1.17a), and $D$ would have an irreducible factor of compact type.

For the converse, let $D$ be the set of $G(\mathbb{R})^+$-conjugates of $u$. The centralizer $K_u$ of $u$ in $G(\mathbb{R})^+$ is contained in $\{g \in G(\mathbb{C}) \mid g = u(-1)\cdot \overline{g}\cdot u(-1)^{-1}\}$, which, according to (b), is compact. As $K_u$ is closed, it also is compact. The equality $D = (G(\mathbb{R})^+/K_u)\cdot u$ endows $D$ with the structure of smooth (even real-analytic) manifold. For this structure, the tangent space to $D$ at $u$,

$$T_uD = \mathrm{Lie}(G)/\,\mathrm{Lie}(K_u),$$

which, because of (a), can be identified with the subspace of $\mathrm{Lie}(G)_{\mathbb{C}}$ on which $u(z)$ acts as $z$ (see (12)). This endows $T_uD$ with a $\mathbb{C}$-vector space structure for which $u(z)$, $z \in U_1$, acts as multiplication by $z$. Because $D$ is homogeneous, this gives it the structure of an almost-complex manifold, which can be shown to integrable (Wolf 1984, 8.7.9). The action of $K_u$ on $D$ defines an action of it on $T_uD$. Because $K_u$ is compact, there is a $K_u$-invariant positive definite form on $T_uD$ (see 1.18), and because $J = u(i) \in K_u$, any such form will have the hermitian property (7). Choose one, and use the homogeneity of $D$ to move it to each tangent space. This will make $D$ into a hermitian symmetric space, which will be a hermitian symmetric domain because each simple factor of its automorphism group is a noncompact semisimple group (because of (b,c)). □

COROLLARY 1.22. *There is a natural one-to-one correspondence between isomorphism classes of pointed hermitian symmetric domains and pairs $(G, u)$ consisting of a real adjoint Lie group and a nontrivial homomorphism $u\colon U_1 \to G(\mathbb{R})$ satisfying (a), (b), (c).*

EXAMPLE 1.23. Let $u\colon U_1 \to \mathrm{PSL}_2(\mathbb{R})$ be as in (1.10). Then $u(-1) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and we saw in 1.15 that $\mathrm{ad}u(-1)$ is a Cartan involution of $\mathrm{SL}_2$, hence also of $\mathrm{PSL}_2$.

**Classification of hermitian symmetric domains in terms of dynkin diagrams.** Let $G$ be a simple adjoint group over $\mathbb{R}$, and let $u$ be a homomorphism $U_1 \to G$ satisfying (a) and (b) of Theorem 1.21. By base extension, we get an adjoint group $G_{\mathbb{C}}$, which is simple because it is an inner form of its compact form, and a cocharacter $\mu = u_{\mathbb{C}}$ of $G_{\mathbb{C}}$ satisfying the following condition:

> (*) in the action of $\mathbb{G}_m$ on $\mathrm{Lie}(G_{\mathbb{C}})$ defined by $\mathrm{ad} \circ \mu$, only the characters $z, 1, z^{-1}$ occur.

PROPOSITION 1.24. *The map $(G, u) \mapsto (G_{\mathbb{C}}, u_{\mathbb{C}})$ defines a bijection between the sets of isomorphism classes of pairs consisting of*

(a) *a simple adjoint group over $\mathbb{R}$ and a conjugacy class of $u\colon U_1 \to H$ satisfying (1.21a,b), and*
(b) *a simple adjoint group over $\mathbb{C}$ and a conjugacy class of cocharacters satisfying (*).*

PROOF. Let $(G, \mu)$ be as in (b), and let $g \mapsto \overline{g}$ denote complex conjugation on $G(\mathbb{C})$ relative to the unique compact real form of $G$ (cf. 1.16). There is a real form $H$ of $G$ such that complex conjugation on $H(\mathbb{C}) = G(\mathbb{C})$ is $g \mapsto \mu(-1) \cdot \overline{g} \cdot \mu(-1)^{-1}$, and $u =_{\mathrm{df}} \mu|U_1$ takes values in $H(\mathbb{R})$. The pair $(H, u)$ is as in (a), and the map $(G, \mu) \to (H, u)$ is inverse to $(H, u) \mapsto (H_{\mathbb{C}}, u_{\mathbb{C}})$ on isomorphism classes. □

Let $G$ be a simple algebraic group $\mathbb{C}$. Choose a maximal torus $T$ in $G$ and a base $(\alpha_i)_{i \in I}$ for the roots of $G$ relative to $T$. Recall, that the nodes of the dynkin diagram of $(G, T)$ are indexed by $I$. Recall also (Bourbaki 1981, VI 1.8) that there is a unique ***(highest) root*** $\tilde{\alpha} = \sum n_i \alpha_i$ such that, for any other root $\sum m_i \alpha_i$, $n_i \geq m_i$ all $i$. An $\alpha_i$ (or the associated node) is said to be ***special*** if $n_i = 1$.

Let $M$ be a conjugacy class of nontrivial cocharacters of $G$ satisfying (*). Because all maximal tori of $G$ are conjugate, $M$ has a representative in $X_*(T) \subset X_*(G)$, and because the Weyl group acts simply transitively on the Weyl chambers (Humphreys 1972, 10.3) there is a unique representative $\mu$ for $M$ such that

$\langle\alpha_i,\mu\rangle \geq 0$ for all $i \in I$. The condition (*) is that[4] $\langle\alpha,\mu\rangle \in \{1,0,-1\}$ for all roots $\alpha$. Since $\mu$ is nontrivial, not all the values $\langle\alpha,\mu\rangle$ can be zero, and so this condition implies that $\langle\alpha_i,\mu\rangle = 1$ for exactly one $i \in I$, which must in fact be special (otherwise $\langle\tilde{\alpha},\mu\rangle > 1$). Thus, the $M$ satisfying (*) are in one-to-one correspondence with the special nodes of the dynkin diagram. In conclusion:

THEOREM 1.25. *The isomorphism classes of irreducible hermitian symmetric domains are classified by the special nodes on connected dynkin diagrams.*

The special nodes can be read off from the list of dynkin diagrams in, for example, Helgason 1978, p477. In the following table, we list the number of special nodes for each type:

| Type | $A_n$ | $B_n$ | $C_n$ | $D_n$ | $E_6$ | $E_7$ | $E_8$ | $F_4$ | $G_2$ |
|---|---|---|---|---|---|---|---|---|---|
| | $n$ | 1 | 1 | 3 | 2 | 1 | 0 | 0 | 0 |

In particular, there are no irreducible hermitian symmetric domains of type $E_8$, $F_4$, or $G_2$ and, up to isomorphism, there are exactly 2 of type $E_6$ and 1 of type $E_7$. It should be noted that not every simple real algebraic group arises as the automorphism group of a hermitian symmetric domain. For example, $\mathrm{PGL}_n$ arises in this way only for $n = 2$.

NOTES. For introductions to smooth manifolds and riemannian manifolds, see Boothby 1975 and Lee 1997. The ultimate source for hermitian symmetric domains is Helgason 1978, but Wolf 1984 is also very useful, and Borel 1998 gives a succinct treatment close to that of the pioneers. The present account has been influenced by Deligne 1973*a* and Deligne 1979.

## 2. Hodge structures and their classifying spaces

We describe various objects and their parameter spaces. Our goal is a description of hermitian symmetric domains as the parameter spaces for certain special hodge structures.

**Reductive groups and tensors.** Let $G$ be a reductive group over a field $k$ of characteristic zero, and let $\rho\colon G \to \mathrm{GL}(V)$ be a representation of $G$. The ***contragredient*** or ***dual*** $\rho^\vee$ of $\rho$ is the representation of $G$ on the dual vector space $V^\vee$ defined by

$$(\rho^\vee(g)\cdot f)(v) = f(\rho(g^{-1})\cdot v), \quad g \in G,\ f \in V^\vee,\ v \in V.$$

A representation is said to be ***self-dual*** if it is isomorphic to its contragredient.

An $r$***-tensor*** of $V$ is a multilinear map

$$t\colon V \times \cdots \times V \to k \quad (r\text{-copies of } V).$$

For an $r$-tensor $t$, the condition

$$t(gv_1,\ldots,gv_r) = (v_1,\ldots,v_r), \quad \text{all } v_i \in V,$$

on $g$ defines a closed subgroup of $\mathrm{GL}(V)_t$ of $\mathrm{GL}(V)$. For example, if $t$ is a nondegenerate symmetric bilinear form $V \times V \to k$, then $\mathrm{GL}(V)_t$ is the orthogonal group. For a set $T$ of tensors of $V$, $\bigcap_{t\in T}\mathrm{GL}(V)_t$ is called the ***subgroup of*** $\mathrm{GL}(V)$ ***fixing the*** $t \in T$.

[4]The $\mu$ with this property are sometimes said to be ***minuscule*** (cf. Bourbaki 1981, pp226–227).

PROPOSITION 2.1. *For any faithful self-dual representation $G \to \mathrm{GL}(V)$ of $G$, there exists a finite set $T$ of tensors of $V$ such that $G$ is the subgroup of $\mathrm{GL}(V)$ fixing the $t \in T$.*

PROOF. In Deligne 1982, 3.1, it is shown there exists a possibly infinite set $T$ with this property, but, because $G$ is noetherian as a topological space (i.e., it has the descending chain condition on closed subsets), a finite subset will suffice. □

PROPOSITION 2.2. *Let $G$ be the subgroup of $\mathrm{GL}(V)$ fixing the tensors $t \in T$. Then*

$$\mathrm{Lie}(G) = \left\{ g \in \mathrm{End}(V) \;\middle|\; \textstyle\sum_j t(v_1, \ldots, gv_j, \ldots, v_r) = 0, \quad \text{all } t \in T,\ v_i \in V \right\}.$$

PROOF. The Lie algebra of an algebraic group $G$ can be defined to be the kernel of $G(k[\varepsilon]) \to G(k)$. Here $k[\varepsilon]$ is the $k$-algebra with $\varepsilon^2 = 0$. Thus $\mathrm{Lie}(G)$ consists of the endomorphisms $1 + g\varepsilon$ of $V(k[\varepsilon])$ such that

$$t((1+g\varepsilon)v_1, (1+g\varepsilon)v_2, \ldots) = t(v_1, v_2, \ldots), \quad \text{all } t \in T,\ v_i \in V.$$

On expanding this and cancelling, we obtain the assertion. □

**Flag varieties.** Fix a vector space $V$ of dimension $n$ over a field $k$.

*The projective space* $\mathbb{P}(V)$. The set $\mathbb{P}(V)$ of one-dimensional subspaces $L$ of $V$ has a natural structure of an algebraic variety: the choice of a basis for $V$ determines a bijection $\mathbb{P}(V) \to \mathbb{P}^{n-1}$, and the structure of an algebraic variety inherited by $\mathbb{P}(V)$ from the bijection is independent of the choice of the basis.

*Grassmann varieties.* Let $G_d(V)$ be the set of $d$-dimensional subspaces of $V$, some $0 < d < n$. Fix a basis for $V$. The choice of a basis for $W$ then determines a $d \times n$ matrix $A(W)$ whose rows are the coordinates of the basis elements. Changing the basis for $W$ multiplies $A(W)$ on the left by an invertible $d \times d$ matrix. Thus, the family of minors of degree $d$ of $A(W)$ is well-determined up to multiplication by a nonzero constant, and so determines a point $P(W)$ in $\mathbb{P}^{\binom{n}{d}-1}$. The map $W \mapsto P(W) \colon G_d(V) \to \mathbb{P}^{\binom{n}{d}-1}$ identifies $G_d(V)$ with a closed subvariety of $\mathbb{P}^{\binom{n}{d}-1}$ (AG 6.26). A coordinate-free description of this map is given by

$$W \mapsto \textstyle\bigwedge^d W \colon G_d(V) \to \mathbb{P}(\textstyle\bigwedge^d V). \tag{13}$$

Let $S$ be a subspace of $V$ of complementary dimension $n - d$, and let $G_d(V)_S$ be the set of $W \in G_d(V)$ such that $W \cap S = \{0\}$. Fix a $W_0 \in G_d(V)_S$, so that $V = W_0 \oplus S$. For any $W \in G_d(V)_S$, the projection $W \to W_0$ given by this decomposition is an isomorphism, and so $W$ is the graph of a homomorphism $W_0 \to S$:

$$w \mapsto s \iff (w, s) \in W.$$

Conversely, the graph of any homomorphism $W_0 \to S$ lies in $G_d(V)_S$. Thus,

$$G_d(V)_S \cong \mathrm{Hom}(W_0, S). \tag{14}$$

When we regard $G_d(V)_S$ as an open subvariety of $G_d(V)$, this isomorphism identifies it with the affine space $\mathbb{A}(\mathrm{Hom}(W_0, S))$ defined by the vector space $\mathrm{Hom}(W_0, S)$. Thus, $G_d(V)$ is smooth, and the tangent space to $G_d(V)$ at $W_0$,

$$T_{W_0}(G_d(V)) \cong \mathrm{Hom}(W_0, S) \cong \mathrm{Hom}(W_0, V/W_0). \tag{15}$$

*Flag varieties.* The above discussion extends easily to chains of subspaces. Let $\mathbf{d} = (d_1, \ldots, d_r)$ be a sequence of integers with $n > d_1 > \cdots > d_r > 0$, and let $G_{\mathbf{d}}(V)$ be the set of flags

$$F: \quad V \supset V^1 \supset \cdots \supset V^r \supset 0 \tag{16}$$

with $V^i$ a subspace of $V$ of dimension $d_i$. The map

$$G_{\mathbf{d}}(V) \xrightarrow{F \mapsto (V^i)} \prod\nolimits_i G_{d_i}(V) \subset \prod\nolimits_i \mathbb{P}(\bigwedge\nolimits^{d_i} V)$$

realizes $G_{\mathbf{d}}(V)$ as a closed subset of $\prod_i G_{d_i}(V)$ (Humphreys 1978, 1.8), and so it is a projective variety. The tangent space to $G_{\mathbf{d}}(V)$ at the flag $F$ consists of the families of homomorphisms

$$\varphi^i \colon V^i \to V/V^i, \quad 1 \leq i \leq r, \tag{17}$$

satisfying the compatibility condition

$$\varphi^i | V^{i+1} \equiv \varphi^{i+1} \mod V^{i+1}.$$

ASIDE 2.3. A basis $e_1, \ldots, e_n$ for $V$ is ***adapted to*** the flag $F$ if it contains a basis $e_1, \ldots, e_{j_i}$ for each $V^i$. Clearly, every flag admits such a basis, and the basis then determines the flag. Because $\mathrm{GL}(V)$ acts transitively on the set of bases for $V$, it acts transitively on $G_{\mathbf{d}}(V)$. For a flag $F$, the subgroup $P(F)$ stabilizing $F$ is an algebraic subgroup of $\mathrm{GL}(V)$, and the map

$$g \mapsto gF_0 \colon \mathrm{GL}(V)/P(F_0) \to G_{\mathbf{d}}(V)$$

is an isomorphism of algebraic varieties. Because $G_{\mathbf{d}}(V)$ is projective, this shows that $P(F_0)$ is a parabolic subgroup of $\mathrm{GL}(V)$.

**Hodge structures.**

*Definition.* For a real vector space $V$, complex conjugation on $V(\mathbb{C}) =_{\mathrm{df}} \mathbb{C} \otimes_{\mathbb{R}} V$ is defined by

$$\overline{z \otimes v} = \overline{z} \otimes v.$$

An $\mathbb{R}$-basis $e_1, \ldots, e_m$ for $V$ is also a $\mathbb{C}$-basis for $V(\mathbb{C})$ and $\overline{\sum a_i e_i} = \sum \overline{a_i} e_i$.

A ***hodge decomposition*** of a real vector space $V$ is a decomposition

$$V(\mathbb{C}) = \bigoplus_{p,q \in \mathbb{Z} \times \mathbb{Z}} V^{p,q}$$

such that $V^{q,p}$ is the complex conjugate of $V^{p,q}$. A ***hodge structure*** is a real vector space together with a hodge decomposition. The set of pairs $(p, q)$ for which $V^{p,q} \neq 0$ is called the ***type*** of the hodge structure. For each $n$, $\bigoplus_{p+q=n} V^{p,q}$ is stable under complex conjugation, and so is defined over $\mathbb{R}$, i.e., there is a subspace $V_n$ of $V$ such that $V_n(\mathbb{C}) = \bigoplus_{p+q=n} V^{p,q}$ (see AG 16.7). Then $V = \bigoplus_n V_n$ is called the ***weight decomposition*** of $V$. If $V = V_n$, then $V$ is said to have weight $n$.

An ***integral*** (resp. ***rational***) ***hodge structure*** is a free $\mathbb{Z}$-module of finite rank $V$ (resp. $\mathbb{Q}$-vector space) together with a hodge decomposition of $V(\mathbb{R})$ such that the weight decomposition is defined over $\mathbb{Q}$.

EXAMPLE 2.4. Let $J$ be a complex structure on a real vector space $V$, and define $V^{-1,0}$ and $V^{0,-1}$ to be the $+i$ and $-i$ eigenspaces of $J$ acting on $V(\mathbb{C})$. Then $V(\mathbb{C}) = V^{-1,0} \oplus V^{0,-1}$ is a hodge structure of type $(-1, 0), (0, -1)$, and every real hodge structure of this type arises from a (unique) complex structure. Thus, to give a rational hodge structure of type $(-1, 0), (0, -1)$ amounts to giving a $\mathbb{Q}$-vector space $V$ and a complex structure on $V(\mathbb{R})$, and to give an integral hodge structure of type $(-1,0),(0,-1)$ amounts to giving a $\mathbb{C}$-vector space $V$ and a lattice $\Lambda \subset V$ (i.e., a $\mathbb{Z}$-submodule generated by an $\mathbb{R}$-basis for $V$).

EXAMPLE 2.5. Let $X$ be a nonsingular projective algebraic variety over $\mathbb{C}$. Then $H = H^n(X,\mathbb{Q})$ has a hodge structure of weight $n$ for which $H^{p,q} \subset H^n(X,\mathbb{C})$ is canonically isomorphic to $H^q(X,\Omega^p)$ (Voisin 2002, 6.1.3).

EXAMPLE 2.6. Let $\mathbb{Q}(m)$ be the hodge structure of weight $-2m$ on the vector space $\mathbb{Q}$. Thus, $(\mathbb{Q}(m))(\mathbb{C}) = \mathbb{Q}(m)^{-m,-m}$. Define $\mathbb{Z}(m)$ and $\mathbb{R}(m)$ similarly.[5]

*The hodge filtration.* The ***hodge filtration*** associated with a hodge structure of weight $n$ is

$$F^\bullet\colon \quad \cdots \supset F^p \supset F^{p+1} \supset \cdots, \quad F^p = \bigoplus\nolimits_{r\geq p} V^{r,s} \subset V(\mathbb{C}).$$

Note that for $p+q=n$,

$$\overline{F^q} = \bigoplus\nolimits_{s\geq q} \overline{V^{s,r}} = \bigoplus\nolimits_{s\geq q} V^{r,s} = \bigoplus\nolimits_{r\leq p} V^{r,s}$$

and so

$$V^{p,q} = F^p \cap \overline{F^q}. \tag{18}$$

EXAMPLE 2.7. For a hodge structure of type $(-1,0),(0,-1)$, the hodge filtration is

$$(F^{-1} \supset F^0 \supset F^2) = (V(\mathbb{C}) \supset V^{0,-1} \supset 0).$$

The obvious $\mathbb{R}$-linear isomorphism $V \to V(\mathbb{C})/F^0$ defines the complex structure on $V$ noted in (2.4).

*Hodge structures as representations of* $\mathbb{S}$. Let $\mathbb{S}$ be $\mathbb{C}^\times$ regarded as a torus over $\mathbb{R}$. It can be identified with the closed subgroup of $\mathrm{GL}_2(\mathbb{R})$ of matrices of the form $\left(\begin{smallmatrix} a & b \\ -b & a\end{smallmatrix}\right)$. Then $\mathbb{S}(\mathbb{C}) \approx \mathbb{C}^\times \times \mathbb{C}^\times$ with complex conjugation acting by the rule $\overline{(z_1,z_2)} = (\overline{z_2},\overline{z_1})$. We fix the isomorphism $\mathbb{S}_\mathbb{C} \cong \mathbb{G}_m \times \mathbb{G}_m$ so that $\mathbb{S}(\mathbb{R}) \to \mathbb{S}(\mathbb{C})$ is $z \mapsto (z,\overline{z})$, and we define the weight homomorphism $w\colon \mathbb{G}_m \to \mathbb{S}$ so that $\mathbb{G}_m(\mathbb{R}) \xrightarrow{w} \mathbb{S}(\mathbb{R})$ is $r \mapsto r^{-1}\colon \mathbb{R}^\times \to \mathbb{C}^\times$.

The characters of $\mathbb{S}_\mathbb{C}$ are the homomorphisms $(z_1,z_2) \mapsto z_1^p z_2^q$, $(r,s) \in \mathbb{Z}\times\mathbb{Z}$. Thus, $X^*(\mathbb{S}) = \mathbb{Z}\times\mathbb{Z}$ with complex conjugation acting as $(p,q)\mapsto(q,p)$, and to give a representation of $\mathbb{S}$ on a real vector space $V$ amounts to giving a $\mathbb{Z}\times\mathbb{Z}$-grading of $V(\mathbb{C})$ such that $\overline{V^{p,q}} = V^{q,p}$ for all $p,q$ (see p276). Thus, to give a representation of $\mathbb{S}$ on a real vector space $V$ is the same as to give a hodge structure on $V$. Following Deligne 1979, 1.1.1.1, we normalize the relation as follows: the homomorphism $h\colon \mathbb{S} \to \mathrm{GL}(V)$ corresponds to the hodge structure on $V$ such that

$$h_\mathbb{C}(z_1,z_2)v = z_1^{-p} z_2^{-q} v \text{ for } v \in V^{p,q}. \tag{19}$$

In other words,

$$h(z)v = z^{-p}\overline{z}^{-q} v \text{ for } v \in V^{p,q}. \tag{20}$$

Note the minus signs! The associated weight decomposition has

$$V_n = \{v \in V \mid w_h(r)v = r^n\}, \quad w_h = h\circ w. \tag{21}$$

[5] It would be a little more canonical to take the underlying vector space of $\mathbb{Q}(m)$ to be $(2\pi i)^m\mathbb{Q}$ because this makes certain relations invariant under a change of the choice of $i=\sqrt{-1}$ in $\mathbb{C}$.

Let $\mu_h$ be the cocharacter of $\mathrm{GL}(V)$ defined by

$$\mu_h(z) = h_{\mathbb{C}}(z, 1). \tag{22}$$

Then the elements of $F_h^p V$ are sums of $v \in V(\mathbb{C})$ satisfying $\mu_h(z)\, v = z^{-r} v$ for some $r \geq p$.

To give a hodge structure on a $\mathbb{Q}$-vector space $V$ amounts to giving a homomorphism $h\colon \mathbb{S} \to \mathrm{GL}(V(\mathbb{R}))$ such that $w_h$ is defined over $\mathbb{Q}$.

EXAMPLE 2.8. By definition, a complex structure on a real vector space is a homomorphism $h\colon \mathbb{C} \to \mathrm{End}_{\mathbb{R}}(V)$ of $\mathbb{R}$-algebras. Then $h|\mathbb{C}^\times\colon \mathbb{C}^\times \to \mathrm{GL}(V)$ is a hodge structure of type $(-1,0),(0,-1)$ whose associated complex structure (see 2.4) is that defined by $h$.[6]

EXAMPLE 2.9. The hodge structure $\mathbb{Q}(m)$ corresponds to the homomorphism $h\colon \mathbb{S} \to \mathbb{G}_{m\mathbb{R}}$, $h(z) = (z\bar{z})^m$.

*The Weil operator.* For a hodge structure $(V,h)$, the $\mathbb{R}$-linear map $C = h(i)$ is called the ***Weil operator***. Note that $C$ acts as $i^{q-p}$ on $V^{p,q}$ and that $C^2 = h(-1)$ acts as $(-1)^n$ on $V_n$.

EXAMPLE 2.10. If $V$ is of type $(-1,0)$, $(0,-1)$, then $C$ coincides with the $J$ of (2.4). The functor $(V,(V^{-1,0},V^{0,-1})) \mapsto (V,C)$ is an equivalence from the category of real hodge structures of type $(-1,0),(0,-1)$ to the category of complex vector spaces.

*Hodge structures of weight* 0.. Let $V$ be a hodge structure of weight 0. Then $V^{0,0}$ is invariant under complex conjugation, and so $V^{0,0} = V^{00}(\mathbb{C})$, where $V^{00} = V^{0,0} \cap V$ (see AG 16.7). Note that

$$V^{00} = \mathrm{Ker}(V \to V(\mathbb{C})/F^0). \tag{23}$$

*Tensor products of hodge structures.* The ***tensor product of hodge structures*** $V$ and $W$ of weight $m$ and $n$ is a hodge structure of weight $m+n$:

$$V \otimes W, \quad (V \otimes W)^{p,q} = \bigoplus\nolimits_{r+r'=p,s+s'=q} V^{r,s} \otimes V^{r',s'}.$$

In terms of representations of $\mathbb{S}$,

$$(V, h_V) \otimes (W, h_W) = (V \otimes W, h_V \otimes h_W).$$

*Morphisms of hodge structures.* A morphism of hodge structures is a linear map $V \to W$ sending $V^{p,q}$ into $W^{p,q}$ for all $p,q$. In other words, it is a morphism $(V,h_V) \to (W,h_W)$ of representations of $\mathbb{S}$.

*Hodge tensors.* Let $R = \mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, and let $(V,h)$ be an $R$-hodge structure of weight $n$. A multilinear form $t\colon V^r \to R$ is a ***hodge tensor*** if the map

$$V \otimes V \otimes \cdots \otimes V \to R(-nr/2)$$

it defines is a morphism of hodge structures. In other words, $t$ is a hodge tensor if

$$t(h(z)v_1, h(z)v_2, \ldots) = (z\bar{z})^{-nr/2} \cdot t_{\mathbb{R}}(v_1, v_2, \ldots), \text{ all } z \in \mathbb{C},\ v_i \in V(\mathbb{R}),$$

or if

$$\textstyle\sum p_i \neq \sum q_i \Rightarrow t_{\mathbb{C}}(v_1^{p_1,q_1}, v_2^{p_2,q_2}, \ldots) = 0, \quad v_i^{p_i,q_i} \in V^{p_i,q_i}. \tag{24}$$

[6]This partly explains the signs in (19); see also Deligne 1979, 1.1.6. Following Deligne 1973*b*, 8.12, and Deligne 1979, 1.1.1.1, $h_{\mathbb{C}}(z_1,z_2)v^{p,q} = z_1^{-p} z_2^{-q} v^{p,q}$ has become the standard convention in the theory of Shimura varieties. Following Deligne 1971*a*, 2.1.5.1, the convention $h_{\mathbb{C}}(z_1,z_2)v^{p,q} = z_1^p z_2^q v^{p,q}$ is commonly used in hodge theory (e.g., Voisin 2002, p147).

Note that, for a hodge tensor $t$,

$$t(Cv_1, Cv_2, \ldots) = t(v_1, v_2, \ldots).$$

EXAMPLE 2.11. Let $(V, h)$ be a hodge structure of type $(-1,0),(0,-1)$. A bilinear form $t\colon V \times V \to \mathbb{R}$ is a hodge tensor if and only if $t(Ju, Jv) = t(u, v)$ for all $u, v \in V$.

*Polarizations.* Let $(V, h)$ be a hodge structure of weight $n$. A ***polarization*** of $(V, h)$ is a hodge tensor $\psi\colon V \times V \to \mathbb{R}$ such that $\psi_C(u, v) =_{\text{df}} \psi(u, Cv)$ is symmetric and positive definite. Then $\psi$ is symmetric or alternating according as $n$ is even or odd, because

$$\psi(v, u) = \psi(Cv, Cu) = \psi_C(Cv, u) = \psi_C(u, Cv) = \psi(u, C^2 v) = (-1)^n \psi(u, v).$$

More generally, let $(V, h)$ be an $R$-hodge structure of weight $n$ where $R$ is $\mathbb{Z}$ or $\mathbb{Q}$. A ***polarization*** of $(V, h)$ is a bilinear form $\psi\colon V \times V \to R$ such that $\psi_{\mathbb{R}}$ is a polarization of $(V(\mathbb{R}), h)$.

EXAMPLE 2.12. Let $(V, h)$ be an $R$-hodge structure of type $(-1,0),(0,-1)$ with $R = \mathbb{Z}$, $\mathbb{Q}$, or $\mathbb{R}$, and let $J = h(i)$. A polarization of $(V, h)$ is an alternating bilinear form $\psi\colon V \times V \to R$ such that, for $u, v \in V(\mathbb{R})$,

$$\psi_{\mathbb{R}}(Ju, Jv) = \psi(u, v), \text{ and}$$
$$\psi_{\mathbb{R}}(u, Ju) > 0 \text{ if } u \neq 0.$$

(These conditions imply that $\psi_{\mathbb{R}}(u, Jv)$ is symmetric.)

EXAMPLE 2.13. Let $X$ be a nonsingular projective variety over $\mathbb{C}$. The choice of an embedding $X \hookrightarrow \mathbb{P}^N$ determines a polarization on the primitive part of $H^n(X, \mathbb{Q})$ (Voisin 2002, 6.3.2).

**Variations of hodge structures.** Fix a real vector space $V$, and let $S$ be a connected complex manifold. Suppose that, for each $s \in S$, we have a hodge structure $h_s$ on $V$ of weight $n$ (independent of $s$). Let $V_s^{p,q} = V_{h_s}^{p,q}$ and $F_s^p = F_s^p V = F_{h_s}^p V$.

The family of hodge structures $(h_s)_{s \in S}$ on $V$ is said to be ***continuous*** if, for fixed $p$ and $q$, the subspace $V_s^{p,q}$ varies continuously with $s$. This means that the dimension $d(p, q)$ of $V_s^{p,q}$ is constant and the map

$$s \mapsto V_s^{p,q}\colon S \to G_{d(p,q)}(V)$$

is continuous.

A continuous family of hodge structures $(V_s^{p,q})_s$ is said to be ***holomorphic*** if the hodge filtration $F_s^\bullet$ varies holomorphically with $s$. This means that the map $\varphi$,

$$s \mapsto F_s^\bullet\colon S \to G_{\mathbf{d}}(V)$$

is holomorphic. Here $\mathbf{d} = (\ldots, d(p), \ldots)$ where $d(p) = \dim F_s^p V = \sum_{r \geq p} d(r, q)$. Then the differential of $\varphi$ at $s$ is a $\mathbb{C}$-linear map

$$d\varphi_s\colon T_s S \to T_{F_s^\bullet}(G_{\mathbf{d}}(V)) \overset{(17)}{\subset} \bigoplus\nolimits_p \operatorname{Hom}(F_s^p, V/F_s^p).$$

If the image of $d\varphi_s$ is contained in

$$\bigoplus\nolimits_p \operatorname{Hom}(F_s^p, F_s^{p-1}/F_s^p),$$

for all $s$, then the holomorphic family is called a ***variation of hodge structures on*** $S$.

Now let $T$ be a family of tensors on $V$ including a nondegenerate bilinear form $t_0$, and let $d\colon \mathbb{Z}\times\mathbb{Z}\to\mathbb{N}$ be a function such that

$$\begin{aligned} &d(p,q)=0 \text{ for almost all } p,q;\\ &d(q,p)=d(p,q);\\ &d(p,q)=0 \text{ unless } p+q=n. \end{aligned}$$

Define $S(d,T)$ to be the set of all hodge structures $h$ on $V$ such that

- $\dim V_h^{p,q}=d(p,q)$ for all $p,q$;
- each $t\in T$ is a hodge tensor for $h$;
- $t_0$ is a polarization for $h$.

Then $S(d,T)$ acquires a topology as a subspace of $\prod_{d(p,q)\neq 0} G_{d(p,q)}(V)$.

THEOREM 2.14. *Let $S^+$ be a connected component of $S(d,T)$.*

(a) *If nonempty, $S^+$ has a unique complex structure for which $(h_s)$ is a holomorphic family of hodge structures.*
(b) *With this complex structure, $S^+$ is a hermitian symmetric domain if $(h_s)$ is a variation of hodge structures.*
(c) *Every irreducible hermitian symmetric domain is of the form $S^+$ for a suitable $V$, $d$, and $T$.*

PROOF (SKETCH). (a) Let $S^+=S(d,T)^+$. Because the hodge filtration determines the hodge decomposition (see (18)), the map $x\mapsto F_s^\bullet\colon S^+\xrightarrow{\varphi} G_{\mathbf{d}}(V)$ is injective. Let $G$ be the smallest algebraic subgroup of $\mathrm{GL}(V)$ such that

$$h(\mathbb{S})\subset G,\quad \text{all } h\in S^+ \tag{25}$$

(take $G$ to be the intersection of the algebraic subgroups of $\mathrm{GL}(V)$ with this property), and let $h_o\in S^+$. For any $g\in G(\mathbb{R})^+$, $gh_og^{-1}\in S^+$, and it can be shown that the map $g\mapsto g\cdot h_o\cdot g^{-1}\colon G(\mathbb{R})^+\to S^+$ is surjective:

$$S^+=G(\mathbb{R})^+\cdot h_o.$$

The subgroup $K_o$ of $G(\mathbb{R})^+$ fixing $h_o$ is closed, and so $G(\mathbb{R})^+/K_o$ is a smooth (in fact, real analytic) manifold. Therefore, $S^+$ acquires the structure of a smooth manifold from

$$S^+=(G(\mathbb{R})^+/K_o)\cdot h_o\cong G(\mathbb{R})^+/K_o.$$

Let $\mathfrak{g}=\mathrm{Lie}(G)$. From $\mathbb{S}\xrightarrow{h_o}G\xrightarrow{\mathrm{Ad}}\mathfrak{g}\subset\mathrm{End}(V)$, we obtain hodge structures on $\mathfrak{g}$ and $\mathrm{End}(V)$. Clearly, $\mathfrak{g}^{00}=\mathrm{Lie}(K_o)$ and so $T_{h_o}S^+\cong\mathfrak{g}/\mathfrak{g}^{00}$. In the diagram,

$$\begin{tikzcd} T_{h_o}S^+\cong \mathfrak{g}/\mathfrak{g}^{00} \arrow[r, hook] \arrow[d, "(23)"', "\cong"] & \mathrm{End}(V)/\mathrm{End}(V)^{00} \arrow[d, "(23)"', "\cong"] \\ \mathfrak{g}(\mathbb{C})/F^0 \arrow[r, hook] & \mathrm{End}(V(\mathbb{C}))/F^0 \cong T_{h_o}G_{\mathbf{d}}(V). \end{tikzcd} \tag{26}$$

the map from top-left to bottom-right is $(d\varphi)_{h_o}$, which therefore maps $T_{h_o}S^+$ onto a complex subspace of $T_{h_o}G_{\mathbf{d}}(V)$. Since this is true for all $h_o\in S^+$, we see that $\varphi$ identifies $S^+$ with an almost-complex submanifold $G_{\mathbf{d}}(V)$. It can be shown that this almost-complex structure is integrable, and so provides $S^+$ with a complex structure for which $\varphi$ is holomorphic. Clearly, this is the only (almost-)complex structure for which this is true.

(b) See Deligne 1979, 1.1.

(c) Given an irreducible hermitian symmetric domain $D$, choose a faithful self-dual representation $G \to \mathrm{GL}(V)$ of the algebraic group $G$ associated with $D$ (as in 1.7). Because $V$ is self-dual, there is a nondegenerate bilinear form $t_0$ on $V$ fixed by $G$. Apply Theorem 2.1 to find a set of tensors $T$ such that $G$ is the subgroup of $\mathrm{GL}(V)$ fixing the $t \in T$. Let $h_o$ be the composite $\mathbb{S} \xrightarrow{z \mapsto z/\bar{z}} U_1 \xrightarrow{u_o} \mathrm{GL}(V)$ with $u_o$ as in (1.9). Then, $h_o$ defines a hodge structure on $V$ for which the $t \in T$ are hodge tensors and $t_o$ is a polarization. One can check that $D$ is naturally identified with the component of $S(d,T)^+$ containing this hodge structure. □

REMARK 2.15. The map $S^+ \to G_{\mathbf{d}}(V)$ in the proof is an embedding of smooth manifolds (injective smooth map that is injective on tangent spaces and maps $S^+$ homeomorphically onto its image). Therefore, if a smooth map $T \to G_{\mathbf{d}}(V)$ factors into

$$T \xrightarrow{\alpha} S^+ \longrightarrow G_{\mathbf{d}}(V),$$

then $\alpha$ will be smooth. Moreover, if the map $T \to G_{\mathbf{d}}(V)$ is defined by a holomorphic family of hodge structures on $T$, and it factors through $S^+$, then $\alpha$ will be holomorphic.

ASIDE 2.16. As we noted in (2.5), for a nonsingular projective variety $V$ over $\mathbb{C}$, the cohomology group $H^n(V(\mathbb{C}),\mathbb{Q})$ has a natural hodge structure of weight $n$. Now consider a regular map $\pi\colon V \to S$ of nonsingular varieties whose fibres $V_s$ ($s \in S$) are nonsingular projective varieties of constant dimension. The vector spaces $H^n(V_s,\mathbb{Q})$ form a local system of $\mathbb{Q}$-vector spaces on $S$, and Griffiths showed that the hodge structures on them form a variation of hodge structures in a slightly more general sense than that defined above (Voisin 2002, Proposition 10.12).

NOTES. Theorem 2.14 is taken from Deligne 1979.

## 3. Locally symmetric varieties

In this section, we study quotients of hermitian symmetric domains by certain discrete groups.

### Quotients of hermitian symmetric domains by discrete groups.

PROPOSITION 3.1. *Let $D$ be a hermitian symmetric domain, and let $\Gamma$ be a discrete subgroup of $\mathrm{Hol}(D)^+$. If $\Gamma$ is torsion free, then $\Gamma$ acts freely on $D$, and there is a unique complex structure on $\Gamma\backslash D$ for which the quotient map $\pi\colon D \to \Gamma\backslash D$ is a local isomorphism. Relative to this structure, a map $\varphi$ from $\Gamma\backslash D$ to a second complex manifold is holomorphic if and only if $\varphi \circ \pi$ is holomorphic.*

PROOF. Let $\Gamma$ be a discrete subgroup of $\mathrm{Hol}(D)^+$. According to (1.5, 1.6), the stabilizer $K_p$ of any point $p \in D$ is compact and $g \mapsto gp\colon \mathrm{Hol}(D)^+/K_p \to D$ is a homeomorphism, and so (MF, 2.5):

(a) for any $p \in D$, $\{g \in \Gamma \mid gp = p\}$ is finite;
(b) for any $p \in D$, there exists a neighbourhood $U$ of $p$ such that, for $g \in \Gamma$, $gU$ is disjoint from $U$ unless $gp = p$;
(c) for any points $p, q \in D$ not in the same $\Gamma$-orbit, there exist neighbourhoods $U$ of $p$ and $V$ of $q$ such that $gU \cap V = \emptyset$ for all $g \in \Gamma$.

Assume $\Gamma$ is torsion free. Then the group in (a) is trivial, and so $\Gamma$ acts freely on $D$. Endow $\Gamma\backslash D$ with the quotient topology. If $U$ and $V$ are as in (c) , then $\pi U$ and $\pi V$ are disjoint neighbourhoods of $\pi p$ and $\pi q$, and so $\Gamma\backslash D$ is separated. Let $q \in \Gamma\backslash D$, and let $p \in \pi^{-1}(q)$. If $U$ is as in (b), then the restriction of $\pi$ to $U$ is a homeomorphism $U \to \pi U$, and it follows that $\Gamma\backslash D$ a manifold.

Define a $\mathbb{C}$-valued function $f$ on an open subset $U$ of $\Gamma\backslash D$ to be holomorphic if $f \circ \pi$ is holomorphic on $\pi^{-1}U$. The holomorphic functions form a sheaf on $\Gamma\backslash D$ for which $\pi$ is a local isomorphism of ringed spaces. Therefore, the sheaf defines a complex structure on $\Gamma\backslash D$ for which $\pi$ is a local isomorphism of complex manifolds.

Finally, let $\varphi\colon \Gamma\backslash D \to M$ be a map such that $\varphi \circ \pi$ is holomorphic, and let $f$ be a holomorphic function on an open subset $U$ of $M$. Then $f \circ \varphi$ is holomorphic because $f \circ \varphi \circ \pi$ is holomorphic, and so $\varphi$ is holomorphic. □

When $\Gamma$ is torsion free, we often write $D(\Gamma)$ for $\Gamma\backslash D$ regarded as a complex manifold. In this case, $D$ is the universal covering space of $D(\Gamma)$ and $\Gamma$ is the group of covering transformations; moreover, for any point $p$ of $D$, the map

$$g \mapsto [\text{image under } \pi \text{ of any path from } p \text{ to } gp]\colon \Gamma \to \pi_1(D(\Gamma), \pi p)$$

is an isomorphism (Hatcher 2002, 1.40).

**Subgroups of finite covolume.** We shall only be interested in quotients of $D$ by "big" discrete subgroups $\Gamma$ of $\mathrm{Aut}(D)^+$. This condition is conveniently expressed by saying that $\Gamma\backslash D$ has finite volume. By definition, $D$ has a riemannian metric $g$ and hence a volume element $\Omega$: in local coordinates

$$\Omega = \sqrt{\det(g_{ij}(x))}dx^1 \wedge \ldots \wedge dx^n.$$

Since $g$ is invariant under $\Gamma$, so also is $\Omega$, and so it passes to the quotient $\Gamma\backslash D$. The condition is that $\int_{\Gamma\backslash D} \Omega < \infty$.

For example, let $D = \mathcal{H}_1$ and let $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$. Then

$$F = \{z \in \mathcal{H}_1 \mid |z| > 1, \quad -\tfrac{1}{2} < \Re z < \tfrac{1}{2}\}$$

is a fundamental domain for $\Gamma$ and

$$\int_{\Gamma\backslash D} \Omega = \iint_F \frac{dxdy}{y^2} \leq \int_{\sqrt{3}/2}^{\infty} \int_{-1/2}^{1/2} \frac{dxdy}{y^2} = \int_{\sqrt{3}/2}^{\infty} \frac{dy}{y^2} < \infty.$$

On the other hand, the quotient of $\mathcal{H}_1$ by the group of translations $z \mapsto z + n$, $n \in \mathbb{Z}$, has infinite volume, as does the quotient of $\mathcal{H}_1$ by the trivial group.

A real Lie group $G$ has a left invariant volume element, which is unique up to a positive constant (cf. Boothby 1975, VI 3.5). A discrete subgroup $\Gamma$ of $G$ is said to have ***finite covolume*** if $\Gamma\backslash G$ has finite volume. For a torsion free discrete subgroup $\Gamma$ of $\mathrm{Hol}(D)^+$, an application of Fubini's theorem shows that $\Gamma\backslash\mathrm{Hol}(D)^+$ has finite volume if and only if $\Gamma\backslash D$ has finite volume (Witte 2001, Exercise 1.27).

**Arithmetic subgroups.** Two subgroups $S_1$ and $S_2$ of a group $H$ are ***commensurable*** if $S_1 \cap S_2$ has finite index in both $S_1$ and $S_2$. For example, two infinite cyclic subgroups $\mathbb{Z}a$ and $\mathbb{Z}b$ of $\mathbb{R}$ are commensurable if and only if $a/b \in \mathbb{Q}^\times$. Commensurability is an equivalence relation.

Let $G$ be an algebraic group over $\mathbb{Q}$. A subgroup $\Gamma$ of $G(\mathbb{Q})$ is ***arithmetic*** if it is commensurable with $G(\mathbb{Q}) \cap \mathrm{GL}_n(\mathbb{Z})$ for some embedding $G \hookrightarrow \mathrm{GL}_n$. It is then commensurable with $G(\mathbb{Q}) \cap \mathrm{GL}_{n'}(\mathbb{Z})$ for every embedding $G \hookrightarrow \mathrm{GL}_{n'}$ (Borel 1969, 7.13).

PROPOSITION 3.2. *Let $\rho\colon G \to G'$ be a surjective homomorphism of algebraic groups over $\mathbb{Q}$. If $\Gamma \subset G(\mathbb{Q})$ is arithmetic, then so also is $\rho(\Gamma) \subset G'(\mathbb{Q})$.*

PROOF. Borel 1969, 8.9, 8.11, or Platonov and Rapinchuk 1994, Theorem 4.1, p204. □

An arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$ is obviously discrete in $G(\mathbb{R})$, but it need not have finite covolume; for example, $\Gamma = \{\pm 1\}$ is an arithmetic subgroup of $\mathbb{G}_m(\mathbb{Q})$ of infinite covolume in $\mathbb{R}^\times$. Thus, if $\Gamma$ is to have finite covolume, there can be no nonzero homomorphism $G \to \mathbb{G}_m$. For reductive groups, this condition is also sufficient.

THEOREM 3.3. *Let $G$ be a reductive group over $\mathbb{Q}$, and let $\Gamma$ be an arithmetic subgroup of $G(\mathbb{Q})$.*

(a) *The space $\Gamma\backslash G(\mathbb{R})$ has finite volume if and only if* $\mathrm{Hom}(G, \mathbb{G}_m) = 0$ *(in particular, $\Gamma\backslash G(\mathbb{R})$ has finite volume if $G$ is semisimple).*[7]
(b) *The space $\Gamma\backslash G(\mathbb{R})$ is compact if and only if* $\mathrm{Hom}(G, \mathbb{G}_m) = 0$ *and $G(\mathbb{Q})$ contains no unipotent element (other than 1).*

PROOF. Borel 1969, 13.2, 8.4, or Platonov and Rapinchuk 1994, Theorem 4.13, p213, Theorem 4.12, p210. [The intuitive reason for the condition in (b) is that the rational unipotent elements correspond to cusps (at least in the case of $\mathrm{SL}_2$ acting on $\mathcal{H}_1$), and so no rational unipotent elements means no cusps.] □

EXAMPLE 3.4. Let $B$ be a quaternion algebra over $\mathbb{Q}$ such that $B \otimes_{\mathbb{Q}} \mathbb{R} \approx M_2(\mathbb{R})$, and let $G$ be the algebraic group over $\mathbb{Q}$ such that $G(\mathbb{Q})$ is the group of elements in $B$ of norm 1. The choice of an isomorphism $B \otimes_{\mathbb{Q}} \mathbb{R} \to M_2(\mathbb{R})$ determines an isomorphism $G(\mathbb{R}) \to \mathrm{SL}_2(\mathbb{R})$, and hence an action of $G(\mathbb{R})$ on $\mathcal{H}_1$. Let $\Gamma$ be an arithmetic subgroup of $G(\mathbb{Q})$.

If $B \approx M_2(\mathbb{Q})$, then $G \approx \mathrm{SL}_2$, which is semisimple, and so $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ (hence also $\Gamma\backslash\mathcal{H}_1$) has finite volume. However, $\mathrm{SL}_2(\mathbb{Q})$ contains the unipotent element $\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)$, and so $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ is not compact.

If $B \not\approx M_2(\mathbb{Q})$, it is a division algebra, and so $G(\mathbb{Q})$ contains no unipotent element $\neq 1$ (for otherwise $B^\times$ would contain a nilpotent element). Therefore, $\Gamma\backslash G(\mathbb{R})$ (hence also $\Gamma\backslash\mathcal{H}_1$) is compact

Let $k$ be a subfield of $\mathbb{C}$. An automorphism $\alpha$ of a $k$-vector space $V$ is said to be ***neat*** if its eigenvalues in $\mathbb{C}$ generate a torsion free subgroup of $\mathbb{C}^\times$ (which implies that $\alpha$ does not have finite order). Let $G$ be an algebraic group over $\mathbb{Q}$. An element $g \in G(\mathbb{Q})$ is ***neat*** if $\rho(g)$ is neat for one faithful representation $G \hookrightarrow \mathrm{GL}(V)$, in which case $\rho(g)$ is neat for every representation $\rho$ of $G$ defined over a subfield of $\mathbb{C}$ (apply Waterhouse 1979, 3.5). A subgroup of $G(\mathbb{Q})$ is ***neat*** if all its elements are.

PROPOSITION 3.5. *Let $G$ be an algebraic group over $\mathbb{Q}$, and let $\Gamma$ be an arithmetic subgroup of $G(\mathbb{Q})$. Then, $\Gamma$ contains a neat subgroup $\Gamma'$ of finite index. Moreover, $\Gamma'$ can be defined by congruence conditions (i.e., for some embedding $G \hookrightarrow \mathrm{GL}_n$ and integer $N$, $\Gamma' = \{g \in \Gamma \mid g \equiv 1 \bmod N\}$).*

[7] Recall (cf. the Notations) that $\mathrm{Hom}(G, \mathbb{G}_m) = 0$ means that there is no nonzero homomorphism $G \to \mathbb{G}_m$ *defined over* $\mathbb{Q}$.

PROOF. Borel 1969, 17.4. □

Let $H$ be a connected real Lie group. A subgroup $\Gamma$ of $H$ is ***arithmetic*** if there exists an algebraic group $G$ over $\mathbb{Q}$ and an arithmetic subgroup $\Gamma_0$ of $G(\mathbb{Q})$ such that $\Gamma_0 \cap G(\mathbb{R})^+$ maps onto $\Gamma$ under a surjective homomorphism $G(\mathbb{R})^+ \to H$ with compact kernel.

PROPOSITION 3.6. *Let $H$ be a semisimple real Lie group that admits a faithful finite-dimensional representation. Every arithmetic subgroup $\Gamma$ of $H$ is discrete of finite covolume, and it contains a torsion free subgroup of finite index.*

PROOF. Let $\alpha\colon G(\mathbb{R})^+ \twoheadrightarrow H$ and $\Gamma_0 \subset G(\mathbb{Q})$ be as in the definition of arithmetic subgroup. Because $\mathrm{Ker}(\alpha)$ is compact, $\alpha$ is proper (Bourbaki 1989, I 10.3) and, in particular, closed. Because $\Gamma_0$ is discrete in $G(\mathbb{R})$, there exists an open $U \subset G(\mathbb{R})^+$ whose intersection with $\Gamma_0$ is exactly the kernel of $\Gamma_0 \cap G(\mathbb{R})^+ \to \Gamma$. Now $\alpha(G(\mathbb{R})^+ \smallsetminus U)$ is closed in $H$, and its complement intersects $\Gamma$ in $\{1_\Gamma\}$. Therefore, $\Gamma$ is discrete in $H$. It has finite covolume because $\Gamma_0 \backslash G(\mathbb{R})^+$ maps onto $\Gamma \backslash H$ and we can apply (3.3a). Let $\Gamma_1$ be a neat subgroup of $\Gamma_0$ of finite index (3.5). The image of $\Gamma_1$ in $H$ has finite index in $\Gamma$, and its image under any faithful representation of $H$ is torsion free. □

REMARK 3.7. There are many nonarithmetic discrete subgroup in $\mathrm{SL}_2(\mathbb{R})$ of finite covolume. According to the Riemann mapping theorem, every compact riemann surface of genus $g \geq 2$ is the quotient of $\mathcal{H}_1$ by a discrete subgroup of $\mathrm{PGL}_2(\mathbb{R})^+$ acting freely on $\mathcal{H}_1$. Since there are continuous families of such riemann surfaces, this shows that there are uncountably many discrete cocompact subgroups in $\mathrm{PGL}_2(\mathbb{R})^+$ (therefore also in $\mathrm{SL}_2(\mathbb{R})$), but there only countably many arithmetic subgroups.

The following (Fields medal) theorem of Margulis shows that $\mathrm{SL}_2$ is exceptional in this regard: let $\Gamma$ be a discrete subgroup of finite covolume in a noncompact simple real Lie group $H$; then $\Gamma$ is arithmetic unless $H$ is isogenous to $\mathrm{SO}(1,n)$ or $\mathrm{SU}(1,n)$ (see Witte 2001, 6.21 for a discussion of the theorem). Note that, because $\mathrm{SL}_2(\mathbb{R})$ is isogenous to $\mathrm{SO}(1,2)$, the theorem doesn't apply to it.

**Brief review of algebraic varieties.** Let $k$ be a field. An ***affine $k$-algebra*** is a finitely generated $k$-algebra $A$ such that $A \otimes_k k^{\mathrm{al}}$ is reduced (i.e., has no nilpotents). Such an algebra is itself reduced, and when $k$ is perfect every reduced finitely generated $k$-algebra is affine.

Let $A$ be an affine $k$-algebra. Define $\mathrm{specm}(A)$ to be the set of maximal ideals in $A$ endowed with the topology having as basis $D(f)$, $D(f) = \{\mathfrak{m} \mid f \notin \mathfrak{m}\}$, $f \in A$. There is a unique sheaf of $k$-algebras $\mathcal{O}$ on $\mathrm{specm}(A)$ such that $\mathcal{O}(D(f)) = A_f$ for all $f$. Here $A_f$ is the algebra obtained from $A$ by inverting $f$. Any ringed space isomorphic to a ringed space of the form

$$\mathrm{Specm}(A) = (\mathrm{specm}(A), \mathcal{O})$$

is called an ***affine variety*** over $k$. The stalk at $\mathfrak{m}$ is the local ring $A_\mathfrak{m}$, and so $\mathrm{Specm}(A)$ is a locally ringed space.

This all becomes much more familiar when $k$ is algebraically closed. When we write $A = k[X_1, \ldots, X_n]/\mathfrak{a}$, the space $\mathrm{specm}(A)$ becomes identified with the zero set of $\mathfrak{a}$ in $k^n$ endowed with the zariski topology, and $\mathcal{O}$ becomes identified with the sheaf of $k$-valued functions on $\mathrm{specm}(A)$ locally defined by polynomials.

A topological space $V$ with a sheaf of $k$-algebras $\mathcal{O}$ is a ***prevariety*** over $k$ if there exists a finite covering $(U_i)$ of $V$ by open subsets such that $(U_i, \mathcal{O}|U_i)$ is an affine variety over $k$ for all $i$. A ***morphism of prevarieties over*** $k$ is simply a morphism of ringed spaces of $k$-algebras. A prevariety $V$ over $k$ is ***separated*** if, for all pairs of morphisms of $k$-prevarieties $\alpha, \beta \colon Z \rightrightarrows V$, the subset of $Z$ on which $\alpha$ and $\beta$ agree is closed. A ***variety over*** $k$ is a separated prevariety over $k$.

Alternatively, the varieties over $k$ are precisely the ringed spaces obtained from geometrically-reduced separated schemes of finite type over $k$ by deleting the nonclosed points.

A morphism of algebraic varieties is also called a ***regular map***, and the elements of $\mathcal{O}(U)$ are called the ***regular functions*** on $U$.

For the variety approach to algebraic geometry, see AG, and for the scheme approach, see Hartshorne 1977.

**Algebraic varieties versus complex manifolds.**

*The functor from nonsingular algebraic varieties to complex manifolds.* For a nonsingular variety $V$ over $\mathbb{C}$, $V(\mathbb{C})$ has a natural structure as a complex manifold. More precisely:

PROPOSITION 3.8. *There is a unique functor $(V, \mathcal{O}_V) \mapsto (V^{an}, \mathcal{O}_{V^{an}})$ from nonsingular varieties over $\mathbb{C}$ to complex manifolds with the following properties:*

(a) *as sets, $V = V^{an}$, every zariski-open subset is open for the complex topology, and every regular function is holomorphic;*
(b) *if $V = \mathbb{A}^n$, then $V^{an} = \mathbb{C}^n$ with its natural structure as a complex manifold;*
(c) *if $\varphi \colon V \to W$ is étale, then $\varphi^{an} \colon V^{an} \to W^{an}$ is a local isomorphism.*

PROOF. A regular map $\varphi \colon V \to W$ is étale if the map $d\varphi_p \colon T_pV \to T_pW$ is an isomorphism for all $p \in V$. Note that conditions (a,b,c) determine the complex-manifold structure on any open subvariety of $\mathbb{A}^n$ and also on any variety $V$ that admits an étale map to an open subvariety of $\mathbb{A}^n$. Since every nonsingular variety admits a zariski-open covering by such $V$ (AG 5.27), this shows that there exists at most one functor satisfying (a,b,c), and suggests how to define it. □

Obviously, a regular map $\varphi \colon V \to W$ is determined by $\varphi^{\mathrm{an}} \colon V^{\mathrm{an}} \to W^{\mathrm{an}}$, but not every holomorphic map $V^{\mathrm{an}} \to W^{\mathrm{an}}$ is regular. For example, $z \mapsto e^z \colon \mathbb{C} \to \mathbb{C}$ is not regular. Moreover, a complex manifold need not arise from a nonsingular algebraic variety, and two nonsingular varieties $V$ and $W$ can be isomorphic as complex manifolds without being isomorphic as algebraic varieties (Shafarevich 1994, VIII 3.2). In other words, the functor $V \mapsto V^{\mathrm{an}}$ is faithful, but it is neither full nor essentially surjective on objects.

REMARK 3.9. The functor $V \mapsto V^{\mathrm{an}}$ can be extended to all algebraic varieties once one has the notion of a "complex manifold with singularities". This is called a ***complex space***. For holomorphic functions $f_1, \ldots, f_r$ on a connected open subset $U$ of $\mathbb{C}^n$, let $V(f_1, \ldots, f_r)$ denote the set of common zeros of the $f_i$ in $U$; one endows $V(f_1, \ldots, f_r)$ with a natural structure of ringed space, and then defines a complex space to be a ringed space $(S, \mathcal{O}_S)$ that is locally isomorphic to one of this form (Shafarevich 1994, VIII 1.5).

*Necessary conditions for a complex manifold to be algebraic.*

3.10. Here are two necessary conditions for a complex manifold $M$ to arise from an algebraic variety.

(a) It must be possible to embed $M$ as an open submanifold of a compact complex manfold $M^*$ in such a way that the boundary $M^* \smallsetminus M$ is a finite union of manifolds of dimension $\dim M - 1$.
(b) If $M$ is compact, then the field of meromorphic functions on $M$ must have transcendence degree $\dim M$ over $\mathbb{C}$.

The necessity of (a) follows from Hironaka's theorem on the resolution of singularities, which shows that every nonsingular variety $V$ can be embedded as an open subvariety of a complete nonsingular variety $V^*$ in such a way that the boundary $V^* \smallsetminus V$ is a divisor with normal crossings (see p293), and the necessity of (b) follows from the fact that, when $V$ is complete and nonsingular, the field of meromorphic functions on $V^{\text{an}}$ coincides with the field of rational functions on $V$ (Shafarevich 1994, VIII 3.1).

Here is one positive result: the functor

$$\{\text{projective nonsingular curves over } \mathbb{C}\} \to \{\text{compact riemann surfaces}\}$$

is an equivalence of categories (see MF, pp88-91, for a discussion of this theorem). Since the proper zariski-closed subsets of algebraic curves are the finite subsets, we see that for riemann surfaces the condition (3.10a) is also sufficient: a riemann surface $M$ is algebraic if and only if it is possible to embed $M$ in a compact riemann surface $M^*$ in such a way that the boundary $M^* \smallsetminus M$ is finite. The maximum modulus principle (Cartan 1963, VI 4.4) shows that a holomorphic function on a connected compact riemann surface is constant. Therefore, if a connected riemann surface $M$ is algebraic, then every bounded holomorphic function on $M$ is constant. We conclude that $\mathcal{H}_1$ does not arise from an algebraic curve, because the function $z \mapsto \frac{z-i}{z+i}$ is bounded, holomorphic, and nonconstant.

For any lattice $\Lambda$ in $\mathbb{C}$, the Weierstrass $\wp$ function and its derivative embed $\mathbb{C}/\Lambda$ into $\mathbb{P}^2(\mathbb{C})$ (as an elliptic curve). However, for a lattice $\Lambda$ in $\mathbb{C}^2$, the field of meromorphic functions on $\mathbb{C}^2/\Lambda$ will usually have transcendence degree $< 2$, and so $\mathbb{C}^2/\Lambda$ is not an algebraic variety. For quotients of $\mathbb{C}^g$ by a lattice $\Lambda$, condition (3.10b) is sufficient for algebraicity (Mumford 1970, p35).

*Projective manifolds and varieties.* A complex manifold (resp. algebraic variety) is ***projective*** if it is isomorphic to a closed submanifold (resp. closed subvariety) of a projective space. The first truly satisfying theorem in the subject is the following:

THEOREM 3.11 (Chow 1949). *Every projective complex manifold has a unique structure of a nonsingular projective algebraic variety, and every holomorphic map of projective complex manifolds is regular for these structures. (Moreover, a similar statement holds for complex spaces.)*

PROOF. See Shafarevich 1994, VIII 3.1 (for the manifold case). □

In other words, the functor $V \mapsto V^{\text{an}}$ is an equivalence from the category of (nonsingular) projective algebraic varieties to the category of projective complex (manifolds) spaces.

### The theorem of Baily and Borel.

THEOREM 3.12 (Baily and Borel 1966). *Let* $D(\Gamma) = \Gamma\backslash D$ *be the quotient of a hermitian symmetric domain by a torsion free arithmetic subgroup* $\Gamma$ *of* $\mathrm{Hol}(D)^+$.

*Then $D(\Gamma)$ has a canonical realization as a zariski-open subset of a projective algebraic variety $D(\Gamma)^*$. In particular, it has a canonical structure as an algebraic variety.*

Recall the proof for $D = \mathcal{H}_1$. Set $\mathcal{H}_1^* = \mathcal{H}_1 \cup \mathbb{P}^1(\mathbb{Q})$ (rational points on the real axis plus the point $i\infty$). Then $\Gamma$ acts on $\mathcal{H}_1^*$, and the quotient $\Gamma\backslash\mathcal{H}_1^*$ is a compact riemann surface. One can then show that the modular forms of a sufficiently high weight embed $\Gamma\backslash\mathcal{H}_1^*$ as a closed submanifold of a projective space. Thus $\Gamma\backslash\mathcal{H}_1^*$ is algebraic, and as $\Gamma\backslash\mathcal{H}_1$ omits only finitely many points of $\Gamma\backslash\mathcal{H}_1^*$, it is automatically a zariski-open subset of $\Gamma\backslash\mathcal{H}_1^*$. The proof in the general case is similar, but is much more difficult. Briefly, $D(\Gamma)^* = \Gamma\backslash D^*$ where $D^*$ is the union of $D$ with certain "rational boundary components" endowed with the Satake topology; again, the automorphic forms of a sufficiently high weight map $\Gamma\backslash D^*$ isomorphically onto a closed subvariety of a projective space, and $\Gamma\backslash D$ is a zariski-open subvariety of $\Gamma\backslash D^*$.

For the Siegel upper half space $\mathcal{H}_g$, the compactification $\mathcal{H}_g^*$ was introduced by Satake (1956) in order to give a geometric foundation to certain results of Siegel (1939), for example, that the space of holomorphic modular forms on $\mathcal{H}_g$ of a fixed weight is finite dimensional, and that the meromorphic functions on $\mathcal{H}_g$ obtained as the quotient of two modular forms of the same weight form an algebraic function field of transcendence degree $g(g+1)/2 = \dim\mathcal{H}_g$ over $\mathbb{C}$.

That the quotient $\Gamma\backslash\mathcal{H}_g^*$ of $\mathcal{H}_g^*$ by an arithmetic group $\Gamma$ has a projective embedding by modular forms, and hence is a projective variety, was proved in Baily 1958, Cartan 1958, and Satake and Cartan 1958.

The construction of $\mathcal{H}_g^*$ depends on the existence of fundamental domains for the arithmetic group $\Gamma$ acting on $\mathcal{H}_g$. Weil (1958) used reduction theory to construct fundamental sets (a notion weaker than fundamental domain) for the domains associated with certain classical groups (groups of automorphisms of semsimple $\mathbb{Q}$-algebras with, or without, involution), and Satake (1960) applied this to construct compactifications of these domains. Borel and Harish-Chandra developed a reduction theory for general semisimple groups (Borel and Harish-Chandra 1962; Borel 1962), which then enabled Baily and Borel (1966) to obtain the above theorem in complete generality.

The only source for the proof is the original paper, although some simplifications to the proof are known.

REMARK 3.13. (a) The variety $D(\Gamma)^*$ is usually very singular. The boundary $D(\Gamma)^* \smallsetminus D(\Gamma)$ has codimension $\geq 2$, provided $\mathrm{PGL}_2$ is not a quotient of the $\mathbb{Q}$-group $G$ giving rise to $\Gamma$.

(b) The variety $D(\Gamma)^* = \mathrm{Proj}(\bigoplus_{n\geq 0} A_n)$ where $A_n$ is the vector space of automorphic forms for the $n^{\text{th}}$ power of the canonical automorphy factor (Baily and Borel 1966, 10.11). It follows that, if $\mathrm{PGL}_2$ is not a quotient of $G$, then $D(\Gamma)^* = \mathrm{Proj}(\bigoplus_{n\geq 0} H^0(D(\Gamma), \omega^n))$ where $\omega$ is the sheaf of algebraic differentials of maximum degree on $D(\Gamma)$. Without the condition on $G$, there is a similar description of $D(\Gamma)^*$ in terms of differentials with logarithmic poles (Brylinski 1983, 4.1.4; Mumford 1977).

(b) When $D(\Gamma)$ is compact, Theorem 3.12 follows from the Kodaira embedding theorem (Wells 1980, VI 4.1, 1.5). Nadel and Tsuji (1988, 3.1) extended this to those $D(\Gamma)$ having boundary of dimension 0, and Mok and Zhong (1989) give an alternative proof of Theorem 3.12, but without the information on the boundary given by the original proof.

An algebraic variety $D(\Gamma)$ arising as in the theorem is called a ***locally symmetric variety*** (or an ***arithmetic locally symmetric variety***, or an ***arithmetic variety***, but not yet a Shimura variety).

**The theorem of Borel.**

THEOREM 3.14 (Borel 1972). *Let $D(\Gamma)$ and $D(\Gamma)^*$ be as in (3.12) — in particular, $\Gamma$ is torsion free and arithmetic. Let $V$ be a nonsingular quasi-projective variety over $\mathbb{C}$. Then every holomorphic map $f\colon V^{an} \to D(\Gamma)^{an}$ is regular.*

The key step in Borel's proof is the following result:

LEMMA 3.15. *Let $\mathcal{D}_1^{\times}$ be the punctured disk $\{z \mid 0 < |z| < 1\}$. Then every holomorphic map*[8] *$\mathcal{D}_1^{\times r} \times \mathcal{D}_1^{s} \to D(\Gamma)$ extends to a holomorphic map $\mathcal{D}_1^{r+s} \to D(\Gamma)^*$ (of complex spaces).*

The original result of this kind is the big Picard theorem, which, interestingly, was first proved using elliptic modular functions. Recall that the theorem says that if a function $f$ has an essential singularity at a point $p \in \mathbb{C}$, then on any open disk containing $p$, $f$ takes every complex value except possibly one. Therefore, if a holomorphic function $f$ on $\mathcal{D}_1^{\times}$ omits two values in $\mathbb{C}$, then it has at worst a pole at 0, and so extends to a holomorphic function $\mathcal{D}_1 \to \mathbb{P}^1(\mathbb{C})$. This can be restated as follows: every holomorphic function from $\mathcal{D}_1^{\times}$ to $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3 \text{ points}\}$ extends to a holomorphic function from $\mathcal{D}_1$ to the natural compactification $\mathbb{P}^1(\mathbb{C})$ of $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3 \text{ points}\}$. Over the decades, there were various improvements made to this theorem. For example, Kwack (1969) replaced $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{3 \text{ points}\}$ with a more general class of spaces. Borel (1972) verified that Kwack's theorem applies to $D(\Gamma) \subset D(\Gamma)^*$, and extended the result to maps from a product $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^{s}$.

Using the lemma, we can prove the theorem. According Hironaka's (Fields medal) theorem on the resolution of singularities (Hironaka 1964; see also Bravo et al. 2002), we can realize $V$ as an open subvariety of a projective nonsingular variety $V^*$ in such a way that $V^* \smallsetminus V$ is a divisor with normal crossings. This means that, locally for the complex topology, the inclusion $V \hookrightarrow V^*$ is of the form $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^{s} \hookrightarrow \mathcal{D}_1^{r+s}$. Therefore, the lemma shows that $f\colon V^{\mathrm{an}} \to D(\Gamma)^{\mathrm{an}}$ extends to a holomorphic map $V^{*\mathrm{an}} \to D(\Gamma)^*$, which is regular by Chow's theorem (3.11).

COROLLARY 3.16. *The structure of an algebraic variety on $D(\Gamma)$ is unique.*

PROOF. Let $D(\Gamma)$ denote $\Gamma\backslash D$ with the canonical algebraic structure provided by Theorem 3.12, and suppose $\Gamma\backslash D = V^{\mathrm{an}}$ for a second variety $V$. Then the identity map $f\colon V^{\mathrm{an}} \to D(\Gamma)$ is a regular bijective map of nonsingular varieties in characteristic zero, and is therefore an isomorphism (cf. AG 8.19). □

The proof of the theorem shows that the compactification $D(\Gamma) \hookrightarrow D(\Gamma)^*$ has the following property: for any compactification $D(\Gamma) \to D(\Gamma)^{\dagger}$ with $D(\Gamma)^{\dagger} \smallsetminus D(\Gamma)$ a divisor with normal crossings, there is a unique regular map $D(\Gamma)^{\dagger} \to D(\Gamma)^*$ making

[8] Recall that $\mathcal{D}_1$ is the open unit disk. The product $\mathcal{D}_1^{\times r} \times \mathcal{D}_1^{s}$ is obtained from $\mathcal{D}_1^{r+s}$ by removing the first $r$ coordinate hyperplanes.

$$\begin{tikzcd} & D(\Gamma)^{\dagger} \arrow[dd] \\ D(\Gamma) \arrow[ur] \arrow[dr] & \\ & D(\Gamma)^{*} \end{tikzcd}$$

commute. For this reason, $D(\Gamma) \hookrightarrow D(\Gamma)^*$ is often called the ***minimal*** compactification. Other names: ***standard***, ***Satake-Baily-Borel***, ***Baily-Borel.***

ASIDE 3.17. (a) Theorem 3.14 also holds for singular $V$ — in fact, it suffices to show that $f$ becomes regular when restricted to an open dense set of $V$, which we may take to be the complement of the singular locus.

(b) Theorem 3.14 definitely fails without the condition that $\Gamma$ be torsion free. For example, it is false for $\Gamma\backslash\mathcal{H}_1 = \mathbb{A}^1$ — consider $z \mapsto e^z : \mathbb{C} \to \mathbb{C}$.

**Finiteness of the group of automorphisms of $D(\Gamma)$.**

DEFINITION 3.18. A semisimple group $G$ over $\mathbb{Q}$ is said to be of ***compact type*** if $G(\mathbb{R})$ is compact, and it is of ***noncompact type*** if it does not contain a nonzero normal subgroup of compact type.

A semisimple group over $\mathbb{Q}$ is an almost direct product of its minimal connected normal subgroups, and it will be of noncompact type if and only if none of these subgroups is of compact type. In particular, a simply connected or adjoint group is of noncompact type if and only if it has no simple factor of compact type.

We shall need one last result about arithmetic subgroups.

THEOREM 3.19 (Borel density theorem). *Let $G$ be a semisimple group over $\mathbb{Q}$ of noncompact type. Then every arithmetic subgroup $\Gamma$ of $G(\mathbb{Q})$ is zariski-dense in $G$.*

PROOF. Borel 1969, 15.12, or Platonov and Rapinchuk 1994, Theorem 4.10, p205. □

COROLLARY 3.20. *For $G$ as in (3.19), the centralizer of $\Gamma$ in $G(\mathbb{R})$ is $Z(\mathbb{R})$, where $Z$ is the centre of $G$ (as an algebraic group over $\mathbb{Q}$).*

PROOF. The theorem implies that the centralizer of $\Gamma$ in $G(\mathbb{C})$ is $Z(\mathbb{C})$, and $Z(\mathbb{R}) = Z(\mathbb{C}) \cap G(\mathbb{R})$. □

THEOREM 3.21. *Let $D(\Gamma)$ be the quotient of a hermitian symmetric domain $D$ by a torsion free arithmetic group $\Gamma$. Then $D(\Gamma)$ has only finitely many automorphisms.*

PROOF. As $\Gamma$ is a torsion free, $D$ is the universal covering space of $\Gamma\backslash D$ and $\Gamma$ is the group of covering transformations (see p287). An automorphism $\alpha : \Gamma\backslash D \to \Gamma\backslash D$ lifts to an automorphism $\tilde{\alpha} : D \to D$. For any $\gamma \in \Gamma$, $\tilde{\alpha}\gamma\tilde{\alpha}^{-1}$ is a covering transformation, and so lies in $\Gamma$. Conversely, an automorphism of $D$ normalizing $\Gamma$ defines an automorphism of $\Gamma\backslash D$. Thus,

$$\mathrm{Aut}(\Gamma\backslash D) = N/\Gamma, \quad N = \text{normalizer of } \Gamma \text{ in } \mathrm{Aut}(D).$$

The corollary implies that the map $\mathrm{ad}\colon N \to \mathrm{Aut}(\Gamma)$ is injective. The group $\Gamma$ is countable because it is a discrete subgroup of a group that admits a countable basis for its open subsets, and so $N$ is also countable. Because $\Gamma$ is closed in $\mathrm{Aut}(D)$, so also is $N$. Write $N$ as a countable union of its finite subsets. According to the Baire category theorem (MF 1.3) one of the finite sets must have an interior point, and this implies that $N$ is discrete. Because $\Gamma\backslash \mathrm{Aut}(D)$ has finite volume (3.3a), this implies that $\Gamma$ has finite index in $N$.

Alternatively, there is a geometric proof, at least when $\Gamma$ is neat. According to Mumford 1977, Proposition 4.2, $D(\Gamma)$ is then an algebraic variety of logarithmic general type, which implies that its automorphism group is finite (Iitaka 1982, 11.12). □

ASIDE 3.22. In most of this section we have considered only quotients $\Gamma\backslash D$ with $\Gamma$ torsion free. In particular, we disallowed $\Gamma(1)\backslash\mathcal{H}_1$. Typically, if $\Gamma$ has torsion, then $\Gamma\backslash D$ will be singular and some of the above statements will fail for $\Gamma\backslash D$.

NOTES. Borel 1969, Raghunathan 1972, and (eventually) Witte 2001 contain good expositions on discrete subgroups of Lie groups. There is a large literature on the various compactifications of locally symmetric varieties. For overviews, see Satake 2001 and Goresky 2003, and for a detailed description of the construction of toroidal compactifications, which, in contrast to the Baily-Borel compactification, may be smooth and projective, see Ash et al. 1975.

## 4. Connected Shimura varieties

**Congruence subgroups.** Let $G$ be a reductive algebraic group over $\mathbb{Q}$. Choose an embedding $G \hookrightarrow \mathrm{GL}_n$, and define

$$\Gamma(N) = G(\mathbb{Q}) \cap \{g \in \mathrm{GL}_n(\mathbb{Z}) \mid g \equiv I_n \bmod N\}.$$

For example, if $G = \mathrm{SL}_2$, then

$$\Gamma(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) \mid ad - bc = 1, \quad a, d \equiv 1, \quad b, c \equiv 0 \mod N \right\}.$$

A ***congruence subgroup of*** $G(\mathbb{Q})$ is any subgroup containing some $\Gamma(N)$ as a subgroup of finite index. Although $\Gamma(N)$ depends on the choice the embedding, this definition does not (see 4.1 below).

With this terminology, a subgroup of $G(\mathbb{Q})$ is arithmetic if it is commensurable with $\Gamma(1)$. The classical congruence subgroup problem for $G$ asks whether every arithmetic subgroup of $G(\mathbb{Q})$ is congruence, i.e., contains some $\Gamma(N)$. For split simply connected groups other than $\mathrm{SL}_2$, the answer is yes (Matsumoto 1969), but $\mathrm{SL}_2$ and all nonsimply connected groups have many noncongruence arithmetic subgroups (for a discussion of the problem, see Platonov and Rapinchuk 1994, section 9.5). In contrast to arithmetic subgroups, the image of a congruence subgroup under an isogeny of algebraic groups need not be a congruence subgroup.

The ***ring of finite adèles*** is the restricted topological product

$$\mathbb{A}_f = \prod (\mathbb{Q}_\ell \colon \mathbb{Z}_\ell)$$

where $\ell$ runs over the finite primes of $\ell$ (that is, we omit the factor $\mathbb{R}$). Thus, $\mathbb{A}_f$ is the subring of $\prod \mathbb{Q}_\ell$ consisting of the $(a_\ell)$ such that $a_\ell \in \mathbb{Z}_\ell$ for almost all $\ell$, and it is endowed with the topology for which $\prod \mathbb{Z}_\ell$ is open and has the product topology.

Let $V = \operatorname{Specm} A$ be an affine variety over $\mathbb{Q}$. The set of points of $V$ with coordinates in a $\mathbb{Q}$-algebra $R$ is

$$V(R) = \operatorname{Hom}_{\mathbb{Q}}(A, R).$$

When we write

$$A = \mathbb{Q}[X_1, \ldots, X_m]/\mathfrak{a} = \mathbb{Q}[x_1, \ldots, x_m],$$

the map $P \mapsto (P(x_1), \ldots, P(x_m))$ identifies $V(R)$ with

$$\{(a_1, \ldots, a_m) \in R^m \mid f(a_1, \ldots, a_m) = 0, \quad \forall f \in \mathfrak{a}\}.$$

Let $\mathbb{Z}[x_1, \ldots, x_m]$ be the $\mathbb{Z}$-subalgebra of $A$ generated by the $x_i$, and let

$$V(\mathbb{Z}_\ell) = \operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}[x_1, \ldots, x_m], \mathbb{Z}_\ell) = V(\mathbb{Q}_\ell) \cap \mathbb{Z}_\ell^m \qquad (\text{inside } \mathbb{Q}_\ell^m).$$

This set depends on the choice of the generators $x_i$ for $A$, but if $A = \mathbb{Q}[y_1, \ldots, y_n]$, then the $y_i$'s can be expressed as polynomials in the $x_i$ with coefficients in $\mathbb{Q}$, and vice versa. For some $d \in \mathbb{Z}$, the coefficients of these polynomials lie in $\mathbb{Z}[\frac{1}{d}]$, and so

$$\mathbb{Z}[\tfrac{1}{d}][x_1, \ldots, x_m] = \mathbb{Z}[\tfrac{1}{d}][y_1, \ldots, y_n] \qquad (\text{inside } A).$$

It follows that for $\ell \nmid d$, the $y_i$'s give the same set $V(\mathbb{Z}_\ell)$ as the $x_i$'s. Therefore,

$$V(\mathbb{A}_f) = \prod (V(\mathbb{Q}_\ell) \colon V(\mathbb{Z}_\ell))$$

is independent of the choice of generators for[9] $A$.

For an algebraic group $G$ over $\mathbb{Q}$, we define

$$G(\mathbb{A}_f) = \prod (G(\mathbb{Q}_\ell) \colon G(\mathbb{Z}_\ell))$$

similarly. For example,

$$\mathbb{G}_m(\mathbb{A}_f) = \prod (\mathbb{Q}_\ell^\times : \mathbb{Z}_\ell^\times) = \mathbb{A}_f^\times.$$

PROPOSITION 4.1. *For any compact open subgroup $K$ of $G(\mathbb{A}_f)$, $K \cap G(\mathbb{Q})$ is a congruence subgroup of $G(\mathbb{Q})$, and every congruence subgroup arises in this way.*

PROOF. Fix an embedding $G \hookrightarrow \mathrm{GL}_n$. From this we get a surjection $\mathbb{Q}[\mathrm{GL}_n] \to \mathbb{Q}[G]$ (of $\mathbb{Q}$-algebras of regular functions), i.e., a surjection

$$\mathbb{Q}[X_{11}, \ldots, X_{nn}, T]/(\det(X_{ij})T - 1) \to \mathbb{Q}[G],$$

and hence $\mathbb{Q}[G] = \mathbb{Q}[x_{11}, \ldots, x_{nn}, t]$. For this presentation of $\mathbb{Q}[G]$,

$$G(\mathbb{Z}_\ell) = G(\mathbb{Q}_\ell) \cap \mathrm{GL}_n(\mathbb{Z}_\ell) \qquad (\text{inside } \mathrm{GL}_n(\mathbb{Q}_\ell)).$$

For an integer $N > 0$, let

$$K(N) = \textstyle\prod_\ell K_\ell, \quad \text{where} \quad K_\ell = \begin{cases} G(\mathbb{Z}_\ell) & \text{if} \quad \ell \nmid N \\ \{g \in G(\mathbb{Z}_\ell) \mid g \equiv I_n \bmod \ell^{r_\ell}\} & \text{if} \quad r_\ell = \operatorname{ord}_\ell(N). \end{cases}$$

Then $K(N)$ is a compact open subgroup of $G(\mathbb{A}_f)$, and

$$K(N) \cap G(\mathbb{Q}) = \Gamma(N).$$

It follows that the compact open subgroups of $G(\mathbb{A}_f)$ containing $K(N)$ intersect $G(\mathbb{Q})$ exactly in the congruence subgroups of $G(\mathbb{Q})$ containing $\Gamma(N)$. Since every

[9] In a more geometric language, let $\alpha \colon V \hookrightarrow \mathbb{A}_{\mathbb{Q}}^m$ be a closed immersion. The zariski closure $V_\alpha$ of $V$ in $\mathbb{A}_{\mathbb{Z}}^m$ is a model of $V$ flat over $\operatorname{Spec} \mathbb{Z}$. A different closed immersion $\beta$ gives a different flat model $V_\beta$, but for some $d$, the isomorphism $(V_\alpha)_{\mathbb{Q}} \cong V \cong (V_\beta)_{\mathbb{Q}}$ on generic fibres extends to an isomorphism $V_\alpha \to V_\beta$ over $\operatorname{Spec} \mathbb{Z}[\frac{1}{d}]$. For the primes $\ell$ not dividing $d$, the subgroups $V_\alpha(\mathbb{Z}_\ell)$ and $V_\beta(\mathbb{Z}_\ell)$ of $V(\mathbb{Q}_\ell)$ will coincide.

compact open subgroup of $G(\mathbb{A}_f)$ contains $K(N)$ for some $N$, this completes the proof. $\square$

REMARK 4.2. There is a topology on $G(\mathbb{Q})$ for which the congruence subgroups form a fundamental system of neighbourhoods. The proposition shows that this topology coincides with that defined by the diagonal embedding $G(\mathbb{Q}) \subset G(\mathbb{A}_f)$.

EXERCISE 4.3. Show that the image in $\mathrm{PGL}_2(\mathbb{Q})$ of a congruence subgroup in $\mathrm{SL}_2(\mathbb{Q})$ need not be congruence.

**Connected Shimura data.**

DEFINITION 4.4. A ***connected Shimura datum*** is a pair $(G, D)$ consisting of a semisimple algebraic group $G$ over $\mathbb{Q}$ and a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class $D$ of homomorphisms $u\colon U_1 \to G^{\mathrm{ad}}_{\mathbb{R}}$ satisfying the following conditions:

SU1**:** for $u \in D$, only the characters $z, 1, z^{-1}$ occur in the representation of $U_1$ on $\mathrm{Lie}(G^{\mathrm{ad}})_{\mathbb{C}}$ defined by $u$;
SU2**:** for $u \in D$, $\mathrm{ad}u(-1)$ is a Cartan involution on $G^{\mathrm{ad}}$;
SU3**:** $G^{\mathrm{ad}}$ has no $\mathbb{Q}$-factor $H$ such that $H(\mathbb{R})$ is compact.

EXAMPLE 4.5. Let $u\colon U_1 \to \mathrm{PGL}_2(\mathbb{R})$ be the homomorphism sending $z = (a+bi)^2$ to $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \bmod \pm I_2$ (cf. 1.10), and let $D$ be the set of conjugates of this homomorphism, i.e., $D$ is the set of homomorphisms $U_1 \to \mathrm{PGL}_2(\mathbb{R})$ of the form

$$z = (a+bi)^2 \mapsto A \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) A^{-1} \bmod \pm I_2, \quad A \in \mathrm{SL}_2(\mathbb{R}).$$

Then $(\mathrm{SL}_2, D)$ is a Shimura datum (here $\mathrm{SL}_2$ is regarded as a group over $\mathbb{Q}$).

REMARK 4.6. (a) If $u\colon U_1 \to G^{\mathrm{ad}}(\mathbb{R})$ satisfies the conditions SU1,2, then so does any conjugate of it by an element of $G^{\mathrm{ad}}(\mathbb{R})^+$. Thus a pair $(G, u)$ satisfying SU1,2,3 determines a connected Shimura datum. Our definition of connected Shimura datum was phrased so as to avoid $D$ having a distinguished point.

(b) Condition SU3 says that $G$ is of noncompact type (3.18). It is fairly harmless to assume this, because replacing $G$ with its quotient by a connected normal subgroup $N$ such that $N(\mathbb{R})$ is compact changes little. Assuming it allows us to apply the strong approximation theorem when $G$ is simply connected (see 4.16 below).

LEMMA 4.7. *Let $H$ be an adjoint real Lie group, and let $u\colon U_1 \to H$ be a homomorphism satisfying SU1,2. Then the following conditions on $u$ are equivalent:*

(a) $u(-1) = 1$;
(b) *$u$ is trivial, i.e., $u(z) = 1$ for all $z$;*
(c) *$H$ is compact.*

PROOF. (a)$\Leftrightarrow$(b). If $u(-1) = 1$, then $u$ factors through $U_1 \xrightarrow{2} U_1$, and so $z^{\pm 1}$ can not occur in the representation of $U_1$ on $\mathrm{Lie}(H)_{\mathbb{C}}$. Therefore $U_1$ acts trivially on $\mathrm{Lie}(H)_{\mathbb{C}}$, which implies (b). The converse is trivial.

(a)$\Leftrightarrow$(c). We have

$$H \text{ is compact} \overset{1.17a}{\iff} \mathrm{ad}u(-1) = 1 \overset{Z(H)=1}{\iff} u(-1) = 1. \quad \square$$

PROPOSITION 4.8. *To give a connected Shimura datum is the same as to give*

- *a semisimple algebraic group $G$ over $\mathbb{Q}$ of noncompact type,*
- *a hermitian symmetric domain $D$, and*

∘ *an action of* $G(\mathbb{R})^+$ *on* $D$ *defined by a surjective homomorphism* $G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ *with compact kernel.*

PROOF. Let $(G, D)$ be a connected Shimura datum, and let $u \in D$. Decompose $G^{\mathrm{ad}}_{\mathbb{R}}$ into a product of its simple factors: $G^{\mathrm{ad}}_{\mathbb{R}} = H_1 \times \cdots \times H_s$. Correspondingly, $u = (u_1, \ldots, u_s)$ where $u_i$ is the projection of $u$ into $H_i(\mathbb{R})$. Then $u_i = 1$ if $H_i$ is compact (4.7), and otherwise there is an irreducible hermitian symmetric domain $D_i'$ such that $H_i(\mathbb{R})^+ = \mathrm{Hol}(D_i')^+$ and $D_i'$ is in natural one-to-one correspondence with the set $D_i$ of $H_i(\mathbb{R})^+$-conjugates of $u_i$ (see 1.21). The product $D'$ of the $D_i'$ is a hermitian symmetric domain on which $G(\mathbb{R})^+$ acts via a surjective homomorphism $G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+$ with compact kernel. Moreover, there is a natural identification of $D' = \prod D_i'$ with $D = \prod D_i$.

Conversely, let $(G, D, G(\mathbb{R})^+ \to \mathrm{Hol}(D)^+)$ satisfy the conditions in the proposition. Decompose $G^{\mathrm{ad}}_{\mathbb{R}}$ as before, and let $H_{\mathrm{c}}$ (resp. $H_{\mathrm{nc}}$) be the product of the compact (resp. noncompact) factors. The action of $G(\mathbb{R})^+$ on $D$ defines an isomorphism $H_{\mathrm{nc}}(\mathbb{R})^+ \cong \mathrm{Hol}(D)^+$, and $\{u_p \mid p \in D\}$ is an $H_{\mathrm{nc}}(\mathbb{R})^+$-conjugacy class of homomorphisms $U_1 \to H_{\mathrm{nc}}(\mathbb{R})^+$ satisfying SU1,2 (see 1.21). Now

$$\{(1, u_p)\colon U_1 \to H_{\mathrm{c}}(\mathbb{R}) \times H_{\mathrm{nc}}(\mathbb{R}) \mid p \in D\},$$

is a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class of homomorphisms $U_1 \to G^{\mathrm{ad}}(\mathbb{R})$ satisfying SU1,2. □

PROPOSITION 4.9. *Let* $(G, D)$ *be a connected Shimura datum, and let* $X$ *be the* $G^{\mathrm{ad}}(\mathbb{R})$*-conjugacy class of homomorphisms* $\mathbb{S} \to G_{\mathbb{R}}$ *containing* $D$. *Then* $D$ *is a connected component of* $X$, *and the stabilizer of* $D$ *in* $G^{\mathrm{ad}}(\mathbb{R})$ *is* $G^{\mathrm{ad}}(\mathbb{R})^+$.

PROOF. The argument in the proof of (1.5) shows that $X$ is a disjoint union of orbits $G^{\mathrm{ad}}(\mathbb{R})^+h$, each of which is both open and closed in $X$. In particular, $D$ is a connected component of $X$.

Let $H_{\mathrm{c}}$ (resp. $H_{\mathrm{nc}}$) be the product of the compact (resp. noncompact) simple factors of $G_{\mathbb{R}}$. Then $H_{\mathrm{nc}}$ is a connected algebraic group over $\mathbb{R}$ such that $H_{\mathrm{nc}}(\mathbb{R})^+ = \mathrm{Hol}(D)$, and $G(\mathbb{R})^+$ acts on $D$ through its quotient $H_{\mathrm{nc}}(\mathbb{R})^+$. As $H_{\mathrm{c}}(\mathbb{R})$ is connected (Borel 1991, p277), the last part of the proposition follows from (1.7). □

**Definition of a connected Shimura variety.** Let $(G, D)$ be a connected Shimura datum, and regard $D$ as a hermitian symmetric domain with $G(\mathbb{R})^+$ acting on it as in (4.8). Because $G^{\mathrm{ad}}(\mathbb{R})^+ \to \mathrm{Aut}(D)^+$ has compact kernel, the image $\overline{\Gamma}$ of any arithmetic subgroup $\Gamma$ of $G^{\mathrm{ad}}(\mathbb{Q})^+$ in $\mathrm{Aut}(D)^+$ will be arithmetic (this is the definition p289). The kernel of $\Gamma \to \overline{\Gamma}$ is finite. If $\Gamma$ is torsion free, then $\Gamma \cong \overline{\Gamma}$, and so the Baily-Borel and Borel theorems (3.12, 3.14) apply to

$$D(\Gamma) \stackrel{\mathrm{df}}{=} \Gamma\backslash D = \overline{\Gamma}\backslash D.$$

In particular, $D(\Gamma)$ is an algebraic variety, and, for any $\Gamma \supset \Gamma'$, the natural map

$$D(\Gamma) \leftarrow D(\Gamma')$$

is regular.

DEFINITION 4.10. The ***connected Shimura variety*** $\mathrm{Sh}^{\circ}(G, D)$ is the inverse system of locally symmetric varieties $(D(\Gamma))_{\Gamma}$ where $\Gamma$ runs over the torsion-free arithmetic subgroups of $G^{\mathrm{ad}}(\mathbb{Q})^+$ whose inverse image in $G(\mathbb{Q})^+$ is a congruence subgroup.

REMARK 4.11. An element $g$ of $G^{\mathrm{ad}}(\mathbb{Q})^{+}$ defines a holomorphic map $g\colon D \to D$, and hence a map

$$\Gamma\backslash D \to g\Gamma g^{-1}\backslash D.$$

This is again holomorphic (3.1), and hence is regular (3.14). Therefore the group $G^{\mathrm{ad}}(\mathbb{Q})^{+}$ acts on the family $\mathrm{Sh}^{\circ}(G,D)$ (but not on the individual $D(\Gamma)$'s).

LEMMA 4.12. *Write $\pi$ for the homomorphism $G(\mathbb{Q})^{+} \to G^{\mathrm{ad}}(\mathbb{Q})^{+}$. The following conditions on an arithmetic subgroup $\Gamma$ of $G^{\mathrm{ad}}(\mathbb{Q})^{+}$ are equivalent:*

(a) *$\pi^{-1}(\Gamma)$ is a congruence subgroup of $G(\mathbb{Q})^{+}$;*
(b) *$\pi^{-1}(\Gamma)$ contains a congruence subgroup of $G(\mathbb{Q})^{+}$;*
(c) *$\Gamma$ contains the image of a congruence subgroup of $G(\mathbb{Q})^{+}$.*

*Therefore, the varieties $\Gamma\backslash D$ with $\Gamma$ a congruence subgroup of $G(\mathbb{Q})^{+}$ such $\pi(\Gamma)$ is torsion free are cofinal in the family $\mathrm{Sh}^{\circ}(G,D)$.*

PROOF. (a) $\Longrightarrow$ (b). Obvious.

(b) $\Longrightarrow$ (c). Let $\Gamma'$ be a congruence subgroup of $G(\mathbb{Q})^{+}$ contained in $\pi^{-1}(\Gamma)$. Then

$$\Gamma \supset \pi(\pi^{-1}(\Gamma)) \supset \pi(\Gamma').$$

(c) $\Longrightarrow$ (a). Let $\Gamma'$ be a congruence subgroup of $G(\mathbb{Q})^{+}$ such that $\Gamma \supset \pi(\Gamma')$, and consider

$$\pi^{-1}(\Gamma) \supset \pi^{-1}\pi(\Gamma') \supset \Gamma'.$$

Because $\pi(\Gamma')$ is arithmetic (3.2), it is of finite index in $\Gamma$, and it follows that $\pi^{-1}\pi(\Gamma')$ is of finite index in $\pi^{-1}(\Gamma)$. Because $Z(\mathbb{Q})\cdot\Gamma' \supset \pi^{-1}\pi(\Gamma')$ and $Z(\mathbb{Q})$ is finite ($Z$ is the centre of $G$), $\Gamma'$ is of finite index in $\pi^{-1}\pi(\Gamma')$. Therefore, $\Gamma'$ is of finite index in $\pi^{-1}(\Gamma)$, which proves that $\pi^{-1}(\Gamma)$ is congruence. □

REMARK 4.13. The homomorphism $\pi\colon G(\mathbb{Q})^{+} \to G^{\mathrm{ad}}(\mathbb{Q})^{+}$ is usually far from surjective. Therefore, $\pi\pi^{-1}(\Gamma)$ is usually not equal to $\Gamma$, and the family $D(\Gamma)$ with $\Gamma$ a congruence subgroup of $G(\mathbb{Q})^{+}$ is usually much smaller than $\mathrm{Sh}^{\circ}(G,D)$.

EXAMPLE 4.14. (a) $G = \mathrm{SL}_2$, $D = \mathcal{H}_1$. Then $\mathrm{Sh}^{\circ}(G,D)$ is the family of elliptic modular curves $\Gamma\backslash\mathcal{H}_1$ with $\Gamma$ a torsion-free arithmetic subgroup of $\mathrm{PGL}_2(\mathbb{R})^{+}$ containing the image of $\Gamma(N)$ for some $N$.

(b) $G = \mathrm{PGL}_2$, $D = \mathcal{H}_1$. The same as (a), except that now the $\Gamma$ are required to be congruence subgroups of $\mathrm{PGL}_2(\mathbb{Q})$ — there are *many fewer* of these (see 4.3).

(c) Let $B$ be a quaternion algebra over a totally real field $F$. Then

$$B \otimes_{\mathbb{Q}} \mathbb{R} \cong \prod\nolimits_{v\colon F\hookrightarrow\mathbb{R}} B \otimes_{F,v} \mathbb{R}$$

and each $B \otimes_{F,v} \mathbb{R}$ is isomorphic either to the usual quaternions $\mathbb{H}$ or to $M_2(\mathbb{R})$. Let $G$ be the semisimple algebraic group over $\mathbb{Q}$ such that

$$G(\mathbb{Q}) = \mathrm{Ker}(\mathrm{Nm}\colon B^{\times} \to F^{\times}).$$

Then

$$G(\mathbb{R}) \approx \mathbb{H}^{\times 1} \times \cdots \times \mathbb{H}^{\times 1} \times \mathrm{SL}_2(\mathbb{R}) \times \cdots \times \mathrm{SL}_2(\mathbb{R}) \tag{27}$$

where $\mathbb{H}^{\times 1} = \mathrm{Ker}(\mathrm{Nm}\colon \mathbb{H}^{\times} \to \mathbb{R}^{\times})$. Assume that at least one $\mathrm{SL}_2(\mathbb{R})$ occurs (so that $G$ is of noncompact type), and let $D$ be a product of copies of $\mathcal{H}_1$, one for each copy of $\mathrm{SL}_2(\mathbb{R})$. The choice of an isomorphism (27) determines an action of $G(\mathbb{R})$ on $D$ which satisfies the conditions of (4.8), and hence defines a connected Shimura datum. In this case, $D(\Gamma)$ has dimension equal to the number of copies of $M_2(\mathbb{R})$

in the decomposition of $B \otimes_{\mathbb{Q}} \mathbb{R}$. If $B \approx M_2(F)$, then $G(\mathbb{Q})$ has unipotent elements, e.g., $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and so $D(\Gamma)$ is not compact (3.3). In this case the varieties $D(\Gamma)$ are called ***Hilbert modular varieties***. On the other hand, if $B$ is a division algebra, $G(\mathbb{Q})$ has no unipotent elements, and so the $D(\Gamma)$ are compact (as manifolds, hence they are projective as algebraic varieties).

ASIDE 4.15. In the definition of $\mathrm{Sh}^{\circ}(G, D)$, why do we require the inverse images of the $\Gamma$'s in $G(\mathbb{Q})^{+}$ to be congruence? The arithmetic properties of the quotients of hermitian symmetric domains by noncongruence arithmetic subgroups are not well understood even for $D = \mathcal{H}_1$ and $G = \mathrm{SL}_2$. Also, the congruence subgroups turn up naturally when we work adèlically.

**The strong approximation theorem.** Recall that a semisimple group $G$ is said to be simply connected if any isogeny $G' \to G$ with $G'$ connected is an isomorphism. For example, $\mathrm{SL}_2$ is simply connected, but $\mathrm{PGL}_2$ is not.

THEOREM 4.16 (Strong Approximation). *Let $G$ be an algebraic group over $\mathbb{Q}$. If $G$ is semisimple, simply connected, and of noncompact type, then $G(\mathbb{Q})$ is dense in $G(\mathbb{A}_f)$.*

PROOF. Platonov and Rapinchuk 1994, Theorem 7.12, p427. □

REMARK 4.17. Without the conditions on $G$, the theorem fails, as the following examples illustrate:

(a) $\mathbb{G}_m$: the group $\mathbb{Q}^{\times}$ is not dense in $\mathbb{A}_f^{\times}$.
(b) $\mathrm{PGL}_2$: the determinant defines surjections

$$\mathrm{PGL}_2(\mathbb{Q}) \to \mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$$
$$\mathrm{PGL}_2(\mathbb{A}_f) \to \mathbb{A}_f^{\times}/\mathbb{A}_f^{\times 2}$$

and $\mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ is not dense in $\mathbb{A}_f^{\times}/\mathbb{A}_f^{\times 2}$.
(c) $G$ of compact type: because $G(\mathbb{Z})$ is discrete in $G(\mathbb{R})$ (see 3.3), it is finite, and so it is not dense in $G(\hat{\mathbb{Z}})$, which implies that $G(\mathbb{Q})$ is not dense in $G(\mathbb{A}_f)$.

**An adèlic description of $D(\Gamma)$.**

PROPOSITION 4.18. *Let $(G, D)$ be a connected Shimura datum with $G$ simply connected. Let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$, and let*

$$\Gamma = K \cap G(\mathbb{Q})$$

*be the corresponding congruence subgroup of $G(\mathbb{Q})$. The map $x \mapsto [x, 1]$ defines a bijection*

$$\Gamma \backslash D \cong G(\mathbb{Q}) \backslash D \times G(\mathbb{A}_f)/K. \tag{28}$$

*Here $G(\mathbb{Q})$ acts on both $D$ and $G(\mathbb{A}_f)$ on the left, and $K$ acts on $G(\mathbb{A}_f)$ on the right:*

$$q \cdot (x, a) \cdot k = (qx, qak), \quad q \in G(\mathbb{Q}), \quad x \in D, \quad a \in G(\mathbb{A}_f), \quad k \in K.$$

*When we endow $D$ with its usual topology and $G(\mathbb{A}_f)$ with the adèlic topology (or the discrete topology), this becomes a homeomorphism.*

PROOF. Because $K$ is open, $G(\mathbb{A}_f) = G(\mathbb{Q}) \cdot K$ (strong approximation theorem). Therefore, every element of $G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K$ is represented by an element of the form $[x, 1]$. By definition, $[x, 1] = [x', 1]$ if and only if there exist $q \in G(\mathbb{Q})$ and $k \in K$ such that $x' = qx$, $1 = qk$. The second equation implies that $q = k^{-1} \in \Gamma$, and so $[x, 1] = [x', 1]$ if and only if $x$ and $x'$ represent the same element in $\Gamma\backslash D$.

Consider

$$\begin{array}{ccc} D & \xrightarrow{x\mapsto(x,[1])} & D \times (G(\mathbb{A}_f)/K) \\ \downarrow & & \downarrow \\ \Gamma\backslash D & \xrightarrow{[x]\mapsto[x,1]} & G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K. \end{array}$$

As $K$ is open, $G(\mathbb{A}_f)/K$ is discrete, and so the upper map is a homeomorphism of $D$ onto its image, which is open. It follows easily that the lower map is a homeomorphism. □

What happens when we pass to the inverse limit over $\Gamma$? The obvious map

$$D \to \varprojlim \Gamma\backslash D,$$

is injective because each $\Gamma$ acts freely on $D$ and $\bigcap \Gamma = \{1\}$. Is the map surjective? The example

$$\mathbb{Z} \to \varprojlim \mathbb{Z}/m\mathbb{Z} = \hat{\mathbb{Z}}$$

is not encouraging — it suggests that $\varprojlim \Gamma\backslash D$ might be some sort of completion of $D$ relative to the $\Gamma$'s. This is correct: $\varprojlim \Gamma\backslash D$ is much larger than $D$. In fact, when we pass to the limit on the right in (28), we get the obvious answer:

PROPOSITION 4.19. *In the limit,*

$$\varprojlim{}_K G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)/K = G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f) \tag{29}$$

*(adèlic topology on $G(\mathbb{A}_f)$).*

Before proving this, we need a lemma.

LEMMA 4.20. *Let $G$ be a topological group acting continuously on a topological space $X$, and let $(G_i)_{i\in I}$ be a directed family of subgroups of $G$. The canonical map $X/\bigcap G_i \to \varprojlim X/G_i$ is injective if the $G_i$ are compact, and it is surjective if in addition the orbits of the $G_i$ in $X$ are separated.*

PROOF. We shall use that a directed intersection of nonempty compact sets is nonempty, which has the consequence that a directed inverse limit of nonempty compact sets is nonempty.

Assume that each $G_i$ is compact, and let $x, x' \in X$. For each $i$, let

$$G_i(x, x') = \{g \in G_i \mid xg = x'\}.$$

If $x$ and $x'$ have the same image in $\varprojlim X/G_i$, then the $G_i(x, x')$ are all nonempty. Since each is compact, their intersection is nonempty. For any $g$ in the intersection, $xg = x'$, which shows that $x$ and $x'$ have the same image in $X/\bigcap G_i$.

Now assume that each orbit is separated and hence compact. For any $(x_iG_i)_{i\in I} \in \varprojlim X/G_i$, $\varprojlim x_iG_i$ is nonempty. If $x \in \varprojlim x_iG_i$, then $x \cdot \bigcap G_i$ maps to $(x_iG_i)_{i\in I}$. □

PROOF OF 4.19. Let $(x, a) \in D \times G(\mathbb{A}_f)$, and let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$. In order to be able to apply the lemma, we have to show that the image of the orbit $(x, a)K$ in $G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)$ is separated for $K$ sufficiently small. Let $\Gamma = G(\mathbb{Q}) \cap aKa^{-1}$ — we may assume that $\Gamma$ is torsion free (3.5). There exists an open neighbourhood $V$ of $x$ such that $gV \cap V = \emptyset$ for all $g \in \Gamma \smallsetminus \{1\}$ (see the proof of 3.1). For any $(x, b) \in (x, a)K$, $g(V \times aK) \cap (V \times bK) = \emptyset$ for all $g \in G(\mathbb{Q}) \smallsetminus \{1\}$, and so the images of $V \times Ka$ and $V \times Kb$ in $G(\mathbb{Q})\backslash D \times G(\mathbb{A}_f)$ separate $(x, a)$ and $(x, b)$. □

ASIDE 4.21. (a) Why replace the single coset space on the left of (28) with the more complicated double coset space on the right? One reason is that it makes transparent that (in this case) there is an action of $G(\mathbb{A}_f)$ on the inverse system $(\Gamma\backslash D)_\Gamma$, and hence, for example, on

$$\varinjlim H^i(\Gamma\backslash D, \mathbb{Q}).$$

Another reason will be seen presently — we use double cosets to define Shimura varieties. Double coset spaces are pervasive in work on the Langlands program.

(b) The inverse limit of the $D(\Gamma)$ exists as a scheme — it is even locally noetherian and regular (cf. 5.30 below).

**Alternative definition of connected Shimura data.** Recall that $\mathbb{S}$ is the real torus such that $\mathbb{S}(\mathbb{R}) = \mathbb{C}^\times$. The exact sequence

$$0 \to \mathbb{R}^\times \xrightarrow{r \mapsto r^{-1}} \mathbb{C}^\times \xrightarrow{z \mapsto z/\bar{z}} U_1 \to 0$$

arises from an exact sequence of tori

$$0 \to \mathbb{G}_m \xrightarrow{w} \mathbb{S} \longrightarrow U_1 \to 0.$$

Let $H$ be a semisimple real algebraic group with trivial centre. A homomorphism $u\colon U_1 \to H$ defines a homomorphism $h\colon \mathbb{S} \to H$ by the rule $h(z) = u(z/\bar{z})$, and $U_1$ will act on $\mathrm{Lie}(H)_\mathbb{C}$ through the characters $z, 1, z^{-1}$ if and only if $\mathbb{S}$ acts on $\mathrm{Lie}(H)_\mathbb{C}$ through the characters $z/\bar{z}, 1, \bar{z}/z$. Conversely, let $h$ be a homomorphism $\mathbb{S} \to H$ for which $\mathbb{S}$ acts on $\mathrm{Lie}(H)_\mathbb{C}$ through the characters $z/\bar{z}, 1, \bar{z}/z$. Then $w(\mathbb{G}_m)$ acts trivially on $\mathrm{Lie}(H)_\mathbb{C}$, which implies that $h$ is trivial on $w(\mathbb{G}_m)$ because the adjoint representation $H \to \mathrm{Lie}(H)$ is faithful. Thus, $h$ arises from a $u$.

Now let $G$ be a semisimple algebraic group over $\mathbb{Q}$. From the above remark, we see that to give a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class $D$ of homomorphisms $u\colon U_1 \to G^{\mathrm{ad}}_\mathbb{R}$ satisfying SU1,2 is the same as to give a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class $X^+$ of homomorphisms $h\colon \mathbb{S} \to G^{\mathrm{ad}}_\mathbb{R}$ satisfying the following conditions:

SV1: for $h \in X^+$, only the characters $z/\bar{z}, 1, \bar{z}/z$ occur in the representation of $\mathbb{S}$ on $\mathrm{Lie}(G^{\mathrm{ad}})_\mathbb{C}$ defined by $h$;
SV2: $\mathrm{ad}h(i)$ is a Cartan involution on $G^{\mathrm{ad}}$.

DEFINITION 4.22. A ***connected Shimura datum*** is a pair $(G, X^+)$ consisting of a semisimple algebraic group over $\mathbb{Q}$ and a $G^{\mathrm{ad}}(\mathbb{R})^+$-conjugacy class of homomorphisms $h\colon \mathbb{S} \to G^{\mathrm{ad}}_\mathbb{R}$ satisfying SV1, SV2, and

SV3: $G^{\mathrm{ad}}$ has no $\mathbb{Q}$-factor on which the projection of $h$ is trivial.

In the presence of the other conditions, SV3 is equivalent to SU3 (see 4.7). Thus, because of the correspondence $u \leftrightarrow h$, this is essentially the same as Definition 4.4.

Definition 4.4 is more convenient when working with only connected Shimura varieties, while Definition 4.22 is more convenient when working with both connected and nonconnected Shimura varieties.

NOTES. Connected Shimura varieties were defined *en passant* in Deligne 1979, 2.1.8.

## 5. Shimura varieties

Connected Shimura varieties are very natural objects, so why do we need anything more complicated? There are two main reasons. From the perspective of the Langlands program, we should be working with reductive groups, not semisimple groups. More fundamentally, the varieties $D(\Gamma)$ making up a connected Shimura variety $\mathrm{Sh}^\circ(G, D)$ have models over number fields, but the models depend a realization of $G$ as the derived group of a reductive group. Moreover, the number field depends on $\Gamma$ — as $\Gamma$ shrinks the field grows. For example, the modular curve $\Gamma(N)\backslash\mathcal{H}_1$ is naturally defined over $\mathbb{Q}[\zeta_N]$, $\zeta_N = e^{2\pi i/N}$. Clearly, for a canonical model we would like all the varieties in the family to be defined over the same field.[10]

How can we do this? Consider the line $Y + i = 0$. This is naturally defined over $\mathbb{Q}[i]$, not $\mathbb{Q}$. On the other hand, the variety $Y^2 + 1 = 0$ is naturally defined over $\mathbb{Q}$, and over $\mathbb{C}$ it decomposes into a disjoint pair of conjugate lines $(Y - i)(Y + i) = 0$. So we have managed to get our variety defined over $\mathbb{Q}$ at the cost of adding other connected components. It is always possible to lower the field of definition of a variety by taking the disjoint union of it with its conjugates. Shimura varieties give a systematic way of doing this for connected Shimura varieties.

**Notations for reductive groups.** Let $G$ be a reductive group over $\mathbb{Q}$, and let $G \xrightarrow{\mathrm{ad}} G^{\mathrm{ad}}$ be the quotient of $G$ by its centre $Z$. We let $G(\mathbb{R})_+$ denote the group of elements of $G(\mathbb{R})$ whose image in $G^{\mathrm{ad}}(\mathbb{R})$ lies in its identity component $G^{\mathrm{ad}}(\mathbb{R})^+$, and we let $G(\mathbb{Q})_+ = G(\mathbb{Q}) \cap G(\mathbb{R})_+$. For example, $\mathrm{GL}_2(\mathbb{Q})_+$ consists of the $2 \times 2$ matrices with rational coefficients having positive determinant.

For a reductive group $G$ (resp. for $\mathrm{GL}_n$), there are exact sequences

$$
\begin{array}{ll}
1 \longrightarrow G^{\mathrm{der}} \longrightarrow G \xrightarrow{\nu} T \longrightarrow 1 & 1 \longrightarrow \mathrm{SL}_n \longrightarrow \mathrm{GL}_n \xrightarrow{\det} \mathbb{G}_m \longrightarrow 1 \\
1 \longrightarrow Z \longrightarrow G \xrightarrow{\mathrm{ad}} G^{\mathrm{ad}} \longrightarrow 1 & 1 \longrightarrow \mathbb{G}_m \longrightarrow \mathrm{GL}_n \xrightarrow{\mathrm{ad}} \mathrm{PGL}_n \longrightarrow 1 \\
1 \longrightarrow Z' \longrightarrow Z \longrightarrow T \longrightarrow 1 & 1 \longrightarrow \mu_n \longrightarrow \mathbb{G}_m \xrightarrow{x \mapsto x^n} \mathbb{G}_m \longrightarrow 1
\end{array}
$$

Here $T$ (a torus) is the largest commutative quotient of $G$, and $Z' =_{\mathrm{df}} Z \cap G^{\mathrm{der}}$ (a finite algebraic group) is the centre of $G^{\mathrm{der}}$.

**The real points of algebraic groups.**

PROPOSITION 5.1. *For a surjective homomorphism $\varphi\colon G \to H$ of algebraic groups over $\mathbb{R}$, $G(\mathbb{R})^+ \to H(\mathbb{R})^+$ is surjective.*

[10]In fact, Shimura has an elegant way of describing a canonical model in which the varieties in the family are defined over different fields, but this doesn't invalidate my statement. Incidentally, Shimura also requires a reductive (not a semisimple) group in order to have a canonical model over a number field. For an explanation of Shimura's point of view in the language of these notes, see Milne and Shih 1981.

PROOF. The map $\varphi(\mathbb{R})\colon G(\mathbb{R})^+ \to H(\mathbb{R})^+$ can be regarded as a smooth map of smooth manifolds. As $\varphi$ is surjective on the tangent spaces at 1, the image of $\varphi(\mathbb{R})$ contains an open neighbourhood of 1 (Boothby 1975, II 7.1). This implies that the image itself is open because it is a group. It is therefore also closed, and this implies that it equals $H(\mathbb{R})^+$. □

Note that $G(\mathbb{R}) \to H(\mathbb{R})$ need not be surjective. For example, $\mathbb{G}_m \xrightarrow{x\mapsto x^n} \mathbb{G}_m$ is surjective as a map of algebraic groups, but the image of $\mathbb{G}_m(\mathbb{R}) \xrightarrow{n} \mathbb{G}_m(\mathbb{R})$ is $\mathbb{G}_m(\mathbb{R})^+$ or $\mathbb{G}_m(\mathbb{R})$ according as $n$ is even or odd. Also $\mathrm{SL}_2 \to \mathrm{PGL}_2$ is surjective, but the image of $\mathrm{SL}_2(\mathbb{R}) \to \mathrm{PGL}_2(\mathbb{R})$ is $\mathrm{PGL}_2(\mathbb{R})^+$.

For a simply connected algebraic group $G$, $G(\mathbb{C})$ is simply connected as a topological space, but $G(\mathbb{R})$ need not be. For example, $\mathrm{SL}_2(\mathbb{R})$ is not simply connected.

THEOREM 5.2 (Cartan 1927). *For a simply connected group $G$ over $\mathbb{R}$, $G(\mathbb{R})$ is connected.*

PROOF. See Platonov and Rapinchuk 1994, Theorem 7.6, p407. □

COROLLARY 5.3. *For a reductive group $G$ over $\mathbb{R}$, $G(\mathbb{R})$ has only finitely many connected components (for the real topology).*[11]

PROOF. Because of (5.1), an exact sequence of real algebraic groups

$$1 \to N \to G' \to G \to 1 \tag{30}$$

with $N \subset Z(G')$ gives rise to an exact sequence

$$\pi_0(G'(\mathbb{R})) \to \pi_0(G(\mathbb{R})) \to H^1(\mathbb{R}, N).$$

Let $\tilde{G}$ be the universal covering group of $G^{\mathrm{der}}$. As $G$ is an almost direct product of $Z = Z(G)$ and $G^{\mathrm{der}}$, there is an exact sequence (30) with $G' = Z \times \tilde{G}$ and $N$ finite. Now

- $\pi_0(\tilde{G}(\mathbb{R})) = 0$ because $\tilde{G}$ is simply connected,
- $\pi_0(Z(\mathbb{R}))$ is finite because $Z^\circ$ has finite index in $Z$ and $Z^\circ$ is a quotient (by a finite group) of a product of copies of $U_1$ and $\mathbb{G}_m$, and
- $H^1(\mathbb{R}, N)$ is finite because $N$ is finite. □

For example, $\mathbb{G}_m^d(\mathbb{R}) = (\mathbb{R}^\times)^d$ has $2^d$ connected components, and each of $\mathrm{PGL}_2(\mathbb{R})$ and $\mathrm{GL}_2(\mathbb{R})$ has 2 connected components.

THEOREM 5.4 (real approximation). *For any connected algebraic group $G$ over $\mathbb{Q}$, $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$.*

PROOF. See Platonov and Rapinchuk 1994, Theorem 7.7, p415. □

**Shimura data.**

DEFINITION 5.5. A ***Shimura datum*** is a pair $(G, X)$ consisting of a reductive group $G$ over $\mathbb{Q}$ and a $G(\mathbb{R})$-conjugacy class $X$ of homomorphisms $h\colon \mathbb{S} \to G_{\mathbb{R}}$ satisfying the conditions SV1, SV2, and SV3 (see p302).

[11] This also follows from the theorem of Whitney 1957: for an algebraic variety $V$ over $\mathbb{R}$, $V(\mathbb{R})$ has only finitely many connected components (for the real topology) — see Platonov and Rapinchuk 1994, Theorem 3.6, p119.

Note that, in contrast to a connected Shimura datum, $G$ is reductive (not semisimple), the homomorphisms $h$ have target $G_{\mathbb{R}}$ (not $G^{\mathrm{ad}}_{\mathbb{R}}$), and $X$ is the full $G(\mathbb{R})$-conjugacy class (not a connected component).

EXAMPLE 5.6. Let $G = \mathrm{GL}_2$ (over $\mathbb{Q}$) and let $X$ be the set of $\mathrm{GL}_2(\mathbb{R})$-conjugates of the homomorphism $h_o\colon \mathbb{S} \to \mathrm{GL}_{2\mathbb{R}}$, $h_o(a+ib) = \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$. Then $(G, X)$ is a Shimura datum. Note that there is a natural bijection $X \to \mathbb{C} \smallsetminus \mathbb{R}$, namely, $h_o \mapsto i$ and $gh_og^{-1} \mapsto gi$. More intrinsically, $h \leftrightarrow z$ if and only if $h(\mathbb{C}^{\times})$ is the stabilizer of $z$ in $\mathrm{GL}_2(\mathbb{R})$ and $h(z)$ acts on the tangent space at $z$ as multiplication by $z/\overline{z}$ (rather than $\overline{z}/z$).

PROPOSITION 5.7. *Let $G$ be a reductive group over $\mathbb{R}$. For a homomorphism $h\colon \mathbb{S} \to G$, let $\overline{h}$ be the composite of $h$ with $G \to G^{\mathrm{ad}}$. Let $X$ be a $G(\mathbb{R})$-conjugacy class of homomorphisms $\mathbb{S} \to G$, and let $\overline{X}$ be the $G^{\mathrm{ad}}(\mathbb{R})$-conjugacy class of homomorphisms $\mathbb{S} \to G^{\mathrm{ad}}$ containing the $\overline{h}$ for $h \in X$.*

(a) *The map $h \mapsto \overline{h}\colon X \to \overline{X}$ is injective and its image is a union of connected components of $\overline{X}$.*

(b) *Let $X^+$ be a connected component of $X$, and let $\overline{X}^+$ be its image in $\overline{X}$. If $(G, X)$ satisfies the axioms SV1–3 then $(G^{\mathrm{der}}, \overline{X}^+)$ satisfies the axioms SV1–3; moreover, the stabilizer of $X^+$ in $G(\mathbb{R})$ is $G(\mathbb{R})_+$ (i.e., $gX^+ = X^+ \iff g \in G(\mathbb{R})_+$).*

PROOF. (a) A homomorphism $h\colon \mathbb{S} \to G$ is determined by its projections to $T$ and $G^{\mathrm{ad}}$, because any other homomorphism with the same projections will be of the form $he$ for some regular map $e\colon \mathbb{S} \to Z'$ and $e$ is trivial because $\mathbb{S}$ is connected and $Z'$ is finite. The elements of $X$ all have the same projection to $T$, because $T$ is commutative, which proves that $h \mapsto \overline{h}\colon X \to \overline{X}$ is injective. For the second part of the statement, use that $G^{\mathrm{ad}}(\mathbb{R})^+$ acts transitively on each connected component of $\overline{X}$ (see 1.5) and $G(\mathbb{R})^+ \to G^{\mathrm{ad}}(\mathbb{R})^+$ is surjective.

(b) The first assertion is obvious. In (a) we showed that $\pi_0(X) \subset \pi_0(\overline{X})$. The stabilizer in $G^{\mathrm{ad}}(\mathbb{R})$ of $[\overline{X}^+]$ is $G^{\mathrm{ad}}(\mathbb{R})^+$ (see 4.9), and so its stabilizer in $G(\mathbb{R})$ is the inverse image of $G^{\mathrm{ad}}(\mathbb{R})^+$ in $G(\mathbb{R})$. □

COROLLARY 5.8. *Let $(G, X)$ be a Shimura datum, and let $X^+$ be a connected component of $X$ regarded as a $G(\mathbb{R})^+$-conjugacy class of homomorphisms $\mathbb{S} \to G^{\mathrm{ad}}_{\mathbb{R}}$ (5.7). Then $(G^{\mathrm{der}}, X^+)$ is a connected Shimura datum. In particular, $X$ is a finite disjoint union of hermitian symmetric domains.*

PROOF. Apply Proposition 5.7 and Proposition 4.8. □

Let $(G, X)$ be a Shimura datum. For every $h\colon \mathbb{S} \to G(\mathbb{R})$ in $X$, $\mathbb{S}$ acts on $\mathrm{Lie}(G)_{\mathbb{C}}$ through the characters $z/\overline{z}$, 1, $\overline{z}/z$. Thus, for $r \in \mathbb{R}^{\times} \subset \mathbb{C}^{\times}$, $h(r)$ acts trivially on $\mathrm{Lie}(G)_{\mathbb{C}}$. As the adjoint action of $G$ on $\mathrm{Lie}(G)$ factors through $G^{\mathrm{ad}}$ and $\mathrm{Ad}\colon G^{\mathrm{ad}} \to \mathrm{GL}(\mathrm{Lie}(G))$ is injective, this implies that $h(r) \in Z(\mathbb{R})$ where $Z$ is the centre of $G$. Thus, $h|\mathbb{G}_m$ is independent of $h$ — we denote its reciprocal by $w_X$ (or simply $w$) and we call $w_X$ the ***weight homomorphism***. For any representation $\rho\colon G_{\mathbb{R}} \to \mathrm{GL}(V)$, $\rho \circ w_X$ defines a decomposition of $V = \bigoplus V_n$ which is the weight decomposition of the hodge structure $(V, \rho \circ h)$ for every $h \in X$.

PROPOSITION 5.9. *Let $(G, X)$ be a Shimura datum. Then $X$ has a unique structure of a complex manifold such that, for every representation $\rho\colon G_{\mathbb{R}} \to \mathrm{GL}(V)$,*

*$(V,\rho\circ h)_{h\in X}$ is a holomorphic family of hodge structures. For this complex structure, each family $(V,\rho\circ h)_{h\in X}$ is a variation of hodge structures, and so $X$ is a finite disjoint union of hermitian symmetric domains.*

PROOF. Let $\rho\colon G_{\mathbb{R}}\to \mathrm{GL}(V)$ be a faithful representation of $G_{\mathbb{R}}$. The family of hodge structures $(V,\rho\circ h)_{h\in X}$ is continuous, and a slight generalization of (a) of Theorem 2.14 shows that $X$ has a unique structure of a complex manifold for which this family is holomorphic. It follows from Waterhouse 1979, 3.5, that the family of hodge structures defined by every representation is then holomorphic for this complex structure. The condition SV1 implies that $(V,\rho\circ h)_h$ is a variation of hodge structures, and so we can apply (b) of Theorem 2.14. □

Of course, the complex structures defined on $X$ by (5.8) and (5.9) coincide.

ASIDE 5.10. Let $(G,X)$ be a Shimura datum. The maps $\pi_0(X)\to\pi_0(\overline{X})$ and $G(\mathbb{R})/G(\mathbb{R})_+\to G^{\mathrm{ad}}(\mathbb{R})/G^{\mathrm{ad}}(\mathbb{R})^+$ are injective, and the second can be identified with the first once an $h\in X$ has been chosen. In general, the maps will not be surjective unless $H^1(\mathbb{R},Z)=0$.

**Shimura varieties.** Let $(G,X)$ be a Shimura datum.

LEMMA 5.11. *For any connected component $X^+$ of $X$, the natural map*

$$G(\mathbb{Q})_+\backslash X^+\times G(\mathbb{A}_f)\to G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f)$$

*is a bijection.*

PROOF. Because $G(\mathbb{Q})$ is dense in $G(\mathbb{R})$ (see 5.4) and $G(\mathbb{R})$ acts transitively on $X$, every $x\in X$ is of the form $qx^+$ with $q\in G(\mathbb{Q})$ and $x^+\in X^+$. This shows that the map is surjective.

Let $(x,a)$ and $(x',a')$ be elements of $X^+\times G(\mathbb{A}_f)$. If $[x,a]=[x',a']$ in $G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f)$, then

$$x'=qx,\quad a'=qa,\quad \text{some } q\in G(\mathbb{Q}).$$

Because $x$ and $x'$ are both in $X^+$, $q$ stabilizes $X^+$ and so lies in $G(\mathbb{R})_+$ (see 5.7). Therefore, $[x,a]=[x',a']$ in $G(\mathbb{Q})_+\backslash X\times G(\mathbb{A}_f)$. □

LEMMA 5.12. *For any open subgroup $K$ of $G(\mathbb{A}_f)$, the set $G(\mathbb{Q})_+\backslash G(\mathbb{A}_f)/K$ is finite.*

PROOF. Since $G(\mathbb{Q})_+\backslash G(\mathbb{Q})\to G^{\mathrm{ad}}(\mathbb{R})^+\backslash G^{\mathrm{ad}}(\mathbb{R})$ is injective and the second group is finite (5.3), it suffices to show that $G(\mathbb{Q})\backslash G(\mathbb{A}_f)/K$ is finite. Later (Theorem 5.17) we shall show that this follows from the strong approximation theorem if $G^{\mathrm{der}}$ is simply connected, and the general case is not much more difficult. □

For $K$ a compact open subgroup $K$ of $G(\mathbb{A}_f)$, consider the double coset space

$$\mathrm{Sh}_K(G,X)=G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f)/K$$

in which $G(\mathbb{Q})$ acts on $X$ and $G(\mathbb{A}_f)$ on the left, and $K$ acts on $G(\mathbb{A}_f)$ on the right:

$$q(x,a)k=(qx,qak),\quad q\in G(\mathbb{Q}),\quad x\in X,\quad a\in G(\mathbb{A}_f),\quad k\in K.$$

LEMMA 5.13. *Let $\mathcal{C}$ be a set of representatives for the double coset space $G(\mathbb{Q})_+\backslash G(\mathbb{A}_f)/K$, and let $X^+$ be a connected component of $X$. Then*

$$G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f)/K\cong\textstyle\bigsqcup_{g\in\mathcal{C}}\Gamma_g\backslash X^+$$

*where $\Gamma_g$ is the subgroup $gKg^{-1} \cap G(\mathbb{Q})_+$ of $G(\mathbb{Q})_+$. When we endow $X$ with its usual topology and $G(\mathbb{A}_f)$ with its adèlic topology (equivalently, the discrete topology), this becomes a homeomorphism.*

PROOF. It is straightforward to prove that, for $g \in \mathcal{C}$, the map

$$[x] \mapsto [x, g] \colon \Gamma_g \backslash X^+ \to G(\mathbb{Q})_+ \backslash X^+ \times G(\mathbb{A}_f)/K$$

is injective, and that $G(\mathbb{Q})_+ \backslash X^+ \times G(\mathbb{A}_f)/K$ is the disjoint union of the images of these maps. Thus, the first statement follows from (5.11). The second statement can be proved in the same way as the similar statement in (4.18). □

Because $\Gamma_g$ is a congruence subgroup of $G(\mathbb{Q})$, its image in $G^{\mathrm{ad}}(\mathbb{Q})$ is arithmetic (3.2), and so (by definition) its image in $\mathrm{Aut}(X^+)$ is arithmetic. Moreover, when $K$ is sufficiently small, $\Gamma_g$ will be neat for all $g \in \mathcal{C}$ (apply 3.5) and so its image in $\mathrm{Aut}(X^+)^+$ will also be neat and hence torsion free. Then $\Gamma_g \backslash X^+$ is an arithmetic locally symmetric variety, and $\mathrm{Sh}_K(G, X)$ is finite disjoint of such varieties. Moreover, for an inclusion $K' \subset K$ of sufficiently small compact open subgroups of $G(\mathbb{A}_f)$, the natural map $\mathrm{Sh}_{K'}(G, X) \to \mathrm{Sh}_K(G, X)$ is regular. Thus, when we vary $K$ (sufficiently small), we get an inverse system of algebraic varieties $(\mathrm{Sh}_K(G, X))_K$. There is a natural action of $G(\mathbb{A}_f)$ on the system: for $g \in G(\mathbb{A}_f)$, $K \mapsto g^{-1}Kg$ maps compact open subgroups to compact open subgroups, and

$$\mathcal{T}(g) \colon \mathrm{Sh}_K(G, X) \to \mathrm{Sh}_{g^{-1}Kg}(G, X)$$

acts on points as

$$[x, a] \mapsto [x, ag] \colon G(\mathbb{Q}) \backslash X \otimes G(\mathbb{A}_f)/K \to G(\mathbb{Q}) \backslash X \times G(\mathbb{A}_f)/g^{-1}Kg.$$

Note that this is a right action: $\mathcal{T}(gh) = \mathcal{T}(h) \circ \mathcal{T}(g)$.

DEFINITION 5.14. The ***Shimura variety*** $\mathrm{Sh}(G, X)$ ***attached to the Shimura datum*** $(G, X)$ is the inverse system of varieties $(\mathrm{Sh}_K(G, X))_K$ endowed with the action of $G(\mathbb{A}_f)$ described above. Here $K$ runs through the sufficiently small compact open subgroups of $G(\mathbb{A}_f)$.

**Morphisms of Shimura varieties.**

DEFINITION 5.15. Let $(G, X)$ and $(G', X')$ be Shimura data.

(a) A ***morphism of Shimura data*** $(G, X) \to (G', X')$ is a homomorphism $G \to G'$ of algebraic groups sending $X$ into $X'$.
(b) A ***morphism of Shimura varieties*** $\mathrm{Sh}(G, X) \to \mathrm{Sh}(G', X')$ is an inverse system of regular maps of algebraic varieties compatible with the action of $G(\mathbb{A}_f)$.

THEOREM 5.16. *A morphism of Shimura data $(G, X) \to (G', X')$ defines a morphism $\mathrm{Sh}(G, X) \to \mathrm{Sh}(G', X')$ of Shimura varieties, which is a closed immersion if $G \to G'$ is injective.*

PROOF. The first part of the statement is obvious from (3.14), and the second is proved in Theorem 1.15 of Deligne 1971*b*. □

**The structure of a Shimura variety.** By the structure of $\mathrm{Sh}(G,X)$, I mean the structure of the set of connected components and the structure of each connected component. This is worked out in general in Deligne 1979, 2.1.16, but the result there is complicated. When $G^{\mathrm{der}}$ is simply connected,[12] it is possible to prove a more pleasant result: the set of connected components is a "zero-dimensional Shimura variety", and each connected component is a connected Shimura variety.

Let $(G,X)$ be a Shimura datum. As on p303, $Z$ is the centre of $G$ and $T$ the largest commutative quotient of $G$. There are homomorphisms $Z \hookrightarrow G \xrightarrow{\nu} T$, and we define

$$\begin{aligned} T(\mathbb{R})^{\dagger} &= \mathrm{Im}(Z(\mathbb{R}) \to T(\mathbb{R})), \\ T(\mathbb{Q})^{\dagger} &= T(\mathbb{Q}) \cap T(\mathbb{R})^{\dagger}. \end{aligned}$$

Because $Z \to T$ is surjective, $T(\mathbb{R})^{\dagger} \supset T(\mathbb{R})^{+}$ (see 5.1), and so $T(\mathbb{R})^{\dagger}$ and $T(\mathbb{Q})^{\dagger}$ are of finite index in $T(\mathbb{R})$ and $T(\mathbb{Q})$ (see 5.3). For example, for $G = \mathrm{GL}_2$, $T(\mathbb{Q})^{\dagger} = T(\mathbb{Q})^{+} = \mathbb{Q}_{>0}$.

Theorem 5.17. *Assume $G^{der}$ is simply connected. For $K$ sufficiently small, the natural map*

$$G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \to T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K)$$

*defines an isomorphism*

$$\pi_0(\mathrm{Sh}_K(G,X)) \cong T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K).$$

*Moreover, $T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K)$ is finite, and the connected component over* [1] *is canonically isomorphic to $\Gamma\backslash X^{+}$ for some congruence subgroup $\Gamma$ of $G^{der}(\mathbb{Q})$ containing $K \cap G^{der}(\mathbb{Q})$.*

In Lemma 5.20 below, we show that $\nu(G(\mathbb{Q})_{+}) \subset T(\mathbb{Q})^{\dagger}$. The "natural map" in the theorem is

$$G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \overset{5.11}{\cong} G(\mathbb{Q})_{+}\backslash X^{+} \times G(\mathbb{A}_f)/K \xrightarrow{[x,g]\mapsto[\nu(g)]} T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K).$$

The theorem gives a diagram

$$\begin{tikzcd} G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K \arrow[d] & \Gamma\backslash X^{+}, \arrow[l, hook'] \arrow[d] \\ T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K) & {[1]} \arrow[l, hook'] \end{tikzcd}$$

in which $T(\mathbb{Q})^{\dagger}\backslash T(\mathbb{A}_f)/\nu(K)$ is finite and discrete, the left hand map is continuous and onto with connected fibres, and $\Gamma\backslash X^{+}$ is the fibre over [1].

Lemma 5.18. *Assume $G^{der}$ is simply connected. Then $G(\mathbb{R})_{+} = G^{der}(\mathbb{R}) \cdot Z(\mathbb{R})$.*

[12] The Shimura varieties with simply connected derived group are the most important — if one knows everything about them, then one knows everything about all Shimura varieties (because the remainder are quotients of them). However, there are naturally occurring Shimura varieties for which $G^{\mathrm{der}}$ is not simply connected, and so we should not ignore them.

PROOF. Because $G^{\mathrm{der}}$ is simply connected, $G^{\mathrm{der}}(\mathbb{R})$ is connected (5.2) and so $G^{\mathrm{der}}(\mathbb{R}) \subset G(\mathbb{R})_+$. Hence $G(\mathbb{R})_+ \supset G^{\mathrm{der}}(\mathbb{R}) \cdot Z(\mathbb{R})$. For the converse, we use the exact commutative diagram:

$$\begin{tikzcd}
1 \arrow[r] & Z'(\mathbb{R}) \arrow[r, "z\mapsto(z^{-1},z)"] \arrow[d, equal] & Z(\mathbb{R})\times G^{\mathrm{der}}(\mathbb{R}) \arrow[r, "(z,g)\mapsto zg"] \arrow[d, "(z,g)\mapsto g"] & G(\mathbb{R}) \arrow[r] \arrow[d] & H^1(\mathbb{R},Z') \arrow[d, equal] \\
1 \arrow[r] & Z'(\mathbb{R}) \arrow[r] & G^{\mathrm{der}}(\mathbb{R}) \arrow[r] & G^{\mathrm{ad}}(\mathbb{R}) \arrow[r] & H^1(\mathbb{R},Z').
\end{tikzcd}$$

As $G^{\mathrm{der}} \to G^{\mathrm{ad}}$ is surjective, so also is $G^{\mathrm{der}}(\mathbb{R}) \to G^{\mathrm{ad}}(\mathbb{R})^+$ (see 5.1). Therefore, an element $g$ of $G(\mathbb{R})$ lies in $G(\mathbb{R})_+$ if and only if its image in $G^{\mathrm{ad}}(\mathbb{R})$ lifts to $G^{\mathrm{der}}(\mathbb{R})$. Thus,

$$\begin{aligned} g \in G(\mathbb{R})_+ &\iff g \mapsto 0 \text{ in } H^1(\mathbb{R},Z') \\ &\iff g \text{ lifts to } Z(\mathbb{R})\times G^{\mathrm{der}}(\mathbb{R}) \\ &\iff g \in Z(\mathbb{R})\cdot G^{\mathrm{der}}(\mathbb{R}) \end{aligned}$$ □

LEMMA 5.19. *Let $H$ be a simply connected semisimple algebraic group $H$ over $\mathbb{Q}$.*

(a) *For every finite prime, the group $H^1(\mathbb{Q}_\ell, H) = 0$.*
(b) *The map $H^1(\mathbb{Q},H) \to \prod_{l\le\infty} H^1(\mathbb{Q}_l,H)$ is injective (Hasse principle).*

PROOF. (a) See Platonov and Rapinchuk 1994, Theorem 6.4, p284.
(b) See ibid., Theorem 6.6, p286. □

Both statements fail for groups that are not simply connected.

LEMMA 5.20. *Assume $G^{der}$ is simply connected, and let $t \in T(\mathbb{Q})$. Then $t \in T(\mathbb{Q})^\dagger$ if and only if $t$ lifts to an element of $G(\mathbb{Q})_+$.*

PROOF. Lemma 5.19 implies that the vertical arrow at right in the following diagram is injective:

$$\begin{tikzcd}
1 \arrow[r] & G^{\mathrm{der}}(\mathbb{Q}) \arrow[r] \arrow[d] & G(\mathbb{Q}) \arrow[r, "\nu"] \arrow[d] & T(\mathbb{Q}) \arrow[r] \arrow[d] & H^1(\mathbb{Q},G^{\mathrm{der}}) \arrow[d, "\text{injective}"] \\
1 \arrow[r] & G^{\mathrm{der}}(\mathbb{R}) \arrow[r] & G(\mathbb{R}) \arrow[r, "\nu"] & T(\mathbb{R}) \arrow[r] & H^1(\mathbb{R},G^{\mathrm{der}})
\end{tikzcd}$$

Let $t \in T(\mathbb{Q})^\dagger$. By definition, the image $t_\mathbb{R}$ of $t$ in $T(\mathbb{R})$ lifts to an element $z \in Z(\mathbb{R}) \subset G(\mathbb{R})$. From the diagram, we see that this implies that $t$ maps to the trivial element in $H^1(\mathbb{Q},G^{\mathrm{der}})$ and so it lifts to an element $g \in G(\mathbb{Q})$. Now $g_\mathbb{R}\cdot z^{-1} \mapsto t_\mathbb{R}\cdot t_\mathbb{R}^{-1} = 1$ in $T(\mathbb{R})$, and so $g_\mathbb{R} \in G^{\mathrm{der}}(\mathbb{R})\cdot z \subset G^{\mathrm{der}}(\mathbb{R})\cdot Z(\mathbb{R}) \subset G(\mathbb{R})_+$. Therefore, $g \in G(\mathbb{Q})_+$.

Let $t$ be an element of $T(\mathbb{Q})$ lifting to an element $a$ of $G(\mathbb{Q})_+$. According to 5.18, $a_\mathbb{R} = gz$ for some $g \in G^{\mathrm{der}}(\mathbb{R})$ and $z \in Z(\mathbb{R})$. Now $a_\mathbb{R}$ and $z$ map to the same element in $T(\mathbb{R})$, namely, to $t_\mathbb{R}$, and so $t \in T(\mathbb{Q})^\dagger$ □

The lemma allows us to write

$$T(\mathbb{Q})^\dagger\backslash T(\mathbb{A}_f)/\nu(K) = \nu(G(\mathbb{Q})_+)\backslash T(\mathbb{A}_f)/\nu(K).$$

We now study the fibre over [1] of the map

$$G(\mathbb{Q})_+\backslash X^+ \times G(\mathbb{A}_f)/K \xrightarrow{[x,g]\mapsto[\nu(g)]} \nu(G(\mathbb{Q})_+)\backslash T(\mathbb{A}_f)/\nu(K).$$

Let $g \in G(\mathbb{A}_f)$. If $[\nu(g)] = [1]_K$, then $\nu(g) = \nu(q)\nu(k)$ some $q \in G(\mathbb{Q})_+$ and $k \in K$. It follows that $\nu(q^{-1}gk^{-1}) = 1$, that $q^{-1}gk^{-1} \in G^{\mathrm{der}}(\mathbb{A}_f)$, and that $g \in G(\mathbb{Q})_+ \cdot G^{\mathrm{der}}(\mathbb{A}_f) \cdot K$. Hence every element of the fibre over $[1]$ is represented by an element $(x, a)$ with $a \in G^{\mathrm{der}}(\mathbb{A}_f)$. But, according to the strong approximation theorem (4.16), $G^{\mathrm{der}}(\mathbb{A}_f) = G^{\mathrm{der}}(\mathbb{Q}) \cdot (K \cap G^{\mathrm{der}}(\mathbb{A}_f))$, and so the fibre over $[1]$ is a quotient of $X^+$; in particular, it is connected. More precisely, it equals $\Gamma\backslash X^+$ where $\Gamma$ is the image of $K \cap G(\mathbb{Q})_+$ in $G^{\mathrm{ad}}(\mathbb{Q})^+$. This $\Gamma$ is an arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{Q})^+$ containing the image of the congruence subgroup $K \cap G^{\mathrm{der}}(\mathbb{Q})$ of $G^{\mathrm{der}}(\mathbb{Q})$. Moreover, arbitrarily small such $\Gamma$'s arise in this way. Hence, the inverse system of fibres over $[1]$ (indexed by the compact open subgroups $K$ of $G(\mathbb{A}_f)$) is equivalent to the inverse system $\mathrm{Sh}^\circ(G^{\mathrm{der}}, X^+) = (\Gamma\backslash X^+)$.

The study of the fibre over $[t]$ will be similar once we show that there exists an $a \in G(\mathbb{A}_f)$ mapping to $t$ (so that the fibre is nonempty). This follows from the next lemma.

LEMMA 5.21. *Assume $G^{der}$ is simply connected. Then the map $\nu\colon G(\mathbb{A}_f) \to T(\mathbb{A}_f)$ is surjective and sends compact open subgroups to compact open subgroups.*

PROOF. We have to show:

(a) the homomorphism $\nu\colon G(\mathbb{Q}_\ell) \to T(\mathbb{Q}_\ell)$ is surjective for all finite $\ell$;
(b) the homomorphism $\nu\colon G(\mathbb{Z}_\ell) \to T(\mathbb{Z}_\ell)$ is surjective for almost all $\ell$.

(a) For each prime $\ell$, there is an exact sequence

$$1 \to G^{\mathrm{der}}(\mathbb{Q}_\ell) \to G(\mathbb{Q}_\ell) \xrightarrow{\nu} T(\mathbb{Q}_\ell) \to H^1(\mathbb{Q}_\ell, G^{\mathrm{der}})$$

and so (5.19a) shows that $\nu\colon G(\mathbb{Q}_\ell) \to T(\mathbb{Q}_\ell)$ is surjective.

(b) Extend the homomorphism $G \to T$ to a homomorphism of group schemes $\mathcal{G} \to \mathcal{T}$ over $\mathbb{Z}[\frac{1}{N}]$ for some integer $N$. After $N$ has been enlarged, this map will be a smooth morphism of group schemes and its kernel $\mathcal{G}'$ will have nonsingular connected fibres. On extending the base ring to $\mathbb{Z}_\ell$, $\ell \nmid N$, we obtain an exact sequence

$$0 \to \mathcal{G}'_\ell \to \mathcal{G}_\ell \xrightarrow{\nu} \mathcal{T}_\ell \to 0$$

of group schemes over $\mathbb{Z}_\ell$ such that $\nu$ is smooth and $(\mathcal{G}'_\ell)_{\mathbb{F}_\ell}$ is nonsingular and connected. Let $P \in \mathcal{T}_\ell(\mathbb{Z}_\ell)$, and let $Y = \nu^{-1}(P) \subset \mathcal{G}_\ell$. We have to show that $Y(\mathbb{Z}_\ell)$ is nonempty. By Lang's lemma (Springer 1998, 4.4.17), $H^1(\mathbb{F}_\ell, (\mathcal{G}'_\ell)_{\mathbb{F}_\ell}) = 0$, and so

$$\nu\colon \mathcal{G}_\ell(\mathbb{F}_\ell) \to \mathcal{T}_\ell(\mathbb{F}_\ell)$$

is surjective. Therefore $Y(\mathbb{F}_\ell)$ is nonempty. Because $Y$ is smooth over $\mathbb{Z}_\ell$, an argument as in the proof of Newton's lemma (e.g., ANT 7.22) now shows that a point $Q_0 \in Y(\mathbb{F}_\ell)$ lifts to a point $Q \in Y(\mathbb{Z}_\ell)$. □

It remains to show that $T(\mathbb{Q})^\dagger\backslash T(\mathbb{A}_f)/\nu(K)$ is finite. Because $T(\mathbb{Q})^\dagger$ has finite index in $T(\mathbb{Q})$, it suffices to prove that $T(\mathbb{Q})\backslash T(\mathbb{A}_f)/\nu(K)$ is finite. But $\nu(K)$ is open, and so this follows from the next lemma.

LEMMA 5.22. *For any torus $T$ over $\mathbb{Q}$, $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ is compact.*

PROOF. Consider first the case $T = \mathbb{G}_m$. Then

$$T(\mathbb{A}_f)/T(\hat{\mathbb{Z}}) = \mathbb{A}_f^\times/\hat{\mathbb{Z}}^\times \cong \bigoplus\nolimits_{\ell \text{ finite}} \mathbb{Q}_\ell^\times/\mathbb{Z}_\ell^\times \xrightarrow[\cong]{\oplus \mathrm{ord}_\ell} \bigoplus\nolimits_{\ell \text{ finite}} \mathbb{Z},$$

which is the group of fractional ideals of $\mathbb{Z}$. Therefore, $\mathbb{Q}^\times\backslash\mathbb{A}_f^\times/\hat{\mathbb{Z}}^\times$ is the ideal class group of $\mathbb{Z}$, which is trivial: $\mathbb{A}_f^\times = \mathbb{Q}^\times \cdot \hat{\mathbb{Z}}^\times$. Hence $\mathbb{Q}^\times\backslash\mathbb{A}_f^\times$ is a quotient of $\hat{\mathbb{Z}}^\times$, which is compact.

For a number field $F$, the same argument using the finiteness of the class number of $F$ shows that $F^\times\backslash\mathbb{A}_{F,f}^\times$ is compact. Here $\mathbb{A}_{F,f}^\times = \prod_{v \text{ finite}}(F_v^\times : \mathcal{O}_v^\times)$.

An arbitrary torus $T$ over $\mathbb{Q}$ will split over some number field, say, $T_F \approx \mathbb{G}_m^{\dim(T)}$. Then $T(F)\backslash T(\mathbb{A}_{F,f}) \approx (F^\times\backslash\mathbb{A}_{F,f}^\times)^{\dim(T)}$, which is compact, and $T(\mathbb{Q})\backslash T(\mathbb{A}_f)$ is a closed subset of it. □

REMARK 5.23. One may ask whether the fibre over [1] equals

$$\Gamma\backslash X^+ = G^{\mathrm{der}}(\mathbb{Q})\backslash X^+ \times G^{\mathrm{der}}(\mathbb{A}_f)/K \cap G^{\mathrm{der}}(\mathbb{A}_f), \quad \Gamma = K \cap G^{\mathrm{der}}(\mathbb{Q}),$$

rather than quotient of $X^+$ by some larger group than $\Gamma$. This will be true if $Z'$ satisfies the Hasse principle for $H^1$ (for then every element in $G(\mathbb{Q})_+ \cap K$ with $K$ sufficiently small will lie in $G^{\mathrm{der}}(\mathbb{Q}) \cdot Z(\mathbb{Q})$). It is known that $Z'$ satisfies the Hasse principle for $H^1$ when $G^{\mathrm{der}}$ has no isogeny factors of type $A$, but not in general otherwise (Milne 1987). This is one reason why, in the definition of $\mathrm{Sh}^\circ(G^{\mathrm{der}}, X^+)$, we include quotients $\Gamma\backslash X^+$ in which $\Gamma$ is an arithmetic subgroup of $G^{\mathrm{ad}}(\mathbb{Q})^+$ containing, but not necessarily equal to, the image of congruence subgroup of $G^{\mathrm{der}}(\mathbb{Q})$.

**Zero-dimensional Shimura varieties.** Let $T$ be a torus over $\mathbb{Q}$. According to Deligne's definition, every homomorphism $h\colon \mathbb{C}^\times \to T(\mathbb{R})$ defines a Shimura variety $\mathrm{Sh}(T, \{h\})$ — in this case the conditions SV1,2,3 are vacuous. For any compact open $K \subset T(\mathbb{A}_f)$,

$$\mathrm{Sh}_K(T, \{h\}) = T(\mathbb{Q})\backslash\{h\} \times T(\mathbb{A}_f)/K \cong T(\mathbb{Q})\backslash T(\mathbb{A}_f)/K$$

(finite discrete set). We should extend this definition a little. Let $Y$ be a finite set on which $T(\mathbb{R})/T(\mathbb{R})^+$ acts transitively. Define $\mathrm{Sh}(T, Y)$ to be the inverse system of finite sets

$$\mathrm{Sh}_K(T, Y) = T(\mathbb{Q})\backslash Y \times T(\mathbb{A}_f)/K,$$

with $K$ running over the compact open subgroups of $T(\mathbb{A}_f)$. Call such a system a ***zero-dimensional Shimura variety.***

Now let $(G, X)$ be a Shimura datum with $G^{\mathrm{der}}$ simply connected, and let $T = G/G^{\mathrm{der}}$. Let $Y = T(\mathbb{R})/T(\mathbb{R})^\dagger$. Because $T(\mathbb{Q})$ is dense in $T(\mathbb{R})$ (see 5.4), $Y \cong T(\mathbb{Q})/T(\mathbb{Q})^\dagger$ and

$$T(\mathbb{Q})^\dagger\backslash T(\mathbb{A}_f)/K \cong T(\mathbb{Q})\backslash Y \times T(\mathbb{A}_f)/K$$

Thus, we see that if $G^{\mathrm{der}}$ is simply connected, then

$$\pi_0(\mathrm{Sh}_K(G, X)) \cong \mathrm{Sh}_{\nu(K)}(T, Y).$$

In other words, the set of connected components of the Shimura variety is a zero-dimensional Shimura variety (as promised).

**Additional axioms.** The weight homomorphism $w_X$ is a homomorphism $\mathbb{G}_m \to G_{\mathbb{R}}$ over $\mathbb{R}$ of algebraic groups that are defined over $\mathbb{Q}$. It is therefore defined over $\mathbb{Q}^{\mathrm{al}}$. Some simplifications to the theory occur when some of the following conditions hold:

**SV4:** The weight homomorphism $w_X\colon \mathbb{G}_m \to G_{\mathbb{R}}$ is defined over $\mathbb{Q}$ (we then say that ***the weight is rational).***

**SV5:** The group $Z(\mathbb{Q})$ is discrete in $Z(\mathbb{A}_f)$.

**SV6:** The torus $Z^{\circ}$ splits over a CM-field (see p334 for the notion of a CM-field).

Let $G \to \mathrm{GL}(V)$ be a representation of $G$ (meaning, of course, a $\mathbb{Q}$-representation). Each $h \in X$ defines a hodge structure on $V(\mathbb{R})$. When SV4 holds, these are rational hodge structures (p283). It is hoped that these hodge structures all occur in the cohomology of algebraic varieties and, moreover, that the Shimura variety is a moduli variety for motives when SV4 holds and a fine moduli variety when additionally SV5 holds. This will be discussed in more detail later. In Theorem 5.26 below, we give a criterion for SV5 to hold.

Axiom SV6 makes some statements more natural. For example, when SV6 holds, $w$ is defined over a totally real field.

EXAMPLE 5.24. Let $B$ be a quaternion algebra over a totally real field $F$, and let $G$ be the algebraic group over $\mathbb{Q}$ with $G(\mathbb{Q}) = B^{\times}$. Then, $B \otimes_{\mathbb{Q}} F = \prod_v B \otimes_{F,v} \mathbb{R}$ where $v$ runs over the embeddings of $F$ into $\mathbb{R}$. Thus,

$$\begin{array}{ccccccccccc}
B \otimes_{\mathbb{Q}} \mathbb{R} & \approx & \mathbb{H} & \times & \cdots & \times & \mathbb{H} & \times & M_2(\mathbb{R}) & \times & \cdots & \times & M_2(\mathbb{R}) \\
G(\mathbb{R}) & \approx & \mathbb{H}^{\times} & \times & \cdots & \times & \mathbb{H}^{\times} & \times & \mathrm{GL}_2(\mathbb{R}) & \times & \cdots & \times & \mathrm{GL}_2(\mathbb{R}) \\
h(a+ib) & = & 1 & & \cdots & & 1 & & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) & & \cdots & & \left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right) \\
w(r) & = & 1 & & \cdots & & 1 & & r^{-1}I_2 & & \cdots & & r^{-1}I_2
\end{array}$$

Let $X$ be the $G(\mathbb{R})$-conjugacy class of $h$. Then $(G, X)$ satisfies SV1 and SV2, and so it is a Shimura datum if $B$ splits at at least one real prime of $F$. Let $I = \mathrm{Hom}(F, \mathbb{Q}^{\mathrm{al}}) = \mathrm{Hom}(F, \mathbb{R})$, and let $I_{\mathrm{nc}}$ be the set of $v$ such that $B \otimes_{F,v} \mathbb{R}$ is split. Then $w$ is defined over the subfield of $\mathbb{Q}^{\mathrm{al}}$ fixed by the automorphisms of $\mathbb{Q}^{\mathrm{al}}$ stabilizing $I_{\mathrm{nc}}$. This field is always totally real, and it equals $\mathbb{Q}$ if and only if $I = I_{\mathrm{nc}}$.

**Arithmetic subgroups of tori.** Let $T$ be a torus over $\mathbb{Q}$, and let $T(\mathbb{Z})$ be an arithmetic subgroup of $T(\mathbb{Q})$, for example,

$$T(\mathbb{Z}) = \mathrm{Hom}(X^*(T), \mathcal{O}_L^{\times})^{\mathrm{Gal}(L/\mathbb{Q})},$$

where $L$ is some galois splitting field of $T$. The congruence subgroup problem is known to have a positive answer for tori (Serre 1964, 3.5), i.e., every subgroup of $T(\mathbb{Z})$ of finite index contains a congruence subgroup. Thus the topology induced on $T(\mathbb{Q})$ by that on $T(\mathbb{A}_f)$ has the following description: $T(\mathbb{Z})$ is open, and the induced topology on $T(\mathbb{Z})$ is the profinite topology. In particular,

$$T(\mathbb{Q}) \text{ is discrete } \iff T(\mathbb{Z}) \text{ is discrete } \iff T(\mathbb{Z}) \text{ is finite.}$$

EXAMPLE 5.25. (a) Let $T = \mathbb{G}_m$. Then $T(\mathbb{Z}) = \{\pm 1\}$, and so $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$. This, of course, can be proved directly.

(b) Let $T(\mathbb{Q}) = \{a \in \mathbb{Q}[\sqrt{-1}]^{\times} \mid \mathrm{Nm}(a) = 1\}$. Then $T(\mathbb{Z}) = \{\pm 1, \pm\sqrt{-1}\}$, and so $T(\mathbb{Q})$ is discrete.

(c) Let $T(\mathbb{Q}) = \{a \in \mathbb{Q}[\sqrt{2}]^{\times} \mid \mathrm{Nm}(a) = 1\}$. Then $T(\mathbb{Z}) = \{\pm(1+\sqrt{2})^n \mid n \in \mathbb{Z}\}$, and so neither $T(\mathbb{Z})$ nor $T(\mathbb{Q})$ is discrete.

THEOREM 5.26. *Let $T$ be a torus over $\mathbb{Q}$, and let $T^a = \bigcap_{\chi} \mathrm{Ker}(\chi\colon T \to \mathbb{G}_m)$ (characters $\chi$ of $T$ rational over $\mathbb{Q}$). Then $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if $T^a(\mathbb{R})$ is compact.*

PROOF. According to a theorem of Ono (Serre 1968, pII-39), $T(\mathbb{Z})\cap T^a(\mathbb{Q})$ is of finite index in $T(\mathbb{Z})$, and the quotient $T^a(\mathbb{R})/T(\mathbb{Z})\cap T^a(\mathbb{Q})$ is compact. Now $T(\mathbb{Z})\cap T^a(\mathbb{Q})$ is an arithmetic subgroup of $T^a(\mathbb{Q})$, and hence is discrete in $T^a(\mathbb{R})$. It follows that $T(\mathbb{Z})\cap T^a(\mathbb{Q})$ is finite if and only if $T^a(\mathbb{R})$ is compact. □

For example, in (5.25)(a), $T^a = 1$ and so certainly $T^a(\mathbb{R})$ is compact; in (b), $T^a(\mathbb{R}) = U_1$, which is compact; in (c), $T^a = T$ and $T(\mathbb{R}) = \{(a,b)\in\mathbb{R}\times\mathbb{R} \mid ab = 1\}$, which is not compact.

REMARK 5.27. A torus $T$ over a field $k$ is said to be ***anisotropic*** if there are no characters $\chi\colon T\to\mathbb{G}_m$ defined over $k$. A real torus is anisotropic if and only if it is compact. The torus $T^a =_{\text{df}} \bigcap \operatorname{Ker}(\chi\colon T\to\mathbb{G}_m)$ is the largest anisotropic subtorus of $T$. Thus (5.26) says that $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if the largest anisotropic subtorus of $T$ remains anisotropic over $\mathbb{R}$.

Note that SV5 holds if and only if $(Z^{\circ a})_{\mathbb{R}}$ is anisotropic.

Let $T$ be a torus that splits over CM-field $L$. In this case there is a torus $T^+\subset T$ such that $T^+_L = \bigcap_{\iota\chi=-\chi}\operatorname{Ker}(\chi\colon T_L\to\mathbb{G}_m)$. Then $T(\mathbb{Q})$ is discrete in $T(\mathbb{A}_f)$ if and only if $T^+$ is split, i.e., if and only if the largest subtorus of $T$ that splits over $\mathbb{R}$ is already split over $\mathbb{Q}$.

**Passage to the limit.** Let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$, and let $Z(\mathbb{Q})^-$ be the closure of $Z(\mathbb{Q})$ in $Z(\mathbb{A}_f)$. Then $Z(\mathbb{Q})\cdot K = Z(\mathbb{Q})^-\cdot K$ (in $G(\mathbb{A}_f)$) and

$$\begin{aligned}\operatorname{Sh}_K(G,X) &=_{\text{df}} G(\mathbb{Q})\backslash X\times(G(\mathbb{A}_f)/K)\\ &\cong \frac{G(\mathbb{Q})}{Z(\mathbb{Q})}\Big\backslash X\times(G(\mathbb{A}_f)/Z(\mathbb{Q})\cdot K)\\ &\cong \frac{G(\mathbb{Q})}{Z(\mathbb{Q})}\Big\backslash X\times(G(\mathbb{A}_f)/Z(\mathbb{Q})^-\cdot K).\end{aligned}$$

THEOREM 5.28. *For any Shimura datum $(G,X)$,*

$$\varprojlim_K \operatorname{Sh}_K(G,X) = \frac{G(\mathbb{Q})}{Z(\mathbb{Q})}\Big\backslash X\times(G(\mathbb{A}_f)/Z(\mathbb{Q})^-).$$

*When SV5 holds,*

$$\varprojlim_K \operatorname{Sh}_K(G,X) = G(\mathbb{Q})\backslash X\times G(\mathbb{A}_f).$$

PROOF. The first equality can be proved by the same argument as (4.19), and the second follows from the first (cf. Deligne 1979, 2.1.10, 2.1.11). □

REMARK 5.29. Put $S_K = \operatorname{Sh}_K(G,X)$. For varying $K$, the $S_K$ form a variety (scheme) with a right action of $G(\mathbb{A}_f)$ in the sense of Deligne 1979, 2.7.1. This means the following:

(a) the $S_K$ form an inverse system of algebraic varieties indexed by the compact open subgroups $K$ of $G(\mathbb{A}_f)$ (if $K\subset K'$, there is an obvious quotient map $S_{K'}\to S_K$);
(b) there is an action $\rho$ of $G(\mathbb{A}_f)$ on the system $(S_K)_K$ defined by isomorphisms (of algebraic varieties) $\rho_K(a)\colon S_K\to S_{g^{-1}Kg}$ (on points, $\rho_K(a)$ is $[x,a']\mapsto [x,a'a]$);

(c) for $k \in K$, $\rho_K(k)$ is the identity map; therefore, for $K'$ normal in $K$, there is an action of the finite group $K/K'$ on $S_{K'}$; the variety $S_K$ is the quotient of $S_{K'}$ by the action of $K/K'$.

REMARK 5.30. When we regard the $\mathrm{Sh}_K(G,X)$ as schemes, the inverse limit of the system $\mathrm{Sh}_K(G,X)$ exists:

$$S = \varprojlim \mathrm{Sh}_K(G,X).$$

This is a scheme over $\mathbb{C}$, not(!) of finite type, but it is locally noetherian and regular (cf. Milne 1992, 2.4). There is a right action of $G(\mathbb{A}_f)$ on $S$, and, for $K$ a compact open subgroup of $G(\mathbb{A}_f)$,

$$\mathrm{Sh}_K(G,X) = S/K$$

(Deligne 1979, 2.7.1). Thus, the system $(\mathrm{Sh}_K(G,X))_K$ together with its right action of $G(\mathbb{A}_f)$ can be recovered from $S$ with its right action of $G(\mathbb{A}_f)$. Moreover,

$$S(\mathbb{C}) \cong \varprojlim \mathrm{Sh}_K(G,X)(\mathbb{C}) = \varprojlim G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K.$$

NOTES. Axioms SV1, SV2, SV3, and SV4 are respectively the conditions (2.1.1.1), (2.1.1.2), (2.1.1.3), and (2.1.1.4) of Deligne 1979. Axiom SV5 is weaker than the condition (2.1.1.5) ibid., which requires that $\mathrm{ad}h(i)$ be a Cartan involution on $(G/w(\mathbb{G}_m))_{\mathbb{R}}$, i.e., that $(Z^{\circ}/w(\mathbb{G}_m))_{\mathbb{R}}$ be anisotropic.

## 6. The Siegel modular variety

In this section, we study the most important Shimura variety, namely, the Siegel modular variety.

**Dictionary.** Let $V$ be an $\mathbb{R}$-vector space. Recall (2.4) that to give a $\mathbb{C}$-structure $J$ on $V$ is the same as to give a hodge structure $h_J$ on $V$ of type $(-1,0),(0,-1)$. Here $h_J$ is the restriction to $\mathbb{C}^{\times}$ of the homomorphism

$$a+bi \mapsto a+bJ\colon \mathbb{C} \to \mathrm{End}_{\mathbb{R}}(V).$$

For the hodge decompostion $V(\mathbb{C}) = V^{-1,0} \oplus V^{-1,0}$,

| | $V^{-1,0}$ | $V^{0,-1}$ |
|---|---|---|
| $J$ acts as | $+i$ | $-i$ |
| $h_J(z)$ acts as | $z$ | $\bar{z}$ |

Let $\psi$ be a nondegenerate $\mathbb{R}$-bilinear alternating form on $V$. A direct calculation shows that

$$\psi(Ju,Jv) = \psi(u,v) \iff \psi(zu,zv) = |z|^2\psi(u,v) \text{ for all } z \in \mathbb{C}.$$

Let $\psi_J(u,v) = \psi(u,Jv)$. Then

$$\psi(Ju,Jv) = \psi(u,v) \iff \psi_J \text{ is symmetric}$$

and

$$\begin{array}{l}\psi(Ju,Jv)=\psi(u,v) \text{ and}\\ \psi_J \text{ is positive definite}\end{array} \overset{(2.12)}{\iff} \begin{array}{l}\psi \text{ is a polarization of the}\\ \text{hodge structure } (V,h_J).\end{array}$$

**Symplectic spaces.** Let $k$ be a field of characteristic $\neq 2$, and let $(V, \psi)$ be a ***symplectic space*** of dimension $2n$ over $k$, i.e., $V$ is a $k$-vector space of dimension $2n$ and $\psi$ is a nondegenerate alternating form $\psi$. A subspace $W$ of $V$ is ***totally isotropic*** if $\psi(W, W) = 0$. A ***symplectic basis*** of $V$ is a basis $(e_{\pm i})_{1 \leq i \leq n}$ such that

$$\psi(e_i, e_{-i}) = 1 \text{ for } 1 \leq i \leq n,$$
$$\psi(e_i, e_j) = 0 \text{ for } \quad j \neq \pm i.$$

LEMMA 6.1. *Let $W$ be a totally isotropic subspace of $V$. Then any basis of $W$ can be extended to a symplectic basis for $V$. In particular, $V$ has symplectic bases (and two symplectic spaces of the same dimension are isomorphic).*

PROOF. Standard. □

Thus, a maximal totally isotropic subspace of $V$ will have dimension $n$. Such subspaces are called ***lagrangians***.

Let $\mathrm{GSp}(\psi)$ be the group of ***symplectic similitudes*** of $(V, \psi)$, i.e., the group of automorphisms of $V$ preserving $\psi$ up to a scalar. Thus

$$\mathrm{GSp}(\psi)(k) = \{g \in \mathrm{GL}(V) \mid \psi(gu, gv) = \nu(g) \cdot \psi(u, v) \text{ some } \nu(g) \in k^\times\}.$$

Define $\mathrm{Sp}(\psi)$ by the exact sequence

$$1 \to \mathrm{Sp}(\psi) \to \mathrm{GSp}(\psi) \xrightarrow{\nu} \mathbb{G}_m \to 1.$$

Then $\mathrm{GSp}(\psi)$ has derived group $\mathrm{Sp}(\psi)$, centre $\mathbb{G}_m$, and adjoint group $\mathrm{GSp}(\psi)/\mathbb{G}_m = \mathrm{Sp}(\psi)/\pm I$.

For example, when $V$ has dimension 2, there is only one nondegenerate alternating form on $V$ up to scalars, which must therefore be preserved up to scalars by any automorphism, and so $\mathrm{GSp}(\psi) = \mathrm{GL}_2$ and $\mathrm{Sp}(\psi) = \mathrm{SL}_2$.

The group $\mathrm{Sp}(\psi)$ acts simply transitively on the set of symplectic bases: if $(e_{\pm i})$ and $(f_{\pm i})$ are bases of $V$, then there is a unique $g \in \mathrm{GL}_{2n}(k)$ such that $ge_{\pm i} = f_{\pm i}$, and if $(e_{\pm i})$ and $(f_{\pm i})$ are both symplectic, then $g \in \mathrm{Sp}(\psi)$.

**The Shimura datum attached to a symplectic space.** Fix a symplectic space $(V, \psi)$ over $\mathbb{Q}$, and let $G = \mathrm{GSp}(\psi)$ and $S = \mathrm{Sp}(\psi) = G^{\mathrm{der}}$.

Let $J$ be a complex structure on $V(\mathbb{R})$ such that $\psi(Ju, Jv) = \psi(u, v)$. Then $J \in S(\mathbb{R})$, and $h_J(z)$ lies in $G(\mathbb{R})$ (and in $S(\mathbb{R})$ if $|z| = 1$) — see the dictionary. We say that $J$ is ***positive*** (resp. ***negative***) if $\psi_J(u, v) =_{\mathrm{df}} \psi(u, Jv)$ is positive definite (resp. negative definite).

Let $X^+$ (resp. $X^-$) denote the set of positive (resp. negative) complex structures on $V(\mathbb{R})$, and let $X = X^+ \sqcup X^-$. Then $G(\mathbb{R})$ acts on $X$ according to the rule

$$(g, J) \mapsto gJg^{-1},$$

and the stabilizer in $G(\mathbb{R})$ of $X^+$ is

$$G(\mathbb{R})^+ = \{g \in G(\mathbb{R}) \mid \nu(g) > 0\}.$$

For a symplectic basis $(e_{\pm i})$ of $V$, define $J$ by $Je_{\pm i} = \pm e_{\mp i}$, i.e.,

$$e_i \overset{J}{\longmapsto} e_{-i} \overset{J}{\longmapsto} -e_i, \quad 1 \leq i \leq n.$$

Then $J^2 = -1$ and $J \in X^+$ — in fact, $(e_i)_i$ is an orthonormal basis for $\psi_J$. Conversely, if $J \in X^+$, then $J$ has this description relative to any orthonormal basis for the positive definite form $\psi_J$. The map from symplectic bases to $X^+$ is equivariant for the actions of $S(\mathbb{R})$. Therefore, $S(\mathbb{R})$ acts transitively on $X^+$, and $G(\mathbb{R})$ acts transitively on $X$.

For $J \in X$, let $h_J$ be the corresponding homomorphism $\mathbb{C}^\times \to G(\mathbb{R})$. Then $h_{gJg^{-1}}(z) = gh_J(z)g^{-1}$. Thus $J \mapsto h_J$ identifies $X$ with a $G(\mathbb{R})$-conjugacy class of homomorphisms $h\colon \mathbb{C}^\times \to G(\mathbb{R})$. We check that $(G, X)$ satisfies the axioms SV1–SV6.

(SV1). For $h \in X$, let $V^+ = V^{-1,0}$ and $V^- = V^{0,-1}$, so that $V(\mathbb{C}) = V^+ \oplus V^-$ with $h(z)$ acting on $V^+$ and $V^-$ as multiplication by $z$ and $\overline{z}$ respectively. Then

$$\begin{array}{ccccccccc} \mathrm{Hom}(V(\mathbb{C}), V(\mathbb{C})) & = & \mathrm{Hom}(V^+, V^+) & \oplus & \mathrm{Hom}(V^+, V^-) & \oplus & \mathrm{Hom}(V^-, V^+) & \oplus & \mathrm{Hom}(V^-, V^-) \\ h(z) \text{ acts as} & & 1 & & z/\overline{z} & & \overline{z}/z & & 1 \end{array}$$

The Lie algebra of $G$ is the subspace

$$\mathrm{Lie}(G) = \{f \in \mathrm{Hom}(V, V) \mid \psi(f(u), v) + \psi(u, f(v)) = 0\},$$

of $\mathrm{End}(V)$, and so SV1 holds.

(SV2). We have to show that $\mathrm{ad}J$ is a Cartan involution on $G^{\mathrm{ad}}$. But, $J^2 = -1$ lies in the centre of $S(\mathbb{R})$ and $\psi$ is a $J$-polarization for $S_{\mathbb{R}}$ in the sense of (1.20), which shows that $\mathrm{ad}J$ is a Cartan involution for $S$.

(SV3). In fact, $G^{\mathrm{ad}}$ is $\mathbb{Q}$-simple, and $G^{\mathrm{ad}}(\mathbb{R})$ is not compact.

(SV4). For $r \in \mathbb{R}^\times$, $w_h(r)$ acts on both $V^{-1,0}$ and $V^{0,-1}$ as $v \mapsto rv$. Therefore, $w_X$ is the homomorphism $\mathbb{G}_{m\mathbb{R}} \to \mathrm{GL}(V(\mathbb{R}))$ sending $r \in \mathbb{R}^\times$ to multplication by $r$. This is defined over $\mathbb{Q}$.

(SV5). The centre of $G$ is $\mathbb{G}_m$, and $\mathbb{Q}^\times$ is discrete in $\mathbb{A}_f^\times$ (see 5.25).

(SV6). The centre of $G$ is split already over $\mathbb{Q}$.

We often write $(G(\psi), X(\psi))$ for the Shimura datum defined by a symplectic space $(V, \psi)$, and $(S(\psi), X(\psi)^+)$ for the connected Shimura datum.

Exercise 6.2. (a) Show that for any $h \in X(\psi)$, $\nu(h(z)) = z\overline{z}$. [Hint: for nonzero $v^+ \in V^+$ and $v^- \in V^-$, compute $\psi_{\mathbb{C}}(h(z)v^+, h(z)v^-)$ in two different ways.]

(b) Show that the choice of a symplectic basis for $V$ identifies $X^+$ with $\mathcal{H}_g$ as an $\mathrm{Sp}(\psi)$-set (see 1.2).

**The Siegel modular variety.** Let $(G, X) = (G(\psi), X(\psi))$ be the Shimura datum defined by a symplectic space $(V, \psi)$ over $\mathbb{Q}$. The ***Siegel modular variety attached to*** $(V, \psi)$ is the Shimura variety $\mathrm{Sh}(G, X)$.

Let $V(\mathbb{A}_f) = \mathbb{A}_f \otimes_{\mathbb{Q}} V$. Then $G(\mathbb{A}_f)$ is the group of $\mathbb{A}_f$-linear automorphisms of $V(\mathbb{A}_f)$ preserving $\psi$ up to multiplication by an element of $\mathbb{A}_f^\times$.

Let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$, and let $\mathcal{H}_K$ be the set of triples $((W, h), s, \eta K)$ where

- $(W, h)$ is a rational hodge structure of type $(-1, 0), (0, -1)$;
- $\pm s$ is a polarization for $(W, h)$;
- $\eta K$ is a $K$-orbit of $\mathbb{A}_f$-linear isomorphisms $V(\mathbb{A}_f) \to W(\mathbb{A}_f)$ sending $\psi$ to an $\mathbb{A}_f^\times$-multiple of $s$.

An ***isomorphism***

$$((W, h), s, \eta K) \to ((W', h'), s', \eta' K)$$

of triples is an isomorphism $b\colon (W, h) \to (W', h')$ of rational hodge structures such that $b(s) = cs'$ some $c \in \mathbb{Q}^\times$ and $b \circ \eta = \eta' \mod K$.

Note that to give an element of $\mathcal{H}_K$ amounts to giving a symplectic space $(W, s)$ over $\mathbb{Q}$, a complex structure on $W$ that is positive or negative for $s$, and $\eta K$. The existence of $\eta$ implies that $\dim W = \dim V$, and so $(W, s)$ and $(V, \psi)$ are isomorphic. Choose an isomorphism $a\colon W \to V$ sending $\psi$ to a $\mathbb{Q}^{\times}$-multiple of $s$. Then

$$ah =_{\mathrm{df}} (z \mapsto a \circ h(z) \circ a^{-1})$$

lies in $X$, and

$$V(\mathbb{A}_f) \xrightarrow{\eta} W(\mathbb{A}_f) \xrightarrow{a} V(\mathbb{A}_f)$$

lies in $G(\mathbb{A}_f)$. Any other isomorphism $a'\colon W \to V$ sending $\psi$ to a multiple of $s$ differs from $a$ by an element of $G(\mathbb{Q})$, say, $a' = q \circ a$ with $q \in G(\mathbb{Q})$. Replacing $a$ with $a'$ only replaces $(ah, a \circ \eta)$ with $(qah, qa \circ \eta)$. Similarly, replacing $\eta$ with $\eta k$ replaces $(ah, a \circ \eta)$ with $(ah, a \circ \eta k)$. Therefore, the map

$$(W \ldots) \mapsto [ah, a \circ \eta]_K \colon \mathcal{H}_K \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$$

is well-defined.

Proposition 6.3. *The set* $\mathrm{Sh}_K(G, X)$ *classifies the triples in* $\mathcal{H}_K$ *modulo isomorphism. More precisely, the map* $(W, \ldots) \mapsto [ah, a \circ \eta]_K$ *defines a bijection*

$$\mathcal{H}_K/{\approx} \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K.$$

Proof. It is straightforward to check that the map sends isomorphic triples to the same class, and that two triples are isomorphic if they map to the same class. The map is onto because $[h, g]$ is the image of $((V, h), \psi, gK)$. □

**Complex abelian varieties.** An ***abelian variety*** $A$ over a field $k$ is a connected projective algebraic variety over $k$ together with a group structure given by regular maps. A one-dimensional abelian variety is an elliptic curve. Happily, a theorem, whose origins go back to Riemann, reduces the study of abelian varieties over $\mathbb{C}$ to multilinear algebra.

Recall that a lattice in a real or complex vector space $V$ is the $\mathbb{Z}$-module generated by an $\mathbb{R}$-basis for $V$. For a lattice $\Lambda$ in $\mathbb{C}^n$, make $\mathbb{C}^n/\Lambda$ into a complex manifold by endowing it with the quotient structure. A ***complex torus*** is a complex manifold isomorphic to $\mathbb{C}^n/\Lambda$ for some lattice $\Lambda$ in $\mathbb{C}^n$.

Note that $\mathbb{C}^n$ is the universal covering space of $M = \mathbb{C}^n/\Lambda$ with $\Lambda$ as its group of covering transformations, and $\pi_1(M, 0) = \Lambda$ (Hatcher 2002, 1.40). Therefore, (ib. 2A.1)

$$H_1(M, \mathbb{Z}) \cong \Lambda \tag{31}$$

and (Greenberg 1967, 23.14)

$$H^1(M, \mathbb{Z}) \cong \mathrm{Hom}(\Lambda, \mathbb{Z}). \tag{32}$$

Proposition 6.4. *Let* $M = \mathbb{C}^n/\Lambda$. *There is a canonical isomorphism*

$$H^n(M, \mathbb{Z}) \cong \mathrm{Hom}(\textstyle\bigwedge^n \Lambda, \mathbb{Z}),$$

*i.e.,* $H^n(M, \mathbb{Z})$ *is canonically isomorphic to the set of* $n$*-alternating forms* $\Lambda \times \cdots \times \Lambda \to \mathbb{Z}$.

Proof. From (32), we see that

$$\textstyle\bigwedge^n H^1(M, \mathbb{Z}) \cong \bigwedge^n \mathrm{Hom}(\Lambda, \mathbb{Z}).$$

Since[13]

$$\bigwedge^n \mathrm{Hom}(\Lambda,\mathbb{Z}) \cong \mathrm{Hom}(\bigwedge^n \Lambda,\mathbb{Z}),$$

we see that it suffices to show that cup-product defines an isomorphism

$$\bigwedge^n H^1(M,\mathbb{Z}) \to H^n(M,Z). \tag{33}$$

Let $\mathcal{T}$ be the class of topological manifolds $M$ whose cohomology groups are free $\mathbb{Z}$-modules of finite rank and for which the maps (33) are isomorphisms for all $n$. Certainly, the circle $S^1$ is in $\mathcal{T}$ (its cohomology groups are $\mathbb{Z}$, $\mathbb{Z}$, $0,\ldots$), and the Künneth formula (Hatcher 2002, 3.16 et seq.) shows that if $M_1$ and $M_2$ are in $\mathcal{T}$, then so also is $M_1 \times M_2$. As a topological manifold, $\mathbb{C}^n/\Lambda \approx (S^1)^{2n}$, and so $M$ is in $\mathcal{T}$. □

PROPOSITION 6.5. *A linear map $\alpha\colon \mathbb{C}^n \to \mathbb{C}^{n'}$ such that $\alpha(\Lambda) \subset \Lambda'$ defines a holomorphic map $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ sending $0$ to $0$, and every holomorphic map $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ sending $0$ to $0$ is of this form (for a unique $\alpha$).*

PROOF. The map $\mathbb{C}^n \xrightarrow{\alpha} \mathbb{C}^{n'} \to \mathbb{C}^{n'}/\Lambda'$ is holomorphic, and it factors through $\mathbb{C}^n/\Lambda$. Because $\mathbb{C}/\Lambda$ has the quotient structure, the resulting map $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ is holomorphic. Conversely, let $\varphi\colon \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ be a holomorphic map such that $\varphi(0) = 0$. Then $\mathbb{C}^n$ and $\mathbb{C}^{n'}$ are universal covering spaces of $\mathbb{C}^n/\Lambda$ and $\mathbb{C}^{n'}/\Lambda'$, and a standard result in topology (Hatcher 2002, 1.33, 1.34) shows that $\varphi$ lifts uniquely to a continuous map $\tilde{\varphi}\colon \mathbb{C}^n \to \mathbb{C}^{n'}$ such that $\tilde{\varphi}(0) = 0$:

$$\begin{tikzcd}
\mathbb{C}^n \arrow[r, "\tilde{\varphi}"] \arrow[d] & \mathbb{C}^{n'} \arrow[d] \\
\mathbb{C}^n/\Lambda \arrow[r, "\varphi"] & \mathbb{C}^{n'}/\Lambda'.
\end{tikzcd}$$

Because the vertical arrows are local isomorphisms, $\tilde{\varphi}$ is automatically holomorphic. For any $\omega \in \Lambda$, the map $z \mapsto \tilde{\varphi}(z+\omega) - \tilde{\varphi}(z)$ is continuous and takes values in $\Lambda' \subset \mathbb{C}$. Because $\mathbb{C}^n$ is connected and $\Lambda'$ is discrete, it must be constant. Therefore, for each $j$, $\frac{\partial\tilde{\varphi}}{\partial z_j}$ is a doubly periodic function, and so defines a holomorphic function $\mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}$, which must be constant (because $\mathbb{C}^n/\Lambda$ is compact). Write $\tilde{\varphi}$ as an $n'$-tuple $(\tilde{\varphi}_1,\ldots,\tilde{\varphi}_{n'})$ of holomorphic functions $\tilde{\varphi}_i$ in $n$ variables. Because $\tilde{\varphi}_i(0) = 0$ and $\frac{\partial\tilde{\varphi}_i}{\partial z_j}$ is constant for each $j$, the power series expansion of $\tilde{\varphi}_i$ at 0 is of the form $\sum a_{ij}z_j$. Now $\tilde{\varphi}_i$ and $\sum a_{ij}z_j$ are holomorphic functions on $\mathbb{C}^n$ that coincide on a neighbourhood of 0, and so are equal on the whole of $\mathbb{C}^n$. We have shown that

$$\tilde{\varphi}(z_1,\ldots,z_n) = (\textstyle\sum a_{1j}z_j,\ldots,\sum a_{n'j}z_j). \qquad \square$$

ASIDE 6.6. The proposition shows that every holomorphic map $\varphi\colon \mathbb{C}^n/\Lambda \to \mathbb{C}^{n'}/\Lambda'$ such that $\varphi(0) = 0$ is a homomorphism. A similar statement is true for abelian varieties over any field $k$: a regular map $\varphi\colon A \to B$ of abelian varieties such that $\varphi(0) = 0$ is a homomorphism (AG, 7.14). For example, the map sending an element to its inverse is a homomorphism, which implies that the group law on $A$

[13]For a free $\mathbb{Z}$-module $\Lambda$ of finite rank, the pairing

$$\bigwedge^n \Lambda^\vee \times \bigwedge^n \Lambda \to \mathbb{Z}$$

determined by

$$(f_1 \wedge \cdots \wedge f_n, v_1 \otimes \cdots \otimes v_n) = \det(f_i(v_j))$$

is nondegenerate (since it is modulo $p$ for every $p$ — see Bourbaki 1958, §8).

is commutative. Also, the group law on an abelian variety is uniquely determined by the zero element.

Let $M = \mathbb{C}^n/\Lambda$ be a complex torus. The isomorphism $\mathbb{R} \otimes \Lambda \cong \mathbb{C}^n$ defines a complex structure $J$ on $\mathbb{R} \otimes \Lambda$. A ***riemann form*** for $M$ is an alternating form $\psi\colon \Lambda \times \Lambda \to \mathbb{Z}$ such that $\psi_{\mathbb{R}}(Ju, Jv) = \psi_{\mathbb{R}}(u, v)$ and $\psi_{\mathbb{R}}(u, Ju) > 0$ for $u \neq 0$. A complex torus $\mathbb{C}^n/\Lambda$ is said to be ***polarizable*** if there exists a riemann form.

THEOREM 6.7. *The complex torus $\mathbb{C}^n/\Lambda$ is projective if and only if it is polarizable.*

PROOF. See Mumford 1970, Chapter I, (or Murty 1993, 4.1, for the "if" part). Alternatively, one can apply the Kodaira embedding theorem (Voisin 2002, Th. 7.11, 7.2.2). $\square$

Thus, by Chow's theorem (3.11), a polarizable complex torus is a projective algebraic variety, and holomorphic maps of polarizable complex tori are regular. Conversely, it is easy to see that the complex manifold associated with an abelian variety is a complex torus: let $\mathrm{Tgt}_0 A$ be the tangent space to $A$ at 0; then the exponential map $\mathrm{Tgt}_0 A \to A(\mathbb{C})$ is a surjective homomorphism of Lie groups with kernel a lattice $\Lambda$, which induces an isomorphism $(\mathrm{Tgt}_0 A)/\Lambda \cong A(\mathbb{C})$ of complex manifolds (Mumford 1970, p2).

For a complex torus $M = \mathbb{C}^n/\Lambda$, the isomorphism $\Lambda \otimes_{\mathbb{Z}} \mathbb{R} \cong \mathbb{C}^n$ endows $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ with a complex structure, and hence endows $\Lambda \cong H_1(M, \mathbb{Z})$ with an integral hodge structure of weight $-1$. Note that a riemann form for $M$ is nothing but a polarization of the integral hodge structure $\Lambda$.

THEOREM 6.8 (Riemann's theorem). [14]*The functor $A \mapsto H_1(A, \mathbb{Z})$ is an equivalence from the category* $\mathsf{AV}$ *of abelian varieties over $\mathbb{C}$ to the category of polarizable integral hodge structures of type $(-1, 0), (0, -1)$.*

PROOF. We have functors

$$\mathsf{AV} \xrightarrow{A \mapsto A^{\mathrm{an}}} \{\text{category of polarizable complex tori }\}$$
$$\xrightarrow{M \mapsto H_1(M,\mathbb{Z})} \{\text{category of polarizable integral hodge structures of type}(-1,0),(0,-1)\}.$$

The first is fully faithful by Chow's theorem (3.11), and it is essentially surjective by Theorem 6.7; the second is fully faithful by Proposition 6.5, and it is obviously essentially surjective. $\square$

Let $\mathsf{AV}^0$ be the category whose objects are abelian varieties over $\mathbb{C}$ and whose morphisms are

$$\mathrm{Hom}_{\mathsf{AV}^0}(A, B) = \mathrm{Hom}_{\mathsf{AV}}(A, B) \otimes \mathbb{Q}.$$

COROLLARY 6.9. *The functor $A \mapsto H_1(A, \mathbb{Q})$ is an equivalence from the category* $\mathsf{AV}^0$ *to the category of polarizable rational hodge structures of type $(-1, 0), (0, -1)$.*

PROOF. Immediate consequence of the theorem. $\square$

[14]In fact, it should be called the "theorem of Riemann, Frobenius, Weierstrass, Poincaré, Lefschetz, et al." (see Shafarevich 1994, Historical Sketch, 5), but "Riemann's theorem" is shorter.

REMARK 6.10. Recall that in the dictionary between complex structures $J$ on a real vector space $V$ and hodge structures of type $(-1,0),(0,-1)$,

$$(V,J) \cong V(\mathbb{C})/V^{-1,0} = V(\mathbb{C})/F^0.$$

Since the hodge structure on $H_1(A,\mathbb{R})$ is defined by the isomorphism $\mathrm{Tgt}_0(A) \cong H_1(A,\mathbb{R})$, we see that

$$\mathrm{Tgt}_0(A) \cong H_1(A,\mathbb{C})/F^0 \tag{34}$$

(isomorphism of complex vector spaces).

**A modular description of the points of the Siegel variety.** Let $\mathcal{M}_K$ be the set of triples $(A,s,\eta K)$ in which $A$ is an abelian variety over $\mathbb{C}$, $s$ is an alternating form on $H_1(A,\mathbb{Q})$ such that $s$ or $-s$ is a polarization on $H_1(A,\mathbb{Q})$, and $\eta$ is an isomorphism $V(\mathbb{A}_f) \to V_f(\mathbb{A}_f)$ sending $\psi$ to a multiple of $s$ by an element of $\mathbb{A}_f^{\times}$. An isomorphism from one triple $(A,s,\eta K)$ to a second $(A',s',\eta' K)$ is an isomorphism $A \to A'$ (as objects in $\mathsf{AV}^0$) sending $s$ to a multiple of $s'$ by an element of $\mathbb{Q}^{\times}$ and $\eta K$ to $\eta' K$.

THEOREM 6.11. *The set $\mathrm{Sh}_K(G,X)$ classifies the triples $(A,s,\eta K)$ in $\mathcal{M}_K$ modulo isomorphism, i.e., there is a canonical bijection $\mathcal{M}_K/\approx \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$.*

PROOF. Combine (6.9) with (6.3). □

## 7. Shimura varieties of hodge type

In this section, we examine one important generalization of Siegel modular varieties.

DEFINITION 7.1. A Shimura datum $(G,X)$ is of ***hodge type*** if there exists a symplectic space $(V,\psi)$ over $\mathbb{Q}$ and an injective homomorphism $\rho\colon G \hookrightarrow G(\psi)$ carrying $X$ into $X(\psi)$. The Shimura variety $\mathrm{Sh}(G,X)$ is then said to be of ***hodge type***. Here $(G(\psi),X(\psi))$ denotes the Shimura datum defined by $(V,\psi)$.

The composite of $\rho$ with the character $\nu$ of $G(\psi)$ is a character of $G$, which we again denote by $\nu$. Let $\mathbb{Q}(r)$ denote the vector space $\mathbb{Q}$ with $G$ acting by $r\nu$, i.e., $g \cdot v = \nu(g)^r \cdot v$. For each $h \in X$, $(\mathbb{Q}(r), h \circ \nu)$ is a rational hodge structure of type $(-r,-r)$ (apply 6.2a), and so this notation is consistent with that in (2.6).

LEMMA 7.2. *There exist multilinear maps $t_i\colon V \times \cdots \times V \to \mathbb{Q}(r_i)$, $1 \le i \le n$, such that $G$ is the subgroup of $G(\psi)$ fixing the $t_i$.*

PROOF. According to Deligne 1982, 3.1, there exist tensors $t_i$ in $V^{\otimes r_i} \otimes V^{\vee \otimes s_i}$ such that this is true. But $\psi$ defines an isomorphism $V \cong V^{\vee} \otimes \mathbb{Q}(1))$, and so

$$V^{\otimes r_i} \otimes V^{\vee\otimes s_i} \cong V^{\vee \otimes (r_i+s_i)} \otimes \mathbb{Q}(r_i) \cong \mathrm{Hom}(V^{\otimes(r_i+s_i)}, \mathbb{Q}(r_i)). \qquad \square$$

Let $(G,X)$ be of hodge type. Choose an embedding of $(G,X)$ into $(G(\psi),X(\psi))$ for some symplectic space $(V,\psi)$ and multilinear maps $t_1,\ldots,t_n$ as in the lemma. Let $\mathcal{H}_K$ be the set of triples $((W,h),(s_i)_{0\le i\le n},\eta K)$ in which

- $(W,h)$ is a rational hodge structure of type $(-1,0),(0,-1)$,
- $\pm s_0$ is a polarization for $(W,h)$,
- $s_1,\ldots,s_n$ are multilinear maps $s_i\colon W \times \cdots \times W \to \mathbb{Q}(r_i)$, and

- $\eta K$ is a $K$-orbit of isomorphisms $V(\mathbb{A}_f) \to W(\mathbb{A}_f)$ sending $\psi$ onto an $\mathbb{A}_f^\times$-multiple of $s_0$ and each $t_i$ to $s_i$,

satisfying the following condition:

> (*) there exists an isomorphism $a\colon W \to V$ sending $s_0$ to a $\mathbb{Q}^\times$-multiple of $\psi$, $s_i$ to $t_i$ each $i \geq 1$, and $h$ onto an element of $X$.

An isomorphism from one triple $(W, \ldots)$ to a second $(W', \ldots)$ is an isomorphism $(W, h) \to (W', h')$ of rational hodge structures sending $s_0$ to a $\mathbb{Q}^\times$-multiple of $s_0'$, $s_i$ to $s_i'$ for $i > 0$, and $\eta K$ to $\eta' K$.

PROPOSITION 7.3. *The set* $\mathrm{Sh}_K(G, X)(\mathbb{C})$ *classifies the triples in* $\mathcal{H}_K$ *modulo isomorphism.*

PROOF. Choose an isomorphism $a\colon W \to V$ as in (*), and consider the pair $(ah, a \circ \eta)$. By assumption $ah \in X$ and $a \circ \eta$ is a symplectic similitude of $(V(\mathbb{A}_f), \psi)$ fixing the $t_i$, and so $(ah, a \circ \eta) \in X \times G(\mathbb{A}_f)$. The isomorphism $a$ is determined up to composition with an element of $G(\mathbb{Q})$ and $\eta$ is determined up to composition with an element of $K$. It follows that the class of $(ah, a \circ \eta)$ in $G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$ is well-defined. The proof that $(W, \ldots) \mapsto [ah, a \circ \eta]_K$ gives a bijection from the set of isomorphism classes of triples in $\mathcal{H}_K$ onto $\mathrm{Sh}_K(G, X)(\mathbb{C})$ is now routine (cf. the proof of 6.3). □

Let $t\colon V \times \cdots \times V \to \mathbb{Q}(r)$ ($m$-copies of $V$) be a multilinear form fixed by $G$, i.e., such that

$$t(gv_1, \ldots, gv_m) = \nu(g)^r \cdot t(v_1, \ldots, v_m), \text{ for all } v_1, \ldots, v_m \in V, \quad g \in G(\mathbb{Q}).$$

For $h \in X$, this equation shows that $t$ defines a morphism of hodge structures $(V, h)^{\otimes m} \to \mathbb{Q}(r)$. On comparing weights, we see that if $t$ is nonzero, then $m = 2r$.

Now let $A$ be an abelian variety over $\mathbb{C}$, and let $V = H_1(A, \mathbb{Q})$. Then (see 6.4)

$$H^m(A, \mathbb{Q}) \cong \mathrm{Hom}(\textstyle\bigwedge^m V, \mathbb{Q}).$$

We say that $t \in H^{2r}(A, \mathbb{Q})$ is a ***hodge tensor for*** $A$ if the corresponding map

$$V^{\otimes 2r} \to \textstyle\bigwedge^{2r} V \to \mathbb{Q}(r)$$

is a morphism of hodge structures.

Let $(G, X) \hookrightarrow (G(\psi), X(\psi))$ and $t_1, \ldots, t_n$ be as above. Let $\mathcal{M}_K$ be the set of triples $(A, (s_i)_{0 \leq i \leq n}, \eta K)$ in which

- $A$ is a complex abelian variety,
- $\pm s_0$ is a polarization for the rational hodge structure $H_1(A, \mathbb{Q})$,
- $s_1, \ldots, s_n$ are hodge tensors for $A$ or its powers, and
- $\eta K$ is a $K$-orbit of $\mathbb{A}_f$-linear isomorphisms $V(\mathbb{A}_f) \to V_f(A)$ sending $\psi$ onto an $\mathbb{A}_f^\times$-multiple of $s_0$ and each $t_i$ to $s_i$,

satisfying the following condition:

> (**) there exists an isomorphism $a\colon H_1(A, \mathbb{Q}) \to V$ sending $s_0$ to a $\mathbb{Q}^\times$-multiple of $\psi$, $s_i$ to $t_i$ each $i \geq 1$, and $h$ to an element of $X$.

An isomorphism from one triple $(A, (s_i)_i, \eta K)$ to a second $(A', (s_i'), \eta' K)$ is an isomorphism $A \to A'$ (as objects of $\mathsf{AV}^0$) sending $s_0$ to a multiple of $s_0'$ by an element of $\mathbb{Q}^\times$, each $s_i$ to $s_i'$, and $\eta$ to $\eta'$ modulo $K$.

THEOREM 7.4. *The set* $\mathrm{Sh}_K(G,X)(\mathbb{C})$ *classifies the triples in* $\mathcal{M}_K$ *modulo isomorphism.*

PROOF. Combine Propositions 7.3 and 6.9. □

The problem with Theorem 7.4 is that it is difficult to check whether a triple satisfies the condition (**). In the next section, we show that when the hodge tensors are endomorphisms of the abelian variety, then it is sometimes possible to replace (**) by a simpler trace condition.

REMARK 7.5. When we write $A(\mathbb{C}) = \mathbb{C}^g/\Lambda$, then (see 6.4),

$$H^m(A,\mathbb{Q}) \cong \mathrm{Hom}(\textstyle\bigwedge^m \Lambda, \mathbb{Q})$$

Now $\Lambda \otimes \mathbb{C} \cong T \oplus \overline{T}$ where $T = \mathrm{Tgt}_0(A)$. Therefore,

$$H^m(A,\mathbb{C}) \cong \mathrm{Hom}(\textstyle\bigwedge^m(\Lambda\otimes\mathbb{C}),\mathbb{C}) \cong \mathrm{Hom}(\displaystyle\bigoplus_{p+q=m} \textstyle\bigwedge^p T \otimes \bigwedge^q \overline{T}, \mathbb{C}) \cong \displaystyle\bigoplus_{p+q=m} H^{p,q}$$

where

$$H^{p,q} = \mathrm{Hom}(\textstyle\bigwedge^p T \otimes \bigwedge^q \overline{T}, \mathbb{C}).$$

This rather ad hoc construction of the hodge structure on $H^m$ does agree with the usual construction (2.5) — see Mumford 1970, Chapter I. A hodge tensor on $A$ is an element of

$$H^{2r}(A,\mathbb{Q}) \cap H^{r,r} \quad (\text{intersection inside } H^{2r}(A,\mathbb{C})).$$

The Hodge conjecture predicts that all hodge tensors are the cohomology classes of algebraic cycles with $\mathbb{Q}$-coefficients. For $r=1$, this is known even over $\mathbb{Z}$. The exponential sequence

$$0 \to \mathbb{Z} \longrightarrow \mathcal{O}_A \xrightarrow{z\mapsto \exp(2\pi i z)} \mathcal{O}_A^\times \to 0$$

gives a cohomology sequence

$$H^1(A,\mathcal{O}_A^\times) \to H^2(A,\mathbb{Z}) \to H^2(A,\mathcal{O}_A).$$

The cohomology group $H^1(A,\mathcal{O}_A^\times)$ classifies the divisors on $A$ modulo linear equivalence, i.e., $\mathrm{Pic}(A) \cong H^1(A,\mathcal{O}_A^\times)$, and the first arrow maps a divisor to its cohomology class. A class in $H^2(A,\mathbb{Z})$ maps to zero in $H^2(A,\mathcal{O}_A) = H^{0,2}$ if and only if it maps to zero in its complex conjugate $H^{2,0}$. Therefore, we see that

$$\mathrm{Im}(\mathrm{Pic}(A)) = H^2(A,\mathbb{Z}) \cap H^{1,1}.$$

## 8. PEL Shimura varieties

Throughout this section, $k$ is a field of characteristic zero. Bilinear forms are always nondegenerate.

**Algebras with involution.** By a ***k-algebra*** I mean a ring $B$ containing $k$ in its centre and finite dimensional over $k$. A $k$-algebra $A$ is ***simple*** if it contains no two-sided ideals except 0 and $A$. For example, every matrix algebra $M_n(D)$ over a division algebra $D$ is simple, and conversely, Wedderburn's theorem says that every simple algebra is of this form (CFT, IV 1.9). Up to isomorphism, a simple $k$-algebra has only one simple module (ibid, IV 1.15). For example, up to isomorphism, $D^n$ is the only simple $M_n(D)$-module.

Let $B = B_1 \times \cdots \times B_n$ be a product of simple $k$-algebras (a ***semisimple k-algebra)***. A simple $B_i$-module $M_i$ becomes a simple $B$-module when we let $B$

act through the quotient map $B \to B_i$. These are the only simple $B$-modules, and every $B$-module is a direct sum of simple modules. A $B$-module $M$ defines a $k$-linear map

$$b \mapsto \mathrm{Tr}_k(b|M) \colon B \to k$$

which we call the ***trace map*** of $M$.

PROPOSITION 8.1. *Let $B$ be a semisimple $k$-algebra. Two $B$-modules are isomorphic if and only if they have the same trace map.*

PROOF. Let $B_1, \ldots, B_n$ be the simple factors of $B$, and let $M_i$ be a simple $B_i$-module. Then every $B$-module is isomorphic to a direct sum $\bigoplus_j r_j M_j$ with $r_j M_j$ the direct sum of $r_j$ copies of $M_i$. We have to show that the trace map determines the multiplicities $r_j$. But for $e_i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots)$,

$$\mathrm{Tr}_k(e_i|\textstyle\sum r_j M_j) = r_i \dim_k M_i\,. \qquad \square$$

REMARK 8.2. The lemma fails when $k$ has characteristic $p$, because the trace map is identically zero on $pM$.

An ***involution*** of a $k$-algebra $B$ is a $k$-linear map $b \mapsto b^* \colon B \to B$ such that $(ab)^* = b^* a^*$ and $b^{**} = b$. Note that then $1^* = 1$ and so $c^* = c$ for $c \in k$.

PROPOSITION 8.3. *Let $k$ be an algebraically closed field, and let $(B, *)$ be a semisimple $k$-algebra with involution. Then $(B, *)$ is isomorphic to a product of pairs of the following types:*

- **(A):** $M_n(k) \times M_n(k)$, $\quad (a, b)^* = (b^t, a^t)$;
- **(C):** $M_n(k)$, $\quad b^* = b^t$;
- **(BD):** $M_n(k)$, $\quad b^* = J \cdot b^t \cdot J^{-1}$, $\quad J = \left(\begin{smallmatrix} 0 & -I \\ I & 0 \end{smallmatrix}\right)$.

PROOF. The decomposition $B = B_1 \times \cdots \times B_r$ of $B$ into a product of simple algebras $B_i$ is unique up to the ordering of the factors (Farb and Dennis 1993, 1.13). Therefore, $*$ permutes the set of $B_i$, and $B$ is a product of semisimple algebras with involution each of which is either (i) simple or (ii) the product of two simple algebras interchanged by $*$.

Let $(B, *)$ be as in (i). Then $B$ is isomorphic to $M_n(k)$ for some $n$, and the Noether-Skolem theorem (CFT, 2.10) shows that $b^* = u \cdot b^t \cdot u^{-1}$ for some $u \in M_n(k)$. Then $b = b^{**} = (u^t u^{-1})^{-1} b (u^t u^{-1})$ for all $b \in B$, and so $u^t u^{-1}$ lies in the centre $k$ of $M_n(k)$. Denote it by $c$, so that $u^t = cu$. Then $u = u^{tt} = c^2 u$, and so $c^2 = 1$. Therefore, $u^t = \pm u$, and $u$ is either symmetric or skew-symmetric. Relative to a suitable basis, $u$ is $I$ or $J$, and so $(B, *)$ is of type (C) or (BD).

Let $(B, *)$ be as in (ii). Then $*$ is an isomorphism of the opposite of the first factor onto the second. The Noether-Skolem theorem then shows that $(B, *)$ is isomorphic to $M_n(k) \times M_n(k)^{\mathrm{opp}}$ with the involution $(a, b) \mapsto (b, a)$. Now use that $a \leftrightarrow a^t \colon M_n(k)^{\mathrm{opp}} \cong M_n(k)$ to see that $(B, *)$ is of type (A). $\square$

The following is a restatement of the proposition.

PROPOSITION 8.4. *Let $(B, *)$ and $k$ be as in (8.3). If the only elements of the centre of $B$ invariant under $*$ are those in $k$, then $(B, *)$ is isomorphic to one of the following:*

- **(A):** $\mathrm{End}_k(W) \times \mathrm{End}_k(W^\vee)$, $\quad (a, b)^* = (b^t, a^t)$;
- **(C):** $\mathrm{End}_k(W)$, *$b^*$ the transpose of $b$ with respect to a symmetric bilinear form on $W$;*

**(BD):** $\mathrm{End}_k(W)$, *$b^*$ the transpose of $b$ with respect to an alternating bilinear form on $W$.*

**Symplectic modules and the associated algebraic groups.** Let $(B,*)$ be a semisimple $k$-algebra with involution $*$, and let $(V,\psi)$ be a ***symplectic*** $(B,*)$-***module***, i.e., a $B$-module $V$ endowed with an alternating $k$-bilinear form $\psi\colon V\times V\to k$ such that

$$\psi(bu,v)=\psi(u,b^*v) \text{ for all } b\in B, u,v\in V. \tag{35}$$

Let $F$ be the centre of $B$, and let $F_0$ be the subalgebra of invariants of $*$ in $F$. Assume that $B$ and $V$ are free over $F$ and that for all $k$-homomorphisms $\rho\colon F_0\to k^{\mathrm{al}}$, $(B\otimes_{F_0,\rho}k^{\mathrm{al}},*)$ is of the same type (A), (C), or (BD). This will be the case, for example, if $F$ is a field. Let $G$ be the subgroup of $\mathrm{GL}(V)$ such that

$$G(\mathbb{Q})=\{g\in\mathrm{Aut}_B(V)\mid \psi(gu,gv)=\mu(g)\psi(u,v) \text{ some } \mu(g)\in k^\times\},$$

and let

$$G'=\mathrm{Ker}(\mu)\cap\mathrm{Ker}(\det).$$

Example 8.5. (Type A.) Let $F$ be $k\times k$ or a field of degree 2 over $k$, and let $B=\mathrm{End}_F(W)$ equipped with the involution $*$ defined by a hermitian form[15] $\phi\colon W\times W\to F$. Then $(B,*)$ is of type A. Let $V_0$ be an $F$-vector space, and let $\psi_0$ be a skew-hermitian form $V_0\times V_0\to F$. The bilinear form $\psi$ on $V=W\otimes_F V_0$ defined by

$$\psi(w\otimes v,w'\otimes v')=\mathrm{Tr}_{F/k}(\phi(w,w')\psi_0(v,v')) \tag{36}$$

is alternating and satisfies (35): $(V,\psi)$ is a symplectic $(B,*)$-module. Let $C=\mathrm{End}_B(V)$ (the centralizer of $B$ in $\mathrm{End}_F(V)$). Then $C$ is stable under the involution $*$ defined by $\psi$, and

$$G(k)=\{c\in C^\times\mid cc^*\in k^\times\} \tag{37}$$

$$G'(k)=\{c\in C^\times\mid cc^*=1,\quad \det(c)=1\}. \tag{38}$$

In fact, $C\cong\mathrm{End}_F(V_0)$ and $*$ is transposition with respect to $\psi_0$. Therefore, $G$ is the group of symplectic similitudes of $\psi_0$ whose multiplier lies in $k$, and $G'$ is the special unitary group of $\psi_0$.

Conversely, let $(B,*)$ be of type A, and assume

(a) the centre $F$ of $B$ is of degree 2 over $k$ (so $F$ is a field or $k\times k$);
(b) $B$ is isomorphic to a matrix algebra over $F$ (when $F$ is a field, this just means that $B$ is simple and split over $F$).

Then I claim that $(B,*,V,\psi)$ arises as in the last paragraph. To see this, let $W$ be a simple $B$-module — condition (b) implies that $B\cong\mathrm{End}_F(W)$ and that $*$ is defined by a hermitian form $\phi\colon W\times W\to F$. As a $B$-module, $V$ is a direct sum of copies of $W$, and so $V=W\otimes_F V_0$ for some $F$-vector space $V_0$. Choose an element $f$ of $F\smallsetminus k$ whose square is in $k$. Then $f^*=-f$, and

$$\psi(v,v')=\mathrm{Tr}_{F/k}\left(f\Psi(v,v')\right)$$

for a unique hermitian form $\Psi\colon V\times V\to F$ (Deligne 1982, 4.6), which has the property that $\Psi(bv,v')=\Psi(v,b^*v')$. The form $(v,v')\mapsto f\Psi(v,v')$ is skew-hermitian,

[15]There is a unique involution of $F$ fixing $k$, which we again denote $*$. To say that $\phi$ is hermitian means that it is $F$-linear in one variable and satisfies $\phi(w,v)=\phi(v,w)^*$.

and can be[16] written $f\Psi = \phi \otimes \psi_0$ with $\psi_0$ skew-hermitian on $V_0$. Now $\psi, \phi, \psi_0$ are related by (36).

EXAMPLE 8.6. (Type C.) Let $B = \mathrm{End}_k(W)$ equipped with the involution $*$ defined by a symmetric bilinear form $\phi\colon W \times W \to k$. Let $V_0$ be a $k$-vector space, and let $\psi_0$ be an alternating form $V_0 \times V_0 \to k$. The bilinear form $\psi$ on $V = W \otimes V_0$ defined by

$$\psi(w \otimes v, w' \otimes v') = \phi(w, w')\psi_0(v, v')$$

is alternating and satisfies (35). Let $C = \mathrm{End}_B(V)$. Then $C$ is stable under the involution $*$ defined by $\psi$, and $G(k)$ and $G'(k)$ are described by the equations (37) and (38). In fact, $C \cong \mathrm{End}_k(V_0)$ and $*$ is transposition with respect to $\psi_0$. Therefore $G = \mathrm{GSp}(V_0, \psi_0)$ and $G' = \mathrm{Sp}(V_0, \psi_0)$. Every system $(B, *, V, \psi)$ with $B$ simple and split over $k$ arises in this way (cf. 8.5).

PROPOSITION 8.7. *For $(B, *)$ of type A or C, the group $G$ is reductive (in particular, connected), and $G'$ is semisimple and simply connected.*

PROOF. It suffices to prove this after extending the scalars to the algebraic closure of $k$. Then $(B, *, V, \psi)$ decomposes into quadruples of the types considered in Examples 8.5 and 8.6, and so the proposition follows from the calculations made there. □

REMARK 8.8. Assume $B$ is simple, and let $m$ be the ***reduced dimension*** of $V$,

$$m = \frac{\dim_F(V)}{[B\colon F]^{\frac{1}{2}}}.$$

In case (A), $G'_{\mathbb{Q}^{\mathrm{al}}} \approx (\mathrm{SL}_m)^{[F_0:\mathbb{Q}]}$ and in case (C), $G'_{\mathbb{Q}^{\mathrm{al}}} \approx (\mathrm{Sp}_m)^{[F_0:\mathbb{Q}]}$.

REMARK 8.9. In case (BD), the group $G$ is not connected ($G'$ is a special orthogonal group) although its identity component is reductive.

**Algebras with positive involution.** Let $C$ be a semisimple $\mathbb{R}$-algebra with an involution $*$, and let $V$ be a $C$-module. In the next proposition, by a ***hermitian form*** on $V$ we mean a symmetric bilinear form $\psi\colon V \times V \to \mathbb{R}$ satisfying (35). Such a form is said to be ***positive definite*** if $\psi(v, v) > 0$ for all nonzero $v \in V$.

PROPOSITION 8.10. *Let $C$ be a semisimple algebra over $\mathbb{R}$. The following conditions on an involution $*$ of $C$ are equivalent:*

(a) *some faithful $C$-module admits a positive definite hermitian form;*
(b) *every $C$-module admits a positive definite hermitian form;*
(c) $\mathrm{Tr}_{C/\mathbb{R}}(c^*c) > 0$ *for all nonzero $c \in C$.*

PROOF. (a) $\Longrightarrow$ (b). Let $V$ be a faithful $C$-module. Then every $C$-module is a direct summand of a direct sum of copies of $V$ (see p323). Hence, if $V$ carries a positive definite hermitian form, then so does every $C$-module.

(b) $\Longrightarrow$ (c). Let $V$ be a $C$-module with a positive definite hermitian form $(\,|\,)$, and choose an orthonormal basis $e_1, \ldots, e_n$ for $V$. Then

$$\mathrm{Tr}_{\mathbb{R}}(c^*c|V) = \textstyle\sum_i (e_i|c^*ce_i) = \sum_i (ce_i|ce_i),$$

[16]Probably the easiest way to prove things like this is use the correspondence between involutions on algebras and (skew-)hermitian forms (up to scalars) — see Knus et al. 1998, I 4.2. The involution on $\mathrm{End}_F(V)$ defined by $\psi$ stabilizes $C$ and corresponds to a skew-hermitian form on $V_0$.

which is $>0$ unless $c$ acts as the zero map on $V$. On applying this remark with $V=C$, we obtain (c).

(c) $\Longrightarrow$ (a). The condition (c) is that the hermitian form $(c,c')\mapsto \mathrm{Tr}_{C/\mathbb{R}}(c^*c')$ on $C$ is positive definite. □

DEFINITION 8.11. An involution satisfying the equivalent conditions of (8.10) is said to be ***positive***.

PROPOSITION 8.12. *Let $B$ be a semisimple $\mathbb{R}$-algebra with a positive involution $*$ of type A or C. Let $(V,\psi)$ be a symplectic $(B,*)$-module, and let $C$ be the centralizer of $B$ in $\mathrm{End}_{\mathbb{R}}(V)$. Then there exists a homomorphism of $\mathbb{R}$-algebras $h\colon \mathbb{C}\to C$, unique up to conjugation by an element $c$ of $C^\times$ with $cc^*=1$, such that*

- *$h(\overline{z})=h(z)^*$ and*
- *$u,v\mapsto \psi(u,h(i)v)$ is positive definite and symmetric.*

PROOF. To give an $h$ satisfying the conditions amounts to giving an element $J$ $(=h(i))$ of $C$ such that

$$J^2=-1,\quad \psi(Ju,Jv)=\psi(u,v),\quad \psi(v,Jv)>0 \text{ if } v\neq 0. \tag{39}$$

Suppose first that $(B,*)$ is of type A. Then $(B,*,V,\psi)$ decomposes into systems arising as in (8.5). Thus, we may suppose $B=\mathrm{End}_F(W)$, $V=W\otimes V_0$, etc., as in (8.5). We then have to classify the $J\in C\cong \mathrm{End}_{\mathbb{C}}(V_0)$ satisfying (39) with $\psi$ replaced by $\psi_0$. There exists a basis $(e_j)$ for $V_0$ such that

$$(\psi_0(e_j,e_k))_{j,k}=\mathrm{diag}(i,\ldots,\underset{r}{i},-i,\ldots,-i),\quad i=\sqrt{-1}.$$

Define $J$ by $J(e_j)=-\psi_0(e_j,e_j)e_j$. Then $J$ satisfies the required conditions, and it is uniquely determined up to conjugation by an element of the unitary group of $\psi_0$. This proves the result for type A, and type C is similar. (For more details, see Zink 1983, 3.1). □

REMARK 8.13. Let $(B,*)$ and $(V,\psi)$ be as in the proposition. For an $h$ satisfying the conditions of the proposition, define

$$t(b)=\mathrm{Tr}_{\mathbb{C}}(b|V/F_h^0V),\quad b\in B.$$

Then, $t$ is independent of the choice of $h$, and in fact depends only on the isomorphism class of $(V,\psi)$ as a $B$-module. Conversely, $(V,\psi)$ is determined up to $B$-isomorphism by its dimension and $t$. For example, if $V=W\otimes_{\mathbb{C}}V_0$, $\phi$, $\psi_0$, etc. are as in the above proof, then

$$\mathrm{Tr}_k(b|V)=r\cdot \mathrm{Tr}_k(b|W),$$

and $r$ and $\dim V_0$ determine $(V_0,\psi_0)$ up to isomorphism. Since $W$ and $\phi$ are determined (up to isomorphism) by the requirement that $W$ be a simple $B$-module and $\phi$ be a hermitian form giving $*$ on $B$, this proves the claim for type A.

**PEL data.** Let $B$ be a simple $\mathbb{Q}$-algebra with a positive involution $*$ (meaning that it becomes positive on $B\otimes_{\mathbb{Q}}\mathbb{R}$), and let $(V,\psi)$ be a symplectic $(B,*)$-module. Throughout this subsection, we assume that $(B,*)$ is of type A or C.

PROPOSITION 8.14. *There is a unique $G(\mathbb{R})$-conjugacy class $X$ of homomorphisms $h\colon \mathbb{S}\to G_{\mathbb{R}}$ such that each $h\in X$ defines a complex structure on $V(\mathbb{R})$ that is positive or negative for $\psi$. The pair $(G,X)$ satisfies the conditions SV1–4.*

PROOF. The first statement is an immediate consequence of (8.12). The composite of $h$ with $G \hookrightarrow G(\psi)$ lies in $X(\psi)$, and therefore satisfies SV1, SV2, SV4. As $h$ is nontrivial, SV3 follows from the fact that $G^{\mathrm{ad}}$ is simple. $\square$

DEFINITION 8.15. The Shimura data arising in this way are called ***simple PEL data of type A or C***.

The simple refers to the fact that (for simplicity), we required $B$ to be simple (which implies that $G^{\mathrm{ad}}$ is simple).

REMARK 8.16. Let $b \in B$, and let $t_b$ be the tensor $(x,y) \mapsto \psi(x,by)$ of $V$. An element $g$ of $G(\psi)$ fixes $t_b$ if and only if it commutes with $b$. Let $b_1,\ldots,b_s$ be a set of generators for $B$ as a $\mathbb{Q}$-algebra. Then $(G,X)$ is the Shimura datum of hodge type associated with the system $(V,\{\psi,t_{b_1},\ldots,t_{b_s}\})$.

**PEL Shimura varieties.**

THEOREM 8.17. *Let $(G,X)$ be a simple PEL datum of type A or C associated with $(B,*,V,\psi)$ as in the last subsection, and let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$. Then $\mathrm{Sh}_K(G,X)(\mathbb{C})$ classifies the isomorphism classes of quadruples $(A,s,i,\eta K)$ in which*

- ◦ *$A$ is a complex abelian variety,*
- ◦ *$\pm s$ is a polarization of the hodge structure $H_1(A,\mathbb{Q})$,*
- ◦ *$i$ is a homomorphism $B \to \mathrm{End}(A)\otimes\mathbb{Q}$, and*
- ◦ *$\eta K$ is a $K$-orbit of $B\otimes\mathbb{A}_f$-linear isomorphisms $\eta\colon V(\mathbb{A}_f) \to H^1(A,\mathbb{Q})\otimes\mathbb{A}_f$ sending $\psi$ to an $\mathbb{A}_f^\times$-multiple of $s$,*

*satisfying the following condition:*

*(**) there exists a $B$-linear isomorphism $a\colon H_1(A,\mathbb{Q}) \to V$ sending $s$ to a $\mathbb{Q}^\times$-multiple of $\psi$.*

PROOF. In view of the dictionary $b \leftrightarrow t_b$ between endomorphisms and tensors (8.16), Theorem 7.4 shows that $\mathrm{Sh}_K(G,X)(\mathbb{C})$ classifies the quadruples $(A,i,t,\eta K)$ with the additional condition that $ah \in X$, but $ah$ defines a complex structure on $V(\mathbb{R})$ that is positive or negative for $\psi$, and so (8.14) shows that $ah$ automatically lies in $X$. $\square$

Let $(G,X)$ be the Shimura datum arising from $(B,*)$ and $(V,\psi)$. For $h \in X$, we have a trace map
$$b \mapsto \mathrm{Tr}(b|V(\mathbb{C})/F_h^0)\colon B \to \mathbb{C}.$$
Since this map is independent of the choice of $h$ in $X$, we denote it by $\mathrm{Tr}_X$.

REMARK 8.18. Consider a triple $(A,s,i,\eta K)$ as in the theorem. The existence of the isomorphism $a$ in (**) implies that

(a) $s(bu,v) = s(u,b^*v)$, and
(b) $\mathrm{Tr}(i(b)|\mathrm{Tgt}_0A) = \mathrm{Tr}_X(b)$ for all $b \in B\otimes\mathbb{C}$.

The first is obvious, because $\psi$ has this property, and the second follows from the $B$-isomorphisms
$$\mathrm{Tgt}_0(A) \overset{(34)}{\cong} H_1(A,\mathbb{C})/F^0 \xrightarrow{a} V(\mathbb{C})/F_h^0.$$

We now divide the type A in two, depending on whether the reduced dimension of $V$ is even or odd.

Proposition 8.19. *For types Aeven and C, the condition (**) of Theorem 8.17 is implied by conditions (a) and (b) of (8.18).*

Proof. Let $W = H_1(A,\mathbb{Q})$. We have to show that there exists a $B$-linear isomorphism $\alpha\colon W \to V$ sending $s$ to a $\mathbb{Q}^\times$-multiple of $\psi$. The existence of $\eta$ shows that $W$ has the same dimension as $V$, and so there exists a $B\otimes_{\mathbb{Q}}\mathbb{Q}^{\mathrm{al}}$-isomorphism $\alpha\colon V(\mathbb{Q}^{\mathrm{al}}) \to W(\mathbb{Q}^{\mathrm{al}})$ sending $t$ to a $\mathbb{Q}^{\mathrm{al}\times}$-multiple of $\psi$. For $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ write $\sigma\alpha = \alpha\circ a_\sigma$ with $a_\sigma \in G(\mathbb{Q}^{\mathrm{al}})$. Then $\sigma \mapsto a_\sigma$ is a one-cocycle. If its class in $H^1(\mathbb{Q},G)$ is trivial, say, $a_\sigma = a^{-1}\cdot\sigma a$, then $\alpha\circ a^{-1}$ is fixed by all $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$, and is therefore defined over $\mathbb{Q}$.

Thus, it remains to show that the class of $(a_\sigma)$ in $H^1(\mathbb{Q},G)$ is trivial. The existence of $\eta$ shows that the image of the class in $H^1(\mathbb{Q}_\ell,G)$ is trivial for all finite primes $\ell$, and (8.13) shows that its image in $H^1(\mathbb{R},G)$ is trivial, and so the statement follows from the next two lemmas. □

Lemma 8.20. *Let $G$ be a reductive group with simply connected derived group, and let $T = G/G^{\mathrm{der}}$. If $H^1(\mathbb{Q},T)\to\prod_{l\le\infty}H^1(\mathbb{Q}_l,T)$ is injective, then an element of $H^1(\mathbb{Q},G)$ that becomes trivial in $H^1(\mathbb{Q}_l,G)$ for all $l$ is itself trivial.*

Proof. Because $G^{\mathrm{der}}$ is simply connected, $H^1(\mathbb{Q}_l,G^{\mathrm{der}}) = 0$ for $l\neq\infty$ and $H^1(\mathbb{Q},G^{\mathrm{der}})\to H^1(\mathbb{R},G^{\mathrm{der}})$ is injective (5.19). Using this, we obtain a commutative diagram with exact rows

$$\begin{array}{ccccccccc} & & T(\mathbb{Q}) & \longrightarrow & H^1(\mathbb{Q},G^{\mathrm{der}}) & \longrightarrow & H^1(\mathbb{Q},G) & \longrightarrow & H^1(\mathbb{Q},T)\\ & & \downarrow & & \downarrow{\scriptstyle\text{injective}} & & \downarrow & & \downarrow{\scriptstyle\text{injective}}\\ G(\mathbb{R}) & \longrightarrow & T(\mathbb{R}) & \longrightarrow & H^1(\mathbb{R},G^{\mathrm{der}}) & \to & \prod_l H^1(\mathbb{Q}_l,G) & \longrightarrow & \prod_l H^1(\mathbb{Q}_l,T).\end{array}$$

If an element $c$ of $H^1(\mathbb{Q},G)$ becomes trivial in all $H^1(\mathbb{Q}_l,G)$, then a diagram chase shows that it arises from an element $c'$ of $H^1(\mathbb{Q},G^{\mathrm{der}})$ whose image $c'_\infty$ in $H^1(\mathbb{R},G^{\mathrm{der}})$ maps to the trivial element in $H^1(\mathbb{R},G)$. The image of $G(\mathbb{R})$ in $T(\mathbb{R})$ contains $T(\mathbb{R})^+$ (see 5.1), and the real approximation theorem (5.4) shows that $T(\mathbb{Q})\cdot T(\mathbb{R})^+ = T(\mathbb{R})$. Therefore, there exists a $t\in T(\mathbb{Q})$ whose image in $H^1(\mathbb{R},G^{\mathrm{der}})$ is $c'_\infty$. Then $t\mapsto c'$ in $H^1(\mathbb{Q},G^{\mathrm{der}})$, which shows that $c$ is trivial. □

Lemma 8.21. *Let $(G,X)$ be a simple PEL Shimura datum of type Aeven or C, and let $T = G/G^{\mathrm{der}}$. Then $H^1(\mathbb{Q},T)\to\prod_{l\le\infty}H^1(\mathbb{Q}_l,T)$ is injective.*

Proof. For $G$ of type Aeven, $T = \mathrm{Ker}((\mathbb{G}_m)_F \xrightarrow{\mathrm{Nm}_{F/k}} (\mathbb{G}_m)_{F_0})\times\mathbb{G}_m$. The group $H^1(\mathbb{Q},\mathbb{G}_m) = 0$, and the map on $H^1$'s of the first factor is

$$F_0^\times/\operatorname{Nm}F^\times \to \textstyle\prod_v F_{0v}^\times/\operatorname{Nm}F_v^\times.$$

This is injective (CFT, VIII 1.4).

For $G$ of type $C$, $T = \mathbb{G}_m$, and so $H^1(\mathbb{Q},T) = 0$. □

**PEL modular varieties.** Let $B$ be a semisimple algebra over $\mathbb{Q}$ with a positive involution $*$, and let $(V,\psi)$ be a symplectic $(B,*)$-module. Let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$. There exists an algebraic variety $M_K$ over $\mathbb{C}$ classifying the isomorphism classes of quadruples $(A,s,i,\eta K)$ satisfying (a) and (b) of (8.18) (but not necessarily condition (**)), which is called the ***PEL modular variety attached to*** $(B,*,V,\psi)$. In the simple cases (Aeven) and (C), Proposition 8.17

shows that $M_K$ coincides with $\mathrm{Sh}_K(G, X)$, but in general it is a finite disjoint union of Shimura varieties.

NOTES. The theory of Shimura varieties of PEL-type is worked out in detail in several papers of Shimura, for example, Shimura 1963, but in a language somewhat different from ours. The above account follows Deligne 1971*c*, §§5,6. See also Zink 1983 and Kottwitz 1992, §§1–4.

## 9. General Shimura varieties

**Abelian motives.** Let $\mathsf{Hod}(\mathbb{Q})$ be the category of polarizable rational hodge structures. It is an abelian subcategory of the category of all rational hodge structures closed under the formation of tensor products and duals.

Let $V$ be a variety over $\mathbb{C}$ whose connected components are abelian varieties, say $V = \bigsqcup V_i$ with $V_i$ an abelian variety. Recall that for manifolds $M_1$ and $M_2$,

$$H^r(M_1 \sqcup M_2, \mathbb{Q}) \cong H^r(M_1, \mathbb{Q}) \oplus H^r(M_2, \mathbb{Q}).$$

For each connected component $V^\circ$ of $V$,

$$H^*(V^\circ, \mathbb{Q}) \cong \bigwedge H^1(V^\circ, \mathbb{Q}) \cong \mathrm{Hom}_{\mathbb{Q}}(\bigwedge H_1(V^\circ, \mathbb{Q}), \mathbb{Q})$$

(see 6.4). Therefore, $H^*(V, \mathbb{Q})$ acquires a polarizable hodge structure from that on $H_1(V, \mathbb{Q})$. We write $H^*(V, \mathbb{Q})(m)$ for the hodge structure $H^*(V, \mathbb{Q}) \otimes \mathbb{Q}(m)$ (see 2.6).

Let $(W, h)$ be a rational hodge structure. An endomorphism $e$ of $(W, h)$ is an ***idempotent*** if $e^2 = e$. Then

$$(W, h) = \mathrm{Im}(e) \oplus \mathrm{Im}(1 - e)$$

(direct sum of rational hodge structures).

An ***abelian motive*** over $\mathbb{C}$ is a triple $(V, e, m)$ in which $V$ is a variety over $\mathbb{C}$ whose connected components are abelian varieties, $e$ is an idempotent in $\mathrm{End}(H^*(V, \mathbb{Q}))$, and $m \in \mathbb{Z}$. For example, let $A$ be an abelian variety; then the projection

$$H^*(A, \mathbb{Q}) \to H^i(A, \mathbb{Q}) \subset H^*(A, \mathbb{Q})$$

is an idempotent $e^i$, and we denote $(A, e^i, 0)$ by $h^i(A)$.

Define $\mathrm{Hom}((V, e, m), (V', e', m'))$ to be the set of maps $H^*(V, \mathbb{Q}) \to H^*(V', \mathbb{Q})$ of the form $e' \circ f \circ e$ with $f$ a homomorphism $H^*(V, \mathbb{Q}) \to H^*(V', \mathbb{Q})$ of degree $d = m' - m$. Moreover, define

$$\begin{aligned}
(V, e, m) \oplus (V', e', m) &= (V \sqcup V', e \oplus e', m) \\
(V, e, m) \otimes (V', e', m) &= (V \times V', e \otimes e', m + m') \\
(V, e, m)^\vee &= (V, e^t, d - m) \text{ if } V \text{ is purely } d\text{-dimensional.}
\end{aligned}$$

For an abelian motive $(V, e, m)$ over $\mathbb{C}$, let $H(V, e, m) = eH^*(V, \mathbb{Q})(m)$. Then $(V, e, m) \mapsto H(V, e, m)$ is a functor from the category of abelian motives $\mathsf{AM}$ to $\mathsf{Hod}(\mathbb{Q})$ commuting with $\oplus$, $\otimes$, and $^\vee$. We say that a rational hodge structure is ***abelian*** if it is in the essential image of this functor, i.e., if it is isomorphic to $H(V, e, m)$ for some abelian motive $(V, e, m)$. Every abelian hodge structure is polarizable.

PROPOSITION 9.1. *Let* $\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ *be the full subcategory of* $\mathsf{Hod}(\mathbb{Q})$ *of abelian hodge structures. Then* $\mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ *is the smallest strictly full subcategory of* $\mathsf{Hod}(\mathbb{Q})$ *containing* $H_1(A,\mathbb{Q})$ *for each abelian variety* $A$ *and closed under the formation of direct sums, subquotients, duals, and tensor products; moreover,* $H\colon \mathsf{AM} \to \mathsf{Hod}^{\mathrm{ab}}(\mathbb{Q})$ *is an equivalence of categories.*

PROOF. Straightforward from the definitions. □

For a description of the essential image of $H$, see Milne 1994, 1.27.

**Shimura varieties of abelian type.** Recall (§6) that a symplectic space $(V,\psi)$ over $\mathbb{Q}$ defines a connected Shimura datum $(S(\psi), X(\psi)^+)$ with $S(\psi) = \mathrm{Sp}(\psi)$ and $X(\psi)^+$ the set of complex structures on $V(\mathbb{R}),\psi)$.

DEFINITION 9.2. (a) A connected Shimura datum $(H, X^+)$ with $H$ simple is of ***primitive abelian type*** if there exists a symplectic space $(V,\psi)$ and an injective homomorphism $H \to S(\psi)$ carrying $X^+$ into $X(\psi)^+$.

(b) A connected Shimura datum $(H, X^+)$ is of ***abelian type*** if there exist pairs $(H_i, X_i^+)$ of primitive abelian type and an isogeny $\prod_i H_i \to H$ carrying $\prod_i X_i^+$ into $X$.

(b) A Shimura datum $(G,X)$ is of ***abelian type*** if $(G^{\mathrm{der}}, X^+)$ is of abelian type.

(c) The (connected) Shimura variety attached to a (connected) Shimura datum of abelian type is said to be of ***abelian type***.

PROPOSITION 9.3. *Let* $(G,X)$ *be a Shimura datum, and assume*

(a) *the weight* $w_X$ *is rational SV4 and* $Z(G)^\circ$ *splits over a CM-field SV6, and*
(b) *there exists a homomorphism* $\nu\colon G \to \mathbb{G}_m$ *such that* $\nu \circ w_X = -2$.

*If* $G$ *is of abelian type, then* $(V, h\circ\rho)$ *is an abelian hodge structure for all representations* $(V,\rho)$ *of* $G$ *and all* $h \in X$*; conversely, if there exists a faithful representation* $\rho$ *of* $G$ *such that* $(V, h\circ\rho)$ *is an abelian hodge structure for all* $h$*, then* $(G,X)$ *is of abelian type.*

PROOF. See Milne 1994, 3.12. □

Let $(G,X)$ be a Shimura datum of abelian type satisfying (a) and (b) of the proposition, and let $\rho\colon G \to \mathrm{GL}(V)$ be a faithful representation of $G$. Assume that there exists a pairing $\psi\colon V \times V \to \mathbb{Q}$ such that

(a) $g\psi = \nu(g)^m\psi$ for all $g \in G$,
(b) $\psi$ is a polarization of $(V, h\circ\rho)$ for all $h \in X$.

There exist multilinear maps $t_i\colon V \times \cdots \times V \to \mathbb{Q}(r_i)$, $1 \le i \le n$, such that $G$ is the subgroup of $\mathrm{GL}(V)$ whose elements satisfy (a) and fix $t_1, \dots t_n$ (cf. 7.2).

THEOREM 9.4. *With the above notations,* $\mathrm{Sh}(G,X)$ *classifies the isomorphism classes of triples* $(A, (s_i)_{0\le i\le n}, \eta K)$ *in which*

- $A$ *is an abelian motive,*
- $\pm s_0$ *is a polarization for the rational hodge structure* $H(A)$,
- $s_1, \dots, s_n$ *are tensors for* $A$*, and*
- $\eta K$ *is a* $K$*-orbit of* $\mathbb{A}_f$*-linear isomorphisms* $V(\mathbb{A}_f) \to V_f(A)$ *sending* $\psi$ *to an* $\mathbb{A}_f^\times$*-multiple of* $s_0$ *and each* $t_i$ *to* $s_i$,

*satisfying the following condition:*

*(**) there exists an isomorphism* $a\colon H(A) \to V$ *sending* $s_0$ *to a* $\mathbb{Q}^{\times}$*-multiple of* $\psi$*, each* $s_i$ *to* $t_i$*, and* $h$ *onto an element of* $X$*.*

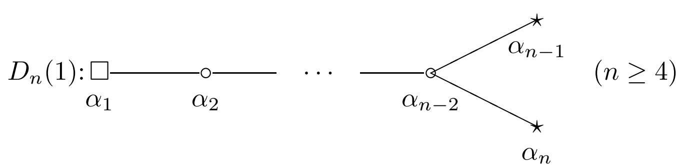
PROOF. With $A$ replaced by a hodge structure, this can be proved by an elementary argument (cf. 6.3, 7.3), but (9.3) shows that the hodge structures arising are abelian, and so can be replaced by abelian motives (9.1). For more details, see Milne 1994, Theorem 3.31. □

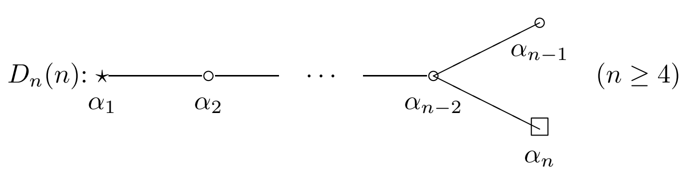
**Classification of Shimura varieties of abelian type.** Deligne (1979) classifies the connected Shimura data of abelian type. Let $(G, X^+)$ be a connected Shimura datum with $G$ simple. If $G^{\mathrm{ad}}$ is of type A, B, or C, then $(G, X^+)$ is of abelian type. If $G^{\mathrm{ad}}$ is of type $\mathrm{E}_6$ or $\mathrm{E}_7$, then $(G, X^+)$ is not of abelian type. If $G^{\mathrm{ad}}$ is of type $D$, $(G, X^+)$ may or may not be of abelian type. There are two problems that may arise.

(a) Let $G$ be the universal covering group of $G^{\mathrm{ad}}$. There may exist homomorphisms $(G, X^+) \to (S(\psi), X(\psi)^+)$ but no injective such homomorphism, i.e., there may be a nonzero finite algebraic subgroup $N \subset G$ that is in the kernel of all homomorphisms $G \to S(\psi)$ sending $X^+$ into $X(\psi)^+$. Then $(G/N', X^+)$ is of abelian type for all $N' \supset N$, but $(G, X^+)$ is not of abelian type.

(b) There may not exist a homomorphism $G \to S(\psi)$ at all.

This last problem arises for the following reason. Even when $G^{\mathrm{ad}}$ is $\mathbb{Q}$-simple, it may decompose into a product of simple group $G_{\mathbb{R}}^{\mathrm{ad}} = G_1 \times \cdots \times G_r$ over $\mathbb{R}$. For each $i$, $G_i$ has a dynkin diagram of the shape shown below:




$D_n(n-1)$: Same as $D_n(n)$ by with $\alpha_{n-1}$ and $\alpha_n$ interchanged (rotation about the horizontal axis).

Nodes marked by squares are special (p278), and nodes marked by stars correspond to symplectic representations. The number in parenthesis indicates the position of the special node. As is explained in §1, the projection of $X^+$ to a conjugacy class of homomorphisms $\mathbb{S} \to G_i$ corresponds to a node marked with a □. Since $X^+$ is defined over $\mathbb{R}$, the nodes can be chosen independently for each $i$. On the other hand, the representations $G_{i\mathbb{R}} \to S(\psi)_{\mathbb{R}}$ correspond to nodes marked with a $*$. Note that the $*$ has to be at the opposite end of the diagram from the □. In order for a family of representations $G_{i\mathbb{R}} \to S(\psi)_{\mathbb{R}}$, $1 \leq i \leq r$, to arise from a symplectic representation over $\mathbb{Q}$, the $*$'s must be all in the same position since a galois group must permute the dynkin diagrams of the $G_i$. Clearly, this is impossible if the □'s occur at different ends. (See Deligne 1979, 2.3, for more details.)

**Shimura varieties not of abelian type.** It is hoped (Deligne 1979, p248) that all Shimura varieties with rational weight classify isomorphism classes of motives with additional structure, but this is not known for those not of abelian type. More precisely, from the choice of a rational representation $\rho\colon G \to \mathrm{GL}(V)$, we obtain a family of hodge structures $h \circ \rho_{\mathbb{R}}$ on $V$ indexed by $X$. When the weight of $(G, X)$ is defined over $\mathbb{Q}$, it is hoped that these hodge structures always occur (in a natural way) in the cohomology of algebraic varieties. When the weight of $(G, X)$ is not defined over $\mathbb{Q}$ they obviously can not.

**Example: simple Shimura varieties of type $A_1$.** Let $(G, X)$ be the Shimura datum attached to a $B$ be a quaternion algebra over a totally real field $F$, as in (5.24). With the notations of that example,

$$G(\mathbb{R}) \approx \prod_{v \in I_c} \mathbb{H}^{\times} \times \prod_{v \in I_{nc}} \mathrm{GL}_2(\mathbb{R}).$$

(a) If $B = M_2(F)$, then $(G, X)$ is of PEL-type, and $\mathrm{Sh}_K(G, X)$ classifies isomorphism classes of quadruples $(A, i, t, \eta K)$ in which $A$ is an abelian variety of dimension $d = [F\colon \mathbb{Q}]$ and $i$ is a homomorphism homomorphism $i\colon F \to \mathrm{End}(A) \otimes \mathbb{Q}$. These Shimura varieties are called ***Hilbert*** (or ***Hilbert-Blumenthal***) varieties, and whole books have been written about them.

(b) If $B$ is a division algebra, but $I_c = \emptyset$, then $(G, X)$ is again of PEL-type, and $\mathrm{Sh}_K(G, X)$ classifies isomorphism classes of quadruples $(A, i, t, \eta K)$ in which $A$ is an abelian variety of dimension $2[F\colon \mathbb{Q}]$ and $i$ is a homomorphism $i\colon B \to \mathrm{End}(A) \otimes \mathbb{Q}$. In this case, the varieties are projective. These varieties have also been extensively studied.

(c) If $B$ is a division algebra and $I_c \neq \emptyset$, then $(G, X)$ is of abelian type, but the weight is not defined over $\mathbb{Q}$. Over $\mathbb{R}$, the weight map $w_X$ sends $a \in \mathbb{R}$ to the element of $(F \otimes \mathbb{R})^{\times} \cong \prod_{v\colon F \to \mathbb{R}} \mathbb{R}$ with component 1 for $v \in I_c$ and component $a$ for $v \in I_{nc}$. Let $T$ be the torus over $\mathbb{Q}$ with $T(\mathbb{Q}) = F^{\times}$. Then $w_X : \mathbb{G}_m \to T_{\mathbb{R}}$ is defined over the subfield $L$ of $\overline{\mathbb{Q}}$ whose fixed group is the subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ stabilizing $I_c \subset I_c \sqcup I_{nc}$. On choosing a rational representation of $G$, we find that $\mathrm{Sh}_K(G, X)$ classifies certain isomorphism classes of hodge structures with additional structure, but the hodge structures are not motivic — they do not arise in the cohomology of algebraic varieties (they are not rational hodge structures).

## 10. Complex multiplication: the Shimura-Taniyama formula

**Where we are headed.** Let $V$ be a variety over $\mathbb{Q}$. For any $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ and $P \in V(\mathbb{Q}^{\mathrm{al}})$, the point $\sigma P \in V(\mathbb{Q}^{\mathrm{al}})$. For example, if $V$ is the subvariety of $\mathbb{A}^n$ defined by equations

$$f(X_1, \ldots, X_n) = 0, \quad f \in \mathbb{Q}[X_1, \ldots, X_n],$$

then

$$f(a_1, \ldots, a_n) = 0 \Longrightarrow f(\sigma a_1, \ldots, \sigma a_n) = 0$$

(apply $\sigma$ to the first equality). Therefore, if we have a variety $V$ over $\mathbb{Q}^{\mathrm{al}}$ that we suspect is actually defined over $\mathbb{Q}$, then we should be able to describe an action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ on its points $V(\mathbb{Q}^{\mathrm{al}})$.

Let $E$ be a number field contained in $\mathbb{C}$, and let $\mathrm{Aut}(\mathbb{C}/E)$ denote the group of automorphisms of $\mathbb{C}$ (as an abstract field) fixing the elements of $E$. Then a similar remark applies: if a variety $V$ over $\mathbb{C}$ is defined by equations with coefficients in $E$, then $\mathrm{Aut}(\mathbb{C}/E)$ will act on $V(\mathbb{C})$. Now, I claim that all Shimura varieties are defined (in a natural way) over specific number fields, and so I should be able to describe an action of a big subgroup of $\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ on their points. If, for example, the Shimura variety is of hodge type, then there is a set $\mathcal{M}_K$ whose elements are abelian varieties plus additional data and a map

$$(A,\ldots)\mapsto P(A,\ldots)\colon \mathcal{M}_K\to \mathrm{Sh}_K(G,X)(\mathbb{C})$$

whose fibres are the isomorphism classes in $\mathcal{M}_K$. On applying $\sigma\in\mathrm{Aut}(\mathbb{C}/\mathbb{Q})$ to the coefficients of the polynomials defining $A,\ldots$, we get a new triple $(\sigma A,\ldots)$ which may or may not lie in $\mathcal{M}_K$. When it does we define ${}^{\sigma}P(A,\ldots)$ to be $P(\sigma A,\ldots)$. Our task will be to show that, for some specific field $E$, this does give an action of $\mathrm{Aut}(\mathbb{C}/E)$ on $\mathrm{Sh}_K(G,X)$ and that the action does arise from a model of $\mathrm{Sh}_K(G,X)$ over $E$.

For example, for $P\in\Gamma(1)\backslash\mathcal{H}_1$, ${}^{\sigma}P$ is the point such that $j({}^{\sigma}P)=\sigma\left(j(P)\right)$. If $j$ were a polynomial with coefficients in $\mathbb{Z}$ (rather than a power series with coefficients in $\mathbb{Z}$), we would have $j(\sigma P)=\sigma j(P)$ with the obvious meaning of $\sigma P$, but this is definitely false (if $\sigma$ is not complex conjugation, then it is not continuous, nor even measurable).

You may complain that we fail to explicitly describe the action of $\mathrm{Aut}(\mathbb{C}/E)$ on $\mathrm{Sh}(G,X)(\mathbb{C})$, but I contend that there can not exist a completely explicit description of the action. What are the elements of $\mathrm{Aut}(\mathbb{C}/E)$? To construct them, we can choose a transcendence basis $B$ for $\mathbb{C}$ over $E$, choose a permutation of the elements of $B$, and extend the resulting automorphism of $\mathbb{Q}(B)$ to its algebraic closure $\mathbb{C}$. But proving the existence of transcendence bases requires the axiom of choice (e.g., FT, 8.13), and so we can have no explicit description of, or way of naming, the elements of $\mathrm{Aut}(\mathbb{C}/E)$, and hence no completely explicit description of the action is possible.

However, all is not lost. Abelian class field theory names the elements of $\mathrm{Gal}(E^{\mathrm{ab}}/E)$, where $E^{\mathrm{ab}}$ is a maximal abelian extension of $E$. Thus, if we suspect that a point $P$ has coordinates in $E^{\mathrm{ab}}$, the action of $\mathrm{Aut}(\mathbb{C}/E)$ on it will factor through $\mathrm{Gal}(E^{\mathrm{ab}}/E)$, and we may hope to be able to describe the action of $\mathrm{Aut}(\mathbb{C}/E)$ explicitly. This the theory of complex multiplication allows us to do for certain special points $P$.

**Review of abelian varieties.** The theory of abelian varieties is very similar to that of elliptic curves — just replace $E$ with $A$, 1 with $g$ (the dimension of $A$), and, whenever $E$ occurs twice, replace one copy with the dual $A^{\vee}$ of $A$.

Thus, for any $m$ not divisible by the characteristic of the ground field $k$,

$$A(k^{\mathrm{al}})_m\approx(\mathbb{Z}/m\mathbb{Z})^{2g}. \tag{40}$$

Here $A(k^{\mathrm{al}})_m$ consists of the elements of $A(k^{\mathrm{al}})$ killed by $m$. Hence, for $\ell\neq\mathrm{char}(k)$,

$$T_\ell A\overset{\mathrm{df}}{=}\varprojlim A(k^{\mathrm{al}})_{\ell^n}$$

is a free $\mathbb{Z}_\ell$-module of rank $2g$, and

$$V_\ell(A)\overset{\mathrm{df}}{=}T_\ell A\otimes_{\mathbb{Z}_\ell}\mathbb{Q}_\ell$$

is a $\mathbb{Q}_\ell$-vector space of dimension $2g$. In characteristic zero, we set

$$T_fA=\prod T_\ell A=\varprojlim_m A(k^{\mathrm{al}})_m,$$
$$V_fA=T_f\otimes_{\mathbb{Z}}\mathbb{Q}=\prod(V_\ell A\colon T_\ell A)\ \text{(restricted topological product)}.$$

They are, respectively, a free $\hat{\mathbb{Z}}$-module of rank $2g$ and a free $\mathbb{A}_f$-module of rank $2g$. The galois group $\mathrm{Gal}(k^{\mathrm{al}}/k)$ acts continuously on these modules.

For an endomorphism $a$ of an abelian variety $A$, there is a unique monic polynomial $P_a(T)$ with integer coefficients (the ***characteristic polynomial of*** $a$) such that $|P_a(n)| = \deg(a-n)$ for all $n \in \mathbb{Z}$. Moreover, $P_a$ is the characteristic polynomial of $a$ acting on $V_\ell A$ ($\ell \neq \mathrm{char}(k)$).

For an abelian variety $A$ over a field $k$, the tangent space $\mathrm{Tgt}_0(A)$ to $A$ at 0 is a vector space over $k$ of dimension $g$. As we noted in §6, when $k = \mathbb{C}$, the exponential map defines a surjective homomorphism $\mathrm{Tgt}_0(A) \to A(\mathbb{C})$ whose kernel is a lattice $\Lambda$ in $\mathrm{Tgt}_0(A)$. Thus $A(\mathbb{C})_m \cong \frac{1}{m}\Lambda/\Lambda \cong \Lambda/m\Lambda$, and

$$T_\ell A \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Z}_\ell, \quad V_\ell A \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{Q}_\ell, \quad T_f A = \Lambda \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}, \quad V_f A = \Lambda \otimes_{\mathbb{Z}} \mathbb{A}_f. \tag{41}$$

An endomorphism $a$ of $A$ defines a $\mathbb{C}$-linear endomorphism $(da)_0 = \alpha$ of $\mathrm{Tgt}_0(A)$ such that $\alpha(\Lambda) \subset \Lambda$ (see 6.5), and $P_a(T)$ is the characteristic polynomial of $\alpha$ on $\Lambda$.

For abelian varieties $A, B$, $\mathrm{Hom}(A,B)$ is a torsion free $\mathbb{Z}$-module of finite rank. We let $\mathsf{AV}(k)$ denote the category of abelian varieties and homomorphisms over $k$ and $\mathsf{AV}^0(k)$ the category with the same objects but with

$$\mathrm{Hom}_{\mathsf{AV}^0(k)}(A,B) = \mathrm{Hom}^0(A,B) = \mathrm{Hom}_{\mathsf{AV}(k)}(A,B) \otimes \mathbb{Q}.$$

An ***isogeny*** of abelian varieties is a surjective homomorphism with finite kernel. A homomorphism of abelian varieties is an isogeny if and only if it becomes an isomorphism in the category $\mathsf{AV}^0$. Two abelian varieties are said to be ***isogenous*** if there is an isogeny from one to the other — this is an equivalence relation.

An abelian variety $A$ over a field $k$ is ***simple*** if it contains no nonzero proper abelian subvariety. Every abelian variety is isogenous to a product of simple abelian varieties. If $A$ and $B$ are simple, then every nonzero homomorphism from $A$ to $B$ is an isogeny. It follows that $\mathrm{End}^0(A)$ is a division algebra when $A$ is simple and a semisimple algebra in general.

NOTES. For a detailed account of abelian varieties over algebraically closed fields, see Mumford 1970, and for a summary over arbitrary fields, see Milne 1986.

**CM fields.** A number field $E$ is a ***CM*** (or ***complex multiplication***) ***field*** if it is a quadratic totally imaginary extension of a totally real field $F$. Let $a \mapsto a^*$ denote the nontrivial automorphism of $E$ fixing $F$. Then $\rho(a^*) = \overline{\rho(a)}$ for every $\rho\colon E \hookrightarrow \mathbb{C}$. We have the following picture:

$$\begin{array}{ccc} E \otimes_{\mathbb{Q}} \mathbb{R} & \approx & \mathbb{C} \times \cdots \times \mathbb{C} \\ | & & | \\ F \otimes_{\mathbb{Q}} \mathbb{R} & \approx & \mathbb{R} \times \cdots \times \mathbb{R} \end{array} \tag{42}$$

The involution $*$ is positive (in the sense of 8.11), because we can compute $\mathrm{Tr}_{E\otimes_{\mathbb{Q}}\mathbb{R}/F\otimes_{\mathbb{Q}}\mathbb{R}}(b^*b)$ on each factor on the right, where it becomes $\mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(\overline{z}z) = 2|z|^2 > 0$. Thus, we are in the PEL situation considered in §8.

Let $E$ be a CM-field with largest real subfield $F$. Each embedding of $F$ into $\mathbb{R}$ will extend to two conjugate embeddings of $E$ into $\mathbb{C}$. A ***CM-type*** $\Phi$ for $E$ is a choice of one element from each conjugate pair $\{\varphi, \overline{\varphi}\}$. In other words, it is a subset $\Phi \subset \mathrm{Hom}(E,\mathbb{C})$ such that

$$\mathrm{Hom}(E,\mathbb{C}) = \Phi \sqcup \overline{\Phi} \qquad (\text{disjoint union, } \overline{\Phi} = \{\overline{\varphi} \mid \varphi \in \Phi\}).$$

Because $E$ is quadratic over $F$, $E = F[\alpha]$ with $\alpha$ a root of a polynomial $X^2 + aX + b$. On completing the square, we obtain an $\alpha$ such that $\alpha^2 \in F^\times$. Then $\alpha^* = -\alpha$. Such an element $\alpha$ of $E$ is said to be ***totally imaginary*** (its image in $\mathbb{C}$ under every embedding is purely imaginary).

**Abelian varieties of CM-type.** Let $E$ be a CM-field of degree $2g$ over $\mathbb{Q}$. Let $A$ be an abelian variety of dimension $g$ over $\mathbb{C}$, and let $i$ be a homomorphism $E \to \mathrm{End}^0(A)$. If

$$\mathrm{Tr}(i(a) \mid \mathrm{Tgt}_0(A)) = \textstyle\sum_{\varphi\in\Phi}\varphi(a), \quad \text{all } a \in E, \tag{43}$$

for some CM-type $\Phi$ of $E$, then $(A, i)$ is said to be of ***CM-type*** $(E, \Phi)$.

REMARK 10.1. (a) In fact, $(A, i)$ will always be of CM-type for some $\Phi$. Recall (p319) that $A(\mathbb{C}) \cong \mathrm{Tgt}_0(A)/\Lambda$ with $\Lambda$ a lattice in $\mathrm{Tgt}_0(A)$ (so $\Lambda \otimes \mathbb{R} \cong \mathrm{Tgt}_0(A)$). Moreover,

$$\begin{aligned}
&\Lambda \otimes \mathbb{Q} \cong H_1(A, \mathbb{Q})\\
&\Lambda \otimes \mathbb{R} \cong H_1(A, \mathbb{R}), \cong \mathrm{Tgt}_0(A)\\
&\Lambda \otimes \mathbb{C} = H_1(A, \mathbb{C}) \cong H^{-1,0} \oplus H^{0,-1} \cong \mathrm{Tgt}_0(A) \oplus \overline{\mathrm{Tgt}_0(A)}.
\end{aligned}$$

Now $H_1(A, \mathbb{Q})$ is a one-dimensional vector space over $E$, so $H_1(A, \mathbb{C}) \cong \bigoplus_{\varphi\colon E\to\mathbb{C}}\mathbb{C}_\varphi$ where $\mathbb{C}_\varphi$ denotes a 1-dimensional vector space with $E$ acting through $\varphi$. If $\varphi$ occurs in $\mathrm{Tgt}_0(A)$, then $\overline{\varphi}$ occurs in $\overline{\mathrm{Tgt}_0(A)}$, and so $\mathrm{Tgt}_0(A) \cong \bigoplus_{\varphi\in\Phi}\mathbb{C}_\varphi$ with $\Phi$ a CM-type for $E$.

(b) A field $E$ of degree $2g$ over $\mathbb{Q}$ acting on a complex abelian variety $A$ of dimension $g$ need not be be CM unless $A$ is simple.

Let $\Phi$ be a CM-type on $E$, and let $\mathbb{C}^\Phi$ be a direct sum of copies of $\mathbb{C}$ indexed by $\Phi$. Denote by $\Phi$ again the homomorphism $\mathcal{O}_E \to \mathbb{C}^\Phi$, $a \mapsto (\varphi a)_{\varphi\in\Phi}$.

PROPOSITION 10.2. *The image $\Phi(\mathcal{O}_E)$ of $\mathcal{O}_E$ in $\mathbb{C}^\Phi$ is a lattice, and the quotient $\mathbb{C}^\Phi/\Phi(\mathcal{O}_E)$ is an abelian variety $A_\Phi$ of CM-type $(E, \Phi)$ for the natural homomorphism $i_\Phi\colon E \to \mathrm{End}^0(A_\Phi)$. Any other pair $(A, i)$ of CM-type $(E, \Phi)$ is $E$-isogenous to $(A_\Phi, i_\Phi)$.*

PROOF. We have

$$\mathcal{O}_E \otimes_{\mathbb{Z}} \mathbb{R} \cong \mathcal{O}_E \otimes_{\mathbb{Z}} \mathbb{Q} \otimes_{\mathbb{Q}} \mathbb{R} \cong E \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow[\cong]{e\otimes r\mapsto(\ldots,r\cdot\varphi e,\ldots)} \mathbb{C}^\Phi,$$

and so $\Phi(\mathcal{O}_E)$ is a lattice in $\mathbb{C}^\Phi$.

To show that the quotient is an abelian variety, we have to exhibit a riemann form (6.7). Let $\alpha$ be a totally imaginary element of $E$. The weak approximation theorem allows us to choose $\alpha$ so that $\Im(\varphi\alpha) > 0$ for $\varphi \in \Phi$, and we can multiply it by an integer (in $\mathbb{N}$) to make it an algebraic integer. Define

$$\psi(u, v) = \mathrm{Tr}_{E/\mathbb{Q}}(\alpha u v^*), \qquad u, v \in \mathcal{O}_E.$$

Then $\psi(u, v) \in \mathbb{Z}$. The remaining properties can be checked on the right of (42). Here $\psi$ takes the form $\psi = \sum_{\varphi\in\Phi}\psi_\varphi$, where

$$\psi_\varphi(u, v) = \mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(\alpha_\varphi \cdot u \cdot \overline{v}), \quad \alpha_\varphi = \varphi(\alpha), \quad u, v \in \mathbb{C}.$$

Because $\alpha$ is totally imaginary,

$$\psi_\varphi(u, v) = \alpha_\varphi(u\overline{v} - \overline{u}v) \in \mathbb{R},$$

from which it follows that $\psi_\varphi(u,u)=0$, $\psi_\varphi(iu,iv)=\psi_\varphi(u,v)$, and $\psi_\varphi(u,iu)>0$ for $u\neq 0$. Thus, $\psi$ is a riemann form and $A_\Phi$ is an abelian variety.

An element $\alpha\in\mathcal{O}_E$ acts on $\mathbb{C}^\Phi$ as muliplication by $\Phi(\alpha)$. This preserves $\Phi(\mathcal{O}_E)$, and so defines a homomorphism $\mathcal{O}_E\to\mathrm{End}(A_\Phi)$. On tensoring this with $\mathbb{Q}$, we obtain the homomorphism $i_\Phi$. The map $\mathbb{C}^\Phi\to\mathbb{C}^\Phi/\Phi(\mathcal{O}_E)$ defines an isomorphism $\mathbb{C}^\Phi=\mathrm{Tgt}_0(\mathbb{C}^\Phi)\to\mathrm{Tgt}_0(A_\Phi)$ compatible with the actions of $E$. Therefore, $(A_\Phi,i_\Phi)$ is of CM-type $(E,\Phi)$.

Finally, let $(A,i)$ be of CM-type $(E,\Phi)$. The condition (43) means that $\mathrm{Tgt}_0(A)$ is isomorphic to $\mathbb{C}^\Phi$ as an $E\otimes_{\mathbb{Q}}\mathbb{C}$-module. Therefore, $A(\mathbb{C})$ is a quotient of $\mathbb{C}^\Phi$ by a lattice $\Lambda$ such that $\mathbb{Q}\Lambda$ is stable under the action of $E$ on $\mathbb{C}^\Phi$ given by $\Phi$ (see 6.7 et seq.). This implies that $\mathbb{Q}\Lambda=\Phi(E)$, and so $\Lambda=\Phi(\Lambda')$ where $\Lambda'$ is a lattice in $E$. Now, $N\Lambda'\subset\mathcal{O}_E$ for some $N$, and we have $E$-isogenies

$$\mathbb{C}^\Phi/\Lambda\xrightarrow{N}\mathbb{C}^\Phi/N\Lambda\leftarrow\mathbb{C}^\Phi/\Phi(\mathcal{O}_E).\qquad\square$$

PROPOSITION 10.3. *Let $(A,i)$ be an abelian variety of CM-type $(E,\Phi)$ over $\mathbb{C}$. Then $(A,i)$ has a model over $\mathbb{Q}^{al}$, uniquely determined up to isomorphism.*

PROOF. Let $k\subset\Omega$ be algebraically closed fields of characteristic zero. For an abelian variety $A$ over $k$, the torsion points in $A(k)$ are zariski dense, and the map on torsion points $A(k)_{\mathrm{tors}}\to A(\Omega)_{\mathrm{tors}}$ is bijective (see (40)), and so every regular map $A_\Omega\to W_\Omega$ ($W$ a variety over $k$) is fixed by the automorphisms of $\Omega/k$ and is therefore defined over $k$ (AG 16.9; see also 13.1 below). It follows that $A\mapsto A_\Omega\colon\mathsf{AV}(k)\to\mathsf{AV}(\Omega)$ is fully faithful.

It remains to show that every abelian variety $(A,i)$ of CM-type over $\mathbb{C}$ arises from a pair over $\mathbb{Q}^{\mathrm{al}}$. The polynomials defining $A$ and $i$ have coefficients in some subring $R$ of $\mathbb{C}$ that is finitely generated over $\mathbb{Q}^{\mathrm{al}}$. According to the Hilbert Nullstellensatz, a maximal ideal $\mathfrak{m}$ of $R$ will have residue field $\mathbb{Q}^{\mathrm{al}}$, and the reduction of $(A,i)$ mod $\mathfrak{m}$ is called a ***specialization*** of $(A,i)$. Any specialization $(A',i')$ of $(A,i)$ to a pair over $\mathbb{Q}^{\mathrm{al}}$ with $A'$ nonsingular will still be of CM-type $(E,\Phi)$, and therefore (see 10.2) there exists an isogeny $(A',i')_{\mathbb{C}}\to(A,i)$. The kernel $H$ of this isogeny is a subgroup of $A'(\mathbb{C})_{\mathrm{tors}}=A'(\mathbb{Q}^{\mathrm{al}})_{\mathrm{tors}}$, and $(A'/H,i)$ will be a model of $(A,i)$ over $\mathbb{Q}^{\mathrm{al}}$. $\square$

REMARK 10.4. The proposition implies that, in order for an elliptic curve $A$ over $\mathbb{C}$ to be of CM-type, its $j$-invariant must be algebraic.

Let $A$ be an abelian variety of dimension $g$ over a subfield $k$ of $\mathbb{C}$, and let $i\colon E\to\mathrm{End}^0(A)$ be a homomorphism with $E$ a CM-field of degree $2g$. Then $\mathrm{Tgt}_0(A)$ is a $k$-vector space of dimension $g$ on which $E$ acts $k$-linearly, and, provided $k$ is large enough to contain all conjugates of $E$, it will decompose into one-dimensional $k$-subspaces indexed by a subset $\Phi$ of $\mathrm{Hom}(E,k)$. When we identify $\Phi$ with a subset of $\mathrm{Hom}(E,\mathbb{C})$, it becomes a CM-type, and we again say $(A,i)$ is of ***CM-type*** $(E,\Phi)$.

Let $A$ be an abelian variety over a number field $K$. We say that $A$ has ***good reduction at*** $\mathfrak{P}$ if it extends to an abelian scheme over $\mathcal{O}_{K,\mathfrak{P}}$, i.e., a smooth proper scheme over $\mathcal{O}_{K,\mathfrak{P}}$ with a group structure. In down-to-earth terms this means the following: embed $A$ as a closed subvariety of some projective space $\mathbb{P}^n_K$; for each polynomial $P(X_0,\ldots,X_n)$ in the homogeneous ideal $\mathfrak{a}$ defining $A\subset\mathbb{P}^n_K$, multiply $P$ by an element of $K$ so that it (just) lies in $\mathcal{O}_{K,\mathfrak{P}}[X_0,\ldots,X_n]$ and let $\overline{P}$ denote the reduction of $P$ modulo $\mathfrak{P}$; the $\overline{P}$'s obtained in this fashion generate a homogeneous $\overline{\mathfrak{a}}$ ideal in $k[X_0,\ldots,X_n]$ where $k=\mathcal{O}_K/\mathfrak{P}$; the abelian variety $A$ has good reduction at $\mathfrak{P}$ if it is possible to choose the projective embedding of $A$ so that the zero set of $\bar{\mathfrak{a}}$ is an abelian variety $\overline{A}$ over $k$. Then $\overline{A}$ is called ***the reduction of** $A$ **at** $\mathfrak{P}$*. It can be shown that, up to a canonical isomorphism, $\overline{A}$ is independent of all choices. For $\ell \neq \text{char}(k)$, $V_\ell(A) \cong V_\ell(\overline{A})$. There is an injective homorphism $\text{End}(A) \to \text{End}(\overline{A})$ compatible with $V_\ell(A) \cong V_\ell(\overline{A})$ (both are reduction maps).

PROPOSITION 10.5. *Let $(A, i)$ be an abelian variety of CM-type $(E, \Phi)$ over a number field $K \subset \mathbb{C}$, and let $\mathfrak{P}$ be a prime ideal in $\mathcal{O}_K$. Then, after possibly replacing $K$ by a finite extension, $A$ will have good reduction at $\mathfrak{P}$.*

PROOF. We use the Néron (alias, Ogg-Shafarevich) criterion (Serre and Tate 1968, Theorem 1):

> an abelian variety over a number field $K$ has good reduction at $\mathfrak{P}$ if for some prime $\ell \neq \text{char}(\mathcal{O}_K/\mathfrak{P})$, the inertia group $I$ at $\mathfrak{P}$ acts trivially on $T_\ell A$.

In our case, $V_\ell A$ is a free $E \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$-module of rank 1 because $H_1(A_{\mathbb{C}}, \mathbb{Q})$ is a one-dimensional vector space over $E$ and $V_\ell A \cong H_1(A_{\mathbb{C}}, \mathbb{Q}) \otimes \mathbb{Q}_\ell$ (see (41)). Therefore, $E \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$ is its own centralizer in $\text{End}_{\mathbb{Q}_\ell}(V_\ell A)$ and the representation of $\text{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})$ on $V_\ell A$ has image in $(E \otimes \mathbb{Q}_\ell)^\times$, and, in fact, in a compact subgroup of $(E \otimes \mathbb{Q}_\ell)^\times$. But such a subgroup will have a pro-$\ell$ subgroup of finite index. Since $I$ has a pro-$p$ subgroup of finite index (ANT, 7.5), this shows that image of $I$ is finite. After $K$ has been replaced by a finite extension, the image of $I$ will be trivial, and Néron's criterion applies. □

**Abelian varieties over a finite field.** Let $\mathbb{F}$ be an algebraic closure of the field $\mathbb{F}_p$ of $p$-elements, and let $\mathbb{F}_q$ be the subfield of $\mathbb{F}$ with $q = p^m$ elements. An element $a$ of $\mathbb{F}$ lies in $\mathbb{F}_q$ if and only if $a^q = a$. Recall that, in characteristic $p$, $(X + Y)^p = X^p + Y^p$. Therefore, if $f(X_1, \ldots, X_n)$ has coefficients in $\mathbb{F}_q$, then

$$f(X_1, \ldots, X_n)^q = f(X_1^q, \ldots, X_n^q), \quad f(a_1, \ldots, a_n)^q = f(a_1^q, \ldots, a_n^q), \quad a_i \in \mathbb{F}.$$

In particular,

$$f(a_1, \ldots, a_n) = 0 \implies f(a_1^q, \ldots, a_n^q) = 0, \quad a_i \in \mathbb{F}.$$

PROPOSITION 10.6. *There is a unique way to attach to every variety $V$ over $\mathbb{F}_q$ a regular map $\pi_V : V \to V$ such that*

(a) *for any regular map $\alpha: V \to W$, $\alpha \circ \pi_V = \pi_W \circ \alpha$;*
(b) *$\pi_{\mathbb{A}^n}$ is the map $(a_1, \ldots, a_n) \mapsto (a_1^q, \ldots, a_n^q)$.*

PROOF. For an affine variety $V = \text{Specm}A$, define $\pi_V$ be the map corresponding to the $\mathbb{F}_q$-homomorphism $x \mapsto x^q : A \to A$. The rest of the proof is straightforward. □

The map $\pi_V$ is called the ***Frobenius map of** $V$*.

THEOREM 10.7 (Weil 1948). *For an abelian variety $A$ over $\mathbb{F}_q$, $\text{End}^0(A)$ is a finite-dimensional semisimple $\mathbb{Q}$-algebra with $\pi_A$ in its centre. For every embedding $\rho: \mathbb{Q}[\pi_A] \to \mathbb{C}$, $|\rho(\pi_A)| = q^{\frac{1}{2}}$.*

PROOF. See, for example, Milne 1986, 19.1. □

If $A$ is simple, $\mathbb{Q}[\pi_A]$ is a field (p334), and $\pi_A$ is an algebraic integer in it (p334). An algebraic integer $\pi$ such that $|\rho(\pi)| = q^{\frac{1}{2}}$ for all embeddings $\rho: \mathbb{Q}[\pi] \to \mathbb{C}$ is called a ***Weil** $q$**-integer*** (formerly, Weil $q$-number).

For a Weil $q$-integer $\pi$,

$$\rho(\pi)\cdot\overline{\rho(\pi)} = q = \rho(\pi)\cdot\rho(q/\pi), \quad \text{all } \rho\colon \mathbb{Q}[\pi]\to\mathbb{C},$$

and so $\rho(q/\pi) = \overline{\rho(\pi)}$. It follows that the field $\rho(\mathbb{Q}[\pi])$ is stable under complex conjugation and that the automorphism of $\mathbb{Q}[\pi]$ induced by complex conjugation sends $\pi$ to $q/\pi$ and is independent of $\rho$. This implies that $\mathbb{Q}[\pi]$ is a CM-field (the typical case), $\mathbb{Q}$, or $\mathbb{Q}[\sqrt{p}]$.

LEMMA 10.8. *Let $\pi$ and $\pi'$ be Weil $q$-integers lying in the same field $E$. If $\mathrm{ord}_v(\pi) = \mathrm{ord}_v(\pi')$ for all $v|p$, then $\pi' = \zeta\pi$ for some root of 1 in $E$.*

PROOF. As noted above, there is an automorphism of $\mathbb{Q}[\pi]$ sending $\pi$ to $q/\pi$. Therefore $q/\pi$ is also an algebraic integer, and so $\mathrm{ord}_v(\pi) = 0$ for every finite $v \nmid p$. Since the same is true for $\pi'$, we find that $|\pi|_v = |\pi'|_v$ for all $v$. Hence $\pi/\pi'$ is a unit in $\mathcal{O}_E$ such that $|\pi/\pi'|_v = 1$ for all $v|\infty$. But in the course of proving the unit theorem, one shows that such a unit has to be root of 1 (ANT, 5.6). □

**The Shimura-Taniyama formula.**

LEMMA 10.9. *Let $(A, i)$ be an abelian variety of CM-type $(E, \Phi)$ over a number field $k \subset \mathbb{C}$ having good reduction at $\mathfrak{P} \subset \mathcal{O}_k$ to $(\overline{A}, \bar{\imath})$ over $\mathcal{O}_k/\mathfrak{P} = \mathbb{F}_q$. Then the Frobenius map $\pi_{\overline{A}}$ of $\overline{A}$ lies in $\bar{\imath}(E)$.*

PROOF. Let $\pi = \pi_{\overline{A}}$. It suffices to check that $\pi$ lies in $\bar{\imath}(E)$ after tensoring with $\mathbb{Q}_\ell$. As we saw in the proof of (10.5), $V_\ell A$ is a free $E\otimes_{\mathbb{Q}}\mathbb{Q}_\ell$-module of rank 1. It follows that $V_\ell\overline{A}$ is also a free $E\otimes_{\mathbb{Q}}\mathbb{Q}_\ell$-module of rank 1 (via $\bar{\imath}$). Therefore, any endomorphism of $V_\ell\overline{A}$ commuting with the action of $E\otimes\mathbb{Q}_\ell$ will lie in $E\otimes\mathbb{Q}_\ell$. □

Thus, from $(A, i)$ and a prime $\mathfrak{P}$ of $k$ at which $A$ has good reduction, we get a Weil $q$-integer $\pi \in E$.

THEOREM 10.10 (Shimura-Taniyama). *In the situation of the lemma, assume that $k$ is galois over $\mathbb{Q}$ and contains all conjugates of $E$. Then for all primes $v$ of $E$ dividing $p$,*

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)} = \frac{|\Phi\cap H_v|}{|H_v|} \tag{44}$$

*where $H_v = \{\rho\colon E\to k \mid \rho^{-1}(\mathfrak{P}) = \mathfrak{p}_v\}$ and $|S|$ denotes the order of a set $S$.*

REMARK 10.11. (a) According to (10.8), the theorem determines $\pi$ up to a root of 1. Note that the formula depends only on $(E, \Phi)$. It is possible to see directly that different pairs $(A, i)$ over $k$ of CM-type $(E, \Phi)$ can give different Frobenius elements, but they will differ only by a root of 1.

(b) Let $*$ denote complex conjugation on $\mathbb{Q}[\pi]$. Then $\pi\pi^* = q$, and so

$$\mathrm{ord}_v(\pi) + \mathrm{ord}_v(\pi^*) = \mathrm{ord}_v(q). \tag{45}$$

Moreover,

$$\mathrm{ord}_v(\pi^*) = \mathrm{ord}_{v^*}(\pi)$$

and

$$\Phi\cap H_{v^*} = \overline{\Phi}\cap H_v.$$

Therefore, (44) is consistent with (45):

$$\frac{\mathrm{ord}_v(\pi)}{\mathrm{ord}_v(q)} + \frac{\mathrm{ord}_v(\pi^*)}{\mathrm{ord}_v(q)} \overset{(44)}{=} \frac{|\Phi\cap H_v| + |\Phi\cap H_{v^*}|}{|H_v|} = \frac{|(\Phi\cup\overline{\Phi})\cap H_v|}{|H_v|} = 1.$$

In fact, (44) is the only obvious formula for $\mathrm{ord}_v(\pi)$ consistent with (45), which is probably a more convincing argument for its validity than the proof sketched below.

*The $\mathcal{O}_E$-structure of the tangent space.* Let $R$ be a Dedekind domain. Any finitely generated torsion $R$-module $M$ can be written as a direct sum $\bigoplus_i R/\mathfrak{p}_i^{r_i}$ with each $\mathfrak{p}_i$ an ideal in $R$, and the set of pairs $(\mathfrak{p}_i, r_i)$ is uniquely determined by $M$. Define $|M|_R = \prod \mathfrak{p}_i^{r_i}$. For example, for $R = \mathbb{Z}$, $M$ is a finite abelian group and $|M|_{\mathbb{Z}}$ is the ideal in $\mathbb{Z}$ generated by the order of $M$.

For Dedekind domains $R \subset S$ with $S$ finite over $R$, there is a norm homomorphism sending fractional ideals of $S$ to fractional ideals of $R$ (ANT, p58). It is compatible with norms of elements, and

$$\mathrm{Nm}(\mathfrak{P}) = \mathfrak{p}^{f(\mathfrak{P}/\mathfrak{p})}, \quad \mathfrak{P} \text{ prime}, \ \mathfrak{p} = \mathfrak{P} \cap R$$

Clearly,

$$|S/\mathfrak{A}|_R = \mathrm{Nm}(\mathfrak{A}) \tag{46}$$

since this is true for prime ideals, and both sides are multiplicative.

PROPOSITION 10.12. *Let $A$ be an abelian variety of dimension $g$ over $\mathbb{F}_q$, and let $i$ be a homomorphism from the ring of integers $\mathcal{O}_E$ of a field $E$ of degree $2g$ over $\mathbb{Q}$ into* $\mathrm{End}(A)$. *Then*

$$|\operatorname{Tgt}_0 A|_{\mathcal{O}_E} = (\pi_A).$$

PROOF. Omitted (for a scheme-theoretic proof, see Giraud 1968, Théorème 1). □

*Sketch of the proof the Shimura-Taniyama formula.* We return to the situation of the Theorem 10.10. After replacing $A$ with an isogenous variety, we may assume $i(\mathcal{O}_E) \subset \mathrm{End}(A)$. By assumption, there exists an abelian scheme $\mathcal{A}$ over $\mathcal{O}_{k,\mathfrak{P}}$ with generic fibre $A$ and special fibre an abelian variety $\overline{A}$. Because $\mathcal{A}$ is smooth over $\mathcal{O}_{k,\mathfrak{P}}$, the relative tangent space of $\mathcal{A}/\mathcal{O}_{k,\mathfrak{P}}$ is a free $\mathcal{O}_{k,\mathfrak{P}}$-module $T$ of rank $g$ endowed with an action of $\mathcal{O}_E$ such that

$$T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} k = \operatorname{Tgt}_0(A), \quad T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} \mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P} = \operatorname{Tgt}_0(\overline{A}).$$

Therefore,

$$(\pi) \overset{10.12}{=} \left|\operatorname{Tgt}_0 \overline{A}\right|_{\mathcal{O}_E} = \left|T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} (\mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P})\right|_{\mathcal{O}_E}. \tag{47}$$

For simplicity, assume that $(p) =_{\mathrm{df}} \mathfrak{P} \cap \mathbb{Z}$ is unramified in $E$. Then the isomorphism of $E$-modules

$$T \otimes_{\mathcal{O}_{k,\mathfrak{P}}} k \approx k^{\Phi}$$

induces an isomorphism of $\mathcal{O}_E$-modules

$$T \approx \mathcal{O}_{k,\mathfrak{P}}^{\Phi} \tag{48}$$

In other words, $T$ is a direct sum of copies of $\mathcal{O}_{k,\mathfrak{P}}$ indexed by the elements of $\Phi$, and $\mathcal{O}_E$ acts on the $\varphi^{\text{th}}$ copy through the map

$$\mathcal{O}_E \xrightarrow{\varphi} \mathcal{O}_k \subset \mathcal{O}_{k,\mathfrak{P}}.$$

As $\mathcal{O}_k/\mathfrak{P} \cong \mathcal{O}_{k,\mathfrak{P}}/\mathfrak{P}$ (ANT, 3.11), the contribution of the $\varphi^{\text{th}}$ copy to $(\pi)$ in (47) is

$$|\mathcal{O}_k/\mathfrak{P}|_{\mathcal{O}_E} \overset{(46)}{=} \varphi^{-1}(\mathrm{Nm}_{k/\varphi E}\,\mathfrak{P}).$$

Thus,

$$(\pi)=\prod\nolimits_{\varphi\in\Phi}\varphi^{-1}(\mathrm{Nm}_{k/\varphi E}\,\mathfrak{P}). \tag{49}$$

It is only an exercise to derive (44) from (49).

NOTES. The original formulation of the Shimura-Taniyama theorem is in fact (49). It is proved in Shimura and Taniyama 1961, III.13, in the unramified case using spaces of differentials rather than tangent spaces. The proof sketched above is given in detail in Giraud 1968, and there is a proof using $p$-divisible groups in Tate 1969, §5. See also Serre 1968, pII-28.

## 11. Complex multiplication: the main theorem

**Review of class field theory.** Classical class field theory classifies the abelian extensions of a number field $E$, i.e., the galois extensions $L/E$ such $\mathrm{Gal}(L/E)$ is commutative. Let $E^{\mathrm{ab}}$ be the composite of all the finite abelian extensions of $E$ inside some fixed algebraic closure $E^{\mathrm{al}}$ of $E$. Then $E^{\mathrm{ab}}$ is an infinite galois extension of $E$.

According to class field theory, there exists a continuous surjective homomorphism (the ***reciprocity*** or ***Artin map***)

$$\mathrm{rec}_E\colon \mathbb{A}_E^\times\to\mathrm{Gal}(E^{\mathrm{ab}}/E)$$

such that, for every finite extension $L$ of $E$ contained in $E^{\mathrm{ab}}$, $\mathrm{rec}_E$ gives rise to a commutative diagram

$$\begin{tikzcd}
E^\times\backslash\mathbb{A}_E^\times \arrow[r, "\mathrm{rec}_E", "\text{onto}"'] \arrow[d] & \mathrm{Gal}(E^{\mathrm{ab}}/E) \arrow[d, "\sigma\mapsto\sigma|L"] \\
E^\times\backslash\mathbb{A}_E^\times/\mathrm{Nm}_{L/E}(\mathbb{A}_L^\times) \arrow[r, "\mathrm{rec}_{L/E}", "\cong"'] & \mathrm{Gal}(L/E).
\end{tikzcd}$$

It is determined by the following two properties:

(a) $\mathrm{rec}_{L/E}(u)=1$ for every $u=(u_v)\in\mathbb{A}_E^\times$ such that
  i) if $v$ is unramified in $L$, then $u_v$ is a unit,
  ii) if $v$ is ramified in $L$, then $u_v$ is sufficiently close to 1 (depending only on $L/E$), and
  iii) if $v$ is real but becomes complex in $L$, then $u_v>0$.

(b) For every prime $v$ of $E$ unramified in $L$, the idèle

$$\alpha=(1,\ldots,1,\underset{v}{\pi},1,\ldots),\quad \pi \text{ a prime element of } \mathcal{O}_{E_v},$$

maps to the Frobenius element $(v,L/E)\in\mathrm{Gal}(L/E)$.

Recall that if $\mathfrak{P}$ is a prime ideal of $L$ lying over $\mathfrak{p}_v$, then $(v,L/E)$ is the automorphism of $L/E$ fixing $\mathfrak{P}$ and acting as $x\mapsto x^{(\mathcal{O}_E\,:\,\mathfrak{p}_v)}$ on $\mathcal{O}_L/\mathfrak{P}$.

To see that there is at most one map satisfying these conditions, let $\alpha\in\mathbb{A}_E^\times$, and use the weak approximation theorem to choose an $a\in E^\times$ that is close to $\alpha_v$ for all primes $v$ that ramify in $L$ or become complex. Then $\alpha=au\beta$ with $u$ an idèle as in (a) and $\beta$ a finite product of idèles as in (b). Now $\mathrm{rec}_{L/E}(\alpha)=\mathrm{rec}_{L/E}(\beta)$, which can be computed using (b).

Note that, because $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ is totally disconnected, the identity component of $E^\times\backslash\mathbb{A}_E^\times$ is contained in the kernel of $\mathrm{rec}_E$. In particular, the identity component of $\prod_{v|\infty} E_v^\times$ is contained in the kernel, and so, when $E$ is totally imaginary, $\mathrm{rec}_E$ factors through $E^\times \backslash \mathbb{A}_{E,f}^\times$.

For $E = \mathbb{Q}$, the reciprocity map factors through $\mathbb{Q}^\times \backslash \{\pm\} \times \mathbb{A}_f^\times$, and every element in this quotient is uniquely represented by an element of $\hat{\mathbb{Z}}^\times \subset \mathbb{A}_f^\times$. In this case, we get the diagram

$$\begin{tikzcd} \hat{\mathbb{Z}}^\times \arrow[r, "\mathrm{rec}_{\mathbb{Q}}", "\cong"'] \arrow[d] & \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) \arrow[r, equal] \arrow[d, "\mathrm{restrict}"] & \bigcup \mathbb{Q}[\zeta_N] \\ (\mathbb{Z}/N\mathbb{Z})^\times \arrow[r, "{[a]\mapsto(\zeta_N\mapsto\zeta_N^a)}", "\cong"'] & \mathrm{Gal}(\mathbb{Q}[\zeta_N]/\mathbb{Q}) & \end{tikzcd} \tag{50}$$

which *commutes with an inverse.* This can be checked by writing an idèle $\alpha$ in the form $au\beta$ as above, but it is more instructive to look at an example. Let $p$ be a prime not dividing $N$, and let

$$\alpha = p \cdot (1, \ldots, 1, \underset{p}{p^{-1}}, 1, \ldots) \in \mathbb{Z} \cdot \mathbb{A}_f^\times = \mathbb{A}_f^\times.$$

Then $\alpha \in \hat{\mathbb{Z}}^\times$ and has image $[p]$ in $\mathbb{Z}/N\mathbb{Z}$, which acts as $(p, \mathbb{Q}[\zeta_N]/\mathbb{Q})$ on $\mathbb{Q}[\zeta_N]$. On the other hand, $\mathrm{rec}_{\mathbb{Q}}(\alpha) = \mathrm{rec}_{\mathbb{Q}}((1, \ldots, p^{-1}, \ldots))$, which acts as $(p, \mathbb{Q}[\zeta_N]/\mathbb{Q})^{-1}$.

NOTES. For the proofs of the above statements, see Tate 1967 or my notes CFT.

**Convention for the (Artin) reciprocity map.** It simplifies the formulas in Langlands theory if one replaces the reciprocity map with its reciprocal. For $\alpha \in \mathbb{A}_E^\times$, write

$$\mathrm{art}_E(\alpha) = \mathrm{rec}_E(\alpha)^{-1}. \tag{51}$$

Now, the diagram (50) commutes. In other words,

$$\mathrm{art}_{\mathbb{Q}}(\chi(\sigma)) = \sigma, \quad \text{for } \sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}),$$

where $\chi$ is the cyclotomic character $\mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) \to \hat{\mathbb{Z}}^\times$, which is characterized by

$$\sigma\zeta = \zeta^{\chi(\sigma)}, \quad \zeta \text{ a root of 1 in } \mathbb{C}^\times.$$

**The reflex field and norm of a CM-type.** Let $(E, \Phi)$ be a CM-type.

DEFINITION 11.1. The ***reflex field*** $E^*$ of $(E, \Phi)$ is the subfield of $\mathbb{Q}^{\mathrm{al}}$ characterized by any one of the following equivalent conditions:

(a) $\sigma \in \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ fixes $E^*$ if and only if $\sigma\Phi = \Phi$; here $\sigma\Phi = \{\sigma \circ \varphi | \varphi \in \Phi\}$;
(b) $E^*$ is the field generated over $\mathbb{Q}$ by the elements $\sum_{\varphi\in\Phi} \varphi(a)$, $a \in E$;
(c) $E^*$ is the smallest subfield $k$ of $\mathbb{Q}^{\mathrm{al}}$ such that there exists a $k$-vector space $V$ with an action of $E$ for which

$$\mathrm{Tr}_k(a|V) = \textstyle\sum_{\varphi\in\Phi} \varphi(a), \quad \text{all } a \in E.$$

Let $V$ be an $E^*$-vector space with an action of $E$ such that $\mathrm{Tr}_{E^*}(a|V) = \sum_{\varphi\in\Phi} \varphi(a)$ for all $a \in E$. We can regard $V$ as an $E^* \otimes_{\mathbb{Q}} E$-space, or as an $E$-vector space with a $E$-linear action of $E^*$. The ***reflex norm*** is the homomorphism $N_{\Phi^*} \colon (\mathbb{G}_m)_{E^*/\mathbb{Q}} \to (\mathbb{G}_m)_{E/\mathbb{Q}}$ such that

$$N_{\Phi^*}(a) = \det{}_E(a|V), \quad \text{all } a \in E^{*\times}.$$

This definition is independent of the choice of $V$ because $V$ is unique up to an isomorphism respecting the actions of $E$ and $E^*$.

Let $(A, i)$ be an abelian variety of CM-type $(E, \Phi)$ defined over $\mathbb{C}$. According to (11.1c) applied to $\mathrm{Tgt}_0(A)$, any field of definition of $(A, i)$ contains $E^*$.

**Statement of the main theorem of complex multiplication.** A homomorphism $\sigma\colon k \to \Omega$ of fields defines a functor $V \mapsto \sigma V$, $\alpha \mapsto \sigma\alpha$, "extension of the base field" from varieties over $k$ to varieties over $\Omega$. In particular, an abelian variety $A$ over $k$ equipped with a homomorphism $i\colon E \to \mathrm{End}^0(A)$ defines a similar pair $\sigma(A, i) = (\sigma A, {}^\sigma i)$ over $\Omega$. Here ${}^\sigma i\colon E \to \mathrm{End}(\sigma A)$ is defined by

$$ {}^\sigma i(a) = \sigma(i(a)). $$

A point $P \in A(k)$ gives a point $\sigma P \in A(\Omega)$, and so $\sigma$ defines a homomorphism $\sigma\colon V_f(A) \to V_f(\sigma A)$ provided that $k$ and $\Omega$ are algebraically closed (otherwise one would have to choose an extension of $k$ to a homomorphism $k^{\mathrm{al}} \to \Omega^{\mathrm{al}}$).

THEOREM 11.2. *Let $(A, i)$ be an abelian variety of CM-type $(E, \Phi)$ over $\mathbb{C}$, and let $\sigma \in \mathrm{Aut}(\mathbb{C}/E^*)$. For any $s \in \mathbb{A}^\times_{E^*,f}$ with $\mathrm{art}_{E^*}(s) = \sigma|E^{*ab}$, there is a unique $E$-linear isogeny $\alpha\colon A \to \sigma A$ such that $\alpha(N_{\Phi^*}(s) \cdot x) = \sigma x$ for all $x \in V_f A$.*

PROOF. Formation of the tangent space commutes with extension of the base field, and so

$$ \mathrm{Tgt}_0(\sigma A) = \mathrm{Tgt}_0(A) \otimes_{\mathbb{C},\sigma} \mathbb{C} $$

as an $E \otimes_{\mathbb{Q}} \mathbb{C}$-module. Therefore, $(\sigma A, {}^\sigma i)$ is of CM type $\sigma\Phi$. Since $\sigma$ fixes $E^*$, $\sigma\Phi = \Phi$, and so there exists an $E$-linear isogeny $\alpha\colon A \to \sigma A$ (10.2). The map

$$ V_f(A) \xrightarrow{\sigma} V_f(\sigma A) \xrightarrow{V_f(\alpha)^{-1}} V_f(A) $$

is $E \otimes_{\mathbb{Q}} \mathbb{A}_f$-linear. As $V_f(A)$ is free of rank one over $E \otimes_{\mathbb{Q}} \mathbb{A}_f = \mathbb{A}_{E,f}$, this map must be multiplication by an element of $a \in \mathbb{A}^\times_{E,f}$. When the choice of $\alpha$ is changed, then $a$ is changed only by an element of $E^\times$, and so we have a well-defined map

$$ \sigma \mapsto aE^\times\colon\ \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*) \to \mathbb{A}^\times_{E,f}/E^\times, $$

which one checks to be a homomorphism. The map factors through $\mathrm{Gal}(E^{*\mathrm{ab}}/E^*)$, and so, when composed with the reciprocity map $\mathrm{art}_{E^*}$, it gives a homomorphism

$$ \eta\colon \mathbb{A}^\times_{E^*,f}/E^{*\times} \to \mathbb{A}^\times_{E,f}/E^\times. $$

We have to check that $\eta$ is the homomorphism defined by $N_{\Phi^*}$, but it can be shown that this follows from the Shimura-Taniyama formula (Theorem 10.10). The uniqueness follows from the faithfulness of the functor $A \mapsto V_f(A)$. □

REMARK 11.3. (a) If $s$ is replaced by $as$, $a \in E^{*\times}$, then $\alpha$ must be replaced by $\alpha \circ N_{\Phi^*}(a)^{-1}$.

(b) The theorem is a statement about the $E$-isogeny class of $(A, i)$. If $\beta\colon (A, i) \to (B, j)$ is an $E$-linear isogeny, and $\alpha$ satisfies the conditions of the theorem for $(A, i)$, then $(\sigma\beta) \circ \alpha \circ \beta^{-1}$ satisfies the conditions for $(B, j)$.

ASIDE 11.4. What happens in (11.2) when $\sigma$ is not assumed to fix $E^*$? This also is known, thanks to Deligne and Langlands. For a discussion of this, and much else concerning complex multiplication, see my notes Milne 1979.

## 12. Definition of canonical models

We attach to each Shimura datum $(G, X)$ an algebraic number field $E(G, X)$, and we define the canonical model of $\mathrm{Sh}(G, X)$ to be an inverse system of varieties over $E(G, X)$ that is characterized by reciprocity laws at certain special points.

**Models of varieties.** Let $k$ be a subfield of a field $\Omega$, and let $V$ be a variety over $\Omega$. A ***model*** of $V$ over $k$ (or a $k$-***structure*** on $V$) is a variety $V_0$ over $k$ together with an isomorphism $\varphi\colon V_{0\Omega} \to V$. We often omit the map $\varphi$ and regard a model as a variety $V_0$ over $k$ such that $V_{0\Omega} = V$.

Consider an affine variety $V$ over $\mathbb{C}$ and a subfield $k$ of $\mathbb{C}$. An embedding $V \hookrightarrow \mathbb{A}^n_{\mathbb{C}}$ defines a model of $V$ over $k$ if the ideal $I(V)$ of polynomials zero on $V$ is generated by polynomials in $k[X_1, \ldots, X_n]$, because then $I_0 =_{\mathrm{df}} I(V) \cap k[X_1, \ldots, X_n]$ is a radical ideal, $k[X_1, \ldots, X_n]/I_0$ is an affine $k$-algebra, and $V(I_0) \subset \mathbb{A}^n_k$ is a model of $V$. Moreover, every model $(V_0, \varphi)$ arises in this way because every model of an affine variety is affine. However, different embeddings in affine space will usually give rise to different models. For example, the embeddings

$$\mathbb{A}^2_{\mathbb{C}} \xleftarrow{(x,y) \longleftarrow\!\shortmid (x,y)} V(X^2 + Y^2 - 1) \xrightarrow{(x,y) \longmapsto (x, y/\sqrt{2})} \mathbb{A}^2_{\mathbb{C}}$$

define the $\mathbb{Q}$-structures

$$X^2 + Y^2 = 1, \quad X^2 + 2Y^2 = 1$$

on the curve $X^2 + Y^2 = 1$. These are not isomorphic.

Similar remarks apply to projective varieties.

In general, a variety over $\mathbb{C}$ will not have a model over a number field, and when it does, it will have many. For example, an elliptic curve $E$ over $\mathbb{C}$ has a model over a number field if and only if its $j$-invariant $j(E)$ is an algebraic number, and if $Y^2Z = X^3 + aXZ^2 + bZ^3$ is one model of $E$ over a number field $k$ (meaning, $a, b \in k$), then $Y^2Z = X^3 + ac^2XZ^2 + bc^3Z^3$ is a second, which is isomorphic to the first only if $c$ is a square in $k$.

**The reflex field.** For a reductive group $G$ over $\mathbb{Q}$ and a subfield $k$ of $\mathbb{C}$, we write $\mathcal{C}(k)$ for the set of $G(k)$-conjugacy classes of cocharacters of $G_k$ defined over $k$:

$$\mathcal{C}(k) = G(k)\backslash \operatorname{Hom}(\mathbb{G}_m, G_k).$$

A homomorphism $k \to k'$ induces a map $\mathcal{C}(k) \to \mathcal{C}(k')$; in particular, $\operatorname{Aut}(k'/k)$ acts on $\mathcal{C}(k')$.

LEMMA 12.1. *Assume $G$ splits over $k$, so that it contains a split maximal torus $T$, and let $W$ be the Weyl group $N_{G(k)}(T)/C_{G(k)}(T)$ of $T$. Then the map*

$$W\backslash \operatorname{Hom}(\mathbb{G}_m, T_k) \to G(k)\backslash \operatorname{Hom}(\mathbb{G}_m, G_k)$$

*is bijective.*

PROOF. As any two maximal split tori are conjugate (Springer 1998, 15.2.6), the map is surjective. Let $\mu$ and $\mu'$ be cocharacters of $T$ that are conjugate by an element of $G(k)$, say, $\mu = \mathrm{ad}(g) \circ \mu'$ with $g \in G(k)$. Then $\mathrm{ad}(g)(T)$ and $T$ are both maximal split tori in the centralizer $C$ of $\mu(\mathbb{G}_m)$, which is a connected reductive group (ibid., 15.3.2). Therefore, there exists a $c \in C(k)$ such that $\mathrm{ad}(cg)(T) = T$. Now $cg$ normalizes $T$ and $\mathrm{ad}(cg) \circ \mu' = \mu$, which proves that $\mu$ and $\mu'$ are in the same $W$-orbit. $\square$

Let $(G,X)$ be a Shimura datum. For each $x\in X$, we have a cocharacter $\mu_x$ of $G_{\mathbb{C}}$:

$$\mu_x(z)=h_{x\mathbb{C}}(z,1).$$

A different $x\in X$ will give a conjugate $\mu_x$, and so $X$ defines an element $c(X)$ of $\mathcal{C}(\mathbb{C})$. Neither $\operatorname{Hom}(\mathbb{G}_m,T_{\mathbb{Q}^{\mathrm{al}}})$ nor $W$ changes when we replace $\mathbb{C}$ with the algebraic closure $\mathbb{Q}^{\mathrm{al}}$ of $\mathbb{Q}$ in $\mathbb{C}$, and so the lemma shows that $c(X)$ contains a $\mu$ defined over $\mathbb{Q}^{\mathrm{al}}$ and that the $G(\mathbb{Q}^{\mathrm{al}})$-conjugacy class of $\mu$ is independent of the choice of $\mu$. This allows us to regard $c(X)$ as an element of $\mathcal{C}(\mathbb{Q}^{\mathrm{al}})$.

DEFINITION 12.2. The ***reflex (***or ***dual) field*** $E(G,X)$ is the field of definition of $c(X)$, i.e., it is the fixed field of the subgroup of $\operatorname{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ fixing $c(X)$ as an element of $\mathcal{C}(\mathbb{Q}^{\mathrm{al}})$ (or stabilizing $c(X)$ as a subset of $\operatorname{Hom}(\mathbb{G}_m,G_{\mathbb{Q}^{\mathrm{al}}})$).

Note that the reflex field a subfield of $\mathbb{C}$.

REMARK 12.3. (a) Any subfield $k$ of $\mathbb{Q}^{\mathrm{al}}$ splitting $G$ contains $E(G,X)$. This follows from the lemma, because $W\backslash\operatorname{Hom}(\mathbb{G}_m,T)$ does not change when we pass from $k$ to $\mathbb{Q}^{\mathrm{al}}$. If follows that $E(G,X)$ has finite degree over $\mathbb{Q}$.

(b) If $c(X)$ contains a $\mu$ defined over $k$, then $k\supset E(G,X)$. Conversely, if $G$ is quasi-split over $k$ and $k\supset E(G,X)$, then $c(X)$ contains a $\mu$ defined over $k$ (Kottwitz 1984, 1.1.3).

(c) Let $(G,X)\overset{i}{\hookrightarrow}(G',X')$ be an inclusion of Shimura data. Suppose $\sigma$ fixes $c(X)$, and let $\mu\in c(X)$. Then $\sigma\mu=g\cdot\mu\cdot g^{-1}$ for some $g\in G(\mathbb{Q}^{\mathrm{al}})$, and so, for any $g'\in G'(\mathbb{Q}^{\mathrm{al}})$,

$$\sigma(g'\cdot(i\circ\mu)\cdot g'^{-1})=(\sigma g')(i(g))\cdot i\circ\mu\cdot(i(g))^{-1}(\sigma g')^{-1}\in c(X').$$

Hence $\sigma$ fixes $c(X')$, and we have shown that

$$E(G,X)\supset E(G',X').$$

EXAMPLE 12.4. (a) Let $T$ be a torus over $\mathbb{Q}$, and let $h$ be a homomorphism $\mathbb{S}\to T_{\mathbb{R}}$. Then $E(T,h)$ is the field of definition of $\mu_h$, i.e., the smallest subfield of $\mathbb{C}$ over which $\mu_h$ is defined.

(b) Let $(E,\Phi)$ be a CM-type, and let $T$ be the torus $(\mathbb{G}_m)_{E/\mathbb{Q}}$, so that $T(\mathbb{Q})=E^{\times}$ and

$$T(\mathbb{R})=(E\otimes_{\mathbb{Q}}\mathbb{R})^{\times}\cong(\mathbb{C}^{\Phi})^{\times},\quad (e\otimes r)\mapsto(\varphi(e)\cdot r)_{\varphi\in\Phi}.$$

Define $h_\Phi\colon\mathbb{C}^{\times}\to T(\mathbb{R})$ to be $z\mapsto(z,\ldots,z)$. The corresponding cocharacter $\mu_\Phi$ is

$$\begin{array}{ccccc}\mathbb{C}^{\times} & \to & T(\mathbb{C}) & \cong & (\mathbb{C}^{\Phi})^{\times}\times(\mathbb{C}^{\overline{\Phi}})^{\times}\\ z & \mapsto & & & (z,\ldots,z,1,\ldots,1)\end{array}$$

Therefore, $\sigma\mu_\Phi=\mu_\Phi$ if and only if $\sigma$ stabilizes $\Phi$, and so $E(T,h_\Phi)$ is the reflex field of $(E,\Phi)$ defined in (11.1).

(c) If $(G,X)$ is a simple PEL datum of type (A) or (C), then $E(G,X)$ is the field generated over $\mathbb{Q}$ by $\{\operatorname{Tr}_X(b)\mid b\in B\}$ (Deligne 1971*c*, 6.1).

(d) Let $(G,X)$ be the Shimura datum attached to a quaternion algebra $B$ over a totally real number field $F$, as in Example 5.24. Then $c(X)$ is represented by the cocharacter $\mu$:

$$\begin{array}{ccccc}G(\mathbb{C}) & \approx & \mathrm{GL}_2(\mathbb{C})^{I_{\mathrm{c}}} & \times & \mathrm{GL}_2(\mathbb{C})^{I_{\mathrm{nc}}}\\ \mu(z) & = & (1,\ldots,1) & \times & \left(\left(\begin{smallmatrix} z & 0\\ 0 & 1\end{smallmatrix}\right),\ldots,\left(\begin{smallmatrix} z & 0\\ 0 & 1\end{smallmatrix}\right)\right).\end{array}$$

Therefore, $E(G,X)$ is the fixed field of the stabilizer in $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ of $I_{\mathrm{nc}} \subset I$. For example, if $I_{\mathrm{nc}}$ consists of a single element $v$ (so we have a Shimura curve), then $E(G,X) = v(F)$.

(e) When $G$ is adjoint, $E(G,X)$ can be described as follows. Choose a maximal torus $T$ in $G_{\mathbb{Q}^{\mathrm{al}}}$ and a base $(\alpha_i)_{i\in I}$ for the roots. Recall that the nodes of the dynkin diagram $\Delta$ of $(G,T)$ are indexed by $I$. The galois group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ acts on $\Delta$. Each $c \in \mathcal{C}(\mathbb{Q}^{\mathrm{al}})$ contains a $\mu\colon \mathbb{G}_m \to G_{\mathbb{Q}^{\mathrm{al}}}$ such that $\langle \alpha_i, \mu\rangle \geq 0$ for all $i$ (cf. 1.25), and the map

$$c \mapsto (\langle \alpha_i, \mu\rangle)_{i\in I} \colon \mathcal{C}(\mathbb{Q}^{\mathrm{al}}) \to \mathbb{N}^I \qquad \text{(copies of } \mathbb{N} \text{ indexed by } I\text{)}$$

is a bijection. Therefore, $E(G,X)$ is the fixed field of the subgroup of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ fixing $(\langle \alpha_i, \mu\rangle)_{i\in I} \in \mathbb{N}^I$. It is either totally real or CM (Deligne 1971*b*, p139).

(f) Let $(G,X)$ be a Shimura datum, and let $G \xrightarrow{\nu} T$ be the quotient of $G$ by $G^{\mathrm{der}}$. From $(G,X)$, we get Shimura data $(G^{\mathrm{ad}}, X^{\mathrm{ad}})$ and $(T,h)$ with $h = \nu \circ h_x$ for all $x \in X$. Then $E(G,X) = E(G^{\mathrm{ad}}, X^{\mathrm{ad}}) \cdot E(T,h)$ (Deligne 1971*b*, 3.8).

(g) It follows from (e) and (f) that if $(G,X)$ satisfies SV6, then $E(G,X)$ is either a totally real field or a CM-field.

**Special points.**

Definition 12.5. A point $x \in X$ is said to be ***special*** if there exists a torus $T \subset G$ such that $h_x(\mathbb{C}^\times) \subset T(\mathbb{R})$. We then call $(T,x)$, or $(T,h_x)$, a ***special pair*** in $(G,X)$. When the weight is rational and $Z(G)^\circ$ splits over a CM-field (i.e., SV4 and SV6 hold), the special points and special pairs are called ***CM points*** and ***CM pairs***.

Remark 12.6. Let $T$ be a maximal torus of $G$ such that $T(\mathbb{R})$ fixes $x$, i.e., such that $\mathrm{ad}(t) \circ h_x = h_x$ for all $t \in T(\mathbb{R})$. Because $T_{\mathbb{R}}$ is its own centralizer in $G_{\mathbb{R}}$, this implies that $h_x(\mathbb{C}^\times) \subset T(\mathbb{R})$, and so $x$ is special. Conversely, if $(T,x)$ is special, then $T(\mathbb{R})$ fixes $x$.

Example 12.7. Let $G = \mathrm{GL}_2$ and let $\mathcal{H}_1^\pm = \mathbb{C} \smallsetminus \mathbb{R}$. Then $G(\mathbb{R})$ acts on $\mathcal{H}_1^\pm$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d}.$$

Suppose that $z \in \mathbb{C}\smallsetminus\mathbb{R}$ generates a quadratic imaginary extension $E$ of $\mathbb{Q}$. Using the $\mathbb{Q}$-basis $\{1,z\}$ for $E$, we obtain an embedding $E \hookrightarrow M_2(\mathbb{Q})$, and hence a maximal subtorus $(\mathbb{G}_m)_{E/\mathbb{Q}} \subset G$. As $(\mathbb{G}_m)_{E/\mathbb{Q}}(\mathbb{R})$ fixes $z$, this shows that $z$ is special. Conversely, if $z \in \mathcal{H}_1^\pm$ is special, then $\mathbb{Q}[z]$ is a field of degree 2 over $\mathbb{Q}$.

**The homomorphism $r_x$.** Let $T$ be a torus over $\mathbb{Q}$ and let $\mu$ be a cocharacter of $T$ defined over a finite extension $E$ of $\mathbb{Q}$. For $Q \in T(E)$, the element $\sum_{\rho\colon E\to\mathbb{Q}^{\mathrm{al}}} \rho(Q)$ of $T(\mathbb{Q}^{\mathrm{al}})$ is stable under $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ and hence lies in $T(\mathbb{Q})$. Let $r(T,\mu)$ be the homomorphism $(\mathbb{G}_m)_{E/\mathbb{Q}} \to T$ such that

$$r(T,\mu)(P) = \sum_{\rho\colon E\to\mathbb{Q}^{\mathrm{al}}} \rho(\mu(P)), \quad \text{all } P \in E^\times. \tag{52}$$

Let $(T,x) \subset (G,X)$ be a special pair, and let $E(x)$ be the field of definition of $\mu_x$. We define $r_x$ to be the homomorphism

$$\mathbb{A}_{E(x)}^\times \xrightarrow{r(T,\mu)} T(\mathbb{A}_{\mathbb{Q}}) \xrightarrow{\text{project}} T(\mathbb{A}_{\mathbb{Q},f}). \tag{53}$$

Let $a \in \mathbb{A}_{E(x)}^{\times}$, and write $a = (a_\infty, a_f) \in (E(x) \otimes_{\mathbb{Q}} \mathbb{R})^{\times} \times \mathbb{A}_{E(x),f}^{\times}$; then

$$r_x(a) = \sum_{\rho\colon E \to \mathbb{Q}^{\mathrm{al}}} \rho(\mu_x(a_f)).$$

**Definition of a canonical model.** For a special pair $(T,x) \subset (G,X)$, we have homomorphisms ((51),(53)),

$$\begin{aligned} \mathrm{art}_{E(x)}\colon \mathbb{A}_{E(x)}^{\times} &\twoheadrightarrow \mathrm{Gal}(E(x)^{\mathrm{ab}}/E(x)) \\ r_x\colon \mathbb{A}_{E(x)}^{\times} &\to T(\mathbb{A}_f). \end{aligned}$$

DEFINITION 12.8. Let $(G,X)$ be a Shimura datum, and let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$. A model $M_K(G,X)$ of $\mathrm{Sh}_K(G,X)$ over $E(G,X)$ is ***canonical*** if, for every special pair $(T,x) \subset (G,X)$ and $a \in G(\mathbb{A}_f)$, $[x,a]_K$ has coordinates in $E(x)^{\mathrm{ab}}$ and

$$\sigma[x,a]_K = [x, r_x(s)a]_K, \tag{54}$$

for all

$$\left.\begin{array}{l} \sigma \in \mathrm{Gal}(E(x)^{\mathrm{ab}}/E(x)) \\ s \in \mathbb{A}_{E(x)}^{\times} \end{array}\right\} \text{ with } \mathrm{art}_{E(x)}(s) = \sigma.$$

In other words, $M_K(G,X)$ is canonical if every automorphism $\sigma$ of $\mathbb{C}$ fixing $E(x)$ acts on $[x,a]_K$ according to the rule (54) where $s$ is any idèle such that $\mathrm{art}_{E(x)}(s) = \sigma|E(x)^{\mathrm{ab}}$.

REMARK 12.9. Let $(T_1,x)$ and $(T_2,x)$ be special pairs in $(G,X)$ (with the same $x$). Then $(T_1 \cap T_2, x)$ is also a special pair, and if the condition in (54) holds for one of $(T_1 \cap T_2, x)$, $(T_1,x)$, or $(T_2,x)$, then it holds for all three. Therefore, in stating the definition, we could have considered only special pairs $(T,x)$ with, for example, $T$ minimal among the tori such that $T_{\mathbb{R}}$ contains $h_x(\mathbb{S})$.

DEFINITION 12.10. Let $(G,X)$ be a Shimura datum.

(a) A ***model*** of $\mathrm{Sh}(G,X)$ over a subfield $k$ of $\mathbb{C}$ is an inverse system $M(G,X) = (M_K(G,X))_K$ of varieties over $k$ endowed with a right action of $G(\mathbb{A}_f)$ such that $M(G,X)_{\mathbb{C}} = \mathrm{Sh}(G,X)$ (with its $G(\mathbb{A}_f)$ action).

(b) A model $M(G,X)$ of $\mathrm{Sh}(G,X)$ over $E(G,X)$ is ***canonical*** if each $M_K(G,X)$ is canonical.

**Examples: Shimura varieties defined by tori.** For a field $k$ of characteristic zero, the functor $V \mapsto V(k^{\mathrm{al}})$ is an equivalence from the category of zero-dimensional varieties over $k$ to the category of finite sets endowed with a continuous action of $\mathrm{Gal}(k^{\mathrm{al}}/k)$. Continuous here just means that the action factors through $\mathrm{Gal}(L/k)$ for some finite galois extension $L$ of $k$ contained in $k^{\mathrm{al}}$. In particular, to give a zero-dimensional variety over an algebraically closed field of characteristic zero is just to give a finite set. Thus, a zero-dimensional variety over $\mathbb{C}$ can be regarded as a zero-dimensional variety over $\mathbb{Q}^{\mathrm{al}}$, and to give a model of $V$ over a number field $E$ amounts to giving a continuous action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ on $V(\mathbb{C})$.

*Tori.* Let $T$ be a torus over $\mathbb{Q}$, and let $h$ be a homomorphism $\mathbb{S} \to T_{\mathbb{R}}$. Then $(T,h)$ is a Shimura datum, and $E =_{\mathrm{df}} E(T,h)$ is the field of definition of $\mu_h$. In this case

$$\mathrm{Sh}_K(T,h) = T(\mathbb{Q})\backslash\{h\} \times T(\mathbb{A}_f)/K$$

is a finite set (see 5.22), and (54) defines a continuous action of $\mathrm{Gal}(E^{\mathrm{ab}}/E)$ on $\mathrm{Sh}_K(T,h)$. This action defines a model of $\mathrm{Sh}_K(T,h)$ over $E$, which, by definition, is canonical.

*CM-tori.* Let $(E,\Phi)$ be a CM-type, and let $(T,h_\Phi)$ be the Shimura pair defined in (12.4b). Then $E(T,h_\Phi)=E^*$, and $r(T,\mu_\Phi)\colon(\mathbb{G}_m)_{E^*/\mathbb{Q}}\to(\mathbb{G}_m)_{E/\mathbb{Q}}$ is the reflex norm $N_{\Phi^*}$.

Let $K$ be a compact open subgroup of $T(\mathbb{A}_f)$. The Shimura variety $\mathrm{Sh}_K(T,h_\Phi)$ classifies isomorphism classes of triples $(A,i,\eta K)$ in which $(A,i)$ is an abelian variety over $\mathbb{C}$ of CM-type $(E,\Phi)$ and $\eta$ is an $E\otimes\mathbb{A}_f$-linear isomorphism $V(\mathbb{A}_f)\to V_f(A)$. An isomorphism $(A,i,\eta K)\to(A',i',\eta' K)$ is an $E$-linear isomorphism $A\to A'$ in $\mathsf{AV}^0(\mathbb{C})$ sending $\eta K$ to $\eta' K$. To see this, let $V$ be a one-dimensional $E$-vector space. The action of $E$ on $V$ realizes $T$ as a subtorus of $\mathrm{GL}(V)$. If $(A,i)$ is of CM-type $(E,\Phi)$, then there exists an $E$-homomorphism $a\colon H_1(A,\mathbb{Q})\to V$ carrying $h_A$ to $h_\Phi$ (see 10.2). Now the isomorphism

$$V(\mathbb{A}_f)\xrightarrow{\eta}V_f(A)\xrightarrow{a}V(\mathbb{A}_f)$$

is $E\otimes\mathbb{A}_f$-linear, and hence is multiplication by an element $g$ of $(E\otimes\mathbb{A}_f)^\times=T^E(\mathbb{A}_f)$. The map $(A,i,\eta)\mapsto[g]$ gives the bijection.

In (10.3) and its proof, we showed that the functor $(A,i)\mapsto(A_\mathbb{C},i_\mathbb{C})$ defines an equivalence from the category of abelian varieties over $\mathbb{Q}^{\mathrm{al}}$ of CM-type $(E,\Phi)$ to the similar category over $\mathbb{C}$ (the abelian varieties are to be regarded as objects of $\mathsf{AV}^0$). Therefore, $\mathrm{Sh}_K(T^E,h_\Phi)$ classifies isomorphism classes of triples $(A,i,\eta K)$ where $(A,i)$ is now an abelian variety over $\mathbb{Q}^{\mathrm{al}}$ of CM-type $(E,\Phi)$.

The group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$ acts on the set $\mathcal{M}_K$ of such triples: let $(A,i,\eta)\in\mathcal{M}_K$; for $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$, define $\sigma(A,i,\eta K)$ to be the triple $(\sigma A,{}^\sigma i,{}^\sigma\eta K)$ where ${}^\sigma\eta$ is the composite

$$V(\mathbb{A}_f)\xrightarrow{\eta}V_f(A)\xrightarrow{\sigma}V_f(\sigma A);\tag{55}$$

because $\sigma$ fixes $E^*$, $(\sigma A,\sigma i)$ is again of CM-type $(E,\Phi)$.

The group $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$ acts on $\mathrm{Sh}_K(T^E,h_\Phi)$ by the rule (54):

$$\sigma[g]=[r_{h_\Phi}(s)g]_K,\quad \mathrm{art}_{E^*}(s)=\sigma|E^*.$$

PROPOSITION 12.11. *The map $(A,i,\eta)\mapsto[a\circ\eta]_K\colon\mathcal{M}_K\to\mathrm{Sh}_K(T^E,h_\Phi)$ commutes with the actions of* $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$.

PROOF. Let $(A,i,\eta)\in\mathcal{M}_K$ map to $[a\circ\eta]_K$ for an appropriate isomorphism $a\colon H_1(A,\mathbb{Q})\to V$, and let $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E^*)$. According to the main theorem of complex multiplication (11.2), there exists an isomorphism $\alpha\colon A\to\sigma A$ such that $\alpha(N_{\Phi^*}(s)\cdot x)=\sigma x$ for $x\in V_f(A)$, where $s\in\mathbb{A}_{E^*}$ is such that $\mathrm{art}_{E^*}(s)=\sigma|E^*$. Then $\sigma(A,i,\eta)\mapsto[a\circ H_1(\alpha)^{=1}\circ\sigma\circ\eta]_K$. But

$$V_f(\alpha)^{=1}\circ\sigma=N_{\Phi^*}(s)=r_{h_\Phi}(s),$$

and so

$$[a\circ H_1(\alpha)^{-1}\circ\sigma\circ\eta]_K=[r_{h_\Phi}(s)\cdot(a\circ\eta)]_K$$

as required. □

NOTES. Our definitions coincide with those of Deligne 1979, except that we have corrected a sign error there (it is necessary to delete "*inverse*" in ibid. 2.2.3, p269, line 10, and in 2.6.3, p284, line 21).

## 13. Uniqueness of canonical models

In this section, I sketch a proof that a Shimura variety has at most one canonical model (up to a unique isomorphism).

**Extension of the base field.**

PROPOSITION 13.1. *Let $k$ be a subfield of an algebraically closed field $\Omega$ of characteristic zero. If $V$ and $W$ are varieties over $k$, then a regular map $V_\Omega \to W_\Omega$ commuting with the actions of $\mathrm{Aut}(\Omega/k)$ on $V(\Omega)$ and $W(\Omega)$ arises from a unique regular map $V \to W$. In other words, the functor*

$$V \mapsto V_\Omega + \textit{action of } \mathrm{Aut}(\Omega/k) \textit{ on } V(\Omega)$$

*is fully faithful.*

PROOF. See AG 16.9. [The first step is to show that the $\Omega^{\mathrm{Aut}(\Omega/k)} = k$, which requires Zorn's lemma in general.] □

COROLLARY 13.2. *A variety $V$ over $k$ is uniquely determined (up to a unique isomorphism) by $V_\Omega$ and the action of $\mathrm{Aut}(\Omega/k)$ on $V(\Omega)$.*

**Uniqueness of canonical models.** Let $(G, X)$ be a Shimura datum.

LEMMA 13.3. *There exists a special point in $X$.*

PROOF (SKETCH). Let $x \in X$, and let $T$ be a maximal torus in $G_\mathbb{R}$ containing $h_x(\mathbb{C})$. Then $T$ is the centralizer of any regular element $\lambda$ of $\mathrm{Lie}(T)$. If $\lambda_0 \in \mathrm{Lie}(G)$ is chosen sufficiently close to $\lambda$, then the centralizer $T_0$ of $\lambda_0$ in $G$ will be a maximal torus in $G$ (Borel 1991, 18.1, 18.2), and $T_0$ will become conjugate[17] to $T$ over $\mathbb{R}$:

$$T_{0\mathbb{R}} = gTg^{-1}, \text{ some } g \in G(\mathbb{R}).$$

Now $h_{gx}(\mathbb{S}) =_{\mathrm{df}} ghg^{-1}(\mathbb{S}) \subset T_{0\mathbb{R}}$, and so $gx$ is special. □

LEMMA 13.4 (Key Lemma). *For any finite extension $L$ of $E(G, X)$ in $\mathbb{C}$, there exists a special point $x_0$ such that $E(x_0)$ is linearly disjoint from $L$.*

PROOF. See Deligne 1971*b*, 5.1. [The basic idea is the same as that of the proof of 13.3 above, but requires the Hilbert irreducibility theorem.] □

If $G = \mathrm{GL}_2$, the lemma just says that, for any finite extension $L$ of $\mathbb{Q}$ in $\mathbb{C}$, there exists a quadratic imaginary extension $E$ over $\mathbb{Q}$ linearly disjoint from $L$. This is obvious — for example, take $E = \mathbb{Q}[\sqrt{-p}]$ for any prime $p$ unramified in $L$.

LEMMA 13.5. *For any $x \in X$, $\{[x, a]_K \mid a \in G(\mathbb{A}_f)\}$ is dense in $\mathrm{Sh}_K(G, X)$ (in the zariski topology).*

PROOF. Write

$$\mathrm{Sh}_K(G, X)(\mathbb{C}) = G(\mathbb{Q})\backslash X \times (G(\mathbb{A}_f)/K)$$

and note that the real approximation theorem (5.4) implies that $G(\mathbb{Q})x$ is dense in $X$ for the complex topology, and, a fortiori, the zariski topology. □

[17] Any element sufficiently close to a regular element will also be regular, which implies that $T_0$ is a maximal torus. Not all maximal tori in $G_{/\mathbb{R}}$ are conjugate — rather, they fall into several connected components, from which the second statement can be deduced.

Let $g \in G(\mathbb{A}_f)$, and let $K$ and $K'$ be compact open subgroups such that $K' \supset g^{-1}Kg$. Then the map $\mathcal{T}(g)$

$$[x,a]_K \mapsto [x,ag]_{K'}\colon \mathrm{Sh}_K(\mathbb{C}) \to \mathrm{Sh}_{K'}(\mathbb{C})$$

is well-defined.

THEOREM 13.6. *If* $\mathrm{Sh}_K(G,X)$ *and* $\mathrm{Sh}_{K'}(G,X)$ *have canonical models over* $E(G,X)$*, then* $\mathcal{T}(g)$ *is defined over* $E(G,X)$*.*

PROOF. After (13.1), it suffices to show that $\sigma(\mathcal{T}(g)) = \mathcal{T}(g)$ for all automorphisms $\sigma$ of $\mathbb{C}$ fixing $E(G,X)$. Let $x_0 \in X$ be special. Then $E(x_0) \supset E(G,X)$ (see 12.3b), and we first show that $\sigma(\mathcal{T}(g)) = \mathcal{T}(g)$ for those $\sigma$'s fixing $E(x_0)$. Choose an $s \in \mathbb{A}_{E_0}^{\times}$ such that $\mathrm{art}(s) = \sigma | E(x_0)^{\mathrm{ab}}$. For $a \in G(\mathbb{A}_f)$,

$$\begin{tikzcd}
{[x_0,a]_K} \arrow[r, mapsto, "\mathcal{T}(g)"] \arrow[d, mapsto, "\sigma"] & {[x_0,ag]_{K'}} \arrow[d, mapsto, "\sigma"] \\
{[x_0,r_{x_0}(s)a]_K} \arrow[r, mapsto, "\mathcal{T}(g)"] & {[x_0,r_{x_0}(s)ag]_{K'}}
\end{tikzcd}$$

commutes. Thus, $\mathcal{T}(g)$ and $\sigma(\mathcal{T}(g))$ agree on $\{[x_0,a] \mid a \in G(\mathbb{A}_f)\}$, and hence on all of $\mathrm{Sh}_K$ by Lemma 13.5. We have shown that $\sigma(\mathcal{T}(g)) = \mathcal{T}(g)$ for all $\sigma$ fixing the reflex field of any special point, but Lemma 13.4 shows that these $\sigma$'s generate $\mathrm{Aut}(\mathbb{C}/E(G,X))$. □

THEOREM 13.7. *(a) A canonical model of* $\mathrm{Sh}_K(G,X)$ *(if it exists) is unique up to a unique isomorphism.*

*(b) If, for all compact open subgroups* $K$ *of* $G(\mathbb{A}_f)$*,* $\mathrm{Sh}_K(G,X)$ *has a canonical model, then so also does* $\mathrm{Sh}(G,X)$*, and it is unique up to a unique isomorphism.*

PROOF. (a) Take $K = K'$ and $g = 1$ in (13.6).

(b) Obvious from (13.6). □

In more detail, let $(M_K(G,X),\varphi)$ and $(M'_K(G,X),\varphi')$ be canonical models of $\mathrm{Sh}_K(G,X)$ over $E(G,X)$. Then the composite

$$M_K(G,X)_{\mathbb{C}} \xrightarrow{\varphi} \mathrm{Sh}_K(G,X) \xrightarrow{\varphi'^{-1}} M'_K(G,X)_{\mathbb{C}}$$

is fixed by all automorphisms of $\mathbb{C}$ fixing $E(G,X)$, and is therefore defined over $E(G,X)$.

REMARK 13.8. In fact, one can prove more. Let $a\colon (G,X) \to (G',X')$ be a morphism of Shimura data, and suppose $\mathrm{Sh}(G,X)$ and $\mathrm{Sh}(G',X')$ have canonical models $M(G,X)$ and $M(G',X')$. Then the morphism $\mathrm{Sh}(a)\colon \mathrm{Sh}(G,X) \to \mathrm{Sh}(G',X')$ is defined over $E(G,X) \cdot E(G',X')$.

**The galois action on the connected components.** A canonical model for $\mathrm{Sh}_K(G,X)$ will define an action of $\mathrm{Aut}(\mathbb{C}/E(G,X))$ on the set $\pi_0(\mathrm{Sh}_K(G,X))$. In the case that $G^{\mathrm{der}}$ is simply connected, we saw in §5 that

$$\pi_0(\mathrm{Sh}_K(G,X)) \cong T(\mathbb{Q})\backslash Y \times T(\mathbb{A}_f)/\nu(K)$$

where $\nu\colon G \to T$ is the quotient of $G$ by $G^{\mathrm{der}}$ and $Y$ is the quotient of $T(\mathbb{R})$ by the image $T(\mathbb{R})^{\dagger}$ of $Z(\mathbb{R})$ in $T(\mathbb{R})$. Let $h = \nu \circ h_x$ for any $x \in X$. Then $\mu_h$ is certainly defined over $E(G,X)$. Therefore, it defines a homomorphism

$$r = r(T,\mu_h)\colon \mathbb{A}_{E(G,X)}^{\times} \to T(\mathbb{A}_{\mathbb{Q}}).$$

The action of $\sigma \in \mathrm{Aut}(\mathbb{C}/E(G,X))$ on $\pi_0(\mathrm{Sh}_K(G,X))$ can be described as follows: let $s \in \mathbb{A}^{\times}_{E(G,X)}$ be such that $\mathrm{art}_{E(G,X)}(s) = \sigma|E(G,X)^{\mathrm{ab}}$, and let $r(s) = (r(s)_\infty, r(s)_f) \in T(\mathbb{R}) \times T(\mathbb{A}_f)$; then

$$\sigma[y,a]_K = [r(s)_\infty y, r(s)_f \cdot a]_K, \text{ for all } y \in Y, \quad a \in T(\mathbb{A}_f). \tag{56}$$

When we use (56) to define the notion a canonical model of a zero-dimensional Shimura variety, we can say that $\pi_0$ of the canonical model of $\mathrm{Sh}_K(G,X)$ is the canonical model of $\mathrm{Sh}(T,Y)$.

If $\sigma$ fixes a special $x_0$ mapping to $y$, then (56) follows from (54), and a slight improvement of (13.4) shows that such $\sigma$'s generate $\mathrm{Aut}(\mathbb{C}/E(G,X))$.

NOTES. The proof of uniqueness follows Deligne 1971*b*, §3, except that I am more unscrupulous in my use of the Zorn's lemma.

## 14. Existence of canonical models

Canonical models are known to exist for all Shimura varieties. In this section, I explain some of the ideas that go into the proof.

**Descent of the base field.** Let $k$ be a subfield of an algebraically closed field $\Omega$ of characteristic zero, and let $\mathcal{A} = \mathrm{Aut}(\Omega/k)$. In (13.1) we observed that the functor

$$\{\text{varieties over } k\} \to \{\text{varieties } V \text{ over } \Omega + \text{action of } \mathcal{A} \text{ on } V(\Omega)\},$$

is fully faithful. In this subsection, we find conditions on a pair $(V, \cdot)$ that ensure that it is in the essential image of the functor, i.e., that it arises from a variety over $k$. We begin by listing two necessary conditions.

*The regularity condition.* Obviously, the action $\cdot$ should recognize that $V(\Omega)$ is not just a set, but rather the set of points of an algebraic variety. Recall that, for $\sigma \in \mathcal{A}$, $\sigma V$ is obtained from $V$ by applying $\sigma$ to the coefficients of the polynomials defining $V$, and $\sigma P \in (\sigma V)(\Omega)$ is obtained from $P \in V(\Omega)$ by applying $\sigma$ to the coordinates of $P$.

DEFINITION 14.1. An action $\cdot$ of $\mathcal{A}$ on $V(\Omega)$ is ***regular*** if the map

$$\sigma P \mapsto \sigma \cdot P \colon (\sigma V)(\Omega) \to V(\Omega)$$

is a regular isomorphism for all $\sigma$.

A priori, this is only a map of sets. The condition requires that it be induced by a regular map $f_\sigma \colon \sigma V \to V$. If $(V, \cdot)$ arises from a variety over $k$, then $\sigma V = V$ and $\sigma P = \sigma \cdot P$, and so the condition is clearly necessary.

REMARK 14.2. (a) When regular, the maps $f_\sigma$ are automatically isomorphisms provided $V$ is nonsingular.

(b) The maps $f_\sigma$ satisfy the cocycle condition $f_\sigma \circ \sigma f_\tau = f_{\sigma\tau}$. Conversely, every family $(f_\sigma)_{\sigma \in \mathcal{A}}$ of regular isomorphisms satisfying the cocycle condition arises from an action of $\mathcal{A}$ satisfying the regularity condition. Such families $(f_\sigma)_{\sigma \in \mathcal{A}}$ are called ***descent data***, and normally one expresses descent theory in terms of them rather than actions of $\mathcal{A}$.

*The continuity condition.*

DEFINITION 14.3. An action $\cdot$ of $\mathcal{A}$ on $V(\Omega)$ is ***continuous*** if there exists a subfield $L$ of $\Omega$ finitely generated over $k$ and a model $V_0$ of $V$ over $L$ such that the action of $\mathrm{Aut}(\Omega/L)$ on $V(\Omega)$ defined by $V_0$ is $\cdot$.

More precisely, the condition requires that there exist a model $(V_0, \varphi)$ of $V$ over $L$ such that $\varphi(\sigma P) = \sigma \cdot \varphi(P)$ for all $P \in V_0(\Omega)$ and $\sigma \in \mathrm{Aut}(\mathbb{C}/L)$. Clearly this condition is necessary.

PROPOSITION 14.4. *A regular action $\cdot$ of $\mathcal{A}$ on $V(\Omega)$ is continuous if there exist points $P_1, \ldots, P_n \in V(\Omega)$ such that*

(a) *the only automorphism of $V$ fixing every $P_i$ is the identity map;*
(b) *there exists a subfield $L$ of $\Omega$ finitely generated over $k$ such that $\sigma \cdot P_i = P_i$ for all $\sigma$ fixing $L$.*

PROOF. Let $(V_0, \varphi)$ be a model of $V$ over a subfield $L$ of $\Omega$ finitely generated over $k$. After possibly enlarging $L$, we may assume that $\varphi^{-1}(P_i) \in V_0(L)$ and that $\sigma \cdot P_i = P_i$ for all $\sigma$ fixing $L$ (because of (b)). For such a $\sigma$, $f_\sigma$ and $\varphi \circ (\sigma\varphi)^{-1}$ are regular maps $\sigma V \to V$ sending $\sigma P_i$ to $P_i$ for each $i$, and so they are equal (because of (a)). Hence

$$\varphi(\sigma P) = f_\sigma((\sigma\varphi)(\sigma P)) = f_\sigma(\sigma(\varphi(P))) = \sigma \cdot \varphi(P)$$

for all $P \in V_0(\Omega)$, and so the action of $\mathrm{Aut}(\mathbb{C}/L)$ on $V(\Omega)$ defined by $(V_0, \varphi)$ is $\cdot$. $\square$

*A sufficient condition for descent.*

THEOREM 14.5. *If $V$ is quasiprojective and $\cdot$ is regular and continuous, then $(V, \cdot)$ arises from a variety over $k$.*

PROOF. This is a restatement of the results of Weil 1956 (see Milne 1999, 1.1). $\square$

COROLLARY 14.6. *The pair $(V, \cdot)$ arises from a variety over $k$ if*

(a) *$V$ is quasiprojective,*
(b) *$\cdot$ is regular, and*
(c) *there exists points $P_1, \ldots, P_n$ in $V(\Omega)$ satisfying the conditions (a) and (b) of (14.4).*

PROOF. Immediate from (14.5) and (14.6). $\square$

For an elementary proof of the corollary, not using the results of Weil 1956, see AG 16.33.

**Review of local systems and families of abelian varieties.** Let $S$ be a topological manifold. A ***local system of $\mathbb{Z}$-modules on*** $S$ is a sheaf $F$ on $S$ that is locally isomorphic to the constant sheaf $\mathbb{Z}^n$ ($n \in \mathbb{N}$).

Let $F$ be a local system of $\mathbb{Z}$-modules on $S$, and let $o \in S$. There is an action of $\pi_1(S, o)$ on $F_o$ that can be described as follows: let $\gamma\colon [0,1] \to S$ be a loop at $o$; because $[0,1]$ is simply connected, there is an isomorphism from $\gamma^*F$ to the constant sheaf defined by a group $M$ say; when we choose such an isomorphism, we obtain isomorphisms $(\gamma^*F)_i \to M$ for all $i \in [0,1]$; now $(\gamma^*F)_i = F_{\gamma(i)}$ and $\gamma(0) = o = \gamma(1)$, and so we get two isomorphisms $F_o \to M$; these isomorphisms differ by an automorphism of $F_o$, which depends only the homotopy class of $\gamma$.

PROPOSITION 14.7. *If $S$ is connected, then $F \mapsto (F_o, \rho_o)$ defines an equivalence from the category of local systems of $\mathbb{Z}$-modules on $S$ to the category of finitely generated $\mathbb{Z}$-modules endowed with an action of $\pi_1(S, o)$.*

PROOF. This is well known; cf. Deligne 1970, I 1. □

Let $F$ be a local system of $\mathbb{Z}$-modules on $S$. Let $\pi\colon \tilde{S} \to S$ be the universal covering space of $S$, and choose a point $o \in \tilde{S}$. We can identify $\pi^*F$ with the constant sheaf defined by $F_{\pi(o)}$. Suppose that we have a hodge structure $h_s$ on $F_s \otimes \mathbb{R}$ for every $s \in S$. We say that $F$, together with the hodge structures, is a ***variation of integral hodge structures on*** $S$ if $s \mapsto h_{\pi(s)}$ (hodge structure on $F_{\pi(o)} \otimes \mathbb{R}$) is a variation of hodge structures on $\tilde{S}$. A ***polarization*** of a variation of hodge structures $(F, (h_s))$ is a pairing $\psi\colon F \times F \to \mathbb{Z}$ such that $\psi_s$ is a polarization of $(F_s, h_s)$ for every $s$.

Let $V$ be a nonsingular algebraic variety over $\mathbb{C}$. A ***family of abelian varieties over*** $V$ is a regular map $f\colon A \to V$ of nonsingular varieties plus a regular multiplication $A \times_V A \to A$ over $V$ such that the fibres of $f$ are abelian varieties of constant dimension (in a different language, $A$ is an abelian scheme over $V$).

THEOREM 14.8. *Let $V$ be a nonsingular variety over $\mathbb{C}$. There is an equivalence $(A, f) \mapsto (R^1 f_*\mathbb{Z})^\vee$ from the category of families of abelian varieties over $V$ to the category of polarizable integral variations of hodge structures of type* $(-1, 0)$, $(0, -1)$ *on $S$.*

This is a generalization of Riemann's theorem (6.8) — see Deligne 1971*a*, 4.4.3.

**The Siegel modular variety.** Let $(V, \psi)$ be a symplectic space over $\mathbb{Q}$, and let $(G, X) = (\mathrm{GSp}(\psi), X(\psi))$ be the associated Shimura datum (§6). We also denote $\mathrm{Sp}(\psi)$ by $S$. We abbreviate $\mathrm{Sh}_K(G, X)$ to $\mathrm{Sh}_K$.

*The reflex field.* Consider the set of pairs $(L, L')$ of complementary lagrangians in $V(\mathbb{C})$:

$$V(\mathbb{C}) = L \oplus L', \quad L, L' \text{ totally isotropic.} \tag{57}$$

Every symplectic basis for $V(\mathbb{C})$ defines such a pair, and every such pair arises from a symplectic basis. Therefore, $G(\mathbb{C})$ (even $S(\mathbb{C})$) acts transitively on the set of pairs $(L, L')$ of complementary lagrangians. For such a pair, let $\mu_{(L,L')}$ be the homomorphism $\mathbb{G}_m \to \mathrm{GL}(V)$ such that $\mu(z)$ acts as $z$ on $L$ and as 1 on $L'$. Then, $\mu_{(L,L')}$ takes values in $G_{\mathbb{C}}$, and as $(L, L')$ runs through the set of pairs of complementary lagrangians in $V(\mathbb{C})$, $\mu_{(L,L')}$ runs through $c(X)$ (notation as on p343). Since $V$ itself has symplectic bases, there exist pairs of complementary lagrangians in $V$. For such a pair, $\mu_{(L,L')}$ is defined over $\mathbb{Q}$, and so $c(X)$ has a representative defined over $\mathbb{Q}$. This shows that the reflex field $E(G, X) = \mathbb{Q}$.

*The special points.* Let $K$ be a compact open subgroup of $G(\mathbb{A}_f)$, and, as in §6, let $\mathcal{M}_K$ be the set of triples $(A, s, \eta K)$ in which $A$ is an abelian variety over $\mathbb{C}$, $s$ is an alternating form on $H_1(A, \mathbb{Q})$ such that $\pm s$ is a polarization, and $\eta$ is an isomorphism $V(\mathbb{A}_f) \to V_f(A)$ sending $\psi$ to a multiple of $s$. Recall (6.11) that there is a natural map $\mathcal{M}_K \to \mathrm{Sh}_K(\mathbb{C})$ whose fibres are the isomorphism classes.

In this subsubsection we answer the question: which triples $(A, s, \eta K)$ correspond to points $[x, a]$ with $x$ special?

DEFINITION 14.9. A ***CM-algebra*** is a finite product of CM-fields. An abelian variety $A$ over $\mathbb{C}$ is ***CM*** if there exists a CM-algebra $E$ and a homomorphism $E \to \operatorname{End}^0(A)$ such that $H_1(A, \mathbb{Q})$ is a free $E$-module of rank 1.

Let $E \to \operatorname{End}^0(A)$ be as in the definition, and let $E$ be a product of CM-fields $E_1, \ldots, E_m$. Then $A$ is isogenous to a product of abelian varieties $A_1 \times \cdots \times A_m$ with $A_i$ of CM-type $(E_i, \Phi_i)$ for some $\Phi_i$.

Recall that, for an abelian variety $A$ over $\mathbb{C}$, there is a homomorphism $h_A \colon \mathbb{C}^\times \to \mathrm{GL}(H_1(A, \mathbb{R}))$ describing the natural complex structure on $H_1(A, \mathbb{R})$ (see §6).

PROPOSITION 14.10. *An abelian variety $A$ over $\mathbb{C}$ is CM if and only if there exists a torus $T \subset \mathrm{GL}(H_1(A, \mathbb{Q}))$ such that $h_A(\mathbb{C}^\times) \subset T(\mathbb{R})$.*

PROOF. See Mumford 1969, §2, or Deligne 1982, §3. □

COROLLARY 14.11. *If $(A, s, \eta K) \mapsto [x, a]_K$ under $\mathcal{M}_K \to \mathrm{Sh}_K(G, X)$, then $A$ is of CM-type if and only if $x$ is special.*

PROOF. Recall that if $(A, s, \eta K) \mapsto [x, a]_K$, then there exists an isomorphism $H_1(A, \mathbb{Q}) \to V$ sending $h_A$ to $h_x$. Thus, the statement follows from the proposition. □

*A criterion to be canonical.* We now define an action of $\mathrm{Aut}(\mathbb{C})$ on $\mathcal{M}_K$. Let $(A, s, \eta K) \in \mathcal{M}_K$. Then $s \in H^2(A, \mathbb{Q})$ is a hodge tensor, and therefore equals $r[D]$ for some $r \in \mathbb{Q}^\times$ and divisor $D$ on $A$ (see 7.5). We let ${}^\sigma s = r[\sigma D]$. The condition that $\pm s$ be positive definite is equivalent to an algebro-geometric condition on $D$ (Mumford 1970, pp29–30) which is preserved by $\sigma$. Therefore, $\pm {}^\sigma s$ is a polarization for $H_1(A, \mathbb{Q})$. We define $\sigma(A, s, \eta K)$ to be $(\sigma A, {}^\sigma s, {}^\sigma \eta K)$ with ${}^\sigma \eta$ as in (55).

PROPOSITION 14.12. *Suppose that* $\mathrm{Sh}_K$ *has a model* $M_K$ *over* $\mathbb{Q}$ *for which the map*

$$\mathcal{M}_K \to M_K(\mathbb{C})$$

*commutes with the actions of* $\mathrm{Aut}(\mathbb{C})$. *Then* $M_K$ *is canonical.*

PROOF. For a special point $[x, a]_K$ corresponding to an abelian variety $A$ with complex multiplication by a field $E$, the condition (54) is an immediate consequence of the main theorem of complex multiplication (cf. 12.11). For more general special points, it also follows from the main theorem of complex multiplication, but not quite so immediately. □

*Outline of the proof of the existence of a canonical model.* Since the action of $\mathrm{Aut}(\mathbb{C})$ on $\mathcal{M}_K$ preserves the isomorphism classes, from the map $\mathcal{M}_K \to \mathrm{Sh}_K(\mathbb{C})$, we get an action of $\mathrm{Aut}(\mathbb{C})$ on $\mathrm{Sh}_K(\mathbb{C})$. If this action satisfies the conditions of hypotheses of Corollary 14.6, then $\mathrm{Sh}_K(G, X)$ has a model over $\mathbb{Q}$, which Proposition 14.12 will show to be canonical.

Condition (a) of (14.6). We know that $\mathrm{Sh}_K(G, X)$ is quasi-projective from (3.12).

Condition (b) of (14.6). We have to show that the map

$$\sigma P \mapsto \sigma \cdot P \colon \sigma\, \mathrm{Sh}_K(\mathbb{C}) \xrightarrow{f_\sigma} \mathrm{Sh}_K(\mathbb{C})$$

is regular. It suffices to do this for $K$ small, because if $K' \supset K$, then $\mathrm{Sh}_{K'}(G, X)$ is a quotient of $\mathrm{Sh}_K(G, X)$.

Recall (5.17) that $\pi_0(\mathrm{Sh}_K) \cong \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$. Let $\varepsilon \in \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$, and let $\mathrm{Sh}_K^\varepsilon$ be the corresponding connected component of $\mathrm{Sh}_K$. Then $\mathrm{Sh}_K^\varepsilon = \Gamma_\varepsilon\backslash X^+$ where $\Gamma_\varepsilon = G(\mathbb{Q}) \cap K_\varepsilon$ for some conjugate $K_\varepsilon$ of $K$ (see 5.17, 5.23)

Let $(A, s, \eta K) \in \mathcal{M}_K$ and choose an isomorphism $a\colon H_1(A, \mathbb{Q}) \to V$ sending $s$ to a multiple of $\psi$. Then the image of $(A, s, \eta K)$ in $\mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$ is represented by $\nu(a \circ \eta)$ where $a \circ \eta\colon V(\mathbb{A}_f) \to V(\mathbb{A}_f)$ is to be regarded as an element of $G(\mathbb{A}_f)$. Write $\mathcal{M}_K^\varepsilon$ for the set of triples with $\nu(a \circ \eta) \in \varepsilon$. Define $\mathcal{H}_K^\varepsilon$ similarly.

The map $\mathcal{M}_K \to \mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$ is equivariant for the action of $\mathrm{Aut}(\mathbb{C})$ when we let $\mathrm{Aut}(\mathbb{C})$ act on $\mathbb{Q}_{>0}\backslash\mathbb{A}_f^\times/\nu(K)$ through the cyclotomic character, i.e.,

$$\sigma[\alpha] = [\chi(\sigma)\alpha] \text{ where } \chi(\sigma) \in \hat{\mathbb{Z}}^\times,\ \zeta^{\chi(\sigma)} = \sigma\zeta,\ \zeta \text{ a root of } 1.$$

Write $X^+(\Gamma_\varepsilon)$ for $\Gamma_\varepsilon\backslash X^+$ regarded as an algebraic variety, and let $\sigma(X^+(\Gamma_\varepsilon))$ be the algebraic variety obtained from $X^+(\Gamma_\varepsilon)$ by change of base field $\sigma\colon \mathbb{C} \to \mathbb{C}$. Consider the diagram:

$$\begin{array}{ccc}
X^+ & \xleftarrow{\alpha} & U \\
\downarrow & & \downarrow \\
X^+(\Gamma_{\sigma\varepsilon}) & \xleftarrow{f_\sigma} & \sigma(X^+(\Gamma_\varepsilon)) \\
\mathcal{M}_K^{\sigma\varepsilon} & \xleftarrow{\sigma} & \mathcal{M}_K^\varepsilon
\end{array}$$

The map $\sigma$ sends $(A, \ldots)$ to $\sigma(A, \ldots)$, and the map $f_\sigma$ is the map of sets $\sigma P \mapsto \sigma \cdot P$. The two maps are compatible. The map $U \to \sigma(X^+(\Gamma_\varepsilon))$ is the universal covering space of the complex manifold $(\sigma(X^+(\Gamma_\varepsilon)))^{\mathrm{an}}$.

Fix a lattice $\Lambda$ in $V$ that is stable under the action of $\Gamma_\varepsilon$. From the action of $\Gamma_\varepsilon$ on $\Lambda$, we get a local system of $\mathbb{Z}$-modules $M$ on $X^+(\Gamma_\varepsilon)$ (see 14.7), which, in fact, is a polarized integral variation of hodge structures $F$. According to Theorem 14.8, this variation of hodge structures arises from a polarized family of abelian varieties $f\colon \mathcal{A} \to X^+(\Gamma_\varepsilon)$. As $f$ is a regular map of algebraic varieties, we can apply $\sigma$ to it, and obtain a polarized family of abelian varieties $\sigma f\colon \sigma\mathcal{A} \to \sigma(X^+(\Gamma_\varepsilon))$. Then $(R^1(\sigma f)_*\mathbb{Z})^\vee$ is a polarized integral hodge structure on $\sigma(X^+(\Gamma_\varepsilon))$. On pulling this back to $U$ and tensoring with $\mathbb{Q}$, we obtain a variation of polarized rational hodge structures over the space $U$, whose underlying local system can identified with the constant sheaf defined by $V$. When this identification is done correctly, each $u \in U$ defines a complex structure on $V$ that is positive for $\psi$, i.e., a point $x$ of $X^+$, and the map $u \mapsto x$ makes the diagram commute. Now (2.15) shows that $u \mapsto x$ is holomorphic. It follows that $f_\sigma$ is holomorphic, and Borel's theorem (3.14) shows that it is regular.

Condition (c) of (14.6). For any $x \in X$, the set $\{[x, a]_K \mid a \in G(\mathbb{A}_f)\}$ has the property that only the identity automorphism of $\mathrm{Sh}_K(G, X)$ fixes its elements (see 13.5). But, there are only finitely many automorphisms of $\mathrm{Sh}_K(G, X)$ (see 3.21), and so a finite sequence of points $[x, a_1], \ldots, [x, a_n]$ will have this property. When we choose $x$ to be special, the main theorem of complex multiplication (11.2) tells us that $\sigma \cdot [x, a_i] = [x, a_i]$ for all $\sigma$ fixing some fixed finite extension of $E(x)$, and so condition (c) holds for these points.

**Simple PEL Shimura varieties of type A or C.** The proof is similar to the Siegel case. Here $\mathrm{Sh}_K(G, X)$ classifies quadruples $(A, i, s, \eta K)$ satisfying certain conditions. One checks that if $\sigma$ fixes the reflex field $E(G, X)$, then $\sigma(A, i, s, \eta K)$

lies in the family again (see 12.7). Again the special points correspond to CM abelian varieties, and the Shimura-Taniyama theorem shows that, if $\mathrm{Sh}_K(G,X)$ has a model $M_K$ over $E(G,X)$ for which the action of $\mathrm{Aut}(\mathbb{C}/E(G,X))$ on $M_K(\mathbb{C}) = \mathrm{Sh}_K(G,X)(\mathbb{C})$ agrees with its action on the quadruples, then it is canonical.

**Shimura varieties of hodge type.** In this case, $\mathrm{Sh}_K(G,X)$ classifies isomorphism classes of triples $(A,(s_i)_{0\leq i\leq n},\eta K)$ where the $s_i$ are hodge tensors. A proof similar to that in the Siegel case will apply once we have defined ${}^{\sigma}s$ for $s$ a hodge tensor on an abelian variety.

If the Hodge conjecture is true, then $s$ is the cohomology class of some algebraic cycle $Z$ on $A$ (i.e., formal $\mathbb{Q}$-linear combination of integral subvarieties of $A$). Then we could define ${}^{\sigma}s$ to be the cohomology class of $\sigma Z$ on $\sigma A$. Unfortunately, a proof of the Hodge conjecture seems remote, even for abelian varieties. Deligne succeeded in defining ${}^{\sigma}s$ without the Hodge conjecture. It is important to note that there is no natural map between $H^n(A,\mathbb{Q})$ and $H^n(\sigma A,\mathbb{Q})$ (unless $\sigma$ is continuous, and hence is the identity or complex conjugation). However, there is a natural isomorphism $\sigma\colon H^n(A,\mathbb{A}_f)\to H^n(\sigma A,\mathbb{A}_f)$ coming from the identification

$$H^n(A,\mathbb{A}_f)\cong \mathrm{Hom}(\bigwedge^n\Lambda,\mathbb{A}_f)\cong \mathrm{Hom}(\bigwedge^n(\Lambda\otimes\mathbb{A}_f),\mathbb{A}_f)\cong \mathrm{Hom}(\bigwedge^n V_fA,\mathbb{A}_f)$$

(or, equivalently, from identifying $H^n(A,\mathbb{A}_f)$ with étale cohomology).

THEOREM 14.13. *Let $s$ be a hodge tensor on an abelian variety $A$ over $\mathbb{C}$, and let $s_{\mathbb{A}_f}$ be the image of $s$ the $\mathbb{A}_f$-cohomology. For any automorphism $\sigma$ of $\mathbb{C}$, there exists a hodge tensor ${}^{\sigma}s$ on $\sigma A$ (necessarily unique) such that $({}^{\sigma}s)_{\mathbb{A}_f}=\sigma(s_{\mathbb{A}_f})$.*

PROOF. This is the main theorem of Deligne 1982. [Interestingly, the theory of locally symmetric varieties is used in the proof.] □

As an alternative to using Deligne's theorem, one can apply the following result (note, however, that the above approach has the advantage of giving a description of the points of the canonical model with coordinates in any field containing the reflex field).

PROPOSITION 14.14. *Let $(G,X)\hookrightarrow(G',X')$ be an inclusion of Shimura data; if $\mathrm{Sh}(G',X')$ has canonical model, so also does $\mathrm{Sh}(G,X)$.*

PROOF. This follows easily from 5.16. □

**Shimura varieties of abelian type.** Deligne (1979, 2.7.10) defines the notion of a canonical model of a *connected* Shimura variety $\mathrm{Sh}^{\circ}(G,X)$. This is an inverse system of connected varieties over $\mathbb{Q}^{\mathrm{al}}$ endowed with the action of a large group (a mixture of a galois group and an adèlic group). A key result is the following.

THEOREM 14.15. *Let $(G,X)$ be a Shimura datum and let $X^+$ be a connected component of $X$. Then $\mathrm{Sh}(G,X)$ has a canonical model if and only if $\mathrm{Sh}^{\circ}(G^{\mathrm{der}},X^+)$ has a canonical model.*

PROOF. See Deligne 1979, 2.7.13. □

Thus, for example, if $(G_1,X_1)$ and $(G_2,X_2)$ are Shimura data such that $(G_1^{\mathrm{der}},X_1^+)\approx(G_2^{\mathrm{der}},X_2^+)$, and one of $\mathrm{Sh}(G_1,X_1)$ or $\mathrm{Sh}(G_2,X_2)$ has a canonical model, then they both do.

The next result is more obvious (ibid. 2.7.11).

Proposition 14.16. *(a) Let* $(G_i, X_i)$ $(1 \leq i \leq m)$ *be connected Shimura data. If each connected Shimura variety* $\mathrm{Sh}^\circ(G_i, X_i)$ *has a canonical model* $M^\circ(G_i, X_i)$*, then* $\prod_i M^\circ(G_i, X_i)$ *is a canonical model for* $\mathrm{Sh}^\circ(\prod_i G_i, \prod_i X_i)$*.*

*(b) Let* $(G_1, X_1) \to (G_2, X_2)$ *be an isogeny of connected Shimura data. If* $\mathrm{Sh}^\circ(G_1, X_1)$ *has a canonical model, then so also does* $\mathrm{Sh}^\circ(G_2, X_2)$*.*

More precisely, in case (b) of the theorem, let $G^{\mathrm{ad}}(\mathbb{Q})_1^+$ and $G^{\mathrm{ad}}(\mathbb{Q})_2^+$ be the completions of $G^{\mathrm{ad}}(\mathbb{Q})^+$ for the topologies defined by the images of congruence subgroups in $G_1(\mathbb{Q})^+$ and $G_2(\mathbb{Q})^+$ respectively; then the canonical model for $\mathrm{Sh}^\circ(G_2, X_2)$ is the quotient of the canonical model for $\mathrm{Sh}^\circ(G_2, X_2)$ by the kernel of $G^{\mathrm{ad}}(\mathbb{Q})_1^+ \to G^{\mathrm{ad}}(\mathbb{Q})_2^+$.

We can now prove the existence of canonical models for all Shimura varieties of abelian type. For a connected Shimura variety of primitive type, the existence follows from (14.15) and the existence of canonical models for Shimura varieties of hodge type (see above). Now (14.16) proves the existence for all connected Shimura varieties of abelian type, and (14.16) proves the existence for all Shimura varieties of abelian type.

Remark 14.17. The above proof is only an existence proof: it gives little information about the canonical model. For the Shimura varieties it treats, Theorem 9.4 can be used to construct canonical models and give a description of the points of the canonical model in any field containing the reflex field.

**General Shimura varieties.** There is an approach that proves the existence of canonical models for all Shimura varieties, and is largely independent of that discussed above except that it assumes the existence[18] of canonical models for Shimura varieties of type $A_1$ (and it uses (14.15) and (14.16)).

The essential idea is the following. Let $(G, X)$ be a connected Shimura datum with $G$ the group over $\mathbb{Q}$ obtained from a simple group $H$ over a totally real field $F$ by restriction of scalars.

Assume first that $H$ splits over a CM-field of degree 2 over $F$. Then there exist many homomorphisms $H_i \to H$ from groups of type $A_1$ into $H$. From this, we get many inclusions

$$\mathrm{Sh}^\circ(G_i, X_i) \hookrightarrow \mathrm{Sh}^\circ(G, X)$$

where $G_i$ is the restriction of scalars of $H_i$. From this, and the existence of canonical models for the $\mathrm{Sh}^\circ(G_i, X_i)$, it is possible to prove the existence of the canonical model for $\mathrm{Sh}^\circ(G, X)$.

In the general case, there will be a totally real field $F'$ containing $F$ and such that $H_{F'}$ splits over a CM-field of degree 2 over $F$. Let $G_*$ be the restriction of scalars of $H_{F'}$. Then there is an inclusion $(G, X) \hookrightarrow (G_*, X_*)$ of connected Shimura data, and the existence of a canonical model for $\mathrm{Sh}^\circ(G_*, X_*)$ implies the existence of a canonical model for $\mathrm{Sh}^\circ(G, X)$ (cf. 14.14).

For the details, see Borovoi 1984, 1987 and Milne 1983.

**Final remark: rigidity.** One might expect that if one modified the condition (54), for example, by replacing $r_x(s)$ with $r_x(s)^{-1}$, then one would arrive at a modified notion of canonical model, and the same theorems would hold. This is not true: the condition (54) is the *only* one for which canonical models can exist.

[18] In fact, the approach assumes a stronger statement for Shimura varieties of type $A_1$, namely, Langlands's conjugation conjecture, and it proves Langlands's conjecture for all Shimura varieties.

In fact, if $G$ is adjoint, then the Shimura variety $\mathrm{Sh}(G,X)$ has no automorphisms commuting with the action of $G(\mathbb{A}_f)$ (Milne 1983, 2.7), from which it follows that the canonical model is the *only* model of $\mathrm{Sh}(G,X)$ over $E(G,X)$, and we know that for the canonical model the reciprocity law at the special points is given by (54).

NOTES. The concept of a canonical model characterized by reciprocity laws at special points is due to Shimura, and the existence of such models was proved for major families by Shimura, Miyake, and Shih. Shimura recognized that to have a canonical model it is necessary to have a reductive group, but for him the semisimple group was paramount: in our language, given a connected Shimura datum $(H,Y)$, he asked for Shimura datum $(G,X)$ such that $(G^{\mathrm{der}},X^{+})=(H,Y)$ and $\mathrm{Sh}(G,X)$ has a canonical model (see his talk at the 1970 International Congress Shimura 1971). In his Bourbaki report on Shimura's work (1971b), Deligne placed the emphasis on reductive groups, thereby enlarging the scope of the field.

## 15. Abelian varieties over finite fields

For each Shimura datum $(G,X)$, we now have a canonical model $\mathrm{Sh}(G,X)$ of the Shimura variety over its reflex field $E(G,X)$. In order, for example, to understand the zeta function of the Shimura variety or the galois representations occurring in its cohomology, we need to understand the points on the canonical model when we reduce it modulo a prime of $E(G,X)$. After everything we have discussed, it would be natural to do this in terms of abelian varieties (or motives) over the finite field plus additional structure. However, such a description will not be immediately useful — what we want is something more combinatorial, which can be plugged into the trace formula. The idea of Langlands and Rapoport (1987) is to give an elementary definition of a category of "fake" abelian varieties (better, abelian motives) over the algebraic closure of a finite field that looks just like the true category, and to describe the points in terms of it. In this section, I explain how to define such a category.

**Semisimple categories.** An object of an abelian category $\mathsf{M}$ is ***simple*** if it has no proper nonzero subobjects. Let $F$ be a field. By an ***$F$-category***, I mean an additive category in which the Hom-sets $\mathrm{Hom}(x,y)$ are finite dimensional $F$-vector spaces and composition is $F$-bilinear. An $F$-category $\mathsf{M}$ is said to be ***semisimple*** if it is abelian and every object is a direct sum (necessarily finite) of simple objects.

If $e$ is simple, then a nonzero morphism $e\to e$ is an isomorphism. Therefore, $\mathrm{End}(e)$ is a division algebra over $F$. Moreover, $\mathrm{End}(re)\cong M_r(\mathrm{End}(e))$. Here $re$ denotes the direct sum of $r$ copies of $e$. If $e'$ is a second simple object, then either $e\approx e'$ or $\mathrm{Hom}(e,e')=0$. Therefore, if $x=\sum r_ie_i$ $(r_i\geq 0)$ and $y=\sum s_ie_i$ $(s_i\geq 0)$ are two objects of $\mathsf{M}$ expressed as sums of copies of simple objects $e_i$ with $e_i\not\approx e_j$ for $i\neq j$, then

$$\mathrm{Hom}(x,y)=\prod M_{s_i,r_i}(\mathrm{End}(e_i)).$$

Thus, the category $\mathsf{M}$ is described up to equivalence by:

(a) the set $\Sigma(\mathsf{M})$ of isomorphism classes of simple objects in $\mathsf{M}$;
(b) for each $\sigma\in\Sigma$, the isomorphism class $[D_\sigma]$ of the endomorphism algebra $D_\sigma$ of a representative of $\sigma$.

We call $(\Sigma(\mathsf{M}),([D_\sigma])_{\sigma\in\Sigma(\mathsf{M})})$ the ***numerical invariants*** of $\mathsf{M}$.

**Division algebras; the Brauer group.** We shall need to understand what the set of isomorphism classes of division algebras over a field $F$ look like.

Recall the definitions: by an $F$-***algebra***, we mean a ring $A$ containing $F$ in its centre and finite dimensional as $F$-vector space; if $F$ equals the centre of $A$, then $A$ is called a ***central*** $F$-algebra; a ***division algebra*** is an algebra in which every nonzero element has an inverse; an $F$-algebra $A$ is ***simple*** if it contains no two-sided ideals other than 0 and $A$. By a theorem of Wedderburn, the simple $F$-algebras are the matrix algebras over division $F$-algebras.

EXAMPLE 15.1. (a) If $F$ is algebraically closed or finite, then every central division algebra is isomorphic to $F$.

(b) Every central division algebra over $\mathbb{R}$ is isomorphic either to $\mathbb{R}$ or to the (usual) quaternion algebra:

$$\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j, \quad j^2 = -1, \quad jzj^{-1} = \overline{z} \quad (z \in \mathbb{C}).$$

(c) Let $F$ be a $p$-adic field (finite extension of $\mathbb{Q}_p$), and let $\pi$ be a prime element of $\mathcal{O}_F$. Let $L$ be an unramified extension field of $F$ of degree $n$, and let $\sigma$ denote the Frobenius generator of $\mathrm{Gal}(L/F)$ — $\sigma$ acts as $x \mapsto x^p$ on the residue field. For each $i$, $1 \leq i \leq n$, define

$$D_{i,n} = L \oplus La \oplus \cdots \oplus La^{n-1}, \quad a^n = \pi^i, \quad aza^{-1} = \sigma(z) \quad (z \in L).$$

Then $D_{i,n}$ is a central simple algebra over $F$, which is a division algebra if and only if $\gcd(i,n) = 1$. Every central division algebra over $F$ is isomorphic to $D_{i,n}$ for exactly one relatively prime pair $(i,n)$ (CFT, IV 4.2).

If $B$ and $B'$ are central simple $F$-algebras, then so also is $B \otimes_F B'$ (CFT, 2.8). If $D$ and $D'$ are central division algebras, then Wedderburn's theorem shows that $D \otimes_F D' \approx M_r(D'')$ for some $r$ and some central division algebra $D''$ well-defined up to isomorphism, and so we can set

$$[D][D'] = [D''].$$

This law of composition is obviously, and $[F]$ is an identity element. Let $D^{\mathrm{opp}}$ denote the opposite algebra to $D$ (the same algebra but with the multiplication reversed: $a^{\mathrm{opp}}b^{\mathrm{opp}} = (ba)^{\mathrm{opp}}$). Then (CFT, IV 2.9)

$$D \otimes_F D^{\mathrm{opp}} \cong \mathrm{End}_{F\text{-linear}}(D) \approx M_r(F),$$

and so $[D][D^{\mathrm{opp}}] = [F]$. Therefore, the isomorphism classes of central division algebras over $F$ (equivalently, the isomorphism classes of central simple algebras over $F$) form a group, called the ***Brauer group*** of $F$.

EXAMPLE 15.2. (a) The Brauer group of an algebraically closed field or a finite field is zero.

(b) The Brauer group $\mathbb{R}$ has order two: $\mathrm{Br}(\mathbb{R}) \cong \frac{1}{2}\mathbb{Z}/\mathbb{Z}$.

(c) For a $p$-adic field $F$, the map $[D_{n,i}] \mapsto \frac{i}{n} \mod \mathbb{Z}$ is an isomorphism $\mathrm{Br}(F) \cong \mathbb{Q}/\mathbb{Z}$.

(d) For a number field $F$ and a prime $v$, write $\mathrm{inv}_v$ for the canonical homomorphism $\mathrm{Br}(F_v) \to \mathbb{Q}/\mathbb{Z}$ given by (a,b,c) (so $\mathrm{inv}_v$ is an isomorphism except when $v$ is real or complex, in which case it has image $\frac{1}{2}\mathbb{Z}/\mathbb{Z}$ or 0). For a central simple algebra $B$ over $F$, $[B \otimes_F F_v] = 0$ for almost all $v$, and the sequence

$$0 \longrightarrow \mathrm{Br}(F) \xrightarrow{[B]\mapsto[B\otimes_F F_v]} \oplus \mathrm{Br}(F_v) \xrightarrow{\sum \mathrm{inv}_v} \mathbb{Q}/\mathbb{Z} \longrightarrow 0.$$

is exact.

Statement (d) is shown in the course of proving the main theorem of class field theory by the cohomological approach (CFT, VIII 2.2). It says that to give a division algebra over $F$ (up to isomorphism) is the same as to give a family $(i_v) \in \bigoplus_{v \text{ finite}} \mathbb{Q}/\mathbb{Z} \oplus \bigoplus_{v \text{ real}} \frac{1}{2}\mathbb{Z}/\mathbb{Z}$ such that $\sum i_v = 0$.

The key tool in computing Brauer groups is an isomorphism

$$\mathrm{Br}(F) \cong H^2(F, \mathbb{G}_m) \stackrel{\text{df}}{=} H^2(\mathrm{Gal}(F^{\mathrm{al}}/F), F^{\mathrm{al}\times}) \stackrel{\text{df}}{=} \varinjlim H^2(\mathrm{Gal}(L/F), L^\times).$$

The last limit is over the fields $L \subset F^{\mathrm{al}}$ of finite degree and galois over $\mathbb{Q}$. This isomorphism can be most elegantly defined as follows. Let $D$ be a central simple division of degree $n^2$ over $F$, and assume that $D$ contains a subfield $L$ of degree $n$ over $F$ and galois over $F$. Then each $\beta \in D$ normalizing $L$ defines an element $x \mapsto \beta x \beta^{-1}$ of $\mathrm{Gal}(L/F)$, and the Noether-Skolem theorem (CFT, IV 2.10) shows that every element of $\mathrm{Gal}(L/F)$ arises in this way. Because $L$ is its own centralizer (ibid., 3.4), the sequence

$$1 \to L^\times \to N(L) \to \mathrm{Gal}(L/F) \to 1$$

is exact. For each $\sigma \in \mathrm{Gal}(L/F)$, choose an $s_\sigma \in N(L)$ mapping to $\sigma$, and let

$$s_\sigma \cdot s_\tau = d_{\sigma,\tau} \cdot s_{\sigma\tau}, \quad d_{\sigma,\tau} \in L^\times.$$

Then $(d_{\sigma,\tau})$ is a 2-cocycle whose cohomology class is independent of the choice of the family $(s_\sigma)$. Its class in $H^2(\mathrm{Gal}(L/F), L^\times) \subset H^2(F, \mathbb{G}_m)$ is the cohomology class of $[D]$.

EXAMPLE 15.3. Let $L$ be the completion of $\mathbb{Q}_p^{\mathrm{un}}$ (equal to the field of fractions of the ring of Witt vectors with coefficients in $\mathbb{F}$), and let $\sigma$ be the automorphism of $L$ inducing $x \mapsto x^p$ on its residue field. An ***isocrystal*** is a finite dimensional $L$-vector space $V$ equipped with a $\sigma$-linear isomorphism $F\colon V \to V$. The category $\mathsf{Isoc}$ of isocrystals is a semisimple $\mathbb{Q}_p$-linear category with $\Sigma(\mathsf{Isoc}) = \mathbb{Q}$, and the endomorphism algebra of a representative of the isomorphism class $\lambda$ is a division algebra over $\mathbb{Q}_p$ with invariant $\lambda$. If $\lambda \geq 0$, $\lambda = r/s$, $\gcd(r,s) = 1$, $s > 0$, then $E^\lambda$ can be taken to be $(\mathbb{Q}_p/(T^r - p^s)) \otimes_{\mathbb{Q}_p} L$, and if $\lambda < 0$, $E^\lambda$ can be taken to be the dual of $E^{-\lambda}$. See Demazure 1972, Chap. IV.

**Abelian varieties.** Recall (p334) that $\mathsf{AV}^0(k)$ is the category whose objects are the abelian varieties over $k$, but whose homs are $\mathrm{Hom}^0(A,B) = \mathrm{Hom}(A,B) \otimes \mathbb{Q}$. It follows from results of Weil that $\mathsf{AV}^0(k)$ is a semisimple $\mathbb{Q}$-category with the simple abelian varieties (see p334) as its simple objects. Amazingly, when $k$ is finite, we know its numerical invariants.

*Abelian varieties over* $\mathbb{F}_q$, $q = p^n$. Recall that a Weil $q$-integer is an algebraic integer such that, for every embedding $\rho\colon \mathbb{Q}[\pi] \to \mathbb{C}$, $|\rho\pi| = q^{\frac{1}{2}}$. Two Weil $q$-integers $\pi$ and $\pi'$ are ***conjugate*** if there exists an isomorphism $\mathbb{Q}[\pi] \to \mathbb{Q}[\pi']$ sending $\pi$ to $\pi'$.

THEOREM 15.4 (Honda-Tate). *The map $A \mapsto \pi_A$ defines a bijection from $\Sigma(\mathsf{AV}(\mathbb{F}_q))$ to the set of conjugacy classes of Weil $q$-integers. For any simple $A$, the centre of $D =_{df} \mathrm{End}^0(A)$ is $F = \mathbb{Q}[\pi_A]$, and for a prime $v$ of $F$,*

$$\mathrm{inv}_v(D) = \begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ \frac{\mathrm{ord}_v(\pi_A)}{\mathrm{ord}_v(q)}[F_v\colon \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

*Moreover,* $2\dim A = [D\colon F]^{\frac{1}{2}} \cdot [F\colon \mathbb{Q}]$.

In fact, $\mathbb{Q}[\pi]$ can only have a real prime if $\pi = \sqrt{p^n}$. Let $W_1(q)$ be the set of Weil $q$-integers in $\mathbb{Q}^{\mathrm{al}} \subset \mathbb{C}$. Then the theorem gives a bijection

$$\Sigma(\mathsf{AV}^0(\mathbb{F}_q)) \to \Gamma\backslash W_1(q), \quad \Gamma = \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}).$$

NOTES. Except for the statement that every $\pi_A$ arises from an $A$, the theorem is due to Tate. That every Weil $q$-integer arises from an abelian variety was proved (using 10.10) by Honda. See Tate 1969 for a discussion of the theorem.

*Abelian varieties over* $\mathbb{F}$. We shall need a similar result for an algebraic closure $\mathbb{F}$ of $\mathbb{F}_p$.

If $\pi$ is a Weil $p^n$-integer, then $\pi^m$ is a Weil $p^{mn}$-integer, and so we have a homomorphism $\pi \mapsto \pi^m\colon W_1(p^n) \to W_1(p^{nm})$. Define

$$W_1 = \varinjlim W_1(p^n).$$

If $\pi \in W_1$ is represented by $\pi_n \in W_1(p^n)$, then $\pi_n^m \in W_1(p^{nm})$ also represents $\pi$, and $\mathbb{Q}[\pi_n] \supset \mathbb{Q}[\pi_n^m]$. Define $\mathbb{Q}\{\pi\}$ to be the field of smallest degree over $\mathbb{Q}$ generated by a representative of $\pi$.

Every abelian variety over $\mathbb{F}$ has a model defined over a finite field, and if two abelian varieties over a finite field become isomorphic over $\mathbb{F}$, then they are isomorphic already over a finite field. Let $A$ be an abelian variety over $\mathbb{F}_q$. When we regard $A$ as an abelian variety over $\mathbb{F}_{q^m}$, then the Frobenius map is raised to the $m^{\mathrm{th}}$-power (obviously): $\pi_{A_{\mathbb{F}_{q^m}}} = \pi_A^m$.

Let $A$ be an abelian variety defined over $\mathbb{F}$, and let $A_0$ be a model of $A$ over $\mathbb{F}_q$. The above remarks show that $s_A(v) =_{\mathrm{df}} \frac{\mathrm{ord}_v(\pi_{A_0})}{\mathrm{ord}_v(q)}$ is independent of the choice of $A_0$. Moreover, for any $\rho\colon \mathbb{Q}[\pi_{A_0}] \hookrightarrow \mathbb{Q}^{\mathrm{al}}$, the $\Gamma$-orbit of the element $\pi_A$ of $W_1$ represented by $\rho\pi_{A_0}$ depends only on $A$.

THEOREM 15.5. *The map $A \mapsto \Gamma\pi_A$ defines a bijection $\Sigma(\mathsf{AV}^0(\mathbb{F})) \to \Gamma\backslash W_1$. For any simple $A$, the centre of $D =_{df} \mathrm{End}^0(A)$ is isomorphic to $F = \mathbb{Q}\{\pi_A\}$, and for any prime $v$ of $F$,*

$$\mathrm{inv}_v(D) = \begin{cases} \frac{1}{2} & \text{if } v \text{ is real} \\ s_A(v) \cdot [F_v\colon \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.} \end{cases}$$

PROOF. This follows from the Honda-Tate theorem and the above discussion. □

Our goal in the remainder of this section is to give an elementary construction of a semisimple $\mathbb{Q}$-category that contains, in a natural way, a category of "fake abelian varieties over $\mathbb{F}$" with the same numerical invariants as $\mathsf{AV}^0(\mathbb{F})$.

For the remainder of this section $F$ is a field of characteristic zero.

**Tori and their representations.** Let $T$ be a torus over $F$ split by a galois extension $L/F$ with galois group $\Gamma$. As we noted on p276, to give a representation $\rho$ of $T$ on an $F$-vector space $V$ amounts to giving an $X^*(T)$-grading $V(L)=\bigoplus_{\chi\in X^*(T)}V_\chi$ of $V(L)$ with the property that $\sigma V_\chi = V_{\sigma\chi}$ for all $\sigma\in\Gamma$ and $\chi\in X^*(T)$. In this, $L/F$ can be an infinite galois extension.

PROPOSITION 15.6. *Let $\Gamma=\mathrm{Gal}(F^{\mathrm{al}}/F)$. The category of representations $\mathsf{Rep}(T)$ of $T$ on $F$-vector spaces is semisimple. The set of isomorphism classes of simple objects is in natural one-to-one correspondence with the orbits of $\Gamma$ acting on $X^*(T)$, i.e., $\Sigma(\mathsf{Rep}(T))=\Gamma\backslash X^*(T)$. If $V_{\Gamma\chi}$ is a simple object corresponding to $\Gamma\chi$, then $\dim(V_{\Gamma\chi})$ is the order of $\Gamma\chi$, and*

$$\mathrm{End}(V_\chi)\approx F(\chi)$$

*where $F(\chi)$ is the fixed field of the subgroup $\Gamma(\chi)$ of $\Gamma$ fixing $\chi$.*

PROOF. Follows easily from the preceding discussion. □

REMARK 15.7. Let $\chi\in X^*(T)$, and let $\Gamma(\chi)$ and $F(\chi)$ be as in the proposition. Then $\mathrm{Hom}(F(\chi),F^{\mathrm{al}})\cong\Gamma/\Gamma(\chi)$, and so $X^*((\mathbb{G}_m)_{F(\chi)/F})=\mathbb{Z}^{\Gamma/\Gamma(\chi)}$. The map

$$\textstyle\sum n_\sigma\sigma\mapsto\sum n_\sigma\sigma\chi\colon\mathbb{Z}^{\Gamma/\Gamma(\chi)}\to X^*(T)$$

defines a homorphism

$$T\to(\mathbb{G}_m)_{F(\chi)/F}. \tag{58}$$

From this, we get a homomorphism of cohomology groups

$$H^2(F,T)\to H^2(F,(\mathbb{G}_m)_{F(\chi)/F}).$$

But Shapiro's lemma (CFT, II 1.11) shows that $H^2(F,(\mathbb{G}_m)_{F(\chi)/F})\cong H^2(F(\chi),\mathbb{G}_m)$, which is the Brauer group of $F(\chi)$. On composing these maps, we get a homomorphism

$$H^2(F,T)\to\mathrm{Br}(F(\chi)). \tag{59}$$

The proposition gives a natural construction of a semisimple category $\mathsf{M}$ with $\Sigma(\mathsf{M})=\Gamma\backslash N$, where $N$ is any finitely generated $\mathbb{Z}$-module equipped with a continuous action of $\Gamma$. However, the simple objects have commutative endomorphism algebras. To go further, we need to look at new type of structure.

**Affine extensions.** Let $L/F$ be a Galois extension of fields with Galois group $\Gamma$, and let $G$ be an algebraic group over $F$. In the following, we consider only extensions

$$1\to G(L)\to E\to\Gamma\to 1$$

in which the action of $\Gamma$ on $G(L)$ defined by the extension is the natural action, i.e.,

$$\text{if } e_\sigma\mapsto\sigma\text{, then } e_\sigma g e_\sigma^{-1}=\sigma g\quad(e_\sigma\in E,\ \sigma\in\Gamma,\ g\in T(F^{\mathrm{al}})).$$

For example, there is always the ***split extension*** $E_G=_{\mathrm{df}}G(L)\rtimes\Gamma$.

An extension $E$ is ***affine*** if its pull-back to some open subgroup of $\Gamma$ is split. Equivalently, if for the $\sigma$ in some open subgroup of $\Gamma$, there exist $e_\sigma\mapsto\sigma$ such that $e_{\sigma\tau}=e_\sigma e_\tau$. We sometimes call such an $E$ an $L/F$-affine extension with ***kernel*** $G$.

Consider an extension

$$1\to T\to E\to\Gamma\to 1$$

with $T$ commutative. If $E$ is affine, then it is possible to choose the $e_\sigma$'s so that the 2-cocycle $d\colon \Gamma \times \Gamma \to T(L)$ defined by

$$e_\sigma e_\tau = d_{\sigma,\tau} e_\sigma e_\tau, \quad d_{\sigma,\tau} \in T(F^{\mathrm{al}}).$$

is continuous. Thus, in this case $E$ defines a class $cl(E) \in H^2(F,T)$.

A ***homomorphism*** of affine extensions is a commutative diagram

$$\begin{tikzcd}
1 \arrow[r] & G_1(L) \arrow[r] \arrow[d] & E_1 \arrow[r] \arrow[d, "\phi"] & \Gamma \arrow[r] \arrow[d, equal] & 1 \\
1 \arrow[r] & G_2(L) \arrow[r] & E_2 \arrow[r] & \Gamma \arrow[r] & 1
\end{tikzcd}$$

such that the restriction of the homomorphism $\phi$ to $G_1(L)$ is defined by a homomorphism of algebraic groups (over $L$). A ***morphism*** $\phi \to \phi'$ of homomorphisms $E_1 \to E_2$ is an element of $g$ of $G_2(L)$ such that $\mathrm{ad}(g) \circ \phi = \phi'$, i.e., such that

$$g \cdot \phi(e) \cdot g^{-1} = \phi'(e), \quad \text{all } e \in E_1.$$

For a vector space $V$ over $F$, let $E_V$ be the split affine extension defined by the algebraic group $\mathrm{GL}(V)$. A ***representation*** of an affine extension $E$ is a homomorphism $E \to E_V$.

REMARK 15.8. To give a representation of $E_G$ on $E_V$ is the same as to give a representation of $G$ on $V$. More precisely, the functor $\mathsf{Rep}(G) \to \mathsf{Rep}(E_G)$ is an equivalence of categories. The proof of this uses that $H^1(F, \mathrm{GL}(V)) = 1$.

PROPOSITION 15.9. *Let $E$ be an $L/F$-affine extension whose kernel is a torus $T$ split by $L$. The category $\mathsf{Rep}(E)$ is a semisimple $F$-category with $\Sigma(\mathsf{Rep}(E)) = \Gamma\backslash X^*(T)$. Let $V_{\Gamma\chi}$ be a simple representation of $E$ corresponding to $\Gamma\chi \in \Gamma\backslash X^*(T)$. Then, $D = \mathrm{End}(V_{\Gamma_\chi})$ has centre $F(\chi)$, and its class in $\mathrm{Br}(F(\chi))$ is the image of $cl(E)$ under the homomorphism (59).*

PROOF. Omitted (but it is not difficult). □

We shall also need to consider affine extensions in which the kernel is allowed to be a protorus, i.e., the limit of an inverse system of tori. For $T = \varprojlim T_i$, $X^*(T) = \varinjlim X^*(T_i)$, and $T \mapsto X^*(T)$ defines an equivalence from the category of protori to the category of free $\mathbb{Z}$-modules with a continuous action of $\Gamma$. Here continuous means that every element of the module is fixed by an open subgroup of $\Gamma$. Let $L = F^{\mathrm{al}}$. By an ***affine extension with kernel*** $T$, we mean an exact sequence

$$1 \to T(F^{\mathrm{al}}) \to E \to \Gamma \to 1$$

whose push-out

$$1 \to T_i(F^{\mathrm{al}}) \to E_i \to \Gamma \to 1$$

by $T(F^{\mathrm{al}}) \to T_i(F^{\mathrm{al}})$ is an affine extension in the previous sense. A representation of such an extension is defined exactly as before.

REMARK 15.10. Let

$$\begin{tikzcd}
L \arrow[r, dash, "\subset"] \arrow[d, dash, "\Gamma"] & L' \arrow[d, dash, "\Gamma'"] \\
F \arrow[r, dash, "\subset"] & F'
\end{tikzcd}$$

be a diagram of fields in which $L'/F'$ is Galois with group $\Gamma'$. From an $L/F$-affine extension

$$1 \to G(L) \to E \to \Gamma \to 1$$

with kernel $G$ we obtain an $L'/F'$-affine extension

$$1 \to G(L') \to E' \to \Gamma' \to 1$$

with kernel $G_{F'}$ by pulling back by $\sigma \mapsto \sigma|L\colon \Gamma' \to \Gamma$ and pushing out by $G(L) \to G(L')$).

EXAMPLE 15.11. Let $\mathbb{Q}_p^{\mathrm{un}}$ be a maximal unramified extension of $\mathbb{Q}_p$, and let $L_n$ be the subfield of $\mathbb{Q}_p^{\mathrm{un}}$ of degree $n$ over $\mathbb{Q}_p$. Let $\Gamma_n = \mathrm{Gal}(L_n/\mathbb{Q}_p)$, let $D_{1,n}$ be the division algebra in (15.1c), and let

$$1 \to L_n^\times \to N(L_n^\times) \to \Gamma_n \to 1$$

be the corresponding extension. Here $N(L_n^\times)$ is the normalizer of $L_n^\times$ in $D_{1,n}$:

$$N(L_n^\times) = \textstyle\bigsqcup_{0 \le i \le n-1} L_n^\times a^i.$$

This is an $L_n/\mathbb{Q}_p$-affine extension with kernel $\mathbb{G}_m$. On pulling back by $\Gamma \to \Gamma_n$ and pushing out by $L_n^\times \to \mathbb{Q}_p^{\mathrm{un}\times}$, we obtain a $\mathbb{Q}_p^{\mathrm{un}\times}/\mathbb{Q}_p$-affine extension $D_n$ with kernel $\mathbb{G}_m$. From a representation of $D_n$ we obtain a vector space $V$ over $\mathbb{Q}_p^{\mathrm{un}}$ equipped with a $\sigma$-linear map $F$ (the image of $(1,a)$ is $(F,\sigma)$). On tensoring this with the completion $L$ of $\mathbb{Q}_p^{\mathrm{un}}$, we obtain an isocrystal that can be expressed as a sum of $E^\lambda$'s with $\lambda \in \frac{1}{n}\mathbb{Z}$.

Note that there is a canonical section to $N(L_n^\times) \to \Gamma_n$, namely, $\sigma^i \mapsto a^i$, which defines a canonical section to $D_n \to \Gamma$.

There is a homomorphism $D_{nm} \to D_n$ whose restriction to the kernel is multiplication by $m$. The inverse limit of this system is a $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-affine extension $D$ with kernel $\mathbb{G} =_{\mathrm{df}} \varprojlim \mathbb{G}_m$. Note that $X^*(\mathbb{G}) = \varinjlim \frac{1}{n}\mathbb{Z}/\mathbb{Z} = \mathbb{Q}$. There is a natural functor from $\mathsf{Rep}(D)$ to the category of isocrystals, which is faithful and essentially surjective on objects but not full. We call $D$ the ***Dieudonné affine extension***.

**The affine extension $\mathfrak{P}$.** Let $W(p^n)$ be the subgroup of $\mathbb{Q}^{\mathrm{al}\times}$ generated by $W_1(p^n)$, and let $W = \varinjlim W(p^n)$. Then $W$ is a free $\mathbb{Z}$-module of infinite rank with a continuous action of $\Gamma = \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$. For $\pi \in W$, we define $\mathbb{Q}\{\pi\}$ to be the smallest field generated by a representative of $\pi$. If $\pi$ is represented by $\pi_n \in W(p^n)$ and $|\rho(\pi_n)| = (p^n)^{m/2}$, we say that $\pi$ has weight $m$ and we write

$$s_\pi(v) = \frac{\mathrm{ord}_v(\pi_n)}{\mathrm{ord}_v(q)}.$$

THEOREM 15.12. *Let $P$ be the protorus over $\mathbb{Q}$ with $X^*(P) = W$. Then there exists an affine extension*

$$1 \to P(\mathbb{Q}^{\mathrm{al}}) \to \mathfrak{P} \to \Gamma \to 1$$

*such that*

(a) $\Sigma(\mathsf{Rep}(\mathfrak{P})) = \Gamma\backslash W$;
(b) *for $\pi \in W$, let $D(\pi) = \mathrm{End}(V_{\Gamma\pi})$ where $V_{\Gamma\pi}$ is a representation corresponding to $\Gamma\pi$; then $D(\pi)$ is isomorphic to the division algebra $D$ with*

*centre* $\mathbb{Q}\{\pi\}$ *and the invariants*

$$\operatorname{inv}_v(D) = \begin{cases} (\frac{1}{2})^{wt(\pi)} & \text{if } v \text{ is real} \\ s_\pi(v)\cdot[\mathbb{Q}\{\pi\}_v\colon \mathbb{Q}_p] & \text{if } v|p \\ 0 & \text{otherwise.}\end{cases}$$

*Moreover,* $\mathfrak{P}$ *is unique up to isomorphism.*

PROOF. Let $c(\pi)$ denote the class in $\operatorname{Br}(\mathbb{Q}\{\pi\})$ of the division algebra $D$ in (b). To prove the result, we have to show that there exists a unique class in $H^2(\mathbb{Q},P)$ mapping to $c(\pi)$ in $\operatorname{Br}(\mathbb{Q}\{\pi\})$ for all $\pi$:

$$c\mapsto (c(\pi))\colon H^2(\mathbb{Q},P)\overset{(59)}{\to}\textstyle\prod_{\Gamma\pi\in\Gamma\backslash W}\operatorname{Br}(\mathbb{Q}\{\pi\}).$$

This is an exercise in galois cohomology, which, happily, is easier than it looks. □

We call a representation of $\mathfrak{P}$ a ***fake motive*** over $\mathbb{F}$, and a ***fake abelian variety*** if its simple summands correspond to $\pi\in\Gamma\backslash W_1$. Note that the category of fake abelian varieties is a semisimple $\mathbb{Q}$-category with the same numerical invariants as $\mathsf{AV}^0(\mathbb{F})$.

*The local form* $\mathfrak{P}_l$ *of* $\mathfrak{P}$. Let $l$ be a prime of $\mathbb{Q}$, and choose a prime $w_l$ of $\mathbb{Q}^{\mathrm{al}}$ dividing $l$. Let $\mathbb{Q}_l^{\mathrm{al}}$ be the algebraic closure of $\mathbb{Q}_l$ in the completion of $\mathbb{Q}^{\mathrm{al}}$ at $w_l$. Then $\Gamma_l =_{\mathrm{df}} \operatorname{Gal}(\mathbb{Q}_l^{\mathrm{al}}/\mathbb{Q}_l)$ is a closed subgroup of $\Gamma =_{\mathrm{df}} \operatorname{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$, and we have a diagram

(60)
$$\begin{tikzcd} \mathbb{Q}^{\mathrm{al}} \arrow[r, dash] \arrow[d, dash, "\Gamma"] & \mathbb{Q}_l^{\mathrm{al}} \arrow[d, dash, "\Gamma_l"] \\ \mathbb{Q} \arrow[r, dash] & \mathbb{Q}_l. \end{tikzcd}$$

From $\mathfrak{P}$ we obtain a $\mathbb{Q}_l^{\mathrm{al}}/\mathbb{Q}_l$-affine extension $\mathfrak{P}(l)$ by pulling back by $\Gamma_l\to\Gamma$ and pushing out by $P(\mathbb{Q}^{\mathrm{al}})\to P(\mathbb{Q}_l^{\mathrm{al}})$ (cf. 15.10).

*The* $\mathbb{Q}_\ell$*-space attached to a fake motive.* Let $\ell\neq p,\infty$ be a prime of $\mathbb{Q}$.

PROPOSITION 15.13. *There exists a continuous homomorphism* $\zeta_\ell$ *making*

$$\begin{tikzcd} & & & \Gamma_\ell \arrow[dl, "\zeta_\ell"'] \arrow[d, equal] & \\ 1 \arrow[r] & P(\mathbb{Q}_\ell^{\mathrm{al}}) \arrow[r] & \mathfrak{P}(\ell) \arrow[r] & \Gamma_\ell \arrow[r] & 1 \end{tikzcd}$$

*commute.*

PROOF. To prove this, we have to show that the cohomology class of $\mathfrak{P}$ in $H^2(\mathbb{Q},P)$ maps to zero in $H^2(\mathbb{Q}_\ell,P)$, but this is not difficult. □

Fix a homomorphism $\zeta_\ell\colon\Gamma_\ell\to\mathfrak{P}(\ell)$ as in the diagram. Let $\rho\colon\mathfrak{P}\to E_V$ be a fake motive. From $\rho$, we get a homomorphism

$$\rho(\ell)\colon\mathfrak{P}(\ell)\to\mathrm{GL}(V(\mathbb{Q}_\ell^{\mathrm{al}}))\rtimes\Gamma_\ell.$$

For $\sigma\in\Gamma_\ell$, let $(\rho(\ell)\circ\zeta_\ell)(\sigma)=(e_\sigma,\sigma)$. Because $\zeta_\ell$ is a homomorphism, the automorphisms $e_\sigma$ of $V(\mathbb{Q}_\ell^{\mathrm{al}})$ satisfy

$$e_\sigma\circ\sigma e_\tau = e_{\sigma\tau},\quad \sigma,\tau\in\Gamma_\ell,$$

and so

$$\sigma\cdot v = e_\sigma(\sigma v)$$

is an action of $\Gamma_\ell$ on $V(\mathbb{Q}_\ell^{\mathrm{al}})$, which one can check to be continuous. Therefore (AG 16.14), $V_\ell(\rho) =_{\mathrm{df}} V(\mathbb{Q}_\ell^{\mathrm{al}})^{\Gamma_\ell}$ is a $\mathbb{Q}_\ell$-structure on $V(\mathbb{Q}_\ell^{\mathrm{al}})$. In this way, we get a functor $\rho \mapsto V_\ell(\rho)$ from the category of fake motives over $\mathbb{F}$ to vector spaces over $\mathbb{Q}_\ell$.

The $\zeta_\ell$ can be chosen in such a way that the spaces $V_\ell(\rho)$ contain lattices $\Lambda_\ell(\rho)$ that are well-defined for almost all $\ell \neq p$, which makes it possible to define

$$V_f^p(\rho) = \prod\nolimits_{\ell\neq p,\infty}(V_\ell(\rho)\colon \Lambda_\ell(\rho)).$$

It is a free module over $\mathbb{A}_f^p =_{\mathrm{df}} \prod_{\ell\neq p,\infty}(\mathbb{Q}_\ell\colon \mathbb{Z}_\ell)$.

*The isocrystal of a fake motive.* Choose a prime $w_p$ of $\mathbb{Q}^{\mathrm{al}}$ dividing $p$, and let $\mathbb{Q}_p^{\mathrm{un}}$ and $\mathbb{Q}_p^{\mathrm{al}}$ denote the subfields of the completion of $\mathbb{Q}^{\mathrm{al}}$ at $w_p$. Then $\Gamma_p =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{al}}/\mathbb{Q}_p)$ is a closed subgroup of $\Gamma =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ and $\Gamma_p^{\mathrm{un}} =_{\mathrm{df}} \mathrm{Gal}(\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p)$ is a quotient of $\Gamma_p$.

PROPOSITION 15.14. *(a) The affine extension $\mathfrak{P}(p)$ arises by pull-back and push-out from a $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-affine extension $\mathfrak{P}(p)^{\mathrm{un}}$.*

*(b) There is a homomorphism of $\mathbb{Q}_p^{\mathrm{un}}/\mathbb{Q}_p$-extensions $D \to \mathfrak{P}(p)^{\mathrm{un}}$ whose restriction to the kernels, $\mathbb{G} \to P_{\mathbb{Q}_p}$, corresponds to the map on characters $\pi \mapsto s_\pi(w_p)\colon W \to \mathbb{Q}$.*

PROOF. (a) This follows from the fact that the image of the cohomology class of $\mathfrak{P}$ in $H^2(\Gamma_p, P(\mathbb{Q}_p^{\mathrm{al}}))$ arises from a cohomology class in $H^2(\Gamma_p^{\mathrm{un}}, P(\mathbb{Q}_p^{\mathrm{un}}))$.

(b) This follows from the fact that the homomorphism $H^2(\mathbb{Q}_p, \mathbb{G}) \to H^2(\mathbb{Q}_p, P_{\mathbb{Q}_p})$ sends the cohomology class of $D$ to that of $\mathfrak{P}(p)^{\mathrm{un}}$. $\square$

In summary:

$$\begin{array}{ccccccccc}
1 & \longrightarrow & \mathbb{G}_m(\mathbb{Q}_p^{\mathrm{un}}) & \longrightarrow & D & \longrightarrow & \Gamma_p^{\mathrm{un}} & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \| & & \\
1 & \longrightarrow & P(\mathbb{Q}_p^{\mathrm{un}}) & \longrightarrow & \mathfrak{P}(p)^{\mathrm{un}} & \longrightarrow & \Gamma_p^{\mathrm{un}} & \longrightarrow & 1 \\
 & & \downarrow & & \vdots & & \uparrow & & \\
1 & \longrightarrow & P(\mathbb{Q}_p^{\mathrm{al}}) & \longrightarrow & \mathfrak{P}(p) & \longrightarrow & \Gamma_p & \longrightarrow & 1
\end{array}$$

A fake motive $\rho\colon \mathfrak{P} \to E_V$ gives rise to a representation of $\mathfrak{P}(p)$, which arises from a representation of $\mathfrak{P}(p)^{\mathrm{un}}$ (cf. the argument in the preceding subsubsection). On composing this with the homomorphism $D \to \mathfrak{P}(p)^{\mathrm{un}}$, we obtain a representation of $D$, which gives rise to an isocrystal $D(\rho)$ as in (15.11).

*Abelian varieties of CM-type and fake abelian varieties.* We saw in (10.5) that an abelian variety of CM-type over $\mathbb{Q}^{\mathrm{al}}$ defines an abelian variety over $\mathbb{F}$. Does it also define a fake abelian variety? The answer is yes.

PROPOSITION 15.15. *Let $T$ be a torus over $\mathbb{Q}$ split by a CM-field, and let $\mu$ be a cocharacter of $T$ such that $\mu + \iota\mu$ is defined over $\mathbb{Q}$ (here $\iota$ is complex conjugation). Then there is a homomorphism, well defined up to isomorphism,*

$$\phi_\mu\colon \mathfrak{P} \to E_T.$$

PROOF. Omitted. $\square$

Let $A$ be an abelian variety of CM-type $(E,\Phi)$ over $\mathbb{Q}^{\mathrm{al}}$, and let $T=(\mathbb{G}_m)_{E/\mathbb{Q}}$. Then $\Phi$ defines a cocharacter $\mu_\Phi$ of $T$ (see 12.4(b)), which obviously satisfies the conditions of the proposition. Hence we obtain a homomorphism $\phi\colon \mathfrak{P}\to E_T$. Let $V=H_1(A,\mathbb{Q})$. From $\phi$ and the representation $\rho$ of $T$ on $V$ we obtain a fake abelian variety $\rho\circ\phi$ such that $V_\ell(\rho\circ\phi)=H_1(A,\mathbb{Q}_\ell)$ (obvious) and $D(\rho)$ is isomorphic to the Dieudonné module of the reduction of $A$ (restatement of the Shimura-Taniyama formula).

Aside 15.16. The category of fake abelian varieties has similar properties to $\mathsf{AV}^0(\mathbb{F})$. By using the $\mathbb{Q}_\ell$-spaces and the isocrystals attached to a fake abelian variety, it is possible to define a $\mathbb{Z}$-linear category with properties similar to $\mathsf{AV}(\mathbb{F})$.

Notes. The affine extension $\mathfrak{P}$ is defined in Langlands and Rapoport 1987, §§1–3, where it is called "die pseudomotivische Galoisgruppe". There an affine extension is called a Galoisgerbe although, rather than a gerbe, it can more accurately be described as a concrete realizations of a groupoid. See also Milne 1992. In the above, I have ignored uniqueness questions, which can be difficult (see Milne 2003).

## 16. The good reduction of Shimura varieties

We now write $\mathrm{Sh}_K(G,X)$, or just $\mathrm{Sh}_K$, for the canonical model of the Shimura variety over its reflex field.

**The points of the Shimura variety with coordinates in the algebraic closure of the rational numbers.** When we have a description of the points of the Shimura variety over $\mathbb{C}$ in terms of abelian varieties or motives plus additional data, then the same description holds over $\mathbb{Q}^{\mathrm{al}}$. For example, for the Siegel modular variety attached to a symplectic space $(V,\psi)$, $\mathrm{Sh}_K(\mathbb{Q}^{\mathrm{al}})$ classifies the isomorphism classes of triples $(A,s,\eta K)$ in which $A$ is an abelian variety defined over $\mathbb{Q}^{\mathrm{al}}$, $s$ is an element of $\mathrm{NS}(A)\otimes\mathbb{Q}$ containing a $\mathbb{Q}^\times$-multiple of an ample divisor, and $\eta$ is a $K$-orbit of isomorphisms $V(\mathbb{A}_f)\to V_f(A)$ sending $\psi$ to an $\mathbb{A}_f^\times$-multiple of the pairing defined by $s$. Here $\mathrm{NS}(A)$ is the Nèron-Severi group of $A$ (divisor classes modulo algebraic equivalence).

On the other hand, I do not know a description of $\mathrm{Sh}_K(\mathbb{Q}^{\mathrm{al}})$ when, for example, $G^{\mathrm{ad}}$ has factors of type $E_6$ or $E_7$ or mixed type $D$. In these cases, the proof of the existence of a canonical model is quite indirect.

**The points of the Shimura variety with coordinates in the reflex field.** Over $E=E(G,X)$ the following additional problem arises. Let $A$ be an abelian variety over $\mathbb{Q}^{\mathrm{al}}$. Suppose we know that $\sigma A$ is isomorphic to $A$ for all $\sigma\in\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E)$. Does this imply that $A$ is defined over $E$? Choose an isomorphism $f_\sigma\colon \sigma A\to A$ for each $\sigma$ fixing $E$. A necessary condition that the $f_\sigma$ arise from a model over $E$ is that they satisfy the cocycle condition: $f_\sigma\circ\sigma f_\tau=f_{\sigma\tau}$. Of course, if the cocycle condition fails for one choice of the $f_\sigma$'s, we can try another, but there is an obstruction to obtaining a cocycle which lies in the cohomology set $H^2(\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E),\mathrm{Aut}(A))$.

Certainly, this obstruction would vanish if $\mathrm{Aut}(A)$ were trivial. One may hope that the automorphism group of an abelian variety (or motive) plus data in the family classified by $\mathrm{Sh}_K(G,X)$ is trivial, at least when $K$ is small. This is so when condition SV5 holds, but not otherwise.

In the Siegel case, the centre of $G$ is $\mathbb{G}_m$ and so SV5 holds. Therefore, provided $K$ is sufficiently small, for any field $L$ containing $E(G,X)$, $\mathrm{Sh}_K(L)$ classifies triples $(A, s, \eta K)$ satisfying the same conditions as when $L = \mathbb{Q}^{\mathrm{al}}$. Here $A$ an abelian variety over $L$, $s \in \mathrm{NS}(A) \otimes \mathbb{Q}$, and $\eta$ is an isomorphism $V(\mathbb{A}_f) \to V_f(A)$ such that $\eta K$ is stable under the action of $\mathrm{Gal}(L^{\mathrm{al}}/L)$.

In the Hilbert case (4.14), the centre of $G$ is $(\mathbb{G}_m)_{F/\mathbb{Q}}$ for $F$ a totally real field and SV5 fails: $F^\times$ is not discrete in $\mathbb{A}_{F,f}^\times$ because every nonempty open subgroup of $\mathbb{A}_{F,f}^\times$ will contain infinitely many units. In this case, one has a description of $\mathrm{Sh}_K(L)$ when $L$ is algebraically closed, but otherwise all one can say is that $\mathrm{Sh}_K(L) = \mathrm{Sh}_K(L^{\mathrm{al}})^{\mathrm{Gal}(L^{\mathrm{al}}/L)}$.

**Hyperspecial subgroups.** The modular curve $\Gamma_0(N)\backslash\mathcal{H}_1$ is defined over $\mathbb{Q}$, and it has good reduction at the primes not dividing the level $N$ and bad reduction at the others. Before explaining what is known in general, we need to introduce the notion of a hyperspecial subgroup.

Definition 16.1. Let $G$ be a reductive group over $\mathbb{Q}$ (over $\mathbb{Q}_p$ will do). A subgroup $K \subset G(\mathbb{Q}_p)$ is ***hyperspecial*** if there exists a flat group scheme $\mathcal{G}$ over $\mathbb{Z}_p$ such that

- $\circ$ $\mathcal{G}_{\mathbb{Q}_p} = G$ (i.e., $\mathcal{G}$ extends $G$ to $\mathbb{Z}_p$);
- $\circ$ $\mathcal{G}_{\mathbb{F}_p}$ is a connected reductive group (necessarily of the same dimension as $G$ because of flatness);
- $\circ$ $\mathcal{G}(\mathbb{Z}_p) = K$.

Example 16.2. Let $G = \mathrm{GSp}(V,\psi)$. Let $\Lambda$ be a lattice in $V(\mathbb{Q}_p)$, and let $K_p$ be the stabilizer of $\Lambda$. Then $K_p$ is hyperspecial if the restriction of $\psi$ to $\Lambda \times \Lambda$ takes values in $\mathbb{Z}_p$ and is perfect (i.e., induces an isomorphism $\Lambda \to \Lambda^\vee$; equivalently, induces a nondegenerate pairing $\Lambda/p\Lambda \times \Lambda/p\Lambda \to \mathbb{F}_p$). In this case, $\mathcal{G}_{\mathbb{F}_p}$ is again a group of symplectic similitudes over $\mathbb{F}_p$ (at least if $p \neq 2$).

Example 16.3. In the PEL-case, in order for there to exist a hyperspecial group, the algebra $B$ must be unramified above $p$, i.e., $B \otimes_{\mathbb{Q}} \mathbb{Q}_p$ must be a product of matrix algebras over unramified extensions of $\mathbb{Q}_p$. When this condition holds, the description of the hyperspecial groups is similar to that in the Siegel case.

There exists a hyperspecial subgroup in $G(\mathbb{Q}_p)$ if and only if $G$ is ***unramified*** over $\mathbb{Q}_p$, i.e., quasisplit over $\mathbb{Q}_p$ and split over an unramified extension.

*For the remainder of this section we fix a hyperspecial subgroup $K_p \subset G(\mathbb{Q}_p)$, and we write $\mathrm{Sh}_p(G,X)$ for the family of varieties $\mathrm{Sh}_{K^p \times K_p}(G,X)$ with $K^p$ running over the compact open subgroups of $G(\mathbb{A}_f^p)$. The group $G(\mathbb{A}_f^p)$ acts on the family $\mathrm{Sh}_p(G,X)$.*

**The good reduction of Shimura varieties.** Roughly speaking, there are two reasons a Shimura variety may have bad reduction at a prime dividing $p$: the reductive group itself may be ramified at $p$ or $p$ may divide the level. For example, the Shimura curve defined by a quaternion algebra $B$ over $\mathbb{Q}$ will have bad reduction at a prime $p$ dividing the discriminant of $B$, and (as we noted above) $\Gamma_0(N)\backslash\mathcal{H}_1$ has bad reduction at a prime dividing $N$. The existence of a hyperspecial subgroup $K_p$ forces $G$ to be unramified at $p$, and by considering only the varieties $\mathrm{Sh}_{K^p K_p}(G,X)$ we avoid the second problem.

THEOREM 16.4. *Let* $\mathrm{Sh}_p(G,X)$ *be the inverse system of varieties over* $E(G,X)$ *defined by a Shimura datum* $(G,X)$ *of abelian type and a hyperspecial subgroup* $K_p \subset G(\mathbb{Q}_p)$. *Then, except possibly for some small set of primes* $p$ *depending only on* $(G,X)$, $\mathrm{Sh}_p(G,X)$ *has canonical good reduction at every prime* $\mathfrak{p}$ *of* $E(G,X)$ *dividing* $p$, .

REMARK 16.5. Let $E_\mathfrak{p}$ be the completion of $E$ at $\mathfrak{p}$, let $\hat{\mathcal{O}}_\mathfrak{p}$ be the ring of integers in $E_\mathfrak{p}$, and let $k(\mathfrak{p})$ be the residue field $\hat{\mathcal{O}}_\mathfrak{p}/\mathfrak{p}$.

(a) By $\mathrm{Sh}_p(G,X)$ having good reduction $\mathfrak{p}$, we mean that the inverse system

$$(\mathrm{Sh}_{K^pK_p}(G,X))_{K^p}, \quad K^p \subset G(\mathbb{A}_f^p) \text{ compact open, } K_p \text{ fixed,}$$

extends to an inverse system of flat schemes $\mathcal{S}_p = (\mathcal{S}_{K^p})$ over $\hat{\mathcal{O}}_\mathfrak{p}$ whose reduction modulo $\mathfrak{p}$ is an inverse system of varieties $(\overline{\mathrm{Sh}}_{K^pK_p}(G,X))_{K^p}$ over $k(\mathfrak{p})$ such that, for $K^p \supset K'^p$ sufficiently small,

$$\overline{\mathrm{Sh}}_{K^pK_p} \leftarrow \overline{\mathrm{Sh}}_{K'^pK_p}$$

is an étale map of smooth varieties. We require also that the action of $G(\mathbb{A}_f^p)$ on $\mathrm{Sh}_p$ extends to an action on $\mathcal{S}_p$.

(b) A variety over $E_\mathfrak{p}$ may not have good reduction to a smooth variety over $k(\mathfrak{p})$ — this can already be seen for elliptic curves — and, when it does it will generally have good reduction to many different smooth varieties, none of which is obviously the best. For example, given one good reduction, one can obtain another by blowing up a point in its closed fibre. By $\mathrm{Sh}_p(G,X)$ having *canonical* good reduction at $\mathfrak{p}$, I mean that, for any formally smooth scheme $T$ over $\hat{\mathcal{O}}_\mathfrak{p}$,

$$\mathrm{Hom}_{\hat{\mathcal{O}}_\mathfrak{p}}(T, \varprojlim_{K^p} \mathcal{S}_{K^p}) \cong \mathrm{Hom}_{E_\mathfrak{p}}(T_{E_\mathfrak{p}}, \varprojlim_{K^p} \mathrm{Sh}_{K^pK_p}). \tag{61}$$

A smooth scheme is formally smooth, and an inverse limit of schemes étale over a smooth scheme is formally smooth. As $\varprojlim \mathcal{S}_{K^p}$ is formally smooth over $\hat{\mathcal{O}}_\mathfrak{p}$, (61) characterizes it uniquely up to a unique isomorphism (by the Yoneda lemma).

(c) In the Siegel case, Theorem 16.4 was proved by Mumford (his Fields medal theorem; Mumford 1965). In this case, the $\mathcal{S}_{K^p}$ and $\overline{\mathrm{Sh}}_{K^pK_p}$ are moduli schemes. The PEL-case can be considered folklore in that several authors have deduced it from the Siegel case and published sketches of proof, the most convincing of which is in Kottwitz 1992. In this case, $\mathcal{S}_p(G,X)$ is the zariski closure of $\mathrm{Sh}_p(G,X)$ in $\mathcal{S}_p(G(\psi),X(\psi))$ (cf. 5.16), and it is a moduli scheme. The hodge case[19] was proved by Vasiu (1999) except for a small set of primes. In this case, $\mathcal{S}_p(G,X)$ is the normalization of the zariski closure of $\mathrm{Sh}_p(G,X)$ in $\mathcal{S}_p(G(\psi),X(\psi))$. The case of abelian type follows easily from the hodge case.

(d) That $\mathrm{Sh}_p$ should have good reduction when $K_p$ is hyperspecial was conjectured in Langlands 1976, p411. That there should be a canonical model characterized by a condition like that in (b) was conjectured in Milne 1992, §2.

[19]Over the reflex field, Shimura varieties of hodge type are no more difficult than Shimura varieties of PEL-type, but when one reduces modulo a prime they become much more difficult for two reasons: general tensors are more difficult to work with than endomorphisms, and little is known about hodge tensors in characteristic $p$.

**Definition of the Langlands-Rapoport set.** Let $(G,X)$ be a Shimura datum for which SV4,5,6 hold, and let

$$\mathrm{Sh}_p(\mathbb{C})=\mathrm{Sh}(\mathbb{C})/K_p=\varprojlim_{K^p}\mathrm{Sh}_{K^pK_p}(G,X)(\mathbb{C}).$$

For $x\in X$, let $I(x)$ be the subgroup $G(\mathbb{Q})$ fixing $x$, and let

$$S(x)=I(x)\backslash X^p(x)\times X_p(x),\quad X^p(x)=G(\mathbb{A}_f^p),\quad X_p(x)=G(\mathbb{Q}_p)/K_p.$$

One sees easily that there is a canonical bijection of sets with $G(\mathbb{A}_f^p)$-action

$$\bigsqcup S(x)\to\mathrm{Sh}_p(\mathbb{C})$$

where the left hand side is the disjoint union over a set of representatives for $G(\mathbb{Q})\backslash X$. This decomposition has a modular interpretation. For example, in the case of a Shimura variety of hodge type, the set $S(x)$ classifies the family of isomorphism classes of triples $(A,(s_i),\eta K)$ with $(A,(s_i))$ isomorphic to a fixed pair.

Langlands and Rapoport (1987, 5e) conjecture that $\overline{\mathrm{Sh}}_p(\mathbb{F})$ has a similar description except that now the left hand side runs over a set of isomorphism classes of homomorphisms $\phi\colon\mathfrak{P}\to E_G$. Recall that an isomorphism from one $\phi$ to a second $\phi'$ is an element $g$ of $G(\mathbb{Q}^{\mathrm{al}})$ such that

$$\phi'(p)=g\cdot\phi(p)\cdot g^{-1},\quad\text{all }p\in\mathfrak{P}.$$

Such a $\phi$ should be thought of as a "pre fake abelian motive with tensors". Specifically, if we fix a faithful representation $G\hookrightarrow\mathrm{GL}(V)$ and tensors $t_i$ for $V$ such that $G$ is the subgroup of $\mathrm{GL}(V)$ fixing the $t_i$, then each $\phi$ gives a representation $\mathfrak{P}\to\mathrm{GL}(V(\mathbb{Q}^{\mathrm{al}}))\rtimes\Gamma$ (i.e., a fake abelian motive) plus tensors.

*Definition of the set $S(\phi)$.* We now fix a homomorphism $\phi\colon\mathfrak{P}\to E_G$ and define a set $S(\phi)$ equipped with a right action of $G(\mathbb{A}_f^p)$ and a commuting Frobenius operator $\Phi$.

Definition of the group $I(\phi)$. The group $I(\phi)$ is defined to be the group of automorphisms of $\phi$,

$$I(\phi)=\{g\in G(\mathbb{Q}^{\mathrm{al}})\mid\mathrm{ad}(g)\circ\phi=\phi\}.$$

Note that if $\rho\colon G\to\mathrm{GL}(V)$ is a faithful representation of $G$, then $\rho\circ\phi$ is a fake motive and $I(\phi)\subset\mathrm{Aut}(\rho\circ\phi)$ (here we have abbreviated $\rho\rtimes 1$ to $\rho$).

Definition of $X^p(\phi)$. Let $\ell$ be a prime $\neq p,\infty$. We choose a prime $w_\ell$ of $\mathbb{Q}^{\mathrm{al}}$ dividing $\ell$, and define $\mathbb{Q}_\ell^{\mathrm{al}}$ and $\Gamma_\ell\subset\Gamma$ as on p364.

Regard $\Gamma_\ell$ as an $\mathbb{Q}_\ell^{\mathrm{al}}/\mathbb{Q}_\ell$-affine extension with trivial kernel, and write $\xi_\ell$ for the homomorphism

$$\sigma\mapsto(1,\sigma)\colon\Gamma_\ell\to E_G(\ell),\quad E_G(\ell)=G(\mathbb{Q}_\ell^{\mathrm{al}})\rtimes\Gamma_\ell.$$

From $\phi$ we get a homomorphism $\phi(\ell)\colon\mathfrak{P}(\ell)\to E_G(\ell)$, and, on composing this with the homomorphism $\zeta_\ell\colon\Gamma_\ell\to\mathfrak{P}(\ell)$ provided by (15.13), we get a second homomorphism $\Gamma_\ell\to E_G(\ell)$.

Define

$$X_\ell(\phi)=\mathrm{Isom}(\xi_\ell,\zeta_\ell\circ\phi(\ell)).$$

Clearly, $\mathrm{Aut}(\xi_\ell)=G(\mathbb{Q}_\ell)$ acts on $X_\ell(\phi)$ on the right, and $I(\phi)$ acts on the left. If $X_\ell(\phi)$ is nonempty, then the first action makes $X_\ell(\phi)$ into a principal homogeneous space for $G(\mathbb{Q}_\ell)$.

Note that if $\rho\colon G\to\mathrm{GL}(V)$ is a faithful representation of $G$, then

$$X_\ell(\phi)\subset\mathrm{Isom}(V(\mathbb{Q}_\ell),V_\ell(\rho\circ\phi)).\tag{62}$$

By choosing the $\zeta_\ell$ judiciously (cf. p365), we obtain compact open subspaces of the $X_\ell(\phi)$, and we can define $X^p(\phi)$ to be the restricted product of the $X_\ell(\phi)$. If nonempty, it is a principal homogeneous space for $G(\mathbb{A}_f^p)$.

Definition of $X_p(\phi)$. We choose a prime $w_p$ of $\mathbb{Q}^{\mathrm{al}}$ dividing $p$, and we use the notations of p365. We let $L$ denote the completion of $\mathbb{Q}_p^{\mathrm{un}}$, and we let $\mathcal{O}_L$ denote the ring of integers in $L$ (it is the ring of Witt vectors with coefficients in $\mathbb{F}$). We let $\sigma$ Frobenius automorphism of $\mathbb{Q}_p^{\mathrm{un}}$ or $L$ that acts as $x \mapsto x^p$ on the residue field.

From $\phi$ and (15.14), we have homomorphisms

$$D \longrightarrow \mathfrak{P}(p)^{\mathrm{un}} \xrightarrow{\phi(p)^{\mathrm{un}}} G(\mathbb{Q}_p^{\mathrm{un}}) \rtimes \Gamma_p^{\mathrm{un}}.$$

For some $n$, the composite factors through $D_n$. There is a canonical element in $D_n$ mapping to $\sigma$, and we let $(b, \sigma)$ denote its image in $G(\mathbb{Q}_p^{\mathrm{un}}) \rtimes \Gamma_p^{\mathrm{un}}$. The image $b(\phi)$ of $b$ in $G(L)$ is well-defined up to $\sigma$-conjugacy, i.e., if $b(\phi)'$ also arises in this way, then $b(\phi)' = g^{-1} \cdot b(\phi) \cdot \sigma g$.

Note that if $\rho\colon G \to \mathrm{GL}(V)$ is a faithful representation of $G$, then $D(\phi \circ \rho)$ is $V(L)$ with $F$ acting as $v \mapsto b(\phi)\sigma v$.

Recall p344 that we have a well-defined $G(\mathbb{Q}^{\mathrm{al}})$-conjugacy class $c(X)$ of cocharacters of $G_{\mathbb{Q}^{\mathrm{al}}}$. We can transfer this to conjugacy class of cocharacters of $G_{\mathbb{Q}_p^{\mathrm{al}}}$, which contains an element $\mu$ defined over $\mathbb{Q}_p^{\mathrm{un}}$ (see 12.3; $G$ splits over $\mathbb{Q}_p^{\mathrm{un}}$ because we are assuming it contains a hyperspecial group). Let

$$C_p = G(\mathcal{O}_L) \cdot \mu(p) \cdot G(\mathcal{O}_L).$$

Here we are writing $G(\mathcal{O}_L)$ for $\mathcal{G}(\mathcal{O}_L)$ with $\mathcal{G}$ as in the definition of hyperspecial.

Define

$$X_p(\phi) = \{g \in G(L)/G(\mathcal{O}_L) \mid g^{-1} \cdot b(\phi) \cdot g \in C_p\}.$$

There is a natural action of $I(\phi)$ on this set.

Definition of the Frobenius element $\Phi$. For $g \in X_p(\phi)$, define

$$\Phi(g) = b(\phi) \cdot \sigma b(\phi) \cdot \dots \cdot \sigma^{m-1} b(\phi) \cdot \sigma^m g$$

where $m = [E_v\colon \mathbb{Q}_p]$.

The set $S(\phi)$. Let

$$S(\phi) = I(\phi) \backslash X^p(\phi) \times X_p(\phi).$$

Since $I(\phi)$ acts on both $X^p(\phi)$ and $X_p(\phi)$, this makes sense. The group $G(\mathbb{A}_f^p)$ acts on $S(\phi)$ through its action on $X^p(\phi)$ and $\Phi$ acts through its action on $X_p(\phi)$.

*The admissibility condition.* The homomorphisms $\phi\colon \mathfrak{P} \to E_G$ contributing to the Langlands-Rapoport set must satisfy an admissibility condition at each prime plus one global condition.

The condition at $\infty$. Let $E_\infty$ be the extension

$$1 \to \mathbb{C}^\times \to E_\infty \to \Gamma_\infty \to 1, \quad \Gamma_\infty = \mathrm{Gal}(\mathbb{C}/\mathbb{R}) = \langle \iota \rangle$$

associated with the quaternion algebra $\mathbb{H}$, and regard it as an affine extension with kernel $\mathbb{G}_m$. Note that $E_\infty = \mathbb{C}^\times \sqcup \mathbb{C}^\times j$ and $jzj^{-1} = \bar{z}$.

From the diagram (60) with $l = \infty$, we obtain a $\mathbb{C}/\mathbb{R}$-affine extension

$$1 \to P(\mathbb{C}) \to \mathfrak{P}(\infty) \to \Gamma_\infty \to 1.$$

LEMMA 16.6. *There is a homomorphism $\zeta_\infty\colon E_\infty \to \mathfrak{P}(\infty)$ whose restriction to the kernels, $\mathbb{G}_m \mapsto P_{\mathbb{C}}$, corresponds to the map on characters $\pi \mapsto wt(\pi)$.*

PROOF. This follows from the fact that the homomorphism $H^2(\Gamma_\infty, \mathbb{G}_m) \to H^2(\Gamma_\infty, P_\mathbb{R})$ sends the cohomology class of $E_\infty$ to that of $\mathfrak{P}(\infty)$. $\square$

LEMMA 16.7. *For any $x \in X$, the formulas*

$$\xi_x(z) = (w_X(z), 1), \quad \xi_x(j) = (\mu_x(-1)^{-1}, \iota)$$

*define a homomorphism $E_\infty \to \mathfrak{P}(\infty)$. Replacing $x$ with a different point, replaces the homomorphism with an isomorphic homomorphism.*

PROOF. Easy exercise. $\square$

Write $\xi_X$ for the isomorphism class of homomorphisms defined in (16.7). Then the admissibility condition at $\infty$ is that $\zeta_\infty \circ \phi(\infty) \in \xi_X$.

The condition at $\ell \neq p$. The admissibility condition at $\ell \neq p$ is that the set $X_\ell(\phi)$ be nonempty, i.e., that $\zeta_\ell \circ \phi(\ell)$ be isomorphic to $\xi_\ell$.

The condition at $p$. The condition at $p$ is that the set $X_p(\phi)$ be nonempty.

The global condition. Let $\nu\colon G \to T$ be the quotient of $G$ by its derived group. From $X$ we get a conjugacy class of cocharacters of $G_\mathbb{C}$, and hence a well defined cocharacter $\mu$ of $T$. Under our hypotheses on $(G, X)$, $\mu$ satisfies the conditions of (15.15), and so defines a homomorphism $\phi_\mu\colon \mathfrak{P} \to E_T$. The global condition is that $\nu \circ \phi$ be isomorphic to $\phi_\mu$.

*The Langlands-Rapoport set.* The Langlands-Rapoport set

$$\mathrm{LR}(G, X) = \bigsqcup S(\phi)$$

where the disjoint union is over a set of representatives for the isomorphism classes of admissible homomorphism $\phi\colon \mathfrak{P} \to E_G$. There are commuting actions of $G(\mathbb{A}_f^p)$ and of the Frobenius operator $\Phi$ on $\mathrm{LR}(G, X)$.

**The conjecture of Langlands and Rapoport.**

CONJECTURE 16.8 (Langlands and Rapoport 1987). *Let $(G, X)$ be a Shimura datum satisfying SV4,5,6 and such that $G^{der}$ is simply connected, and let $K_p$ be a hyperspecial subgroup of $G(\mathbb{Q}_p)$. Let $\mathfrak{p}$ be a prime of $E(G, X)$ dividing $p$, and assume that* $\mathrm{Sh}_p$ *has canonical good reduction at $\mathfrak{p}$. Then there is a bijection of sets*

$$\mathrm{LR}(G, X) \to \overline{\mathrm{Sh}}_p(G, X)(\mathbb{F}) \tag{63}$$

*compatible with the actions $G(\mathbb{A}_f^p)$ and the Frobenius elements.*

REMARK 16.9. (a) The conditions SV5 and SV6 are not in the original conjecture — I included them to simplify the statement of the conjecture.

(b) There is also a conjecture in which one does not require SV4, but this requires that $\mathfrak{P}$ be replaced by a more complicated affine extension $\mathfrak{Q}$.

(c) The conjecture as originally stated is definitely wrong without the assumption that $G^{\mathrm{der}}$ is simply connected. However, when one replaces the "admissible homomorphisms" in the statement with another notion, that of "special homomorphisms", one obtains a statement that should be true for all Shimura varieties. In fact, it is known that the statement with $G^{\mathrm{der}}$ simply connected then implies the general statement (see Milne 1992, §4, for the details and a more precise statement).

(d) It is possible to state, and prove, a conjecture similar to (16.8) for zero-dimensional Shimura varieties. The map $(G, X) \to (T, Y)$ (see p311) defines a map of the associated Langlands-Rapoport sets, and we should add to the conjecture that

$$\begin{array}{ccc} LR(G,X) & \longrightarrow & \mathrm{Sh}_p(G,X)(\mathbb{F}) \\ \downarrow & & \downarrow \\ LR(T,Y) & \longrightarrow & \mathrm{Sh}_p(T,Y)(\mathbb{F}) \end{array}$$

commutes.

## 17. A formula for the number of points

A reader of the last two sections may be sceptical of the value of a description like (63), even if proved. In this section we briefly explain how it leads to a very explicit, and useful, formula for the number of points on the reduction of a Shimura variety with values in a finite field.

Throughout, $(G,X)$ is a Shimura datum satisfying SV4,5,6 and $K_p$ is a hyperspecial subgroup of $G(\mathbb{Q}_p)$. We assume that $G^{\mathrm{der}}$ simply connected and that $\mathrm{Sh}_p(G,X)$ has canonical good reduction at a prime $\mathfrak{p}|p$ of the reflex field $E = E(G,X)$. Other notations are as in the last section; for example, $L_n$ is the subfield of $\mathbb{Q}_p^{\mathrm{un}}$ of degree $n$ over $\mathbb{Q}_p$ and $L$ is the completion of $\mathbb{Q}_p^{\mathrm{un}}$. We fix a field $\mathbb{F}_q \supset k(\mathfrak{p}) \supset \mathbb{F}_p$, $q = p^n$.

**Triples.** We consider triples $(\gamma_0;\gamma,\delta)$ where

- $\gamma_0$ is a semisimple element of $G(\mathbb{Q})$ that is contained in an elliptic torus of $G_{\mathbb{R}}$ (i.e., a torus that is anisotropic modulo the centre of $G_{\mathbb{R}}$),
- $\gamma = (\gamma(\ell))_{\ell\neq p,\infty}$ is an element of $G(\mathbb{A}_f^p)$ such that, for all $\ell$, $\gamma(\ell)$ becomes conjugate to $\gamma_0$ in $G(\mathbb{Q}_\ell^{\mathrm{al}})$,
- $\delta$ is an element of $G(L_n)$ such that

$$\mathcal{N}\delta \overset{\mathrm{df}}{=} \delta\cdot\sigma\delta\cdot\ldots\cdot\sigma^{n-1}\delta,$$

  becomes conjugate to $\gamma_0$ in $G(\mathbb{Q}_p^{\mathrm{al}})$.

Two triples $(\gamma_0;\gamma,\delta)$ and $(\gamma_0';\gamma',\delta')$ are said to be ***equivalent***, $(\gamma_0;\gamma,\delta)\sim(\gamma_0';\gamma',\delta')$, if $\gamma_0$ is conjugate to $\gamma_0'$ in $G(\mathbb{Q})$, $\gamma(\ell)$ is conjugate to $\gamma'(\ell)$ in $G(\mathbb{Q}_\ell)$ for each $\ell\neq p,\infty$, and $\delta$ is $\sigma$-conjugate to $\delta'$ in $G(L_n)$.

Given such a triple $(\gamma_0;\gamma,\delta)$, we set:

- $I_0 = G_{\gamma_0}$, the centralizer of $\gamma_0$ in $G$; it is connected and reductive;
- $I_\infty =$ the inner form of $I_{0\mathbb{R}}$ such that $I_\infty/Z(G)$ is anisotropic;
- $I_\ell =$ the centralizer of $\gamma(\ell)$ in $G_{\mathbb{Q}_\ell}$;
- $I_p =$ the inner form of $G_{\mathbb{Q}_p}$ such that $I_p(\mathbb{Q}_p) = \{x\in G(L_n) \mid x^{-1}\cdot\delta\cdot\sigma x = \delta\}$.

We need to assume that the triple satisfies the following condition:

> (*) there exists an inner form $I$ of $I_0$ such that $I_{\mathbb{Q}_\ell}$ is isomorphic to $I_\ell$ for all $\ell$ (including $p$ and $\infty$).

Because $\gamma_0$ and $\gamma_\ell$ are stably conjugate, there exists an isomorphism $a_\ell\colon I_{0,\mathbb{Q}_\ell^{\mathrm{al}}}\to I_{\ell,\mathbb{Q}_\ell^{\mathrm{al}}}$, well-defined up to an inner automorphism of $I_0$ over $\mathbb{Q}_\ell^{\mathrm{al}}$. Choose a system $(I,a,(j_\ell))$ consisting of a $\mathbb{Q}$-group $I$, an inner twisting $a\colon I_0\to I$ (isomorphism over $\mathbb{Q}^{\mathrm{al}}$), and isomorphisms $j_\ell\colon I_{\mathbb{Q}_\ell}\to I_\ell$ over $\mathbb{Q}_\ell$ for all $\ell$, unramified for almost all $\ell$, such that $j_\ell\circ a$ and $a_\ell$ differ by an inner automorphism — our assumption

(*) guarantees the existence of such a system. Moreover, any other such system is isomorphic to one of the form $(I, a, (j_\ell \circ \mathrm{ad} h_\ell))$ where $(h_\ell) \in I^{\mathrm{ad}}(\mathbb{A})$.

Let $dx$ denote the Haar measure on $G(\mathbb{A}_f^p)$ giving measure 1 to $K^p$. Choose a Haar measure $di^p$ on $I(\mathbb{A}_f^p)$ that gives rational measure to compact open subgroups of $I(\mathbb{A}_f^p)$, and use the isomorphisms $j_\ell$ to transport it to a measure on $G(\mathbb{A}_f^p)_\gamma$ (the centralizer of $\gamma$ in $G(\mathbb{A}_f^p)$). The resulting measure does not change if $(j_\ell)$ is modified by an element of $I^{\mathrm{ad}}(\mathbb{A})$. Write $d\overline{x}$ for the quotient of $dx$ by $di^p$. Let $f$ be an element of the Hecke algebra $\mathcal{H}$ of locally constant $K$-bi-invariant $\mathbb{Q}$-valued functions on $G(\mathbb{A}_f)$, and assume that $f = f^p \cdot f_p$ where $f^p$ is a function on $G(\mathbb{A}_f^p)$ and $f_p$ is the characteristic function of $K_p$ in $G(\mathbb{Q}_p)$ divided by the measure of $K_p$. Define

$$O_\gamma(f^p) = \int_{G(\mathbb{A}_f^p)_\gamma \backslash G(\mathbb{A}_f^p)} f^p(x^{-1}\gamma x)\ d\overline{x}$$

Let $dy$ denote the Haar measure on $G(L_n)$ giving measure 1 to $G(\mathcal{O}_{L_n})$. Choose a Haar measure $di_p$ on $I(\mathbb{Q}_p)$ that gives rational measure to the compact open subgroups, and use $j_p$ to transport the measure to $I_p(\mathbb{Q}_p)$. Again the resulting measure does not change if $j_p$ is modified by an element of $I^{\mathrm{ad}}(\mathbb{Q}_p)$. Write $d\overline{y}$ for the quotient of $dy$ by $di_p$. Proceeding as on p370, we choose a cocharacter $\mu$ in $c(X)$ well-adapted to the hyperspecial subgroup $K_p$ and defined over $L_n$, and we let $\varphi$ be the characteristic function of the coset $G(\mathcal{O}_{L_n}) \cdot \mu(p) \cdot G(\mathcal{O}_{L_n})$. Define

$$TO_\delta(\varphi) = \int_{I(\mathbb{Q}_p)\backslash G(L_n)} \varphi(y^{-1}\delta\sigma(y))d\overline{y}$$

Since $I/Z(G)$ is anisotropic over $\mathbb{R}$, and since we are assuming SV5, $I(\mathbb{Q})$ is a discrete subgroup of $I(\mathbb{A}_f^p)$, and we can define the volume of $I(\mathbb{Q})\backslash I(\mathbb{A}_f)$. It is a rational number because of our assumption on $di^p$ and $di_p$. Finally, define

$$I(\gamma_0;\gamma,\delta) = I(\gamma_0;\gamma,\delta)(f^p,r) = \mathrm{vol}(I(\mathbb{Q})\backslash I(\mathbb{A}_f)) \cdot O_\gamma(f^p) \cdot TO_\delta(\phi_r).$$

The integral $I(\gamma_0;\gamma,\delta)$ is independent of the choices made, and

$$(\gamma_0;\gamma,\delta) \sim (\gamma_0';\gamma',\delta') \implies I(\gamma_0;\gamma,\delta) = I(\gamma_0';\gamma',\delta').$$

**The triple attached to an admissible pair $(\phi,\varepsilon)$.** An ***admissible pair*** $(\phi,\gamma_0)$ is an admissible homomorphism $\phi\colon \mathfrak{P} \to E_G$ and a $\gamma \in I_\phi(\mathbb{Q})$ such that $\gamma_0 x = \Phi^r x$ for some $x \in X_p(\phi)$. Here $r = [k(\mathfrak{p})\colon \mathbb{F}_p]$. An ***isomorphism*** $(\phi,\gamma_0) \to (\phi',\gamma_0')$ of admissible pairs is an isomorphism $\phi \to \phi'$ sending $\gamma$ to $\gamma'$, i.e., it is a $g \in G(\mathbb{Q}^{\mathrm{al}})$ such that

$$\mathrm{ad}(g)\circ\phi = \phi', \quad \mathrm{ad}(g)(\gamma) = \gamma'.$$

Let $(T,x) \subset (G,X)$ be a special pair. Because of our assumptions on $(G,X)$, the cocharacter $\mu_x$ of $T$ satisfies the conditions of (15.15) and so defines a homomorphism $\phi_x\colon \mathfrak{P} \to E_T$. Langlands and Rapoport (1987, 5.23) show that every admissible pair is isomorphic to a pair $(\phi,\gamma)$ with $\phi = \phi_x$ and $\gamma \in T(\mathbb{Q})$. For such a pair $(\phi,\gamma)$, $b(\phi)$ is represented by a $\delta \in G(L_n)$ which is well-defined up to conjugacy.

Let $\gamma$ be the image of $\gamma_0$ in $G(\mathbb{A}_f^p)$. Then the triple $(\gamma_0;\gamma,\delta)$ satisfies the conditions in the last subsection. A triple arising in this way from an admissible pair will be called ***effective.***

**The formula.** For a triple $(\gamma_0 \dots)$, the kernel of

$$H^1(\mathbb{Q}, I_0) \to H^1(\mathbb{Q}, G) \oplus \prod_l H^1(\mathbb{Q}_l, I_0)$$

is finite — we denote its order by $c(\gamma_0)$.

THEOREM 17.1. *Let $(G, X)$ be a Shimura datum satisfying the hypotheses of (16.8). If that conjecture is true, then*

$$\# \operatorname{Sh}_p(\mathbb{F}_q) = \sum_{(\gamma_0;\gamma,\delta)} c(\gamma_0) \cdot I(\gamma_0; \gamma, \delta) \tag{64}$$

*where the sum is over a set of representatives for the effective triples.*

PROOF. See Milne 1992, 6.13. □

NOTES. Early versions of (64) can be found in papers of Langlands, but the first precise general statement of such a formula is in Kottwitz 1990. There Kottwitz attaches a cohomological invariant $\alpha(\gamma_0; \gamma, \delta)$ to a triple $(\gamma_0; \gamma, \delta)$, and conjectures that the formula (64) holds if the sum is taken over a set of representatives for the triples with $\alpha = 1$ (ibid. §3). Milne (1992, 7.9) proves that, among triples contributing to the sum, $\alpha = 1$ if and only if the triple is effective, and so the conjecture of Langlands and Rapoport implies Kottwitz's conjecture.[20] On the other hand, Kottwitz (1992) proves his conjecture for Shimura varieties of simple PEL type A or C unconditionally (without however proving the conjecture of Langlands and Rapoport for these varieties).

### References

Ash, A., Mumford, D., Rapoport, M. and Tai, Y. (1975), *Smooth compactification of locally symmetric varieties*, Math. Sci. Press, Brookline, Mass.

Baily, Jr., W. L. and Borel, A. (1966), 'Compactification of arithmetic quotients of bounded symmetric domains', *Ann. of Math. (2)* **84**, 442–528.

Baily, W. L. (1958), 'Satake's compactification of $V_n$', *Amer. J. Math.* **80**, 348–364.

Boothby, W. M. (1975), *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, New York-London.

Borel, A. (1962), Ensembles fondamentaux pour les groupes arithmétiques, *in* 'Colloq. Théorie des Groupes Algébriques (Bruxelles, 1962)', Librairie Universitaire, Louvain, pp. 23–40.

Borel, A. (1969), *Introduction aux groupes arithmétiques*, Publications de l'Institut de Mathématique de l'Université de Strasbourg, XV. Actualités Scientifiques et Industrielles, No. 1341, Hermann, Paris.

Borel, A. (1972), 'Some metric properties of arithmetic quotients of symmetric spaces and an extension theorem', *J. Differential Geometry* **6**, 543–560.

Borel, A. (1991), *Linear algebraic groups*, Vol. 126 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Borel, A. (1998), *Semisimple groups and Riemannian symmetric spaces*, Vol. 16 of *Texts and Readings in Mathematics*, Hindustan Book Agency, New Delhi.

Borel, A. and Harish-Chandra (1962), 'Arithmetic subgroups of algebraic groups', *Ann. of Math. (2)* **75**, 485–535.

Borovoĭ, M. V. (1984), 'Langlands' conjecture concerning conjugation of connected Shimura varieties', *Selecta Math. Soviet.* **3**(1), 3–39.

Borovoĭ, M. V. (1987), The group of points of a semisimple group over a totally real-closed field, *in* 'Problems in group theory and homological algebra (Russian)', Matematika, Yaroslav. Gos. Univ., pp. 142–149.

[20]At least in the case that the weight is rational — Kottwitz does not make this assumption.

Bourbaki, N. (1958), *Éléments de mathématique. I: Les structures fondamentales de l'analyse. Fascicule VII. Livre II: Algèbre. Chapitre 3: Algèbre multilinéaire*, Nouvelle édition. Actualités Scientifiques et Industrielles, No. 1044, Hermann, Paris.

Bourbaki, N. (1981), *Groupes et Algèbres de Lie. Chapitres 4,5 et 6*, Elements of Mathematics, Masson, Paris.

Bourbaki, N. (1989), *General topology. Chapters 1–4*, Elements of Mathematics, Springer-Verlag, Berlin.

Bravo, A., Encinas, S. and Villamayor, O. (2002), 'A Simplified Proof of Desingularization and Applications'. arXiv:math.AG/0206244.

Brylinski, J.-L. (1983), "1-motifs" et formes automorphes (théorie arithmétique des domaines de Siegel), *in* 'Conference on automorphic theory (Dijon, 1981)', Vol. 15 of *Publ. Math. Univ. Paris VII*, Univ. Paris VII, Paris, pp. 43–106.

Cartan, H. (1958), 'Prolongement des espaces analytiques normaux', *Math. Ann.* **136**, 97–110.

Cartan, H. (1963), *Elementary theory of analytic functions of one or several complex variables*, Éditions Scientifiques Hermann, Paris.

Chow, W.-L. (1949), 'On compact complex analytic varieties', *Amer. J. Math.* **71**, 893–914.

Deligne, P. (1970), *Équations différentielles à points singuliers réguliers*, Springer-Verlag, Berlin.

Deligne, P. (1971*a*), 'Théorie de Hodge. II', *Inst. Hautes Études Sci. Publ. Math.* **40**, 5–57.

Deligne, P. (1971*b*), Travaux de Shimura, *in* 'Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389', Springer, Berlin, pp. 123–165. Lecture Notes in Math., Vol. 244.

Deligne, P. (1971*c*), 'Travaux de Shimura'. Preliminary unpublished version of Deligne 1971b.

Deligne, P. (1973*a*), *Espaces hermitiens symétriques, Spring 1973*, Transcription available at www.jmilne.org/math/.

Deligne, P. (1973*b*), Les constantes des équations fonctionnelles des fonctions $L$, *in* 'Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)', Springer, Berlin, pp. 501–597. Lecture Notes in Math., Vol. 349.

Deligne, P. (1979), Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques, *in* 'Automorphic forms, representations and $L$-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2', Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., pp. 247–289.

Deligne, P. (1982), Hodge cycles on abelian varieties (notes by J.S. Milne), *in* 'Hodge cycles, motives, and Shimura varieties', Vol. 900 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, pp. 9–100. Available at www.jmilne.org/math/.

Demazure, M. (1972), *Lectures on p-divisible groups*, Springer-Verlag, Berlin.

Farb, B. and Dennis, R. K. (1993), *Noncommutative algebra*, Vol. 144 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Giraud, J. (1968), 'Remarque sur une formule de Shimura-Taniyama', *Invent. Math.* **5**, 231–236.

Goresky, M. (2003), 'Compactifications of modular varieties'. Preliminary version of article for this volume.

Greenberg, M. J. (1967), *Lectures on algebraic topology*, W. A. Benjamin, Inc., New York-Amsterdam.

Hartshorne, R. (1977), *Algebraic geometry*, Springer-Verlag, New York.

Hatcher, A. (2002), *Algebraic topology*, Cambridge University Press, Cambridge. Available at http://www.math.cornell.edu/ hatcher/.

Helgason, S. (1978), *Differential geometry, Lie groups, and symmetric spaces*, Vol. 80 of *Pure and Applied Mathematics*, Academic Press Inc., New York.

Hironaka, H. (1964), 'Resolution of singularities of an algebraic variety over a field of characteristic zero. I, II', *Ann. of Math. (2) 79 (1964), 109–203; ibid. (2)* **79**, 205–326.

Humphreys, J. E. (1972), *Introduction to Lie algebras and representation theory*, Springer-Verlag, New York.

Humphreys, J. E. (1978), *Introduction to Lie algebras and representation theory*, Vol. 9 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Ihara, Y. (1968), *On congruence monodromy problems. Vol. 1*, Lecture Notes, No. 1, Department of Mathematics, University of Tokyo, Tokyo.

Iitaka, S. (1982), *Algebraic geometry*, Vol. 76 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Knus, M.-A., Merkurjev, A., Rost, M. and Tignol, J.-P. (1998), *The book of involutions*, Vol. 44 of *American Mathematical Society Colloquium Publications*, American Mathematical Society, Providence, RI.

Kottwitz, R. E. (1984), 'Shimura varieties and twisted orbital integrals', *Math. Ann.* **269**(3), 287–300.

Kottwitz, R. E. (1990), Shimura varieties and $\lambda$-adic representations, *in* 'Automorphic forms, Shimura varieties, and $L$-functions, Vol. I (Ann Arbor, MI, 1988)', Vol. 10 of *Perspect. Math.*, Academic Press, Boston, MA, pp. 161–209.

Kottwitz, R. E. (1992), 'Points on some Shimura varieties over finite fields', *J. Amer. Math. Soc.* **5**(2), 373–444.

Krantz, S. G. (1982), *Function theory of several complex variables*, John Wiley & Sons Inc., New York.

Kwack, M. H. (1969), 'Generalization of the big Picard theorem', *Ann. of Math. (2)* **90**, 9–22.

Langlands, R. P. (1976), Some contemporary problems with origins in the Jugendtraum, *in* 'Mathematical developments arising from Hilbert problems (Proc. Sympos. Pure Math., Vol. XXVIII, Northern Illinois Univ., De Kalb, Ill., 1974)', Amer. Math. Soc., Providence, R. I., pp. 401–418.

Langlands, R. P. (1977), 'Shimura varieties and the Selberg trace formula', *Canad. J. Math.* **29**(6), 1292–1299.

Langlands, R. P. and Rapoport, M. (1987), 'Shimuravarietäten und Gerben', *J. Reine Angew. Math.* **378**, 113–220.

Lee, J. M. (1997), *Riemannian manifolds*, Vol. 176 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Mac Lane, S. (1998), *Categories for the working mathematician*, Vol. 5 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York.

Matsumoto, H. (1969), 'Sur les sous-groupes arithmétiques des groupes semi-simples déployés', *Ann. Sci. École Norm. Sup. (4)* **2**, 1–62.

Milne, J. S. (1979), Complex multiplication for pedestrians. September 19, 1979; revised June 7, 1998, available at www.jmilne.org/math/.

Milne, J. S. (1983), The action of an automorphism of $\mathbf{C}$ on a Shimura variety and its special points, *in* 'Arithmetic and geometry, Vol. I', Vol. 35 of *Progr. Math.*, Birkhäuser Boston, Boston, MA, pp. 239–265.

Milne, J. S. (1986), Abelian varieties, *in* 'Arithmetic geometry (Storrs, Conn., 1984)', Springer, New York, pp. 103–150.

Milne, J. S. (1987), The (failure of the) Hasse principle for centres of semisimple groups. 6th June, 1987, available at www.jmilne.org/math/.

Milne, J. S. (1992), The points on a Shimura variety modulo a prime of good reduction, *in* 'The zeta functions of Picard modular surfaces', Univ. Montréal, Montreal, QC, pp. 151–253.

Milne, J. S. (1994), Shimura varieties and motives, *in* 'Motives (Seattle, WA, 1991)', Vol. 55 of *Proc. Sympos. Pure Math.*, Amer. Math. Soc., Providence, RI, pp. 447–523.

Milne, J. S. (1999), 'Descent for Shimura varieties', *Michigan Math. J.* **46**(1), 203–208.

Milne, J. S. (2003), 'Gerbes and abelian motives'. Preprint available at www.jmilne.org/math/.

Milne, J. S. and Shih, K.-y. (1981), 'Automorphism groups of Shimura varieties and reciprocity laws', *Amer. J. Math.* **103**(5), 911–935.

Mok, N. and Zhong, J. Q. (1989), 'Compactifying complete Kähler-Einstein manifolds of finite topological type and bounded curvature', *Ann. of Math. (2)* **129**(3), 427–470.

Mumford, D. (1965), *Geometric invariant theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Band 34, Springer-Verlag, Berlin.

Mumford, D. (1969), 'A note of Shimura's paper "Discontinuous groups and abelian varieties"', *Math. Ann.* **181**, 345–351.

Mumford, D. (1970), *Abelian varieties*, Tata Institute of Fundamental Research Studies in Mathematics, No. 5, Published for the Tata Institute of Fundamental Research, Bombay.

Mumford, D. (1977), 'Hirzebruch's proportionality theorem in the noncompact case', *Invent. Math.* **42**, 239–272.

Murty, V. K. (1993), *Introduction to abelian varieties*, Vol. 3 of *CRM Monograph Series*, American Mathematical Society, Providence, RI.

Nadel, A. and Tsuji, H. (1988), 'Compactification of complete Kähler manifolds of negative Ricci curvature', *J. Differential Geom.* **28**(3), 503–512.

Platonov, V. and Rapinchuk, A. (1994), *Algebraic groups and number theory*, Vol. 139 of *Pure and Applied Mathematics*, Academic Press Inc., Boston, MA.

Raghunathan, M. S. (1972), *Discrete subgroups of Lie groups*, Springer-Verlag, New York.

Satake, I. (1956), 'On the compactification of the Siegel space', *J. Indian Math. Soc. (N.S.)* **20**, 259–281.

Satake, I. (1960), 'On compactifications of the quotient spaces for arithmetically defined discontinuous groups', *Ann. of Math. (2)* **72**, 555–580.

Satake, I. (1980), *Algebraic structures of symmetric domains*, Vol. 4 of *Kanô Memorial Lectures*, Iwanami Shoten, Tokyo.

Satake, I. (2001), 'Compactifications, old and new [translation of Sūgaku **51** (1999), no. 2, 129–141; MR 2000j:32035]', *Sugaku Expositions* **14**(2), 175–189.

Satake, I. and Cartan, H. (1958), Démonstration du théorème fondamental, *in* 'Séminaire Henri Cartan; 10e année: 1957/58. Fonctions Automorphes, Exposé 15', Secrétariat mathématique, ENS, Paris, p. 12.

Serre, J.-P. (1964), 'Sur les groupes de congruence des variétés abéliennes', *Izv. Akad. Nauk SSSR Ser. Mat.* **28**, 3–20.

Serre, J.-P. (1968), *Abelian l-adic representations and elliptic curves*, McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute, W. A. Benjamin, Inc., New York-Amsterdam.

Serre, J.-P. and Tate, J. (1968), 'Good reduction of abelian varieties', *Ann. of Math. (2)* **88**, 492–517.

Shafarevich, I. R. (1994), *Basic algebraic geometry. 1,2*, Springer-Verlag, Berlin.

Shimura, G. (1963), 'On analytic families of polarized abelian varieties and automorphic functions', *Ann. of Math. (2)* **78**, 149–192.

Shimura, G. (1971), On arithmetic automorphic functions, *in* 'Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2', Gauthier-Villars, Paris, pp. 343–348.

Shimura, G. and Taniyama, Y. (1961), *Complex multiplication of abelian varieties and its applications to number theory*, Vol. 6 of *Publications of the Mathematical Society of Japan*, The Mathematical Society of Japan, Tokyo.

Siegel, C. L. (1939), 'Einführung in die Theorie der Modulfunktionen $n$-ten Grades', *Math. Ann.* **116**, 617–657.

Springer, T. A. (1998), *Linear algebraic groups*, Vol. 9 of *Progress in Mathematics*, Birkhäuser Boston Inc., Boston, MA.

Tate, J. T. (1967), Global class field theory, *in* 'Algebraic Number Theory (Proc. Instructional Conf., Brighton, 1965)', Thompson, Washington, D.C., pp. 162–203.

Tate, J. T. (1969), 'Classes d'isogénie des variétés abéliennes sur un corps fini'. Séminaire Bourbaki.

Taylor, J. L. (2002), *Several complex variables with connections to algebraic geometry and Lie groups*, Vol. 46 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI.

Vasiu, A. (1999), 'Integral canonical models of Shimura varieties of preabelian type', *Asian J. Math.* **3**(2), 401–518.

Voisin, C. (2002), *Théorie de Hodge et géométrie algébrique complexe*, Vol. 10 of *Cours Spécialisés*, Société Mathématique de France, Paris. English translation available from Cambridge University Press.

Waterhouse, W. C. (1979), *Introduction to affine group schemes*, Vol. 66 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Weil, A. (1948), *Variétés abéliennes et courbes algébriques*, Actualités Sci. Ind., no. 1064 = Publ. Inst. Math. Univ. Strasbourg 8 (1946), Hermann & Cie., Paris.

Weil, A. (1956), 'The field of definition of a variety', *Amer. J. Math.* **78**, 509–524.

Weil, A. (1958), *Introduction à l'étude des variétés kählériennes*, Publications de l'Institut de Mathématique de l'Université de Nancago, VI. Actualités Sci. Ind. no. 1267, Hermann, Paris.

Wells, R. O. (1980), *Differential analysis on complex manifolds*, Vol. 65 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

Whitney, H. (1957), 'Elementary structure of real algebraic varieties', *Ann. of Math. (2)* **66**, 545–556.

Witte, D. (2001), *Introduction to Arithmetic Groups*, arXiv:math.DG/0106063.

Wolf, J. A. (1984), *Spaces of constant curvature*, Publish or Perish Inc., Houston, TX.

Zink, T. (1983), 'Isogenieklassen von Punkten von Shimuramannigfaltigkeiten mit Werten in einem endlichen Körper', *Math. Nachr.* **112**, 103–124.

*E-mail address*: `math@jmilne.org`

Clay Mathematics Proceedings
Volume 4, 2005

# Linear Algebraic Groups

Fiona Murnaghan

Abstract. We give a summary, without proofs, of basic properties of linear algebraic groups, with particular emphasis on reductive algebraic groups.

## 1. Algebraic groups

Let $K$ be an algebraically closed field. An *algebraic $K$-group* $\mathbf{G}$ is an algebraic variety over $K$, and a group, such that the maps $\mu : \mathbf{G} \times \mathbf{G} \to \mathbf{G}$, $\mu(x, y) = xy$, and $\iota : \mathbf{G} \to \mathbf{G}$, $\iota(x) = x^{-1}$, are morphisms of algebraic varieties. For convenience, in these notes, we will fix $K$ and refer to an algebraic $K$-group as an algebraic group. If the variety $\mathbf{G}$ is affine, that is, $\mathbf{G}$ is an algebraic set (a Zariski-closed set) in $K^n$ for some natural number $n$, we say that $\mathbf{G}$ is a *linear algebraic group.* If $\mathbf{G}$ and $\mathbf{G}'$ are algebraic groups, a map $\varphi : \mathbf{G} \to \mathbf{G}'$ is a *homomorphism of algebraic groups* if $\varphi$ is a morphism of varieties and a group homomorphism. Similarly, $\varphi$ is an *isomorphism of algebraic groups* if $\varphi$ is an isomorphism of varieties and a group isomorphism.

A closed subgroup of an algebraic group is an algebraic group. If $\mathbf{H}$ is a closed subgroup of a linear algebraic group $\mathbf{G}$, then $\mathbf{G}/\mathbf{H}$ can be made into a quasi-projective variety (a variety which is a locally closed subset of some projective space). If $\mathbf{H}$ is normal in $\mathbf{G}$, then $\mathbf{G}/\mathbf{H}$ (with the usual group structure) is a linear algebraic group.

Let $\varphi : \mathbf{G} \to \mathbf{G}'$ be a homomorphism of algebraic groups. Then the kernel of $\varphi$ is a closed subgroup of $\mathbf{G}$ and the image of $\varphi$ is a closed subgroup of $\mathbf{G}$.

Let $X$ be an affine algebraic variety over $K$, with affine algebra (coordinate ring) $K[X] = K[x_1, \ldots, x_n]/I$. If $k$ is a subfield of $K$, we say that $X$ is *defined over* $k$ if the ideal $I$ is generated by polynomials in $k[x_1, \ldots, x_n]$, that is, $I$ is generated by $I_k := I \cap k[x_1, \ldots, x_n]$. In this case, the $k$-subalgebra $k[X] := k[x_1, \ldots, x_n]/I_k$ of $K[X]$ is called a *$k$-structure* on $X$, and $K[X] = k[X] \otimes_k K$. If $X$ and $X'$ are algebraic varieties defined over $k$, a morphism $\varphi : X \to X'$ is *defined over $k$* (or is a *$k$-morphism*) if there is a homomorphism $\varphi_k^* : k[X'] \to k[X]$ such that the algebra homomorphism $\varphi^* : K[X'] \to K[X]$ defining $\varphi$ is $\varphi_k^* \times id$. Equivalently, the coordinate functions of $\varphi$ all have coefficients in $k$. The set $X(k) := X \cap k^n$ is called the *$K$-rational* points of $X$.

2000 *Mathematics Subject Classification.* 20G15.

If $k$ is a subfield of $K$, we say that a linear algebraic group $\mathbf{G}$ is *defined over $k$* (or is a *$k$-group*) if the variety $\mathbf{G}$ is defined over $k$ and the homomorphisms $\mu$ and $\iota$ are defined over $k$. Let $\varphi : \mathbf{G} \to \mathbf{G}'$ be a $k$-homomorphism of $k$-groups. Then the image of $\varphi$ is defined over $k$ but the kernel of $\varphi$ might not be defined over $k$.

An algebraic variety $X$ over $K$ is *irreducible* if it cannot be expressed as the union of two proper closed subsets. Any algebraic variety $X$ over $K$ can be expressed as the union of finitely many irreducible closed subsets:

$$X = X_1 \cup X_2 \cup \cdots \cup X_r,$$

where $X_i \not\subset X_j$ if $j \neq i$. This decomposition is unique and the $X_i$ are the maximal irreducible subsets of $X$ (relative to inclusion). The $X_i$ are called the *irreducible components* of $X$.

Let $\mathbf{G}$ be an algebraic group. Then $\mathbf{G}$ has a unique irreducible component $\mathbf{G}^0$ containing the identity element. The irreducible component $\mathbf{G}^0$ is a closed normal subgroup of $\mathbf{G}$. The cosets of $\mathbf{G}^0$ in $\mathbf{G}$ are the irreducible components of $\mathbf{G}$, and $\mathbf{G}^0$ is the connected component of the identity in $\mathbf{G}$. Also, if $\mathbf{H}$ is a closed subgroup of $\mathbf{G}$ of finite index in $\mathbf{G}$, then $\mathbf{H} \supset \mathbf{G}^0$. For a linear algebraic group, connectedness is equivalent to irreducibility. It is usual to refer to an irreducible algebraic group as a connected algebraic group.

If $\varphi : \mathbf{G} \to \mathbf{G}'$ is a homomorphism of algebraic groups, then $\varphi(\mathbf{G}^0) = \varphi(\mathbf{G})^0$. If $k$ is a subfield of $K$ and $\mathbf{G}$ is defined over $k$, then $\mathbf{G}^0$ is defined over $k$.

The *dimension* of $\mathbf{G}$ is the dimension of the variety $\mathbf{G}^0$. That is, the dimension of $\mathbf{G}$ is the transcendence degree of the field $K(\mathbf{G}^0)$ over $K$.

If $\mathbf{G}$ is a linear algebraic group, then $\mathbf{G}$ is isomorphic, as an algebraic group, to a closed subgroup of $\mathbf{GL}_n(K)$ for some natural number $n$.

Example 1.1. $\mathbf{G} = K$, with $\mu(x, y) = x+y$ and $\iota(x) = -x$. The usual notation for this group is $\mathbf{G}_a$. It is connected and has dimension 1.

Example 1.2. Let $n$ be a positive integer and let $M_n(K)$ be the set of $n \times n$ matrices with entries in $K$. The general linear group $\mathbf{G} = \mathbf{GL}_n(K)$ is the group of matrices in $M_n(K)$ that have nonzero determinant. Note that $\mathbf{G}$ can be identified with the closed subset $\{(g, x) \mid g \in M_n(K),\ x \in K,\ (\det g)x = 1\}$ of $K^{n^2} \times K = K^{n^2+1}$. Then $K[\mathbf{G}] = K[\,x_{ij},\, 1 \le i, j \le n,\, \det(x_{ij})^{-1}]$. The dimension of $\mathbf{GL}_n(K)$ is $n^2$, and it is connected. In the case $n = 1$, the usual notation for $\mathbf{GL}_1(K)$ is $\mathbf{G}_m$. The only connected algebraic groups of dimension 1 are $\mathbf{G}_a$ and $\mathbf{G}_m$.

Example 1.3. Let $n$ be a positive integer and let $I_n$ be the $n \times n$ identity matrix. The $2n \times 2n$ matrix $J = \left[\begin{smallmatrix} 0 & I_n \\ -I_n & 0 \end{smallmatrix}\right]$ is invertible and satisfies ${}^tJ = -J$, where ${}^tJ$ denotes the transpose of $J$. The $2n \times 2n$ symplectic group $\mathbf{G} = \mathbf{Sp}_{2n}(K)$ is defined by $\{\, g \in M_{2n}(K) \mid {}^tgJg = J \,\}$.

## 2. Jordan decomposition in linear algebraic groups

Recall that a matrix $x \in M_n(K)$ is *semisimple* if $x$ is diagonalizable: there is a $g \in \mathbf{GL}_n(K)$ such that $gxg^{-1}$ is a diagonal matrix. Also, $x$ is unipotent if $x - I_n$ is nilpotent: $(x - I_n)^k = 0$ for some natural number $k$. Given $x \in \mathbf{GL}_n(K)$, there exist elements $x_s$ and $x_u$ in $\mathbf{GL}_n(K)$ such that $x_s$ is semisimple, $x_u$ is unipotent, and $x = x_s x_u = x_u x_s$. Furthermore, $x_s$ and $x_u$ are uniquely determined.

Now suppose that $\mathbf{G}$ is a linear algebraic group. Choose $n$ and an injective homomorphism $\varphi : \mathbf{G} \to \mathbf{GL}_n(K)$ of algebraic groups. If $g \in \mathbf{G}$, the semisimple and unipotent parts $\varphi(g)_s$ and $\varphi(g)_u$ of $\varphi(g)$ lie in $\varphi(\mathbf{G})$, and the elements $g_s$ and $g_u$ such that $\varphi(g_s) = \varphi(g)_s$ and $\varphi(g_u) = \varphi(g)_u$ depend only on $g$ and not on the choice of $\varphi$ (or $n$). The elements $g_s$ and $g_u$ are called the semisimple and unipotent part of $g$, respectively. An element $g \in \mathbf{G}$ is *semisimple* if $g = g_s$ (and $g_u = 1$), and *unipotent* if $g = g_u$ (and $g_s = 1$).

*Jordan decomposition.*

(1) If $g \in \mathbf{G}$, there exist elements $g_s$ and $g_u$ in $\mathbf{G}$ such that $g = g_s g_u = g_u g_s$, $g_s$ is semisimple, and $g_u$ is unipotent. Furthermore, $g_s$ and $g_u$ are uniquely determined by the above conditions.
(2) If $k$ is a perfect subfield of $K$ and $\mathbf{G}$ is a $k$-group, then $g \in \mathbf{G}(k)$ implies $g_s,\ g_u \in \mathbf{G}(k)$.

Jordan decompositions are preserved by homomorphisms of algebraic groups. Suppose that $\mathbf{G}$ and $\mathbf{G}'$ are linear algebraic groups and $\varphi : \mathbf{G} \to \mathbf{G}'$ is a homomorphism of linear algebraic groups. Let $g \in \mathbf{G}$. Then $\varphi(g)_s = \varphi(g_s)$ and $\varphi(g)_u = \varphi(g_u)$.

## 3. Lie algebras

Let $\mathbf{G}$ be a linear algebraic group. The tangent bundle $T(\mathbf{G})$ of $\mathbf{G}$ is the set $\mathrm{Hom}_{K-alg}(K[\mathbf{G}], K[t]/(t^2))$ of $K$-algebra homomorphisms from the affine algebra $K[\mathbf{G}]$ of $\mathbf{G}$ to the algebra $K[t]/(t^2)$. If $g \in \mathbf{G}$, the evaluation map $f \mapsto f(g)$ from $K[\mathbf{G}]$ to $K$ is a $K$-algebra isomorphism. This results in a bijection betweeen $\mathbf{G}$ and $\mathrm{Hom}_{K-alg}(K[\mathbf{G}], K)$. Composing elements of $T(\mathbf{G})$ with the map $a + bt + (t^2) \mapsto a$ from $K[t]/(t^2)$ to $K$ results in a map from $T(\mathbf{G})$ to $\mathbf{G} = \mathrm{Hom}_{K-alg}(K[\mathbf{G}], K)$. The tangent space $T_1(\mathbf{G})$ of $\mathbf{G}$ at the identity element 1 of $\mathbf{G}$ is the fibre of $T(\mathbf{G})$ over 1. If $X \in T_1(\mathbf{G})$ and $f \in K[\mathbf{G}]$, then $X(f) = f(1) + t\, d_X(f) + (t^2)$ for some $d_X(f) \in K$. This defines a map $d_X : K[\mathbf{G}] \to K$ which satisfies:

$$d_X(f_1 f_2) = d_X(f_1) f_2(1) + f_1(1) d_X(f_2), \qquad f_1,\ f_2 \in K[\mathbf{G}].$$

Let $\mu^* : K[\mathbf{G}] \to K[\mathbf{G}] \otimes_K K[\mathbf{G}]$ be the $K$-algebra homomorphism which corresponds to the multiplication map $\mu : \mathbf{G} \times \mathbf{G} \to \mathbf{G}$. Set $\delta_X = (1 \otimes d_X) \circ \mu^*$. The map $\delta_X : K[\mathbf{G}] \to K[\mathbf{G}]$ is a $K$-linear map and a *derivation*:

$$\delta_X(f_1 f_2) = \delta_X(f_1) f_2 + f_1 \delta_X(f_2), \qquad f_1,\ f_2 \in K[\mathbf{G}].$$

Furthermore, $\delta_X$ is left-invariant: $\ell_g \delta_X = \delta_X \ell_g$ for all $g \in \mathbf{G}$, where $(\ell_g f)(g') = f(g^{-1} g')$, $f \in K[\mathbf{G}]$. The map $X \mapsto \delta_X$ is a $K$-linear isomorphism of $T_1(\mathbf{G})$ onto the vector space of $K$-linear maps from $K[\mathbf{G}]$ to $K[\mathbf{G}]$ which are left-invariant derivations.

Let $\mathfrak{g} = T_1(\mathbf{G})$. Define $[X, Y] \in \mathfrak{g}$ by $\delta_{[X,Y]} = \delta_X \circ \delta_Y - \delta_Y \circ \delta_X$. Then $\mathfrak{g}$ is a vector space over $K$ and the map $[\cdot, \cdot]$ satisfies:

(1) $[\cdot, \cdot]$ is linear in both variables
(2) $[X, X] = 0$ for all $X \in \mathfrak{g}$
(3) $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ for all $X,\ Y,\ X \in \mathfrak{g}$. (*Jacobi identity*)

Therefore $\mathfrak{g}$ is a Lie algebra over $K$. We call it the Lie algebra of $\mathbf{G}$.

Example 3.1. If $\mathbf{G} = \mathbf{GL}_n(K)$, then $\mathfrak{g}$ is isomorphic to the Lie algebra $\mathfrak{gl}_n(K)$ which is $M_n(K)$ equipped with the Lie bracket $[X, Y] = XY - YX$, $X,\ Y \in M_n(K)$.

EXAMPLE 3.2. If $\mathbf{G} = \mathbf{Sp}_{2n}(K)$, then $\mathfrak{g}$ is isomorphic to the Lie algebra $\{ X \in M_{2n}(K) \mid {}^tXJ + JX = 0 \}$, with bracket $[X, Y] = XY - YX$.

Let $\varphi : \mathbf{G} \to \mathbf{G}'$ be a homomorphism of linear algebraic groups. Composition with the algebra homomorphism $\varphi^* : K[\mathbf{G}'] \to K[\mathbf{G}]$ results in a map $T(\varphi) : T(\mathbf{G}) \to T(\mathbf{G}')$. The *differential* $d\varphi$ of $\varphi$ is the restriction $d\varphi = T(\varphi)\,|_{\mathfrak{g}}$ of $T(\varphi)$ to $\mathfrak{g}$. It is a $K$-linear map from $\mathfrak{g}$ to $\mathfrak{g}'$, and satisfies

$$d\varphi([X, Y]) = [d\varphi(X), d\varphi(Y)], \qquad X,\, Y \in \mathfrak{g}.$$

That is, $d\varphi$ is a *homomorphism of Lie algebras.* If $\varphi$ is bijective, then $\varphi$ is an isomorphism if and only if $d\varphi$ is an isomorphism of Lie algebras. If $K$ has characteristic zero, any bijective homomorphism of linear algebraic groups is an isomorphism.

If $\mathbf{H}$ is a closed subgroup of a linear algebraic group $\mathbf{G}$, then (via the differential of inclusion) the Lie algebra $\mathfrak{h}$ of $\mathbf{H}$ is isomorphic to a Lie subalgebra of $\mathfrak{g}$. And $\mathbf{H}$ is a normal subgroup of $\mathbf{G}$ if and only if $\mathfrak{h}$ is an ideal in $\mathfrak{g}$ ($[X, Y] \in \mathfrak{h}$ whenever $X \in \mathfrak{g}$ and $Y \in \mathfrak{h}$).

If $g \in \mathbf{G}$, then $\mathrm{Int}_g : \mathbf{G} \to \mathbf{G}$, $\mathrm{Int}_g = gg_0g^{-1}$, $g_0 \in \mathbf{G}$, is an isomorphism of algebraic groups, so $\mathrm{Ad}\, g := d(\mathrm{Int}_g) : \mathfrak{g} \to \mathfrak{g}$ is an isomorphism of Lie algebras. Note that $(\mathrm{Ad}\, g)^{-1} = \mathrm{Ad}\, g^{-1}$, $g \in \mathbf{G}$, and $\mathrm{Ad}\,(g_1g_2) = \mathrm{Ad}\, g_1 \circ \mathrm{Ad}\, g_2$, $g_1,\, g_2 \in \mathbf{G}$. The map $\mathrm{Ad} : \mathbf{G} \to \mathbf{GL}(\mathfrak{g})$ is a homomorphism of algebraic groups, called the *adjoint representation* of $\mathbf{G}$.

If $\mathbf{G}$ is a $k$-group, then its Lie algebra $\mathfrak{g}$ has a natural $k$-structure $\mathfrak{g}(k)$, with $\mathfrak{g} \simeq K \otimes_k \mathfrak{g}(k)$. Also, Ad is defined over $k$.

*Jordan decomposition in the Lie algebra.* We can define semisimple and nilpotent elements in $\mathfrak{g}$ in manner analogous to definitions of semisimple and unipotent elements in $\mathbf{G}$ (as $\mathfrak{g}$ is isomorphic to a Lie subalgebra of $\mathfrak{gl}_n(K)$ for some $n$). If $X \in \mathfrak{g}$, there exist unique elements $X_s$ and $X_n \in \mathfrak{g}$ such that $X = X_s + X_n$, $[X_s, X_n] = 0$, $X_s$ is semisimple, and $X_n$ is nilpotent. If $\varphi : \mathbf{G} \to \mathbf{G}'$ is a homomorphism of algebraic groups, then $d\varphi(X)_s = d\varphi(X_s)$ and $d\varphi(X)_n = d\varphi(X_n)$ for all $X \in \mathfrak{g}$.

## 4. Tori

A *torus* is a linear algebraic group which is isomorphic to the direct product $\mathbf{G}_m^d = \mathbf{G}_m \times \cdots \times \mathbf{G}_m$ ($d$ times), where $d$ is a positive integer. A linear algebraic group $\mathbf{G}$ is a torus if and only if $\mathbf{G}$ is connected and abelian, and every element of $\mathbf{G}$ is semisimple.

A *character* of a torus $\mathbf{T}$ is a homomorphism of algebraic groups from $\mathbf{T}$ to $\mathbf{G}_m$. The product of two characters of $\mathbf{T}$ is a character of $\mathbf{T}$, the inverse of a character of $\mathbf{T}$ is a character of $\mathbf{T}$, and characters of $\mathbf{T}$ commute with each other, so the set $X(\mathbf{T})$ of characters of $\mathbf{T}$ is an abelian group. A *one-parameter* subgroup of $\mathbf{T}$ is a homomorphism of algebraic groups from $\mathbf{G}_m$ to $\mathbf{T}$. The set $Y(\mathbf{T})$ of one-parameter subgroups is an abelian group. If $\mathbf{T} \simeq \mathbf{G}_m$, then $X(\mathbf{T}) = Y(\mathbf{T})$ is just the set of maps $x \mapsto x^r$, as $r$ varies over $\mathbb{Z}$. In general, $\mathbf{T} \simeq \mathbf{G}_m^d$ for some positive integer $d$, so $X(\mathbf{T}) \simeq X(\mathbf{G}_m)^d \simeq \mathbb{Z}^d \simeq Y(\mathbf{T})$. We have a pairing

$$\begin{aligned} &\langle \cdot, \cdot \rangle : X(\mathbf{T}) \times Y(\mathbf{T}) \to \mathbb{Z} \\ &\langle \chi, \eta \rangle \mapsto r \text{ where } \chi \circ \eta(x) = x^r,\ x \in \mathbf{G}_m. \end{aligned}$$

Let $k$ be a subfield of $K$. A torus $\mathbf{T}$ is a *$k$-torus* if $\mathbf{T}$ is defined over $k$. Let $\mathbf{T}$ be a $k$-torus. Let $X(\mathbf{T})_k$ be the subgroup of $X(\mathbf{T})$ made up of those characters of $\mathbf{T}$ which are defined over $k$. We say that $\mathbf{T}$ is *$k$-split* (or *splits over $k$*) whenever $X(\mathbf{T})_k$ spans $k[\mathbf{T}]$, or, equivalently, whenever $\mathbf{T}$ is $k$-isomorphic to $\mathbf{G}_m \times \cdots \times \mathbf{G}_m$ ($d$ times, $d = \dim \mathbf{T}$). In this case, $\mathbf{T}(k) \simeq k^\times \times \cdots \times k^\times$. If $X(\mathbf{T})_k = 0$, then we say that $\mathbf{T}$ is *$k$-anisotropic*. There exists a finite Galois extension of $k$ over which $\mathbf{T}$ splits. There exist unique tori $\mathbf{T}_{spl}$ and $\mathbf{T}_{an}$ of $\mathbf{T}$, both defined over $k$, such that $\mathbf{T} = \mathbf{T}_{spl}\mathbf{T}_{an}$, $\mathbf{T}_{spl}$ is $k$-split and $\mathbf{T}_{an}$ is $k$-anisotropic. Also, $\mathbf{T}_{an}$ is the identity component of $\cap_{\chi \in X(\mathbf{T})_k} \ker \chi$.

Example 4.1. Let $\mathbf{T}$ be the subgroup of $\mathbf{GL}_n(K)$ consisting of diagonal matrices in $\mathbf{GL}_n(K)$. Then $\mathbf{T}$ is a $k$-split $k$-torus for any subfield $k$ of $K$.

Example 4.2. Let $\mathbf{T}$ be the closed subgroup of $\mathbf{GL}_2(\mathbb{C})$ defined by

$$\mathbf{T} = \left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \mid a,\, b \in \mathbb{C},\ a^2 + b^2 \neq 0 \right\}.$$

Then $\mathbf{T}$ is an $\mathbb{R}$-torus and is $\mathbb{R}$-anisotropic.

## 5. Reductive groups, root systems and root data–the absolute case

Let $\mathbf{G}$ be a linear algebraic group which contains at least one torus. Then the set of tori in $\mathbf{G}$ has maximal elements, relative to inclusion. Such maximal elements are called *maximal tori* of $\mathbf{G}$. All of the maximal tori in $\mathbf{G}$ are conjugate. The *rank* of $\mathbf{G}$ is defined to be the dimension of a maximal torus in $\mathbf{G}$.

Now suppose that $\mathbf{G}$ is a linear algebraic group and $\mathbf{T}$ is a torus in $\mathbf{G}$. Recall that the adjoint representation $\mathrm{Ad} : \mathbf{G} \to \mathbf{GL}(\mathfrak{g})$ is a homomorphism of algebraic groups. Therefore $\mathrm{Ad}\,(\mathbf{T})$ consists of commuting semisimple elements, and so is diagonalizable. Given $\alpha \in X(\mathbf{T})$, let $\mathfrak{g}_\alpha = \{\, X \in \mathfrak{g} \mid \mathrm{Ad}\,(t)X = \alpha(t)X,\ \forall\, t \in \mathbf{T}\,\}$. The nonzero $\alpha \in X(\mathbf{T})$ such that $\mathfrak{g}_\alpha \neq 0$ are the *roots* of $\mathbf{G}$ relative to $\mathbf{T}$. The set of roots of $\mathbf{G}$ relative to $\mathbf{T}$ will be denoted by $\Phi(\mathbf{G}, \mathbf{T})$.

The centralizer $Z_{\mathbf{G}}(\mathbf{T})$ of $\mathbf{T}$ in $\mathbf{G}$ is the identity component of the normalizer $N_{\mathbf{G}}(\mathbf{T})$ of $\mathbf{T}$ in $\mathbf{G}$. The *Weyl group* $W(\mathbf{G}, \mathbf{T})$ of $\mathbf{T}$ in $\mathbf{G}$ is the (finite) quotient $N_{\mathbf{G}}(\mathbf{T})/Z_{\mathbf{G}}(\mathbf{T})$. Because $W(\mathbf{G}, \mathbf{T})$ acts on $\mathbf{T}$, $W(\mathbf{G}, \mathbf{T})$ also acts on $X(\mathbf{T})$, and $W(\mathbf{G}, \mathbf{T})$ permutes the roots of $\mathbf{T}$ in $\mathbf{G}$. Since any two maximal tori in $\mathbf{G}$ are conjugate, their Weyl groups are isomorphic. The Weyl group of any maximal torus is referred to as the Weyl group of $\mathbf{G}$.

An algebraic group $\mathbf{G}$ contains a unique maximal normal solvable subgroup, and this subgroup is closed. Its identity component is called the *radical* of $\mathbf{G}$, written $R(\mathbf{G})$. The set $R_u(\mathbf{G})$ of unipotent elements in $R(\mathbf{G})$ is a normal closed subgroup of $\mathbf{G}$, and is called the *unipotent radical* of $\mathbf{G}$. If $\mathbf{G}$ is a linear algebraic group such that the radical $R(\mathbf{G}^0)$ of $\mathbf{G}^0$ is trivial, then $\mathbf{G}$ is *semisimple.* In fact, $\mathbf{G}$ is semisimple if and only if $\mathbf{G}$ has no nontrivial connected abelian normal subgroups. If $R_u(\mathbf{G}^0)$ is trivial, then $\mathbf{G}$ is *reductive.* The *semisimple rank* of $\mathbf{G}$ is defined to be the rank of $\mathbf{G}/R(\mathbf{G})$, and the *reductive rank* of $\mathbf{G}$ is the rank of $\mathbf{G}/R_u(\mathbf{G})$.

The derived group $\mathbf{G}_{der}$ of $\mathbf{G}$ is a closed subgroup of $\mathbf{G}$, and is connected when $\mathbf{G}$ is connected. Suppose that $\mathbf{G}$ is connected and reductive. Then

(1) $\mathbf{G}_{der}$ is semisimple.
(2) $R(\mathbf{G}) = Z(\mathbf{G})^0$, where $Z(\mathbf{G})$ is the centre of $\mathbf{G}$, and $R(\mathbf{G})$ is a torus.
(3) $R(\mathbf{G}) \cap \mathbf{G}_{der}$ is finite, and $\mathbf{G} = R(\mathbf{G})\mathbf{G}_{der}$.

For the rest of this section, assume that $\mathbf{G}$ is a connected reductive group. Let $\mathbf{T}$ be a torus in $\mathbf{G}$. Then $Z_{\mathbf{G}}(\mathbf{T})$ is reductive. This fact is useful for inductive arguments. Now assume that $\mathbf{T}$ is maximal. Let $\mathfrak{t}$ be the Lie algebra of $\mathbf{T}$ and let $\Phi = \Phi(\mathbf{G}, \mathbf{T})$. Then

(1) $\mathfrak{g} = \mathfrak{t} \oplus \bigoplus_{\alpha\in\Phi} \mathfrak{g}_\alpha$ and $\dim \mathfrak{g}_\alpha = 1$ for all $\alpha \in \Phi$.
(2) If $\alpha \in \Phi$, let $\mathbf{T}_\alpha = (\operatorname{Ker}\alpha)^0$. Then $\mathbf{T}_\alpha$ is a torus, of codimension one in $\mathbf{T}$.
(3) If $\alpha \in \Phi$, let $\mathbf{Z}_\alpha = Z_{\mathbf{G}}(\mathbf{T}_\alpha)$. Then $\mathbf{Z}_\alpha$ is a reductive group of semisimple rank 1, and the Lie algebra $\mathfrak{z}_\alpha$ of $\mathbf{Z}_\alpha$ satisfies $\mathfrak{z}_\alpha = \mathfrak{t} \oplus \mathfrak{g}_\alpha \oplus \mathfrak{g}_{-\alpha}$. The group $\mathbf{G}$ is generated by the subgroups $\mathbf{Z}_\alpha$, $\alpha \in \Phi$.
(4) The centre $Z(\mathbf{G})$ of $\mathbf{G}$ is equal to $\cap_{\alpha\in\Phi}\mathbf{T}_\alpha$.
(5) If $\alpha \in \Phi$, there exists a unique connected $\mathbf{T}$-stable (relative to conjugation by $\mathbf{T}$) subgroup $\mathbf{U}_\alpha$ of $\mathbf{G}$ having Lie algebra $\mathfrak{g}_\alpha$. Also, $\mathbf{U}_\alpha \subset \mathbf{Z}_\alpha$.
(6) Let $n \in N_{\mathbf{G}}(\mathbf{T})$, and let $w$ be the corresponding element of $W = W(\mathbf{G}, \mathbf{T})$. Then $n\mathbf{U}_\alpha n^{-1} = \mathbf{U}_{w(\alpha)}$ for all $\alpha \in \Phi$.
(7) Let $\alpha \in \Phi$. Then there exists an isomorphism $\varepsilon_\alpha : \mathbf{G}_a \to \mathbf{U}_\alpha$ such that $t\,\varepsilon_\alpha(x)t^{-1} = \varepsilon_\alpha(\alpha(t)x)$, $t \in \mathbf{T}$, $x \in \mathbf{G}_a$.
(8) The groups $\mathbf{U}_\alpha$, $\alpha \in \Phi$, together with $\mathbf{T}$, generate the group $\mathbf{G}$.

Let $\langle\Phi\rangle$ be the subgroup of $X(\mathbf{T})$ generated by $\Phi$ and let $V = \langle\Phi\rangle \otimes_{\mathbb{Z}} \mathbb{R}$. Then the set $\Phi$ is a subset of the vector space $V$ and is a root system. In general an *abstract root system* in a finite dimensional real vector space $V$, is a subset $\Phi$ of $V$ that satisfies the following axioms:

**(R1):** $\Phi$ is finite, $\Phi$ spans $V$, and $0 \notin \Phi$.
**(R2):** If $\alpha \in \Phi$, there exists a reflection $s_\alpha$ relative to $\alpha$ such that $s_\alpha(\Phi) \subset \Phi$. (A *reflection* relative to $\alpha$ is a linear transformation sending $\alpha$ to $-\alpha$ that restricts to the identity map on a subspace of codimension one).
**(R3):** If $\alpha$, $\beta \in \Phi$, then $s_\alpha(\beta) - \beta$ is an integer multiple of $\alpha$.

A root system is *reduced* if it has the property that if $\alpha \in \Phi$, then $\pm\alpha$ are the only multiples of $\alpha$ which belong to $\Phi$.

The *rank* of $\Phi$ is defined to be $\dim V$. The *abstract Weyl group* $W(\Phi)$ is the subgroup of $\mathbf{GL}(V)$ generated by the set $\{\, s_\alpha \mid \alpha \in \Phi \,\}$.

If $\mathbf{T}$ is a maximal torus in $\mathbf{G}$, then $\Phi = \Phi(\mathbf{G}, \mathbf{T})$ is a root system in $V = \langle\Phi\rangle \otimes_{\mathbb{Z}} \mathbb{R}$, and it is reduced. The rank of $\Phi$ is equal to the semisimple rank of $\mathbf{G}$, and the abstract Weyl group $W(\Phi)$ is isomorphic to $W = W(\mathbf{G}, \mathbf{T})$.

A *base* of $\Phi$ is a subset $\Delta = \{\, \alpha_1, \ldots, \alpha_\ell \}$, $\ell = \operatorname{rank}(\Phi)$, such that $\Delta$ is a basis of $V$ and each $\alpha \in \Phi$ is uniquely expressed in the form $\alpha = \sum_{i=1}^{\ell} c_i\alpha_i$, where the $c_i$'s are all integers, no two of which have different signs. The elements of $\Delta$ are called *simple roots*. The set of *positive roots* $\Phi^+$ is the set of $\alpha \in \Phi$ such that the coefficients of the simple roots in the expression for $\alpha$, as a linear combination of simple roots, are all nonnegative. Similarly, $\Phi^-$ consists of those $\alpha \in \Phi$ such that the coefficients are all nonpositive. Clearly $\Phi$ is the disjoint union of $\Phi^+$ and $\Phi^-$. Given $\alpha \in \Phi$, there exists a base containing $\alpha$. Given a base $\Delta$, the set $\{\, s_\alpha \mid \alpha \in \Delta \,\}$ generates $W = W(\Phi)$. The subgroups $\mathbf{Z}_\alpha$, $\alpha \in \Delta$, generate $\mathbf{G}$. Equivalently, the subgroups $\mathbf{U}_\alpha$, $\alpha \in \Delta$, and $\mathbf{T}$, generate $\mathbf{G}$.

There is an inner product $(\cdot,\cdot)$ on $V$ with respect to which each $w \in W$ is an orthogonal linear transformation. If $\alpha$, $\beta \in \Phi$, then $s_\alpha(\beta) = \beta - (2(\beta,\alpha)/(\alpha,\alpha))\alpha$. A *Weyl chamber* in $V$ is a connected component in the complement of the union of the hyperplanes orthogonal to the roots. The set of Weyl chambers in $V$ and the set of bases of $\Phi$ correspond in a natural way, and $W$ permutes each of them simply transitively.

If $\alpha \in \Phi$, there exists a unique $\alpha^\vee \in Y(\mathbf{T})$ such that $\langle \beta, \alpha^\vee \rangle = 2(\beta,\alpha)/(\alpha,\alpha)$ for all $\beta \in \Phi$. The set $\Phi^\vee$ of elements $\alpha^\vee$ (called co-roots) forms a root system in $\langle \Phi^\vee \rangle \otimes_{\mathbb{Z}} \mathbb{R}$, called the *dual* of $\Phi$. The Weyl group $W(\Phi^\vee)$ is isomorphic to $W(\Phi)$, via the map $s_\alpha \mapsto s_{\alpha^\vee}$.

A root system $\Phi$ is said to be *irreducible* if $\Phi$ cannot be expressed as the union of two mutually orthogonal proper subsets. In general, $\Phi$ can be partitioned uniquely into a union of irreducible root systems in subspaces of $V$. The group $\mathbf{G}$ is *simple* (or *almost simple*) if $\mathbf{G}$ contains no proper nontrivial closed connected normal subgroup. When $\mathbf{G}$ is semisimple and connected, then $\mathbf{G}$ is simple if and only if $\Phi$ is irreducible.

The reduced irreducible root systems are those of type $A_n$, $n \geq 1$, $B_n$, $n \geq 1$, $C_n$, $n \geq 3$, $D_n$, $n \geq 4$, $E_6$, $E_7$, $E_8$, $F_4$, and $G_2$. For each $n \geq 1$ there is one irreducible nonreduced root system, $BC_n$. (These root systems are described in many of the references). If $n \geq 2$, the root system of $\mathbf{GL}_n(K)$ (relative to any maximal torus) is of type $A_{n-1}$. The root system of $\mathbf{Sp}_{2n}(K)$ is of type $C_n$, if $n \geq 3$, and of type $A_1$ and $B_2$ for $n = 1$ and 2, respectively.

The quadruple $\Psi(\mathbf{G},\mathbf{T}) = (X, Y, \Phi, \Phi^\vee) = (X(\mathbf{T}), Y(\mathbf{T}), \Phi(\mathbf{G},\mathbf{T}), \Phi^\vee(\mathbf{G},\mathbf{T}))$ is a *root datum*. An abstract root datum is a quadruple $\Psi = (X, Y, \Phi, \Phi^\vee)$, where $X$ and $Y$ are free abelian groups such that there exists a bilinear mapping $\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{Z}$ inducing isomorphisms $X \simeq \mathrm{Hom}(Y, \mathbb{Z})$ and $Y \simeq \mathrm{Hom}(X, \mathbb{Z})$, and $\Phi \subset X$ and $\Phi^\vee \subset Y$ are finite subsets, and there exists a bijection $\alpha \mapsto \alpha^\vee$ of $\Phi$ onto $\Phi^\vee$. The following two axioms must be satisfied:

**(RD1):** $\langle \alpha, \alpha^\vee \rangle = 2$

**(RD2):** If $s_\alpha : X \to X$ and $s_{\alpha^\vee} : Y \to Y$ are defined by $s_\alpha(x) = x - \langle x, \alpha^\vee \rangle \alpha$ and $s_{\alpha^\vee}(y) = y - \langle \alpha, y \rangle \alpha^\vee$, then $s_\alpha(\Phi) \subset \Phi$ and $s_{\alpha^\vee}(\Phi^\vee) \subset \Phi^\vee$ (for all $\alpha \in \Phi$).

The axiom **(RD2)** may be replaced by the equivalent axiom:

**(RD2'):** If $\alpha \in \Phi$, then $s_\alpha(\Phi) \subset \Phi$, and the $s_\alpha$, $\alpha \in \Phi$, generate a finite group.

If $\Phi \neq \emptyset$, then $\Phi$ is a root system in $V := \langle \Phi \rangle \otimes_{\mathbb{Z}} \mathbb{R}$, where $\langle \Phi \rangle$ is the subgroup of $X$ generated by $\Phi$. The set $\Phi^\vee$ is the dual of the root system $\Phi$.

The quadruple $\Psi^\vee = (Y, X, \Phi^\vee, \Phi)$ is also a root datum, called the *dual* of $\Psi$. A root datum is *reduced* if it satisfies a third axiom

**(RD3):** $\alpha \in \Phi \Longrightarrow 2\alpha \notin \Phi$.

The root datum $\Psi(\mathbf{G},\mathbf{T})$ is reduced.

An isomorphism of a root datum $\Psi = (X, Y, \Phi, \Phi^\vee)$ onto a root datum $\Psi' = (X', Y', \Phi', \Phi'^\vee)$ is a group isomorphism $f : X \to X'$ which induces a bijection of $\Phi$ onto $\Phi'$ and whose dual induces a bijection of $\Phi'^\vee$ onto $\Phi^\vee$. If $\mathbf{G}'$ is a linear algebraic group which is isomorphic to $\mathbf{G}$, and $\mathbf{T}'$ is a maximal torus in $\mathbf{G}'$, then the root data $\Psi(\mathbf{G},\mathbf{T})$ and $\Psi(\mathbf{G}',\mathbf{T}')$ are isomorphic.

If $\Psi$ is a reduced root datum, there exists a connected reductive $K$-group $\mathbf{G}$ and a maximal torus $\mathbf{T}$ in $\mathbf{G}$ such that $\Psi = \Psi(\mathbf{G},\mathbf{T})$. The pair $(\mathbf{G},\mathbf{T})$ is unique up to isomorphism.

## 6. Parabolic subgroups

Let $\mathbf{G}$ be a connected linear algebraic group. The set of connected closed solvable subgroups of $\mathbf{G}$, ordered by inclusion, contains maximal elements. Such a maximal element is called a *Borel subgroup* of $\mathbf{G}$. If $\mathbf{B}$ is a Borel subgroup, then $\mathbf{G}/\mathbf{B}$ is a projective variety and any other Borel subgroup is conjugate to $\mathbf{B}$. If $\mathbf{P}$ is a closed subgroup of $\mathbf{G}$, then $\mathbf{G}/\mathbf{P}$ is a projective variety if and only if $\mathbf{P}$ contains a Borel subgroup. Such a subgroup is called a *parabolic subgroup*. If $\mathbf{P}$ is a parabolic subgroup, then $\mathbf{P}$ is connected and the normalizer $N_{\mathbf{G}}(\mathbf{P})$ of $\mathbf{P}$ in $\mathbf{G}$ is $\mathbf{P}$. If $\mathbf{P}$ and $\mathbf{P}'$ are parabolic subgroups containing a Borel subgroup $\mathbf{B}$, and $\mathbf{P}$ and $\mathbf{P}'$ are conjugate, then $\mathbf{P} = \mathbf{P}'$.

Now assume that $\mathbf{G}$ is a connected reductive linear algebraic group. Let $\mathbf{T}$ be a maximal torus in $\mathbf{G}$. Then $\mathbf{T}$ lies inside some Borel subgroup $\mathbf{B}$ of $\mathbf{G}$. Let $\mathbf{U} = R_u(\mathbf{B})$ be the unipotent radical of $\mathbf{B}$. There exists a unique base $\Delta$ of $\Phi = \Phi(\mathbf{G}, \mathbf{T})$ such that $\mathbf{U}$ is generated by the groups $\mathbf{U}_\alpha$, $\alpha \in \Phi^+$, and $\mathbf{B} = \mathbf{T} \ltimes \mathbf{U}$. Conversely if $\Delta$ is a base of $\Phi$, then the group generated by $\mathbf{T}$ and by the groups $\mathbf{U}_\alpha$, $\alpha \in \Phi^+$, is a Borel subgroup of $\mathbf{G}$. Hence the set of Borel subgroups of $\mathbf{G}$ which contain $\mathbf{T}$ is in one to one correspondence with the set of bases of $\Phi$. The Weyl group $W$ permutes the set of Borel subgroups containing $\mathbf{T}$ simply transitively. The set of Borel subgroups containing $\mathbf{T}$ generates $\mathbf{G}$.

*The Bruhat decomposition.* Let $\mathbf{B}$ be a Borel subgroup of $\mathbf{G}$, and let $\mathbf{T}$ be a maximal torus of $\mathbf{G}$ contained in $\mathbf{B}$. Then $\mathbf{G}$ is the disjoint union of the double cosets $\mathbf{B}w\mathbf{B}$, as $w$ ranges over a set of representatives in $N_{\mathbf{G}}(\mathbf{T})$ of the Weyl group $W$ ($BwB = Bw'B$ if and only if $w = w'$ in $W$).

Let $\mathbf{G}$, $\mathbf{B}$ and $\mathbf{T}$ be as above. Let $\Delta$ be the base of $\Phi(\mathbf{G}, \mathbf{T})$ corresponding to $\mathbf{B}$. If $I$ is a subset of $\Delta$, let $W_I$ be the subgroup of $W$ generated by the subset $S_I = \{\, s_\alpha \mid \alpha \in I \,\}$ of $I$. Let $\mathbf{P}_I = \mathbf{B}W_I\mathbf{B}$ (note that $\mathbf{P}_\emptyset = \mathbf{B}$). Then $\mathbf{P}_I$ is a parabolic subgroup of $\mathbf{G}$ (containing $\mathbf{B}$). A subgroup of $\mathbf{G}$ containing $\mathbf{B}$ is equal to $\mathbf{P}_I$ for some subset $I$ of $\Delta$. If $I$ and $J$ are subsets of $\Delta$ then $W_I \subset W_J$ implies $I \subset J$ and $\mathbf{P}_I \subset \mathbf{P}_J$ implies $I \subset J$. Also, $\mathbf{P}_I$ is conjugate to $\mathbf{P}_J$ if and only if $I = J$. A parabolic subgroup is called *standard* if it contains $\mathbf{B}$. Any parabolic subgroup $\mathbf{P}$ is conjugate to some standard parabolic subgroup.

Let $I \subset \Delta$. The set $\Phi_I$ of $\alpha \in \Phi$ such that $\alpha$ is an integral linear combination of elements of $I$ forms a root system, with Weyl group $W_I$. The set of roots $\Phi(\mathbf{P}_I, \mathbf{T})$ of $\mathbf{P}_I$ relative to $\mathbf{T}$ is equal to $\Phi^+ \cup (\Phi^- \cap \Phi_I)$. Let $\mathbf{N}_I = R_u(\mathbf{P}_I)$. Then $\mathbf{N}_I$ is a $\mathbf{T}$-stable subgroup of $\mathbf{U} = \mathbf{B}_u$, and is generated by those $\mathbf{U}_\alpha$ which are contained in $\mathbf{N}_I$, that is, by those $\mathbf{U}_\alpha$ such that $\alpha \in \Phi^+$ and $\alpha \notin \Phi_I$. Let $\mathbf{T}_I = (\cap_{\alpha \in I} \operatorname{Ker} \alpha)^0$, and let $\mathbf{M}_I = Z_{\mathbf{G}}(\mathbf{T}_I)$. The set $\Phi_I$ coincides with the set of roots in $\Phi$ which are trivial on $\mathbf{T}_I$. The group $\mathbf{M}_I$ is reductive and is generated by $\mathbf{T}$ and by the set of $\mathbf{U}_\alpha$, $\alpha \in \Phi_I$, $\mathbf{T}_I$ is the identity component of the centre of $\mathbf{M}_I$, and $\Phi(\mathbf{M}_I, \mathbf{T}) = \Phi_I$. The Lie algebra of $\mathbf{M}_I$ is equal to $\mathfrak{t} \oplus \bigoplus_{\alpha \in \Phi_I} \mathfrak{g}_\alpha$ (here $\mathfrak{t}$ is the Lie algebra of $\mathbf{T}$). The group $\mathbf{M}_I$ normalizes $\mathbf{N}_I$ and $\mathbf{P}_I = \mathbf{M}_I \ltimes \mathbf{N}_I$. A *Levi factor* (or *Levi component*) of $\mathbf{P}_I$ is a reductive group $\mathbf{M}$ such that $\mathbf{P}_I = \mathbf{M} \ltimes \mathbf{N}_I$, and the decomposition $\mathbf{P}_I = \mathbf{M} \ltimes \mathbf{N}_I$ is called a *Levi decomposition* of $\mathbf{P}_I$. If $\mathbf{M}$ is a Levi factor of $\mathbf{P}_I$, then there exists $n \in \mathbf{N}_I$ such that $\mathbf{M} = n\mathbf{M}_I n^{-1}$. It is possible for $\mathbf{M}_I$ and $\mathbf{M}_J$ to be conjugate for distinct subsets $I$ and $J$ of $\Delta$. More generally, if $\mathbf{P}$ is any parabolic subgroup of $\mathbf{G}$, $\mathbf{P}$ has Levi decompositions (which we can obtain via conjugation from Levi decompositions of a standard parabolic subgroup to which $\mathbf{P}$ is conjugate).

Note that if $\mathbf{P}$ is a proper parabolic subgroup of $\mathbf{G}$, then the semisimple rank of a Levi factor of $\mathbf{P}$ is strictly less than the semisimple rank of $\mathbf{G}$. This fact is often used in inductive arguments.

## 7. Reductive groups - relative theory

Let $k$ be a subfield of $K$. Throughout this section, we assume that $\mathbf{G}$ is a connected reductive $k$-group. Then $\mathbf{G}$ has a maximal torus which is defined over $k$. We say that $\mathbf{G}$ is *$k$-split* if $\mathbf{G}$ has a maximal torus $\mathbf{T}$ which is $k$-split. If $\mathbf{G}$ is $k$-split and $\mathbf{T}$ is such a torus, then each $\mathbf{U}_\alpha$, $\alpha \in \Phi(\mathbf{G}, \mathbf{T})$, is defined over $k$, and the associated isomorphism $\varepsilon_\alpha : \mathbf{G}_a \to \mathbf{U}_\alpha$ can be taken to be defined over $k$. If $\mathbf{G}$ contains no $k$-split tori, then $\mathbf{G}$ is said to be *$k$-anisotropic*. There exists a finite separable extension of $k$ over which $\mathbf{G}$ splits.

Suppose that $\mathbf{G}$ and $\mathbf{G}'$ are connected reductive $k$-split $k$-groups which are isomorphic. Then $\mathbf{G}$ and $\mathbf{G}'$ are $k$-isomorphic.

The centralizer $Z_{\mathbf{G}}(\mathbf{T})$ of a $k$-torus $\mathbf{T}$ in $\mathbf{G}$ is reductive and defined over $k$, and if $\mathbf{T}$ is $k$-split, $Z_{\mathbf{G}}(\mathbf{T})$ is the Levi factor of a parabolic $k$-subgroup of $\mathbf{G}$. (Here, we say a closed subgroup $\mathbf{H}$ of $\mathbf{G}$ is a $k$-subgroup of $\mathbf{G}$ if $\mathbf{H}$ is a $k$-group). Any $k$-torus in $\mathbf{G}$ is contained in some maximal torus which is defined over $k$. If $k$ is infinite, then $\mathbf{G}(k)$ is Zariski dense in $\mathbf{G}$.

The maximal $k$-split tori of $\mathbf{G}$ are all conjugate under $\mathbf{G}(k)$. Let $\mathbf{S}$ be a maximal $k$-split torus in $\mathbf{G}$. The *$k$-rank* of $\mathbf{G}$ is the dimension of $\mathbf{S}$. The *semisimple $k$-rank* of $\mathbf{G}$ is the $k$-rank of $\mathbf{G}/R(\mathbf{G})$. The finite group ${}_kW = N_{\mathbf{G}}(\mathbf{S})/Z_{\mathbf{G}}(\mathbf{S})$ is called the *$k$-Weyl group*. The set ${}_k\Phi = \Phi(\mathbf{G}, \mathbf{S})$ of roots of $\mathbf{G}$ relative to $\mathbf{S}$ is called the *$k$-roots* of $\mathbf{G}$. The $k$-roots form an abstract root system, which is not necessarily reduced, with Weyl group isomorphic to ${}_kW$. The rank of ${}_k\Phi$ is equal to the semisimple $k$-rank of $\mathbf{G}$.

A Borel subgroup $\mathbf{B}$ of $\mathbf{G}$ might not be defined over $k$. We say that $\mathbf{G}$ is *$k$-quasisplit* if $\mathbf{G}$ has a Borel subgroup that is defined over $k$. If $\mathbf{P}$ is a parabolic $k$-subgroup of $\mathbf{G}$, then $R_u(\mathbf{P})$ is defined over $k$. A Levi factor $\mathbf{M}$ of a parabolic $k$-subgroup is called a Levi $k$-factor of $\mathbf{P}$ if $\mathbf{M}$ is a $k$-group. Any two Levi $k$-factors of $\mathbf{P}$ are conjugate by a unique element of $R_u(\mathbf{P})(k)$. If two parabolic $k$-subgroups of $\mathbf{G}$ are conjugate by an element of $\mathbf{G}$ then they are conjugate by an element of $\mathbf{G}(k)$. The group $\mathbf{G}$ contains a proper parabolic $k$-subgroup if and only if $\mathbf{G}$ contains a noncentral $k$-split torus, that is, if the semisimple $k$-rank of $\mathbf{G}$ is positive. The results described in this section give no information in the case where $\mathbf{G}$ has semisimple $k$-rank zero.

Let $\mathbf{P}_0$ be a minimal element of the set of parabolic $k$-subgroups of $\mathbf{G}$ (such an element exists, since the set is nonempty, as it contains $\mathbf{G}$). Any minimal parabolic $k$-subgroup of $\mathbf{G}$ is conjugate to $\mathbf{P}_0$ by an element of $\mathbf{G}(k)$. The group $\mathbf{P}_0$ contains a maximal $k$-split torus $\mathbf{S}$ of $\mathbf{G}$, and $Z_{\mathbf{G}}(\mathbf{S})$ is a $k$-Levi factor of $\mathbf{P}_0$. The semisimple $k$-rank of $Z_{\mathbf{G}}(\mathbf{S})$ is zero. Because $N_{\mathbf{G}}(\mathbf{S}) = N_{\mathbf{G}}(\mathbf{S})(k) \cdot Z_{\mathbf{G}}(\mathbf{S})$, $\mathbf{G}(k)$ contains representatives for all elements of ${}_kW$. The group ${}_kW$ acts simply transitively on the set of minimal parabolic $k$-subgroups containing $Z_{\mathbf{G}}(\mathbf{S})$.

Let $\mathrm{Lie}(Z_{\mathbf{G}}(\mathbf{S}))$ be the Lie algebra of $Z_{\mathbf{G}}(\mathbf{S})$. Then

$$\mathfrak{g} = \mathrm{Lie}(Z_{\mathbf{G}}(\mathbf{S})) \oplus \bigoplus_{\alpha \in {}_k\Phi} \mathfrak{g}_\alpha.$$

If $\alpha \in {}_k\Phi$ and $2\alpha \notin {}_k\Phi$, then $\mathfrak{g}_\alpha$ is a subalgebra of $\mathfrak{g}$. If $\alpha$ and $2\alpha \in {}_k\Phi$, then $\mathfrak{g}_\alpha + \mathfrak{g}_{2\alpha}$ is a subalgebra of $\mathfrak{g}$. For each $\alpha \in {}_k\Phi$, set

$$\mathfrak{g}_{(\alpha)} = \begin{cases} \mathfrak{g}_\alpha, & \text{if } 2\alpha \notin {}_k\Phi \\ \mathfrak{g}_\alpha \oplus \mathfrak{g}_{2\alpha}, & \text{if } 2\alpha \in {}_k\Phi. \end{cases}$$

There exists a unique closed connected unipotent $k$-subgroup $\mathbf{U}_{(\alpha)}$ of $\mathbf{G}$ which is normalized by $Z_{\mathbf{G}}(\mathbf{S})$ and has Lie algebra $\mathfrak{g}_{(\alpha)}$.

Let $\mathbf{P}_0$ be as above. Then there exists a unique base ${}_k\Delta$ of ${}_k\Phi$ such that $R_u(\mathbf{P}_0)$ is generated by the groups $\mathbf{U}_{(\alpha)}$, $\alpha \in {}_k\Phi^+$. The set of standard parabolic $k$-subgroups of $\mathbf{G}$ corresponds bijectively with the set of subsets of ${}_k\Phi$. Fix $I \subset {}_k\Delta$. Let $\mathbf{S}_I = (\cap_{\alpha \in I} \cap \operatorname{Ker} \alpha)^0$ and let ${}_k\Phi_I$ be the set of $\alpha \in {}_k\Phi$ which are integral linear combinations of the roots in $I$. Let ${}_kW_I$ be the subgroup of ${}_kW$ generated by the reflections $s_\alpha$, $\alpha \in I$. The parabolic $k$-subgroup of $\mathbf{G}$ corresponding to $I$ is $\mathbf{P}_I = \mathbf{P}_0 \cdot {}_kW_I \cdot \mathbf{P}_0$. The unipotent radical of $\mathbf{P}_I$ is equal to $\mathbf{N}_I$, the subgroup of $\mathbf{G}$ generated by the groups $\mathbf{U}_{(\alpha)}$, as $\alpha$ ranges over the elements of ${}_k\Phi^+$ which are not in ${}_k\Phi_I$. The $k$-subgroup $\mathbf{M}_I := Z_{\mathbf{G}}(\mathbf{S}_I)$ is a Levi $k$-factor of $\mathbf{P}_I$, $\Phi(\mathbf{M}_I, \mathbf{S}) = {}_k\Phi_I$, and ${}_kW_I = {}_kW(\mathbf{M}_I, \mathbf{S})$.

A parabolic $k$-subgroup of $\mathbf{G}$ is conjugate to exactly one $\mathbf{P}_I$, and it is conjugate to $\mathbf{P}_I$ by an element of $\mathbf{G}(k)$.

*Relative Bruhat decomposition.* Let $\mathbf{U}_0 = R_u(\mathbf{P}_0)$. Then $\mathbf{G}(k) = \mathbf{U}_0(k) \cdot N_{\mathbf{G}}(\mathbf{S})(k) \cdot \mathbf{U}_0(k)$, and $\mathbf{G}(k)$ is the disjoint union of the sets $\mathbf{P}_0(k) w \mathbf{P}_0(k)$, as $w$ ranges over a set of representatives for elements of ${}_kW$ in $N_{\mathbf{G}}(\mathbf{S})(k)$.

A *parabolic subgroup* of $\mathbf{G}(k)$ is a subgroup of the form $\mathbf{P}(k)$, where $\mathbf{P}$ is a parabolic $k$-subgroup of $\mathbf{G}$. A subgroup of $\mathbf{G}(k)$ which contains $\mathbf{P}_0(k)$ is equal to $\mathbf{P}_I(k)$ for some $I \subset {}_k\Delta$. If $I \subset {}_k\Delta$, choosing representatives for ${}_kW_I$ in $N_{\mathbf{G}}(\mathbf{S})(k)$, we have $\mathbf{P}_I(k) = \mathbf{P}_0(k) \cdot {}_kW_I \cdot \mathbf{P}_0(k)$. The group $\mathbf{P}_I(k)$ is equal to its own normalizer in $\mathbf{G}(k)$. The Levi decomposition $\mathbf{P}_I = \mathbf{M}_I \ltimes \mathbf{N}_I$ carries over to the $k$-rational points: $\mathbf{P}_I(k) = \mathbf{M}_I(k) \ltimes \mathbf{N}_I(k)$. If $I$, $J \subset {}_k\Delta$ and $g \in \mathbf{G}(k)$, then $g\mathbf{P}_J(k)g^{-1} \subset \mathbf{P}_I(k)$ if and only if $J \subset I$ and $g \in \mathbf{P}_I(k)$.

## 8. Examples

EXAMPLE 8.1. $\mathbf{G} = \mathbf{GL}_n(K)$, $n \geq 2$.

The group $\mathbf{T} = \{ \operatorname{diag}(t_1, t_2, \ldots, t_n) \mid t_i \in K^\times \}$ is a maximal torus in $\mathbf{G}$. For $1 \leq i \leq n$, let $\ell_i = (0, 0, \ldots, 0, 1, 0, \cdots, 0) \in \mathbf{Z}^n$, with the 1 occurring in the $i$th coordinate. The map $\sum_{i=1}^n k_i\ell_i \mapsto \chi_{\sum_{i=1}^n k_i\ell_i}$, where

$$\chi_{\sum_{i=1}^n k_i\ell_i}(\operatorname{diag}(t_1, \cdots, t_n)) = t_1^{k_1} \cdots t_n^{k_n},$$

is an isomorphism from $\mathbf{Z}^n$ to $X(\mathbf{T})$. If $\mu_{\sum_{i=1}^n k_i\ell_i}(t) = \operatorname{diag}(t^{k_1}, \cdots, t^{k_n})$, $t \in K^\times$, then the map $\sum_{i=1}^n k_i\ell_i \mapsto \mu_{\sum_{i=1}^n k_i\ell_i}$ is an isomorphism from $\mathbf{Z}^n$ to $Y(\mathbf{T})$. Also, $\langle \chi_{\Sigma k_i\ell_i}, \mu_{\Sigma \ell_i e_i} \rangle = \sum_{i=1}^n k_i\ell_i$. The root system $\Phi = \Phi(\mathbf{G}, \mathbf{T}) = \{ \chi_{\ell_i - \ell_j} \mid 1 \leq i \neq j \leq n \}$.

For $1 \leq i \neq j \leq n$, let $E_{ij} \in M_n(K) = \mathfrak{g}$ be the matrix having a 1 in the $ij^{\text{th}}$ entry, and zeros elsewhere. If $\alpha = \chi_{\ell_i - \ell_j}$, $i \neq j$, then $\mathfrak{g}_\alpha$ is spanned by $E_{ij}$, and

$\mathbf{U}_\alpha = \{\, I_n + tE_{ij} \mid t \in K \,\}$. The reflection $s_\alpha$ permutes $\ell_i$ and $\ell_j$, and fixes all $\ell_k$ with $k \notin \{i, j\}$. The co-root $\alpha^\vee$ is $\mu_{\ell_i - \ell_j}$ The Weyl group $W$ is isomorphic to the symmetric group $S_n$. The root system $\Phi \simeq \Phi^\vee$ is of type $A_{n-1}$.

The set $\Delta := \{\, \chi_{\ell_i - \ell_{i+1}} \mid 1 \le i \le n-1 \,\}$ is a base of $\Phi$. The corresponding Borel subgroup $\mathbf{B}$ is the subgroup of $\mathbf{G}$ consisting of upper triangular matrices.

If $I \subset \Delta$, there exists a partition $(n_1, n_2, \cdots, n_r)$ of $n$ ($n_i$ a positive integer, $1 \le i \le r$, $n_1 + n_2 + \cdots + n_r = n$), such that

$$\mathbf{T}_I = \{\, \mathrm{diag}\,(\underbrace{a_1, \ldots, a_1}_{n_1 \text{ times}}, \underbrace{a_2, \ldots, a_2}_{n_2 \text{ times}}, \ldots, \underbrace{a_r, \ldots, a_r}_{n_r \text{ times}}) \mid a_1, a_2, \ldots, a_r \in K^\times \,\}$$

The group $\mathbf{M}_I := \mathbf{Z}_{\mathbf{G}}(\mathbf{T}_I)$ is isomorphic to $\mathbf{GL}_{n_1}(K) \times \mathbf{GL}_{n_2}(K) \times \cdots \times \mathbf{GL}_{n_r}(K)$, $\mathbf{N}_I$ consists of matrices of the form

$$\begin{bmatrix} I_{n_1} & * & * & * \\ & I_{n_2} & * & \vdots \\ 0 & & \ddots & * \\ & & & I_{n_r} \end{bmatrix},$$

and $\mathbf{P}_I = \mathbf{M}_I \ltimes \mathbf{N}_I$.

Example 8.2. $\mathbf{G} = \mathbf{Sp}_4(K)$ (the $4 \times 4$ symplectic group). Let

$$J = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}$$

Then $\mathbf{G} = \{\mathfrak{g} \in \mathbf{GL}_4(K) \mid {}^t gJg = J\}$ and $\mathfrak{g} = \{X \in M_4(K) \mid {}^t XJ + JX = 0\}$.

The group $\mathbf{T} := \{\, \mathrm{diag}\ (a,\, b,\, b^{-1},\, a^{-1}) \mid a,\ b \in K^\times\}$ is a maximal torus in $\mathbf{G}$ and $X(\mathbf{T}) \simeq \mathbf{Z} \times \mathbf{Z}$, via $\chi_{(i,j)} \leftrightarrow (i, j)$, where $\chi_{(i,j)}(\mathrm{diag}\,(a,\, b,\, b^{-1},\, a^{-1})) = a^i b^j$. And $Y(\mathbf{T}) \simeq \mathbf{Z} \times \mathbf{Z}$, via $\mu_{(i,j)} \leftrightarrow (i, j)$, where $\mu_{(i,j)}(t) = (\mathrm{diag}\,(t^i,\, t^j,\, t^{-j},\, t^{-i})$. Note that $\langle \chi_{(i,j)}, \mu_{(k,\ell)} \rangle = ki + j\ell$.

Let $\alpha = \chi_{(1,-1)}$ and $\beta = \chi_{(0,2)}$. Then

$$\Phi = \{\, \pm\alpha,\, \pm\beta,\, \pm(\alpha + \beta),\, \pm(2\alpha + \beta \,\},$$

$\Delta := \{\alpha,\ \beta\}$ is a base of $\Phi = \Phi(\mathbf{G}, \mathbf{T})$, and

$$\begin{aligned} &\mathfrak{g}_\alpha = \mathrm{Span}_K(E_{12} - E_{34}), \quad \mathfrak{g}_{-\alpha} = \mathrm{Span}_K(E_{21} - E_{43}) \quad \mathfrak{g}_\beta = \mathrm{Span}_K E_{23} \\ &\mathfrak{g}_{\alpha+\beta} = \mathrm{Span}_K(E_{13} + E_{24}), \quad \mathfrak{g}_{2\alpha+\beta} = \mathrm{Span}_K E_{14}, \quad \text{etc.} \end{aligned}$$

Identifying $\alpha$ and $\beta$ with $(1, -1)$ and $(0, 2) \in \mathbf{Z} \times \mathbf{Z}$, respectively, we have $s_\alpha(1,\ -1) = (-1,\ 1) = -\alpha$ and $s_\alpha(1,\ 1) = (1,\ 1)$. The corresponding element of $W = N_{\mathbf{G}}(\mathbf{T})/\mathbf{T}$ is represented by the matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

We also have $s_\beta(0,2) = (0,-2) = -\beta$ and $s_\beta(1,0) = (1,0)$. The corresponding element of $W = N_{\mathbf{G}}(\mathbf{T})/\mathbf{T}$ is represented by the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The Weyl group $W = W(\Phi)$ is equal to $\{\, 1,\, s_\alpha,\, s_\beta,\, s_\alpha s_\beta,\, s_\beta s_\alpha,\, s_\beta s_\alpha s_\beta,\, s_\alpha s_\beta s_\alpha,\, (s_\beta s_\alpha)^2 \,\}$ which is isomorphic to the dihedral group of order 8.

The dual root system $\Phi^\vee$ is described by

$$\begin{aligned} \Phi^\vee &= \{\pm\alpha^\vee,\, \pm\beta^\vee,\, \pm(\alpha+\beta)^\vee,\, \pm(2\alpha+\beta)^\vee\} \\ \alpha^\vee &= (1,\, -1) \quad (\alpha+\beta)^\vee = (1,\, 1) \\ \beta^\vee &= (0,\, 1) \quad (2\alpha+\beta)^\vee = (1,\, 0) \end{aligned}$$

The root system $\Phi$ is of type $C_2$ and $\Phi^\vee$ is of type $B_2$, isomorphic to $C_2$.

REMARK 8.3. If $n > 2$ the root system of $\mathbf{Sp}_{2n}(K)$ is of type $C_n$, and the dual is of type $B_n$, and $B_n$ and $C_n$ are not isomorphic.

The Borel subgroup of $\mathbf{G}$ which corresponds to $\Delta$ is the subgroup $\mathbf{B}$ of upper triangular matrices in $\mathbf{G}$. Apart from $\mathbf{G}$ and $\mathbf{B}$, there are two standard parabolic subgroups, $\mathbf{P}_\alpha$ and $\mathbf{P}_\beta$, attached to the subsets $\{\alpha\}$ and $\{\beta\}$ of $\Delta$, respectively. It is easy to check that

$$\mathbf{T}_\alpha = (\mathrm{Ker}\ \alpha)^\circ = \{\mathrm{diag}\ (a,\ a,\ a^{-1}\, a^{-1}) \mid a \in K^\times\}$$

$$\mathbf{M}_\alpha = \mathbf{Z}_{\mathbf{G}}(\mathbf{T}_\alpha) = \left\{ \begin{bmatrix} A & 0 \\ 0 & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} {}^tA^{-1} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{bmatrix} \;\Big|\; A \in \mathbf{GL}_2(K) \right\}$$

$$\mathbf{N}_\alpha = \left\{ \begin{bmatrix} I_2 & B \\ 0 & I_2 \end{bmatrix} \;\Big|\; B \in M_2(K),\ {}^tB = B \right\}$$

$$\mathbf{T}_\beta = (\mathrm{Ker}\ \beta)^\circ = \{\, \mathrm{diag}\, (a,\ 1,\ 1,\ a^{-1}) \mid a \in K^\times \,\}$$

$$\mathbf{M}_\beta = \mathbf{Z}_{\mathbf{G}}(\mathbf{T}_\beta) = \left\{ \begin{bmatrix} d & 0 & 0 & 0 \\ 0 & c_{11} & c_{12} & 0 \\ 0 & c_{21} & c_{22} & 0 \\ 0 & 0 & 0 & d^{-1} \end{bmatrix} \;\Big|\; d \in K^\times,\ c_{11}c_{22} - c_{12}c_{21} = 1 \right\} \simeq \mathbf{SL}_2(K) \times K^\times$$

$$\mathbf{N}_\beta = \left\{ \begin{bmatrix} 1 & x & y & z \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & -x \\ 0 & 0 & 0 & 1 \end{bmatrix} \;\Big|\; x, y, z \in K \right\}$$

## 9. Comments on references

For the basic theory of linear algebraic groups, see [**B1**], [**H**] and [**Sp2**], as well as the survey article [**B2**]. For information on reductive groups defined over non algebraically closed fields, the main reference is [**BoT1**] and [**BoT2**]. Some material appears in [**B1**], and there is a survey of rationality properties at the end of [**Sp2**]. See also the survey article [**Sp1**]. For the classification of semisimple algebraic groups, see [**Sa**] and [**T2**]. For information on reductive groups over local nonarchimedean fields, see [**BrT1**], [**BrT2**], and the article [**T1**]. Adeles and algebraic groups are discussed in [**W**].

## References


[B1] A. Borel, *Linear algebraic groups*, Second Enlarged Edition, Graduate Texts in Mathematics, **126**, Springer-Verlag, New York 1991.

[B2] A. Borel, *Linear algebraic groups*, in Algebraic Groups and Discontinuous Subgroups, Proc. Symp. Pure Math. **9**, (1966), 3–19.

[BoT1] A. Borel and J. Tits, *Groupes réductifs*, Publ. Math. I.H.E.S. **27** (1965), 55-151.

[BoT2] A. Borel and J. Tits, *Compléments à larticle: Groupes réductifs*, Publ. Math. I.H.E.S. **41** (1972), 253–276.

[B] N. Bourbaki, *Groupes et algèbres de Lie*, Hermann, Paris, 1968, Chapters IV, V, VI.

[BrT1] F. Bruhat and J. Tits, *Groupes réductifs sur un corps local I, Données radicielles valuées*, Publ. Math. IHES **41**, (1972), 5–251.

[BrT2] F. Bruhat and J. Tits, *Groupes réductifs sur un corps local II, Schémas en groupes. Existence d'une donnée radicielle valuée*, Publ. Math. IHES **60** (1984), 197–376.

[H] J.E. Humphreys, *Linear algebraic groups*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1975.

[Sa] Satake, *Classification theory of semisimple algebraic groups*, Lecture Notes in Pure and Applied Mathematics, Marcel Dekker, New York 1971.

[Sp1] T.A. Springer, *Reductive groups*, in Automorphic Forms, Representations, and $L$-functions, Proc. Symp. Pure Math. **33**, (1979), part 1, 3–27.

[Sp2] T.A. Springer, *Linear algebraic groups*, Birkhauser, Boston, 1981.

[T1] J. Tits, *Reductive groups over local fields*, Proc. Symp. Pure Math. **33** (1979), part 1, 29–69.

[T2] J. Tits, *Classification of algebraic semisimple groups*, Proc. Symp. Pure Math. **9** (1966), 33-62.

[W] A. Weil, *Adeles and algebraic groups*, Insitute for Advanced Study, Princeton, N.J., 1961.



Department of Mathematics, University of Toronto, Toronto, Canada M5S 3G3
*E-mail address*: `fiona@math.toronto.edu`

Clay Mathematics Proceedings
Volume 4, 2005

# Harmonic Analysis on Reductive $p$-adic Groups and Lie Algebras

Robert E. Kottwitz

Contents

## Introduction

One of the long term goals in the representation theory of reductive groups over $p$-adic fields is the local Langlands conjecture, which classifies irreducible representations in terms of Langlands parameters (and auxiliary data on the Langlands dual group). This goal has been achieved for $GL_n$ (see [**Car00**] for a survey of recent work on this problem), but not in general.

One way to approach the classification problem for classical groups is via the twisted Arthur-Selberg trace formula for $GL_n$, the reason behind this being that all quasi-split classical groups are captured as twisted endoscopic groups for $GL_n$ (strictly speaking, a suitable restriction of scalars of $GL_n$ in the case of unitary groups). The building blocks of representation theory on a $p$-adic group are the supercuspidal representations, and these show up in the trace formula only through their distribution characters. To use the trace formula successfully it is necessary to know some qualitative facts about these characters.

Here we should pause to recall that for any smooth representation $\pi$ of our $p$-adic group $G$ and any $f \in C_c^\infty(G)$ (the space of locally constant and compactly supported functions on $G$) there is an operator $\pi(f)$ on the underlying vector space $V_\pi$ of $\pi$, defined on $v \in V_\pi$ by

$$\pi(f)(v) := \int_G f(g)\pi(g)(v)\, dg, \tag{0.0.1}$$

$dg$ being some fixed Haar measure on $G$.

When $\pi$ is irreducible, it is necessarily admissible (see the survey article [**BZ76**] for this and other basic facts about the representation theory of $p$-adic groups), which guarantees that $\pi(f)$ has finite rank and hence has a trace. The character $\Theta_\pi$ of $\pi$ is the distribution on $G$ defined by

$$\Theta_\pi(f) = \operatorname{tr} \pi(f)$$

on each test function $f \in C_c^\infty(G)$.

It is a deep theorem of Harish-Chandra that the distribution $\Theta_\pi$ can be represented by integration against a locally constant function, still denoted $\Theta_\pi$, on the set $G_{\text{rs}}$ of regular semisimple elements in $G$. Thus, for all $f \in C_c^\infty(G)$ there is an equality

$$\Theta_\pi(f) = \int_G f(g)\Theta_\pi(g)\, dg. \tag{0.0.2}$$

The function $\Theta_\pi$ is independent of the choice of Haar measure, and by comparing equations (0.0.1) and (0.0.2) one sees that formally $\Theta_\pi(g) = \operatorname{tr} \pi(g)$, though of course $\operatorname{tr} \pi(g)$ does not make sense literally when $\pi$ is infinite dimensional, as is usually the case.

In order to extract information about classical groups from the twisted trace formula for $GL_n$ (see [**Art97**]), one must stabilize the twisted trace formula, which is to say that one must express it as a linear combination of stable trace formulas for the elliptic twisted endoscopic groups $H$ of $GL_n$. The stabilization process should then yield identities between (suitable linear combinations of) characters on the classical group $H$ and (suitable linear combinations of) twisted characters on $GL_n$. This has been done in full [**Rog90**] for $GL_3(E)$ and its twisted endoscopic group $U(3)$ (quasi-split unitary group over $F$ coming from a quadratic extension $E/F$), giving a classification for representations of $U(3)$ in terms of the better understood representations of $GL_3(E)$.

Information about the characters of irreducible representations of $p$-adic groups is embedded in the spectral side of the trace formula, but in order to carry out the stabilization one must start with the geometric sides of the relevant trace formulas, and so one must study orbital integrals (invariant integrals over conjugacy classes) as well as Arthur's weighted orbital integrals (obtained from certain non-invariant integrals over conjugacy classes, and more generally from certain limits of these).

Thus, in order to use the trace formula, one needs a good understanding of characters, orbital integrals and weighted orbital integrals, and these are precisely the main objects of study in harmonic analysis on $G$. Many of the most basic (and deepest) results in this area are due to Harish-Chandra, among them his theorem, mentioned above, that the distribution character $\Theta_\pi$ can be represented by a locally integrable function on $G$, locally constant on $G_{\mathrm{rs}}$.

The main step in Harish-Chandra's proof of this result involves passing to the Lie algebra $\mathfrak{g}$ of $G$ and proving an analogous result there. For this reason, among others, one should study harmonic analysis on $\mathfrak{g}$ along with that on $G$, and in this article, after introducing some of the key concepts in harmonic analysis on $G$, we will then concentrate on $\mathfrak{g}$, giving an almost self-contained exposition of Waldspurger's local trace formula on $\mathfrak{g}$ [**Wal95**] as well as many results of Harish-Chandra [**HC78, HC99**]. We should now follow up these rather vague motivational remarks with a more precise discussion of some of the key results in harmonic analysis on $G$ and $\mathfrak{g}$.

Harish-Chandra's first assault [**HC70**] on the problem of representing the distribution character $\Theta_\pi$ by a locally integrable function on $G$ was limited to the case in which $\pi$ is a supercuspidal representation. Given a vector $v$ in the space of $\pi$ and a vector $\tilde{v}$ in the space of the contragredient representation $\tilde{\pi}$, we can define a locally constant function $f_{\tilde{v},v}$ on $G$ by

$$f_{\tilde{v},v}(g) := \langle \tilde{v}, \pi(g)v \rangle, \tag{0.0.3}$$

the pairing denoting the value of the linear functional $\tilde{v}$ on the vector $\pi(g)v$. The function $f_{\tilde{v},v}$ is referred to as a matrix coefficient for $\pi$. For simplicity let us assume for the remainder of the introduction that the center of $G$ is compact. Our assumption that $\pi$ is supercuspidal implies [**BZ76**] that all its matrix coefficients are compactly supported functions on $G$. (Conversely, an irreducible smooth representation whose matrix coefficients are compactly supported is supercuspidal.)

Now choose $v, \tilde{v}$ so that $\langle \tilde{v}, v\rangle$ is the formal degree (see [**HC70**]) of $\pi$, and let $\phi$ denote the matrix coefficient $f_{\tilde{v},v}$. (Thus $\phi(1)$ is the formal degree of $\pi$.) Then as an easy first step Harish-Chandra shows that

$$\begin{aligned} \Theta_\pi(f) &= \int_G \Big[ \int_G \phi(x) f(g^{-1}xg)\, dx \Big]\, dg \\ &= \int_G \Big[ \int_G \phi(gxg^{-1}) f(x)\, dx \Big]\, dg. \end{aligned} \tag{0.0.4}$$

These integrals are convergent only as iterated integrals, and it is not legitimate to interchange the order of integration. However, let's pretend for a moment that we could interchange the order of integration. Then we would conclude that the distribution $\Theta_\pi$ is represented by the function $x \mapsto \int_G \phi(gxg^{-1})\, dg$ on $G$. This is nonsense since the integral $\int_G \phi(gxg^{-1})\, dg$ diverges unless the centralizer of $x$ is compact. Nevertheless Harish-Chandra shows that the restriction of the distribution $\Theta_\pi$ to $G_{\mathrm{rs}}$ is represented by the function

$$x \mapsto \int_G \Big[ \int_K \phi(gkxk^{-1}g^{-1})\,dk \Big]\,dg \tag{0.0.5}$$

for any compact open subgroup $K$ of $G$ that we like, and he then goes on to prove the difficult result that the function $\Theta_\pi$ is locally integrable on $G$ and represents the distribution $\Theta_\pi$ on all of $G$, not just on $G_{\mathrm{rs}}$.

For $x \in G_{\mathrm{rs}}$ the identity component of the centralizer of $x$ is a torus, and when this torus is compact (compact modulo the center, when the center is not assumed to be compact) we say that $x$ is elliptic. For elliptic regular semisimple $x$ Harish-Chandra shows that the order of integration in (0.0.5) can be reversed, so that (for such $x$) the character value at $x$ is given by the orbital integral

$$\Theta_\pi(x) = \int_G \phi(gxg^{-1})\,dg. \tag{0.0.6}$$

Arthur [**Art87**] has generalized the formula (0.0.6) in a very beautiful way: for any $x \in G_{\mathrm{rs}}$ the character value at $x$ is given by the weighted orbital integral

$$\Theta_\pi(x) = (-1)^{\dim A_M} \int_{A_M\backslash G} \phi(g^{-1}xg) v_M(g)\,d\dot{g} \tag{0.0.7}$$

for a weight function $v_M$ described in 12.6 and a suitably normalized invariant measure $d\dot{g}$ on the homogeneous space $A_M\backslash G$. Here $M$ is a Levi subgroup in which $x$ is elliptic, and $A_M$ is the split component of the center of $M$.

What happens when our irreducible smooth representation $\pi$ is not assumed to be supercuspidal? If $\pi$ is obtained by parabolic induction from a supercuspidal representation $\sigma$ of a Levi subgroup $M$ of $G$, then one can easily express the character of $\pi$ in terms of that of $\sigma$, and in this way show that $\Theta_\pi$ is represented by a locally integrable function on $G$. However, even for $GL_2$, the Grothendieck group of representations of $G$ having finite length is not spanned by the classes of representations that are parabolically induced from supercuspidal representations of Levi subgroups, so parabolic induction alone does not solve our problem.

To handle the general case Harish-Chandra [**HC78, HC99**] changed his strategy, no longer treating supercuspidal representations separately, and relying even more heavily on passage to the Lie algebra $\mathfrak{g}$. One goal of this article is to work through the main results of [**HC78, HC99**] concerning harmonic analysis on $\mathfrak{g}$. What serve as the Lie algebra analogs of the invariant distributions $\Theta_\pi$? The answer is quite simple: Fourier transforms of orbital integrals.

Orbital integrals are obtained as follows. For any $X \in \mathfrak{g}$ the adjoint orbit of $X$ can be identified with $G_X\backslash G$ ($G_X$ being the centralizer of $X$ in $G$) and carries a $G$-invariant measure $d\bar{g}$; moreover for any $f \in C_c^\infty(\mathfrak{g})$ the integral

$$O_X(f) := \int_{G_X\backslash G} f(g^{-1}Xg)\,d\bar{g}$$

converges, yielding an invariant distribution $O_X$, called the orbital integral for $X$.

The Fourier transform $f \mapsto \hat{f}$ for $f \in C_c^\infty(\mathfrak{g})$ is reviewed in 8.2. As usual one extends the notion of Fourier transform to distributions $T$ on $\mathfrak{g}$ by the rule

$$\hat{T}(f) = T(\hat{f}).$$

In particular we can consider the Fourier transform $\hat{O}_X$ of any orbital integral $O_X$, and, as mentioned above, $\hat{O}_X$ is analogous to the character of an irreducible representation.

For example, if $G$ is split with split maximal torus $A$ and $X$ is a regular element in $\mathrm{Lie}(A)$, then $\hat{O}_X$ is analogous to a principal series character. If $X = 0$, then $\hat{O}_X$ is analogous to the character of the trivial representation. If $T$ is an elliptic maximal torus and $X$ is regular in $\mathrm{Lie}(T)$, then $\hat{O}_X$ is analogous to the character of a supercuspidal representation of $G$.

Given this analogy, it is perhaps not hard to guess the statement of one of Harish-Chandra's main results in harmonic analysis on $\mathfrak{g}$ (see Theorem 27.8): for every $X \in \mathfrak{g}$ the distribution $\hat{O}_X$ is represented by a locally integrable function on $\mathfrak{g}$, locally constant on $\mathfrak{g}_{\mathrm{rs}}$, the set of regular semisimple elements in $\mathfrak{g}$. In fact Theorem 27.8 says something even more general: for any invariant distribution $I$ on $\mathfrak{g}$ whose support is bounded modulo conjugation (see 15.2) the Fourier transform $\hat{I}$ is represented by a locally integrable function on $\mathfrak{g}$, locally constant on $\mathfrak{g}_{\mathrm{rs}}$.

It is not easy to prove that $\hat{O}_X$ is represented by a function. The essential case is when $X$ lies in $\mathfrak{g}_e$, the open subset of elliptic regular semisimple elements in $\mathfrak{g}$, which Harish-Chandra treats by using Howe's finiteness theorem (see 26.2) to reduce to proving that $\hat{I}_\phi$ is represented by a function for any $\phi \in C_c^\infty(\mathfrak{g}_e)$. Here $I_\phi$ is the invariant distribution on $\mathfrak{g}$ defined for any cusp form $\phi \in C_c^\infty(\mathfrak{g})$ by the iterated integral

$$I_\phi(f) = \int_G \Big[ \int_{\mathfrak{g}} \phi(X) f(g^{-1}Xg)\, dX \Big]\, dg, \tag{0.0.8}$$

the Lie algebra analog of (0.0.4). The integral (0.0.8) is actually convergent as a double integral for the special cusp forms $\phi \in C_c^\infty(\mathfrak{g}_e)$. [Recall that we are assuming that the center of $G$ is compact and hence can be ignored. We should also note that later, when discussing $I_\phi$ systematically (see 25.2), we will find it convenient to build in a harmless factor $|\mathcal{Z}_G|^{-1}$ (see (25.4.3)).] In particular, we see that $\phi$ (and more generally any cusp form on $\mathfrak{g}$) behaves analogously to a cusp form on $G$. (Cusp forms on $G$ turn out to be linear combinations of matrix coefficients of supercuspidal representations of $G$.)

Now (0.0.8) is something that arises naturally in the context of Waldspurger's local trace formula [**Wal95**] on $\mathfrak{g}$, and following Waldspurger we use (see 25.2) the local trace formula to prove that $\hat{I}_\phi$ is represented by a function, as well as to prove Harish-Chandra's Lie algebra analog of (0.0.6) and Waldspurger's Lie algebra analog of (0.0.7).

Waldspurger uses the exponential map to derive the local trace formula on $\mathfrak{g}$ from Arthur's [**Art91a**] local trace formula on $G$. A second goal of this article is to write out a direct proof of the local trace formula on $\mathfrak{g}$. For the most part we follow Arthur's treatment of the geometric side of the one on $G$, the main point being Arthur's key geometric result (Theorem 22.3). However some steps in the proof are handled differently. For example toric varieties are used to pass from weight factors obtained by counting lattice points to weight factors obtained as volumes of convex polytopes; these considerations lead to a variant of the local trace formula taking values in the complexified $K$-theory of the relevant toric variety.

Shalika germs play an important role in Harish-Chandra's proofs and are used elsewhere in harmonic analysis. The last goal of this article is to give a self-contained treatment of them, including Harish-Chandra's deep linear independence result, which is closely tied to the density of regular semisimple orbital integrals (see Theorem 27.5) in the space of all invariant distributions on $\mathfrak{g}$.

It remains to explain the organization of this article. The first ten sections cover roughly the same material as that presented at the summer school. The first section discusses an abstract form of the local trace formula that one has on any compact group. It provides motivation and a first glimpse of how harmonic analysis works, but is not used again later in the article. The second section discusses the basics of integration on l.c.t.d spaces and proves a rather technical lemma used later in semisimple descent for orbital integrals. The third and fourth sections provide background for the fifth section, which aims to give the reader a feel for orbital integrals on $p$-adic groups by calculating lots of them for $GL_2$; a side benefit is that the calculations illustrate the phenomenon of homogeneity that is the subject of DeBacker's article in this volume. The sixth section establishes the existence of the Shalika germ expansion on $G$. The seventh section proves the Weyl integration formula, a simple but important ingredient in the local trace formula. The eighth section begins our discussion of the local trace formula. The ninth and tenth sections prove the local trace formula on $\mathfrak{g}$ for $G = GL_2$ and derive from it the fact that $\hat{I}_\phi$ is represented by a function for any $\phi \in C_c^\infty(\mathfrak{g}_e)$ (see (25.4.3) in order to understand why the function $I_\phi$ considered in section 10 is essentially the same as the one defined earlier in this introduction).

The remainder of the article is less elementary, though it is still almost completely self-contained. To keep the structure theory of $G$ as simple as possible, we usually assume that $G$ is split. The eleventh section (on certain convex cones and polytopes in Euclidean space) is quite technical and should be consulted only as needed while reading later sections. The twelfth section proves some basic facts about the weight factors occurring in Arthur's weighted orbital integrals. Once the definitions have been understood, the reader can move on, returning to the lemmas proved in this section when they are referred to later. The next four sections concern descent, both parabolic and semisimple, which is used to perform reduction steps in later proofs. The seventeenth section proves some relatively easy results about Shalika germs on $\mathfrak{g}$: homogeneity (which lets one define Shalika germs as canonical functions on $\mathfrak{g}_{\mathrm{rs}}$, not just germs of functions) and local boundedness (see Theorem 17.9) of normalized Shalika germs. As a consequence we obtain the boundedness (see Theorem 17.10) of the function $X \mapsto I_X(f)$ on $\mathfrak{g}_{\mathrm{rs}}$, where $I_X$ denotes the normalized orbital integral over the orbit of $X \in \mathfrak{g}_{\mathrm{rs}}$.

In the next two sections we study norms on the set $X(F)$ of $F$-points on a variety $X$ (usually affine) over a field $F$ equipped with a non-trivial absolute value. It is standard practice to use such norms on $\mathfrak{g}$ and $G$, but it seems useful to study them in greater generality, so that one can also take $X$ to be a $G$-orbit in $\mathfrak{g}$, for example. Most of this material is very easy, the one result requiring some work being Proposition 18.3. The twentieth section uses this theory of norms to estimate weighted orbital integrals.

The next four sections prove the local trace formula on $\mathfrak{g}$ (for any split group $G$), including the $K$-theoretic version (see 24.5) as well as the standard one (Theorem 24.1) involving volumes of convex hulls. The formula simplifies (see (24.10.8) and (24.10.9)) when one of the test functions is a cusp form. In the next section we use the local trace formula to prove (see Theorem 25.1) the facts about $\hat{I}_\phi$ mentioned earlier in this introduction.

The next two sections apply Theorem 25.1, Howe's finiteness theorem (see 26.2) and the elementary part of the theory of Shalika germs in order to prove the rest of the main results in harmonic analysis on $\mathfrak{g}$, namely Theorem 27.5 (linear independence of Shalika germs and density of regular semisimple orbital integrals), Theorem 27.8 ($\hat{I}$ is represented by a function when $I$ is an invariant distribution whose support is bounded modulo conjugation), and Theorem 27.12 (the Lie algebra analog of Harish-Chandra's local character expansion for $\Theta_\pi$).

The last section is a guide to some of the notation used in this article.

## 1. Local trace formula for compact groups $G$

In this section $G$ denotes a compact (Hausdorff) topological group and $dg$ denotes the unique Haar measure on $G$ that gives measure 1 to $G$.

**1.1. Finite dimensional representations of $G$.** We need to spend a moment discussing finite dimensional representations $(\pi, V)$ of $G$. In other words we are interested in continuous linear actions $G \times V \to V$, where $V$ is a finite dimensional complex vector space. Continuity means that the map $G \times V \to V$ is continuous; linearity means that for each $g \in G$ the map $v \mapsto gv$ is a linear transformation $\pi(g)$ from $V$ to itself.

We write $(\pi^*, V^*)$ for the contragredient representation of $(\pi, V)$. Here $V^*$ is the vector space dual to $V$, and $\pi^*$ is the obvious representation of $G$ on $V^*$, characterized by the equation

$$\langle gv^*, gv \rangle = \langle v^*, v \rangle \tag{1.1.1}$$

for all $v \in V$, $v^* \in V^*$, $g \in G$, the pairing on both sides of this equation being the canonical one given by evaluating linear functionals on vectors.

**1.2. Group algebra $C(G)$.** For our purposes the best version of the group algebra of $G$ is obtained by taking the vector space $C(G)$ of continuous complex-valued functions on $G$, viewed as a $\mathbb{C}$-algebra using convolution. Recall that the convolution $f_1 * f_2$ of two functions $f_1$, $f_2 \in C(G)$ is the function on $G$ whose value at $x \in G$ is given by

$$\int_G f_1(xg^{-1}) f_2(g)\, dg. \tag{1.2.1}$$

Let $(\pi, V)$ be a finite dimensional representation of $G$. Then the group algebra $C(G)$ acts on $V$ in a natural way. For $f \in C(G)$ we denote by $\pi(f)$ the linear transformation by which $f$ acts on $V$; it is defined by

$$\pi(f)(v) = \int_G f(g) gv\, dg. \tag{1.2.2}$$

**1.3. Characters of finite dimensional representations of $G$.** Let $(\pi, V)$ be a finite dimensional representation of $G$. We write $\Theta_\pi$ for the character of $\pi$, which is by definition the function on $G$ defined by

$$\Theta_\pi(g) = \operatorname{trace} \pi(g). \tag{1.3.1}$$

Similarly, for $f \in C(G)$ we define a complex number $\Theta_\pi(f)$ by

$$\Theta_\pi(f) = \operatorname{trace} \pi(f). \tag{1.3.2}$$

It is clear from the definitions that

$$\Theta_\pi(f) = \int_G f(g)\Theta_\pi(g)\,dg. \tag{1.3.3}$$

**1.4. Function space $L^2(G)$.** We use $dg$ to form the space $L^2(G)$ of square-integrable functions on $G$. The group $G \times G$ acts by unitary transformations on the Hilbert space $L^2(G)$, the action of $(g_1, g_2) \in G \times G$ on $\varphi \in L^2(G)$ being given by the rule

$$\big((g_1, g_2)\varphi\big)(x) = \varphi(g_1^{-1} x g_2). \tag{1.4.1}$$

The $(G \times G)$-module $L^2(G)$ can also be viewed as a $C(G) \otimes_{\mathbb{C}} C(G)$-module, the action of $f_1 \otimes f_2 \in C(G) \otimes_{\mathbb{C}} C(G)$ on $\varphi \in L^2(G)$ being given by the following integrated form of (1.4.1):

$$\big((f_1 \otimes f_2)\varphi\big)(x) = \int_G \int_G f_1(g_1) f_2(g_2) \varphi(g_1^{-1} x g_2)\,dg_1\,dg_2. \tag{1.4.2}$$

In the integral over $g_2 \in G$ we may replace $g_2$ by $x^{-1} g_1 g_2$, obtaining

$$\big((f_1 \otimes f_2)\varphi\big)(x) = \int_G \int_G f_1(g_1) f_2(x^{-1} g_1 g_2) \varphi(g_2)\,dg_1\,dg_2, \tag{1.4.3}$$

which shows that $f_1 \otimes f_2$ acts by an integral operator whose kernel function $K$ is given by

$$K(x, y) = \int_G f_1(g) f_2(x^{-1} g y)\,dg. \tag{1.4.4}$$

Clearly this kernel is a continuous (hence square-integrable) function on $G \times G$, so that the action of $f_1 \otimes f_2$ on $L^2(G)$ is given by a Hilbert-Schmidt operator. Similarly (and even more simply) the left-translation (resp., right-translation) action of $f_1$ (resp., $f_2$) on $L^2(G)$ is given by a continuous kernel, hence by a Hilbert-Schmidt operator; since the product of the Hilbert-Schmidt operators obtained from $f_1$ and $f_2$ separately gives the action of $f_1 \otimes f_2$, we see that $f_1 \otimes f_2$ is a trace class operator whose trace is equal to the integral of the kernel $K$ over the diagonal:

$$\operatorname{trace}(f_1 \otimes f_2; L^2(G)) = \int_G \int_G f_1(g) f_2(x^{-1} g x)\,dg\,dx. \tag{1.4.5}$$

This equation is a preliminary form of the trace formula for the compact group $G$. We will modify both sides of (1.4.5) in order to get the final form of the trace formula for $G$. To rewrite the left side we will use the Peter-Weyl theorem.

**1.5. Peter-Weyl theorem.** The Peter-Weyl theorem (see the book [**Kna86**] by Knapp, for example) tells us that the $(G \times G)$-module $L^2(G)$ is isomorphic to the Hilbert space direct sum

$$\hat{\bigoplus}_{(\pi,V)} V \otimes_{\mathbb{C}} V^*, \tag{1.5.1}$$

where $(\pi, V)$ runs over a set of representatives for the isomorphism classes of irreducible finite dimensional representations of $G$. We have already discussed the $G$-module structure on $V^*$. We regard $V \otimes_{\mathbb{C}} V^*$ as a $(G \times G)$-module by the following rule:

$$(g_1, g_2)(v \otimes v^*) = g_1 v \otimes g_2 v^*. \tag{1.5.2}$$

Therefore the left (spectral) side of the trace formula can be rewritten as

$$\sum_{(\pi,V)} \Theta_\pi(f_1)\Theta_{\pi^*}(f_2). \tag{1.5.3}$$

**1.6. Final form of the trace formula.** Now we manipulate the right (geometric) side of the trace formula. Note that the right side of (1.4.5) can be rewritten as

$$\int_G \int_G f_1(y^{-1}gy) f_2(x^{-1}gx)\, dg\, dx \tag{1.6.1}$$

for any $y \in G$. [Change variables twice: first replace $x$ by $y^{-1}x$, then $g$ by $y^{-1}gy$.] Integrating over $y$ and changing the order of integration, we arrive at the final form of the trace formula:

$$\sum_{(\pi,V)} \Theta_\pi(f_1)\Theta_{\pi^*}(f_2) = \int_G \tilde{f}_1(g)\tilde{f}_2(g)\, dg, \tag{1.6.2}$$

where for any $f \in C(G)$ we define $\tilde{f} \in C(G)$ by

$$\tilde{f}(g) = \int_G f(x^{-1}gx)\, dx. \tag{1.6.3}$$

Thus $\tilde{f}(g)$ is obtained by integrating $f$ over the orbit (or conjugacy class) of $g$; for this reason $\tilde{f}(g)$ is known as an "orbital integral." Obviously the function $\tilde{f}$ is constant on orbits.

**1.7. Algebraic form of the Peter-Weyl theorem.** Consider once again the Peter-Weyl theorem isomorphism (of $(G \times G)$-modules)

$$L^2(G) \cong \hat{\bigoplus}_{(\pi,V)} V \otimes_{\mathbb{C}} V^*. \tag{1.7.1}$$

Inside the Hilbert space direct sum on the right side of (1.7.1) we have the algebraic direct sum, which can be characterized as the set of vectors $u$ such that the $(G\times G)$-module generated by $u$ is finite dimensional. Under the Peter-Weyl isomorphism these correspond to functions $\varphi \in L^2(G)$ that are left and right $G$-finite, in the sense that the $(G \times G)$-submodule of $L^2(G)$ generated by $\varphi$ is finite dimensional; it turns out that such functions are automatically continuous. Thus we obtain the algebraic form of the Peter-Weyl theorem

$$C(G)_0 \cong \bigoplus_{(\pi,V)} V \otimes_{\mathbb{C}} V^*, \tag{1.7.2}$$

where $C(G)_0$ denotes the space of left and right $G$-finite continuous functions on $G$.

We have not yet specified how we are normalizing the Peter-Weyl isomorphism. To do so we note that $V \otimes_{\mathbb{C}} V^*$ is canonically isomorphic to $\mathrm{End}_{\mathbb{C}}(V)$ (even as $(G \times G)$-module). In our normalization of the Peter-Weyl isomorphism a function $f \in C(G)_0$ maps to the family of elements $\pi(f) \in \mathrm{End}_{\mathbb{C}}(V) = V \otimes_{\mathbb{C}} V^*$. In particular for $f \in C(G)_0$ we have $\pi(f) = 0$ (and hence $\Theta_\pi(f) = 0$) for all but finitely many isomorphism classes of irreducible finite dimensional representations $(\pi, V)$.

**1.8. Fourier transforms of orbital integrals.** For any irreducible finite dimensional representation $(\pi, V)$ of $G$ the linear functional $f \mapsto \Theta_\pi(f)$ on $C(G)_0$ is conjugation invariant, and it is clear from the algebraic Peter-Weyl theorem that any conjugation invariant linear functional is an infinite linear combination of these basic ones. In particular this is the case for the orbital integral

$$f \mapsto \int_G f(x^{-1}gx)\,dx. \tag{1.8.1}$$

In fact we have the following formula for any $g \in G$ and $f \in C(G)_0$:

$$\int_G f(x^{-1}gx)\,dx = \sum_{(\pi,V)} \Theta_\pi(g)\Theta_{\pi^*}(f). \tag{1.8.2}$$

How does one prove this formula? Both sides of it are continuous functions of $g$; to prove that they are equal it is enough to show that they have the same integral against an arbitrary continuous function on $G$, and this is just a restatement of the preliminary form (1.4.5) of the trace formula.

**1.9. Plancherel formula.** In the special case $g = 1$ equation (1.8.2) yields the Plancherel formula (valid for any $f \in C(G)_0$)

$$f(1) = \sum_{(\pi,V)} \dim(\pi)\Theta_\pi(f). \tag{1.9.1}$$

(Here we used that $\pi$ and $\pi^*$ have the same dimension.)

**1.10. Matrix coefficients.** Let $(\pi, V)$ be an irreducible finite dimensional representation of $G$. For $v \in V$, $v^* \in V^*$ such that $\langle v^*, v\rangle = 1$ we define functions $f_{v^*,v}$ and $f_{v,v^*}$ on $G$ by $f_{v^*,v}(g) = \langle v^*, gv\rangle$ and $f_{v,v^*}(g) = \langle gv^*, v\rangle$. Both functions lie in $C(G)_0$. The function $f_{v^*,v}$ is a *matrix coefficient* for $\pi$, while $f_{v,v^*}$ is a matrix coefficient for $\pi^*$. The two functions are related by $f_{v,v^*}(g) = f_{v^*,v}(g^{-1})$.

We can use matrix coefficients to give an explicit formula for the inverse $\beta$ of the isomorphism $\alpha$ appearing in the algebraic Peter-Weyl theorem. Recall that $\alpha(f)$ is the element of

$$\bigoplus_{(\pi,V)} \mathrm{End}_{\mathbb{C}}(V) \tag{1.10.1}$$

whose $(\pi, V)$-th component is $\pi(f)$. We will give an explicit formula for $\beta$ on each summand $V \otimes_{\mathbb{C}} V^* = \mathrm{End}_{\mathbb{C}}(V)$. For $v \otimes v^* \in V \otimes_{\mathbb{C}} V^*$ we claim that $\beta(v \otimes v^*) = \dim\pi \cdot f_{v,v^*}$.

Why is this the right formula for $\beta$? The representations $V \otimes_{\mathbb{C}} V^*$ of $G \times G$ are irreducible and pairwise non-isomorphic, and the map $v \otimes v^* \mapsto f_{v,v^*}$ is $(G \times G)$-equivariant and non-zero. Therefore it is clear that there exists a scalar $c_\pi$ such that $\beta(v \otimes v^*) = c_\pi \cdot f_{v,v^*}$. Taking $f = f_{v,v^*}$ in the Plancherel formula (1.9.1), we see that $c_\pi = \dim\pi$.

**1.11. Orbital integrals of matrix coefficients.** Let $(\pi, V)$ be an irreducible finite dimensional representation of $G$. Let $g \in G$. Then from (1.8.2) it follows easily that

$$\int_G f_{v^*,v}(x^{-1}gx)\,dx = (\dim\pi)^{-1} \cdot \Theta_\pi(g). \tag{1.11.1}$$

Thus the orbital integrals of the matrix coefficient $f_{v^*,v}$ for $\pi$ give the character values of $\pi$ (up to the scalar $(\dim \pi)^{-1}$). We have proved this as a consequence of the Peter-Weyl theorem and the trace formula for $G$, but in fact there is a simple direct proof, as the reader may wish to devise as an exercise. [Hint: Consider the endomorphism $\int_G \pi(x^{-1}gx)\,dx$ of $V$.]

**1.12. Comments.** Our goal here has not been to develop harmonic analysis on compact groups in the most efficient way, but rather to concentrate on the trace formula and its relationship to other basic concepts, stressing those, such as orbital integrals, that we will meet again in the non-compact case.

A more standard treatment would emphasize the orthogonality relations (for irreducible characters and for matrix coefficients). The trace formula for $G$ has essentially the same information in it, but packaged in a slightly different way, as we have tried to illustrate.

## 2. Basics of integration

**2.1. l.c.t.d spaces.** What kind of spaces will we be integrating over? In this article we are interested in $p$-adic groups and Lie algebras, so the topological spaces we will encounter will be l.c.t.d spaces (short for locally compact and totally disconnected). Thus an l.c.t.d space is a Hausdorff topological space in which every point has a neighborhood basis of compact open subsets.

On an l.c.t.d space $X$ the most important space of functions is $C_c^\infty(X)$, the space of all locally constant, compactly supported, complex-valued functions on $X$. Any such function can be written as a linear combination of characteristic functions of compact open subsets of $X$. This makes integration rather easy, at least in principle: we just need to assign measures to compact open subsets.

Let $X$ be an l.c.t.d space and let $Y$ be a closed subset with complementary open subset $U$. Then both $Y$ and $U$ are themselves l.c.t.d spaces, and it is an instructive exercise to check that the sequence

$$0 \to C_c^\infty(U) \to C_c^\infty(X) \to C_c^\infty(Y) \to 0 \tag{2.1.1}$$

is exact. (The first map is given by extending by 0, the second by restriction to $Y$.)

Sometimes it is useful to consider vector-valued functions. For any complex vector space $V$ we write $C_c^\infty(X;V)$ for the space of locally constant, compactly supported functions on $X$ with values in $V$. It is easy to check that

$$C_c^\infty(X;V) = C_c^\infty(X) \otimes_{\mathbb{C}} V. \tag{2.1.2}$$

Lemma 2.1. *Let $X$, $Y$ be l.c.t.d topological spaces. Then the product $X \times Y$ is also a l.c.t.d space, and moreover there are equalities*

$$C_c^\infty(X \times Y) = C_c^\infty(X; C_c^\infty(Y)) = C_c^\infty(X) \otimes_{\mathbb{C}} C_c^\infty(Y).$$

We leave the proof to the reader as another exercise.

**2.2. l.c.t.d groups.** An l.c.t.d topological group is by definition a locally compact Hausdorff topological group $G$ in which the identity element has a neighborhood basis of compact open subgroups. Clearly $G$ is then a l.c.t.d topological space. For us a typical example is $G(F)$, where $F$ is a $p$-adic field and $G$ is a linear algebraic group over $F$. To see that $G(F)$ is a l.c.t.d group we reduce to the case of the general linear group $GL_n$ (by choosing an embedding of $G$ in a general linear group); in $GL_n(F)$ the compact open subgroups

$$K_n = \{g \in GL_n(\mathcal{O}) : g \equiv 1 \mod \pi^n\} \tag{2.2.1}$$

give the desired neighborhood basis at the identity element. Here (and throughout the article) we write $\mathcal{O}$ for the valuation ring in $F$ and $\pi$ for a generator of the maximal ideal of $\mathcal{O}$.

**2.3. Unimodular groups.** Any locally compact Hausdorff topological group $G$ admits a left invariant Radon measure $dg$, known as a (left) Haar measure, and $dg$ is unique up to a positive scalar. Since right translations commute with left translations, a right translate of $dg$ is another Haar measure, hence is a positive multiple of $dg$; in this way one obtains the *modulus character* $\delta_G$ (with values in the multiplicative group of positive real numbers) characterized by the property

$$d(gh^{-1}) = \delta_G(h) \cdot dg \tag{2.3.1}$$

or, equivalently,

$$d(hgh^{-1}) = \delta_G(h) \cdot dg. \tag{2.3.2}$$

When the modulus character is trivial, one says that $G$ is *unimodular*. In this case $dg$ is both left and right invariant and $d(g^{-1}) = dg$.

For a reductive group $G$ over our $p$-adic field $F$ the group $G(F)$ is always unimodular. This stems from the fact that $G$ acts trivially on the top exterior power of the Lie algebra of $G$. On the other hand, for any proper parabolic subgroup $P$ of $G$, the group $P(F)$ is not unimodular.

On a l.c.t.d group $G$ integration is particularly simple. Fix some compact open subgroup $K_0$. Then there is a unique Haar measure $dg$ giving $K_0$ measure 1. For any compact open subgroup $K$ of $G$ the measure of $K$ is

$$[K : K \cap K_0] \cdot [K_0 : K \cap K_0]^{-1}. \tag{2.3.3}$$

Moreover for any compact open subset $S$ of $G$ there is a compact open subgroup $K$ small enough that $S$ is a disjoint union of cosets $gK$, so that the measure of $S$ is the number of such cosets times the measure of $K$. That's all we need to know about integration on l.c.t.d groups!

**2.4. Integration on homogeneous spaces.** Let $G$ be a unimodular locally compact Hausdorff topological group and let $H$ be a closed subgroup. Then there exists a right $G$-invariant Radon measure on $H\backslash G$ if and only if $H$ is unimodular. Assume this is so, and assume also that $G$ (and hence $H$) is a l.c.t.d group.

Choose Haar measures $dg$, $dh$ on $G$, $H$ respectively. Then there is a quotient measure (right $G$-invariant) $dg/dh$ on $H\backslash G$ characterized by the formula (integration in stages)

$$\int_G f(g)dg = \int_{H\backslash G} \int_H f(hg)\, dh\, dg/dh, \tag{2.4.1}$$

valid for all $f \in C_c^\infty(G)$.

The reason why the integration in stages formula characterizes the invariant integral on $H\backslash G$ is that any function in $C_c^\infty(H\backslash G)$ lies in the image of the linear map

$$C_c^\infty(G) \to C_c^\infty(H\backslash G) \qquad (2.4.2)$$
$$f \mapsto f^\sharp$$

defined by putting

$$f^\sharp(g) = \int_H f(hg)\, dh. \qquad (2.4.3)$$

Again we can see rather concretely how the measure works. Indeed, any compact open subset of our homogeneous space can be written as a disjoint union of ones of the form $H\backslash HgK$ (for some compact open subgroup $K$ of $G$), and the measure of $H\backslash HgK$ is given by

$$\operatorname{meas}_{dg}(K)/\operatorname{meas}_{dh}(H \cap gKg^{-1}), \qquad (2.4.4)$$

as one sees by applying integration in stages to the characteristic function of $gK$.

**2.5. Integration in stages in reversed order.** We continue with $G$, $H$ as above. Later, when discussing descent for orbital integrals, we will need the integral formula in Lemma 2.3 below. As a warm-up exercise, we first prove a simpler statement, which can be thought of as a version of integration in stages in which the order of integration has in a sense been reversed. This reversed formula involves a compact open subset $C$ of $H\backslash G$. We also need to choose $\alpha \in C_c^\infty(G)$ such that $\alpha^\sharp = 1_C$.

LEMMA 2.2. *For all $f \in C_c^\infty(G)$ such that the image of* $\operatorname{Supp}(f)$ *under $G \twoheadrightarrow H\backslash G$ is contained in $C$ there is an equality*

$$\int_G f(g)\, dg = \int_H \Big(\int_G f(hg)\alpha(g)\, dg\Big)\, dh. \qquad (2.5.1)$$

PROOF. Change variables in the integral on the right, replacing $g$ by $h^{-1}g$, then reverse the order of integration, then replace $h$ by $h^{-1}$, noting that $d(h^{-1}) = dh$, a consequence of the unimodularity of $H$. $\square$

For descent theory we will actually need a variant of (2.5.1), in which we are also given a closed unimodular subgroup $I$ of $H$ and a Haar measure $di$ on $I$. With $C$, $\alpha$ as before we then have the following lemma.

LEMMA 2.3. *For any integrable, locally constant function $f$ on $I\backslash G$ such that the image of* $\operatorname{Supp}(f)$ *in $H\backslash G$ is contained in $C$ there is an equality*

$$\int_{I\backslash G} f(g)\, dg/di = \int_{I\backslash H} \phi(h)\, dh/di \qquad (2.5.2)$$

*where $\phi$ is the integrable, locally constant function on $I\backslash H$ defined by*

$$\phi(h) = \int_G f(hg)\alpha(g)\, dg.$$

PROOF. Let $\beta \in C_c^\infty(G)$ and consider $\beta^\sharp \in C_c^\infty(H\backslash G)$ defined as above by

$$\beta^\sharp(g) = \int_H \beta(hg)\, dh = \int_H \beta(h^{-1}g)\, dh.$$

Then, abbreviating $dg/di$ and $dh/di$ to $d\dot{g}$ and $d\dot{h}$ respectively, we have

$$\begin{aligned}\int_{I\backslash G} f(g)\beta^\sharp(g)\,d\dot{g} &= \int_{I\backslash G} f(g)\int_H \beta(h^{-1}g)\,dh\,d\dot{g}\\ &= \int_{I\backslash G} f(g)\int_{I\backslash H}\int_I \beta(h^{-1}i^{-1}g)\,di\,d\dot{h}\,d\dot{g}\\ &= \int_{I\backslash H}\int_{I\backslash G}\int_I f(g)\beta(h^{-1}ig)\,di\,d\dot{g}\,d\dot{h}\\ &= \int_{I\backslash H}\int_G f(g)\beta(h^{-1}g)\,dg\,d\dot{h}\\ &= \int_{I\backslash H}\int_G f(hg)\beta(g)\,dg\,d\dot{h}.\end{aligned}$$

To see that all these integrals are convergent, replace $f$ and $\beta$ by their absolute values and note that the integral we started with is obviously convergent, since $f$ is integrable and $\beta^\sharp$ is bounded. Fubini's theorem takes care of the rest.

Taking $\beta = \alpha$, we obtain

$$\int_{I\backslash G} f(g)1_C(g)\,d\dot{g} = \int_{I\backslash H}\phi(h)d\dot{h}, \tag{2.5.3}$$

which in view of our assumption on the support of $f$ yields the equality stated in the lemma. We have seen along the way that $\phi$ is integrable, and it is obviously locally constant. □

## 3. Preliminaries about orbital integrals

**3.1. The set-up.** We are going to discuss orbital integrals on $G(F)$, where $G$ is a connected reductive group over a $p$-adic field $F$. We fix an algebraic closure $\bar{F}$ of $F$.

**3.2. Orbits.** Let $\gamma \in G(F)$. We are interested in the orbit $O(\gamma)$ of $\gamma$ for the conjugation action of $G$ on itself. In other words $O(\gamma)$ is the conjugacy class of $\gamma$, a locally closed subset $G$ in the Zariski topology, isomorphic as variety to $G/G_\gamma$, where $G_\gamma$ denotes the centralizer of $\gamma$ in $G$ (see [**Bor91**, Prop. 6.7]). There is an exact sequence of pointed sets

$$1 \to G_\gamma(F) \to G(F) \to (G/G_\gamma)(F) \to H^1(F, G_\gamma) \to H^1(F, G) \tag{3.2.1}$$

where $H^1(F, G)$ denotes the Galois cohomology set $H^1(\mathrm{Gal}(\bar{F}/F), G(\bar{F}))$, and the boundary map in this sequence induces a bijection from the set of $G(F)$-orbits in $O(\gamma)(F)$ to the set

$$\ker[H^1(F, G_\gamma) \to H^1(F, G)]. \tag{3.2.2}$$

Since $H^1(F, G_\gamma)$ is a finite set (see [**Ser02**, Ch. III,§4]), there are in fact only finitely many such orbits. From the theory of $p$-adic manifolds (see [**Ser92**]) one knows first of all that each $G(F)$-orbit is open in $O(\gamma)(F)$, hence also closed in $O(\gamma)(F)$, and second of all that the $G(F)$-orbit of $\gamma$ is isomorphic as $p$-adic manifold (hence also as topological space) to the homogeneous space $G(F)/G_\gamma(F)$. Since $O(\gamma)$ is locally closed in $G$, the set $O(\gamma)(F)$ is locally closed in $G(F)$, and it follows that the same is true of each individual $G(F)$-orbit in $O(\gamma)(F)$.

When $\gamma$ is semisimple, $O(\gamma)$ is closed in $G$ (see [**Bor91**, Thm. 9.2]), hence $O(\gamma)(F)$ and the individual $G(F)$-orbits in it are all closed in $G(F)$.

It is instructive to consider the example of the group $GL_2$. There is map $\alpha$ of algebraic varieties from $GL_2$ to the affine plane $\mathbb{A}^2$, defined by $g \mapsto (\operatorname{trace}(g), \det(g))$. On $\mathbb{A}^2$ we have the discriminant function $D$, defined by $D(b,c) = b^2 - 4c$. All fibers of $\alpha$ are of course closed. The fiber of $\alpha$ over a point $(b,c)$ where $D$ is non-zero consists of a single orbit of regular semisimple elements. (An element in $GL_2$ is regular semisimple if and only if it has distinct eigenvalues.) The fiber of $\alpha$ over a point $(b,c)$ where $D$ vanishes is the union of two $G$-orbits, namely those of the matrices

$$\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \text{ and } \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix},$$

where we have written $a$ for $b/2$. The first matrix is semisimple (but not regular), and its orbit is closed in the fiber. The second matrix is regular (but not semisimple), and its orbit is open in the fiber. A very special feature of the group $GL_2$ (and, more generally, of $GL_n$) is that the Galois cohomology set $H^1(F, G_\gamma)$ is always trivial, so that $O(\gamma)(F)$ always consists of a single $G(F)$-orbit.

A map similar to $\alpha$ exists for any $G$. For $GL_n$ one simply uses all the coefficients of the characteristic polynomial of a matrix to define a map from $GL_n$ to $\mathbb{A}^n$. In general one uses the morphism $G \to G/\operatorname{Int}(G)$, where $G/\operatorname{Int}(G)$ is by definition the affine scheme whose ring of regular functions is the ring of conjugation invariant regular functions on $G$. Later (see 14.2) we will discuss the analogous construction for the Lie algebra of $G$ in greater detail.

**3.3. Definition of orbital integrals.** Let $\gamma \in G(F)$. The orbital integral $O_\gamma(f)$ of a function $f \in C_c^\infty(G(F))$ is by definition the integral

$$O_\gamma(f) := \int_{G_\gamma(F)\backslash G(F)} f(g^{-1}\gamma g)\, d\dot{g} \tag{3.3.1}$$

where $d\dot{g}$ is a right $G(F)$-invariant measure on the homogeneous space over which we are integrating. Thus $O_\gamma$ depends on a choice of measure, but once this choice is made we get a well-defined linear functional on $C_c^\infty(G(F))$. (We are not putting any topology on our function space, so there is no continuity requirement in the definition of linear functional.)

Two comments are needed. First, we need to know that $G_\gamma(F)$ is unimodular in order to ensure that $d\dot{g}$ exists. For semisimple elements $\gamma$ there is no problem, since then $G_\gamma$ is reductive. In general, however, $G_\gamma$ is not reductive, and we need to argue as follows. By the Jordan decomposition (see [**Bor91**]) we can decompose $\gamma$ uniquely as $\gamma = su = us$ with $s$ semisimple and $u$ unipotent. It follows that $u \in G_s$ and that $G_\gamma$ coincides with the centralizer of $u$ in the reductive group $G_s$. This reduces us to the case in which $\gamma$ is unipotent. Then (since the characteristic of our field is 0) we can write $\gamma$ as the exponential of a nilpotent element in the Lie algebra over $G$ over $F$. Using a $G$-invariant non-degenerate symmetric bilinear form to identify the Lie algebra with its dual, we see that it is enough to prove that the stabilizer (for the coadjoint action) of any element in the dual of the Lie algebra is unimodular. This is equivalent to the statement that every coadjoint orbit carries a $G(F)$-invariant measure, which in turn follows from the fact that every coadjoint orbit admits a $G$-invariant structure of symplectic manifold and hence admits a $G$-invariant volume form (which can then be used to construct a

$G(F)$-invariant measure on the $F$-points of the orbit). See 17.3 for a discussion of the symplectic structure on coadjoint orbits.

Second, we need to know that the integral converges. For semisimple elements there is again no problem, since the orbit is closed, which ensures that the integrand is a compactly supported (and locally constant) function on the homogenous space, which is exactly the sort of function we can integrate. For arbitrary $\gamma$ the orbit is only locally closed, and while the integrand is still locally constant, it is not necessarily compactly supported. (Of course the integral is still discrete in nature, but it boils down to an infinite series rather than a finite sum, so that convergence is not obvious.) For a proof of convergence see [**Rao72**], and for a slightly different perspective on the geometry involved see [**Pan91**]. The idea is to reduce to the case of nilpotent orbital integrals and then to show that the $G$-invariant volume form on a nilpotent orbit extends (without singularities) to a suitable desingularization (constructed using the theory of $\mathfrak{sl}(2)$-triples) of the closure of that nilpotent orbit.

**3.4. Orbital integrals of characteristic functions of double cosets.** We continue with $\gamma \in G(F)$. We are going to lighten notation by writing $G$ and $G_\gamma$ instead of $G(F)$ and $G_\gamma(F)$. The material in this subsection will be used in section 5, when we calculate orbital integrals of functions in the spherical Hecke algebra of $GL_2(F)$.

Let $K$ be a compact open subgroup of $G$, and write $X$ for the homogeneous space $G/K$. Since $K$ is open, $X$ has the discrete topology. We write $x_0$ for the base point in $X$ (given by the trivial coset of $K$ in $G$). Given $(x_1, x_2) \in X \times X$ we pick $g_1, g_2 \in G$ such that $x_i = g_i x_0$ for $i = 1, 2$. The double coset $Kg_2^{-1}g_1K$ is well-defined and will be denoted by $\operatorname{inv}(x_1, x_2)$. It follows immediately from these definitions that the map $\operatorname{inv} : X \times X \to K\backslash G/K$ induces a bijection from the set of $G$-orbits on $X \times X$ to $K\backslash G/K$ (with $G$ acting on $X \times X$ by $g(x_1, x_2) = (gx_1, gx_2)$). Here "inv" is short for "invariant". The reason for this name is that $\operatorname{inv}(x_1, x_2)$ is an invariant measuring the relative position of the two points $x_1$ and $x_2$.

For any $a \in G$ we can consider the double coset $KaK$, a compact open subset of $G$, as well as its characteristic function $1_{KaK}$, which lies in $C_c^\infty(G)$. The orbital integrals of $1_{KaK}$ can be understood using the action of $G$ on $X$. Indeed, from (2.4.4) it follows that

$$\int_{G_\gamma\backslash G} 1_{KaK}(g^{-1}\gamma g)\ dg/dg_\gamma = \sum_x \frac{\operatorname{meas}(K)}{\operatorname{meas}(\operatorname{Stab}_{G_\gamma}(x))}, \tag{3.4.1}$$

where the sum on the right side runs over a set of representatives for the $G_\gamma$-orbits on the set of elements $x \in X$ such that $\operatorname{inv}(\gamma x, x) = KaK$, and the measures are taken with respect to the Haar measures $dg$, $dg_\gamma$ on $G$, $G_\gamma$ respectively; $\operatorname{Stab}_{G_\gamma}(x)$ denotes the stabilizer of $x$ in $G_\gamma$, a compact open subgroup of $G_\gamma$.

We may replace $G_\gamma$ by any convenient closed subgroup $G'_\gamma$ such that $G'_\gamma\backslash G_\gamma$ is compact. This multiplies the orbital integral by the factor $\operatorname{meas}(G'_\gamma\backslash G_\gamma)$, but (3.4.1) remains valid, with $G'_\gamma$ replacing $G_\gamma$ everywhere. In particular, when $G_\gamma$ is compact, we may take $G'_\gamma$ to be the trivial subgroup. Then, if we use the Haar measure $dg$ on $G$ giving $K$ measure 1, we have

$$\int_G 1_{KaK}(g^{-1}\gamma g)\ dg = |\{x \in X : \operatorname{inv}(\gamma x, x) = KaK\}|, \tag{3.4.2}$$

where $|S|$ is being used to denote the cardinality of a finite set $S$, showing that our orbital integral is the answer to a simple counting problem involving the action of $G$ on $X$. In section 5 we will solve such counting problems for $G = GL_2(F)$, $K = GL_2(\mathcal{O})$ using the tree for $SL_2(F)$, but first we need to discuss two double coset decompositions.

## 4. Cartan and Iwasawa decompositions

When calculating orbital integrals for $GL_2$, we are going to need both the Cartan and Iwasawa decompositions. Later we will need them for all split groups, as this is the context in which we will discuss the Lie algebra version of the local trace formula. In this section we will state both decompositions for all split groups and will sketch proofs for $GL_n$.

**4.1. Notation.** Let $F$ be a $p$-adic field with valuation ring $\mathcal{O}$ and uniformizing element $\pi$ (so that the valuation of $\pi$ is 1). For $x \in F^\times$ we denote by $\mathrm{val}(x)$ the valuation of $x$.

In this section $G$ denotes a split connected reductive group scheme over $\mathcal{O}$. We need to choose various $\mathcal{O}$-subgroup schemes in $G$: a split maximal torus $A$, and a Borel subgroup $B = AN$ containing $A$ and having unipotent radical $N$. We write $A_G$ for the identity component of the center of $G$. In the case of $G = GL_n$ we make the standard choices: $B$ consists of upper-triangular matrices, $A$ of diagonal matrices, and $N$ of upper-triangular matrices with 1's on the diagonal. We write $W$ for the Weyl group of $A$.

We write $G_{\mathrm{der}}$ for the derived group of the algebraic group $G$, and we write $G_{\mathrm{sc}}$ for its simply connected cover.

We will need the group $X_*(A)$ of cocharacters of $A$ (in other words, the group of homomorphisms from the multiplicative group $\mathbb{G}_m$ to $A$). The cocharacter group is a free abelian group whose rank is equal to the dimension of $A$ over $F$. For $GL_n$ we identify $X_*(A)$ with $\mathbb{Z}^n$ as follows: to the $n$-tuple $(j_1, \dots, j_n)$ corresponds the cocharacter which sends an element $z$ in the multiplicative group to the diagonal matrix whose diagonal entries are $(z^{j_1}, \dots, z^{j_n})$.

To lighten notation we write $K$ for $G(\mathcal{O})$ and then abbreviate $G(F)$, $A(F)$, $B(F)$, $N(F)$ to $G$, $A$, $B$, $N$ respectively.

**4.2. The isomorphism $A/A \cap K \simeq X_*(A)$.** For $GL_n$ the map which sends a diagonal matrix to the $n$-tuple of valuations of the diagonal entries induces an isomorphism from $A/A \cap K$ to $\mathbb{Z}^n$. In general we have a canonical isomorphism $A/A \cap K \simeq X_*(A)$, under which a cocharacter $\mu$ corresponds to the (class of) the element $\pi^\mu \in A$ obtained by applying the cocharacter to the element $\pi$ in the multiplicative group.

**4.3. Cartan decomposition.** The crude version of the Cartan decomposition states simply that $G = KAK$. It is an instructive exercise to prove this for $GL_n$. [Start with an element in $G$ and modify it by row and column operations coming from $K$ until eventually it is transformed into a diagonal matrix.]

For $a, a' \in A$, when are the double cosets $KaK$ and $Ka'K$ the same? The refined version answers this question. First of all it is evident that $KaK = Ka'K$ if $a$, $a'$ have the same image in $A/A \cap K$. Second of all, since we can find representatives in $K$ for all elements of the Weyl group $W$, it is also clear that $KaK = Kw(a)K$ for all $w \in W$ (where $w(a)$ denotes the conjugation action of $W$ on $A$). In fact these two observations turn out to be the end of the story: $KaK = Ka'K$ if and only if the images of $a$, $a'$ in $A/A \cap K$ are conjugate under the Weyl group. In view of the discussion in 4.2 we obtain a natural bijection from the set of $K\backslash G/K$ of $K$-double cosets to the set of orbits of $W$ in $X_*(A)$. The dominant coweights provide a natural set of orbit representatives for this action, so that the set of $K$-double cosets can also be parametrized by dominant coweights. (Recall that a coweight $\mu$ is said to be *dominant* if $\langle \alpha, \mu \rangle \geq 0$ for every positive root $\alpha$ of $A$.)

For $GL_n$ the dominant coweights are the $n$-tuples $(j_1, \dots, j_n) \in \mathbb{Z}^n$ such that $j_1 \geq \cdots \geq j_n$, and we conclude that such coweights parametrize the $K$-double cosets. The following method can be used to prove the refined Cartan decomposition for $GL_n$. The idea is to construct sufficiently many invariants of $K$-double cosets. The first idea is to consider the valuation of the determinant; this invariant of a matrix clearly only depends on its $K$-double coset. A more subtle invariant is to consider the least valuation of all the matrix entries (for this purpose we consider that 0 has valuation $+\infty$). The two procedures can be combined by considering any integer $i$ such that $1 \leq i \leq n$ and considering the least valuation of all the $i \times i$ minors in our matrix. Applying this last invariant to any element in the $K$-double coset containing the diagonal matrix with diagonal entries $(\pi^{j_1}, \dots, \pi^{j_n})$, we obtain the sum of the last $i$ entries in the $n$-tuple $(j_1, \dots, j_n)$. These sums (for all $i$) together determine the dominant coweight uniquely. This proves the refined Cartan decomposition for $GL_n$.

**4.4. Iwasawa decomposition.** The Iwasawa decomposition states that $G = BK$. This reflects the fact that the flag variety $B\backslash G$ is projective over $F$, hence satisfies the valuative criterion for properness. For $GL_n$ it is another instructive exercise to prove the Iwasawa decomposition directly. [Start with an element in $G$ and modify it by column operations coming from $K$ until eventually it is transformed into an upper-triangular matrix.]

**4.5. Definitions of $\Lambda_G$, $H_G : G \to \Lambda_G$, $\mathfrak{a}$, $\mathfrak{a}_G$.** We now introduce various objects that will be used throughout this article. For example, there is an obvious surjective homomorphism $GL_n(F) \to \mathbb{Z}$ defined by $g \mapsto \mathrm{val}(\det g)$, which we need to generalize to all split groups.

Let $\Lambda_G$ denote the quotient of $X_*(A)$ by the coroot lattice for $G$ (by which we mean the subgroup of $X_*(A)$ generated by all the coroots of $A$). There is a surjective homomorphism $H_G : G \to \Lambda_G$, which by virtue of the Cartan decomposition is characterized by the following two properties. The first property is that the restriction of $H_G$ to the subgroup $A$ is equal to the composition

$$A \twoheadrightarrow A/(A \cap K) \cong X_*(A) \twoheadrightarrow \Lambda_G. \tag{4.5.1}$$

The second property is that the restriction of $H_G$ to $K$ is trivial. Moreover it is true that $H_G$ is also trivial on the image in $G$ of the $F$-points of $G_{\mathrm{sc}}$. From this it follows that if $g = ank$ with $a \in A$, $n \in N$, $k \in K$ (Iwasawa decomposition), then $H_G(g) = H_G(a)$.

We write $\mathfrak{a}$ for $X_*(A) \otimes_{\mathbb{Z}} \mathbb{R}$ and $\mathfrak{a}_G$ for $X_*(A_G) \otimes_{\mathbb{Z}} \mathbb{R}$. Thus $\mathfrak{a}_G$ can be viewed as a subspace of $\mathfrak{a}$. Moreover the composition $X_*(A_G) \hookrightarrow X_*(A) \twoheadrightarrow \Lambda_G$ induces an isomorphism

$$\mathfrak{a}_G \cong \Lambda_G \otimes_{\mathbb{Z}} \mathbb{R}. \tag{4.5.2}$$

When $G_{\mathrm{der}}$ is simply connected as algebraic group, the finitely generated abelian group $\Lambda_G$ is torsion-free and is therefore a free abelian group. In this case the natural map $\Lambda_G \to \mathfrak{a}_G$ is injective and identifies $\Lambda_G$ with a lattice in the real vector space $\mathfrak{a}_G$, so that there is no harm in thinking about $H_G$ as being a homomorphism $G \to \mathfrak{a}$ that takes values in the lattice $\Lambda_G$ in $\mathfrak{a}$. When the derived group is not simply connected, $\Lambda_G$ has torsion which is lost when one passes to $\Lambda_G \otimes_{\mathbb{Z}} \mathbb{R} = \mathfrak{a}_G$. In order to avoid confusion the reader should be aware that in Arthur's papers $H_G$ denotes the composition of our $H_G$ with the natural map $\Lambda_G \to \mathfrak{a}_G$.

## 5. Orbital integrals on $GL_2(F)$

**5.1. The goal.** Our goal in this section is to get a better understanding of orbital integrals by calculating lots of them for the group $GL_2(F)$. As we will see, the phenomenon of homogeneity (covered in DeBacker's course in this summer school), shows up very clearly in these calculations. We follow the exposition in [**Lan80**], using the tree for $SL_2(F)$ as our main computational tool.

As before we consider a $p$-adic field $F$ with valuation ring $\mathcal{O}$ and uniformizing element $\pi$. We write $q$ for the cardinality of the residue field $\mathcal{O}/\pi\mathcal{O}$ of $\mathcal{O}$.

We write $G$ for the group $GL_2(F)$ and write $K$ for its compact open subgroup $GL_2(\mathcal{O})$. We will not consider orbital integrals for arbitrary functions in $C_c^\infty(G)$. We will only consider functions lying in the spherical Hecke algebra $\mathcal{H}$, defined as the subspace of $C_c^\infty(G)$ consisting of functions that are both left and right invariant under $K$. The multiplication on $\mathcal{H}$ is given by convolution and turns out to be commutative, and there is a simple description of this commutative $\mathbb{C}$-algebra (using the Satake isomorphism). Important as these facts are, they play no role here. Our limited goal is to understand the linear functionals on $\mathcal{H}$ obtained by restriction from the linear functionals $O_\gamma$ on $C_c^\infty(G)$, and even this will be done only for elements $\gamma \in K$ (which covers orbital integrals for all elements in $G$ whose conjugacy class meets $K$). So, throughout this section $\gamma$ will always denote an element of $K$.

The characteristic functions $1_{KaK}$ of the double cosets $KaK$ of $K$ in $G$ form a basis for the vector space $\mathcal{H}$. By the Cartan decomposition 4.3 there is bijection from $K\backslash G/K$ to $\{(m,n) \in \mathbb{Z}^2 : m \geq n\}$, which associates to the pair $(m,n)$ the double coset containing the diagonal matrix with diagonal entries $(\pi^m, \pi^n)$.

For $(m,n) \in \mathbb{Z}^2$ with $m \geq n$, we write $f_{m,n}$ for the characteristic function of the $K$-double coset corresponding to $(m,n)$. The functions $f_{m,n}$ form a basis for $\mathcal{H}$, so it is enough to calculate the numbers $O_\gamma(f_{m,n})$. Since $\gamma \in K$, the determinant of $\gamma$ has valuation 0, which means that $O_\gamma(f_{m,n})$ vanishes unless $m+n=0$. Therefore, it is enough to consider the functions $f_m$ defined by $f_m := f_{m,-m}$ (for $m \geq 0$). We will now compute, case-by-case, the orbital integrals $O_\gamma(f_m)$ for all conjugacy classes meeting $K$.

To define the orbital integral for $\gamma$ we need an invariant measure $dg/dg'_\gamma$ on $G'_\gamma\backslash G$ (see 3.4). We will always use the Haar measure $dg$ on $G$ that gives $K$ measure 1. We will discuss $dg'_\gamma$ case-by-case below.

**5.2. Some useful subgroups of $G$.** Let $Z$ denote the group of non-zero scalar multiples of the identity matrix; in other words, $Z$ is the center of $G$. Let $B = AN$ be as in 4.1.

**5.3. Tree for $SL_2(F)$.** As mentioned before, our main computational tool will be the tree for $SL_2(F)$, which we now need to discuss. A good reference is [**Ser03**]. A tree is the geometric realization of a 1-dimensional simplicial complex that is both connected and simply connected. It can be specified by giving its set $V$ of vertices and saying which pairs of vertices are joined by an edge. The tree of interest here comes equipped with an action of $G$, and the action is transitive on the set of vertices. In fact the set $V$ of vertices has a base-point $v_0$ whose stabilizer is $KZ$, so that $V$ becomes identified with the homogeneous space $G/KZ$.

Inside the set $V$ we have the orbit of $v_0$ under $A$, which can be identified with $A/(A \cap KZ) = A/(A \cap K)Z$. In fact the group $A/(A \cap K)Z$ is isomorphic to $\mathbb{Z}$, via the isomorphism sending a diagonal matrix to the difference of the valuations of its two diagonal entries. Under this isomorphism an integer $j$ then corresponds to the diagonal matrix

$$\begin{bmatrix} \pi^j & 0 \\ 0 & 1 \end{bmatrix} \in A/(A \cap K)Z, \tag{5.3.1}$$

and we write $v_j$ for the corresponding vertex in $V$. We connect any two successive vertices $v_j$, $v_{j+1}$ by an edge, obtaining a 1-dimensional simplicial complex whose geometric realization is a real line, with vertices placed at each integer point; this copy of the real line is called the *standard apartment* in our tree.

So far we have just described some of the edges in our tree, namely the ones joining vertices in $A/(A \cap K)Z$. We get all the edges in the tree by using the action of $G$ to move around the edges we have already described (ensuring that the $G$-action does preserve the set of edges, as desired). It then turns out that the 1-dimensional simplicial complex we have constructed really is a tree, each vertex of which has $q + 1$ neighbors (where $q$ is the cardinality of the residue field).

Indeed, because of the $G$-action it is enough to show that the base-point $v_0$ has $q + 1$ neighbors. It turns out that $K$ (which fixes $v_0$) acts transitively on the set of neighbors of $v_0$. One of these neighbors is $v_{-1}$, and a simple calculation shows that the stabilizer of $v_{-1}$ in $K$ is

$$\begin{bmatrix} \mathcal{O}^\times & \mathcal{O} \\ \pi O & \mathcal{O}^\times \end{bmatrix}. \tag{5.3.2}$$

Thus the $K$-orbit of $v_{-1}$ is $G(\mathcal{O}/\pi\mathcal{O})/B(\mathcal{O}/\pi\mathcal{O})$, the set of points on the projective line over the residue field, which explains why there are $q + 1$ neighbors of $v_0$.

**5.4. Metric on the tree and its relation to** $\mathrm{inv}(x_1, x_2)$**.** There is an obvious metric $d(y_1, y_2)$ on the tree. It is $G$-invariant, and on the standard apartment discussed above it agrees with the usual metric on the real line. For this metric two neighboring vertices have distance 1, and $d(v_m, v_n) = |m - n|$.

As before we write $X$ for $G/K$ and $x_0$ for the base-point in $X$. Let $x_1, x_2 \in X$. Since the set $V$ of vertices of the tree is equal to $G/KZ = Z\backslash X$, the images of $x_1$, $x_2$ under the canonical surjection $X \to Z\backslash X = V$ are vertices $v_1$, $v_2$ in the tree.

As in 3.4 the relative position of $x_1$, $x_2$ is measured by $\mathrm{inv}(x_1, x_2) \in K\backslash G/K$. By the refined Cartan decomposition for $GL_2$ we can view $\mathrm{inv}(x_1, x_2)$ as a pair $(m, n)$ of integers such that $m \geq n$.

It is easy to see that $m + n$ coincides with the valuation of the determinant of any group element $g$ such that $x_1 = gx_2$, and that $m - n$ coincides with the distance between the vertices $v_1$ and $v_2$ in the tree.

**5.5. Geodesics in the tree, convexity of fixed-point sets.** Any two points $y_1$, $y_2$ in our tree are joined by a unique shortest path $[y_1, y_2]$, called a geodesic. Inside the standard apartment (identified with $\mathbb{R}$) geodesics are closed intervals. We say that a subset $C$ of the tree is *convex* if for every $c_1, c_2 \in C$ the geodesic $[c_1, c_2]$ is contained in $C$.

Any element $\gamma \in G$ takes the geodesic joining $y_1$ and $y_2$ to the geodesic joining $\gamma y_1$ to $\gamma y_2$. Therefore if $y_1$, $y_2$ are fixed by $\gamma$, so is every point of the geodesic joining them. Thus the fixed point set of $\gamma$ in the tree is convex.

For elements $\gamma \in K$ we can say more. First of all such an element obviously fixes the base-vertex $v_0$, so its fixed-point set on the tree is certainly non-empty. Moreover, since the determinant of $\gamma$ is a unit, the distance $d(\gamma v, v)$ is an even integer for any vertex $v \in V$, and it follows that $\gamma$ cannot take a vertex to one of its neighbors, and in particular it cannot interchange the two vertices of an edge. Therefore, if $\gamma$ fixes an interior point of an edge, it actually fixes the entire edge pointwise, and we see that the fixed point set of $\gamma$ in the tree is simply the union of all the edges both of whose vertices are fixed by $\gamma$.

For $\gamma \in K$ these considerations lead to the following simple method for determining $d(\gamma v, v)$ for any vertex $v$. There is a unique geodesic joining $v$ to a vertex $v'$ fixed by $\gamma$ and having the additional property that no vertex along this geodesic other than $v'$ is fixed by $\gamma$. Equivalently, this geodesic is the shortest possible one joining $v$ to some vertex in the fixed-point set of $\gamma$. Applying $\gamma$ to this geodesic, we see that every point moves but $v'$, so that the geodesic together with its transform form a geodesic from $v$ to $\gamma v$. This shows that $d(\gamma v, v) = 2d(v, v')$; *in other words,* $d(\gamma v, v)$ *is twice the distance from* $v$ *to the fixed point set of* $\gamma$. We will use this observation repeatedly below.

**5.6. Unipotent orbital integrals restricted to $\mathcal{H}$ (denoted $L_0$, $L_1$).** In $G$ there are two unipotent orbits, one of them being $\{1\}$, the other being the orbit of the matrix $u$ defined by

$$u = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}. \tag{5.6.1}$$

In this subsection we compute the restrictions to $\mathcal{H}$ of the orbital integrals $O_\gamma$ for $\gamma = 1$ and $\gamma = u$.

For $\gamma = 1$ the problem is trivial, since the orbital integral is just evaluation at the identity. Thus the restriction $L_0$ of $O_1$ to $\mathcal{H}$ takes the value 1 on $f_0$ and vanishes on $f_m$ for $m > 0$.

Next we calculate the restriction $L_1$ of $O_u$ to $\mathcal{H}$. The centralizer $G_u$ is easily seen to be $ZN$. We identify $Z$ with $F^\times$ in the obvious way, sending a scalar $z$ to the corresponding scalar matrix. We use the Haar measure on $Z$ giving $\mathcal{O}^\times$ measure 1. We identify $N$ with the additive group $F$ in the obvious way, using the upper right matrix entry as our coordinate. We use the Haar measure on $F$ that gives $\mathcal{O}$ measure 1. We use the product measure on $ZN$, and hence $K \cap NZ$ has measure 1.

From 3.4 we see that

$$L_1(f_m) = \sum_v \operatorname{meas}(\operatorname{Stab}_N(v))^{-1}, \tag{5.6.2}$$

where the sum runs over a set of representatives for the orbits of $N$ on the set of vertices $v \in V$ such that $d(uv, v) = 2m$. Here we used that $NZ\backslash X = N\backslash V$ (clear) and that $\mathrm{Stab}_{NZ}(x) = \mathrm{Stab}_N(v) \cdot (Z \cap K)$ for $x \in X$ mapping to $v \in V$.

Using the Iwasawa decomposition $G = BK = NAK$, we see that the set $N\backslash V$ of orbits of $N$ on $V$ is $A/Z(A \cap K) \simeq \mathbb{Z}$. The vertices $v_j$ ($j \in \mathbb{Z}$) in the standard apartment are a particularly convenient set of orbit representatives. We need to compute $\mathrm{Stab}_N(v)$ for each orbit representative.

This is a simple computation, the result of which is that $\mathrm{Stab}_N(v_j) = \pi^j\mathcal{O}$ (identifying $N$ with $F$ as above), a group whose measure is $q^{-j}$. Another consequence of this computation is that $v_j$ is fixed by $u$ if and only if $j \leq 0$, and, as we saw in 5.5, this allows us to calculate $d(uv_j, v_j)$ for each $j \in \mathbb{Z}$. Indeed, for $j \leq 0$ the vertex $v_j$ is fixed, so that $d(uv_j, v_j) = 0$. For $j > 0$, the geodesic $[v_j, v_0]$ has $v_0$ as its unique fixed point, so $d(uv_j, v_j) = 2j$.

Putting all these observations together, we see that

$$L_1(f_0) = 1 + q^{-1} + q^{-2} + \cdots = 1/(1 - q^{-1}) \tag{5.6.3}$$

and that for all $m > 0$

$$L_1(f_m) = q^m. \tag{5.6.4}$$

**5.7. $O_\gamma$ for any $\gamma$ that is not regular semisimple.** For $f \in \mathcal{H}$ it is evident that $O_\gamma(f)$ does not change when $\gamma$ is multiplied by $z \in Z \cap K$. Any $\gamma \in K$ which is not regular semisimple is conjugate to an element of the form $z$ or $zu$ (for some $z \in Z \cap K$), and therefore $O_\gamma$ restricted to $\mathcal{H}$ is either $L_0$ or $L_1$, as the case may be. It now remains only to consider regular semisimple elements $\gamma \in K$, in other words those whose eigenvalues are distinct.

**5.8. Hyperbolic orbital integrals.** Next we consider regular semisimple $\gamma$ whose eigenvalues lie in $F$. The conjugacy class of such an element meets $K$ if and only if the two eigenvalues are units, and after replacing $\gamma$ by a conjugate we may assume that

$$\gamma = \begin{bmatrix} a & o \\ 0 & b \end{bmatrix}, \tag{5.8.1}$$

with $a, b$ distinct elements in $\mathcal{O}^\times$. The centralizer $G_\gamma$ is $A$, and the most convenient choice for $G'_\gamma$ is the product of $Z$ and the infinite cyclic subgroup of $A$ generated by

$$\begin{bmatrix} \pi & 0 \\ 0 & 1 \end{bmatrix}. \tag{5.8.2}$$

We use the Haar measure on $Z$ that gives $Z \cap K$ measure 1, and we use the counting measure on the infinite cyclic subgroup $\mathbb{Z}$.

Using 3.4 and 5.5, we see that

$$O_\gamma(f_m) = \sum_v 1, \tag{5.8.3}$$

where the sum runs over a set of representatives for the orbits of $\mathbb{Z}$ on the set of vertices $v \in V$ such that $d(\gamma v, v) = 2m$ (equivalently, such that the distance from $v$ to the fixed point set of $\gamma$ in the tree is equal to $m$). Thus $O_\gamma(f_m)$ is the number of orbits of the infinite cyclic subgroup $\mathbb{Z}$ of $A$ on the set of vertices $v \in V$ at distance $m$ from the fixed point set of $\gamma$ in the tree.

As observed before, since $\gamma \in K$, its fixed point set in the tree is just the union of the edges joining two fixed vertices. Therefore it remains only to understand the set $V^\gamma$ of vertices fixed by $\gamma$. Put $d_\gamma := \mathrm{val}(1 - \frac{a}{b})$, a non-negative integer. We claim that $V^\gamma$ is the set of vertices $v \in V$ whose distance to the standard apartment is less than or equal to $d_\gamma$. By the Iwasawa decomposition we may write $v = anv_0$ with $a \in A$ and $n \in N$. Since the two sets we are trying to prove are equal are both stable under $A$, it is harmless to suppose that $a = 1$. Thus we need only consider $v$ of the form $nv_0$.

Let us determine when $\gamma$ fixes $nv_0$. Since $\gamma$ fixes $v_0$, the condition that $\gamma$ fix $nv_0$ is equivalent to the condition that $\gamma n^{-1}\gamma^{-1}n$ fix $v_0$. But $\gamma n^{-1}\gamma^{-1}n$ lies in $N$ and is easily computed in terms of $n$ and $\gamma$. Indeed, identifying $N$ with $F$ as before, so that $n$ becomes an element $y \in F$, we find that $\gamma n^{-1}\gamma^{-1}n$ becomes the element $(1 - \frac{a}{b})y$ of $F$. Since the stabilizer of $v_0$ is $KZ$, it is now clear that $\gamma$ fixes $nv_0$ if and only if $y \in \pi^{-d_\gamma}\mathcal{O}$.

To finish proving the claim we now need to compute the distance from $nv_0$ to the standard apartment in terms of the valuation of $y$. If $y \in \mathcal{O}$, then $nv_0$ equals $v_0$ and hence has distance 0 to the standard apartment. On the other hand, suppose that the valuation of $y$ is negative, say equal to $-r$ for some positive integer $r$. We saw above that (see 5.6) $\mathrm{Stab}_N(v_j) = \pi^j\mathcal{O}$, from which it follows that the vertex $v_j$ in the standard apartment is fixed by $n$ if and only if $j \le -r$. Therefore the geodesic $n[v_{-r}v_0]$ meets the standard apartment only at its endpoint $v_{-r}$, showing that its other endpoint, namely the point $nv_0$, has distance $r$ to the standard apartment. This completes the proof of the claim.

Having proved the claim, now we can finish the computation of our orbital integral. Working modulo the action of the infinite cyclic subgroup $\mathbb{Z}$, we need to count vertices whose distance to the fixed point set is $m$. The fixed point set consists of all points whose distance to the standard apartment is less than or equal to $d_\gamma$; when $m = 0$ we are simply counting these points. When $m > 0$, a vertex has distance $m$ from the fixed point set if and only if it has distance $m + d_\gamma$ from the standard apartment.

Therefore for any non-negative integer $s$ we need to calculate the number $N(s)$ of orbits of $\mathbb{Z}$ on the set of vertices at distance $s$ from the standard apartment. Clearly $N(s)$ is also equal to the number of vertices $v$ at distance $s$ to the standard apartment and having the additional property that the point in the standard apartment that is closest to $v$ is equal to $v_0$. Elementary reasoning, using that every vertex has $q + 1$ neighbors, shows that $N(0) = 1$ and $N(s) = q^s - q^{s-1}$ for $s > 0$.

Putting everything together, we now see that

$$O_\gamma(f_0) = 1 + (q-1) + (q^2 - q) + \cdots + (q^{d_\gamma} - q^{d_\gamma - 1}) = q^{d_\gamma} \tag{5.8.4}$$

and that for all $m > 0$

$$O_\gamma(f_m) = q^{m+d_\gamma} - q^{m+d_\gamma-1}, \tag{5.8.5}$$

and comparing this with the computation of $L_1$ that we made earlier, we obtain

LEMMA 5.1. *The restriction to $\mathcal{H}$ of the hyperbolic orbital integral $O_\gamma$ is equal to $(1 - q^{-1})q^{d_\gamma} \cdot L_1$.*

**5.9. Elliptic orbital integrals.** The only remaining orbits (among those meeting $K$) are elliptic. The eigenvalues of an elliptic (regular semisimple) element generate a quadratic extension $E$ of $F$. How do such elements sit inside $G$?

Start with a quadratic extension $E/F$. We can view $E$ as a 1-dimensional $E$-vector space and as a 2-dimensional $F$-vector space, and since an $E$-linear map is necessarily $F$-linear, we have $GL_E(E) \subset GL_F(E)$. Choosing an $F$-basis in $E$, this becomes an embedding of $E^\times$ in $G$. The image is the set of $F$-points of a maximal torus $T$ in $GL_2$.

Using this embedding, we view $\gamma \in E^\times$ as an element of $G$. Then its two eigenvalues are $\gamma$, $\bar{\gamma}$ (using bar to denote the non-trivial element in the Galois group of the quadratic extension), and its determinant is the norm $\gamma\bar{\gamma}$ of $\gamma$. If $\gamma$ is conjugate to an element of $K$, then its determinant is a unit, and hence $\gamma$ is a unit in the valuation ring $\mathcal{O}_E$ of $E$. In order to embed $E^\times$ in $G$ we have to choose an $F$-basis in $E$. Let us agree to pick one which is at the same time an $\mathcal{O}$-basis for $\mathcal{O}_E$. Then we will have $E^\times \cap K = \mathcal{O}_E$. In order that $\gamma$ be regular, we need $\gamma \neq \bar{\gamma}$. Thus the elements of interest are those in $(\mathcal{O}_E)^\times$ but not in $\mathcal{O}^\times$.

For such an element $\gamma$ the centralizer is $E^\times$, and we are free to take the group $G'_\gamma$ of 3.4 to be $Z$. It follows from 3.4 and 5.4 that

$$O_\gamma(f_m) = |\{v \in V : d(\gamma v, v) = 2m\}|. \tag{5.9.1}$$

We are going to calculate these orbital integrals in two steps. First we will compute them in terms of the cardinality of the set $V^\gamma$ of vertices fixed by $\gamma$, then we will calculate $|V^\gamma|$.

LEMMA 5.2. *For all elliptic regular semisimple $\gamma \in K$ the restriction of $O_\gamma$ to $\mathcal{H}$ is equal to*

$$\big(2q^{-1} + (1 - q^{-1})|V^\gamma|\big) \cdot L_1 - \frac{2}{q-1} \cdot L_0. \tag{5.9.2}$$

PROOF. We need to compute $O_\gamma(f_m)$ for all $m \geq 0$. Of course

$$O_\gamma(f_0) = |V^\gamma|. \tag{5.9.3}$$

Now assume that $m > 0$. As we have seen in 5.5, the condition $d(\gamma v, v) = 2m$ is equivalent to the condition that the distance $d(v, V^\gamma)$ from $v$ to the fixed point set $V^\gamma$ be $m$. Consider the unique shortest geodesic joining $v$ to the fixed point set, and let $w$ be the unique endpoint of that geodesic lying in $V^\gamma$. Then $v \mapsto w$ is a well-defined retraction of $V$ onto $V^\gamma$, and thus

$$O_\gamma(f_m) = \sum_{w \in V^\gamma} |\{v \in V : d(v, V^\gamma) = m \text{ and } v \mapsto w\}|. \tag{5.9.4}$$

Given $w \in V^\gamma$, an element $v \in V$ satisfies the two conditions $d(v, V^\gamma) = m$ and $v \mapsto w$ if and only if $d(v, w) = m$ and the unique neighboring vertex of $w$ lying on the geodesic $[w, v]$ is not fixed by $\gamma$; the number of such neighbors is $(q + 1) - C_w$, where $C_w$ is the number of neighbors of $w$ fixed by $\gamma$. Therefore

$$O_\gamma(f_m) = \sum_{w \in V^\gamma} q^{m-1}\big((q + 1) - C_w\big). \tag{5.9.5}$$

Summing $C_w$ over all $w \in V^\gamma$, we get $2|E^\gamma|$, where $E^\gamma$ denotes the set of edges in the tree that are fixed by $\gamma$. Now the fixed point set of $\gamma$ in the tree is convex, hence contractible, and therefore its Euler characteristic $|V^\gamma| - |E^\gamma|$ is 1, which means that the sum over $w$ of $C_w$ is $2(|V^\gamma| - 1)$. Thus we have proved that

$$O_\gamma(f_m) = q^{m-1}\big((q - 1)|V^\gamma| + 2\big). \tag{5.9.6}$$

The lemma follows from (5.9.3), (5.9.6) and the formulas for $L_0$ and $L_1$ that we found before. $\square$

Our next task is to calculate $|V^\gamma|$. The answer depends on whether or not the quadratic extension $E/F$ is ramified.

First we consider the case in which $E/F$ is unramified. In particular the cardinality of the residue field of $E$ is $q^2$. The tree for $SL_2(F)$ is a subtree of the tree for $SL_2(E)$, and in this bigger tree every vertex has $q^2+1$ neighbors. Moreover $\mathrm{Gal}(E/F)$ operates on the bigger tree and its set of fixed points is the smaller one.

Inside $GL_2(E)$ our matrix $\gamma$ is conjugate to the diagonal matrix $\gamma'$ with diagonal entries $(\gamma, \bar{\gamma})$, so that we are dealing with a hyperbolic element whose fixed point set we already understand. As in the hyperbolic case we define a non-negative integer

$$d_\gamma := \mathrm{val}(1 - \gamma\bar{\gamma}^{-1}).$$

Choose an element of $GL_2(E)$ that conjugates $\gamma'$ to $\gamma$ and apply it to the standard apartment, obtaining a non-standard apartment in the bigger tree. From previous work we know that the fixed point set of $\gamma$ is the set of vertices in the bigger tree whose distance to the non-standard apartment is less than or equal to $d_\gamma$. The non-trivial element of $\mathrm{Gal}(E/F)$ preserves this non-standard apartment, flipping it end-to-end, and fixes a unique vertex $v'$ in it. The fixed point set of $\gamma$ in the smaller tree is precisely the set of vertices $v \in V$ whose distance to $v'$ is less than or equal to $d_\gamma$, from which one sees easily that

$$|V^\gamma| = 1 + (q+1)(1 + \cdots + q^{d_\gamma - 1}) = 1 + (q+1)\frac{q^{d_\gamma} - 1}{q-1}. \tag{5.9.7}$$

Combining this with Lemma 5.2, we obtain our final formula

$$(1 + q^{-1})q^{d_\gamma} \cdot L_1 - \frac{2}{q-1} \cdot L_0 \tag{5.9.8}$$

for the restriction of $O_\gamma$ to $\mathcal{H}$ in the unramified case.

Next we consider the case in which $E/F$ is ramified. The tree for $SL_2(F)$ still sits inside the tree for $SL_2(E)$, but in a more complicated way. Since the uniformizing element $\pi$ for $F$ has valuation 2 in $E$, the midpoint of each edge in the smaller tree becomes a vertex in the bigger one. Since the residue field does not change, every vertex in the bigger tree still has $q+1$ neighbors. The Galois group of $E/F$ still acts on the bigger tree, and the smaller tree is fixed pointwise by this action, but it does not fill out the whole fixed point set unless $E/F$ is tamely ramified.

Our element $\gamma \in (\mathcal{O}_E)^\times \setminus \mathcal{O}^\times$ still can be diagonalized in $GL_2(E)$, so that we again get a non-standard apartment in the bigger tree. There is a unique edge $e$ in the smaller tree closest to this non-standard apartment, and the shortest path between the edge $e$ and the apartment starts from the midpoint of that edge.

From previous work we know that the fixed point set of $\gamma$ in the bigger tree consists of all points whose distance $d$ to this non-standard apartment is less than or equal to a certain integer (that depends on $\gamma$). For a vertex $v$ in the smaller tree this distance $d$ is twice (because of the subdivision that occurred) the distance from $v$ to the edge $e$ plus a constant (the constant being 1 plus the distance from $e$ to our non-standard apartment). Therefore there exists a non-negative integer $d_\gamma$ such that the fixed point set of $\gamma$ in $V$ consists of all vertices whose distance from $e$

(measured in the smaller tree) is less than or equal to $d_\gamma$. In fact one can show that

$$1 + 2d_\gamma = \sup\{\mathrm{val}_E(\gamma - a) : a \in \mathcal{O}^\times\}, \tag{5.9.9}$$

where the valuation $\mathrm{val}_E$ being used here assigns 1 to uniformizing elements in $E$.

It is then easy to see that

$$|V^\gamma| = 2(1 + \cdots + q^{d_\gamma}) = 2\frac{q^{d_\gamma+1} - 1}{q - 1}. \tag{5.9.10}$$

Combining this with Lemma 5.2, we obtain our final formula

$$2q^{d_\gamma} \cdot L_1 - \frac{2}{q-1} \cdot L_0 \tag{5.9.11}$$

for the restriction of $O_\gamma$ to $\mathcal{H}$ in the ramified case.

**5.10. Homogeneity!** Our computations have revealed something quite remarkable. The restrictions of the orbital integrals for the unipotent elements 1 and $u$ give us two linear functionals $L_0$ and $L_1$ on $\mathcal{H}$. For every other element $\gamma \in K$ the restriction of the orbital integral $O_\gamma$ to $\mathcal{H}$ is a linear combination of $L_0$ and $L_1$. It was by no means obvious *a priori* that this should be the case. In fact this is an example of the deep phenomenon of homogeneity that is the subject of DeBacker's course.

The coefficients of $L_0$, $L_1$ in these linear combinations are very interesting functions of $\gamma$, called *Shalika germs*, and we will discuss them next.

## 6. Shalika germs

**6.1. The goal.** Our goal in this section is to prove the existence of the Shalika germ expansion (see [**Sha72**]). Once the general theory is in place, we will illustrate how it works by re-examining our computations of orbital integrals on $GL_2$. We now work with any connected reductive group over a $p$-adic field $F$, and we once again lighten the notation by writing $G$ for the $F$-points of our group.

**6.2. Notation.** Let $\mathcal{U}$ be the set of unipotent elements in $G$. Then $\mathcal{U}$ is closed, and there are finitely many $G$-orbits $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_r$ in $\mathcal{U}$. We write $\mu_1, \dots, \mu_r$ for the corresponding unipotent orbital integrals (for some choice of invariant measures on the orbits that we prefer not to encode in the notation). Let $T$ be (the set of $F$-points of) a maximal torus in $G$, and let $T_{\mathrm{reg}}$ be the subset of regular elements. We are interested in orbital integrals $O_\gamma$ for variable $\gamma \in T_{\mathrm{reg}}$, so we need a coherent set of choices of invariant measures on the orbits of all such elements $\gamma$. This can be done as follows. Once and for all we choose Haar measures $dg$ and $dt$ on $G$ and $T$ respectively. Then for any $\gamma \in T_{\mathrm{reg}}$ we put (for $f \in C_c^\infty(G)$)

$$O_\gamma(f) = \int_{T\backslash G} f(g^{-1}\gamma g)\, dg/dt. \tag{6.2.1}$$

**6.3. Distributions.** For any l.c.t.d space we have already introduced the space $C_c^\infty(X)$ of locally constant compactly supported functions on $X$. A *distribution* is by definition any linear functional on $C_c^\infty(X)$ (with no continuity hypothesis since there is no topology on our function space). We write $\mathcal{D}(X)$ for the vector space of all distributions on $X$.

Let $Y$ be any closed subset of $X$, and let $U$ be the complementary open subset. Dual to the short exact sequence (2.1.1) is the short exact sequence

$$0 \to \mathcal{D}(Y) \to \mathcal{D}(X) \to \mathcal{D}(U) \to 0. \tag{6.3.1}$$

In other words, given a distribution on $X$, we can restrict it to $U$, and among the distributions on $X$, we have those whose support (see 26.2 for a discussion of the notion of support of a distribution) is contained in $Y$. Now suppose that some group $H$ acts on $X$, preserving $Y$ and $U$. Then, taking invariants under the group action (denoted by a superscript $H$), we get an exact sequence

$$0 \to \mathcal{D}(Y)^H \to \mathcal{D}(X)^H \to \mathcal{D}(U)^H \tag{6.3.2}$$

but there is no guarantee that the restriction map at the right end is surjective.

From the short exact sequence (2.1.1) we also get an exact sequence

$$C_c^\infty(U)_H \to C_c^\infty(X)_H \to C_c^\infty(Y)_H \to 0, \tag{6.3.3}$$

where the subscript $H$ denotes coinvariants for $H$. (For an $H$-module $V$ the space of coinvariants $V_H$ is by definition the biggest quotient of $V$ on which $H$ acts trivially, or, in other words, the quotient of $V$ by the linear span of all vectors of the form $hv - v$ for some $h \in H$, $v \in V$.) The sequence (6.3.2) can also be obtained as the $\mathbb{C}$-dual of the sequence (6.3.3).

**6.4. Existence of the Shalika germ expansion.** Order the unipotent orbits so that their dimensions increase as $i$ does. Of course there can be several orbits of the same dimension; for these the order is immaterial. By dimension of the orbit we really mean the dimension (as algebraic variety) of the $G(\bar{F})$-orbit containing the given $G$-orbit. The purpose of this ordering is to guarantee that

$$\mathcal{O}_1 \cup \cdots \cup \mathcal{O}_i \tag{6.4.1}$$

is closed in $G$ for all $i$.

Inside the space $\mathcal{O}_1 \cup \mathcal{O}_2$ (which is closed in $G$, hence l.c.t.d) we have the closed subset $\mathcal{O}_1$ and complementary open subset $\mathcal{O}_2$. The group $G$ acts by conjugation on all these spaces. Therefore we get an exact sequence

$$0 \to \mathcal{D}(\mathcal{O}_1)^G \to \mathcal{D}(\mathcal{O}_1 \cup \mathcal{O}_2)^G \to \mathcal{D}(\mathcal{O}_2)^G. \tag{6.4.2}$$

The spaces $\mathcal{D}(\mathcal{O}_1)^G$ and $\mathcal{D}(\mathcal{O}_2)^G$ are 1-dimensional, spanned by the invariant integrals on the homogeneous spaces $\mathcal{O}_1$ and $\mathcal{O}_2$. Now we need to recall the non-trivial fact that $\mu_2$ is well-defined on $C_c^\infty(G)$ ("convergence of orbital integrals," discussed in 3.3). Therefore $\mu_2$ gives us an element of $\mathcal{D}(\mathcal{O}_1 \cup \mathcal{O}_2)^G$ that maps to the invariant integral on $\mathcal{O}_2$. We conclude that the map at the right end of the exact sequence above is surjective, and that $\mathcal{D}(\mathcal{O}_1 \cup \mathcal{O}_2)^G$ is 2-dimensional with basis given by (the restrictions of) $\mu_1$ and $\mu_2$. An obvious inductive argument then shows that

$$\mathcal{D}(\mathcal{O}_1 \cup \cdots \cup \mathcal{O}_r)^G$$

is $r$-dimensional with basis given by (the restrictions of) $\mu_1, \ldots, \mu_r$. Now we are ready for the theorem on germ expansions.

THEOREM 6.1 (Shalika [**Sha72**]). *There exist functions $\Gamma_1, \Gamma_2, \ldots, \Gamma_r$ on $T_{\mathrm{reg}}$ having the following property. For every $f \in C_c^\infty(G)$ there exists an open and closed*

*$G$-invariant neighborhood $U_f$ of $1$ in $G$ such that*

$$O_\gamma(f) = \sum_{i=1}^{r} \mu_i(f) \cdot \Gamma_i(\gamma) \tag{6.4.3}$$

*for all $\gamma \in U_f \cap T_{\mathrm{reg}}$. The germs about $1 \in T$ of the functions $\Gamma_1, \dots, \Gamma_r$ are unique. We refer to $\Gamma_i$ as the Shalika germ for the unipotent orbit $\mathcal{O}_i$ and the torus $T$.*

PROOF. Since the unipotent set $\mathcal{U} = \mathcal{O}_1 \cup \dots \cup \mathcal{O}_r$ is closed in $G$, there is a surjective restriction map $C_c^\infty(G) \to C_c^\infty(\mathcal{U})$, and we may choose functions $f_i$ ($i = 1, \dots, r$) such that $\mu_i(f_j) = \delta_{ij}$ (Kronecker $\delta$). Inserting the function $f_i$ into (6.4.3), we see that the germ of $\Gamma_i(\gamma)$ must be equal to the germ of $O_\gamma(f_i)$. This already proves the uniqueness assertion in the theorem. It also shows that we may as well take

$$\Gamma_i(\gamma) := O_\gamma(f_i) \tag{6.4.4}$$

as the definition of $\Gamma_i$. (There is no need to be troubled by the non-uniqueness of the functions $f_i$ since it is only germs that matter in the theorem.)

However we must still show that (6.4.3) is valid for all functions on $G$. So let $f \in C_c^\infty(G)$. The function

$$\sum_{i=1}^{r} \mu_i(f) \cdot f_i \tag{6.4.5}$$

obviously has the same unipotent orbital integrals as $f$ does. In other words all unipotent orbital integrals of

$$\phi := f - \sum_{i=1}^{r} \mu_i(f) \cdot f_i \tag{6.4.6}$$

vanish. Choose a neighborhood $U$ of $\mathcal{U}$ as in Lemma 6.2 below. We claim that (6.4.3) holds for the neighborhood $U_f = U$. Indeed, for $\gamma \in U \cap T_{\mathrm{reg}}$ the orbital integral $O_\gamma(\phi)$ vanishes by Lemma 6.2. In view of how $\phi$ was defined, this establishes (6.4.3). □

LEMMA 6.2. *Let $\phi \in C_c^\infty(G)$ and assume that all unipotent orbital integrals of $\phi$ vanish. Then there is an open and closed conjugation invariant neighborhood $U$ of the unipotent set $\mathcal{U}$ such that $I(\phi) = 0$ for every invariant distribution $I$ supported on $U$.*

PROOF. The dual space to $C_c^\infty(\mathcal{U})_G$ is $\mathcal{D}(\mathcal{U})^G$, which has as basis the unipotent orbital integrals $\mu_1, \dots, \mu_r$, so the vanishing of the unipotent orbital integrals of $\phi$ is equivalent to the vanishing of the image of $\phi$ in $C_c^\infty(\mathcal{U})_G$. In order to construct the desired neighborhood $U$ of $\mathcal{U}$, we use that there exists a continuous map $\alpha$ from $G$ to a l.c.t.d space $\mathbb{A}$ such that every fiber of $\alpha$ is a union of conjugacy classes in $G$ and such that there exists $x \in \mathbb{A}$ for which $\alpha^{-1}(x) = \mathcal{U}$. Therefore by the last statement of Lemma 27.1 there exists an open neighborhood $\omega$ of $x$ in $\mathbb{A}$ such that the image of $\phi$ in $C_c^\infty(\alpha^{-1}\omega)_G$ vanishes. Shrinking $\omega$, we may assume that it is compact as well as open, and then $U := \alpha^{-1}\omega$ is the desired neighborhood of $\mathcal{U}$.

It remains to prove the existence of $\alpha$. In fact the map denoted $\alpha$ in 3.2 does the job. However, we were a bit sketchy about its construction in the general case, and the reader may prefer to use the following cruder version of $\alpha$, which is however sufficient for our current needs. The cruder version is obtained by choosing an embedding (of algebraic groups) of $G$ into some general linear group, constructing $\alpha$ for the general linear group (using the coefficients of the characteristic polynomial, as in 3.2), and then restricting $\alpha$ to the subgroup $G$. This works since an element of $G$ is unipotent if and only if the corresponding matrix is unipotent. $\square$

**6.5. Back to $GL_2$.** Recall that in $G = GL_2$ there are two unipotent classes and hence two unipotent orbital integrals $\mu_1$, $\mu_2$. Our computations of orbital integrals for $GL_2$ revealed that for every regular semisimple $\gamma \in K = GL_2(\mathcal{O})$ there are complex numbers $A_1(\gamma)$, $A_2(\gamma)$ such that for every $f \in \mathcal{H}$ the equality

$$O_\gamma(f) = A_1(\gamma)\mu_1(f) + A_2(\gamma)\mu_2(f) \tag{6.5.1}$$

holds. Our computations also showed that the restrictions of $\mu_1$ and $\mu_2$ to $\mathcal{H}$ are linearly independent. Therefore the numbers $A_1(\gamma)$, $A_2(\gamma)$ are uniquely determined, and, moreover, inside the rather small function space $\mathcal{H}$ we can find functions $f_1$, $f_2$ satisfying $\mu_i(f_j) = \delta_{ij}$. We have seen that the Shalika germs $\Gamma_1$, $\Gamma_2$ are obtained by taking the orbital integrals of $f_1$, $f_2$, which in view of the equality above means that $\Gamma_i$ is the germ of $A_i$ $(i = 1, 2)$.

At first glance it might now seem that Shalika germ theory explains (6.5.1) and hence explains why the restrictions to $\mathcal{H}$ of the orbital integrals $O_\gamma$ for $\gamma \in K$ are all linear combinations of the restrictions of $\mu_1$ and $\mu_2$. This is far from being true, since the Shalika germ expansion is only valid on some (possibly very small) neighborhood of 1 and moreover this neighborhood depends on the function $f$ that we are considering. The amazing thing that has happened here is that there is a *big* neighborhood of 1 (namely $K$) which works for *all* the functions in $\mathcal{H}$.

It is tempting to refer to $A_1$, $A_2$ as "the" Shalika germs for $GL_2$, since among all possible functions having the correct germs, they are singled out naturally by the property (6.5.1).

## 7. Weyl integration formula

In this section we work at first with an arbitrary connected reductive group $G$ over our $p$-adic field $F$; starting with 7.8 we assume that $G$ is split. We write $\mathfrak{g}$ for the Lie algebra of $G$. Before we can get down to work on our next topic, the local trace formula, we need to derive the Weyl integration formula for $\mathfrak{g}$, which, roughly speaking, expresses the integral of a function on $\mathfrak{g}$ as an iterated integral, in which one first integrates over the various (adjoint) orbits in $\mathfrak{g}$ and then integrates over the set of orbits.

**7.1. Remarks on Weyl groups.** When working with maximal tori over non-algebraically closed fields such as $F$, there are three relevant Weyl groups. In order to explain them clearly we continue to make a notational distinction between the algebraic group $G$ and its group $G(F)$ of $F$-points.

Let $T$ be a maximal torus in $G$. The quotient $W := N_G(T)/T$ is a finite algebraic group defined over $F$. (We are writing $N_G(T)$ for the normalizer in $G$ of $T$.) We then have inclusions

$$W_T \subset W(F) \subset W(\bar{F}), \tag{7.1.1}$$

where $W_T$ is by definition the quotient $N_G(T)(F)/T(F)$. Of course $W(\bar{F})$ is the absolute Weyl group, and, up to inner automorphisms, is independent of $T$. The subgroup $W_T$ is the Weyl group needed in the Weyl integration formula.

The main thing we need to know about $W_T$ is that two regular elements $X, X' \in \mathfrak{t}$ are $G(F)$-conjugate if and only if they are conjugate under $W_T$. Indeed, if $g \in G(F)$ conjugates $X$ to $X'$, then it conjugates the centralizer of $X$ to the centralizer of $X'$; since the two elements are regular both centralizers are $T$, and therefore $g$ normalizes $T$, proving the forward implication. The reverse implication is trivial.

**7.2. Calculation of the differential of $\beta$.** Now we return to our usual practice of abbreviating $G(F)$ to $G$, $T(F)$ to $T$, etc. We write $\mathfrak{t}$ for the Lie algebra of $T$.

Consider the map

$$(T\backslash G) \times \mathfrak{t} \xrightarrow{\beta} \mathfrak{g} \tag{7.2.1}$$

defined by $\beta(g, X) = g^{-1}Xg$. For any $X \in \mathfrak{t}$ the differential $d\beta$ of $\beta$ at $(1, X) \in (T\backslash G) \times \mathfrak{t}$ is the map $(\mathfrak{g}/\mathfrak{t}) \times \mathfrak{t} \to \mathfrak{g}$ given by $(Y, Z) \mapsto [X, Y] + Z$. The two tangent spaces $(\mathfrak{g}/\mathfrak{t}) \times \mathfrak{t}$, $\mathfrak{g}$ both sit in short exact sequences with $\mathfrak{t}$ as the subspace and $\mathfrak{g}/\mathfrak{t}$ as the quotient space, and $d\beta$ is the identity on $\mathfrak{t}$. Therefore the top exterior powers of the two tangent spaces are canonically isomorphic, and the determinant of $d\beta$ at $(1, X)$ makes sense and is equal to

$$D(X) := \det\bigl(\mathrm{ad}(X); \mathfrak{g}/\mathfrak{t}\bigr). \tag{7.2.2}$$

The map $\beta$ is $G$-equivariant (for the translation action on $T\backslash G$, the trivial action on $\mathfrak{t}$, and the adjoint action on $\mathfrak{g}$). Choosing a $G$-invariant volume form (*i.e.* non-vanishing differential form of top degree) on $T\backslash G$ is the same as choosing a generator of the top exterior power of $\mathfrak{g}/\mathfrak{t}$. Choosing a translation invariant volume form on $\mathfrak{t}$ is the same as choosing a generator of the top exterior power of $\mathfrak{t}$. Make such choices. From them we get a generator of the top exterior power of $\mathfrak{g}$, which we use to get a translation invariant volume form on $\mathfrak{g}$. In this way we get $G$-invariant volume forms on the source and target of $\beta$, and we may use these volume forms to talk about the determinant of $d\beta$, or, in other words, the Jacobian of $\beta$. In fact the computation we made above, together with the $G$-equivariance of $\beta$, shows that the Jacobian of $\beta$ at any point $(g, X) \in (T\backslash G) \times \mathfrak{t}$ is equal to $D(X)$.

**7.3. Measures obtained from volume forms.** A volume form $\omega$ on a $p$-adic manifold $M$ gives rise to a measure $|\omega|$ on $M$, just as in the real case. In the end it boils down to assigning a measure $|dx|$ on $F$ to the differential form $dx$, where $x$ denotes the standard coordinate on $F$, that is, the identity map on $F$. In the real case the usual convention is of course that $|dx|$ is Lebesgue measure on $\mathbb{R}$. In the $p$-adic case one simply agrees $|dx|$ is some Haar measure on $F$, fixed once and for all. See [**Wei82**] for further details. As in the real case, there is a change of variables formula involving the Jacobian.

The volume forms on $T\backslash G$, $\mathfrak{t}$, $\mathfrak{g}$ chosen above give us a $G$-invariant measure $d\bar{g}$ on $T\backslash G$ and Haar measures $dX$, $dY$ on $\mathfrak{t}$, $\mathfrak{g}$ respectively.

**7.4. Expression for $D(X)$ in terms of roots.** Let $R$ be the set of roots of $T$ in $\mathfrak{g}$. Here we are talking about the absolute root system, a subset of the group $X^*(T)$ of characters on $T$ over $\bar{F}$. The differentials of the roots are linear forms on $\mathfrak{t} \otimes_F \bar{F}$ and hence yield $\bar{F}$-valued functions on $\mathfrak{t}$; these functions on $\mathfrak{t}$ will also be called roots, but no confusion should result from this. It is clear from the definitions that

$$D(X) = \prod_{\alpha \in R} \alpha(X) \tag{7.4.1}$$

for any $X \in \mathfrak{t}$. From this it follows that the differential $d\beta$ is an isomorphism at all points $(g, X)$ such that $X$ is regular. (Recall that a semisimple element in $\mathfrak{g}$ is said to be *regular* if its centralizer in the algebraic group $G$ is a maximal torus, and that $X \in \mathfrak{t}$ is regular in this sense if and only if no root of $T$ vanishes on it.)

**7.5. $D$ as polynomial function on $\mathfrak{g}$.** Let $\ell$ denote the absolute rank of $G$, in other words, the dimension of any maximal torus in $G$. For any $X \in \mathfrak{g}$ we can consider the characteristic polynomial of the endomorphism $\mathrm{ad}(X)$. Each individual coefficient of this characteristic polynomial is a polynomial function of $X$, and since generically $\mathrm{ad}(X)$ has the eigenvalue 0 with multiplicity $\ell$, we see that the lowest non-vanishing coefficient occurs in front of the $\ell$-th power of the variable and is equal to $D(X)$ for $X \in \mathfrak{t}$. In this way we see that the function $D(X)$ defined above for $X \in \mathfrak{t}$ extends to a polynomial function (still denoted by $D$) on all of $\mathfrak{g}$, which explains why we did not include $T$ in the notation. Note that $D(X) \neq 0$ if and only if $X$ is regular semisimple.

**7.6. Decomposition of $\mathfrak{g}_{\mathrm{rs}}$ as a disjoint union of open subsets $\mathfrak{g}_{\mathrm{rs}}^T$.** Let $\mathfrak{t}_{\mathrm{reg}}$ be the set of regular elements in $\mathfrak{t}$, let $\mathfrak{g}_{\mathrm{rs}}$ be the set of regular semisimple elements in $\mathfrak{g}$, and let $\mathfrak{g}_{\mathrm{rs}}^T$ be the subset of $\mathfrak{g}_{\mathrm{rs}}$ consisting of all elements that are conjugate under $G$ to some element of $\mathfrak{t}_{\mathrm{reg}}$. Then the map

$$(T\backslash G) \times \mathfrak{t}_{\mathrm{reg}} \to \mathfrak{g} \tag{7.6.1}$$

(obtained from $\beta$ by restriction) is a local isomorphism of $p$-adic manifolds and its image, namely $\mathfrak{g}_{\mathrm{rs}}^T$, is open in $\mathfrak{g}$. The fiber of $\beta$ through $(g, X) \in (T\backslash G) \times \mathfrak{t}_{\mathrm{reg}}$ has $|W_T|$ elements, namely those of the form $(wg, w(X))$ with $w$ ranging through $W_T$.

To complete our picture of $\mathfrak{g}_{\mathrm{rs}}$, we note that its complement has measure 0, and that

$$\mathfrak{g}_{\mathrm{rs}} = \coprod_T \mathfrak{g}_{\mathrm{rs}}^T, \tag{7.6.2}$$

where the union ranges over a set of representatives $T$ for the set of $G(F)$-conjugacy classes of maximal $F$-tori in $G$.

**7.7. First form of the Weyl integration formula.** These considerations lead to the following formula, known as the Weyl integration formula. Let $f \in C_c^\infty(\mathfrak{g})$. Then

$$\int_{\mathfrak{g}} f(Y)\, dY = \sum_T |W_T|^{-1} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \int_{T\backslash G} f(g^{-1}Xg)\, d\bar{g}\, dX, \tag{7.7.1}$$

where $d\bar{g}$, $dY$, $dX$ are the measures on $T\backslash G$, $\mathfrak{g}$, $\mathfrak{t}$ respectively that were introduced in 7.3, and where the sum ranges over a set of representatives $T$ for the set of $G(F)$-conjugacy classes of maximal $F$-tori in $G$. Since the complement of $\mathfrak{t}_{\mathrm{reg}}$ in $\mathfrak{t}$ has measure 0, we could equally well integrate over $\mathfrak{t}$ instead of $\mathfrak{t}_{\mathrm{reg}}$. Moreover we are not obliged to stick with precisely these measures $d\bar{g}$, $dY$, $dX$. Clearly it is only the product $d\bar{g}\, dX$ that matters in the Weyl integration formula, so we are free to multiply $d\bar{g}$ by a constant as long as we divide $dX$ by the same constant, and we are free to multiply $dY$ by a constant as long as we arrange that the product $d\bar{g}\, dX$

is multiplied by the same constant (for all $T$). For any such choices of measures we say that $d\bar{g}\,dX$ is *compatible* with $dY$.

Actually there are several useful variants of the Weyl integration formula, one of which is the one we will actually use later. For this we need some further preparation. We again need to maintain a notational distinction between an algebraic group and its group of $F$-points. We return to assuming that $G$ is split over $F$.

**7.8. Review of Levi subgroups and the definition of $\mathcal{L}$.** By a Levi subgroup $M$ of $G$ we mean some Levi component of a parabolic $F$-subgroup of $G$. We write $A_M$ for the maximal $F$-split torus in the center of $M$. In particular $A_G$ denotes the maximal $F$-split torus in the center of $G$. A basic fact about Levi subgroups is that $M$ is the centralizer in $G$ of $A_M$.

As usual let us fix a split maximal torus $A$ in $G$. Then $A_M$ is conjugate under $G(F)$ to a subtorus of $A$. Thus, after replacing $M$ by a conjugate, we may assume that $A_M \subset A$. The condition $A_M \subset A$ is equivalent to the condition $M \supset A$. [Use that $M$ is the centralizer of $A_M$ and that $A$ is its own centralizer.] We write $\mathcal{L} = \mathcal{L}(A)$ for the set of Levi subgroups $M$ of $G$ such that $M \supset A$.

**7.9. Definition of $\mathcal{T}_M$.** Let $T$ be a maximal $F$-torus of $G$, and let $A_T$ denote the maximal $F$-split subtorus of $T$. Let $M$ denote the centralizer of $A_T$ in $G$, a Levi subgroup of $G$. We claim that $A_M = A_T$. Indeed, it is obvious that $A_T$ is central in $M$ and hence contained in $A_M$. On the other hand $T$ is contained in $M$ and hence is a maximal torus in $M$, which implies that $T$ contains the center of $M$. Therefore $A_M$ is contained in $T$ and hence in $A_T$.

The reason for introducing $M$ is that $T$ is *elliptic* in $M$, in the sense that $T/A_M$ is an anisotropic torus over $F$ (which implies that $T(F)/A_M(F)$ is compact). We choose a set $\mathcal{T}_M$ of representatives for the $M(F)$-conjugacy classes of elliptic maximal tori $T$ in $M$.

**7.10. Definition of the positive integer $n_T^M$.** Let $M$ be a Levi subgroup of $G$ and let $T$ be a maximal torus in $M$. We write $N_{M(F)}(T)$ for the normalizer in $M(F)$ of $T$. Then $N_{M(F)}(T)/T(F)$ is a finite group, and we write $n_T^M$ for its cardinality.

**7.11. Second form of the Weyl integration formula.** We return to writing $G$ instead of $G(F)$. Let $f \in C_c^\infty(\mathfrak{g})$. Then

$$\int_{\mathfrak{g}} f(Y)\,dY = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \int_{A_M\backslash G} f(g^{-1}Xg)\,d\dot{g}\,dX, \tag{7.11.1}$$

where $W$ (respectively, $W_M$) denotes the Weyl group of $A$ in $G$ (respectively, $M$), and where $d\dot{g}$ is the unique $G$-invariant measure on $A_M\backslash G$ such that

$$\int_{A_M\backslash G} \varphi(g)\,d\dot{g} = \int_{T\backslash G} \varphi(g)\,d\bar{g} \tag{7.11.2}$$

for every $\varphi \in C_c^\infty(T\backslash G)$. (We used here that $A_M\backslash T$ is compact.) Since we have replaced $d\bar{g}$ by $d\dot{g}$, we need to extend the terminology introduced in 7.7 by now saying that the measure $d\dot{g}\,dX$ is compatible with $dY$ (when $d\dot{g}$ has been obtained from $d\bar{g}$ as above, and $d\bar{g}\,dX$ is compatible with $dY$).

**7.12. Derivation of the second form of the Weyl integration formula from the first.** We write $N_{G(F)}(M)$ for the normalizer in $G(F)$ of $M$. The group $N_{G(F)}(M)/M(F)$ is finite, and we denote by $n_M^G$ its cardinality. We need a couple of lemmas in order to derive the second form of the Weyl integration formula from the first.

LEMMA 7.1. *Let $M$ be a Levi subgroup of $G$, and let $T$ be an elliptic maximal torus in $M$. Then the number of $M(F)$-conjugacy classes of maximal tori $T'$ in $M$ such that $T'$ is $G(F)$-conjugate to $T$ is equal to*

$$n_M^G \cdot n_T^M \cdot (n_T^G)^{-1}.$$

PROOF. Let $g \in G(F)$. We claim that $gTg^{-1} \subset M$ if and only if $g \in N_{G(F)}(M)$. Indeed, suppose that $gTg^{-1} \subset M$. Then $gTg^{-1}$ is a maximal torus in $M$, and therefore its split component $gA_Mg^{-1}$ contains $A_M$, hence equals $A_M$ (look at dimensions). Thus $g$ normalizes $A_M$, which implies that it also normalizes the centralizer of $A_M$, namely $M$. This proves the forward implication in the claim; the other implication is trivial. A consequence of the claim is that $N_{G(F)}(T)$ normalizes $M$ and hence normalizes $N_{G(F)}(T) \cap M(F) = N_{M(F)}(T)$.

It follows from the claim we just proved that the set of $M(F)$-conjugacy classes of $T' \subset M$ such that $T'$ is $G(F)$-conjugate to $T$ is in natural bijection with the set

$$M(F)\backslash N_{G(F)}(M)/N_{G(F)}(T),$$

and the cardinality of this set is clearly the index of

$$N_{G(F)}(T)/N_{M(F)}(T),$$

a group of order $n_T^G \cdot (n_T^M)^{-1}$, in

$$N_{G(F)}(M)/M(F),$$

a group of order $n_M^G$. This proves the lemma. □

The next lemma involves the set $\mathcal{L}$ of Levi subgroups of $G$ containing $A$. For $M, M' \in \mathcal{L}$ we write $M \sim M'$ if $M, M'$ are conjugate under $G(F)$. We write $\mathcal{L}/\sim$ for the set of equivalence classes in $\mathcal{L}$ for the equivalence relation $\sim$. Moreover we fix some Borel subgroup $B_0$ containing $A$, and write $\mathcal{P}_0$ for the set of parabolic subgroups of $G$ containing $B_0$ (called *standard* parabolic subgroups). Also, we write $\mathcal{F}(A)$ for the set of parabolic subgroups of $G$ containing $A$, and for $P \in \mathcal{F}(A)$ we write $M_P$ for the unique Levi subgroup of $P$ containing $A$. For a Levi subgroup $M$ we write $\mathcal{P}(M)$ for the set of parabolic subgroups of $G$ having $M$ as Levi component.

LEMMA 7.2. *Let $\psi$ be a function defined on the set $\mathcal{L}$ and assume that $\psi(M) = \psi(M')$ whenever $M \sim M'$. Then*

$$\begin{aligned} \sum_{M \in \mathcal{L}/\sim} (n_M^G)^{-1}\psi(M) &= \sum_{M \in \mathcal{L}} \frac{|W_M|}{|W|}\psi(M) \\ &= \sum_{P \in \mathcal{P}_0} |\mathcal{P}(M_P)|^{-1}\psi(M). \end{aligned} \tag{7.12.1}$$

PROOF. Let $M \in \mathcal{L}$ and let $g \in G(F)$. We claim that $gMg^{-1} \in \mathcal{L}$ if and only if $g \in N_{G(F)}(A) \cdot M(F)$. Indeed, suppose that $gMg^{-1} \in \mathcal{L}$. Then both $g^{-1}Ag$ and $A$ are split maximal tori in $M$, from which it follows that they are conjugate under

$M(F)$. Thus there exists $m \in M(F)$ such that $mg^{-1}$ normalizes $A$. This proves the forward implication in the claim; the reverse implication is clear.

Write $N_{G(F)}(M, A)$ for the intersection of the normalizers in $G(F)$ of $M$ and $A$, and write $N_W(M)$ for $\{w \in W : wMw^{-1} = M\}$, a subgroup of $W$ that contains $W_M$ as a normal subgroup. As a special case of the claim above we see that

$$N_{G(F)}(M) = N_{G(F)}(M, A) \cdot M(F),$$

and from this it follows that

$$N_{G(F)}(M)/M(F) = N_{G(F)}(M, A)/N_{M(F)}(A) = N_W(M)/W_M. \tag{7.12.2}$$

How many $M' \in \mathcal{L}$ are there such that $M' \sim M$? It follows from the claim proved above that the set of such $M'$ is simply the $W$-orbit of $M$ in $\mathcal{L}$, and therefore its cardinality is equal to the index $[W : N_W(M)]$, which by (7.12.2) is equal to

$$(n_M^G)^{-1} \frac{|W|}{|W_M|}. \tag{7.12.3}$$

The first equality in the lemma follows from (7.12.3).

Finally, the second sum in the statement of the lemma can obviously be rewritten as

$$\sum_{P \in \mathcal{F}(A)} \frac{|W_{M_P}|}{|W|} |\mathcal{P}(M_P)|^{-1} \psi(M). \tag{7.12.4}$$

Now any $P \in \mathcal{F}(A)$ is conjugate under $W$ to a unique standard parabolic subgroup, and the stabilizer in $W$ of $P$ is $W_{M_P}$. Therefore (7.12.4) is equal to the third sum in the statement of the lemma. $\square$

Now let's return to the Weyl integration formula. From (7.11.2) and Lemma 7.1 we see that our first form (7.7.1) of that formula can be rewritten as

$$\int_{\mathfrak{g}} f(Y)\, dY = \sum_{M \in \mathcal{L}/\sim} \frac{1}{n_M^G} \sum_{T \in \mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \int_{A_M \backslash G} f(g^{-1}Xg)\, d\dot{g}\, dX, \tag{7.12.5}$$

and then from Lemma 7.2 we see that it can also be rewritten in the form (7.11.1). One could also use Lemma 7.2 to rewrite the Weyl integration formula as a sum over standard parabolic subgroups. Similarly, there are several ways of rewriting the sums in the local trace formula, and the global trace formula as well (see the remarks after Thm. 6.1 in [**Art89a**]).

## 8. Preliminary discussion of the local trace formula

Now that we have a feel for how orbital integrals work on the group $GL_2$, it is time to begin a more systematic treatment. Harish-Chandra [**HC78, HC99, HC70**] developed harmonic analysis on the Lie algebra of $G$ and then used the exponential map to climb back to the group itself, following the same path he had taken for real Lie groups. Now in harmonic analysis on the group the two key objects are orbital integrals and irreducible characters. Orbital integrals still make sense on the Lie algebra (integrate over orbits for the adjoint action of $G$). What about irreducible characters? Have they been irretrievably lost in passing to the Lie algebra? No! Harish-Chandra discovered that the role played by irreducible characters on $G$ is played by Fourier transforms of orbital integrals on the Lie algebra (again he did this first in the real case). Of course Fourier transforms of orbital integrals are in many respects simpler than irreducible characters, and this partly explains why passing to the Lie algebra is so effective.

In any case for most of the rest of this article we are going to work on the Lie algebra rather than the group. We will follow what seems to be the shortest known path through the material, first proving Waldspurger's local trace formula on the Lie algebra [**Wal95**] and then using it as a tool to develop the rest of the theory. This path is not essentially different from the one taken by Harish-Chandra in the papers cited above, and at most key points is exactly the same.

**8.1. Local trace formula on the group.** We began this article by discussing the trace formula on compact groups. Before passing to the Lie algebra, we should briefly discuss Arthur's local trace formula [**Art76, Art87, Art89b, Art91a, Art91b**] on a $p$-adic group $G$. (Actually Arthur also allows real and complex groups.)

Choose a Haar measure $dg$ on $G$. Just as in the compact case, given $f_1, f_2 \in C_c^\infty(G)$, we get an integral operator on $L^2(G)$ with kernel function

$$K(x,y) = \int_G f_1(g) f_2(x^{-1}gy)\, dg, \tag{8.1.1}$$

a locally constant function on $G \times G$. The restriction of the kernel function to the diagonal will be denoted by $K(x)$ and is given by

$$K(x) = \int_G f_1(g) f_2(x^{-1}gx)\, dg, \tag{8.1.2}$$

Next Arthur uses Harish-Chandra's Plancherel theorem to rewrite $K(x)$ in spectral terms. However, since $G$ is no longer assumed to be compact, the kernel function usually fails to be compactly supported, the integral operator is usually not of trace class, and the integral over $G$ of $K(x)$ is usually divergent. As in the global trace formula, Arthur handles these difficulties by truncating both expressions for $K(x)$ before integrating over $G$, obtaining in the end a formula with a geometric side involving weighted orbital integrals (which generalize orbital integrals) and a spectral side involving weighted characters (which generalize characters).

**8.2. First steps towards the local trace formula on $\mathfrak{g}$.** We again write $\mathfrak{g}$ for the Lie algebra of our $p$-adic group $G$ (the $F$-points of a connected reductive group over a $p$-adic field $F$, just as before). Consider a pair of functions $f_1, f_2 \in C_c^\infty(\mathfrak{g})$ and use them to define a locally constant function $K(x)$ on $G$ by

$$K(x) = \int_{\mathfrak{g}} f_1(Y) f_2(x^{-1}Yx)\, dY, \tag{8.2.1}$$

where $dY$ is a Haar measure on $\mathfrak{g}$. (We are using the expression $x^{-1}Yx$ to denote the adjoint action of $x^{-1}$ on $Y$.) Clearly this function is the analog for $\mathfrak{g}$ of the function (8.1.2) above that is the starting point for the local trace formula on $G$. Arthur uses the Plancherel theorem on $G$ to obtain a second expression for (8.1.2); similarly, Fourier theory on the additive group $\mathfrak{g}$ yields an identity

$$\int_{\mathfrak{g}} f_1(Y) f_2(x^{-1}Yx)\, dY = \int_{\mathfrak{g}} \hat{f}_1(Y) \check{f}_2(x^{-1}Yx)\, dY, \tag{8.2.2}$$

where $\hat{f}$ and $\check{f}$ denote the two possible variants of the Fourier transform for $\mathfrak{g}$. In more detail, let us now choose a $G$-invariant non-degenerate symmetric bilinear form $B$ on $\mathfrak{g}$ and a non-trivial additive character $\psi$ on $F$. Then for $f \in C_c^\infty(\mathfrak{g})$ we define the first version of the Fourier transform by

$$\hat{f}(Y) = \int_{\mathfrak{g}} f(Z)\psi(B(Y,Z))\,dZ, \tag{8.2.3}$$

where $dZ$ is a self-dual Haar measure on $\mathfrak{g}$. We define $\check{f}$ by the same formula, except that we replace $\psi$ by $\psi^{-1}$. Thus $f \mapsto \check{f}$ is inverse to $f \mapsto \hat{f}$.

In case $G$ is compact we obtain the local trace formula on $\mathfrak{g}$ simply by regarding both sides of (8.2.2) as functions of $x$ and then integrating over $G$. In general this integral diverges, and we must truncate before integrating. The truncation needed on the Lie algebra is the same as the one Arthur uses on the group.

**8.3. Truncation.** In order to keep the structure theory of $G$ as simple as possible (for expository purposes) we assume from now on that $G$ is a split group, and we use the notation $B = AN$ and $K$ of 4.1. We will eventually need to let $B$ vary through the set $\mathcal{B}(A)$ of all Borel subgroups containing $A$, but it is sometimes convenient to fix one of them, which from now on we will denote by $B_0 = AN_0$. In addition we write $A_G$ for (the $F$-points of) the identity component of the center of our algebraic group $G$.

Recall that $\mu \in X_*(A)$ is said to be *dominant* if $\langle \alpha, x\rangle \geq 0$ for every simple root $\alpha$. There is a standard partial order $\leq$ on $X_*(A)$, defined as follows: $\nu \leq \mu$ means that $\mu - \nu$ is a non-negative integral linear combination of simple coroots. Note that $\nu \leq \mu$ implies that $\mu$ and $\nu$ have the same image in the quotient $\Lambda_G$ of $X_*(A)$ introduced in subsection 4.5. Of course all these notions depend on a choice of Borel subgroup, which determines the sets of simple roots and coroots. When we need to stress which Borel subgroup $B$ is being used, we will say $B$-dominant rather than dominant. However, in the discussion below we will use the fixed Borel subgroup $B_0$ and say dominant rather than $B_0$-dominant.

From the Cartan decomposition discussed in 4.3 we have

$$G = \coprod_{\nu} K\pi^\nu K, \tag{8.3.1}$$

where $\nu$ runs over the set of dominant cocharacters, and where $\pi^\nu$ means (as before) the image of $\pi$ under the homomorphism $F^\times \to A$ obtained from $\nu$. Each $K$-double coset is of course a compact subset of $G$, and therefore the Cartan decomposition gives a very precise way of understanding the non-compactness of $G$.

We are finally in a position to define the function $u^\mu$ that is used to truncate our integral. We need to choose a truncation parameter $\mu$, which is allowed to be any dominant element of $X_*(A)$. Let $G^\mu$ denote the subset of $G$ obtained by taking the union of the double cosets $K\pi^\nu K$ for all dominant cocharacters $\nu$ such that $\nu \leq \mu$. Since $G^\mu$ is a finite union of $K$-double cosets, it is compact.

We write $u^\mu$ for the characteristic function of the subset $G^\mu$ of $G$. For $f_1, f_2 \in C_c^\infty(\mathfrak{g})$ and any truncation parameter $\mu$ we put

$$K^\mu(f_1, f_2) := \int_G u^\mu(g) \int_{\mathfrak{g}} f_1(Y) f_2(g^{-1}Yg)\,dY\,dg. \tag{8.3.2}$$

Here we have written $dg$ for the unique Haar measure on $G$ giving $K$ measure 1. It is evident that the integrand of this double integral is compactly supported as well as locally constant, so that the double integral is convergent and can be manipulated in any way we like.

Multiplying both sides of (8.2.2) by $u^\mu$ and integrating over $G$, we get a very crude first version of the local trace formula on $\mathfrak{g}$, namely the equality

$$K^\mu(f_1, f_2) = K^\mu(\hat{f}_1, \check{f}_2). \tag{8.3.3}$$

Since both sides of the formula have the same shape (unlike what happens on the group), it is enough to analyze the left side.

**8.4. Using the Weyl integration formula to rewrite $K^\mu(f_1, f_2)$.** We now use the Weyl integration formula (7.11.1) to rewrite the inner integral in the expression (8.3.2) defining $K^\mu(f_1, f_2)$, obtaining

$$\begin{aligned}\int_{\mathfrak{g}} f_1(Y) f_2(g^{-1}Yg)\, dY = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)|\cdot \\ \int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}h^{-1}Xhg)\, d\dot{h}\, dX.\end{aligned} \tag{8.4.1}$$

The notation is the same as in the Weyl integration formula (7.11.1), so that in particular $d\dot{h}\, dX$ is compatible with $dY$ in the sense of 7.11. By adjusting both $d\dot{h}$ and $dX$ in such a way that $d\dot{h}\, dX$ remains unchanged, we now assume that $d\dot{h}$ is the quotient of the Haar measure on $G(F)$ giving $K$ measure 1 by the Haar measure $da_M$ on $A_M$ giving $A_M \cap K = A_M(\mathcal{O})$ measure 1.

Substitute (8.4.1) back into (8.3.2), change the order of integration so that the innermost integral becomes the one taken over $G$, change variables by replacing $g$ by $h^{-1}g$, and finally do the integration over $G$ in stages, first integrating over $A_M$ and then integrating over $A_M\backslash G$. This yields

$$\begin{aligned}K^\mu(f_1, f_2) = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)|\cdot \\ \int_{A_M\backslash G}\int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) u_M(h, g; \mu)\, d\dot{h}\, d\dot{g}\, dX,\end{aligned} \tag{8.4.2}$$

where

$$u_M(h, g; \mu) := \int_{A_M} u^\mu(h^{-1}a_M g)\, da_M. \tag{8.4.3}$$

To make further progress on the local trace formula we need to analyze the function $u_M$. In fact, as we will see, the more refined versions of the local trace formula are obtained from our crude one just by replacing $u_M$ by something simpler. We begin by rewriting the definition of $u_M$ in a more convenient form. Recall that we are writing $X$ for the set $G/K$ and $x_0$ for its base-point. Recall also the function inv from 3.4. It takes values in $K\backslash G/K$, which by the Cartan decomposition we have now identified with the set of dominant coweights in $X_*(A)$. It follows from all our various definitions that $a \mapsto u^\mu(h^{-1}ag)$ is the characteristic function of the set of $a \in A$ such that

$$\mathrm{inv}(h^{-1}agx_0, x_0) \le \mu. \tag{8.4.4}$$

Putting $x := gx_0$, $y := hx_0$, we conclude that $u_M(h, g; \mu)$ is the measure of the set of $a \in A_M$ such that

$$\mathrm{inv}(ax, y) \le \mu. \tag{8.4.5}$$

Therefore, in order to understand $u_M$ for any $M$, we need to understand, for fixed $x, y \in X$, the subset of $A$ consisting of all $a \in A$ satisfying (8.4.5). To do so, it is best to begin with the simplest non-trivial example, that of $GL_2$. This will be the topic of the next section.

## 9. Calculation of $u_M$ for $G = GL_2$

In this section $G$ is $GL_2(F)$.

**9.1. Variant of truncation for $GL_2$.** In this special case it seems more convenient to do the truncation slightly differently. Recall the function

$$K(x) = \int_{\mathfrak{g}} f_1(Y) f_2(x^{-1}Yx)\, dY$$

on $G$ that we need to truncate. Our method was to multiply this function by the characteristic function of a compact subset of $G$ and then to integrate over $G$. However, the function $K(x)$ is obviously invariant under translation by $A_G$, so another perfectly good way to proceed is to multiply $K(x)$ by the characteristic function of a compact subset of $G/A_G$ and then integrate over $G/A_G$. This is what we will do for $G = GL_2$.

Our truncation parameter will be a non-negative integer $D$. Given $D$, we then put

$$G_D := \{g \in G : d(gv_0, v_0) \le D\}.$$

Here $v_0$ is the usual base vertex in the tree, and $d$ denotes the usual metric on the tree. Then $G_D$ is the inverse image in $G$ of a compact subset of $G/A_G$. We write $u_D$ for the characteristic function of the subset $G_D$ of $G$. Our weight factor will be

$$u_M(h, g; D) = \int_{A_M/A_G} u_D(h^{-1}ag)\, d\dot{a},$$

where $d\dot{a}$ denotes the quotient $da_M/da_G$ of the Haar measures $da_M$, $da_G$ on $A_M$, $A_G$ respectively that give measure 1 to their intersections with $K$.

Putting $v := gv_0$ and $w := hv_0$, we see that $u(h, g; D)$ is the measure of the set of $a \in A_M/A_G$ such that

$$d(av, w) \le D. \tag{9.1.1}$$

This condition is reminiscent of ones we have seen before and can be understood easily using the geometry of the tree.

**9.2. The case $M = A$.** We now define $d(v)$ to be the distance from $v$ to the standard apartment (and the same for $w$). We warn the reader that later on in the article, when we are working with general split groups, we will use the notation $d(\cdot)$ for a different purpose.

Lemma 9.1. *If $D \ge d(v) + d(w)$, then*

$$u_A(h, g; D) = 2\big(D - d(v) - d(w)\big) + 1. \tag{9.2.1}$$

*If $D < d(v) + d(w)$, then $u_A(h, g; D)$ is a real number between* 0 *and* 1.

Proof. Consider the shortest path in the tree from the vertex $v$ to the standard apartment, let $v'$ denote the other endpoint of this shortest path (so that $v'$ is some vertex in the standard apartment), and note that $d(v) = d(v, v')$. In the same way, from our other vertex $w$, we get $w'$ and $d(w) = d(w, w')$.

So long as $v' \neq w'$ it is clear that

$$d(v,w) = d(v) + d(w) + d(v',w'), \tag{9.2.2}$$

and when $v' = w'$ we at least have the inequality

$$d(v,w) \leq d(v) + d(w) \tag{9.2.3}$$

with strict inequality when there is some overlap between the two shortest paths. Thus $d(v)$, $d(w)$ do not quite determine $d(v,w)$. Nevertheless, we can assert that the condition

$$d(v,w) \leq D \tag{9.2.4}$$

is equivalent to the condition

$$d(v) + d(w) + d(v',w') \leq D \tag{9.2.5}$$

so long as $d(v) + d(w) \leq D$.

Of course it is really $d(av,w)$ that we care about. Since the action of $A$ on the tree preserves the standard apartment, it is clear that $d(av) = d(v)$ and $(av)' = av'$. We conclude that, so long as $d(v) + d(w) \leq D$, the condition (9.1.1) is equivalent to the condition

$$d(av',w') \leq D - d(v) - d(w). \tag{9.2.6}$$

The condition (9.2.6) on $a \in A$ depends only upon the image of $a$ under the surjection $A \to \mathbb{Z}$ sending the diagonal matrix with entries $(a_1, a_2)$ to the integer $\mathrm{val}(a_1) - \mathrm{val}(a_2)$, and therefore the measure of the set of $a \in A/A_G$ satisfying (9.2.6) is equal to the number of lattice points $u'$ in the standard apartment whose distance to $w'$ is less than or equal to $D - d(v) - d(w)$, and this number is obviously equal to

$$2\big(D - d(v) - d(w)\big) + 1. \tag{9.2.7}$$

This proves the lemma when $D \geq d(v) + d(w)$.

On the other hand, when $D < d(v) + d(w)$, the condition (9.1.1) implies that $av' = w'$, so that the measure of the set of all such $a$ is a real number between 0 and 1. □

**9.3. The case $M = G$.** For an elliptic torus $T$ we have $A_T = A_G$, $M = G$, and $u_G(h,g;D) = u_D(h^{-1}g)$, which is equal to 1 if $d(v,w) \leq D$ and is 0 otherwise.

**9.4. The functions $\tilde{v}_A$ and $\tilde{v}_G$.** The explicit computations done above show that for fixed $g$, $h$ and for all sufficiently large $D$, the value of $u_M(h,g;D)$ is given by

$$\begin{aligned} u_A(h,g;D) &= 2\big(D - d(v) - d(w)\big) + 1 \\ u_G(h,g;D) &= 1 \end{aligned} \tag{9.4.1}$$

How large $D$ has to be depends of course on $g$, $h$.

We have already mentioned that more refined versions of the local trace formula on $\mathfrak{g}$ will be obtained by replacing $u_M$ by simpler related functions. To keep our notation consistent with that used later in the general case, we will denote the next weight factor to be considered by $\tilde{v}_M(h,g;D)$. In the case of $GL_2$ we take the right sides of (9.4.1) as our definitions of $\tilde{v}_M$, in other words, we put

$$\begin{aligned} \tilde{v}_A(h,g;D) &= 2\big(D - d(v) - d(w)\big) + 1 \\ \tilde{v}_G(h,g;D) &= 1 \end{aligned} \tag{9.4.2}$$

for all $g$, $h$, $D$.

For later use (in applying Lebesgue's dominated convergence theorem) we note

$$\begin{aligned} |u_A(h,g;D) - \tilde{v}_A(h,g;D)| &\le 2\big(d(v) - d(w)\big) \\ |u_G(h,g;D) - \tilde{v}_G(h,g;D)| &\le 1. \end{aligned} \tag{9.4.3}$$

There is one final comment to make about $\tilde{v}_M$. Until now $D$ has been the non-negative integer $m-n$. However, as we see from (9.4.2), the definition of $\tilde{v}_M(h,g;D)$ still makes sense for any real number $D$.

## 10. The local trace formula for the Lie algebra of $G = GL_2$

In this section $G$ is again $GL_2(F)$.

**10.1. Next form of the local trace formula for $GL_2$.** Our preliminary version of the local trace formula for the Lie algebra of $GL_2$ says that

$$K^D(f_1,f_2) = K^D(\hat{f}_1,\check{f}_2) \tag{10.1.1}$$

with

$$\begin{aligned} K^D(f_1,f_2) = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)|\cdot \\ \int_{A_M\backslash G}\int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) u_M(h,g;D)\,d\dot{h}\,d\dot{g}\,dX. \end{aligned} \tag{10.1.2}$$

In the case of $GL_2$ we have also defined functions $\tilde{v}_M$ that are closely related to the functions $u_M$ appearing in (10.1.2). We now define $J^D(f_1,f_2)$ by the formula

$$\begin{aligned} J^D(f_1,f_2) = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)|\cdot \\ \int_{A_M\backslash G}\int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) \tilde{v}_M(h,g;D)\,d\dot{h}\,d\dot{g}\,dX. \end{aligned} \tag{10.1.3}$$

The only difference between this expression and the previous one is that $u_M$ has been replaced by $\tilde{v}_M$. Recall that $\tilde{v}_M(h,g;D)$ is defined for all $D\in\mathbb{R}$, so the same is true of $J^D(f_1,f_2)$. Looking back at the definition of $\tilde{v}_M$, we see that the convergence of the double integral appearing in (10.1.3) is an immediate consequence of the following two lemmas.

Lemma 10.1. *For any maximal torus $T$ in $G$ the function*

$$X \mapsto |D(X)|^{1/2} \int_{A_T\backslash G} f(g^{-1}Xg)\,d\dot{g} \tag{10.1.4}$$

*on $\mathfrak{t}_{\mathrm{reg}}$ is bounded and locally constant on $\mathfrak{t}_{\mathrm{reg}}$. Moreover this function is compactly supported on $\mathfrak{t}$, in the sense that there exists a compact subset $C$ of $\mathfrak{t}$ such that it vanishes off $C\cap\mathfrak{t}_{\mathrm{reg}}$. In particular the integral*

$$\int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)|^{1/2} \int_{A_T\backslash G} f(g^{-1}Xg)\,d\dot{g}\,dX \tag{10.1.5}$$

*converges.*

In the definition of $\tilde{v}_A(h,g;D)$ we were regarding the function $d(v)$ as a function on $G$. Now we make this more explicit, putting (for $g\in G$) $d(g) := d(gv_0)$, where $v_0$ is the base vertex in the tree.

LEMMA 10.2. *The function*

$$X \mapsto |D(X)|^{1/2} \int_{A\backslash G} f(g^{-1}Xg)d(g)\,d\dot{g} \tag{10.1.6}$$

*on* $\mathrm{Lie}(A)_{\mathrm{reg}}$ *is locally constant on* $\mathrm{Lie}(A)_{\mathrm{reg}}$. *Moreover this function is compactly supported on* $\mathrm{Lie}(A)$*, in the sense that there exists a compact subset* $C$ *of* $\mathrm{Lie}(A)$ *such that it vanishes off* $C \cap \mathrm{Lie}(A)_{\mathrm{reg}}$. *Finally, the integral*

$$\int_{\mathrm{Lie}(A)_{\mathrm{reg}}} |D(X)|^{1/2} \int_{A\backslash G} f(g^{-1}Xg)d(g)\,d\dot{g}\,dX \tag{10.1.7}$$

*converges.*

The first lemma makes sense and is true for any connected reductive $G$. The same will turn out to be true of the second lemma once we have defined a suitable generalization of the function $d(g)$. We prefer to prove these results in general, and will therefore defer their proofs till later (see Theorems 17.10, 17.11 and 20.6).

Granting the two lemmas, we can now state and prove the second form of the local trace formula for the Lie algebra of $GL_2$.

THEOREM 10.3. *For all* $f_1, f_2 \in C_c^\infty(\mathfrak{g})$ *and all* $D \in \mathbb{R}$ *there is an equality*

$$J^D(f_1, f_2) = J^D(\hat{f}_1, \check{f}_2). \tag{10.1.8}$$

PROOF. The functions $f_1$, $f_2$ will remain fixed throughout the proof. The first thing to note is that the function $D \mapsto J^D(f_1, f_2)$ is affine linear, or, in other words, has the form $D \mapsto aD + b$ for suitable real numbers $a$, $b$.

For any non-negative integer $D$ consider the difference $K^D(f_1, f_2) - J^D(f_1, f_2)$, which is given by the expression

$$\sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \cdot \int_{A_M\backslash G} \int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) \cdot \\ \big(u_M(h, g; D) - \tilde{v}_M(h, g; D)\big)\, d\dot{h}\, d\dot{g}\, dX.$$

As $D \to +\infty$ the pointwise limit of the sequence $u_M(h, g; D) - \tilde{v}_M(h, g; D)$ is 0. Moreover the estimate (9.4.3) for $u_M(h, g; D) - \tilde{v}_M(h, g; D)$, in conjunction with the two lemmas above, shows that Lebesgue's dominated convergence theorem can be applied, yielding the conclusion that

$$K^D(f_1, f_2) - J^D(f_1, f_2) \to 0 \text{ as } D \to +\infty. \tag{10.1.9}$$

Now consider the difference $\Delta(D) := J^D(f_1, f_2) - J^D(\hat{f}_1, \check{f}_2)$, which we are trying to prove is zero. On the one hand, we know that $\Delta(D)$ is an affine linear function of $D \in \mathbb{R}$. On the other hand, applying (10.1.9) to both $(f_1, f_2)$ and $(\hat{f}_1, \check{f}_2)$ and using the first form of the local trace formula, we see that the sequence $\Delta(D)$ has limit 0 as $D \to +\infty$. It follows that $\Delta(D)$ is identically zero, as desired. $\square$

**10.2. Final form of the local trace formula for $GL_2$.** In the course of proving the theorem above we saw that $J^D(f_1, f_2)$ is an affine linear function of $D \in \mathbb{R}$. Therefore this theorem is actually an equality between two affine linear functions of $D$, and entails two equalities, one between the linear terms of the two sides, and one between their constant terms. However this is less interesting than it might at first seem, since the equality between linear terms is a consequence of the local trace formula on $\mathrm{Lie}(A)$. (Something similar happens for general $G$, involving the local trace formulas for the Lie algebras of the various Levi subgroups of $G$.)

Thus all the real content of the theorem is in the equality of the constant terms of the two sides. These constant terms are obtained by setting $D$ equal to 0. However, we can capture the same information (modulo the local trace formula on $\mathrm{Lie}(A)$) by setting $D$ equal to any constant we like. The simplest result is obtained by taking $D$ to be $-1/2$, as one sees by looking back at how $\tilde{v}_A$ was defined. Doing so (and denoting the new weight factors by $v_M(h,g) := \tilde{v}_M(h,g;-1/2)$) yields the final form for the local trace formula on the Lie algebra of $GL_2$, namely

THEOREM 10.4. *For all $f_1, f_2 \in C_c^\infty(\mathfrak{g})$ there is an equality*

$$J(f_1,f_2) = J(\hat{f}_1, \check{f}_2), \tag{10.2.1}$$

*where $J(f_1,f_2)$ is defined by*

$$\begin{aligned} J(f_1,f_2) = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \cdot \\ \cdot \int_{A_M\backslash G}\int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) v_M(h,g)\, d\dot{h}\, d\dot{g}\, dX \end{aligned} \tag{10.2.2}$$

*with $v_M(h,g)$ given by*

$$\begin{aligned} v_A(h,g) &= -2\big(d(g)+d(h)\big) \\ v_G(h,g) &= 1. \end{aligned} \tag{10.2.3}$$

**10.3. Invariance versus non-invariance.** When $M = G$, in which case $v_G$ is identically 1, the double integral occurring in (10.2.2) is just the product of the orbital integrals (for $X$) of the functions $f_1$ and $f_2$. When $M = A$, the double integral is still taken over the $(G\times G)$-orbit of $(X,X)$ in $\mathfrak{g}\times\mathfrak{g}$, but we are using $v_M$ times the invariant measure on the orbit. Thus we are dealing with what Arthur calls a *weighted* orbital integral on $\mathfrak{g}\times\mathfrak{g}$. We see from the explicit formula for $v_A(h,g)$ that this weighted orbital integral is the sum of two terms, each term being a product of an orbital integral on one of the two factors of $\mathfrak{g}\times\mathfrak{g}$ and a weighted orbital integral on the other factor. This last phenomenon is an especially simple instance of more complicated *splitting formulas* of Arthur (see [**Art81**, Lemma 6.3] for instance) on general groups $G$ .

We can use these remarks to rewrite $J(f_1,f_2)$ in a form more suited to the application we will make in the next section. For any maximal torus $T$ in $G$ and any $X\in\mathfrak{t}_{\mathrm{reg}}$ we use $O_X$ (as usual) to denote the orbital integral

$$O_X(f) = \int_{A_T\backslash G} f(g^{-1}Xg)\, d\dot{g}. \tag{10.3.1}$$

Similarly, for any $X\in \mathrm{Lie}(A)_{\mathrm{reg}}$ we define the *weighted orbital integral* $WO_X$ by

$$WO_X(f) = \int_{A\backslash G} f(g^{-1}Xg) v_A(g)\, d\dot{g}, \tag{10.3.2}$$

where $v_A(g) := 2d(g)$. (Here we should warn the reader that when we define weight factors for general split groups, we will use a different normalization, in which the weight factor $v_A$ for $GL_2$ turns out to be $d(g)$ rather than $2d(g)$.)

It is then clear from these definitions that

$$J(f_1, f_2) = \sum_{T \in \mathcal{T}_G} |W_T^G|^{-1} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| O_X(f_1) O_X(f_2)\, dX$$
$$(10.3.3) \qquad - |W|^{-1} \int_{(\mathrm{Lie}\, A)_{\mathrm{reg}}} |D(X)| O_X(f_1) W O_X(f_2)\, dX$$
$$- |W|^{-1} \int_{(\mathrm{Lie}\, A)_{\mathrm{reg}}} |D(X)| W O_X(f_1) O_X(f_2)\, dX.$$

**10.4. Application of the local trace formula for the Lie algebra of $GL_2$.** Suppose that $O_X(f_2) = 0$ for all $X \in (\mathrm{Lie}\, A)_{\mathrm{reg}}$. Then the last term in (10.3.3) vanishes, and the remaining terms can be recombined using the Weyl integration formula. We conclude, for such $f_2$, that

$$(10.4.1) \qquad J(f_1, f_2) = \int_{\mathfrak{g}_{\mathrm{reg}}} f_1(X) F_2(X)\, dX,$$

where $F_2$ is the unique conjugation-invariant function on $\mathfrak{g}_{\mathrm{rs}}$ such that

$$(10.4.2) \qquad F_2(X) = \begin{cases} O_X(f_2) & \text{if } X \in \mathfrak{t}_{\mathrm{reg}} \text{ for some } T \in \mathcal{T}_G \\ -W O_X(f_2) & \text{if } X \in \mathrm{Lie}(A)_{\mathrm{reg}}. \end{cases}$$

Since $F_2$ is conjugation-invariant, the distribution $f_1 \mapsto J(f_1, f_2)$ is an invariant distribution on $\mathfrak{g}$. [The group $G$ acts by conjugation on $\mathfrak{g}$, hence on $C_c^\infty(\mathfrak{g})$, hence on $\mathcal{D}(\mathfrak{g})$, and an *invariant* distribution on $\mathfrak{g}$ is one that is fixed by this conjugation action.]

What can we say about the function $F_2$ (under our assumption on $f_2$)? It is a consequence of Lemmas 10.1 and 10.2 that $F_2$ is locally constant on $\mathfrak{g}_{\mathrm{rs}}$. We will often regard $F_2$ as a function on $\mathfrak{g}$ by extending it by 0 on the complement of $\mathfrak{g}_{\mathrm{rs}}$; this extended function is usually not locally constant on $\mathfrak{g}$. It is obvious that the support of $F_2$ is *bounded modulo conjugation*, in the sense that there is a compact subset $\omega$ in $\mathfrak{g}$ such that $F_2(X) = 0$ unless $X \in \mathfrak{g}$ is $G$-conjugate to an element in $\omega$ (any compact set $\omega$ on which $f_2$ is supported will do). Finally, we claim that $F_2$ is a locally integrable function on $\mathfrak{g}$.

We pause to recall that a measurable function $F$ on $\mathfrak{g}$ is said to be *locally integrable* if $F(X)f(X)$ is an integrable function on $\mathfrak{g}$ for all $f \in C_c^\infty(\mathfrak{g})$ (equivalently, for all $f$ obtained by taking characteristic functions of compact open subsets of $\mathfrak{g}$). In our case it is the convergence of the integral (10.4.1) for all $f_1$ which guarantees that $F_2$ is indeed locally integrable.

What is the most obvious source of functions whose hyperbolic orbital integrals vanish? Recall that $X \in \mathfrak{g}_{\mathrm{rs}}$ is said to be *elliptic* if its centralizer in $G$ is an elliptic maximal torus in $G$ (see 7.9 for the definition of elliptic maximal torus). We denote by $\mathfrak{g}_e$ the set of elliptic elements in $\mathfrak{g}_{\mathrm{rs}}$. It follows from the discussion in 7.6 that $\mathfrak{g}_e$ is open in $\mathfrak{g}$. [In the notation of that subsection the set $\mathfrak{g}_e$ is the union of open sets $\mathfrak{g}_{\mathrm{rs}}^T$ for $T$ ranging through $\mathcal{T}_G$.]

Now suppose that $\phi$ is a function lying in the subspace $C_c^\infty(\mathfrak{g}_e)$ of $C_c^\infty(\mathfrak{g})$. It is then obvious that the hyperbolic orbital integrals of $\phi$ vanish. It is less obvious, but true, that the hyperbolic orbital integrals of the Fourier transform $\hat{\phi}$ also vanish. We will prove a suitable generalization of this later (see Lemma 13.5).

As above we now use $\phi$ to define an invariant distribution $I_\phi$ on $\mathfrak{g}$ by putting

$$I_\phi(f) := J(f, \phi).$$

We have seen that $I_\phi$ is represented by the conjugation-invariant function on $\mathfrak{g}_{\mathrm{rs}}$, supported on $\mathfrak{g}_e$, whose value at $X \in \mathfrak{g}_e$ is $O_X(\phi)$. Thus the invariant distribution $I_\phi$ can be thought of as a "continuous linear combination" of elliptic regular semisimple orbital integrals.

Recall that the Fourier transform $\hat{T}$ of a distribution $T$ on $\mathfrak{g}$ is defined by

$$\hat{T}(f) = T(\hat{f}) \tag{10.4.3}$$

for all $f \in C_c^\infty(\mathfrak{g})$. Let's examine the Fourier transform of $I_\phi$. Using the local trace formula for the functions $\hat{f}, \phi$, we see that

$$\hat{I_\phi}(f) = J(\hat{f}, \phi) = J(f, \hat{\phi}),$$

and therefore, since the hyperbolic orbital integrals of $\hat{\phi}$ vanish, we conclude that the distribution $\hat{I_\phi}$ is represented by the locally integrable conjugation-invariant function $F$ on $\mathfrak{g}_{\mathrm{rs}}$ whose values are given by

$$F(X) = \begin{cases} O_X(\phi) & \text{if } X \in \mathfrak{t}_{\mathrm{reg}} \text{ for some } T \in \mathcal{T}_G \\ -W O_X(\phi) & \text{if } X \in \mathrm{Lie}(A)_{\mathrm{reg}}. \end{cases} \tag{10.4.4}$$

This establishes, for the group $GL_2$, a key special case of a more general result of Harish-Chandra (see Theorem 27.8), which says that for any invariant distribution $I$ on $\mathfrak{g}$ whose support is bounded modulo conjugation, the Fourier transform $\hat{I}$ is represented by a locally integrable function on $\mathfrak{g}$ that is locally constant on $\mathfrak{g}_{\mathrm{rs}}$. The formula for $F(X)$ when $X$ is elliptic is also due to Harish-Chandra [**HC78, HC99**]. The formula for $F(X)$ when $X$ is non-elliptic is due to Waldspurger. Indeed, Waldspurger [**Wal95**] proves a similar result for arbitrary $G$, and our approach here follows his.

## 11. Remarks on Euclidean space

Before introducing the weight factors (see section 12) occurring in weighted orbital integrals for general split groups, we need to discuss some elementary (but important) facts about Euclidean space, which will be used again later in the proof of the key geometric result needed for the local trace formula. These results are related to Langlands' combinatorial lemma [**Lan66**], [**Lan76**], [**Art76**, §2], [**Art78**, Lemma 6.3].

**11.1. The abstract set-up.** Throughout this section $V$ will denote a Euclidean space, in other words a finite dimensional real vector space equipped with a positive definite symmetric bilinear form $(v, w)$. We further suppose that $v_1, \dots, v_n$ is a basis for $V$ such that

$$(v_i, v_j) \leq 0 \quad \text{for all } i \neq j, \tag{11.1.1}$$

and we denote by $v_1^*, \dots, v_n^*$ the basis in $V$ dual to $v_1, \dots, v_n$, so that $(v_i, v_j^*) = \delta_{ij}$. (In the example of interest to us $v_1, \dots, v_n$ will be simple roots in a root system.)

We denote by $C$ the cone generated by $v_1^*, \dots, v_n^*$; thus $C$ consists of non-negative linear combinations of the elements $v_1^*, \dots, v_n^*$. We denote by $D$ the cone generated by $v_1, \dots, v_n$. The cones $C$ and $D$ are dual to each other in the sense that $D = \{v \in V : (v, w) \geq 0 \quad \forall\, w \in C\}$ (and the same with the roles of $C$ and $D$ interchanged).

LEMMA 11.1. *Let $V$ and $v_1, \dots, v_n$ be as above. Then*

(1) *For all $i, j$ we have $(v_i^*, v_j^*) \geq 0$. Equivalently, $D$ contains $C$.*
(2) *For any $j$ the vectors*

$$v_1, \dots, v_j, v_{j+1}^*, \dots, v_n^*$$

*form a basis for $V$; moreover for all $k$ the vector $v_k^*$ is a non-negative linear combination of $v_1, \dots, v_j, v_{j+1}^*, \dots, v_n^*$.*

PROOF. We begin by proving the first part of the lemma. Apply the Gram-Schmidt orthonormalization process to $v_1, \dots, v_n$, obtaining an orthonormal basis $e_1, \dots, e_n$. Thus the first basis vector $e_1$ is the unit vector in the same direction as $v_1$, the next basis vector $e_2$ is the unit vector in the same direction as $v_2 - (v_2, e_1)e_1$, and so on.

**Claim 1.** For all $i$ the vector $e_i$ is a non-negative linear combination of $v_1, \dots, v_i$, and the coefficient of $v_i$ in this combination is strictly positive. [To prove this use induction on $i$ together with (11.1.1).]

**Claim 2.** For all $i$ the vector $v_i^*$ is a non-negative linear combination of $e_i, \dots, e_n$, and the coefficient of $e_i$ in this combination is strictly positive. [To prove this note that $v_i^* = \sum_j (v_i^*, e_j)e_j$ and then use Claim 1.]

**Claim 3.** For all $i$ the vector $v_i^*$ is a non-negative linear combination of $v_1, \dots, v_n$. [To prove this combine Claims 1 and 2.]

We are done proving the first statement of the lemma, as it is just a restatement of Claim 3. (For root systems Claim 3 is the familiar fact that the positive Weyl chamber is contained in the cone of elements that are non-negative linear combinations of roots.)

Now we prove the second part of the lemma. The first statement is clear, since $v_{j+1}^*, \dots, v_n^*$ is obviously a basis for the orthogonal complement of the span of $v_1, \dots, v_j$. The second statement is trivial when $k \geq j+1$, so we just need to consider $k$ such that $1 \leq k \leq j$ and show that $v_k^*$ is a non-negative linear combination of $v_1, \dots, v_j, v_{j+1}^*, \dots, v_n^*$.

Let $W$ denote the span of $v_1, \dots, v_j$. We denote by $w_1^*, \dots, w_j^*$ the basis of $W$ dual to $v_1, \dots, v_j$; by the first part of the lemma (in the form of Claim 3) applied to $W$, we know that $w_k^*$ is a non-negative linear combination of $v_1, \dots, v_j$. Clearly $w_k^*$ is the orthogonal projection of $v_k^*$ on $W$, and since (as we have already remarked) $v_{j+1}^*, \dots, v_n^*$ is obviously a basis for the orthogonal complement of $W$, we can write $v_k^*$ as

$$v_k^* = w_k^* + b_{j+1}v_{j+1}^* + \cdots + b_n v_n^* \tag{11.1.2}$$

for real numbers $b_{j+1}, \dots, b_n$. Since we already know that $w_k^*$ is a non-negative linear combination of $v_1, \dots, v_k$, we just need to check that each $b_m$ is non-negative. It follows from (11.1.2) that

$$b_m = (v_k^*, v_m) - (w_k^*, v_m). \tag{11.1.3}$$

The term $(v_k^*, v_m)$ is 0 since $k \neq m$, and $(w_k^*, v_m) \leq 0$ since $w_k^*$ is a non-negative linear combination of $v_1, \dots, v_k$, showing that $b_m \geq 0$, as desired. □

As an immediate consequence of this lemma, we obtain the following corollary about root systems, which will be used later in our proof of the key geometric result needed for the local trace formula. We fix a Borel subgroup $B$ containing $A$ and write $\Delta$ for the set of simple roots of $A$, viewed as linear forms on $\mathfrak{a}$. We also need the fundamental weights $\varpi_\alpha \in (\mathfrak{a}/\mathfrak{a}_G)^* \subset \mathfrak{a}^*$ (indexed by $\alpha \in \Delta$); recall that $\langle \varpi_\alpha, \beta^\vee \rangle = \delta_{\alpha,\beta}$ for all $\alpha, \beta \in \Delta$. Finally we consider a parabolic subgroup $P = MU$ containing $B$ (with Levi component $M$ chosen so that $M \supset A$), and write $\Delta_M$ for the set of simple roots of $A$ in $M$ and $\Delta_U$ for the set of simple roots of $A$ occurring in Lie$(U)$; thus $\Delta$ is the disjoint union of $\Delta_M$ and $\Delta_U$.

LEMMA 11.2. *Let $\mu, \nu \in \mathfrak{a}$ and assume that $\mu$, $\nu$ have the same image under the canonical surjection $\mathfrak{a} \to \mathfrak{a}_G$. Assume further that $\langle \alpha, \nu \rangle \le \langle \alpha, \mu \rangle$ for all $\alpha \in \Delta_M$. Then $\nu \underset{B}{\le} \mu$ if and only if $\langle \varpi_\alpha, \nu \rangle \le \langle \varpi_\alpha, \mu \rangle$ for all $\alpha \in \Delta_U$.*

PROOF. Since we are given that $\mu, \nu$ have the same image in $\mathfrak{a}_G$, the condition $\nu \underset{B}{\le} \mu$ is equivalent to the condition that $\langle \varpi_\alpha, \nu \rangle \le \langle \varpi_\alpha, \mu \rangle$ for all $\alpha \in \Delta$. Therefore we need only show that the inequalities $\langle \alpha, \nu \rangle \le \langle \alpha, \mu \rangle$ for all $\alpha \in \Delta_M$ and $\langle \varpi_\alpha, \nu \rangle \le \langle \varpi_\alpha, \mu \rangle$ for all $\alpha \in \Delta_U$ together imply the inequalities $\langle \varpi_\alpha, \nu \rangle \le \langle \varpi_\alpha, \mu \rangle$ for all $\alpha \in \Delta$. This follows from the second part of Lemma 11.1. □

**11.2. Subspaces $V_I$ of the Euclidean space $V$.** Now let $I$ be a subset of $\{1, \dots, n\}$ and denote by $V_I$ the linear span of $\{v_i^* : i \in I\}$. The orthogonal complement of $V_I$ has basis $\{v_j : j \notin I\}$. We denote by $\pi_I$ the orthogonal projection of $V$ onto $V_I$.

We are going to see that $V_I$ inherits all the structure we have on $V$. As usual the restriction of the inner product on $V$ makes $V_I$ into a Euclidean space. Obviously $\{\pi_I v_i : i \in I\}$ is the basis in $V_I$ dual to $\{v_i^* : i \in I\}$. We denote by $C_I$ (respectively, $D_I$) the cone in $V_I$ generated by $\{v_i^* : i \in I\}$ (respectively, $\{\pi_I v_i : i \in I\}$).

LEMMA 11.3. *The following statements hold.*

(1) $C_I = C \cap V_I = \pi_I C$.
(2) $D_I = D \cap V_I = \pi_I D$.
(3) $(\pi_I v_i, \pi_I v_j) \le 0$ *for all* $i, j \in I$ *with* $i \ne j$.

PROOF. It is clear from the definitions that $C_I = C \cap V_I$ and that $C \cap V_I \subset \pi_I C$. It follows from the second statement of Lemma 11.1 that $C \cap V_I \supset \pi_I C$.

It is clear from the definitions that $D_I = \pi_I D$ and that $D \cap V_I \subset \pi_I D$. It remains only to show that $\pi_I D \subset D$. In other words, we must show that $(\pi_I d, c) \ge 0$ for all $d \in D$ and $c \in C$, and this follows from $(\pi_I d, c) = (d, \pi_I c)$ and $\pi_I C \subset C$.

To prove the last statement of the lemma, we begin by noting that

$$(\pi_I v_i, \pi_I v_j) = (\pi_I v_i, v_j).$$

Next we expand $v_i$ in the basis $v_k^*$, obtaining

$$v_i = \sum_{k=1}^{n} (v_i, v_k) v_k^*$$

and hence

$$(\pi_I v_i, v_j) = \sum_{k=1}^{n} (v_i, v_k)(\pi_I v_k^*, v_j).$$

In this sum the term indexed by $i$ is zero since $\pi_I v_i^* = v_i^*$ and $i \ne j$. Each remaining term is non-positive; indeed its first factor is non-positive (since $k \ne i$) and its second factor is non-negative (since $\pi_I C \subset C$). Thus $(\pi_I v_i, v_j)$ is non-positive, as desired. $\square$

**11.3. The convex polytope $E(x)$.** For $x \in C$ we put $E(x) := C \cap (x - D)$. Here $x - D$ has the obvious meaning: it consists of points of the form $x - d$ with $d \in D$ and is a cone with vertex $x$. Now $E(x)$ is compact (since it is contained in the obviously compact set $D \cap (x - D)$) and is the intersection of finitely many half-spaces; therefore $E(x)$ is a convex polytope. In other words $E(x)$ has finitely many extreme points (also called *vertices*) and is the convex hull of its set of vertices. We are going to determine the set of vertices of $E(x)$, as this is needed in the proof of Lemma 12.2. The reader is encouraged to draw a picture in the 2-dimensional case, where it is evident that $E(x)$ is a quadrilateral.

LEMMA 11.4. *The set of vertices of* $E(x)$ *is* $\{\pi_I x : I \subset \{1, \dots, n\}\}$.

PROOF. To simplify notation we put $x_I := \pi_I x$, an element in $C_I$. We begin by noting that $E(x) \cap V_I = C_I \cap (x_I - D_I)$ for any subset $I \subset \{1, \dots, n\}$, or, in other words, $E(x) \cap V_I$ is the analog $E_I(x_I)$ for $x_I$, $V_I$ of the set $E(x)$ for $x$, $V$. Indeed, using that $C \cap V_I = C_I$ and $\pi_I D = D_I$, we reduce to showing that

$$(x - D) \cap V_I = \pi_I(x - D). \tag{11.3.1}$$

It is clear that the left side is contained in the right side, but we must check that $\pi_I(x - D) \subset x - D$. Using again that $\pi_I D = D_I \subset D$, we see that it is enough to show that $x_I \in x - D$, and this follows from the second part of Lemma 11.1.

Now we prove the assertion of the lemma by induction on the dimension of $V$, the 0-dimensional case being trivial. Let $J \subset \{1, \dots, n\}$ be a subset having $n-1$ elements. Then $E(x)$ lies on one side of the hyperplane $V_J$, so that $E_J(x_J) = E(x) \cap V_J$ is a face of the polytope $E(x)$. By our inductive hypothesis the set of vertices of this face is $\{\pi_I x : I \subset J\}$. We have now accounted for all vertices of $E(x)$ lying in one of the codimension 1 faces $C_J$ of the cone $C$. It remains only to find the vertices of $E(x)$ lying in the interior of the cone $C$. An interior point of $C$ is clearly extreme in $E(x)$ if and only if it is extreme in $x - D$, and, since $x - D$ is a cone with vertex $x$, it has a unique extreme point, namely $x$. Therefore, if $x$ lies in the interior of $C$, there is exactly one vertex of $E(x)$ in the interior of $C$, namely $x$. Otherwise there is no vertex of $E(x)$ lying in the interior of $C$, but in this case $x$ is equal to $x_J$ for some $J$ as above. In either case we conclude that $\{\pi_I x : I \subset \{1, \dots, n\}\}$ is the set of vertices of $E(x)$. $\square$

## 12. Weighted orbital integrals in general

We worked out the local trace formula explicitly for $GL_2$ and found the weight factor $v_A(g) = 2d(g)$ appearing in our weighted orbital integrals. There are similar weight factors $v_M$ in the general case, which we are now going to discuss. Again we prefer to stick to the case of split groups, in order to keep the structure theory of the group as straightforward as possible.

In this section we work with any split connected reductive group $G$ over our $p$-adic field $F$. We use the same notation (e.g. $A$ and $K = G(\mathcal{O})$) as in 4.1. In addition, for any Levi subgroup $M$ in $G$ we write $\mathcal{P}(M)$ for the (finite) set of parabolic $F$-subgroups of $G$ admitting $M$ as Levi component. In the special case $M = A$ we often write $\mathcal{B}(A)$ instead of $\mathcal{P}(A)$; thus $\mathcal{B}(A)$ is the set of Borel subgroups containing $A$.

We will only be considering Levi subgroups $M$ containing $A$. In this case $A_M$ is a subgroup of $A$ and the lattice $X_*(A_M)$ is a subgroup of $X_*(A)$. Recall the definitions $\mathfrak{a} = X_*(A) \otimes_{\mathbb{Z}} \mathbb{R}$ and $\mathfrak{a}_M = X_*(A_M) \otimes_{\mathbb{Z}} \mathbb{R}$.

In the rest of this section $G$ will denote the group of $F$-points of our algebraic group.

**12.1. The maps $H_P : G \to \Lambda_M$.** Let $M$ be a Levi subgroup of $G$ containing $A$, and let $P = MU \in \mathcal{P}(M)$, $U$ being the unipotent radical of $P$. By the Iwasawa decomposition we have $G = MUK$. Therefore we can write any $g \in G$ as $g = muk$ with $m \in M$, $u \in U$, $k \in K$, and the element $m$ we obtain in this way is unique up to right multiplication by an element of $M(\mathcal{O}) = K \cap M$. Recall from 4.5 the homomorphism $H_M : M \to \Lambda_M$, and put $H_P(g) := H_M(m)$. Then $H_P$ is a well-defined map from $G$ to $\Lambda_M$. Clearly

$$H_P(mgk) = H_M(m) + H_P(g) \tag{12.1.1}$$

for all $g \in G$, $m \in M$ and $k \in K$.

The most basic case of this construction occurs when $M = A$, in which case the parabolic subgroup in question is a Borel subgroup $B$ containing $A$, and the map $H_B$ goes from $G$ to $X_*(A)$. It follows from the definitions that there is an important compatibility between the maps $H_P$ and $H_B$ whenever $P$ contains $B$, namely $H_P(g)$ is the image of $H_B(g)$ under the canonical surjection $X_*(A) \twoheadrightarrow \Lambda_M$.

Of course for a given $P = MU$ there are many Borel subgroups $B$ sandwiched between $P$ and $A$; via $B \mapsto B \cap M$ these are in one-to-one correspondence with Borel subgroups of $M$ containing $A$. Moreover, there is another easily verified compatibility, this time between $H_B : G \to X_*(A)$ and $H_{B\cap M} : M \to X_*(A)$, namely for $g = muk$ as above, we have

$$H_B(g) = H_{B\cap M}(m). \tag{12.1.2}$$

This is especially important in the case when $M$ has rank 1 (so that up to isogeny $M$ is the product of $SL_2$ and a split torus), in which case there are two Borel subgroups $B$ sandwiched between $P$ and $A$, corresponding to the two Borel subgroups in $SL_2$ containing the relevant split maximal torus of that group. For this reason, among others, we need to understand $H_B$ for $SL_2$.

So for a moment we consider the case $G = SL_2$. As usual we take $A$ to be the subgroup of diagonal matrices and $B = AN$ (respectively $\bar{B} = A\bar{N}$) to be the subgroup of upper triangular (respectively, lower triangular) matrices. Let $\alpha^\vee$ be the coroot corresponding to the unique root $\alpha$ occurring in Lie$(N)$. In order to understand $H_B$ and $H_{\bar{B}}$ for $SL_2$ it is enough (by (12.1.1) and the Iwasawa decomposition) to compute them on elements $n \in N$. Trivially we have that $H_B(n) = 0$. A simple computation with $2 \times 2$ matrices shows that $H_{\bar{B}}(n)$ is $-r\alpha^\vee$, where $r$ is the following non-negative integer. Look at the upper right matrix entry $y$ of $n$. If $y \in \mathcal{O}$, then $r = 0$. Otherwise, $r$ is the negative of the valuation of $y$. Identifying $\mathfrak{a}$ with the standard apartment (see 5.3) in the tree, we see that $H_B(n) = v_0$, $H_{\bar{B}}(n) = v_{-2r}$. Looking back at 5.8, we see that $v_{-r}$ is the point in the standard apartment that is closest to $nv_0$, and that $r$ is the distance from $nv_0$ to the standard apartment.

From this computation we conclude that for any element $g \in SL_2(F)$

$$H_B(g) - H_{\bar{B}}(g) = r\alpha^\vee \tag{12.1.3}$$

for some non-negative integer $r$, that $r$ is in fact equal to the distance from $gv_0$ to the standard apartment, that the point in the apartment closest to $gv_0$ is the midpoint of the line segment with endpoints at $H_B(g)$ and $H_{\bar{B}}(g)$ (again viewing these as points in the standard apartment), and that the length of this line segment is $2r$. (Since $\langle \alpha, \alpha^\vee \rangle = 2$, translation by $\alpha^\vee$ is translation by 2.) Before going on, we pause to notice that the weight factor $2d(g)$ entering into weighted orbital integrals for $GL_2$ has now been interpreted in terms of $H_B(g)$ and $H_{\bar{B}}(g)$: it is the length of the line segment joining the two points $H_B(g)$ and $H_{\bar{B}}(g)$ in $\mathfrak{a}$. We will see how this generalizes in a moment.

Now we return to our general split group $G$. The computation we just made for $SL_2$, combined with our previous remarks, shows that for any $g \in G$ the family of points $H_B(g)$ indexed by $B \in \mathcal{B}(A)$ is an example of what Arthur calls a positive $(G, A)$-orthogonal set; we discuss this notion next.

**12.2.** $(G, A)$**-orthogonal sets.** A family of points $x_B$ in $X_*(A)$ (respectively, $\mathfrak{a}$), one for each $B \in \mathcal{B}(A)$, is said to be a $(G, A)$-*orthogonal set* in $X_*(A)$ (respectively, $\mathfrak{a}$) if for every pair $B$, $B' \in \mathcal{B}(A)$ of adjacent Borel subgroups (meaning that the corresponding Weyl chambers in $\mathfrak{a}$ are adjacent) there exists an integer (respectively, real number) $r$ such that

$$x_B - x_{B'} = r\alpha^\vee, \tag{12.2.1}$$

where $\alpha^\vee$ is the unique coroot for $A$ that is positive for $B$ and negative for $B'$. (The explanation for the word "orthogonal" is that with respect to a Weyl group invariant inner product on $\mathfrak{a}$, the line segment joining $x_B$ and $x_{B'}$ is orthogonal to the root hyperplane in $\mathfrak{a}$ defined by $\alpha$.) When all the numbers $r$ are non-negative (resp., non-positive), the $(G, A)$-orthogonal set is said to be *positive* (resp., *negative*).

Consider once again a parabolic subgroup $P = MU$ with $M \supset A$. For any $(G, A)$-orthogonal set $(x_B)_{B \in \mathcal{B}(A)}$ it is easy to see that the points

$$(x_B)_{\{B \in \mathcal{B}(A) : B \subset P\}} \tag{12.2.2}$$

form an $(M, A)$-orthogonal set (identifying $\{B \in \mathcal{B}(A) : B \subset P\}$ with $\mathcal{B}^M(A)$, the analog of $\mathcal{B}(A)$ for the group $M$); moreover this $(M, A)$-orthogonal set is positive if the $(G, A)$-orthogonal set we started with is positive.

We see from (12.1.2) that the set of points $H_B(g)$ ($B \in \mathcal{B}(A)$ with $B \subset P$) is the positive $(M, A)$-orthogonal set in $X_*(A)$ attached to the element $m \in M$ obtained from the decomposition $g = muk$.

**12.3. Arthur's weight factor in case** $M = A$**.** Now we define Arthur's weight factor in the case $M = A$. Start with $g \in G$. Obtain from it the positive $(G, A)$-orthogonal set $H_B(g)$. Take the convex hull in $\mathfrak{a}$ of the points $H_B(g)$ ($B \in \mathcal{B}(A)$). Define the weight factor $v_A(g)$ to be the volume of this convex hull. By volume we mean Lebesgue measure in the real affine space consisting of all points in $\mathfrak{a}$ whose image under $\mathfrak{a} \twoheadrightarrow \mathfrak{a}_G$ is the same as the common image of all the points $H_B(g)$, the Lebesgue measure being normalized so that measure 1 is given to any fundamental domain for the (translation) action of the coroot lattice of $G$.

In the case of $GL_2$ the convex hull is the line segment discussed above, and its volume is its length, but measured with respect to the coroot lattice, which has index 2 in $X_*(A)/X_*(A_G)$, so that with our new definition $v_A(g)$ is equal to $d(g)$ instead of $2d(g)$.

Returning to the general case, note that the family $H_B(g)$ depends only on the coset $gK$, and that if $g$ is multiplied on the left by an element $a \in A$, then the whole family is translated by the vector $H_A(a)$, leaving its volume unchanged. It follows that

$$v_A(agk) = v_A(g) \tag{12.3.1}$$

for all $g \in G$, $a \in A$, and $k \in K$.

The convex hulls of positive $(G, A)$-orthogonal sets are very beautiful convex polytopes, about which much can be said. In particular, there is an interesting connection with the theory of toric varieties, as we will see in section 23. In a moment we will see how to get a better picture of the shape of these convex polytopes.

**12.4.** $(G, M)$**-orthogonal sets.** To define the weight factors in the general case we need to generalize the notion of $(G, A)$-orthogonal set. Let $M$ be a Levi subgroup containing $A$. The roots of $A$ in $G$ that are not roots in $M$ have non-zero restrictions to $\mathfrak{a}_M$, hence define hyperplanes (called walls) in $\mathfrak{a}_M$. The connected components of the complement in $\mathfrak{a}_M$ of the union of these hyperplanes are called chambers in $\mathfrak{a}_M$. For $M = A$ these are the usual Weyl chambers. There is a one-to-one correspondence between chambers in $\mathfrak{a}_M$ and the parabolic subgroups $P \in \mathcal{P}(M)$; the chamber corresponding to $P = MU$ is denoted by $\mathfrak{a}_P^+$ and is given by

$$\mathfrak{a}_P^+ := \{x \in \mathfrak{a}_M : \langle \alpha, x \rangle > 0 \quad \forall\, \alpha \in R_U\}, \tag{12.4.1}$$

where $R_U$ denotes the set of roots of $A$ occurring in $\mathfrak{u} := \mathrm{Lie}(U)$.

Consider two adjacent parabolic subgroups $P, P' \in \mathcal{P}(M)$. (By this we mean that the corresponding chambers are adjacent, or, in other words, separated by exactly one wall.) Recall that $\Lambda_M$ is the quotient of $X_*(A)$ by the coroot lattice for $M$. Now consider the collection of elements in $\Lambda_M$ obtained as the images of the coroots $\alpha^\vee$ where $\alpha$ runs through $R_U \cap R_{\bar{U}'}$ (with $\bar{P}' = M\bar{U}'$ denoting the parabolic subgroup in $\mathcal{P}(M)$ opposite $P'$). We define $\beta_{P,P'}$ to be the unique element in this collection such that all other members of the collection are positive integral multiples of $\beta_{P,P'}$. In case $M = A$, so that $P, P'$ are Borel subgroups, $\beta_{P,P'}$ is the unique coroot of $A$ that is positive for $P$ and negative for $P'$.

A family of points $x_P$ in $\Lambda_M$ (respectively, $\mathfrak{a}_M$), one for each $P \in \mathcal{P}(M)$, is said to be a $(G, M)$-*orthogonal set* in $\Lambda_M$ (respectively, $\mathfrak{a}_M$) if for every pair $P, P' \in \mathcal{P}(M)$ of adjacent parabolic subgroups there exists an integer (respectively, real number) $r$ (necessarily unique) such that

$$x_P - x_{P'} = r\beta_{P,P'}.$$

When all the numbers $r$ are non-negative (respectively, non-positive), the $(G, M)$-orthogonal set is said to be *positive* (respectively, *negative*).

It is clear that $\beta_{\bar{P},\bar{P}'} = -\beta_{P,P'}$. Therefore, if $P \mapsto x_P$ is a $(G, M)$-orthogonal set, then so is $P \mapsto x_{\bar{P}}$, and the one is positive if and only if the other is negative.

**12.5. The points** $x_P$ **associated to a** $(G, A)$**-orthogonal set.** Let $P = MU$ be a parabolic subgroup containing $A$, and let $(x_B)$ be a $(G, A)$-orthogonal set in $X_*(A)$. We have already observed that the points $(x_B)_{\{B \in \mathcal{B}(A) : B \subset P\}}$ form an $(M, A)$-orthogonal set in $X_*(A)$. In particular the difference between any two points in this $(M, A)$-orthogonal set is a sum of coroots for $M$, which means that they map to the same point in $\Lambda_M$. Thus we get a well-defined point $x_P \in \Lambda_M$ as the common image in $\Lambda_M$ of all the points $\{x_B : B \subset P\}$.

It is easy to see that the points $(x_P)_{P\in\mathcal{P}(M)}$ form a $(G, M)$-orthogonal set in $\Lambda_M$, and that this $(G, M)$-orthogonal set is positive if $(x_B)$ is positive.

The same things remain true when $X_*(A)$ is replaced by $\mathfrak{a}$ and $\Lambda_M$ is replaced by $\mathfrak{a}_M$.

**12.6. Arthur's weight factor $v_M$.** The weight factor in the general case is defined as follows. Start with $g \in G$. Obtain from it the family of points $H_P(g) \in \Lambda_M$, one for each $P = MU \in \mathcal{P}(M)$. This family $H_P(g)$ $(P \in \mathcal{P}(M))$ is a positive $(G, M)$-orthogonal set in $\Lambda_M$. By definition $v_M(g)$ is the volume of the convex hull of the images in $\mathfrak{a}_M$ of the points $H_P(g)$. (See subsection 24.7 for a precise normalization of the volume.)

The weight factor satisfies

$$v_M(mgk) = v_M(g) \tag{12.6.1}$$

for all $g \in G$, $m \in M$, and $k \in K$.

**12.7. Weighted orbital integrals.** Now we can define weighted orbital integrals for $G$. Let $T$ be a maximal torus such that $A_T$ is contained in $A$. As before, let $M$ be the centralizer of $A_T$, a Levi subgroup containing $A$ for which $A_M = A_T$. For $X \in \mathfrak{t}_{\mathrm{reg}}$ and $f \in C_c^\infty(G)$ put

$$WO_X(f) := \int_{A_M\backslash G} f(g^{-1}Xg)v_M(g)\,d\dot{g}. \tag{12.7.1}$$

Just as for unweighted orbital integrals, the semisimplicity of $X$ ensures that the integrand is locally constant and compactly supported on $A_M\backslash G$, so that the integral makes sense.

For $G = GL_2$ this agrees with our previous definition, apart from the factor of 2 mentioned in 12.3. When $T$ is elliptic, so that $M = G$, the weight factor is 1 and the weighted orbital integral is actually an orbital integral.

**12.8. Weyl group orbits as positive $(G, A)$-orthogonal sets.** As we have seen, each element $g \in G$ gives rise to a positive orthogonal set $H_B(g)$ $(B \in \mathcal{B}(A))$ in $X_*(A)$. However, there is an even simpler way to produce positive orthogonal sets in $X_*(A)$, and this construction is also relevant to the local trace formula.

First recall that an element $\mu \in X_*(A)$ is said to be dominant with respect to a Borel subgroup $B = AN$ if $\langle\alpha, \mu\rangle \geq 0$ for every root of $A$ occurring in Lie$(N)$; since we need to vary $B$ it is best to refer to such an element $x$ as $B$*-dominant*. The set of $B$-dominant elements in $X_*(A)$ is the intersection of $X_*(A)$ with the closure of the Weyl chamber corresponding to $B$ (the Weyl chamber itself being defined as the set of $x \in \mathfrak{a}$ for which all the inequalities $\langle\alpha, x\rangle \geq 0$ are strict).

It is a standard fact about root systems that the closure of any Weyl chamber serves as a fundamental domain for the action of the Weyl group $W$ of $A$ on $\mathfrak{a}$. Thus, given $\mu \in X_*(A)$, for any $B \in \mathcal{B}(A)$ there exists a unique $\mu_B \in X_*(A)$ such that $\mu_B$ lies in the $W$-orbit of $\mu$ and is $B$-dominant. Now suppose that $B_1, B_2 \in \mathcal{B}(A)$ are adjacent. Then there is a unique root $\alpha$ which is positive for $B_1$ and negative for $B_2$, and the corresponding root hyperplane is the unique one separating the

Weyl chambers corresponding to $B_1$, $B_2$. Let $s_\alpha \in W$ be the reflection across this hyperplane. Then $\mu_{B_2} = s_\alpha \mu_{B_1}$, which tells us that

$$\mu_{B_1} - \mu_{B_2} = \langle \alpha, \mu_{B_1} \rangle \alpha^\vee. \tag{12.8.1}$$

Now $\langle \alpha, \mu_{B_1} \rangle \geq 0$ since $\mu_{B_1}$ is $B_1$-dominant and $\alpha$ is positive for $B_1$, from which we conclude that the family $\mu_B$ ($B \in \mathcal{B}(A)$) is a positive $(G, A)$-orthogonal set.

**12.9. Special $(G, M)$-orthogonal sets.** We say that a $(G, A)$-orthogonal set of points $x_B$ is *special* if $x_B$ is $B$-dominant for all $B \in \mathcal{B}(A)$. More generally, we say that a $(G, M)$-orthogonal set of points $x_P$ is *special* if for all $P \in \mathcal{P}(M)$ the image of $x_P$ under $\Lambda_M \to \mathfrak{a}_M$ lies in the closure of the chamber $\mathfrak{a}_P^+$. (This property arises in [**Art91a**] but is not given a name. It seems convenient to have such a name.)

The $(G, A)$-orthogonal set obtained from the Weyl group orbit of $\mu$ as above provides the most obvious example of a special $(G, A)$-orthogonal set. In this example we saw that the $(G, A)$-orthogonal set was positive. This is a general phenomenon: any special $(G, M)$-orthogonal set $(x_P)$ is automatically positive. Indeed, for adjacent $P$, $P'$ we have

$$x_P - x_{P'} = r\beta_{P,P'}. \tag{12.9.1}$$

Let $\alpha$ be any root in $R_U \cap R_{\bar{U}'}$. Evaluating the root $\alpha$ on both sides of this equation, one gets a non-negative number on the left side, and since $\langle \alpha, \beta_{P,P'} \rangle$ is strictly positive, we conclude that $r$ must be non-negative. (To see painlessly that $\langle \alpha, \beta_{P,P'} \rangle$ is strictly positive, it is convenient to use a Weyl group invariant inner product to identify $\mathfrak{a}$ with its dual, so that $\alpha^\vee$, $\alpha$ become positive multiples of each other.)

Consider a special $(G, A)$-orthogonal set $(x_B)$ and a Levi subgroup $M$ containing $A$. Then for any parabolic subgroup $P \in \mathcal{P}(M)$, it is easy to see that the $(M, A)$-orthogonal set (12.2.2) obtained from $(x_B)$ is special. Moreover the $(G, M)$-orthogonal set $(x_P)$ in $\Lambda_M$ obtained from $(x_B)$ (see 12.5) is also special.

**12.10. Shape of the convex hull of a positive $(G, A)$-orthogonal set.** Consider a positive $(G, A)$-orthogonal set of points $x_B$ ($B \in \mathcal{B}(A)$) in $X_*(A)$. Let $B, B' \in \mathcal{B}(A)$. We no longer assume that they are adjacent. However $B$, $B'$ can be joined by a chain of Borel subgroups (all containing $A$) such that each consecutive pair in the chain is a pair of adjacent Borel subgroups. Now assuming that the chain is chosen to be as short as possible, the set of root hyperplanes separating the Weyl chambers of $B$ and $B'$ coincides with the set of hyperplanes separating the various consecutive pairs in our chain. In this way we see that $x_B - x_{B'}$ is a non-negative integral linear combination of the coroots $\alpha^\vee$ that are positive for $B$ and negative for $B'$. In particular we have

$$x_{B'} \underset{B}{\leq} x_B. \tag{12.10.1}$$

What is the meaning of the symbol $B$ below the inequality sign? We have been using the inequality $x \leq y$ to mean that $y - x$ is a non-negative integral linear combination of positive coroots, positive respect to some fixed $B_0 \in \mathcal{B}(A)$. Now we are letting the Borel subgroup vary, and we use the symbol $B$ to indicate which Borel subgroup we are using.

We also need to define $x \underset{B}{\leq} y$ for elements $x, y \in \mathfrak{a}$: say that $x \underset{B}{\leq} y$ if $y - x$ is a non-negative real linear combination of positive coroots.

Since (12.10.1) holds for all $B'$, the convex hull of the points $x_B$, which we denote by

$$\mathrm{Hull}\{x_B : B \in \mathcal{B}(A)\},$$

is contained in the convex cone

$$C_B^* := \{x \in \mathfrak{a} : x \underset{B}{\leq} x_B\}, \tag{12.10.2}$$

and since this is true for all $B$, we conclude that the convex hull is contained in the intersection of the cones $C_B^*$.

LEMMA 12.1. [**Art76**, Lemma 3.2] *Let $(x_B)$ be a positive $(G, A)$-orthogonal set in $\mathfrak{a}$. Then*

$$\mathrm{Hull}\{x_B : B \in \mathcal{B}(A)\} = \{x \in \mathfrak{a} : x \underset{B}{\leq} x_B \ \forall B \in \mathcal{B}(A)\}.$$

PROOF. It remains only to see that the convex hull of the points $x_B$ contains the intersection of the cones $C_B^*$. From the theory of convex sets we know that the convex hull in question is the intersection of all the halfspaces containing it, so it will suffice to show that any such halfspace contains one of the cones $C_B^*$. Say the halfspace is given by the set of points $x \in \mathfrak{a}$ such that $\langle \lambda, x \rangle \leq r$ (with $\lambda \in \mathfrak{a}^*$ and $r \in \mathbb{R}$). Choose $B \in \mathcal{B}(A)$ such that $\lambda$ is dominant for $B$. The halfspace contains the entire convex hull and thus contains $x_B$; the dominance of $\lambda$ then implies that the halfspace contains $C_B^*$. □

**12.11. A property of special $(G, A)$-orthogonal sets.** Let $x_B$ be a special $(G, A)$-orthogonal set of points in $X_*(A)$. In 12.5 we defined points $x_P \in \Lambda_M$, one for each parabolic subgroup $P = MU$ containing $A$. Using the canonical map $\Lambda_M \to \mathfrak{a}_M$, we obtain from $x_P$ an element $\bar{x}_P$ of $\mathfrak{a}_M \subset \mathfrak{a}$. We claim that $\bar{x}_P$ lies in the convex hull of the points $\{x_B : B \subset P\}$.

Indeed, we have seen in 12.2 that the points $x_B$ with $B \subset P$ form an $(M, A)$-orthogonal set. This reduces us to the case in which $P = G$. It is harmless to assume that $G$ is semisimple. Then we must show that the origin lies in the convex hull of the points $x_B$. This follows from Lemma 12.1 since by hypothesis $x_B$ is dominant for $B$, whence $x_B \underset{B}{\geq} 0$.

As an easy exercise the reader may wish to verify that a $(G, A)$-orthogonal set of points $x_B$ is special if and only if it is positive and satisfies the property that for every parabolic subgroup $P$ containing $A$ the point $\bar{x}_P$ lies in the convex hull of the points $\{x_B : B \subset P\}$. (Hint: Use all the parabolic subgroups $P = MU$ for which $M$ has rank 1.)

**12.12. Shape of the convex hull of a special $(G, A)$-orthogonal set.** For special $(G, A)$-orthogonal sets (which are positive, as we have seen), there is an even more useful description of the convex hull, given in terms of the shapes of its intersections with the closures of the various Weyl chambers. We use the same notation $C_B^*$ as above.

LEMMA 12.2. [**Art91a**, Lemma 3.1] *Let $H$ denote the convex hull of a special $(G, A)$-orthogonal set of points $x_B$. Let $B \in \mathcal{B}(A)$ and let $C_B$ be the set of $B$-dominant elements in $\mathfrak{a}$. Then the intersection of $H$ with the cone $C_B$ is equal to the intersection of $C_B^*$ with $C_B$.*

PROOF. It follows from Lemma 12.1 that $H \cap C_B$ is contained in $C_B^* \cap C_B$. It follows from Lemma 11.4 that $C_B^* \cap C_B$ is the convex hull of the points $\{\bar{x}_P : P \supset B\}$; thus, in order to show that $C_B^* \cap C_B$ is contained in $H \cap C_B$, it suffices to show that each $\bar{x}_P$ is in $H$. This was done in 12.11. □

**12.13. Shape of the convex hull of a $(G, M)$-orthogonal set.** Let $M$ be a Levi subgroup of $G$ containing $A$. For $P = MU \in \mathcal{P}(M)$ we now define a partial order $\underset{P}{\leq}$ on $\mathfrak{a}_M$ by saying that $x \underset{P}{\leq} y$ if $y - x$ is a non-negative linear combination of images under $\mathfrak{a} \twoheadrightarrow \mathfrak{a}_M$ of coroots $\alpha^\vee$ associated to roots $\alpha$ of $A$ in Lie$(U)$.

We write $\pi_M$ for the canonical surjection $\mathfrak{a} \twoheadrightarrow \mathfrak{a}_M$. Let $x, y \in \mathfrak{a}$. Clearly, for any $B \in \mathcal{B}(A)$ such that $B \subset P$ we have the implication

$$x \underset{B}{\leq} y \implies \pi_M(x) \underset{P}{\leq} \pi_M(y). \tag{12.13.1}$$

LEMMA 12.3. [**Art76**, Lemma 3.2] *Let $(x_P)$ be a positive $(G, M)$-orthogonal set in $\mathfrak{a}_M$. Then*

$$\mathrm{Hull}\{x_P : P \in \mathcal{P}(M)\} = \{y \in \mathfrak{a}_M : y \underset{P}{\leq} x_P \quad \forall\, P \in \mathcal{P}(M)\}.$$

PROOF. This generalizes Lemma 12.1 and can be proved the same way. □

**12.14. Another property of special orthogonal sets.** We let $M$ and $\pi_M : \mathfrak{a} \twoheadrightarrow \mathfrak{a}_M$ be as in the previous subsection. When $\mathfrak{a}_M$ is identified with a subspace of $\mathfrak{a}$, the map $\pi_M$ becomes orthogonal projection. Let $(x_B)$ be a positive $(G, A)$-orthogonal set in $\mathfrak{a}$, and let $(x_P)$ be the positive $(G, M)$-orthogonal set in $\mathfrak{a}_M$ obtained from $(x_B)$ as in subsection 12.5. Put Hull $:=$ Hull$\{x_B : B \in \mathcal{B}(A)\}$, and also put Hull$_M :=$ Hull$\{x_P : P \in \mathcal{P}(M)\}$. The next result is part of Lemmas 3.1 and 3.2 in [**Art81**].

PROPOSITION 12.1 (Arthur). *The image under* $\pi_M$ *of* Hull *is* Hull$_M$. *Moreover, if $(x_B)$ is special, then so is $(x_P)$, and*

$$\mathfrak{a}_M \cap \mathrm{Hull} = \mathrm{Hull}_M\,.$$

PROOF. We begin by proving that $\pi_M(\mathrm{Hull}) \subset \mathrm{Hull}_M$. By Lemma 12.3 we must show that $\pi_M(x_B) \underset{P}{\leq} x_P$ for all $B \in \mathcal{B}(A)$ and all $P \in \mathcal{P}(M)$. Choose $B' \in \mathcal{B}(A)$ such that $B' \subset P$. Then $x_B \underset{B'}{\leq} x_{B'}$ (see 12.10), and therefore from (12.13.1) it follows that $\pi_M(x_B) \underset{P}{\leq} \pi_M(x_{B'}) = x_P$.

Next we prove that $\mathrm{Hull}_M \subset \pi_M(\mathrm{Hull})$. For this it is enough to show that $x_P \in \pi_M(\mathrm{Hull})$ for all $P \in \mathcal{P}(M)$. This is clear since $x_P = \pi_M(x_B)$ for any $B \in \mathcal{B}(A)$ such that $B \subset P$. We are now done proving that $\pi_M(\mathrm{Hull}) = \mathrm{Hull}_M$.

Now suppose that $(x_B)$ is special, which means simply that for all $B \in \mathcal{B}(A)$ the point $x_B$ is dominant with respect to $B$. Let $P \in \mathcal{P}(M)$ and pick $B \in \mathcal{B}(A)$ such that $B \subset P$. By the first part of Lemma 11.3 we see that the point $x_P = \pi_M(x_B)$ lies in the closure of the chamber $\mathfrak{a}_P^+$ in $\mathfrak{a}_M$. (Note that this closure is the intersection of $\mathfrak{a}_M$ with the closed Weyl chamber for $B$.) Therefore $(x_P)$ is special.

Since we already know that Hull$_M$ is equal to $\pi_M(\mathrm{Hull})$, to prove that $\mathfrak{a}_M \cap \mathrm{Hull}$ coincides with both of them, it is enough to note that

$$\mathfrak{a}_M \cap \mathrm{Hull} \subset \pi_M(\mathrm{Hull})$$

(clear since $\pi_M$ is the identity on the subspace $\mathfrak{a}_M$) and that

$$\mathrm{Hull}_M \subset \mathfrak{a}_M \cap \mathrm{Hull}$$

(clear since each $x_P$ lies in Hull by the discussion in subsection 12.11). □

## 13. Parabolic descent and induction

In this section $G$ denotes a connected reductive group over our $p$-adic field $F$. Let $P = MU$ be a parabolic subgroup with Levi component $M$ and unipotent radical $U$. There always exists some compact open subgroup $K$ of $G$ such that $G = PK$; we fix such a subgroup $K$. We write $\mathfrak{p}$, $\mathfrak{m}$, $\mathfrak{u}$ for the Lie algebras of $P$, $M$, $U$ respectively. Thus $\mathfrak{p} = \mathfrak{m} \oplus \mathfrak{u}$.

Orbital integrals on $\mathfrak{m}$ can be related to orbital integrals on $\mathfrak{g}$ by Harish-Chandra's dual processes of parabolic descent and parabolic induction, as we will see in this section (which follows [**HC99**]). Often parabolic descent is used to prove statements about general maximal tori $T$ by reducing to the case in which $T$ is elliptic. (Take $M$ to be the centralizer of the split component of $T$.) We will encounter an application of this kind in 26.1. Parabolic descent will come up again when we are proving the local trace formula.

**13.1. Definition of $f^P$.** Given $f \in C_c^\infty(\mathfrak{g})$ we define a function $f^P \in C_c^\infty(\mathfrak{m})$ by

$$f^P(Y) := \int_{\mathfrak{u}} f(Y+Z)\, dZ. \tag{13.1.1}$$

Here $Y \in \mathfrak{m}$ and $dZ$ is Haar measure on $\mathfrak{u}$.

**13.2. Definition of cusp forms on $\mathfrak{g}$.** A function $f \in C_c^\infty(\mathfrak{g})$ is said to be a *cusp form* if $f^P$ is identically 0 for every parabolic subgroup $P$ of $G$ such that $P \neq G$. Looking ahead to (13.13.2), we see that the Fourier transform of a cusp form is a cusp form.

**13.3. Definition of $f^{(P)}$.** Given $f \in C_c^\infty(\mathfrak{g})$ we define a function $f^{(P)} \in C_c^\infty(\mathfrak{m})$ by

$$f^{(P)} := \tilde{f}^P, \tag{13.3.1}$$

where $\tilde{f} \in C_c^\infty(\mathfrak{g})$ is defined by

$$\tilde{f}(X) := \int_K f(k^{-1}Xk)\, dk, \tag{13.3.2}$$

$dk$ being the Haar measure on $K$ giving $K$ measure 1. The linear map $f \mapsto f^{(P)}$ depends on the choice of $K$ (and the measure $dZ$).

**13.4. Definition of parabolic induction $i_P^G$.** Let $T_M$ be a distribution on $\mathfrak{m}$. We define a distribution $i_P^G(T_M)$ on $\mathfrak{g}$ as follows: its value on a test function $f \in C_c^\infty(\mathfrak{g})$ is given by

$$i_P^G(T_M)(f) := T_M(f^{(P)}). \tag{13.4.1}$$

LEMMA 13.1. *Suppose that $T_M$ is an invariant distribution on $\mathfrak{m}$. Then $i_P^G(T_M)$ is an invariant distribution on $\mathfrak{g}$ and is independent of the choice of $K$.*

Proof. Recall from before that we use a subscript $G$ to denote coinvariants under $G$. Thus the space of invariant distributions on $\mathfrak{g}$ is the $\mathbb{C}$-linear dual of $C_c^\infty(\mathfrak{g})_G$.

Let $f \in C_c^\infty(\mathfrak{g})$. For any $g \in G$ define ${}^g f \in C_c^\infty(\mathfrak{g})$ by ${}^g f(X) = f(g^{-1}Xg)$. Now consider the map $\varphi : G \to C_c^\infty(\mathfrak{m})_M$ defined by

$$\varphi(g) = ({}^g f)^P. \tag{13.4.2}$$

Note that the vector-valued function $\varphi$ satisfies (for $p \in P$, $g \in G$)

$$\varphi(pg) = \delta_P(p)\varphi(g), \tag{13.4.3}$$

where $\delta_P$ is the modulus character on $P$ (see 2.3), given in this case by

$$\delta_P(p) = |\det(\mathrm{Ad}(p); \mathfrak{u})|.$$

Since $P$ is not unimodular, there is no $G$-invariant measure on the homogeneous space $P\backslash G$. However there is something similar, namely a non-zero $G$-invariant linear form $\oint_{P\backslash G}$ defined on the space of locally constant $\mathbb{C}$-valued functions $\psi$ on $G$ satisfying

$$\psi(pg) = \delta_P(p)\psi(g). \tag{13.4.4}$$

(The reason this works is that $\psi$ gives a measure on $P\backslash G$, and this measure can then be integrated over $P\backslash G$.) The linear form $\oint_{P\backslash G}$ is unique up to a non-zero scalar. Since $P\backslash G = (P \cap K)\backslash K$ and $\oint_{P\backslash G}$ is $G$-invariant (hence $K$-invariant), we see that (for a suitable normalization of $\oint_{P\backslash G}$), we have

$$\oint_{P\backslash G} \psi = \int_K \psi(k)\, dk. \tag{13.4.5}$$

We have the integration-in-stages formula

$$\oint_{P\backslash G} \int_P h(pg)\, dp = \int_G h(g)\, dg \tag{13.4.6}$$

for all $h \in C_c^\infty(G)$, where $dp$ is a left Haar measure on $P$ and $dg$ is a suitable Haar measure on $G$.

We can apply $\oint_{P\backslash G}$ to our vector-valued function $\varphi$, obtaining a well-defined element $\oint_{P\backslash G} \varphi \in C_c^\infty(\mathfrak{m})_M$. Replacing $f$ by a $G$-conjugate replaces $\varphi$ by a right translate, hence leaves $\oint_{P\backslash G} \varphi$ unchanged, so that $f \mapsto \oint_{P\backslash G} \varphi$ is a well-defined linear map $C_c^\infty(\mathfrak{g})_G \to C_c^\infty(\mathfrak{m})_M$.

From (13.4.5) we see that $\oint_{P\backslash G} \varphi$ is equal to the image of $f^{(P)}$ under $C_c^\infty(\mathfrak{m}) \to C_c^\infty(\mathfrak{m})_M$. This concludes the proof (and provides a way to define $f^{(P)}$ as an element of $C_c^\infty(\mathfrak{m})_M$ without having to choose $K$). □

**13.5. A variant.** This variant will not be used in this article and can be safely skipped. Let $\psi$ be as above. In some contexts it is natural to work with smaller compact open subgroups (for instance an Iwahori subgroup). So, just for this subsection, let $K$ be any compact open subgroup such that $\psi$ is right $K$-invariant. Then $P\backslash G/K$ is a finite set. For any $\psi$ as above we have (generalizing (13.4.5))

$$\oint_{P\backslash G} \psi = \sum_{x \in P\backslash G/K} \frac{\mathrm{meas}_{dg}(K)}{\mathrm{meas}_{dp}(P \cap xKx^{-1})} \psi(x) \tag{13.5.1}$$

where $dp$ and $dg$ are as in (13.4.6).

Therefore, if $f \in C_c^\infty(\mathfrak{g})$ is $\mathrm{Ad}(K)$-invariant under some compact open subgroup $K$, then $f^{(P)}$ (viewed in the $M$-coinvariants) is given by

$$f^{(P)} = \sum_{x \in P\backslash G/K} \frac{\mathrm{meas}_{dg}(K)}{\mathrm{meas}_{dp}(P \cap xKx^{-1})} ({}^x f)^P. \tag{13.5.2}$$

**13.6. Dependence on $P$.** We will see later (Corollary 27.7) that for any invariant distribution $T_M$ on $\mathfrak{m}$ the induced distribution $i_P^G(T_M)$ depends only on $M$, not on the choice of parabolic subgroup having Levi component $M$ (provided that one is careful about the choice of measure $dZ$).

**13.7. Analogous construction on $G$.** We are working on the Lie algebra, but there are analogs of $f^{(P)}$ and $i_P^G$ on the group $G$. More precisely

$$f^{(P)}(m) := \delta_P^{1/2}(m) \int_U \tilde{f}(mu)\, du \tag{13.7.1}$$

with $\tilde{f}$ again defined by making $f$ conjugation invariant under $K$.

With a suitable normalization of Haar measures $dg$, $dm$, $du$ one has the following basic fact, which explains the significance of parabolic induction of invariant distributions. Let $\pi_M$ be an irreducible representation of $M$, and let $\Theta_M$ be its distribution character (which depends on $dm$). Let $\pi$ be the representation of $G$ obtained from $\pi_M$ by (unitary) parabolic induction, and let $\Theta$ be its distribution character. Then

$$i_P^G(\Theta_M) = \Theta. \tag{13.7.2}$$

**13.8. Nice conjugation invariant functions on $\mathfrak{g}$.** To state the next result we need the notion of *nice conjugation invariant function* on $\mathfrak{g}$. By this we mean a conjugation invariant function $F$ that is defined and locally constant on $\mathfrak{g}_{\mathrm{rs}}$ and is locally integrable on $\mathfrak{g}$ (after extending it from $\mathfrak{g}_{\mathrm{rs}}$ to $\mathfrak{g}$, say by 0).

The local integrability (a notion reviewed in 10.4) of $F$ guarantees that we get a well-defined distribution

$$f \mapsto \int_{\mathfrak{g}} f(X)F(X)\, dX \tag{13.8.1}$$

on $\mathfrak{g}$, and the conjugation invariance of $F$ implies that this distribution is invariant. We say that $F$ *represents* this distribution. Since nice functions are required to be locally constant on $\mathfrak{g}_{\mathrm{rs}}$, a set whose complement has measure 0, there is at most one nice conjugation invariant function representing a given invariant distribution.

**13.9. Parabolic induction of nice invariant distributions.** Now we return to the parabolic subalgebra $\mathfrak{p} = \mathfrak{m} \oplus \mathfrak{u}$ of Lie algebra $\mathfrak{g}$. Suppose that $F_M$ is a nice conjugation invariant function on $\mathfrak{m}$, and let $T_M$ be the invariant distribution on $\mathfrak{m}$ that it represents.

LEMMA 13.2. *The parabolically induced distribution $i_P^G(T_M)$ on $\mathfrak{g}$ is represented by the nice function*

$$|D^G(X)|^{-1/2} \sum_Y |D^M(Y)|^{1/2} F_M(Y) \tag{13.9.1}$$

*on $\mathfrak{g}_{\mathrm{rs}}$. Here the sum is taken over a set of representatives for the $M$-conjugacy classes of elements $Y \in \mathfrak{m}$ such that $Y$ is $G$-conjugate to $X$. The superscripts on*

*$D$ are used to distinguish between the functions previously denoted by $D(X)$ on $\mathfrak{g}$ and $\mathfrak{m}$.*

PROOF. Use the Weyl integration formula. □

For example suppose that $G$ is a split group with split maximal torus $A$ and Borel subgroup $B$ containing $A$. Let $\chi$ be a quasi-character on $A$, that is, a continuous homomorphism $A \to \mathbb{C}^\times$. Then we can parabolically induce $\chi$, obtaining a principal series representation (possibly reducible, though often irreducible) of $G$, whose character is parabolically induced from $\chi$.

What is the analogous situation on the Lie algebra? The analog of $\chi$ is a continuous homomorphism $\xi : \mathrm{Lie}(A) \to \mathbb{C}^\times$, and it represents a distribution (still call it $\xi$) on $\mathrm{Lie}(A)$ that we can parabolically induce to $\mathfrak{g}$, obtaining an invariant distribution $i_B^G(\xi)$ on $\mathfrak{g}$, which by Lemma 13.2 is represented by the nice function on $\mathfrak{g}_{\mathrm{rs}}$ which vanishes on elements not conjugate to something in $\mathrm{Lie}(A)$ and is given by

$$|D^G(X)|^{-1/2} \sum_{w \in W} \xi(w(X)) \tag{13.9.2}$$

for $X \in \mathrm{Lie}(A)_{\mathrm{reg}}$ (with $W$ denoting the Weyl group of $A$). This formula is completely analogous to the one for the character of a principal series representation, showing that $i_B^G(\xi)$ should be viewed as the Lie algebra analog of the character of a principal series representation.

**13.10. The $MUK$-integration formula.** Let $P = MU$ and $K$ be as before (so that $G = PK = MUK$). We have the integration formula (for $f \in C_c^\infty(G)$)

$$\int_G f(g)\, dg = \int_P \int_K f(pk)\, dk\, dp \tag{13.10.1}$$

for suitably normalized Haar measures $dg$, $dp$, $dk$ on $G$, $P$, $K$ respectively, $dp$ being a left Haar measure.

This formula is not difficult to prove, the main point being that $G$ can be regarded as a homogeneous space for the group $P \times K$ via the action

$$(p, k) \cdot g = pgk^{-1},$$

so that there is a unique (up to a positive constant) measure on $G$ that is left $P$-invariant and right $K$-invariant. Both the left and right sides of the equality (13.10.1) provide such measures on $G$.

Moreover, since $P$ is the semidirect product of $M$ and the normal subgroup $U$, there is another integration formula (for $f \in C_c^\infty(P)$)

$$\int_P f(p)\, dp = \int_M \int_U f(mu)\, du\, dm \tag{13.10.2}$$

for suitable Haar measures $dm$, $du$ on the (unimodular) groups $M$, $U$ respectively. Again one can use that $M \times U$ acts transitively on $P$ via $(m, u) \cdot p = mpu^{-1}$. Note that the order of multiplication matters; if we used $f(um)$ rather than $f(mu)$, we would get a right Haar measure on $P$. Alternatively, (13.10.2) is an instance of integration in stages, but in a more general case than we considered before, since $P$ is not unimodular.

Combining the two integration formulas, we get the $MUK$-integration formula (for $f \in C_c^\infty(G)$)

$$\int_G f(g)\,dg = \int_M \int_U \int_K f(muk)\,dk\,du\,dm. \tag{13.10.3}$$

There is also a useful variant involving a unimodular closed subgroup $H$ of $M$. For any $f' \in C_c^\infty(H\backslash G)$ we then have

$$\int_{H\backslash G} f'(g)\,dg/dh = \int_{H\backslash M} \int_U \int_K f'(muk)\,dk\,du\,dm/dh. \tag{13.10.4}$$

The variant can be derived from the $MUK$-integration formula, using that any $f' \in C_c^\infty(H\backslash G)$ can be obtained from some $f \in C_c^\infty(G)$ as

$$f'(g) = \int_H f(hg)\,dh. \tag{13.10.5}$$

In other words $f' = f^\sharp$ in the notation of 2.4.

**13.11. Parabolic descent for orbital integrals.** For $X \in \mathfrak{g}$ we denote by $\mathfrak{g}_X$ the centralizer of $X$ in $\mathfrak{g}$, or, in other words, the kernel of $\mathrm{ad}(X)$. Then (since we are working in characteristic 0) $\mathfrak{g}_X$ is the Lie algebra of the centralizer $G_X$ of $X$ in $G$.

In this subsection we will need to use the parabolic subgroup $\bar{P} = M\bar{U}$ opposite to $P = MU$ (with Lie algebra $\bar{\mathfrak{p}} = \mathfrak{m} \oplus \bar{\mathfrak{u}}$). We then have the $\mathrm{Ad}(M)$-stable decomposition

$$\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{u} \oplus \bar{\mathfrak{u}}. \tag{13.11.1}$$

Note that as $M$-module $\bar{\mathfrak{u}}$ is contragredient to $\mathfrak{u}$ (via the Killing form on $\mathfrak{g}$).

Now let $X \in \mathfrak{m}$, and let $X_s \in \mathfrak{m}$ be the semisimple part of (the Jordan decomposition of) $X$. Put $D_M^G(X) := \det(\mathrm{ad}(X); \mathfrak{g}/\mathfrak{m})$. Since $\mathrm{ad}(X)$ preserves the decomposition (13.11.1), we see that $\mathfrak{g}_X \subset \mathfrak{m}$ if and only if $D_M^G(X) \neq 0$. Since $D_M^G(X) = D_M^G(X_s)$, we see also that $\mathfrak{g}_X \subset \mathfrak{m}$ if and only if $\mathfrak{g}_{X_s} \subset \mathfrak{m}$. (To put this in context we should recall that $\mathfrak{g}_X \subset \mathfrak{g}_{X_s}$.)

Now assume that $X \in \mathfrak{m}$ does satisfy the condition $D_M^G(X) \neq 0$. Then $X_s$ satisfies the same condition, so that $\mathfrak{g}_{X_s} \subset \mathfrak{m}$. Since $G_{X_s}$ is connected (see [**Ste75**, Cor. 3.11]), we see that $G_{X_s} \subset M$, and since $G_X \subset G_{X_s}$, we conclude that $G_X = M_X$. Choose a Haar measure $dh$ on $H := G_X$. As Haar measure $dZ$ on $\mathfrak{u}$ (the one we used in 13.1 to define $f^P$) we now take the one compatible with the Haar measure $du$ used in the $MUK$-integration formula.

What does compatible mean? For any algebraic group $G$ over our $p$-adic field there is a notion of compatibility of Haar measures on $G$ and $\mathfrak{g}$. This is because Haar measures can be obtained from invariant volume forms, and we can agree that a (left, say) invariant volume form $\omega$ on $G$ is compatible with a translation invariant volume form $\omega'$ on $\mathfrak{g}$ if the value of $\omega$ at $1 \in G$ is equal to the value of $\omega'$ at $0 \in \mathfrak{g}$. (It makes sense to compare the two values since the tangent space in each case is $\mathfrak{g}$.)

LEMMA 13.3. *Assume that $X \in \mathfrak{m}$ satisfies $D_M^G(X) \neq 0$ and put $H := G_X^0$. Let $v$ be any complex-valued function on $G$ that is right invariant under $K$ and left invariant under both $U$ and $H$. Then for all $f \in C_c^\infty(\mathfrak{g})$ there is an equality*

$$|D_M^G(X)|^{1/2} \int_{H\backslash G} f(g^{-1}Xg)v(g)\,d\dot{g} = \int_{H\backslash M} f^{(P)}(m^{-1}Xm)v(m)\,d\dot{m}, \tag{13.11.2}$$

*provided the two integrals converge. Here $d\dot g = dg/dh$ and $d\dot m = dm/dh$. In case $X$ is regular semisimple in $\mathfrak{g}$, say with centralizer $T$, the integrals do converge, and we have*

$$D_M^G(X) = D^G(X)/D^M(X), \tag{13.11.3}$$

*so that the equality above can be rewritten as*

$$|D^G(X)|^{1/2} \int_{T\backslash G} f(g^{-1}Xg)v(g)\,d\dot g = |D^M(X)|^{1/2} \int_{T\backslash M} f^{(P)}(m^{-1}Xm)v(m)\,d\dot m.$$

PROOF. Applying the variant $MUK$-integration formula (13.10.4), we see that we need to compare integrals over the sets $\{u^{-1}Yu : u \in U\}$ and $Y + \mathfrak{u}$, the first using $du$, the second $dZ$, where $Y$ is any $M$-conjugate of $X$. In fact it is enough to show that $U \to (Y + \mathfrak{u}) \cong \mathfrak{u}$ defined by $u \mapsto u^{-1}Yu$ is an isomorphism of algebraic varieties whose Jacobian (with respect to compatible left-invariant volume forms on $U$ and $\mathfrak{u}$) is the non-zero constant $\det(\mathrm{ad}(X); \mathfrak{u})$. (Here one needs to use that $\bar{\mathfrak{u}}$ is contragredient to $\mathfrak{u}$ as $M$-module, so that $D_M^G(X) = (-1)^{\dim(U)} \det(\mathrm{ad}(X); \mathfrak{u})^2$.)

By a theorem of Rosenlicht [**Ros61**] the $\mathrm{Ad}(U)$-orbit of $Y$ in $Y + \mathfrak{u}$ is closed. Since the centralizer of $Y$ is contained in $M$, it intersects $U$ trivially, showing that $u \mapsto u^{-1}Yu$ identifies $U$ with a locally closed subset of $Y + \mathfrak{u}$ having the same dimension as $\mathfrak{u}$ and hence (by closedness) coinciding with $Y + \mathfrak{u}$. Thus our map is a bijective morphism $U \to Y + \mathfrak{u}$, and it remains only to compute its Jacobian.

Identifying all relevant tangent spaces with $\mathfrak{u}$ in the obvious way (using left translations in the case of $U$), we see that the differential of our morphism at the point $u \in U$ is equal to $\mathrm{ad}(Y^u) : \mathfrak{u} \to \mathfrak{u}$, where $Y^u := u^{-1}Yu$. Since $Y^u, X$ are $P$-conjugate, we see that the determinant of the differential is $\det(\mathrm{ad}(X); \mathfrak{u})$, as desired.

Since semisimple orbits are closed, it is clear that the integrals converge for semisimple $X$. When $X$ is not semisimple, convergence will depend on the weight function $v$. As we have made no assumption on the growth rate of $v$, we cannot be sure the integrals converge. □

**13.12. Parabolic induction for orbital integrals.** We continue with the discussion in the last subsection. In particular $X$ still denotes an element of $\mathfrak{m}$ such that $D_M^G(X) \neq 0$. Taking $v = 1$ in Lemma 13.3, we see that

$$i_P^G(O_X^M) = |D_M^G(X)|^{1/2} O_X^G, \tag{13.12.1}$$

where the superscripts on $O_X^G$ and $O_X^M$ are used to distinguish between orbital integrals on $G$ and $M$.

For regular semisimple $X \in \mathfrak{g}$ it is sometimes more convenient to use the *normalized orbital integral* $I_X = I_X^G$ defined by

$$I_X = |D(X)|^{1/2} O_X. \tag{13.12.2}$$

For $X \in \mathfrak{m}$ such that $X$ is regular semisimple in $\mathfrak{g}$ (and hence in $\mathfrak{m}$ as well), Lemma 13.3 yields the especially simple formula

$$i_P^G(I_X^M) = I_X^G. \tag{13.12.3}$$

**13.13. Fourier transform commutes with parabolic induction.** Let $V$ be a finite dimensional vector space over a $p$-adic field, and let $V^*$ be the dual vector space. Then, fixing a non-trivial additive character $\psi$ on $F$, the Fourier transform $f \mapsto \hat{f}$ from $C_c^\infty(V) \to C_c^\infty(V^*)$ is defined by

$$\hat{f}(v^*) = \int_V f(v)\psi(\langle v^*, v\rangle)\, dv; \tag{13.13.1}$$

it depends on the choice of Haar measure $dv$ on $V$. For any linear subspace $W \subset V$ we denote by $W^\perp$ the subspace of $V^*$ consisting of all elements $v^* \in V^*$ such that $\langle v^*, w\rangle = 0$ for all $w \in W$. Now suppose that we have two nested subspaces of $V$, say $V \supset V_1 \supset V_2$. Dually we have nested subspaces $V^* \supset V_2^\perp \supset V_1^\perp$ and a canonical identification $(V_1/V_2)^* \cong V_2^\perp/V_1^\perp$. It is easy to check that the following diagram commutes

$$\begin{array}{ccc} C_c^\infty(V) & \xrightarrow{FT} & C_c^\infty(V^*) \\ \downarrow & & \downarrow \\ C_c^\infty(V_1/V_2) & \xrightarrow{FT} & C_c^\infty(V_2^\perp/V_1^\perp) \end{array}$$

where the horizontal arrows are Fourier transforms, the left vertical arrow is given by restriction to $V_1$ and integration over the cosets of $V_2$, and the right vertical arrow is given by restriction to $V_2^\perp$ and integration over the cosets of $V_1^\perp$. Compatible Haar measures are needed: use dual Haar measures on dual vector spaces and build up the Haar measure on $V$ from Haar measures on $V/V_1$, $V_1/V_2$, and $V_2$.

Now return to $\mathfrak{g}$ and consider the nested subspaces $\mathfrak{g} \supset \mathfrak{p} \supset \mathfrak{u}$. We have agreed (in 8.2) to identify $\mathfrak{g}$ with its dual $\mathfrak{g}^*$ using some fixed $G$-invariant non-degenerate symmetric bilinear form $B(\cdot,\cdot)$ on $\mathfrak{g}$. With this identification we have $\mathfrak{u}^\perp = \mathfrak{p}$ and $\mathfrak{p}^\perp = \mathfrak{u}$. Therefore the following diagram commutes

$$\begin{array}{ccc} C_c^\infty(\mathfrak{g}) & \xrightarrow{FT} & C_c^\infty(\mathfrak{g}) \\ \downarrow & & \downarrow \\ C_c^\infty(\mathfrak{m}) & \xrightarrow{FT} & C_c^\infty(\mathfrak{m}) \end{array} \tag{13.13.2}$$

where the horizontal maps are again Fourier transforms and the two vertical maps are both equal to the map $f \mapsto f^P$ defined in 13.1.

LEMMA 13.4. *The map $f \mapsto f^{(P)}$ commutes with the Fourier transform, or, in other words, the commutative diagram above continues to commute when the vertical arrows are replaced by the map $f \mapsto f^{(P)}$. Parabolic induction $i_P^G$ of invariant distributions also commutes with the Fourier transform, or, in other words, for any invariant distribution $T_M$ on $\mathfrak{m}$ we have*

$$i_P^G(\hat{T}_M) = \widehat{i_P^G(T_M)}. \tag{13.13.3}$$

PROOF. Recall that $f^{(P)} = (\tilde{f})^P$. We have just shown (see (13.13.2)) that $f \mapsto f^P$ commutes with the Fourier transform. It is clear from the definition that $f \mapsto \tilde{f}$ commutes with the Fourier transform. Therefore $f \mapsto f^{(P)}$ commutes with the Fourier transform. Since $i_P^G$ is dual to $f \mapsto f^{(P)}$, it too commutes with the Fourier transform. $\square$

**13.14. Justification of a statement made earlier.** Now we are in a position to prove a statement we needed in 10.4. Recall that $\mathfrak{g}_e$ is the (open) subset of elliptic regular semisimple elements in $\mathfrak{g}$.

LEMMA 13.5. *Assume that $P \neq G$ and that $\phi \in C_c^\infty(\mathfrak{g}_e)$. Then $\phi^{(P)}$ and $(\hat{\phi})^{(P)}$ are identically zero. Moreover, for any $X \in \mathfrak{m}$ such that $D_M^G(X) \neq 0$ the integral*

$$\int_{H\backslash G} \hat{\phi}(g^{-1}Xg)\, d\dot{g} \tag{13.14.1}$$

*vanishes.*

PROOF. Clearly $\tilde{\phi}$ also vanishes off $\mathfrak{g}_e$, and since $\mathfrak{p}$ does not meet $\mathfrak{g}_e$, the function $\phi^{(P)}$ is identically zero. Since $f \mapsto f^{(P)}$ commutes with the Fourier transform, the function $(\hat{\phi})^{(P)}$ is also identically zero. By Lemma 13.3 the integral (13.14.1) vanishes. □

In 10.4 we needed the special case of this lemma in which $G$ is $GL_2$ and $P = AN$ a Borel subgroup, in order to obtain the vanishing of the hyperbolic orbital integrals of the function $\hat{\phi}$ considered in that subsection.

## 14. The map $\pi_G : \mathfrak{g} \to \mathbb{A}_G$ and the geometry behind semisimple descent

We have just discussed parabolic descent. When we begin our systematic treatment of Shalika germs in section 17, we will need semisimple descent, to be discussed in section 16. The purpose of this section and the next is to provide the necessary preparation for semisimple descent, beginning with Chevalley's restriction theorem.

We work with a connected reductive group $G$ over an algebraically closed field $k$ of characteristic 0. Let $T$ be a maximal torus in $G$. We of course write $\mathfrak{g}$ and $\mathfrak{t}$ for the Lie algebras of $G$ and $T$. We denote by $W$ the Weyl group of $T$.

**14.1. Basic polynomial invariants.** The ring of polynomial functions on $\mathfrak{t}$ is the symmetric algebra $S =: S(\mathfrak{t}^*)$. The Weyl group $W$ acts on $S$, and the ring $S^W$ of invariants is the ring of regular functions on the quotient variety $\mathfrak{t}/W$. Both $S$ and $S^W$ are graded (by polynomial degree). Inside $S^W$ we have the (graded) maximal ideal $I$ comprised of all invariant polynomials with constant term 0. The quotient $I/I^2$ is a graded vector space. Choose a basis of homogeneous elements in $I/I^2$ and lift them to homogeneous elements $x_1, \dots, x_n$ in $I$. Then $n = \dim(\mathfrak{t})$ and $S^W$ is isomorphic to the polynomial ring $k[x_1, \dots, x_n]$ (see [**Bou02**, Ch. 5, §5]). Our choice of $x_1, \dots, x_n$ gives us an isomorphism of varieties from $\mathfrak{t}/W$ to the standard affine $n$-space $\mathbb{A}^n$. We will sometimes denote $\mathfrak{t}/W$ by $\mathbb{A}_G$. (The notation $\mathbb{A}_G$ is supposed to remind us that we are dealing with an affine space.)

On any affine space $\mathbb{A}^m$ there is an essentially unique volume form (nowhere-vanishing differential form of top degree), unique, that is, up to an element in $k^\times$. Indeed, it is clear that such forms exist, and uniqueness follows from the fact that there are no units in polynomial rings other than constants. Both $\mathfrak{t}$ and $\mathfrak{t}/W$ are affine spaces of dimension $n = \dim(T)$. Pick volume forms on both affine spaces and look at the Jacobian of the canonical surjection $\mathfrak{t} \to \mathfrak{t}/W$. Up to an element of $k^\times$ it is independent of the choice of volume forms. The Jacobian turns out to be

$$\prod_{\alpha \in R_G^+} \alpha \tag{14.1.1}$$

(see [**Bou02**, Ch. 5, §5.5, Prop. 6]), up to an element of $k^\times$. Here $R_G \subset X^*(T)$ is the set of roots of $T$ in $\mathfrak{g}$, and $R_G^+$ is the subset of positive roots (positive with respect to some Borel subgroup $B$ containing $T$). As before, the differentials of the roots allow us to view them as linear forms on $\mathfrak{t}$, hence as elements in $S$.

**14.2. Chevalley's restriction theorem.** We write $\mathcal{O}_\mathfrak{g}$ for the $k$-algebra of polynomial functions on $\mathfrak{g}$. Inside $\mathcal{O}_\mathfrak{g}$ we have the subalgebra $(\mathcal{O}_\mathfrak{g})^G$ of conjugation invariant polynomial functions on $\mathfrak{g}$. Chevalley's theorem [**SS70**, 3.17] states that by restricting polynomial functions from $\mathfrak{g}$ to $\mathfrak{t}$ we get an isomorphism from $(\mathcal{O}_\mathfrak{g})^G$ to the algebra $S^W$ of $W$-invariant polynomial functions on $\mathfrak{t}$. Thus $\mathrm{Spec}(\mathcal{O}_\mathfrak{g})^G \cong \mathfrak{t}/W = \mathbb{A}_G$.

Dual to the inclusion of $(\mathcal{O}_\mathfrak{g})^G$ in $\mathcal{O}_\mathfrak{g}$ is a surjective morphism

$$\pi_G : \mathfrak{g} \to \mathbb{A}_G = \mathfrak{t}/W \tag{14.2.1}$$

which maps $X \in \mathfrak{g}$ to the unique $W$-orbit in $\mathfrak{t}$ consisting of elements conjugate to the semisimple part of $X$. Therefore $\pi_G(X) = \pi_G(Y)$ if and only if the semisimple parts of $X$, $Y$ are $G$-conjugate. Moreover, the nilpotent cone in $\mathfrak{g}$ equals $\pi_G^{-1}(0)$, where 0 denotes the origin in $\mathbb{A}_G$.

More concretely, we can also view $\pi_G$ as the map $X \mapsto (P_1(X), \dots, P_n(X))$ from $\mathfrak{g}$ to $\mathbb{A}^n$, where $P_i$ is the homogeneous $G$-invariant polynomial on $\mathfrak{g}$ corresponding to the element $x_i$ from before. Letting $d_i$ denote the degree of $P_i$, we define an action of $\mathbb{G}_m$ on $\mathbb{A}^n$ by $\beta \cdot (z_1, \dots, z_n) := (\beta^{d_1} z_1, \beta^{d_2} z_2, \dots, \beta^{d_n} z_n)$ for all $\beta \in \mathbb{G}_m$ and all $(z_1, \dots, z_n) \in \mathbb{A}^n$. For this $\mathbb{G}_m$ action on $\mathbb{A}_G = \mathbb{A}^n$ we have

$$\pi_G(\beta X) = \beta \cdot \pi_G(X). \tag{14.2.2}$$

**14.3. Non-algebraically closed fields.** When the base field $k$ is not algebraically closed, it is better to define $\mathbb{A}_G$ as $\mathrm{Spec}((\mathcal{O}_\mathfrak{g})^G)$, as this avoids having to choose a maximal $k$-torus. Note that $(\mathcal{O}_\mathfrak{g})^G$ is still a polynomial ring over $k$ for which we may choose homogeneous generators $P_1, \dots, P_n$. The morphism $\pi_G$ is defined over $k$ and $\mathbb{G}_m$-equivariant.

**14.4. The subgroup $H$.** Now we come to the geometry behind semisimple descent. For this we consider a connected reductive subgroup $H$ of $G$ such that $T \subset H$. For example $H$ might be a Levi subgroup or the centralizer of a semisimple element in $\mathfrak{g}$ (see [**Bor91**, Prop. 13.19] for the fact that such a centralizer is reductive and see [**Ste75**, Cor. 3.11] for the fact that it is connected when the characteristic of the ground field is 0). Taking Lie algebras of the three groups, we have inclusions

$$\mathfrak{t} \subset \mathfrak{h} \subset \mathfrak{g}. \tag{14.4.1}$$

The normalizer of $T$ in $H$ is a subgroup of the normalizer of $T$ in $G$, so the Weyl group $W_H$ of $T$ in $H$ is a subgroup of the Weyl group $W_G$ of $T$ in $G$.

**14.5. Jacobian of $\rho : \mathbb{A}_H \to \mathbb{A}_G$ and definition of $\mathbb{A}'_H$ and $\mathfrak{h}'$.** Using (14.1.1), we see that the Jacobian of the natural finite morphism

$$\rho : \mathfrak{t}/W_H \twoheadrightarrow \mathfrak{t}/W_G$$

is (up to an element in $k^\times$) equal to

$$\prod_{\alpha \in R_G^+ \setminus R_H^+} \alpha. \tag{14.5.1}$$

Let $\mathfrak{t}'$ be the open $W_H$-stable subvariety of $\mathfrak{t}$ consisting of $X \in \mathfrak{t}$ such that $\alpha(X) \neq 0$ for all $\alpha \in R_G \backslash R_H$, and let $\mathbb{A}'_H$ be the open subvariety of $\mathbb{A}_H$ obtained as the image of $\mathfrak{t}'$ under $\mathfrak{t} \twoheadrightarrow \mathfrak{t}/W_H$. The explicit formula (14.5.1) for the Jacobian of $\rho$ shows that $\mathbb{A}'_H$ is precisely the set of points where the finite morphism $\rho : \mathbb{A}_H \twoheadrightarrow \mathbb{A}_G$ is étale.

Define an open subvariety $\mathfrak{h}'$ in $\mathfrak{h}$ by

$$\begin{aligned} \mathfrak{h}' :&= \{X \in \mathfrak{h} : \det(\mathrm{ad}(X); \mathfrak{g}/\mathfrak{h}) \neq 0\} \\ &= \pi_H^{-1}(\mathbb{A}'_H). \end{aligned} \tag{14.5.2}$$

For any $X \in \mathfrak{h}'$ we have $\mathfrak{g}_{X_s} \subset \mathfrak{h}$, so that $G_{X_s} \subset H$ (by the connectedness of $G_{X_s}$). Since $G_X \subset G_{X_s}$, we also have $G_X \subset H$, so that $H_X = G_X$.

**14.6. The morphism $\beta$.** The morphism $G \times \mathfrak{h}' \to \mathfrak{g}$ defined by $(g, X) \mapsto gXg^{-1}$ is constant on orbits of the right $H$-action on $G \times \mathfrak{h}'$ given by $(g, X) \cdot h = (gh, h^{-1}Xh)$, and therefore descends to a morphism

$$\beta : G \underset{H}{\times} \mathfrak{h}' \to \mathfrak{g} \tag{14.6.1}$$

from the quotient space for this $H$-action to $\mathfrak{g}$. Clearly $\beta$ is $G$-equivariant for the left translation action of $G$ on the first factor in the source and the adjoint action of $G$ on the target.

We claim that $\beta$ is étale. Indeed, by $G$-equivariance it is enough to prove that the differential $d\beta$ is an isomorphism at $(1, X)$ for any $X \in \mathfrak{h}'$. Since $d\beta$ is given by

$$\begin{gathered} \mathfrak{g} \underset{\mathfrak{h}}{\times} \mathfrak{h} \to \mathfrak{g} \\ (\Delta g, \Delta X) \mapsto [\Delta g, X] + \Delta X \end{gathered} \tag{14.6.2}$$

we see that $d\beta$ is surjective (because $\det(\mathrm{ad}(X); \mathfrak{g}/\mathfrak{h}) \neq 0$) and hence an isomorphism (look at dimensions).

**14.7. Factorization of the morphism $\beta$.** The morphism $\beta$ is étale, so general theory says that it can be factorized as an open immersion followed by a finite morphism. In this case there is an obvious way to produce such a factorization. The étale morphism $\mathbb{A}'_H \to \mathbb{A}_G$ factors as

$$\mathbb{A}'_H \xrightarrow{j} \mathbb{A}_H \xrightarrow{\rho} \mathbb{A}_G$$

with $j$ an open immersion and $\rho$ finite. We now perform a base-change, using the morphism $\pi_G : \mathfrak{g} \to \mathbb{A}_G$. We obtain morphisms

$$\mathfrak{g} \times_{\mathbb{A}_G} \mathbb{A}'_H \hookrightarrow \mathfrak{g} \times_{\mathbb{A}_G} \mathbb{A}_H \to \mathfrak{g}$$

with the left arrow an open immersion and the right arrow a finite morphism. We will have the desired factorization of $\beta$ once we identify $G \underset{H}{\times} \mathfrak{h}'$ with $\mathfrak{g} \times_{\mathbb{A}_G} \mathbb{A}'_H$ over $\mathfrak{g}$.

For this purpose we define a morphism

$$\gamma : G \underset{H}{\times} \mathfrak{h}' \to \mathfrak{g} \times_{\mathbb{A}_G} \mathbb{A}'_H$$

over $\mathfrak{g}$ by putting $\gamma(g, X) := (gXg^{-1}, \pi_H(X))$. It then remains to prove the following lemma.

LEMMA 14.1. *The morphism $\gamma$ is an isomorphism. In particular*

$$(g, X) \mapsto (gXg^{-1}, \pi_H(X))$$

*defines a closed immersion*

$$G \underset{H}{\times} \mathfrak{h}' \to \mathfrak{g} \times \mathbb{A}'_H. \tag{14.7.1}$$

PROOF. We have seen that $G \underset{H}{\times} \mathfrak{h}'$ is étale over $\mathfrak{g}$ (via $\beta$). Moreover $\mathfrak{g} \times_{\mathbb{A}_G} \mathbb{A}'_H$ is étale over $\mathfrak{g}$ (since $\rho$ is étale on the open subset $\mathbb{A}'_H$ of $\mathbb{A}_H$). Therefore $\gamma$ is automatically étale, and it is enough to show that it is bijective on $k$-points.

First we check surjectivity. Thus we start with a pair $(X, Y) \in \mathfrak{g} \times_{\mathbb{A}_G} \mathbb{A}'_H$, where $Y \in \mathfrak{t}'$ represents an element in $\mathfrak{t}'/W_H = \mathbb{A}'_H$, and we want to show that this pair is in the image of $\gamma$. Since $X$, $Y$ become the same in $\mathbb{A}_G$, we see that the semisimple part $X_s$ of $X$ is $G$-conjugate to $Y$, and by the obvious $G$-equivariance of $\gamma$ we may assume without loss of generality that $X_s = Y$. Since $Y \in \mathfrak{h}'$, we have $\mathfrak{g}_Y \subset \mathfrak{h}$, and since $X$ commutes with its semisimple part, namely $Y$, we see that $X \in \mathfrak{h}$. In fact $X \in \mathfrak{h}'$, since an element of $\mathfrak{h}$ lies in $\mathfrak{h}'$ if and only if its semisimple part does. The element $(1, X) \in G \underset{H}{\times} \mathfrak{h}'$ maps to $(X, Y)$ under $\gamma$.

Now we check injectivity. Again using the $G$-equivariance of $\gamma$, we see that it is enough to prove that if

$$\gamma(1, X) = \gamma(g, X')$$

then $g \in H$ and $gX'g^{-1} = X$. The second condition is obvious. It follows that $gX'_sg^{-1} = X_s$. Since $X, X' \in \mathfrak{h}'$ become the same in $\mathbb{A}_H$, there exists $h \in H$ such that $hX'_sh^{-1} = X_s$. Therefore $hg^{-1} \in G_{X_s}$. Since $X_s \in \mathfrak{h}'$, we have $G_{X_s} \subset H$, as we noted before. Therefore $hg^{-1} \in H$, proving that $g \in H$, as desired. □

## 15. Harish-Chandra's compactness lemma; boundedness modulo conjugation

Before moving on to semisimple descent, we discuss some related matters that make use of the map $\pi_G : \mathfrak{g} \to \mathbb{A}_G$ of section 14. In this section $F$ is a local field of characteristic 0.

**15.1. Harish-Chandra's compactness lemma.** We return to the situation in 14.4, but now we suppose that $H \subset G$ are defined over $F$. Then we have the following slight generalization of Harish-Chandra's compactness lemma [**HC70**, Lemma 25]. It will be used for semisimple descent.

LEMMA 15.1 (Harish-Chandra). *Let $\omega_{\mathfrak{g}}$ be a compact subset of $\mathfrak{g}$ and let $\omega_H$ be a compact subset of $\mathbb{A}'_H(F)$. Then*

$$\{g \in G(F)/H(F) : \exists X \in \pi_H^{-1}(\omega_H) \text{ such that } gXg^{-1} \in \omega_{\mathfrak{g}}\} \tag{15.1.1}$$

*has compact closure in $G(F)/H(F)$.*

PROOF. Put $Z := G \underset{H}{\times} \mathfrak{h}'$. On $F$-points the closed immersion (14.7.1) yields a closed embedding $Z(F) \hookrightarrow \mathfrak{g} \times \mathbb{A}'_H(F)$ that we will denote by $i$. Projection onto the first factor of $G \times \mathfrak{h}'$ induces a morphism $Z(F) \to (G/H)(F)$ that we will denote by $\eta$. Note that $G(F)/H(F)$ sits inside $(G/H)(F)$ as an open and closed subset. The set (15.1.1) is contained in the compact set $\eta i^{-1}(\omega_{\mathfrak{g}} \times \omega_H)$. □

**15.2. Boundedness modulo conjugation.** Boundedness modulo conjugation came up already in 10.4. We now discuss this notion more systematically. We say that a subset of $\mathbb{A}_G(F)$ is *bounded* if it is contained in some compact subset of $\mathbb{A}_G(F)$. We say that a subset $V$ of $\mathfrak{g}$ is *bounded modulo conjugation* if there exists a compact subset $C$ of $\mathfrak{g}$ such that every element in $V$ is $G$-conjugate to an element in $C$ (in other words $V \subset \mathrm{Ad}(G)(C)$).

LEMMA 15.2. *Let $V$ be a subset of $\mathfrak{g}$. The following three conditions are equivalent.*

(1) *The set $V$ is bounded modulo conjugation.*
(2) *There exists a compact subset $C$ in $\mathfrak{g}$ such that $V$ is contained in the closure of $\mathrm{Ad}(G)(C)$.*
(3) *The image of $V$ under $\pi_G : \mathfrak{g} \to \mathbb{A}_G(F)$ is bounded in $\mathbb{A}_G(F)$.*

PROOF. The implications (1) $\Rightarrow$ (2) $\Rightarrow$ (3) are immediate. It remains to check that (3) $\Rightarrow$ (1). Let $\omega$ be a compact subset of $\mathbb{A}_G(F)$. It is enough to show that $\pi_G^{-1}\omega$ is bounded modulo conjugation.

Let $T$ be a maximal torus in $G$. We say that a linear subspace $\mathfrak{s}$ of $\mathfrak{t}$ is *special* if it arises as the center of the centralizer of some element in $\mathfrak{t}$. Thus $\mathfrak{s}$ is the intersection of the kernels of some subset of the roots of $\mathfrak{t}$; it follows that $\mathfrak{t}$ has only finitely many special subspaces. For each special subspace $\mathfrak{s}$ let $M_{\mathfrak{s}}$ denote the centralizer in $G$ of $\mathfrak{s}$ and let $\mathfrak{m}_{\mathfrak{s}}$ denote the Lie algebra of $M_{\mathfrak{s}}$. Then $\mathfrak{s}$ coincides with the center of $\mathfrak{m}_{\mathfrak{s}}$. Let $\mathcal{N}_{\mathfrak{s}}$ be a set of representatives for the $M_{\mathfrak{s}}$-conjugacy classes of nilpotent elements in $\mathfrak{m}_{\mathfrak{s}}$. Now put

$$C := \bigcup_T \bigcup_{\mathfrak{s}} \bigcup_Y \big((\mathfrak{s} \cap \pi_G^{-1}\omega) + Y\big),$$

where $T$ runs over a set of representatives for the $G$-conjugacy classes of maximal $F$-tori in $G$, $\mathfrak{s}$ runs over the set of special subspaces of $\mathfrak{t} = \mathrm{Lie}(T)$, and $Y$ runs over $\mathcal{N}_{\mathfrak{s}}$. Each set $(\mathfrak{s} \cap \pi_G^{-1}\omega) + Y$ is compact (since the maps $\mathfrak{s} \hookrightarrow \mathfrak{t} \to \mathbb{A}_G(F)$ are proper) and the union is finite, so the set $C$ is compact. It is clear that $\pi_G^{-1}\omega \subset \mathrm{Ad}(G)(C)$. $\square$

It follows from the lemma that $V$ is bounded modulo conjugation if and only if its closure is.

LEMMA 15.3. *Now we work over a $p$-adic field $F$. Let $L$ be a lattice in $\mathfrak{g}$. Then $\mathrm{Ad}(G)(L)$ contains a $G$-invariant open and closed neighborhood $V$ of the nilpotent cone.*

PROOF. The nilpotent cone equals $\pi_G^{-1}(0)$, where 0 denotes the origin in $\mathbb{A}_G(F)$. Pick any compact open neighborhood $\omega$ of 0 in $\mathbb{A}_G(F)$. Then $\pi_G^{-1}(\omega)$ is bounded modulo conjugation, so by the previous lemma there exists an integer $m$ such that $\pi_G^{-1}(\omega) \subset \mathrm{Ad}(G)(\pi^{-m}L)$. Therefore $V := \pi_G^{-1}(\pi^m \cdot \omega)$ does the job. Here we are using the $\mathbb{G}_m$-action on $\mathbb{A}_G$ for which $\pi_G$ is equivariant (see (14.2.2)). $\square$

## 16. Semisimple descent

As usual $F$ is a $p$-adic field. Throughout this section we consider a connected reductive $F$-group $G$ and a connected reductive $F$-subgroup $H$ of $G$ that contains some maximal torus of $G$. Our goal is to compare orbital integrals on $G$ and $H$. As usual we write $G$, $H$ for the $F$-points of these groups. We use notation such as $\mathfrak{h}'$, $\mathbb{A}'_H$, $\pi_H$ from section 14.

**16.1. Associated functions $f$ and $\phi$.** Let $f \in C_c^\infty(\mathfrak{g})$ and let $\omega_H$ be a compact subset of $\mathbb{A}'_H(F)$. By Lemma 15.1 there exists a compact subset $C$ of $H\backslash G$ such that

$$\{g \in G(F)/H(F) : \exists X \in \pi_H^{-1}(\omega_H) \text{ such that } gXg^{-1} \in \operatorname{Supp}(f)\} \tag{16.1.1}$$

is contained in $C$, and since $H\backslash G$ is an l.c.t.d space, we may even assume that $C$ is open as well as compact. Now choose $\alpha \in C_c^\infty(G)$ such that $\alpha^\sharp = 1_C$ (see 2.4), and define $\phi \in C_c^\infty(\mathfrak{h})$ by

$$\phi(Y) := \int_G f(g^{-1}Yg)\alpha(g)\, dg. \tag{16.1.2}$$

Recall (see 14.5) that for $X \in \mathfrak{h}'$ we have $H_X = G_X$. The next result is contained implicitly in the proof of Lemma 29 in [**HC70**].

LEMMA 16.1. *For any $X \in \pi_H^{-1}(\omega_H)$ there is an equality*

$$\int_{G_X\backslash H} \phi(h^{-1}Xh)\, d\dot{h} = \int_{G_X\backslash G} f(g^{-1}Xg)\, d\dot{g}. \tag{16.1.3}$$

PROOF. Use Lemma 2.3, applied to the function $g \mapsto f(g^{-1}Xg)$ on $G_X\backslash G$. $\square$

**16.2. Comparison with parabolic descent.** In the special case when $H$ is a Levi subgroup $M$ of $G$ we have already done much better than Lemma 16.1. Indeed, for any $f \in C_c^\infty(\mathfrak{g})$ we produced a function $f^{(P)}$ on $\mathfrak{m}$ having the same orbital integrals as $f$ (up to a Jacobian factor) for all orbits in $\mathfrak{m}'$. The lemma above produces a function $\phi$ on $\mathfrak{h}$ having the same orbital integrals as $f$ for all orbits in the subset $\pi_H^{-1}(\omega_H)$ of $\mathfrak{h}'$. Of course we are free to take $\omega_H$ as large as we like, provided that it stays compact, but if we change $\omega_H$, we will have to change $\phi$ as well. Nevertheless the lemma above will be enough to prove descent for Shalika germs, as we will see later.

## 17. Basic results on Shalika germs on $\mathfrak{g}$

In this section $F$ is a $p$-adic field and $G$ is a connected reductive $F$-group. As usual we write $G$ for the group of $F$-points of $G$.

We defined Shalika germs on $G$ and calculated them for $GL_2$. Now we begin systematically studying Shalika germs on $\mathfrak{g}$. The discussion will be completed later, in section 27.

**17.1. Orbital integrals $O_X$ for regular semisimple $X$.** Let $\mathcal{T}$ be a set of representatives for the $G$-conjugacy classes of maximal $F$-tori in $G$.

Let $T \in \mathcal{T}$. In order to define our orbital integrals for $X \in \mathfrak{t}_{\text{reg}}$ in a coherent way, we fix Haar measures $dg$, $dt$ on $G$, $T$ respectively and define (for any $X \in \mathfrak{t}_{\text{reg}}$)

$$O_X(f) := \int_{T\backslash G} f(g^{-1}Xg)\, dg/dt. \tag{17.1.1}$$

The Weyl group $W_T$ (defined in subsection 7.1) acts on $T\backslash G$ by left multiplication, preserving the invariant measure on that homogeneous space, and therefore

$$O_{w(X)}(f) = O_X(f) \tag{17.1.2}$$

for all $f \in C_c^\infty(\mathfrak{g})$, all $X \in \mathfrak{t}_{\text{reg}}$ and all $w \in W_T$.

Now let $Y \in \mathfrak{g}_{\text{rs}}$. There exists unique $T \in \mathcal{T}$ such that $Y$ is $G$-conjugate to some $X \in \mathfrak{t}$, and moreover $X$ is unique up to the action of $W_T$. Therefore it is legitimate to put $O_Y := O_X$. Now we have coherent choices for all regular semisimple orbital integrals $O_Y$.

**17.2. Preliminary definition of Shalika germs on $\mathfrak{g}$.** There are finitely many nilpotent $G$-orbits $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_r$ in $\mathfrak{g}$. We write $\mu_1, \dots, \mu_r$ for the corresponding nilpotent orbital integrals. The same reasoning as for unipotent orbital integrals (see 6.4) shows that the distributions $\mu_1, \dots, \mu_r$ are linearly independent.

THEOREM 17.1. *There exist functions $\Gamma_1, \Gamma_2, \dots, \Gamma_r$ on $\mathfrak{g}_{\mathrm{rs}}$ having the following property. For every $f \in C_c^\infty(\mathfrak{g})$ there exists an open neighborhood $U_f$ of $0$ in $\mathfrak{g}$ such that*

$$O_X(f) = \sum_{i=1}^{r} \mu_i(f) \cdot \Gamma_i(X) \tag{17.2.1}$$

*for all $X \in \mathfrak{g}_{\mathrm{rs}} \cap U_f$. The germs about $0 \in \mathfrak{g}$ of the functions $\Gamma_1, \dots, \Gamma_r$ are unique. We refer to $\Gamma_i$ as the provisional Shalika germ for the nilpotent orbit $\mathcal{O}_i$.*

PROOF. This is proved exactly the same way as the analogous result (Theorem 6.1) on the group. There we worked with one $T$ at a time, but the proof provided a set $U_f$ that works for all $T$ at once. □

A Shalika germ is an equivalence class of functions on $\mathfrak{g}_{\mathrm{rs}}$. As we will see next, the homogeneity of Shalika germs makes it possible to single out one particularly nice function $\Gamma_i$ within its equivalence class. Once we have done this, $\Gamma_i$ will from then on denote this function (whose germ about 0 is the old $\Gamma_i$). First we need to understand homogeneity of nilpotent orbital integrals themselves.

**17.3. Coadjoint orbits as symplectic manifolds.** We recall Kirillov's construction of a symplectic structure on coadjoint orbits. We use $\mathrm{Ad}^*$ (respectively, $\mathrm{ad}^*$) to denote the coadjoint action of $G$ (respectively, $\mathfrak{g}$) on $\mathfrak{g}^*$.

For $\lambda \in \mathfrak{g}^*$ we let $G_\lambda$ denote the stabilizer of $\lambda$ in $G$ (for the coadjoint action); the Lie algebra of $G_\lambda$ is $\mathfrak{g}_\lambda := \{X \in \mathfrak{g} : \mathrm{ad}^*(X)\lambda = 0\}$. The tangent space at $\lambda$ to the coadjoint orbit $\mathcal{O}_\lambda$ of $\lambda$ is $\mathfrak{g}/\mathfrak{g}_\lambda$ (since $\lambda$ allows us to identify the orbit with $G/G_\lambda$).

From $\lambda$ we get an alternating form $\omega_\lambda$ on $\mathfrak{g}$, defined by

$$\omega_\lambda(X, Y) := \lambda\big([X, Y]\big) = -\langle \mathrm{ad}^*(X)\lambda, Y\rangle. \tag{17.3.1}$$

It is clear from the equality $\omega_\lambda(X, Y) = -\langle \mathrm{ad}^*(X)\lambda, Y\rangle$ that the kernel of the alternating form $\omega_\lambda$ is $\mathfrak{g}_\lambda$; therefore $\omega_\lambda$ can also be viewed as a non-degenerate alternating bilinear form on $\mathfrak{g}/\mathfrak{g}_\lambda$, or, in other words, on the tangent space to $\mathcal{O}_\lambda$ at $\lambda$.

Now fix a coadjoint orbit $\mathcal{O}$. Letting $\lambda$ vary through the orbit $\mathcal{O}$, the construction above yields a $G$-invariant 2-form $\omega$ on $\mathcal{O}$ whose value at $\lambda$ is $\omega_\lambda$.

Thus $\mathcal{O}$ is a symplectic $G$-manifold, and in particular its dimension is even, say $\dim \mathcal{O} = 2d$. The $d$-fold wedge product $\eta := \omega \wedge \cdots \wedge \omega$ is a $G$-invariant volume form on $\mathcal{O}$. The associated measure $|\eta|$ is $G$-invariant, and consequently $G_\lambda$ is unimodular.

Now suppose that multiplication by $\beta \in F^\times$ preserves the orbit $\mathcal{O}$. (This can happen only if the coadjoint orbit is nilpotent, in the sense that when we identify $\mathfrak{g}^*$ with $\mathfrak{g}$ in the usual way, the corresponding orbit in $\mathfrak{g}$ is nilpotent.) Write $m_\beta : \mathcal{O} \to \mathcal{O}$ for the map $\lambda \mapsto \beta\lambda$.

We need to compute the differential of $m_\beta$ at any point $\lambda \in \mathcal{O}$. This differential is a linear isomorphism from the tangent space $\mathfrak{g}/\mathfrak{g}_\lambda$ at $\lambda$ to the tangent space $\mathfrak{g}/\mathfrak{g}_{\beta\lambda}$ at $\beta\lambda$. Since $m_\beta$ is a $G$-map, we see that its differential $\mathfrak{g}/\mathfrak{g}_\lambda \to \mathfrak{g}/\mathfrak{g}_{\beta\lambda}$ is induced by the identity map on $\mathfrak{g}$ (note that $\mathfrak{g}_{\beta\lambda} = \mathfrak{g}_\lambda$).

Now we can compute the pull-back $m_\beta^*(\omega)$. In fact the computation we just made of the differential of $m_\beta$, together with the fact (obvious from the definition of $\omega$) that $\omega_{\beta\lambda} = \beta\omega_\lambda$, shows that $m_\beta^*(\omega) = \beta\omega$. It follows that $m_\beta^*(\eta) = \beta^d\eta$.

Once again identifying $\mathfrak{g}^*$ with $\mathfrak{g}$, we reach the conclusion that for any adjoint orbit $\mathcal{O}$ there exists a $G$-invariant volume form $\eta$ on $\mathcal{O}$, and that if multiplication by the scalar $\beta$ preserves $\mathcal{O}$, yielding a multiplication map $m_\beta : \mathcal{O} \to \mathcal{O}$, then

$$m_\beta^*(\eta) = \beta^{\dim(\mathcal{O})/2}\eta. \tag{17.3.2}$$

**17.4. Scaling of functions on $\mathfrak{g}$.** For $\beta \in F^\times$ and $f \in C_c^\infty(\mathfrak{g})$ we write $f_\beta$ for the function on $\mathfrak{g}$ defined by

$$f_\beta(X) := f(\beta X). \tag{17.4.1}$$

**17.5. Homogeneity of nilpotent orbital integrals.** Let $\mathcal{O}$ be a nilpotent orbit in $\mathfrak{g}$. Then for any $\alpha \in F^\times$ multiplication by $\alpha^2$ preserves $\mathcal{O}$. Indeed, by the Jacobson-Morozov theorem, it is enough to prove this for the group $G = SL_2$, for which the statement is an easy exercise. Let $\mu_\mathcal{O}$ denote the corresponding nilpotent orbital integral, an invariant distribution on $\mathfrak{g}$ whose homogeneity we will now establish.

LEMMA 17.2. *Let $f \in C_c^\infty(\mathfrak{g})$ and let $\alpha \in F^\times$. Then*

$$\mu_\mathcal{O}(f_{\alpha^2}) = |\alpha|^{-\dim \mathcal{O}}\mu_\mathcal{O}(f). \tag{17.5.1}$$

PROOF. This follows from (17.3.2). □

**17.6. Behavior of regular semisimple orbital integrals under scaling.** It is clear from (17.1.1) that

$$O_X(f_\beta) = O_{\beta X}(f) \tag{17.6.1}$$

for all $X \in \mathfrak{g}_{\mathrm{rs}}$ and all $\beta \in F^\times$.

**17.7. Partial homogeneity of our provisional Shalika germs $\Gamma_i$.** Let $\alpha \in F^\times$. Let $\mathcal{O}_i$ be one of our nilpotent orbits, let $\mu_i$ be the corresponding nilpotent orbital integral, and let $\Gamma_i$ be the corresponding Shalika germ. Put $d_i := \dim \mathcal{O}_i$. We claim that

$$\Gamma_i(X) = |\alpha|^{d_i}\Gamma_i(\alpha^2 X), \tag{17.7.1}$$

where the equality means equality of germs about 0 of functions on $\mathfrak{g}_{\mathrm{rs}}$.

Indeed, as in the proof of existence of Shalika germ expansions, pick a function $f_i \in C_c^\infty(\mathfrak{g})$ such that

$$\mu_j(f_i) = \delta_{ij}. \tag{17.7.2}$$

Then $\Gamma_i(X)$ is the germ about 0 of the function

$$X \mapsto O_X(f_i) \tag{17.7.3}$$

on $\mathfrak{g}_{\mathrm{rs}}$. In fact during the remainder of our discussion of provisional germs, we will use always use (17.7.3) as our choice for a specific function $\Gamma_i$ having the right germ.

In view of the homogeneity of nilpotent orbital integrals established above, the function $|\alpha|^{d_i} \cdot (f_i)_{\alpha^2}$ also satisfies (17.7.2), so that $\Gamma_i(X)$ is also the germ about 0 of the function

$$X \mapsto O_X(|\alpha|^{d_i} \cdot (f_i)_{\alpha^2}) = |\alpha|^{d_i} \cdot O_{\alpha^2 X}(f_i) \tag{17.7.4}$$

on $\mathfrak{g}_{\mathrm{rs}}$. Comparing (17.7.3), (17.7.4), we see that $\Gamma_i(X)$ and $|\alpha|^{d_i}\Gamma_i(\alpha^2 X)$ have the same germ, as desired.

**17.8. Canonical Shalika germs.** Let $\Gamma_i$ be one of our germs. Following Harish-Chandra [**HC78, HC99**], we are going to replace $\Gamma_i$ by another function $\Gamma_i^{\mathrm{new}}$ on $\mathfrak{g}_{\mathrm{rs}}$ that has the same germ about $0 \in \mathfrak{g}$ and is at the same time homogeneous. Along the way we prove a couple of simple properties of $\Gamma_i^{\mathrm{new}}$.

LEMMA 17.3. *There is a unique function $\Gamma_i^{\mathrm{new}}$ on $\mathfrak{g}_{\mathrm{rs}}$ which has the same germ about $0 \in \mathfrak{g}$ as $\Gamma_i$ and which satisfies* (17.7.1) *for all $\alpha \in F^\times$ and all $X \in \mathfrak{g}_{\mathrm{rs}}$. Moreover $\Gamma_i^{\mathrm{new}}$ is real-valued, translation invariant under the center of $\mathfrak{g}$, and invariant by conjugation under $G$.*

PROOF. Choose a lattice $L \subset \mathfrak{g}$ such that (17.7.1) holds for $\alpha = \pi$ (our chosen uniformizing element in $F$) and all $X \in L \cap \mathfrak{g}_{\mathrm{rs}}$. Iterating (17.7.1), we see that

$$\Gamma_i(X) = |\pi^k|^{d_i}\Gamma_i(\pi^{2k}X) \tag{17.8.1}$$

for all $k \geq 0$ and all $X \in L \cap \mathfrak{g}_{\mathrm{rs}}$.

For $X \in \mathfrak{g}_{\mathrm{rs}}$ we define $\Gamma_i^{\mathrm{new}}(X)$ by choosing $k \geq 0$ such that $\pi^{2k}X \in L$ and then putting

$$\Gamma_i^{\mathrm{new}}(X) := |\pi^k|^{d_i}\Gamma_i(\pi^{2k}X); \tag{17.8.2}$$

by (17.8.1) $\Gamma_i^{\mathrm{new}}$ is well-defined. This definition is of course forced on us, so $\Gamma_i^{\mathrm{new}}$ is clearly unique.

Next we show that $\Gamma_i^{\mathrm{new}}$ does satisfy (17.7.1). Let $\alpha \in F^\times$. Let $L'$ be a lattice in $\mathfrak{g}$ such that

$$\Gamma_i(X) = |\alpha|^{d_i}\Gamma_i(\alpha^2 X) \tag{17.8.3}$$

for all $X \in L' \cap \mathfrak{g}_{\mathrm{rs}}$. For a given $X \in \mathfrak{g}_{\mathrm{rs}}$ we may pick $k \geq 0$ such that $\pi^{2k}X \in L \cap L'$ and $\pi^{2k}\alpha^2 X \in L$, and we then have

$$\Gamma_i^{\mathrm{new}}(X) = |\pi^k|^{d_i}\Gamma_i(\pi^{2k}X) = |\pi^k|^{d_i}|\alpha|^{d_i}\Gamma_i(\pi^{2k}\alpha^2 X) = |\alpha|^{d_i}\Gamma_i^{\mathrm{new}}(\alpha^2 X), \tag{17.8.4}$$

as desired.

Looking back at how the functions $f_i$ (satisfying $\mu_j(f_i) = \delta_{ij}$) were shown (in 6.4) to exist, we see that they can be chosen to be real-valued functions. Then $\Gamma_i$ is real-valued and (17.8.2) shows that the same is true of $\Gamma_i^{\mathrm{new}}$.

The function $f_i$ is translation invariant under some lattice in $\mathfrak{g}$ and hence under some lattice in the center of $\mathfrak{g}$. It follows easily that the provisional germ $\Gamma_i(X) = O_X(f_i)$ is translation invariant under this lattice in the center, and hence (from (17.8.2)) that $\Gamma_i^{\mathrm{new}}$ is translation invariant under the center of $\mathfrak{g}$.

The provisional germ $\Gamma_i(X) = O_X(f_i)$ is clearly invariant under conjugation, from which it follows that the same is true of $\Gamma_i^{\mathrm{new}}$. □

From now on we replace the germs $\Gamma_i$ by the functions $\Gamma_i^{\mathrm{new}}$, but we drop the superscript "new."

We also need a slight strengthening of the fact that $\Gamma_i$ is translation invariant under the center $\mathfrak{z}$ of $\mathfrak{g}$. Let $G'$ be the derived group of the algebraic group $G$, and let $Z$ denote the center of $G$. Then $G(\bar{F}) = G'(\bar{F})Z(\bar{F})$, but for $F$-points we have only that $G'Z$ is a normal subgroup of finite index in $G$. We denote by $D$ the finite group $G/G'Z$. Each $G$-orbit $\mathcal{O}$ in $\mathfrak{g} = \mathfrak{g}' \oplus \mathfrak{z}$ decomposes as a finite union of $G'$-orbits $\mathcal{O}'$, permuted transitively by $D$. We normalize the invariant measures on the orbits in such a way that

$$\int_{\mathcal{O}} = \sum_{x \in D} \int_{x^{-1}\mathcal{O}'x}. \tag{17.8.5}$$

For a nilpotent $G$-orbit $\mathcal{O}$ (respectively, nilpotent $G'$-orbit $\mathcal{O}'$) we denote by $\Gamma^G_{\mathcal{O}}$ (respectively, $\Gamma^{G'}_{\mathcal{O}'}$) the corresponding Shalika germ on $\mathfrak{g}_{\mathrm{rs}}$ (respectively, $\mathfrak{g}'_{\mathrm{rs}}$).

Lemma 17.4. *Let $X \in \mathfrak{g}_{\mathrm{rs}}$ and decompose $X$ as $X' + Z$ with $X' \in \mathfrak{g}'_{\mathrm{rs}}$ and $Z \in \mathfrak{z}$. Then*

$$\Gamma^G_{\mathcal{O}}(X) = \sum_{\mathcal{O}' \subset \mathcal{O}} \Gamma^{G'}_{\mathcal{O}'}(X'). \tag{17.8.6}$$

Proof. Let $\mathcal{O}'_1, \dots, \mathcal{O}'_s$ be the nilpotent $G'$-orbits. For each nilpotent $G'$-orbit $\mathcal{O}'_i$ choose $f'_{\mathcal{O}'_i} \in C^\infty_c(\mathfrak{g}')$ such that

$$\int_{\mathcal{O}'_i} f'_{\mathcal{O}'_j} = \delta_{ij}. \tag{17.8.7}$$

Thus the regular semisimple orbital integrals of $f'_{\mathcal{O}'}$ give the provisional Shalika germ $\Gamma^{G'}_{\mathcal{O}'}$.

For a nilpotent $G$-orbit $\mathcal{O}$ put

$$f'_{\mathcal{O}} := |D|^{-1} \sum_{\mathcal{O}' \subset \mathcal{O}} f'_{\mathcal{O}'}. \tag{17.8.8}$$

We extend $f'_{\mathcal{O}}$ to a function $f_{\mathcal{O}} \in C^\infty_c(\mathfrak{g})$ by choosing a lattice $L$ in $\mathfrak{z}$ and putting

$$f(X' + Z) = f'(X')1_L(Z) \tag{17.8.9}$$

for any $X' \in \mathfrak{g}'$ and any $Z \in \mathfrak{z}$. Here $1_L$ denotes the characteristic function of $L$. It is easy to see that

$$\int_{\mathcal{O}_i} f_{\mathcal{O}_j} = \delta_{ij} \tag{17.8.10}$$

for every pair of nilpotent $G$-orbits $\mathcal{O}_i$, $\mathcal{O}_j$, so that the regular semisimple orbital integrals of $f_{\mathcal{O}}$ give the provisional Shalika germ $\Gamma^G_{\mathcal{O}}$, and another easy calculation then shows that the provisional Shalika germs for $G$ and $G'$ are related as in the statement of the lemma. By homogeneity the same is true for the Shalika germs themselves. □

**17.9. Germ expansions about arbitrary central elements in $\mathfrak{g}$.** We have been studying germ expansions about $0 \in \mathfrak{g}$. These involve orbital integrals for the nilpotent orbits $\mathcal{O}_i$. Now we consider germ expansions about an arbitrary element $Z$ in the center of $\mathfrak{g}$. These will involve orbital integrals $\mu_{Z+\mathcal{O}_i}$ for the orbits $Z+\mathcal{O}_i$, but will involve exactly the same germs $\Gamma_i$ as before.

Theorem 17.5. *Let $Z$ be an element in the center of $\mathfrak{g}$. For every $f \in C^\infty_c(\mathfrak{g})$ there exists an open neighborhood $U_f$ of $Z$ in $\mathfrak{g}$ such that*

$$O_X(f) = \sum_{i=1}^{r} \mu_{Z+\mathcal{O}_i}(f) \cdot \Gamma_i(X) \tag{17.9.1}$$

*for all* $X \in U_f \cap \mathfrak{g}_{\mathrm{rs}}$.

Proof. Apply Theorem 17.1 to the translate of $f$ by $Z$ and use that our canonical Shalika germs $\Gamma_i$ are translation invariant under the center. $\square$

**17.10. Germ expansions about arbitrary semisimple elements in $\mathfrak{g}$.** We are going to use the descent theory developed in section 16 in order to obtain germ expansions about an arbitrary semisimple element $S \in \mathfrak{g}$. We fix such an element $S$ and let $H := G_S$ denote the centralizer of $S$, a connected reductive subgroup of $G$. Then $S$ is contained in the open subset $\mathfrak{h}'$ of $\mathfrak{h}$. We write $\mathfrak{h}'_{\mathrm{rs}}$ for the intersection $\mathfrak{h}' \cap \mathfrak{h}_{\mathrm{rs}}$. Then $\mathfrak{h}'_{\mathrm{rs}} \subset \mathfrak{g}_{\mathrm{rs}}$, since a semisimple element in $\mathfrak{h}'$ is regular in $\mathfrak{h}$ if and only if it is regular in $\mathfrak{g}$.

We will also be concerned with all $G$-orbits of elements $X$ such that $X_s = S$ (as usual, $X_s$ denotes the semisimple part of $X$). Such orbits are in one-to-one correspondence with $H$-orbits of nilpotent elements $Y \in \mathfrak{h}$ (with $Y$ corresponding to $X = S + Y$). Let $Y_1, \dots, Y_s$ be a set of representatives for the nilpotent $H$-orbits in $\mathfrak{h}$. Let $\mu_{S+Y_i}$ denote the orbital integral on $\mathfrak{g}$ obtained by integration over the $G$-orbit of $S + Y_i$. Let $\Gamma_i^H$ be the canonical Shalika germ on $\mathfrak{h}_{\mathrm{rs}}$ corresponding to the (nilpotent) $H$-orbit of $Y_i$.

Theorem 17.6. *Let $S$, $H$ be as above. For every $f \in C_c^\infty(\mathfrak{g})$ there exists an open neighborhood $U_f$ of $S$ in $\mathfrak{h}'$ such that*

$$O_X(f) = \sum_{i=1}^{s} \mu_{S+Y_i}(f) \cdot \Gamma_i^H(X) \tag{17.10.1}$$

*for all* $X \in U_f \cap \mathfrak{g}_{\mathrm{rs}} = U_f \cap \mathfrak{h}_{\mathrm{rs}}$.

Proof. Note that $\pi_H(S)$ lies in $\mathbb{A}'_H(F)$. Let $\omega_H$ be a compact open neighborhood of $\pi_H(S)$ in $\mathbb{A}'_H(F)$. By Lemma 16.1 there exists $\phi \in C_c^\infty(\mathfrak{h})$ such that

$$\int_{G_X \backslash H} \phi(h^{-1}Xh)\, d\dot{h} = \int_{G_X \backslash G} f(g^{-1}Xg)\, d\dot{g}. \tag{17.10.2}$$

for any $X \in \pi_H^{-1}(\omega_H)$. Note that $\pi_H^{-1}(\omega_H)$ contains all the elements $S + Y_i$ and is an open neighborhood of $S$ in $\mathfrak{h}'$.

Apply the Shalika germ expansion (Theorem 17.5) to the central element $S \in \mathfrak{h}$ and the function $\phi$. Using (17.10.2) to rewrite this expansion in terms of orbital integrals on $\mathfrak{g}$, we obtain the desired result. $\square$

**17.11. Normalized orbital integrals and Shalika germs.** It is sometimes more convenient (see 13.12, for example) to use the normalized orbital integrals $I_X$ ($X \in \mathfrak{g}_{\mathrm{rs}}$) defined by $I_X = |D(X)|^{1/2} O_X$. When we use $I_X$ instead of $O_X$, we need to use the normalized Shalika germs

$$\bar{\Gamma}_i(X) := |D(X)|^{1/2}\Gamma_i(X)$$

instead of the usual Shalika germs.

Clearly Theorem 17.1 remains valid when $O_X$, $\Gamma_i$ are replaced by $I_X$, $\bar{\Gamma}_i$ respectively. Now consider the germ expansion about an arbitrary semisimple element $S \in \mathfrak{g}$. As usual put $H := G_S$. The function $X \mapsto \det(\mathrm{ad}(X); \mathfrak{g}/\mathfrak{h})$ is non-zero on $\mathfrak{h}'$ and its $p$-adic absolute value is locally constant on $\mathfrak{h}'$. Moreover

$$D^G(X) = D^H(X) \cdot \det(\mathrm{ad}(X); \mathfrak{g}/\mathfrak{h}).$$

Therefore there is a neighborhood of $S$ in $\mathfrak{h}'$ on which

$$|D^G(X)|^{1/2} = |D^H(X)|^{1/2} \cdot |\det(\mathrm{ad}(S); \mathfrak{g}/\mathfrak{h})|^{1/2}.$$

It then follows from Theorem 17.6 that

$$I_X(f) = |\det(\mathrm{ad}(S); \mathfrak{g}/\mathfrak{h})|^{1/2} \sum_{i=1}^{s} \mu_{S+Y_i}(f) \cdot \bar{\Gamma}_i^H(X) \tag{17.11.1}$$

for all $X \in \mathfrak{g}_{\mathrm{rs}}$ in some sufficiently small neighborhood of $S$ in $\mathfrak{h}'$.

Since $D^G(X)$ is a homogeneous polynomial of degree $\dim(G) - \mathrm{rank}(G)$, where $\mathrm{rank}(G)$ denotes the dimension of any maximal torus in $G$, we see immediately that the homogeneity property (17.7.1) of the Shalika germs $\Gamma_i$ implies the following homogeneity property for the normalized Shalika germs $\bar{\Gamma}_i$:

$$\bar{\Gamma}_i(\alpha^2 X) = |\alpha|^{\dim(G_{X_i}) - \mathrm{rank}(G)} \cdot \bar{\Gamma}_i(X) \tag{17.11.2}$$

for all $\alpha \in F^\times$ and all $X \in \mathfrak{g}_{\mathrm{rs}}$. Here we have chosen $X_i \in \mathcal{O}_i$ and introduced its centralizer $G_{X_i}$. Note that the exponent $\dim(G_{X_i}) - \mathrm{rank}(G)$ appearing in (17.11.2) is always non-negative. This simple observation will play an important role in the proof (to be given shortly) of the boundedness of normalized Shalika germs.

**17.12. $\bar{\Gamma}_i$ is a linear combination of functions $\bar{\Gamma}_j^H$ in a neighborhood of $S$.** Again let $S$ be a semisimple element in $\mathfrak{g}$, let $H$ be its centralizer in $G$, and let $T$ be a maximal torus in $H$. Consider one of the normalized Shalika germs $\bar{\Gamma}_i$ for $G$. We are interested in the behavior of $\bar{\Gamma}_i$ on a small neighborhood of $S$ in $\mathfrak{t}$.

LEMMA 17.7. *There exists a neighborhood $V$ of $S$ in $\mathfrak{t}$ such that the restriction of $\bar{\Gamma}_i$ to $V \cap \mathfrak{t}_{\mathrm{reg}}$ is a linear combination of restrictions of normalized Shalika germs for $H$.*

PROOF. Pick $f_i \in C_c^\infty(\mathfrak{g})$ such that $\mu_j(f_i) = \delta_{ij}$. By the Shalika germ expansion for $f_i$ there is a lattice $L$ in $\mathfrak{t}$ small enough that

$$\bar{\Gamma}_i(X) = I_X(f_i) \tag{17.12.1}$$

for all regular $X$ in $L$. Let $\alpha \in F^\times$. By homogeneity of Shalika germs (for both $G$ and $H$) the lemma holds for $S$ if and only if it holds for $\alpha S$. Therefore we may assume (by scaling $S$ suitably) that $S \in L$. For some neighborhood $V$ of $S$ in $L$ we also have (by (17.11.1))

$$I_X(f_i) = |\det(\mathrm{ad}(S); \mathfrak{g}/\mathfrak{h})|^{1/2} \sum_{i=1}^{s} \mu_{S+Y_i}(f_i) \cdot \bar{\Gamma}_i^H(X) \tag{17.12.2}$$

for all regular $X$ in $V$. Combining (17.12.1) and (17.12.2), we get the lemma. □

COROLLARY 17.8. *Let $T$ be any maximal torus in $G$. Each normalized Shalika germ $\bar{\Gamma}_i$ is locally constant on $\mathfrak{t}_{\mathrm{reg}}$.*

PROOF. Apply the lemma above to any regular element $S$ in $\mathfrak{t}$. Then $H = T$. Furthermore, the only nilpotent element in $\mathfrak{h}$ is 0, and its normalized Shalika germ is constant. Therefore $\bar{\Gamma}_i$ is constant in some sufficiently small neighborhood of $S$, as was to be shown. □

**17.13. Locally bounded functions.** We are going to show that the normalized Shalika germs $\bar{\Gamma}_i$ are locally bounded functions on $\mathfrak{t}$. First let's recall what this means. Let $f$ be a complex-valued function on a topological space $X$. We say that $f$ is *locally bounded* on $X$ if every point $x \in X$ has a neighborhood $U_x$ such that $f$ is bounded on $U_x$. When $X$ is a locally compact Hausdorff space, $f$ is locally bounded if and only $f$ is bounded on every compact subset of $X$ (easy exercise!).

**17.14. Local boundedness of normalized Shalika germs.** Let $\bar{\Gamma}_i$ be one of our normalized Shalika germs on $\mathfrak{g}$. Let $T$ be a maximal torus in $G$. We have just seen that $\bar{\Gamma}_i$ is locally constant on $\mathfrak{t}_{\text{reg}}$; therefore it is locally bounded on $\mathfrak{t}_{\text{reg}}$ for trivial reasons. However we now extend $\bar{\Gamma}_i$ by 0 to a function on $\mathfrak{t}$. We are going to show that $\bar{\Gamma}_i$ is locally bounded as a function on $\mathfrak{t}$ (a result of Harish-Chandra). What this means concretely is that for all $S \in \mathfrak{t}$ there is a neighborhood $U$ of $S$ in $\mathfrak{t}$ such that $\bar{\Gamma}_i$ is bounded on $U \cap \mathfrak{t}_{\text{reg}}$.

THEOREM 17.9. [**HC78, HC99**] *Every normalized Shalika germ $\bar{\Gamma}_i$ is a locally bounded function on $\mathfrak{t}$.*

PROOF. We use induction on the dimension of $G$, the case $\dim(G) = 0$ being trivial. By Lemma 17.4 we may assume that the center of $\mathfrak{g}$ is 0.

Let $S$ be any non-zero element in $\mathfrak{t}$. Since $S$ is not central, the induction hypothesis applies to the centralizer $H$ of $S$ in $G$. Now use Lemma 17.7 to conclude that $\bar{\Gamma}_i$ is bounded on some neighborhood of $S$.

We now know that $\bar{\Gamma}_i$ is locally bounded on $\mathfrak{t} \setminus \{0\}$. Choose any lattice $L$ in $\mathfrak{t}$. Then $L \setminus \pi^2 L$ is a compact subset of $\mathfrak{t} \setminus \{0\}$, and therefore $\bar{\Gamma}_i$ is bounded on $L \setminus \pi^2 L$. By homogeneity (17.11.2) it follows that $\bar{\Gamma}_i$ is bounded on $L$ (by the same bound as on $L \setminus \pi^2 L$).

We have shown that $\bar{\Gamma}_i$ is locally bounded everywhere on $\mathfrak{t}$, and the proof is complete. $\square$

As a consequence of the local boundedness of normalized Shalika germs, we get another result of Harish-Chandra, needed for the local trace formula.

THEOREM 17.10. *Let $f \in C_c^\infty(\mathfrak{g})$ and let $T$ be a maximal torus in $G$. Then the function $X \mapsto I_X(f)$ on $\mathfrak{t}_{\text{reg}}$ is bounded and locally constant on $\mathfrak{t}_{\text{reg}}$. When extended by 0 to all of $\mathfrak{t}$, the function $X \mapsto I_X(f)$ is compactly supported as well; in other words, there is a compact subset $C$ of $\mathfrak{t}$ such that $I_X(f)$ vanishes for all regular elements of $\mathfrak{t}$ not lying in $C$. It should be noted that $X \mapsto I_X(f)$ is usually not compactly supported as a function on $\mathfrak{t}_{\text{reg}}$.*

PROOF. Local constancy in a neighborhood of $S \in \mathfrak{t}_{\text{reg}}$ follows from the Shalika germ expansion about $S$ (for which $H = T$, so that there is just one germ, and it is constant).

Boundedness in a sufficiently small neighborhood of any $S \in \mathfrak{t}$ follows from the Shalika germ expansion (17.11.1) about $S$ together with the local boundedness of normalized Shalika germs on $H = G_S$. Thus $X \mapsto I_X(f)$ is locally bounded on $\mathfrak{t}$. Boundedness will then follow once we have proved that $X \mapsto I_X(f)$ is compactly supported on $\mathfrak{t}$.

It now remains only to show that $X \mapsto I_X(f)$ is in fact compactly supported on $\mathfrak{t}$. The support of $f$ is a compact subset of $\mathfrak{g}$, so its image $\omega$ under $\pi_G : \mathfrak{g} \to \mathbb{A}_G(F) = (\mathfrak{t}/W)(F)$ is compact. But since $\mathfrak{t} \to \mathfrak{t}/W$ is a proper morphism of algebraic varieties, the map $\mathfrak{t} \to (\mathfrak{t}/W)(F)$ is a proper map between locally compact

Hausdorff spaces. Therefore the inverse image $C$ of $\omega$ under $\mathfrak{t} \to (\mathfrak{t}/W)(F)$ is a compact subset of $\mathfrak{t}$. Clearly $I_X(f)$ vanishes off $C$. $\square$

For weighted orbital integrals we have the following partial generalization of our last theorem.

THEOREM 17.11. *Let $f \in C_c^\infty(\mathfrak{g})$, let $T$ be a maximal torus in $G$, and let $v$ be a locally constant function on $G$ that is invariant under left translation by $T$. Then the function*

$$X \mapsto \int_{T\backslash G} f(g^{-1}Xg)v(g)\,d\dot{g}$$

*on $\mathfrak{t}_{\mathrm{reg}}$ is locally constant on $\mathfrak{t}_{\mathrm{reg}}$ and, when extended by $0$ to $\mathfrak{t}$, is compactly supported on $\mathfrak{t}$.*

PROOF. Compact support is established just as in the previous result. Local constancy needs to be proved more directly, as we have not developed a theory of Shalika germs for weighted orbital integrals.

Let $Y \in \mathfrak{t}$. We are going to find a neighborhood $U$ of $Y$ in $\mathfrak{t}_{\mathrm{reg}}$ on which our function is constant. Consider the function $\phi$ on $\mathfrak{t} \times (T\backslash G)$ defined by $\phi(X,g) := f(g^{-1}Xg)$. Clearly $\phi$ is locally constant, but it is usually not compactly supported. However, now choosing a compact open neighborhood $\omega_T$ of $Y$ in $\mathfrak{t}_{\mathrm{reg}}$, we see from Lemma 15.1 (Harish-Chandra's compactness lemma, applied to $H = T$ and $\omega_{\mathfrak{g}} = \mathrm{Supp}(f)$) that the restriction of $\phi$ to $\omega_T \times (T\backslash G)$ is compactly supported. By Lemma 2.1 there exists an open neighborhood $U$ of $Y$ in $\omega_T$ such that $\phi(X,g) = \phi(Y,g)$ for all $X \in U$, $g \in T\backslash G$. It follows that

$$\int_{T\backslash G} f(g^{-1}Xg)v(g)\,d\dot{g} = \int_{T\backslash G} f(g^{-1}Yg)v(g)\,d\dot{g}$$

for all $X \in U$. $\square$

## 18. Norms on affine varieties over local fields

The spaces we are working with are usually non-compact, and non-compactly supported functions on them can certainly be unbounded. For various purposes we need a natural way to measure growth rates of such functions. For this we must be able to measure the size of points in the spaces. For instance on the real line one usually uses the absolute value of a real number to measure its size, and one says that a function $f(x)$ on the real line has polynomial growth if there exist $c, R > 0$ such that $|f(x)| \le c|x|^R$ for all $x \in \mathbb{R}$. We want to be able to do something similar on the spaces we are using. For this purpose we now develop a theory of norms on $X(F)$ for any variety (usually affine) over a field $F$ equipped with an absolute value.

Let $F$ be a field equipped with a non-trivial absolute value $|\cdot|$. Thus $|\cdot|$ is a non-negative real-valued function on $F$ such that

(1) $|x| = 0$ if and only if $x = 0$.
(2) $|x+y| \le |x| + |y|$ for all $x, y \in F$.
(3) $|xy| = |x||y|$ for all $x, y \in F$.
(4) There exists $x \in F^\times$ such that $|x| \ne 1$.

As usual $(x,y) \mapsto |x-y|$ defines a metric on $F$ with respect to which $F$ may or may not be complete. Starting with subsection 18.7 we will assume that $F$ is complete.

**18.1. Abstract norms.** By an *abstract norm* on a set $X$ we mean a real-valued function $\|\cdot\|$ on $X$ such that $\|x\| \geq 1$ for all $x \in X$. Given two abstract norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on $X$, we say $\|\cdot\|_2$ *dominates* $\|\cdot\|_1$ and write $\|\cdot\|_1 \prec \|\cdot\|_2$ if there exist real numbers $c > 0$ and $R > 0$ such that

$$\|x\|_1 \leq c\|x\|_2^R$$

for all $x \in X$. The relation of dominance is transitive. We say that two norms are *equivalent* if they dominate each other.

For any abstract norm $\|\cdot\|$ we have (by virtue of our requirement that $\|x\| \geq 1$) the inequality

$$c_1\|x\|^{R_1} \leq c_2\|x\|^{R_2}$$

whenever $0 < c_1 \leq c_2$ and $0 < R_1 \leq R_2$. This allows us to increase the constants $c$, $R$ occurring in the dominance relation whenever it is convenient to do so.

Given two abstract norms $\|x\|_1$ and $\|x\|_2$ on $X$, the three abstract norms

$$\sup\{\|x\|_1, \|x\|_2\}, \ \|x\|_1 + \|x\|_2, \ \|x\|_1 \cdot \|x\|_2$$

on $X$ are equivalent, and their common equivalence class depends only on the equivalence classes of $\|x\|_1$ and $\|x\|_2$.

**18.2. Norms on affine varieties over $F$.** Let $X$ be an affine scheme of finite type over $F$ and write $\mathcal{O}_X$ for its ring of regular functions, a finitely generated $F$-algebra. For any finite set $f_1, \dots, f_m$ of generators for the $F$-algebra $\mathcal{O}_X$, we define an abstract norm $\|\cdot\|$ on $X(F)$ by

$$\|x\| := \sup\{1, |f_1(x)|, \dots, |f_m(x)|\} \tag{18.2.1}$$

for $x \in X(F)$.

Now let $f \in \mathcal{O}_X$. It is easy to see that there exist $c, R \geq 0$ such that

$$|f(x)| \leq c\|x\|^R \tag{18.2.2}$$

for all $x \in X(F)$. [Indeed, writing $f$ as a polynomial in $f_1, \dots, f_m$, we may take for $c$ the sum of the absolute values of the coefficients in the polynomial, and for $R$ the degree of the polynomial.] Since we are free to increase $c, R$, we may choose them so that $c, R > 0$ (or even $\geq 1$) whenever it is convenient to do so.

Using (18.2.2) for all the members of some other generating set for $\mathcal{O}_X$, we see that the equivalence class of the abstract norm (18.2.1) is independent of the choice of generating set, and by a *norm* on $X(F)$ we mean any abstract norm lying in this equivalence class.

EXAMPLE 18.1. On $F^n$, the set of $F$-points of $\mathbb{A}^n$,

$$\|(x_1, \dots, x_n)\| := \sup\{1, |x_1|, \dots, |x_n|\}$$

is a norm. The restriction of this norm to the $F$-points of any closed subscheme of $\mathbb{A}^n$ is a norm on that set.

EXAMPLE 18.2. On $(F^\times)^n$, the set of $F$-points of $(\mathbb{G}_m)^n$,

$$\|(x_1, \dots, x_n)\| := \sup\{|x_1|, |x_1|^{-1} \dots, |x_n|, |x_n|^{-1}\}$$

is a norm.

**18.3. Bounded subsets.** Let $X$ be an affine scheme of finite type over $F$ and let $\|\cdot\|_X$ be a norm on $X(F)$. We say that a subset $B$ of $X(F)$ is *bounded* if the norm function $\|\cdot\|_X$ has an upper bound on $B$. This notion of boundedness is clearly independent of the choice of norm, so it makes sense to talk about bounded subsets of $X(F)$ without specifying any particular norm.

**18.4. Properties of norms.** We need to establish various simple results about norms. It is especially important to compare norms on varieties when a morphism between them is given.

PROPOSITION 18.1. *Let $X$ and $Y$ be affine schemes of finite type over $F$ and let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be norms on $X(F)$ and $Y(F)$ respectively.*

(1) *Let $\phi : Y \to X$ be a morphism and denote by $\phi^*\|\cdot\|_X$ the abstract norm on $Y(F)$ obtained by composing $\|\cdot\|_X$ with $\phi : Y(F) \to X(F)$. Then $\|\cdot\|_Y$ dominates $\phi^*\|\cdot\|_X$. If $\phi$ is finite, then $\|\cdot\|_Y$ is equivalent to $\phi^*\|\cdot\|_X$.*
(2) *Suppose $Y$ is a closed subscheme of $X$. Then the restriction of $\|\cdot\|_X$ to $Y(F)$ is equivalent to $\|\cdot\|_Y$.*
(3) *If $F$ is locally compact, then a subset of $X(F)$ has compact closure if and only if it is bounded.*
(4) *All three of $\sup\{\|x\|_X, \|y\|_Y\}$, $\|x\|_X + \|y\|_Y$, and $\|x\|_X \cdot \|y\|_Y$ are valid norms on $(X \times Y)(F) = X(F) \times Y(F)$.*
(5) *Let $U := X_f$ denote the principal open subset of $X$ determined by a regular function $f$ on $X$, so that $U(F) = \{x \in X(F) : f(x) \neq 0\}$. Then $\|u\|_U := \sup\{\|u\|_X, |f(u)|^{-1}\}$ is a norm on $U(F)$.*
(6) *Suppose we are given a finite cover of $X$ by affine open subsets $U_1, \dots, U_r$ as well as a norm $\|\cdot\|_i$ on $U_i(F)$ for each $i = 1, \dots, r$. For $x \in X(F)$ define $\|x\|$ to be the infinum of the numbers $\|x\|_i$, where $i$ ranges over the set of indices for which $x \in U_i(F)$. Then $\|\cdot\|$ is a norm on $X(F)$.*
(7) *Let $G$ be a group scheme of finite type over $F$, and suppose we are given an action of $G$ on $X$. Let $B$ be a bounded subset of $G(F)$. Then there exist $c, R > 0$ such that $\|bx\|_X \leq c\|x\|_X^R$ for all $b \in B, x \in X(F)$.*

PROOF. We begin by proving the first part of the proposition. Using (18.2.2) for the pull-backs by $\phi$ of the members of a generating set for the $F$-algebra $\mathcal{O}_X$, we see that $\|\cdot\|_Y$ dominates $\phi^*\|\cdot\|_X$.

Now suppose that $\phi$ is finite and let $g \in \mathcal{O}_Y$. Then there exist $n \geq 1$ and $f_1, \dots, f_n \in \mathcal{O}_X$ such that

$$g^n = f_1 g^{n-1} + \dots + f_{n-1} g + f_n.$$

We claim that for all $y \in Y(F)$ we have the inequality

$$|g(y)| \leq \sup\{1, |f_1(\phi(y))| + \dots + |f_n(\phi(y))|\}. \tag{18.4.1}$$

Indeed, this is trivially true if $|g(y)| \leq 1$, and otherwise we have

$$g(y) = f_1(\phi(y)) + f_2(\phi(y))g(y)^{-1} + \dots + f_n(\phi(y))g(y)^{-(n-1)}$$

and hence

$$|g(y)| \leq |f_1(\phi(y))| + |f_2(\phi(y))| + \dots + |f_n(\phi(y))|.$$

Using (18.2.2) for $f_1, \dots, f_n$, we see from (18.4.1) that there exist $c, R > 0$ such that

$$|g(y)| \leq c\|\phi(y)\|_X^R \quad \forall\, y \in Y(F). \tag{18.4.2}$$

Now choose a finite generating set $g_1, \dots, g_m$ for the $F$-algebra $\mathcal{O}_Y$. As our norm $\|\cdot\|_Y$ we are free to take the one obtained from this generating set (see (18.2.1)). Choosing $c, R \geq 1$ large enough that (18.4.2) holds for all the functions $g_1, \dots, g_m$, we see that $\|\cdot\|_Y$ is dominated by $\phi^*\|\cdot\|_X$. Since we already proved dominance in the other direction, we conclude that $\|\cdot\|_Y$ and $\phi^*\|\cdot\|_X$ are equivalent, as desired.

The second part of the proposition follows from the first, because closed immersions are finite morphisms. Of course a more direct proof can also be given.

As for the third part, we use the second part to reduce to the case of affine space $\mathbb{A}^n$, for which the result is obvious.

For the fourth part we note that it is obvious from the definition of norm that

$$\sup\{\|\cdot\|_X, \|\cdot\|_Y\}$$

is a valid norm on $X(F) \times Y(F)$. It then follows from the discussion at the very end of subsection 18.1 that the other two abstract norms are valid norms as well.

For the fifth part note that the equivalence class of $\|u\|_U = \sup\{\|u\|_X, |f(u)|^{-1}\}$ depends only on that of $\|\cdot\|_X$, so we may suppose that $\|\cdot\|_X$ is the norm (18.2.1) obtained from generators $f_1, \dots, f_m$ of the $F$-algebra $\mathcal{O}_X$. Then $f^{-1}, f_1, \dots, f_m$ generate $\mathcal{O}_U$, and the norm obtained from this generating set is precisely $\|\cdot\|_U$.

For the sixth part we must show that $\|\cdot\|_X$ is equivalent to $\|\cdot\|$. First we note that $\|\cdot\|_X$ is dominated by $\|\cdot\|$. Indeed, this follows from the first part of this proposition, applied to the morphism

$$\coprod_{i=1}^{r} U_i \to X.$$

It remains to prove that $\|\cdot\|$ is dominated by $\|\cdot\|_X$. Refine the given open cover $U_1, \dots, U_r$ to get an open cover $V_1, \dots, V_s$ by principal open subsets of $X$ (say $V_j = X_{f_j}$ for $f_j \in \mathcal{O}_X$) such that for each index $j$ there exists an index $i(j)$ such that $V_j \subset U_{i(j)}$. By the fifth part of this proposition

$$\|v\|_j := \sup\{\|v\|_X, |f_j(v)|^{-1}\}$$

is a valid norm on $V_j(F)$. By the first part of this proposition (applied to all the inclusions $V_j \hookrightarrow U_{i(j)}$) there exist $d, S \geq 1$ such that for all $j$

$$\|v\|_{i(j)} \leq d\|v\|_j^S \quad \forall\, v \in V_j(F).$$

Since the principal open subsets $V_j$ cover $X$, there exist $g_1, \dots, g_s \in \mathcal{O}_X$ such that $\sum_{j=1}^{s} f_j g_j = 1$. By (18.2.2) there exist $c, R \geq 1$ such that for all $j$

$$|g_j(x)| \leq c\|x\|_X^R \quad \forall\, x \in X(F).$$

Now let $x \in X(F)$. Then

$$1 = |\sum_{j=1}^{s} f_j(x)g_j(x)| \leq \sum_{j=1}^{s} |f_j(x)| \cdot \big(c\|x\|_X^R\big),$$

and thus there exists $j$ such that $|f_j(x)| \cdot \big(c\|x\|_X^R\big) \geq 1/s$, from which we see that $f_j(x) \neq 0$ (so that $x \in V_j(F) \subset U_{i(j)}(F)$) and that moreover

$$|f_j(x)|^{-1} \leq sc\|x\|_X^R.$$

Thus

$$
\begin{aligned}
\|x\| \leq \|x\|_{i(j)} \leq d\|x\|_j^S &= d[\sup\{\|x\|_X, |f_j(x)|^{-1}\}]^S \\
&\leq d[\sup\{\|x\|_X, sc\|x\|_X^R\}]^S.
\end{aligned}
\tag{18.4.3}
$$

But, since $s, c, R \geq 1$, we have $sc\|x\|_X^R \geq \|x\|_X$, so (18.4.3) becomes

$$\|x\| \leq d(sc)^S \|x\|_X^{RS},$$

showing that $\|\cdot\|$ is dominated by $\|\cdot\|_X$.

Finally, to prove the seventh part one considers the action morphism $G \times X \to X$ and uses the first and fourth parts of this proposition. □

**18.5. Arbitrary schemes of finite type over $F$.** Let $X$ be any scheme of finite type over $F$. Let $U_1, \ldots, U_r$ be any cover of $X$ by affine open subsets. For $i = 1, \ldots, r$ let $\|\cdot\|_i$ be any norm on $U_i(F)$. Define an abstract norm $\|\cdot\|$ on $X(F)$ by

$$\|x\| = \inf\{\|x\|_i : i \text{ such that } x \in U_i(F)\}.$$

It is not difficult to show that the equivalence class of $\|\cdot\|$ is independent of all choices, so that we have defined a canonical equivalence class of norms on $X(F)$. When $X$ is affine, we recover our old notion of norm. When $X$ is projective, the constant function 1 is a valid norm on $X(F)$. The reader may enjoy checking these statements as an exercise, but in this article we will only need norms on affine schemes.

**18.6. Norm descent property.** Let $X$ and $Y$ be affine schemes of finite type over $F$, let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be norms on $X(F)$ and $Y(F)$ respectively, and let $\phi : Y \to X$ be a morphism. We say that $\phi$ has the *norm descent property* if the restriction of $\|\cdot\|_X$ to $\operatorname{im}[Y(F) \to X(F)]$ is equivalent to the abstract norm $\phi_*\|\cdot\|_Y$ on $\operatorname{im}[Y(F) \to X(F)]$ whose value at $x$ is equal to the infinum of the values of $\|\cdot\|_Y$ on the fiber of $\phi : Y(F) \to X(F)$ over $x$. It is easy to see that this condition is independent of the choice of $\|\cdot\|_X$ and $\|\cdot\|_Y$. Moreover, it follows from the first part of Proposition 18.1 that the restriction of $\|\cdot\|_X$ is automatically dominated by $\phi_*\|\cdot\|_Y$; therefore the norm descent property is equivalent to the condition that $\phi_*\|\cdot\|_Y$ be dominated by the restriction of $\|\cdot\|_X$.

Lemma 18.3. *Let $\phi : Y \to X$ be a morphism of affine schemes over $F$, and let $\|\cdot\|_Y$ be a norm on $Y(F)$. Then the following two conditions are equivalent.*

1. *The morphism $\phi$ satisfies the norm descent property.*
2. *There exists a norm $\|\cdot\|_X$ on $X(F)$ such that for all $x \in \operatorname{im}[Y(F) \to X(F)]$ there exists $y \in Y(F)$ such that $\phi(y) = x$ and $\|y\|_Y \leq \|x\|_X$.*

Proof. The second condition trivially implies that $\phi_*\|\cdot\|_Y$ is dominated by the restriction of $\|\cdot\|_X$, hence that $\phi$ has the norm descent property.

Now assume the first condition. Start with any norm $\|\cdot\|_X$ on $X(F)$. Then there exist $c, R \geq 1$ such that $\phi_*\|\cdot\|_Y \leq c\|\cdot\|_X^R$ holds on $\operatorname{im}[Y(F) \to X(F)]$. Increasing $c$, we may improve $\leq$ to $<$, and then replacing $\|\cdot\|_X$ by the equivalent norm $c\|\cdot\|_X^R$, we end up with a norm $\|\cdot\|_X$ for which

$$\phi_*\|\cdot\|_Y < \|\cdot\|_Y$$

holds on $\operatorname{im}[Y(F) \to X(F)]$. It is clear that the second condition is satisfied for this choice of $\|\cdot\|_X$. □

Lemma 18.4. *Consider morphisms* $X \xrightarrow{f} Y \xrightarrow{g} Z$ *of affine schemes of finite type over* $F$. *Put* $h = gf : X \to Z$. *Assume that the map* $f : X(F) \to Y(F)$ *is surjective. Then*

(1) *If* $f, g$ *satisfy the norm descent property, then so does* $h$.
(2) *If* $h$ *satisfies the norm descent property, then so does* $g$.

Proof. The proof of the first statement uses only Lemma 18.3. The proof of the second statement is similar, but also uses the first part of Proposition 18.1, applied to the morphism $f$. Details are left to the reader. □

Proposition 18.2. *Let* $\phi : Y \to X$ *be a morphism of affine schemes of finite type over* $F$. *For open* $U$ *in* $X$ *write* $\phi_U$ *for the morphism* $\phi^{-1}U \to U$ *obtained by restriction from* $\phi$.

(1) *The norm descent property for* $\phi : Y \to X$ *is local with respect to the Zariski topology on* $X$. *In other words, for any cover of* $X$ *by affine open subsets, the morphism* $\phi$ *has the norm descent property if and only if the morphisms* $\phi_U$ *have the norm descent property for every member* $U$ *of the open cover.*
(2) *If the morphism* $\phi : Y \to X$ *admits a section, then* $\phi$ *has the norm descent property. More generally, if* $\phi : Y \to X$ *admits sections locally in the Zariski topology on* $X$, *then* $\phi$ *has the norm descent property.*
(3) *Let* $G$ *be a connected reductive group over* $F$, *and let* $M$ *be a Levi subgroup of* $G$. *Then the canonical morphism* $G \to G/M$ *has the norm descent property.*

Proof. We begin by proving the first part of the proposition in the special case of principal open subsets. So suppose for the moment that $U = X_f$ is the principal open subset of $X$ defined by $f \in \mathcal{O}_X$. Then $\phi^{-1}X_f = Y_g$, where $g$ is the image of $f$ under the homomorphism $\phi^* : \mathcal{O}_X \to \mathcal{O}_Y$. Assuming the norm descent property for $\phi$, we need to prove the norm descent property for $\phi_U : Y_g \to X_f$.

We are free to use any convenient norms on $Y_g(F)$ and $X_f(F)$. Start by picking any norm $\|\cdot\|_Y$ on $Y(F)$. Choose our norm $\|\cdot\|_X$ on $X(F)$ so that the second condition of Lemma 18.3 holds for it. On the principal open subsets $Y_g$, $X_f$ we use the norms $\|\cdot\|_{Y_g}$, $\|\cdot\|_{X_f}$ obtained from $\|\cdot\|_Y$, $\|\cdot\|_X$ by the construction in the fifth part of Proposition 18.1. With these choices, it is easy to check that $\|\cdot\|_{Y_g}$ and $\|\cdot\|_{X_f}$ satisfy the second condition of Lemma 18.3, proving the norm descent property for $\phi_U$.

Next, suppose we have a cover of the affine scheme $X$ by principal affine open subschemes $X_i = X_{f_i}$ $(i = 1, \dots, r)$. Putting $g_i := \phi^*(f_i)$ and $Y_i := Y_{g_i} = \phi^{-1}X_{f_i}$, we get (by restriction from $\phi$) morphisms $\phi_i : Y_i \to X_i$. Assuming that each $\phi_i$ satisfies the norm descent property, we must show that $\phi$ satisfies the norm descent property.

Again we may use any convenient norms on $X(F)$, $Y(F)$. Choose norms $\|\cdot\|_{Y_i}$ on $Y_i(F)$. Choose norms $\|\cdot\|_{X_i}$ on $X_i(F)$ in such a way that the second condition of Lemma 18.3 holds for $\|\cdot\|_{Y_i}$ and $\|\cdot\|_{X_i}$. Use the construction in the sixth part of Proposition 18.1 to get a norm $\|\cdot\|_Y$ on $Y(F)$ (respectively, $\|\cdot\|_X$ on $X(F)$) from the norms $\|\cdot\|_{Y_i}$ (respectively, $\|\cdot\|_{X_i}$). With these choices it is easy to see that the second condition of Lemma 18.3 holds for $\|\cdot\|_Y$ and $\|\cdot\|_X$, proving the norm descent property for $\phi$.

We are now finished with the case of principal affine subsets. Now suppose that we have any cover of $X$ by affine open subsets $U_1, \dots, U_r$. Cover each $U_i$ by principal affine open subsets $V_{ij}$ ($j = 1, \dots, s_i$) in $X$. Of course $V_{ij}$ is then also a principal affine open subset of $U_i$.

By what has already been done we know that the norm descent property for $\phi$ is equivalent to the norm descent property for all the morphisms $\phi^{-1}V_{ij} \to V_{ij}$ and that this in turn is equivalent to the norm descent property for all the morphisms $\phi^{-1}U_i \to U_i$. This completes the proof of the first part of the proposition.

Now we prove the second part. Assume first that $\phi : Y \to X$ has a section $s : X \to Y$. Note that in this case $Y(F) \to X(F)$ is surjective. To show that $\phi$ has the norm descent property we must check that $\phi_*\| \cdot \|_Y$ is dominated by $\| \cdot \|_X$. But this follows from the first part of Proposition 18.1, applied to the morphism $s$. Furthermore, if $\phi$ admits sections Zariski locally, we see from the first part of Proposition 18.2 that $\phi$ has the norm descent property.

Now we prove the third part. Choose a parabolic subgroup $P = MU$ with Levi component $M$ and let $\bar{P} = M\bar{U}$ be the opposite parabolic subgroup. From Bruhat theory we know that multiplication induces an open immersion

$$\bar{U} \times U \times M \hookrightarrow G.$$

Thus $G \to G/M$ has a section over the open subset $\bar{U} \times U$ of $G/M$. Moreover $G/M$ is covered by the $G(F)$-translates of these open sets $\bar{U} \times U$, so we conclude that $G \to G/M$ admits sections locally in the Zariski topology, hence that it has the norm descent property. Note that $G/M$ really is an affine scheme: it can be identified with the $G$-conjugacy class of any sufficiently general element of the center of $M$. $\square$

**18.7. An additional hypothesis.** We now fix an algebraic closure $\bar{F}$ of $F$. Our given absolute value on $F$ can always be extended to $\bar{F}$, and when $F$ is complete this extension is unique (see [**Lan02**]).

From now on we assume that $F$ is complete, and we continue to denote by $|\cdot|$ the unique extension to $\bar{F}$ of our given absolute value on $F$. By uniqueness this extension is fixed by any automorphism of $\bar{F}$ over $F$, and therefore the restriction of our extended absolute value to any finite extension $E$ of $F$ (in $\bar{F}$) of degree $n$ is given by

$$x \mapsto |N_{E/F}(x)|^{1/n}, \tag{18.7.1}$$

where $N_{E/F}$ denotes the usual norm map of field theory.

**18.8. Behavior of norms under algebraic field extensions.** Let $X$ be an affine scheme of finite type over $F$, and let $E$ be a field extension of $F$. Any finite set of generators for the $F$-algebra $\mathcal{O}_X$ can also be regarded as a generating set for the $E$-algebra $E \otimes_F \mathcal{O}_X$ of regular functions on the scheme $X_E$ over $E$ obtained from $X$ by extension of scalars. When $E$ is a subfield of $\bar{F}$ the chosen generating set gives norms on both $X(F)$ and $X(E) = X_E(E)$, and the restriction of the norm on $X(E)$ to the subset $X(F)$ coincides with the norm on $X(F)$, from which it follows that the restriction of any norm on $X(E)$ is a norm on $X(F)$.

For finite separable extensions $E/F$ we use $R_{E/F}$ to denote Weil's restriction of scalars.

Lemma 18.5. *Let $E$ be a finite separable field extension of $F$. Then the abstract norm*

$$x \mapsto \sup\{|N_{E/F}(x)|, |N_{E/F}(x)|^{-1}\} \qquad (x \in E^\times) \tag{18.8.1}$$

*is a norm on $(R_{E/F}\mathbb{G}_m)(F) = E^\times$.*

Proof. Let $I$ be the set of embeddings of $E$ in $\bar{F}$. Since our torus becomes split over $\bar{F}$, we see from Example 18.2 and the discussion preceding this lemma that

$$x \mapsto \sup\{|\sigma(x)|, |\sigma(x)|^{-1} : \sigma \in I\} \qquad (x \in E^\times) \tag{18.8.2}$$

is a norm on $(R_{E/F}\mathbb{G}_m)(F) = E^\times$. But for any $\sigma \in I$ we have (by (18.7.1))

$$|\sigma(x)| = |N_{E/F}(x)|^{1/[E:F]},$$

showing that the norm (18.8.2) is indeed equivalent to the abstract norm in the statement of the lemma. □

Lemma 18.6. *Let $T$ be a torus over $F$, and let $S$ be the biggest split quotient of $T$, so that $X^*(S)$ is the subgroup of $X^*(T)$ consisting of elements fixed by $\mathrm{Gal}(\bar{F}/F)$, and we have a canonical homomorphism $T \to S$. Then the pullback via $T(F) \to S(F)$ of any norm $\|\cdot\|_S$ on $S(F)$ is a norm on $T(F)$.*

Proof. One sees easily from Lemma 18.5 that our current lemma is valid for $T = R_{E/F}T_0$ for any finite separable extension $E/F$ and any split torus $T_0$ over $E$.

Now consider any torus $T$ and choose a finite Galois extension $E/F$ that splits $T$. Then $T$ embeds naturally in $R_{E/F}T_E$, a torus for which the lemma is known to be valid, and thus it now suffices to show that if $T$ is a subtorus of a torus $T'$ for which the lemma is known to be valid, then the lemma is valid for $T$. We of course write $S'$ for the biggest split quotient torus of $T'$.

Then we have a commutative diagram

$$\begin{array}{ccc} T & \longrightarrow & T' \\ \big\downarrow & & \big\downarrow \\ S & \longrightarrow & S'. \end{array}$$

Choose a norm $\|\cdot\|_{S'}$ on $S'(F)$. By our assumption on $T'$ and the second part of Proposition 18.1 (applied to $T \hookrightarrow T'$) we see that the pullback of $\|\cdot\|_{S'}$ to $T(F)$ can serve as our norm on $T(F)$. Going the other way around the commutative square above, we conclude (using the first part of Proposition 18.1 to see that the pullback of $\|\cdot\|_{S'}$ to $S(F)$ is dominated by $\|\cdot\|_S$) that $\|\cdot\|_T$ is dominated by the pullback to $T(F)$ of $\|\cdot\|_S$. But (again by the first part of Proposition 18.1) $\|\cdot\|_T$ also dominates the pullback of $\|\cdot\|_S$, hence is equivalent to it. □

Corollary 18.7. *Let $T$ be a torus over $F$ and let $n$ be a positive integer. Then there exists a bounded subset $B \subset T(F)$ such that $T(F) = B \cdot T(F)^n$. Here we are using the superscript $n$ to indicate that we are looking at the subgroup of all $n$-th powers of elements in $T(F)$.*

Proof. First we treat the split case. So suppose $T = \mathbb{G}_m^r$ and use the norm $\|\cdot\|_T$ in Example 18.2. Pick $\alpha \in F^\times$ such that $|\alpha| \neq 1$ and put $a := |\alpha|$. The bounded set $\{t \in T(F) : \|t\|_T \leq a^n\}$ does the job.

Now we treat the general case. Let $\phi : T \to S$ be the maximal split quotient of $T$, and let $A_T$ denote the maximal split subtorus of $T$. The composition of the inclusion of $A_T$ in $T$ with $\phi$ yields an isogeny $\psi : A_T \to S$, and therefore there exists a positive integer $m$ such that the $m$-th power map on $S$ factors through $\psi$. This guarantees that $\phi(T(F)) \supset S(F)^m$.

By the split case that has already been treated we know that there exists a bounded subset $B_0$ of $S(F)$ such that $S(F) = B_0 \cdot S(F)^{mn}$. Let $B$ denote the inverse image of $B_0$ under $\phi : T(F) \to S(F)$; by Lemma 18.6 the set $B$ is bounded in $T(F)$, and it is immediate that $T(F) = B \cdot T(F)^n$. □

**18.9. Torsors, quotients and a technical lemma.** In this subsection we prove a technical lemma that will be needed in the next subsection. We begin by reviewing some material from SGA 3 on torsors and quotients.

In this subsection absolute values will not appear, and $F$ will denote any field. By a scheme (or morphism) we always mean a scheme (or morphism) over $F$. Given two schemes $X$, $Y$, we denote their product over $F$ simply by $X \times Y$, and we write $X(Y)$ for the set of $Y$-valued points of $X$, or, in other words, the set of morphisms from $Y$ to $X$.

An action of a group scheme $G$ on a scheme $X$ is a morphism $G \times X \to X$ such that for every scheme $T$ the associated map

$$G(T) \times X(T) \to X(T)$$

on $T$-valued points is an action of the group $G(T)$ on the set $X(T)$.

Now suppose that $X$ is a scheme over another scheme $S$, so that it comes equipped with a morphism $p : X \to S$. We say that an action of $G$ on $X$ *preserves the fibers of* $p : X \to S$ (or that $G$ acts on $X$ *over* $S$) if

$$p(gx) = p(x)$$

for all schemes $T$ and all $T$-valued points $g \in G(T)$, $x \in X(T)$. Given an action of this type there is a canonical morphism

$$G \times X \to X \times_S X \tag{18.9.1}$$

given by $(g, x) \mapsto (gx, x)$ on $T$-valued points.

By definition a *$G$-torsor $X$ over $S$* (for the fpqc topology) is a faithfully flat, quasi-compact morphism $p : X \to S$, together with an action of $G$ on $X$ over $S$ for which (18.9.1) is an isomorphism. The significance of (18.9.1) being an isomorphism is easy to understand: it means that for every scheme $T$ either $X(T)$ is empty or else $G(T)$ acts simply transitively on $X(T)$. Considering the fiber product of $p$ with itself one sees that any property of the morphism $G \to \operatorname{Spec} F$ that is stable under base change and faithfully flat descent will be inherited by the morphism $p : X \to S$; for example, if $G$ is smooth over $F$, then any $G$-torsor $X$ over $S$ is smooth over $S$, and so on.

Let $p : X \to S$ be a $G$-torsor. By faithfully flat descent for the morphism $p$, giving a morphism from $S$ to some other scheme $S'$ is the same as giving a morphism $X \to S'$ whose compositions with the two projections $\pi_1, \pi_2 : X \times_S X \to X$ coincide, and since (18.9.1) is an isomorphism, this in turn is the same as giving a morphism $X \to S'$ whose fibers are preserved by the $G$-action. In other words $p : X \to S$ satisfies the universal property one expects of a quotient of $X$ by $G$. In particular, given an action of $G$ on $X$, if there exists a morphism $p : X \to S$ for which $X$ is a

$G$-torsor over $S$, then the morphism $p$ is essentially unique, and we will refer to $S$ as the *quotient* of $X$ by $G$, denoted $G\backslash X$.

An important (and non-trivial) result is that if $G$ is a group scheme of finite type over $F$, and if $H$ is a closed subgroup scheme of $G$, then the quotient $G/H$ does exist (see SGA 3, Exp. $\mathrm{VI_A}$) in the sense just described.

Here is a simple example, which we will need later. Let $T$ be a torus over $F$, and let $n$ be a positive integer. We write $[n] : T \to T$ for the homomorphism given on $S$-valued points by $t \to t^n$, and we write $T_n$ for the kernel of $[n]$, so that

$$T_n(S) = \{t \in T(S) : t^n = 1\}.$$

When $n$ is not invertible in $F$, the scheme $T_n$ is not smooth over $F$. Nevertheless the morphism $[n]$ is faithfully flat, as one can check after extending scalars from $F$ to an algebraic closure $\bar{F}$ of $F$, so that $T$ splits and one is reduced to the obvious fact that for an indeterminate $X$ the ring $\bar{F}[X, X^{-1}]$ is free of rank $n$ as a module over its subring $\bar{F}[X^n, X^{-n}]$. Therefore the morphism $[n]$ makes $T$ into a $T_n$-torsor over $T$ and identifies $T$ with the quotient $T/T_n$, so that the sequence

$$1 \to T_n \to T \xrightarrow{[n]} T \to 1$$

is an exact sequence of sheaves in the fpqc topology. When $n$ is not invertible in $F$, the sequence above is not exact in the étale topology; indeed, the map $T(F_{\mathrm{sep}}) \to T(F_{\mathrm{sep}})$ is not surjective, $F_{\mathrm{sep}}$ being the separable closure of $F$ in $\bar{F}$. This example shows why we are using the fpqc topology.

Now let $G$ be a connected reductive group over $F$, and let $T$ be a torus in $G$, in other words, a closed subgroup scheme that is a torus over $F$. There exists a finite Galois extension $F'/F$ such that $T$ splits over $F'$; put $n := [F' : F]$ and $T_n := \ker([n] : T \to T)$, as above.

Lemma 18.8. *Assume that $F$ is an infinite field. Then the canonical morphism $f : G/T_n \to G/T$ admits sections locally in the Zariski topology on $G/T$.*

Proof. We claim that it is enough to show that there exists a non-empty Zariski open subset $U$ in $G/T$ such that $f : G/T_n \to G/T$ has a section over $U$. Indeed, since $f$ is $G$-equivariant, it will then have sections over all the open sets $gU$ ($g \in G(F)$), so it is enough to show that $V := \cup_{g\in G(F)} gU$ is equal to $G/T$. But $V$ is a non-empty $G(F)$-invariant open subset of $G/T$, so its inverse image $V'$ in $G$ is a non-empty $G(F)$-invariant open subset of $G$, and since $G(F)$ is Zariski dense in $G$ (since $F$ is infinite, see [**Bor91**, Cor. (18.3)]), it follows that $V' = G$ and $V = G/T$.

Now $G/T$ is connected (since $G$ is connected and $G \to G/T$ is surjective) and smooth (see EGA IV(17.7.7)), hence reduced and irreducible. Let $K$ be the function field of $G/T$ and $\xi : \operatorname{Spec} K \to G/T$ the generic point of $G/T$. Since $f : G/T_n \to G/T$ is a morphism of finite type, the existence of a section of $f$ over some non-empty open subset of $G/T$ is equivalent to the existence of a section of $f$ over $\xi$, in other words to the existence of a $K$-point of $G/T_n$ mapping to $\xi$ under

$$(G/T_n)(K) \to (G/T)(K). \tag{18.9.2}$$

Thus it will suffice to show that the map (18.9.2) is surjective. In fact the map (18.9.2) is surjective for any field extension $K$ of $F$, as we will now see.

Since $F'/F$ is a Galois extension of degree $n$, the $K$-algebra $K \otimes_F F'$ has the form $K' \times \cdots \times K'$ for some finite Galois extension $K'$ of $K$ whose degree divides $n$. Thus $T$ splits over $K'$ and the Galois cohomology group $H^1(K, T)$ coincides with

$H^1(\mathrm{Gal}(K'/K), T(K'))$, a group killed by the order of the group $\mathrm{Gal}(K'/K)$ and hence killed by $n$.

When $F$ (and hence $K$) has characteristic 0, we can then use the exact sequence

$$1 \to T_n(\bar{K}) \to T(\bar{K}) \xrightarrow{n} T(\bar{K}) \to 1 \tag{18.9.3}$$

to see that

$$H^1(K, T_n) \to H^1(K, T)$$

is surjective, from which it follows easily that (18.9.2) is surjective. [Use that the set of $G(K)$-orbits on $(G/T)(K)$ can be identified with the kernel of the map $H^1(K,T) \to H^1(K,G)$ of pointed sets.]

When $n$ is not invertible in $K$, we need to argue differently, since (18.9.3) is no longer exact when $\bar{K}$ is replaced by the separable closure $K_{\mathrm{sep}}$. To avoid using flat cohomology, we work directly with torsors. The group $T$ acts on the right of $G$, yielding a $T$-torsor $f' : G \to G/T$ over $G/T$. Moreover $T/T_n$ acts on the right of $G/T_n$, and our morphism $f : G/T_n \to G/T$ is the $T/T_n$-torsor obtained from $f'$ via the canonical homomorphism $T \twoheadrightarrow T/T_n$.

Using the fpqc exact sequence $1 \to T_n \to T \xrightarrow{n} T \to 1$ to identify $T/T_n$ with $T$, we see that the $T = T/T_n$-torsor $f$ is obtained from the $T$-torsor $f'$ via the homomorphism $[n] : T \to T$.

Fortunately, for $T$-torsors (unlike $T_n$-torsors) the difference between the fpqc and étale topologies is unimportant: since $T$ is smooth over $F$, any $T$-torsor over $K$ is automatically smooth over $K$, hence has sections étale locally on $\mathrm{Spec}(K)$. Therefore the group of isomorphism classes of $T$-torsors over $K$ can be identified with the Galois cohomology group $H^1(K,T)$, a group killed by $n$, as we saw above.

Now we can prove that (18.9.2) is surjective. Consider a $K$-point of $G/T$. Pulling back our torsor $f$ to this $K$-point, we get a $T = (T/T_n)$-torsor over $\mathrm{Spec}\, K$, which we just need to show is trivial (so that it has a section). But, as explained above, the class of this torsor lies in the image of multiplication by $n$, and since $H^1(K,T)$ is killed by $n$, every element in the image of multiplication by $n$ is trivial. □

**18.10. Another case of the norm descent property.** We start with the following lemma which will ensure that the varieties we deal with are affine.

Lemma 18.9. *Let $G$ be a connected reductive group over a field $F$ and let $T$ be an $F$-torus in $G$. Then $G/T$ is affine.*

Proof. By EGA IV (2.7.1) the property of being an affine morphism is stable under fpqc descent. So we are free to assume that $F$ is algebraically closed. Choose a maximal torus $T'$ of $G$ containing $T$. Then $G/T'$ can be identified with the $G$-orbit of any suitably regular element in $T'$, so $G/T'$ is affine. Moreover $G/T \to G/T'$ is a $T'/T$-torsor and hence is an affine morphism. This proves that $G/T$ is affine. □

Now we again assume that $F$ is equipped with a non-trivial absolute value and that $F$ is complete as a metric space. The next result is related to the corollary on page 112 of [**HC70**] (see also [**Art91a**, Lemma 4.1]).

Proposition 18.3. *Let $G$ be a connected reductive group over $F$ and let $T$ be an $F$-torus in $G$. Then $G \to G/T$ has the norm descent property. When $T$ is split, the same is true even if $F$ is not complete.*

Proof. Choose a finite Galois extension $F'/F$ that splits $T$, put $n := [F' : F]$ and $C := \ker([n] : T \to T)$. Then $G \to G/T$ factorizes as

$$G \to G/C \to G/T.$$

For $g \in G(\bar{F})$ we write $\bar{g}$ (respectively, $\dot{g}$) for the image of $g$ in $(G/C)(\bar{F})$ (respectively, $(G/T)(\bar{F})$).

Since our absolute value is non-trivial, the field $F$ is infinite, and Lemma 18.8 says that $G/C \to G/T$ has sections Zariski locally and hence satisfies the norm descent property. When $T$ is split we may take $F' = F$, so that $C$ is trivial and we are done.

From now on we assume that $F$ is complete. Since $G \to G/C$ is finite, it too satisfies the norm descent property, as is clear from the first part of Proposition 18.1. So $G \to G/T$ is the composition of two morphisms, both of which satisfy the norm descent property. Nevertheless, since $G(F) \to (G/C)(F)$ need not be surjective, we cannot apply Lemma 18.4, and in fact it requires a bit of effort to prove the norm descent property for $G \to G/T$.

In doing so we are free to use any convenient norms on $G(F)$ and $(G/T)(F)$. We begin by picking any norm $\|\cdot\|_{G/C}$ on $(G/C)(F)$. Since $G \to G/C$ is finite, by the first part of Proposition 18.1 the pullback of $\|\cdot\|_{G/C}$ to $G(F)$ can serve as our norm $\|\cdot\|_G$ on $G(F)$. Thus

$$\|g\|_G = \|\bar{g}\|_{G/C} \qquad \forall\, g \in G(F). \tag{18.10.1}$$

Since $G/C \to G/T$ satisfies the norm descent property, we may (by Lemma 18.3) choose our norm $\|\cdot\|_{G/T}$ on $(G/T)(F)$ so that for all $y$ in the image of $(G/C)(F)$ in $(G/T)(F)$ there exists $z \in (G/C)(F)$ such that $z \mapsto y$ and

$$\|z\|_{G/C} \le \|y\|_{G/T}. \tag{18.10.2}$$

When $T/C$ is identified with $T$ via the fpqc exact sequence

$$1 \to C \to T \xrightarrow{[n]} T \to 1,$$

the canonical map $T \to T/C$ becomes the map $[n] : T \to T$. Thus it follows from Corollary 18.7 that there exists a bounded subset $B \subset (T/C)(F)$ such that $(T/C)(F) = B \cdot (T(F)/C(F))$.

From the seventh part of Proposition 18.1 we see that there exist $d, R > 0$ such that

$$\|\bar{g}b^{-1}\|_{G/C} \le d\|\bar{g}\|^R_{G/C} \qquad \forall\, \bar{g} \in (G/C)(F),\, b \in B. \tag{18.10.3}$$

Now we are ready to prove the norm descent property for $G \to G/T$. For this it will suffice to show that for any $g \in G(F)$ there exists $t \in T(F)$ such that

$$\|gt\|_G \le d\|\dot{g}\|^R_{G/T} \tag{18.10.4}$$

(with $d, R$ as chosen above). By (18.10.2) there exists $s \in (T/C)(F)$ such that

$$\|\bar{g}s\|_{G/C} \le \|\dot{g}\|_{G/T}. \tag{18.10.5}$$

Now write

$$s = tb \tag{18.10.6}$$

for some $t \in T(F)/C(F)$ and $b \in B$. Then, using successively (18.10.1), (18.10.6), (18.10.3), (18.10.5), we see that

$$\|gt\|_G = \|\bar{g}t\|_{G/C} = \|\bar{g}sb^{-1}\|_{G/C} \le d\|\bar{g}s\|_{G/C}^R \le d\|\dot{g}\|_{G/T}^R$$

as desired. $\square$

COROLLARY 18.10. *Again assume the field $F$ is complete. Let $G$ be a connected reductive group over $F$, let $T$ be an $F$-torus in $G$, and let $A_T$ be the maximal split torus in $T$. Consider the canonical morphism $\phi : G/A_T \to G/T$. Then for any norm $\|\cdot\|_{G/T}$ on $(G/T)(F)$ the pullback of $\|\cdot\|_{G/T}$ via $\phi$ is a norm on $(G/A_T)(F)$.*

PROOF. Since $H^1(F, A_T)$ is trivial, the map $G(F) \to (G/A_T)(F)$ is surjective. Moreover, it follows from Proposition 18.3 that $G \to G/T$ satisfies the norm descent property. Therefore, the second part of Lemma 18.4 tells us that $G/A_T \to G/T$ has the norm descent property, which (by Lemma 18.3) means that we can choose our norms $\|\cdot\|_{G/A_T}$ and $\|\cdot\|_{G/T}$ so that for any $y \in (G/T)(F)$ that lies in the image of $(G/A_T)(F)$, there exists $z \in (G/A_T)(F)$ such that $z \mapsto y$ and

$$\|z\|_{G/A_T} \le \|y\|_{G/T}. \tag{18.10.7}$$

Since the biggest split quotient of $T/A_T$ is trivial, we see from Lemma 18.6 that $(T/A_T)(F)$ is bounded. Therefore the seventh part of Proposition 18.1 tells us that there exist $c, R > 0$ such that

$$\|xu\|_{G/A_T} \le c\|x\|_{G/A_T}^R \qquad \forall\, x \in (G/A_T)(F),\, u \in (T/A_T)(F). \tag{18.10.8}$$

Since $(T/A_T)(F)$ acts simply transitively on any non-empty fiber of

$$(G/A_T)(F) \to (G/T)(F),$$

we see from (18.10.7) and (18.10.8) that the inequality $\|\cdot\|_{G/A_T} \le c(\phi^*\|\cdot\|_{G/T})^R$ holds on $(G/A_T)(F)$. Dominance in the other direction follows from the first part of Proposition 18.1. Thus the pullback of $\|\cdot\|_{G/T}$ is equivalent to $\|\cdot\|_{G/A_T}$, as we wished to show. $\square$

**18.11. Norms on split $p$-adic $G$.** We now let $G$ be a split group over a $p$-adic field $F$ (actually over $\mathcal{O}$). As usual we put $K := G(\mathcal{O})$, fix a split maximal torus $A$ over $\mathcal{O}$, and put $\mathfrak{a} := X_*(A) \otimes_{\mathbb{Z}} \mathbb{R}$. We choose a Weyl group invariant inner product on $\mathfrak{a}$, so that $\mathfrak{a}$ becomes a Euclidean space with Euclidean norm $\|\cdot\|_E$. The subscript is supposed to remind us that this is a Euclidean norm, not the sort of norm that we've been discussing in this section.

We define an abstract norm $\|\cdot\|_G$ on $G(F)$ as follows. Let $g \in G(F)$. By the Cartan decomposition there is a unique dominant coweight $\nu$ such that $g \in K\pi^\nu K$. Put $\|g\|_G = \exp(\|\nu\|_E)$.

LEMMA 18.11. *The abstract norm $\|\cdot\|_G$ is a valid norm on $G(F)$.*

PROOF. Pick any norm $\|\cdot\|'_G$ on $G(F)$. We must show that $\|\cdot\|_G$ and $\|\cdot\|'_G$ are equivalent. By the seventh part of Proposition 18.1 there exist positive constants $c, R$ such that

$$\|k_1 g k_2\|'_G \le c(\|g\|'_G)^R \tag{18.11.1}$$

for all $g \in G$, $k_1, k_2 \in K$. Thus for $a \in A(F)$ and $g \in KaK$ there are inequalities

$$\|g\|'_G \le c(\|a\|'_G)^R \tag{18.11.2}$$

and

$$\|a\|'_G \le c(\|g\|'_G)^R \tag{18.11.3}$$

In Example 18.2 we wrote down one valid norm on $A(F)$. It is easy to see that this norm is equivalent to the restriction of $\|\cdot\|_G$ to $A(F)$. Therefore (by the second part of Proposition 18.1) the restrictions of $\|\cdot\|'_G$ and $\|\cdot\|_G$ to $A(F)$ are equivalent. This, together with the inequalities (18.11.2) and (18.11.3), shows that $\|\cdot\|_G$ and $\|\cdot\|'_G$ are equivalent. $\square$

## 19. Another kind of norm on affine varieties over local fields

The norms introduced in section 18 are good for measuring "how big" points are, or, in other words, how close they are to $\infty$, and can therefore be used to measure growth rates of functions. In this section we discuss another kind of norm, the most important of which (namely $\mathbf{N}_{x_0}$) measures how close a point is to some given point $x_0$. In order to prove one of the properties of $\mathbf{N}_{x_0}$ it is useful to introduce a more general variant $N_Y$ which measures how close a point is to some given reduced closed subscheme $Y$.

Again $F$ denotes a field equipped with an absolute value.

**19.1. The norm $N_Y$.** Let $A$ be a finitely generated $F$-algebra and put $X := \operatorname{Spec}(A)$. Let $Y$ be a reduced closed subscheme of $X$, and let $I = I(Y)$ be the corresponding ideal in $A$. Thus $I$ is equal to its radical.

Now choose a finite set $f_1, \dots, f_r$ of generators for the ideal $I$. For $x \in X(F)$ put

$$N_Y(x) := \sup\{|f_1(x)|, \dots, |f_r(x)|\}. \tag{19.1.1}$$

Thus $N_Y(x)$ is a non-negative real-valued function of $x \in X(F)$ which vanishes if and only if $x \in Y(F)$. The size of $N_Y(x)$ measures how far $x$ is from $Y(F)$.

LEMMA 19.1. *Let $f \in I$. There exists a norm $\|\cdot\|_X$ on $X(F)$, of the type considered in the previous section, having the property that*

$$|f(x)| \le \|x\|_X \cdot N_Y(x) \quad \forall\, x \in X(F). \tag{19.1.2}$$

PROOF. Choose elements $g_1, \dots, g_r \in A$ such that $f = g_1 f_1 + \dots + g_r f_r$. Then

$$|f(x)| \le \sum_{i=1}^{r} |g_i(x)||f_i(x)|. \tag{19.1.3}$$

Using (18.2.2) for all the functions $g_i$, we see that there exists a norm $\|\cdot\|_X$ on $X(F)$ such that

$$|g_i(x)| \le r^{-1}\|x\|_X \quad \forall\, i \in \{1, \dots, r\}. \tag{19.1.4}$$

The inequality (19.1.2) follows directly from the inequalities (19.1.3) and (19.1.4). $\square$

Now suppose that we have two affine schemes $X_1$, $X_2$ of finite type over $F$, as well as two reduced closed subschemes $Y_1$, $Y_2$ of $X_1$, $X_2$ respectively. Choose finite generating sets for the ideals $I_1 := I(Y_1)$, $I_2 := I(Y_2)$, obtaining in this way $N_1 := N_{Y_1}$, $N_2 := N_{Y_2}$ on $X_1(F)$, $X_2(F)$ respectively. Suppose further that we are given a morphism $\phi : X_1 \to X_2$ of $F$-schemes such that $\phi(Y_1) \subset Y_2$ (equivalently: $\phi^*(I_2) \subset I_1$).

LEMMA 19.2. *There exists a norm $\|\cdot\|_{X_1}$ on $X_1(F)$, of the type considered in the previous section, having the property that*

$$N_2(\phi(x_1)) \le \|x_1\|_{X_1} \cdot N_1(x_1) \quad \forall\, x_1 \in X_1(F). \tag{19.1.5}$$

Proof. Say $N_1$ is obtained from generators $f_1, \dots, f_r$ of $I_1$ and that $N_2$ is obtained from generators $g_1, \dots, g_s$ of $I_2$. Applying Lemma 19.1 to the functions $\phi^*(g_j)$, we see that there exists a norm $\|\cdot\|_{X_1}$ on $X_1(F)$ such that

$$|g_j(\phi(x_1))| \le \|x_1\|_{X_1} \cdot N_1(x_1) \quad \forall\, x_1 \in X_1(F) \tag{19.1.6}$$

for all $j = 1, \dots, s$; since $N_2(\phi(x_1))$ is the maximum of the quantities appearing on the left side of (19.1.6), the lemma is proved. □

**19.2. The norm $\mathbf{N}_{x_0}$.** We continue with $X = \operatorname{Spec}(A)$ as above. Let $x_0 \in X(F)$ and consider the corresponding reduced closed subscheme $\{x_0\}$ of $X$, whose ideal is the maximal ideal $\mathfrak{m} := \{f \in A : f(x_0) = 0\}$. We are now interested in generating sets of $\mathfrak{m}$ of the following special type. Consider a finite set $f_1, \dots, f_r$ of generators for the $F$-algebra $A$ having the property that each $f_i$ lies in the maximal ideal $\mathfrak{m}$. Such generating sets exist, since we can start with an arbitrary generating set and subtract from each generator its value at $x_0$. It is easy to see that $f_1, \dots, f_r$ necessarily generate the ideal $\mathfrak{m}$. Now define $\mathbf{N}_{x_0}$ to be the function $N_{\{x_0\}}$ obtained from the generating set $f_1, \dots, f_r$. We use the notation $\mathbf{N}_{x_0}$ to keep track of the fact that $f_1, \dots, f_r$ not only generate $\mathfrak{m}$ as ideal, but also $A$ as $F$-algebra.

Since $f_1, \dots, f_r$ generate $A$ as $F$-algebra, they define a closed embedding of $X$ into $\mathbb{A}^r$, and from this one sees easily that the sets

$$\{x \in X(F) : \mathbf{N}_{x_0}(x) < \varepsilon\} \tag{19.2.1}$$

for $\varepsilon > 0$ form a neighborhood base at $x_0$ in $X(F)$.

**19.3. An application.** Now let $G$ be a reduced affine group scheme of finite type over $F$, and let $H$ be a closed subgroup scheme of $G$. We write $e_G$, $e_H$ for the identity elements of $G(F)$, $H(F)$ respectively. It is evident that $g^{-1}hg$ is close to $e_G$ if $h$ is close to $e_H$, but the bigger $g$ is, the closer to $e_H$ we must take $h$ to be. In the proof of the key geometric result needed for the local trace formula we are going to need a quantitative version of this qualitative statement, involving the functions $\mathbf{N}_{x_0}$ we have just introduced.

We write $\mathcal{O}_G$, $\mathcal{O}_H$ for the rings of regular functions on $G$, $H$ respectively. Choose generators $f_1, \dots, f_r$ of the $F$-algebra $\mathcal{O}_H$ such that $f_i(e_H) = 0$ for all $i = 1, \dots, r$, so that $f_1, \dots, f_r$ also generate the maximal ideal obtained from $e_H$, and use these generators to get the function $\mathbf{N}_{e_H}$ on $H(F)$. Similarly, choose generators $g_1, \dots, g_s$ of the $F$-algebra $\mathcal{O}_G$ such that $g_j(e_G) = 0$ for all $j = 1, \dots, s$, and use them to get the function $\mathbf{N}_{e_G}$ on $G(F)$. Finally, let $\|\cdot\|_G$ be any norm on $G(F)$ of the type considered in section 18.

Lemma 19.3. *Let $K$ be a neighborhood of $e_G$ in $G(F)$. Then there exist positive constants $c, R$, depending on $K$, having the following property. For $g \in G(F)$ and $h \in H(F)$ satisfying*

$$\mathbf{N}_{e_H}(h) \le c\|g\|_G^{-R} \tag{19.3.1}$$

*the element $g^{-1}hg$ lies in $K$.*

Proof. Since $f_1, \dots, f_r$ generate the $F$-algebra $\mathcal{O}_H$, we can also use them to get a norm $\|\cdot\|_H$ on $H(F)$ of the type considered in section 18; comparing the definitions of $\|\cdot\|_H$ and $\mathbf{N}_{e_H}$ one sees immediately that

$$\|h\|_H = \sup\{1, \mathbf{N}_{e_H}(h)\}. \tag{19.3.2}$$

Since $G$ is reduced, the pullbacks of the functions $f_1, \dots, f_r$ to $G \times H$ via the second projection map $(g, h) \mapsto h$ generate the ideal of the closed subset $G \times \{e_H\}$ in $G \times H$, so we can use them to define the function $N_{G\times\{e_H\}}$ on $G(F) \times H(F)$, and it is evident from the definitions that

$$N_{G\times\{e_H\}}(g, h) = \mathbf{N}_{e_H}(h). \tag{19.3.3}$$

Noting that $\|g\|_G\|h\|_H$ is a valid norm on $G(F) \times H(F)$ by Proposition 18.1, and applying Lemma 19.2 to the morphism $G \times H \to G$ defined by $(g, h) \mapsto g^{-1}hg$, we see that there exist $c_1, R > 0$ such that

$$\mathbf{N}_{e_G}(g^{-1}hg) \le c_1(\|g\|_G\|h\|_H)^R \mathbf{N}_{e_H}(h). \tag{19.3.4}$$

Now choose $\varepsilon > 0$ small enough that $\mathbf{N}_{e_G}(g) \le \varepsilon$ implies that $g \in K$. Then

$$c_1(\|g\|_G\|h\|_H)^R \mathbf{N}_{e_H}(h) \le \varepsilon \Longrightarrow g^{-1}hg \in K. \tag{19.3.5}$$

From (19.3.2) we see that $\|h\|_H = 1$ when $\mathbf{N}_{e_H}(h) \le 1$, so that

$$\mathbf{N}_{e_H}(h) \le 1 \text{ and } c_1(\|g\|_G)^R \mathbf{N}_{e_H}(h) \le \varepsilon \Longrightarrow g^{-1}hg \in K, \tag{19.3.6}$$

or, in other words,

$$\mathbf{N}_{e_H}(h) \le \inf\{1, \varepsilon c_1^{-1}\|g\|_G^{-R}\} \Longrightarrow g^{-1}hg \in K. \tag{19.3.7}$$

Letting $c$ be the minimum of 1 and $\varepsilon c_1^{-1}$ (and remembering that $\|g\|_G \ge 1$), we see that

$$\mathbf{N}_{e_H}(h) \le c\|g\|_G^{-R} \Longrightarrow g^{-1}hg \in K. \tag{19.3.8}$$

□

**19.4. Special case of the application above.** We will need a more concrete version of the previous lemma that is tailored to the situation we will encounter in proving the local trace formula. We now return to the split reductive group $G$ over the $p$-adic field $F$, and we fix a norm $\|\cdot\|_G$ on $G(F)$ as in section 18.

Consider a parabolic subgroup $P = MU$ containing a Borel subgroup $B = AN$, with $M$ containing $A$. We are going to apply the lemma we just proved to the subgroup $U$ of $G$. As a variety, $U$ is the product of its root subgroups $U_\alpha$, where $\alpha$ runs over $R_U$, the set of roots of $A$ in Lie$(U)$. Fix identifications $U_\alpha \simeq \mathbb{G}_a$, so that we can view an element $u \in U(F)$ as a tuple with components $u_\alpha \in F$, one for each $\alpha \in R_U$. Using the most obvious set of generators, we get $\mathbf{N}_{e_U}$ on $U(F)$, given by

$$\mathbf{N}_{e_U}(u) = \sup\{|u_\alpha| : \alpha \in R_U\}. \tag{19.4.1}$$

Conjugating $u$ by $a \in A(F)$, we get another element $aua^{-1}$ of $U(F)$ whose components are given by $\alpha(a) \cdot u_\alpha$. Therefore

$$\mathbf{N}_{e_U}(aua^{-1}) \le \sup\{|\alpha(a)| : \alpha \in R_U\} \cdot \mathbf{N}_{e_U}(u). \tag{19.4.2}$$

LEMMA 19.4. *Let $K = G(\mathcal{O})$. There exist positive constants $D, R, S$ having the following property. For all $a \in A(F)$, $u \in U(F)$, $g \in G(F)$ satisfying*

$$\inf\{|\alpha(a)|^{-1} : \alpha \in R_U\} \ge D\|u\|_G^R\|g\|_G^S \tag{19.4.3}$$

*the element $aua^{-1}$ lies in $gKg^{-1}$.*

PROOF. Using Lemma 19.3 and (19.4.2), we see that there exist positive constants $c, S$ such that for $a \in A(F)$, $u \in U(F)$, $g \in G(F)$

$$\sup\{|\alpha(a)| : \alpha \in R_U\} \cdot \mathbf{N}_{e_U}(u) \le c\|g\|_G^{-S} \Longrightarrow aua^{-1} \in gKg^{-1}. \tag{19.4.4}$$

As we remarked during the course of the proof of Lemma 19.3, we get a valid norm $\|\cdot\|_U$ on $U(F)$ by putting $\|u\|_U := \sup\{1, \mathbf{N}_{e_U}(u)\}$, and thus (19.4.4) yields

$$\sup\{|\alpha(a)| : \alpha \in R_U\} \le c\|u\|_U^{-1}\|g\|_G^{-S} \Longrightarrow aua^{-1} \in gKg^{-1}. \tag{19.4.5}$$

Since the restriction of $\|\cdot\|_G$ to $U(F)$ is also a valid norm, and moreover any two norms on $U(F)$ are equivalent, we conclude that there exist positive constants $D, R$ such that

$$\sup\{|\alpha(a)| : \alpha \in R_U\} \le D^{-1}\|u\|_G^{-R}\|g\|_G^{-S} \Longrightarrow aua^{-1} \in gKg^{-1}, \tag{19.4.6}$$

from which the conclusion of the lemma follows immediately. $\square$

## 20. Estimates for weighted orbital integrals

In this section we work with a maximal torus $T$ in a connected reductive group $G$ over our $p$-adic field $F$. As usual $D(X)$ is the polynomial function on $\mathfrak{g}$ (see 7.5) that turns up in the Weyl integration formula. We are going to prove estimates for weighted orbital integrals. This will use the theory of norms on affine varieties that was developed in section 18.

**20.1. Local integrability of various functions on $\mathfrak{t}$.** We are interested in the local integrability of various functions on $\mathfrak{t}$ that involve the function $X \mapsto |D(X)|$.

LEMMA 20.1 ([**HC70**, Lemma 44]). *There exists $\epsilon > 0$ such that the function $|D(X)|^{-\epsilon}$ is locally integrable on $\mathfrak{t}$.*

PROOF. The polynomial $D$ is homogeneous of degree $d$, where $d = d_G$ is the number of roots of $T$ in $\mathfrak{g}$. We claim that $|D(X)|^{-\epsilon}$ is locally integrable on $\mathfrak{t}$ provided that $d\epsilon < 1$. We prove this statement by induction on $d$, the case $d = 0$ being trivial.

Now we assume that $d > 0$ and that the statement we are trying to prove is true for all smaller $d$. There is an immediate reduction to the case in which $G$ is semisimple. Let $S$ be any non-zero element in $\mathfrak{t}$, and let $H$ denote its centralizer in $G$. Since the functions $|D^G|$ and $|D^H|$ are positive multiples of each other in some small neighborhood of $S$ in $\mathfrak{t}$, we conclude by our induction hypothesis (using that $d_H < d_G$ and hence that $d_H\epsilon < 1$) that $|D(X)|^{-\epsilon}$ is locally integrable on this neighborhood of $S$. Since this is true for all non-zero $S$, we conclude that $|D(X)|^{-\epsilon}$ is locally integrable on $\mathfrak{t} \setminus \{0\}$.

It remains to show that $|D(X)|^{-\epsilon}$ is integrable on some open neighborhood of 0. For convenience we take this neighborhood to be a lattice $L$. It is enough to show that

$$\int_L |D(X)|^{-\epsilon}\, dX \tag{20.1.1}$$

is finite. Since $D$ is homogeneous of degree $d$, this integral is equal to the product of

$$\int_{L\setminus \pi L} |D(X)|^{-\epsilon}\, dX \tag{20.1.2}$$

and the geometric series with ratio $|\pi|^{\dim(\mathfrak{t})-d\epsilon}$. The geometric series is convergent by our assumption that $d\epsilon < 1$, and the integral (20.1.2) is convergent since $|D(X)|^{-\epsilon}$ is locally integrable away from the origin and hence integrable on the complement of $\pi L$ in $L$.

It is worth noting that this result of Harish-Chandra can also be derived from rather general results of Igusa [**Igu74, Igu77**] and Denef [**Den84**] on integrals of complex powers of absolute values of $p$-adic polynomials. $\square$

The next result involves the function

$$X \mapsto \log(\max\{1, |D(X)|^{-1}\})$$

on $\mathfrak{g}_{\mathrm{rs}}$. This function takes non-negative real values and measures how close $X$ is to the singular set $\mathfrak{g} \setminus \mathfrak{g}_{\mathrm{rs}}$: the larger the function value at $X$, the closer $X$ is to the singular set.

Corollary 20.2. *For every non-negative real number $R$ the function*

$$X \mapsto \big(\log(\max\{1, |D(X)|^{-1}\})\big)^R$$

*is locally integrable on* $\mathfrak{t}$.

Proof. This follows from the previous result together with the following elementary fact: for every $\epsilon > 0$ and every $R \geq 0$ there exists a positive constant $C$ such that

$$\big(\log(\max\{1, y\})\big)^R \leq C y^{\epsilon} \tag{20.1.3}$$

for all $y \geq 0$. $\square$

**20.2. Estimates for orbital integrals with various weight factors.** Now suppose that $M$ is a Levi subgroup of $G$ containing $T$. Then we have $(M\backslash G)(F) = M(F)\backslash G(F)$ and $(A_M\backslash G)(F) = A_M(F)\backslash G(F)$, so no confusion will result from writing $M\backslash G$ for the set of $F$-points of the affine algebraic variety $M\backslash G$, and similarly for $A_M\backslash G$. Let $\|\cdot\|_{M\backslash G}$ and $\|\cdot\|_{A_M\backslash G}$ be any norms (as in 18.2) on $M\backslash G$ and $A_M\backslash G$ respectively.

We are also interested in the affine algebraic variety $T\backslash G$, but here we need to be more careful, since $(T\backslash G)(F)$ can be bigger than $T(F)\backslash G(F)$. We let $\|\cdot\|_{T\backslash G}$ be any norm on $(T\backslash G)(F)$. Having warned the reader of the potential for confusion, we nevertheless now write $T\backslash G$ as a convenient abbreviation for $T(F)\backslash G(F)$, an open and closed subset of $(T\backslash G)(F)$. We will only have occasion to use $\|\cdot\|_{T\backslash G}$ on the subset $T\backslash G$.

Before formulating the next results, let's discuss where we're headed. Let $X \in \mathfrak{t}_{\mathrm{reg}}$ and consider the weighted orbital integral

$$\int_{T\backslash G} f(g^{-1}Xg) v_M(g)\, d\bar{g}.$$

Since $X$ is semisimple, its orbit is closed and hence intersects the support of $f$ in a compact subset of the orbit. Thus there is a compact subset $C$ of $T\backslash G$ such that the integrand vanishes unless $g \in C$. The weight factor is left $M$-invariant, hence left $T$-invariant, hence remains bounded in absolute value on the compact set $C$, say by the positive number $R$. Then the absolute value of the weighted orbital integral is bounded above by $R$ times the orbital integral of $|f|$. Now suppose that we want to estimate the weighted orbital integral for fixed $f$ and variable $X$. Then we will need to control the size of the compact set $C$, which of course depends on $X$. The point is that $C$ grows as $X$ gets closer to the singular set. The theory of norms developed in section 18 makes it easy to get such control, as we will now see.

When we apply the next lemma, the compact set $\omega$ will be the support of $f$, so this is what the reader should have in mind. The lemma is a variant of [**HC70**, Theorem 18] (see also [**Art91a**, Lemma 4.2]).

LEMMA 20.3. *Let $\omega$ be a compact subset of $\mathfrak{g}$. Then there exist positive constants $c_1$, $c_2$, $c_1'$, $c_2'$ having the following property. For all $X \in \mathfrak{t}_{\mathrm{reg}}$ and all $g \in G$ such that $g^{-1}Xg \in \omega$ there are inequalities*

$$\log \|g\|_{T\backslash G} \leq c_1 + c_2 \log \max\{1, |D(X)|^{-1}\}$$

*and*

$$\log \|g\|_{M\backslash G} \leq c_1' + c_2' \log \max\{1, |D(X)|^{-1}\}.$$

PROOF. We again consider the morphism

$$\beta : (T\backslash G) \times \mathfrak{t}_{\mathrm{reg}} \to \mathfrak{g}_{\mathrm{rs}}$$

of affine algebraic varieties defined by $\beta(g, X) = g^{-1}Xg$. Choose a norm $\|\cdot\|_{\mathfrak{g}}$ on the affine variety $\mathfrak{g}$. Now $\mathfrak{g}_{\mathrm{rs}}$ is the principal Zariski open subset of $\mathfrak{g}$ defined by the non-vanishing of the polynomial $D$, so (see Proposition 18.1) as norm $\|\cdot\|_{\mathfrak{g}_{\mathrm{rs}}}$ on $\mathfrak{g}_{\mathrm{rs}}$ we may take

$$\|X\|_{\mathfrak{g}_{\mathrm{rs}}} := \max\{\|X\|_{\mathfrak{g}}, |D(X)|^{-1}\}.$$

Since the morphism $\beta$ is finite, we may take (see Proposition 18.1) as norm on $(T\backslash G) \times \mathfrak{t}_{\mathrm{reg}}$ the pullback of $\|\cdot\|_{\mathfrak{g}_{\mathrm{rs}}}$ by $\beta$. Again by Proposition 18.1 the pullback of the norm $\|\cdot\|_{T\backslash G}$ to $(T\backslash G) \times \mathfrak{t}_{\mathrm{reg}}$ (pull back using the first projection) is dominated by the norm on $(T\backslash G) \times \mathfrak{t}_{\mathrm{reg}}$. Writing out what this means, we see that there are constants $c > 1$ and $R > 0$ such that

$$\|g\|_{T\backslash G} \leq c \max\{\|g^{-1}Xg\|_{\mathfrak{g}}, |D(X)|^{-1}\}^R$$

for all $g \in T\backslash G$ and all $X \in \mathfrak{t}_{\mathrm{reg}}$. Now $\|\cdot\|_{\mathfrak{g}}$ remains bounded on the compact set $\omega$; let's choose $d > 1$ that serves as an upper bound. Thus

$$\|g\|_{T\backslash G} \leq cd^R \max\{1, |D(X)|^{-1}\}^R \tag{20.2.1}$$

for all $g \in T\backslash G$, $X \in \mathfrak{t}_{\mathrm{reg}}$ such that $g^{-1}Xg \in \omega$. Taking the logarithm of both sides of (20.2.1), we get the first inequality of the lemma. The second inequality can be derived from the first since (again by Proposition 18.1) the pullback of the norm $\|\cdot\|_{M\backslash G}$ to $T\backslash G$ is dominated by $\|\cdot\|_{T\backslash G}$. □

Now we use the lemma to estimate weighted orbital integrals. Actually the proof of the local trace formula involves estimating orbital integrals weighted by various factors other than $v_M$, but having the same rough growth rate as $v_M$. Our next result will involve the weight factor $(\log \|\cdot\|_{T\backslash G})^R$, as this will allow us to handle all the weight factors that come up in the proof of the local trace formula.

PROPOSITION 20.1. *Let $f \in C_c^\infty(\mathfrak{g})$ and let $R$ be a non-negative integer. Then the integral*

$$\int_{\mathfrak{t}} |D(X)|^{1/2} \int_{T\backslash G} f(g^{-1}Xg)(\log \|g\|_{T\backslash G})^R \, d\bar{g} \, dX \tag{20.2.2}$$

*converges. If $T$ is elliptic in $M$, then the integral*

$$\int_{\mathfrak{t}} |D(X)|^{1/2} \int_{A_M\backslash G} f(g^{-1}Xg)(\log \|g\|_{A_M\backslash G})^R \, d\dot{g}\, dX \tag{20.2.3}$$

*converges.*

PROOF. The first statement follows from Lemma 20.3 (with $\omega = \mathrm{Supp}(f)$), Corollary 20.2, and the fact that the function $X \mapsto |D(X)|^{1/2} \int_{T\backslash G} |f(g^{-1}Xg)|\, d\bar{g}$ is bounded and compactly supported on $\mathfrak{t}$ (see Theorem 17.10). The second statement follows from the first, together with Corollary 18.10. □

In order to see that the proposition above can be applied to weighted orbital integrals, we need to estimate $v_M(g)$. Before doing so, we discuss the metric on the building of $G$.

**20.3. Metric on $X$, function $d(x)$ on $X$, estimate for $\|H_B(g)\|_E$.** We now assume that $G$ is split, with split maximal torus $A$. As in 18.11 we choose a Weyl group invariant Euclidean norm $\|\cdot\|_E$ on $\mathfrak{a}$. From $\|\cdot\|_E$ we get a metric on $\mathfrak{a}$. Viewing $\mathfrak{a}$ as the standard apartment in the building of $G$, the metric on $\mathfrak{a}$ extends uniquely to a $G$-invariant metric on the building, denoted by $d(x_1, x_2)$.

As usual we put $X = G/K$ and denote by $x_0$ the base-point of $X$. We view $X$ as a subset of the building, so it makes sense to consider the metric $d(x_1, x_2)$ for $x_1, x_2 \in X$. For $x \in X$ we introduce

$$d(x) := d(x, x_0)$$

as a measure of the size of $x$, and for $g \in G$ we also put

$$d(g) := d(gx_0).$$

The next lemma concerns the maps $H_B : G \to X_*(A) \hookrightarrow \mathfrak{a}$ defined in 12.1.

LEMMA 20.4. *Let $B \in \mathcal{B}(A)$. For all $g \in G$ there is an inequality*

$$\|H_B(g)\|_E \le d(g). \tag{20.3.1}$$

PROOF. It follows from [**BT72**, Section 4.4.4] that if $g \in K\pi^\nu K$ for $\nu \in X_*(A)$, then $H_B(g)$ lies in the convex hull of the Weyl group orbit of $\nu$. Therefore

$$\|H_B(g)\|_E \le \|\nu\|_E, \tag{20.3.2}$$

and it is clear from the definition of the function $d$ that $d(g) = \|\nu\|_E$. □

**20.4. Estimate for $v_M$.** Since we have only discussed $v_M$ in the split case, we continue to assume that $G$ is split. Then we have the norm $\|\cdot\|_G$ on $G(F)$ defined in 18.11 using the Euclidean norm $\|\cdot\|_E$ on $\mathfrak{a}$; in terms of the function $d(g)$ introduced above, we have

$$\|g\|_G = \exp(d(g)).$$

By Proposition 18.2 the morphism $G \to M\backslash G$ satisfies the norm descent property. Therefore

$$\|g\|_{M\backslash G} := \inf\{\|mg\|_G : m \in M\} \tag{20.4.1}$$

is a norm on $M\backslash G$. We use this particular norm in the next lemma.

LEMMA 20.5. *Let $M$ be a Levi subgroup of $G$ containing $A$. Then there exists a positive constant $c$ such that*

$$v_M(g) \leq c(\log \|g\|_{M\backslash G})^{\dim(A_M/A_G)} \tag{20.4.2}$$

*for all $g \in G$.*

PROOF. Let $P \in \mathcal{P}(M)$. Choose $B \in \mathcal{B}(A)$ such that $B \subset P$. By Lemma 20.4 we have

$$\|H_B(g)\|_E \leq \log \|g\|_G. \tag{20.4.3}$$

As we saw in 12.1, $H_P(g)$ is obtained as the image of $H_B(g)$ under $\mathfrak{a} \twoheadrightarrow \mathfrak{a}_M$. When we view $\mathfrak{a}_M$ as a subspace of $\mathfrak{a}$, the map $\mathfrak{a} \twoheadrightarrow \mathfrak{a}_M$ is given by orthogonal projection. Therefore

$$\|H_P(g)\|_E \leq \log \|g\|_G. \tag{20.4.4}$$

Since $v_M(g)$ is the volume of the convex hull of the points $H_P(g)$ ($P \in \mathcal{P}(M)$), there is a positive constant $c$ such that

$$v_M(g) \leq c(\log \|g\|_G)^{\dim(A_M/A_G)} \tag{20.4.5}$$

for all $g \in G$. But the function $v_M(g)$ is left-invariant under $M$, so that we may replace $\|g\|_G$ by $\|g\|_{M\backslash G}$ in the last inequality, completing the proof. □

Combining our estimate of $v_M$ with Proposition 20.1, and remembering that the pullback of $\|\cdot\|_{M\backslash G}$ to $T\backslash G$ is dominated by $\|\cdot\|_{T\backslash G}$, we obtain the following result, in which $T$ is any maximal torus in $M$.

THEOREM 20.6. *Let $f \in C_c^\infty(\mathfrak{g})$. Then the integral*

$$\int_{\mathfrak{t}} |D(X)|^{1/2} \int_{T\backslash G} f(g^{-1}Xg) v_M(g)\, d\bar{g}\, dX \tag{20.4.6}$$

*converges.*

**20.5. Estimate for $u_M$.** We continue to assume that $G$ is split. In the proof of the local trace formula we will also need an estimate for the weight factor $u_M$ appearing in our preliminary form of the local trace formula. Recall that

$$u_M(h, g; \mu) := \int_{A_M} u^\mu(h^{-1} a_M g)\, da_M. \tag{20.5.1}$$

We use the Haar measure $da_M$ on $A_M$ giving $A_M \cap K$ measure 1. Just as in 4.2 we have $A_M/(A_M \cap K) = X_*(A_M)$, and for $a \in A_M$ we denote by $\nu_a$ the image of $a$ in $A_M/(A_M \cap K) = X_*(A_M)$. Obviously

$$u_M(h, g; \mu) \leq |\{\nu \in X_*(A_M) : \exists a \in A_M \text{ with } \nu_a = \nu \text{ and } h^{-1}ag \in G^\mu\}|.$$

The previous proof used the left $M$-invariance of $v_M(g)$. Now the function $(h, g) \mapsto u_M(h, g; \mu)$ is left $(A_M \times A_M)$-invariant, but not $(M \times M)$-invariant, and our first step will be to replace $u_M$ by something larger which is $(M \times M)$-invariant. For this we use the injection $X_*(A_M) \hookrightarrow \Lambda_M$ to see that

$$u_M(h, g; \mu) \leq u'_M(h, g; \mu),$$

with $u'_M$ defined by

$$u'_M(h, g; \mu) := |\{\nu \in \Lambda_M : \exists m \in M \text{ with } H_M(m) = \nu \text{ and } h^{-1}mg \in G^\mu\}|.$$

It is evident that this function of $(h, g)$ is left invariant under $M \times M$.

Put $x = gx_0$ and $y = hx_0$. If $h^{-1}mg \in G^\mu$, then $\operatorname{inv}(mx, y) \leq \mu$ and therefore $d(mx, y) \leq \|\mu\|_E$, from which it follows (by the triangle inequality) that

$$d(mx_0) \leq d(x) + d(y) + \|\mu\|_E, \tag{20.5.2}$$

where $d$ is the function defined in 20.3. Writing $\nu$ for $H_M(m)$ and $\bar{\nu}$ for the image of $\nu$ under $\Lambda_M \to \mathfrak{a}_M \subset \mathfrak{a}$, one sees easily from the definitions that $\|\bar{\nu}\|_E \leq d(mx_0)$. Thus we have shown that

$$u'_M(h, g; \mu) \leq |\{\nu \in \Lambda_M : \|\bar{\nu}\|_E \leq d(x) + d(y) + \|\mu\|_E\}|, \tag{20.5.3}$$

from which it is clear that there is a positive constant $c$ such that

$$u'_M(h, g; \mu) \leq c(1 + d(x) + d(y) + \|\mu\|_E)^{\dim A_M}. \tag{20.5.4}$$

Now recalling that $u'_M$ is invariant under $M \times M$, we see that in the last inequality we may replace $d(x)$ by $\inf\{d(mx) : m \in M\}$, and the same for $d(y)$. Thus we have proved

Lemma 20.7. *Let $M$ be a Levi subgroup of $G$ containing $A$. Then there exists a positive constant $c$ such that*

$$u_M(h, g; \mu) \leq c(1 + \log \|g\|_{M\backslash G} + \log \|h\|_{M\backslash G} + \|\mu\|_E)^{\dim(A_M)} \tag{20.5.5}$$

*for all $g, h \in G$.*

Of course the exponent $\dim(A_M)$ could easily be improved to $\dim(A_M/A_G)$.

## 21. Preparation for the key geometric lemma

Now we begin to prepare for the proof of the key geometric result (Theorem 22.3) needed for the local trace formula. Throughout this section we fix a Borel subgroup $B = AN$ containing $A$, which we use to define positive roots, dominance, and so on. As in 20.3 $\|x\|_E$ is a $W$-invariant Euclidean norm on $\mathfrak{a}$, which we use to get the metric $d(x, y)$ on the building as well as the function $d(x)$ on $X$.

**21.1. Retractions of the building with respect to an alcove.** Given an alcove $\mathbf{a}$ in an apartment in the Bruhat-Tits building of our split group $G$, there is a retraction $r_{\mathbf{a}}$ of the building into that apartment. As usual, we are mainly interested in the subset $X = G/K$ of the building and the standard apartment (the one coming from $A$). Inside $G/K$ we have the subset $A/A \cap K$, which we identify with $X_*(A)$ by sending $\mu \in X_*(A)$ to $\pi^\mu$. Let $\mathbf{a}$ be an alcove in the apartment $\mathfrak{a} = X_*(A) \otimes_{\mathbb{Z}} \mathbb{R}$, and let $I_{\mathbf{a}}$ be the corresponding Iwahori subgroup of $G$, defined as the pointwise stabilizer of $\mathbf{a}$ in $G$. From the affine Bruhat decomposition we know that the obvious map $A/A \cap K \to I_{\mathbf{a}}\backslash G/K$ is bijective. We will regard $r_{\mathbf{a}}$ as the retraction of $G/K$ onto $X_*(A)$ defined as follows: given $g \in G/K$ we put $r_{\mathbf{a}}(g) = \mu$ if $g \in I_{\mathbf{a}}\pi^\mu K$. In other words, given a vertex $x$ in the building, $r_{\mathbf{a}}(x)$ is the unique vertex in the standard apartment having the same position relative to $\mathbf{a}$ as $x$ does.

Proposition 21.1. [**BT72**] *The retraction $r_{\mathbf{a}}$ weakly decreases distances. In other words*

$$d(r_{\mathbf{a}}(x_1), r_{\mathbf{a}}(x_2)) \leq d(x_1, x_2) \tag{21.1.1}$$

*for all $x_1, x_2 \in G/K$.*

**21.2. An easy fact about root systems.** In a moment we will need the following easy result.

LEMMA 21.1. *Let $x, y$ be dominant elements in $\mathfrak{a}$, and let $w \in W$. Then*

$$\|x - y\|_E \leq \|x - wy\|_E. \tag{21.2.1}$$

PROOF. Expanding out the two norms, we find that (21.2.1) is equivalent to the inequality $(x, y - wy) \geq 0$, and this inequality is clear since $x$ is dominant and $y - wy$ is a non-negative linear combination of positive coroots (a standard fact that follows from things we discussed in 12.8 and 12.10). □

**21.3. Something like the triangle inequality.** Recall that $X$ denotes $G/K$ and $x_0$ its base-point. Recall also the function inv from 3.4, taking values in $K\backslash G/K$, which by the Cartan decomposition we have identified with the set of dominant coweights in $X_*(A)$. For $x, y \in X$ we can also consider the distance $d(x, y)$, a coarser invariant than $\operatorname{inv}(x, y)$.

LEMMA 21.2. *Let $x, y, x', y' \in X$ and let $\lambda, \lambda'$ be the dominant coweights obtained as $\lambda := \operatorname{inv}(x, y)$ and $\lambda' := \operatorname{inv}(x', y')$. Then*

$$\|\lambda - \lambda'\|_E \leq d(x, x') + d(y, y'). \tag{21.3.1}$$

PROOF. The lemma concerns the effect on $\operatorname{inv}(x, y)$ of replacing $(x, y)$ by $(x', y')$. We can do this replacement in two steps, going from $(x, y)$ to $(x, y')$ to $(x', y')$. Thus in proving the lemma we may assume that $x = x'$ or $y = y'$, and by symmetry (note that $\operatorname{inv}(y, x) = -\operatorname{inv}(x, y)$) we may as well assume that $y = y'$. We are free to transform all our points by any convenient $g \in G$; doing so, we may assume without loss of generality that $x, y$ both lie in the standard apartment. Inside the standard apartment pick an alcove $\mathbf{a}$ containing $y$. Then $\lambda, \lambda'$ are the unique dominant elements in the Weyl group orbits of $r_{\mathbf{a}}(x)$, $r_{\mathbf{a}}(x')$ respectively, and therefore

$$\|\lambda - \lambda'\|_E \leq d(r_{\mathbf{a}}(x), r_{\mathbf{a}}(x')) \leq d(x, x'), \tag{21.3.2}$$

the first inequality coming from Lemma 21.1, the second from Proposition 21.1. □

The next result is tailor-made for use in the proof of the key geometric result needed for the local trace formula.

COROLLARY 21.3. *Let $x_1, x_2 \in X$. Let $\nu$ be a dominant coweight, and let $a \in A$ be an element whose image in $A/A \cap K = X_*(A)$ is $\nu$. Put $\lambda := \operatorname{inv}(ax_2, x_1)$. Then*

$$\|\lambda - \nu\|_E \leq d(x_1) + d(x_2). \tag{21.3.3}$$

PROOF. Since $\nu$ is dominant, we have $\operatorname{inv}(ax_0, x_0) = \nu$. Lemma 21.2 then yields

$$\|\lambda - \nu\|_E \leq d(ax_2, ax_0) + d(x_1, x_0) = d(x_2) + d(x_1). \tag{21.3.4}$$

□

## 22. Key geometric lemma

In this section we are finally going to prove Theorem 22.3, Arthur's key geometric result needed for the local trace formula (see Lemma 4.4 in [**Art91a**]). The reader may wish to skip ahead to the statement of this theorem before trying to digest the lemmas below.

As usual we are working with the split group $G$ over the $p$-adic field $F$. Recall that $X$ denotes the set $G/K$ and that $x_0$ denotes its base-point. As in 20.3, we denote by $d(x,y)$ the metric on the building obtained from some Euclidean norm on $\mathfrak{a}$. For convenience we assume that all roots have length less than or equal to 1 in this Euclidean norm. As in 20.3, for $x \in X$ we put $d(x) := d(x, x_0)$.

**22.1. Main steps in the proof.** The main steps in the proof of Theorem 22.3 are contained in Lemmas 22.1 and 22.2 below.

Consider a Borel subgroup $B = AN$ and a parabolic subgroup $P = MU$ such that $P \supset B$ and $M \supset A$. Put $B_M := B \cap M$, a Borel subgroup of $M$. We write $\Delta$ for the set of simple roots of $A$ (with respect to $B$). Then $\Delta$ is the disjoint union of $\Delta_M$ and $\Delta_U$, where $\Delta_M$ is the set of simple roots of $A$ in $M$, and $\Delta_U$ is the set of simple roots of $A$ that occur in $\mathrm{Lie}(U)$. As usual we write $R_U$ for the set of all roots of $A$ that appear in $\mathrm{Lie}(U)$. We write $\bar{P} = M\bar{U}$ for the parabolic subgroup opposite to $P$.

Recall that we have identified $X_*(A)$ with $A/(A \cap K)$ by sending $\nu$ to $\pi^\nu$. For $a \in A$ we write $\nu_a$ for the element of $X_*(A)$ corresponding to the image of $a$ in $A/(A \cap K)$. For $d \geq 0$ we denote by $A(d)$ the set of elements $a \in A$ such that $\langle \alpha, \nu_a \rangle \geq 0$ for all $\alpha \in \Delta_M$ and $\langle \alpha, \nu_a \rangle \geq d$ for all $\alpha \in \Delta_U$. Note that if $a \in A(d)$, then

$$\langle \alpha, \nu_a \rangle \geq d \quad \forall\, \alpha \in R_U, \tag{22.1.1}$$

as follows from the fact that any $\alpha \in R_U$ is a non-negative integral linear combination of roots in $\Delta$ with some root in $\Delta_U$ occurring with non-zero coefficient.

For $x_1, x_2 \in X$ we have the invariant $\mathrm{inv}(x_1, x_2)$ from 3.4. We now write this invariant as $\mathrm{inv}(x_1, x_2)_B$, to emphasize that within the relevant Weyl group orbit of coweights we are taking the unique one that is dominant with respect to $B$. We write $\mathrm{inv}(x_1, x_2)_P$ for the image of $\mathrm{inv}(x_1, x_2)$ under the canonical surjection $X_*(A) \to \Lambda_M$ (see 4.5 for a discussion of $\Lambda_M$).

Recall also the map $H_P : G \to \Lambda_M$ defined in 12.1. This map descends to a map, also called $H_P$, from $X$ to $\Lambda_M$.

For $x_1, x_2 \in X$ we are going to show that there exists $d \geq 0$ (depending on the points $x_1$, $x_2$), such that

$$\mathrm{inv}(ax_2, x_1)_P = \nu_a + H_P(x_2) - H_{\bar{P}}(x_1) \tag{22.1.2}$$

for all $a \in A(d)$. Here we have abused notation slightly by writing $\nu_a$ when we really mean its image under the canonical surjection $X_*(A) \to \Lambda_M$. This assertion is the main ingredient in the proof of the key geometric result. However, we need some control on how big $d$ needs to be, and in fact we will show that $d$ grows linearly with $d(x_1)$, $d(x_2)$. More precisely, we have the following lemma.

LEMMA 22.1. *There exists $c > 0$ such that for all $x_1, x_2 \in X$*

$$a \in A(d) \Longrightarrow \mathrm{inv}(ax_2, x_1)_P = \nu_a + H_P(x_2) - H_{\bar{P}}(x_1) \tag{22.1.3}$$

*so long as $d \geq c(1 + d(x_1) + d(x_2))$.*

Proof. For each $x \in X$ we choose, once and for all, elements $m_x \in M$, $u_x \in U$, $\bar{m}_x \in M$, $\bar{u}_x \in \bar{U}$ such that

$$u_x m_x x_0 = x = \bar{u}_x \bar{m}_x x_0. \tag{22.1.4}$$

The lemma is an easy consequence of the following two statements. These statements involve elements $x_1, x_2 \in X$. To simplify notation we put $\bar{m}_1 := \bar{m}_{x_1}$, $\bar{u}_1 := \bar{u}_{x_1}$, $m_2 := m_{x_2}$, $u_2 := u_{x_2}$, so that

$$x_1 = \bar{u}_1 \bar{m}_1 x_0 \tag{22.1.5}$$

$$x_2 = u_2 m_2 x_0. \tag{22.1.6}$$

**Statement 1.** There exists $c_1 > 0$ such that for all $x_1, x_2 \in X$

$$a \in A(d) \Longrightarrow \mathrm{inv}(ax_2, x_1)_B = \mathrm{inv}(am_2x_0, \bar{m}_1x_0)_B$$

so long as $d \geq c_1(1 + d(x_1) + d(x_2))$.

**Statement 2.** There exists $c_2 > 0$ such that for all $x_1, x_2 \in X$

$$a \in A(d) \Longrightarrow \mathrm{inv}(am_2x_0, \bar{m}_1x_0)_B = \mathrm{inv}^M(am_2x_0, \bar{m}_1x_0)_{B_M}$$

so long as $d \geq c_2(1 + d(x_1) + d(x_2))$. Here $\mathrm{inv}^M$ is the analog for the group $M$ of inv for $G$.

First we check that the lemma follows from these two statements. Indeed, take $c$ to be the maximum of $c_1$ and $c_2$. Then, so long as $a \in A(d)$ with $d \geq c(1 + d(x_1) + d(x_2))$, we have $\mathrm{inv}(ax_2, x_1)_B = \mathrm{inv}^M(am_2x_0, \bar{m}_1x_0)_{B_M}$. Therefore $\mathrm{inv}(ax_2, x_1)_P$ is the image of $\mathrm{inv}^M((am_2x_0, \bar{m}_1x_0)_{B_M}$ under $X_*(A) \to \Lambda_M$, namely

$$\begin{aligned} H_M(\bar{m}_1^{-1}am_2) &= -H_M(\bar{m}_1) + H_M(a) + H_M(m_2) \\ &= -H_{\bar{P}}(x_1) + \nu_a + H_P(x_2). \end{aligned}$$

Next we prove Statement 1. We begin by observing that

$$\mathrm{inv}(ax_2, x_1)_B = \mathrm{inv}(am_2x_0, \bar{m}_1x_0)_B \tag{22.1.7}$$

so long as

$$au_2a^{-1} \text{ fixes } x_1 \tag{22.1.8}$$

and

$$a^{-1}\bar{u}_1a \text{ fixes } m_2x_0. \tag{22.1.9}$$

Indeed, the first condition ensures that $\mathrm{inv}(ax_2, x_1)_B = \mathrm{inv}(am_2x_0, x_1)_B$, while the second ensures that $\mathrm{inv}(am_2x_0, x_1)_B = \mathrm{inv}(am_2x_0, \bar{m}_1x_0)_B$.

If $a \in A(d)$ with $d \gg 0$, then $au_2a^{-1}$ and $a^{-1}\bar{u}_1a$ will be very close to the identity element, so (22.1.8) and (22.1.9) will hold. Lemma 19.4 allows us to make this rough statement precise. In that lemma appears a norm $\|\cdot\|_G$ on $G(F)$. It is now convenient to take this norm to be of the special type discussed in 18.11, so that

$$\|g\|_G = \exp d(g) = \exp d(gx_0). \tag{22.1.10}$$

Since $P$ is closed in $G$, the restriction of $\|\cdot\|_G$ to $P(F)$ is a valid norm on $P(F)$, and the same is true for $M$ and $U$. Moreover, as a variety $P$ is the product of $M$ and $U$. Thus both $\|mu\|_G$ and $\sup\{\|m\|_G, \|u\|_G\}$ are valid norms on $P(F)$, and therefore there exist positive constants $D_1$, $R_1$ such that

$$\sup\{\|m\|_G, \|u\|_G\} \leq D_1\|mu\|_G^{R_1} \tag{22.1.11}$$

for all $m \in M$ and $u \in U$.

By Lemma 19.4 there exist positive constants $D_2, R_2, S_2$ such that (22.1.8) holds so long as

$$\inf\{|\alpha(a)|^{-1} : \alpha \in R_U\} \geq D_2 \|u_2\|_G^{R_2} \|\bar{u}_1\bar{m}_1\|_G^{S_2}. \tag{22.1.12}$$

Using (22.1.11) and (22.1.12) together, we see that there exist positive constants $D, R, S$ such that (22.1.8) holds so long as

$$\inf\{|\alpha(a)|^{-1} : \alpha \in R_U\} \geq D \|u_2 m_2\|_G^{R} \|\bar{u}_1\bar{m}_1\|_G^{S}. \tag{22.1.13}$$

The logarithm of the right-hand side of (22.1.13) is

$$\log D + Rd(x_2) + Sd(x_1).$$

Bearing in mind (22.1.1), we see that there exists $c_3 > 0$ such that (22.1.8) holds for all $a \in A(d)$ so long as

$$d \geq c_3(1 + d(x_1) + d(x_2)). \tag{22.1.14}$$

A rather similar argument shows that there exists $c_4 > 0$ such that (22.1.9) holds for all $a \in A(d)$ so long as

$$d \geq c_4(1 + d(x_1) + d(x_2)). \tag{22.1.15}$$

It is now clear that Statement 1 holds for $c_1 = \sup\{c_3, c_4\}$.

Finally, we prove Statement 2. Put

$$\lambda := \mathrm{inv}(am_2x_0, \bar{m}_1x_0)_B \in X_*(A) \tag{22.1.16}$$

$$\lambda_M := \mathrm{inv}^M(am_2x_0, \bar{m}_1x_0)_{B_M} \in X_*(A). \tag{22.1.17}$$

We need to prove that $\lambda = \lambda_M$ when $a \in A(d)$ with $d$ sufficiently large. Note that $\lambda$, $\lambda_M$ lie in the same orbit of the Weyl group $W$, and that $\lambda$ is dominant for $B$. Thus, in order to ensure that $\lambda_M = \lambda$, it is enough to ensure that $\lambda_M$ is dominant for $B$. It is automatic that $\lambda_M$ is dominant for $B_M$, so we just need to ensure that

$$\langle \alpha, \lambda_M \rangle \geq 0 \quad \forall\, \alpha \in \Delta_U. \tag{22.1.18}$$

Recall that we have used our chosen Euclidean norm on $\mathfrak{a}$ to get a metric $d(x, y)$ on $X$. Of course this can be done for $M$ as well as $G$, so that we also get a metric $d_M(x_M, y_M)$ on $X_M := M/M \cap K$. The set $X_M$ can be identified with a subset of $X$, and the metric on $X$ extends the one on $X_M$.

Using Lemma 21.2 for the group $M$, we see that $\lambda_M$ lies in the closed ball of radius $d(\bar{m}_1x_0, x_0)$ about $\mathrm{inv}^M(am_2x_0, x_0) = \mathrm{inv}^M(m_2x_0, a^{-1}x_0)$, which in turn lies in the closed ball of radius $d(m_2x_0, x_0)$ about $\mathrm{inv}^M(x_0, a^{-1}x_0) = \nu_a$. Recall that we are assuming that all roots have norm less than or equal to 1 in the Euclidean norm on $\mathfrak{a}$. Therefore for any $\alpha \in \Delta_U$ we have

$$|\langle \alpha, \lambda_M \rangle - \langle \alpha, \nu_a \rangle| \leq d(\bar{m}_1x_0, x_0) + d(m_2x_0, x_0). \tag{22.1.19}$$

Using (22.1.11) (and its analog for $\bar{P}$), we see that there exists $c_2 > 0$ such that

$$|\langle \alpha, \lambda_M \rangle - \langle \alpha, \nu_a \rangle| \leq c_2\big(1 + d(x_1) + d(x_2)\big). \tag{22.1.20}$$

Thus (22.1.18) will hold so long as

$$\langle \alpha, \nu_a \rangle \geq c_2\big(1 + d(x_1) + d(x_2)\big) \tag{22.1.21}$$

for all $\alpha \in \Delta_U$, proving that Statement 2 holds for the constant $c_2$ that we have constructed. □

In the next lemma we use the usual partial order $\leq$ on $X_*(A)$ determined by our choice of Borel subgroup $B$. Thus $x \leq y$ if and only if $y - x$ is a non-negative integral linear combination of simple coroots.

LEMMA 22.2. *There exists $c_B > 0$ having the following property. For any $x_1, x_2 \in X$ and any $\mu \in X_*(A)$ satisfying*

$$\langle \alpha, \mu \rangle \geq c_B\bigl(1 + d(x_1) + d(x_2)\bigr) \quad \forall \alpha \in \Delta \tag{22.1.22}$$

*the following two statements hold.*

(1) *The coweight $\mu - H_B(x_2) + H_{\bar{B}}(x_1)$ is dominant for $B$.*
(2) *For $a \in A$ such that $\nu_a$ is dominant for $B$ the condition $\mathrm{inv}(ax_2, x_1)_B \leq \mu$ is equivalent to the condition $\nu_a \leq \mu - H_B(x_2) + H_{\bar{B}}(x_1)$.*

PROOF. We are going to show that we can take $c_B$ to be $1 + c$, where $c$ is the positive constant appearing in the statement of Lemma 22.1, but chosen large enough to work for all parabolic subgroups $P$ containing $B$. Consider $x_1$, $x_2$, $\mu$ satisfying the hypothesis (22.1.22).

It follows from easily from Lemma 20.4 that the first conclusion of the lemma holds. It remains to verify the second conclusion, so now consider an element $a \in A$ such that $\nu_a$ is dominant for $B$. To simplify notation we put $\lambda := \mathrm{inv}(ax_2, x_1)_B$ and we abbreviate $d(x_1)$, $d(x_2)$ to $d_1$, $d_2$ respectively.

The parabolic subgroups $P$ containing $B$ are in one-to-one correspondence with subsets of $\Delta$ (by making $P = MU$ correspond to the subset $\Delta_M$). Now take $P = MU$ to be the unique parabolic subgroup containing $B$ for which

$$\Delta_M = \{\alpha \in \Delta : \langle \alpha, \nu_a \rangle \leq c(1 + d_1 + d_2)\}, \tag{22.1.23}$$

with $c$ as chosen above. By Lemma 22.1 we then have the equality

$$\lambda = \nu_a + H_P(x_2) - H_{\bar{P}}(x_1) \tag{22.1.24}$$

in $\Lambda_M$ (with $\lambda$ and $\nu_a$ being regarded as elements in $\Lambda_M$ via the canonical surjection $X_*(A) \to \Lambda_M$), or, equivalently,

$$\lambda \text{ and } \nu_a + H_B(x_2) - H_{\bar{B}} \text{ have the same image in } \Lambda_G \tag{22.1.25}$$

and

$$\langle \varpi_\alpha, \lambda \rangle = \langle \varpi_\alpha, \nu_a + H_B(x_2) - H_{\bar{B}}(x_1) \rangle \quad \forall\, \alpha \in \Delta_U. \tag{22.1.26}$$

Here $\varpi_\alpha$ is the fundamental weight corresponding to $\alpha$ (see the discussion preceding Lemma 11.2).

We are going to apply Lemma 11.2, and in order to do so we first need to verify two inequalities:

$$\langle \alpha, \lambda \rangle \leq \langle \alpha, \mu \rangle \quad \forall\, \alpha \in \Delta_M \tag{22.1.27}$$

$$\langle \alpha, \nu_a + H_B(x_2) - H_{\bar{B}}(x_1) \rangle \leq \langle \alpha, \mu \rangle \quad \forall\, \alpha \in \Delta_M. \tag{22.1.28}$$

In view of our hypothesis on $\mu$, it is enough to check that the left sides of both these inequalities are less than or equal to $c_B(1 + d_1 + d_2)$. For the first inequality this follows from Corollary 21.3 and the definition of $\Delta_M$, and for the second inequality it follows from Lemma 20.4 and the definition of $\Delta_M$.

Now using Lemma 11.2 together with (22.1.25), (22.1.26), (22.1.27), (22.1.28), we see that $\lambda \leq \mu$ if and only if $\nu_a + H_B(x_2) - H_{\bar{B}}(x_1) \leq \mu$, which finishes the proof of the lemma. $\square$

**22.2. Key geometric result.** In this section we no longer fix the Borel subgroup $B$, as we need to consider all $B \in \mathcal{B}(A)$ at once. However, it will be convenient to pick one such Borel subgroup and call it $B_0$. We use $B_0$ to define dominance and the partial order $\leq$ on coweights.

In Lemma 22.2 there appears a positive constant $c_B$. We now put $c := \sup\{c_B : B \in \mathcal{B}(A)\}$ in order to get a constant that works for all $B$ at once. This is the positive constant $c$ appearing in our next result, the key geometric result needed for the local trace formula, which, in view of its importance, we give the status of a theorem.

THEOREM 22.3 (Arthur [**Art91a**]). *Let $x_1, x_2 \in X$ and let $\mu$ be a dominant coweight satisfying the inequality*

$$\langle \alpha, \mu \rangle \geq c\big(1 + d(x_1) + d(x_2)\big) \tag{22.2.1}$$

*for every root $\alpha$ of $A$ that is simple for $B_0$. Then $\mu_B - H_B(x_2) + H_{\bar{B}}(x_1)$ is a special $(G, A)$-orthogonal set. Here $\mu_B$ denotes the unique element in the Weyl group orbit of $\mu$ that is dominant with respect to $B$.*

*Moreover for any $a \in A$ the inequality $\operatorname{inv}(ax_2, x_1) \leq \mu$ is satisfied if and only if the following two conditions hold:*

(1) *$\nu_a$ lies in the convex hull $H$ of the set $\{\mu_B - H_B(x_2) + H_{\bar{B}}(x_1) : B \in \mathcal{B}(A)\}$.*
(2) *In $\Lambda_G$ there is an equality $\nu_a = \mu - H_G(x_2) + H_G(x_1)$.*

*In the second condition we have written simply $\nu_a$ and $\mu$ when we really mean their images under the canonical surjection $X_*(A) \twoheadrightarrow \Lambda_G$.*

PROOF. We begin by noting that $\mu_B$ is a positive orthogonal set (see 12.8). Moreover $H_B(x_2)$ is a positive orthogonal set (see 12.1), and $H_{\bar{B}}(x_1)$ is a negative orthogonal set (see the end of 12.4). So $\mu_B - H_B(x_2) + H_{\bar{B}}(x_1)$ is the difference of the positive orthogonal set $\mu_B$ and the negative orthogonal set $H_B(x_2) - H_{\bar{B}}(x_1)$ and in general is neither positive nor negative. However it will be special (hence positive) when $\mu$ is big enough. In fact our assumption on $\mu$ does guarantee that $\mu$ is big enough, since from the first part of Lemma 22.2, we see that for each $B \in \mathcal{B}(A)$ the coweight $\mu_B - H_B(x_2) + H_{\bar{B}}(x_1)$ is dominant for $B$.

Choose $B \in \mathcal{B}(A)$ such that $\nu_a$ is dominant for $B$. From Lemma 12.2 it follows that $\nu_a$ satisfies conditions (1) and (2) in the theorem if and only if

$$\nu_a \underset{B}{\leq} \mu_B - H_B(x_2) + H_{\bar{B}}(x_1), \tag{22.2.2}$$

and by the second part of Lemma 22.2 this happens if and only if

$$\operatorname{inv}(ax_2, x_1)_B \underset{B}{\leq} \mu_B,$$

or, equivalently, if and only if

$$\operatorname{inv}(ax_2, x_1)_{B_0} \underset{B_0}{\leq} \mu_{B_0}.$$

□

## 23. The weight factors $\tilde{u}_M$ and $\tilde{v}_M$

We are almost ready to prove the local trace formula. Before we can do so we need to introduce some more weight factors and relate them to toric varieties.

**23.1. Weight factor $\tilde{u}_A$.** Recall the definition of the weight factor $u_A$ occurring in our preliminary form of the local trace formula: $u_A(g_1, g_2; \mu)$ is the measure of the set of $a \in A$ such that

$$\operatorname{inv}(ax_2, x_1) \leq \mu, \tag{23.1.1}$$

where $x_1 = g_1 x_0$, $x_2 = g_2 x_0$. We use the Haar measure on $A$ giving measure 1 to $A \cap K$. From the key geometric result (Theorem 22.3) we see that when the dominant coweight $\mu$ is big enough relative to $g_1, g_2$, the $(G, A)$-orthogonal set

$$B \mapsto \mu_B - H_B(g_2) + H_{\bar{B}}(g_1)$$

is positive and the weight factor $u_A(g_1, g_2; \mu)$ is equal to the number of coweights $\nu \in X_*(A)$ satisfying the following two conditions:

(1) $\nu$ lies in the convex hull of the points $\mu_B - H_B(g_2) + H_{\bar{B}}(g_1)$,
(2) in $\Lambda_G$ the elements $\nu$ and $\mu - H_G(g_2) + H_G(g_1)$ are equal.

It is well-known that such counting problems for lattice points in convex polyhedra arise naturally in the theory of toric varieties. Fulton's book [**Ful93**] is an excellent reference for all that we will need about toric varieties.

What torus do we need? We write $\hat{A}$ for the Langlands dual torus of $A$; it is a complex torus characterized by the property that $X^*(\hat{A}) = X_*(A)$ (and hence $X_*(\hat{A}) = X^*(A)$). We let $\hat{G} \supset \hat{A}$ be a Langlands dual group for $G$: the roots (respectively, coroots) of $\hat{A}$ in $\hat{G}$ are the coroots (respectively, roots) of $A$ in $G$. We write $Z(\hat{G})$ for the center of $\hat{G}$; thus the adjoint group of $\hat{G}$ is $\hat{G}/Z(\hat{G})$ with maximal torus $\hat{A}/Z(\hat{G})$.

The toric variety $V = V^G$ we need is a toric variety for the torus $\hat{A}/Z(\hat{G})$. To specify $V$ we must say which fan we are using. We take the Weyl fan in $X_*(\hat{A}/Z(\hat{G})) \otimes_{\mathbb{Z}} \mathbb{R}$. This is the fan determined by the root hyperplanes in this vector space. Thus the cones of maximal dimension in our fan are the closures of the Weyl chambers in $X_*(\hat{A}/Z(\hat{G})) \otimes_{\mathbb{Z}} \mathbb{R}$, and there is one cone in the fan for each $P \in \mathcal{F}(A)$, the set of parabolic subgroups $P$ of $G$ such that $P \supset A$. The toric variety $V$ is projective, and since $\hat{G}/Z(\hat{G})$ is adjoint, it is also non-singular.

The torus $\hat{A}/Z(\hat{G})$ acts on $V$ and hence $\hat{A}$ also acts on $V$ (through $\hat{A} \twoheadrightarrow \hat{A}/Z(\hat{G})$). The $\hat{A}$-orbits in $V$ are in one-to-one correspondence with cones in the Weyl fan, that is, with parabolic subgroups $P \in \mathcal{F}(A)$; we write $V_P$ for the orbit of $\hat{A}$ indexed by $P$. Each orbit has a natural base point, and in fact

$$V_P = \hat{A}/Z(\hat{M})$$

where $M$ is *the* Levi component of $P$ (that is, the unique Levi component of $P$ that contains $A$), and $\hat{M}$ is the corresponding Levi subgroup of $\hat{G}$ containing $\hat{A}$ (the one whose roots are the coroots of $M$). The closure $\overline{V}_P$ of $V_P$ is

$$\bigcup_{Q : Q \subset P} V_Q$$

and is the toric variety $V^M$ associated to the Weyl fan for $(\hat{M}/Z(\hat{M}), \hat{A}/Z(\hat{M}))$.

Let $\mathcal{L}$ be an $\hat{A}$-equivariant line bundle on $V$. At each $\hat{A}$-fixed point in $V$ the torus $\hat{A}$ acts by a character on the line (in our line bundle) at that fixed point. There is one fixed point for each $B \in \mathcal{B}(A)$ (namely the single point in the orbit $V_B = \hat{A}/\hat{A}$), so for each $B \in \mathcal{B}(A)$ we get a character $x_B \in X^*(\hat{A})$, or, in other words, a cocharacter $x_B \in X_*(A)$.

Since $Z(\hat{G})$ acts trivially on $V$, there is a single character of $Z(\hat{G})$ by which $Z(\hat{G})$ acts on every line in our line bundle; therefore all the elements $x_B \in X_*(A)$ have the same image in the quotient $\Lambda_G$ of $X_*(A)$. (Note that $\Lambda_G$ can be identified with $X^*(Z(\hat{G}))$. But much more is true. For any $P \in \mathcal{F}(A)$ the restriction of $\mathcal{L}$ to $\overline{V}_P = V^M$ is an equivariant line bundle on the toric variety $V^M$ for $M$; therefore, applying what has already been said to $M$ rather than $G$, we see that the points $x_B$ for all $B \in \mathcal{B}(A)$ such that $B \subset P$ have the same image in $\Lambda_M$; thus $(x_B)$ is a $(G, A)$-orthogonal set in $X_*(A)$. In fact $\mathcal{L} \mapsto (x_B)$ is an isomorphism from the group of isomorphism classes of $\hat{A}$-equivariant line bundles on $V$ to the group of $(G, A)$-orthogonal sets in $X_*(A)$. Restriction of equivariant line bundles from $V$ to $\overline{V}_P = V^M$ corresponds to sending the orthogonal set $(x_B)_{B\in\mathcal{B}(A)}$ to the $(M, A)$-orthogonal set $(x_B)_{B\in\mathcal{B}(A):B\subset P}$, the operation on orthogonal sets discussed in 12.2.

If the orthogonal set is positive, all the higher cohomology groups $H^i(V, \mathcal{L})$ $(i > 0)$ vanish, and as an $\hat{A}$-module $H^0(V, \mathcal{L})$ is multiplicity free and contains the character $x \in X^*(\hat{A}) = X_*(A)$ if and only if the following two conditions hold:

(1) $x$ lies in the convex hull of $\{x_B : B \in \mathcal{B}(A)\}$,

(2) the image of $x$ in $\Lambda_G$ coincides with the common image of the points $x_B$.

For any line bundle $\mathcal{L}$ on $V$ we put

$$EP(\mathcal{L}) := \sum_i (-1)^i \dim H^i(V, \mathcal{L}).$$

For any $g_1, g_2 \in G$ and any dominant coweight $\mu$ we let $\mathcal{L}_{(g_1,g_2;\mu)}$ be an equivariant line bundle on $V$ such that the associated $(G, A)$-orthogonal set in $X_*(A)$ is

$$B \mapsto \mu_B - H_B(g_2) + H_{\bar{B}}(g_1),$$

and we put

$$\tilde{u}_A(g_1, g_2; \mu) := EP(\mathcal{L}_{(g_1,g_2;\mu)}). \tag{23.1.2}$$

It then follows from Theorem 22.3 that when $\mu$ is big compared to $g_1$, $g_2$, the weight factors $\tilde{u}_A(g_1, g_2; \mu)$ and $u_A(g_1, g_2; \mu)$ coincide. In our next version of the local trace formula (see 24.3), the weight factor $u_A$ will be replaced by the more pleasant weight factor $\tilde{u}_A$. Before we can carry this out, we need to introduce modified weight factors $\tilde{u}_M$ for all Levi subgroups $M$ containing $A$.

**23.2. Weight factor $\tilde{u}_M$.** We have treated the case $M = A$. The general case is similar, though slightly more complicated, as we will now see. The toric variety $Y_M = Y_M^G$ we need is a non-singular projective toric variety for the torus $Z(\hat{M})/Z(\hat{G})$. Note that the quotient $Z(\hat{M})/Z(\hat{G})$ really is a torus, although in general $Z(\hat{G})$ and $Z(\hat{M})$ are only diagonalizable groups. Since $Z(\hat{M})/Z(\hat{G})$ is the center of the Levi subgroup $\hat{M}/Z(\hat{G})$ in the adjoint group $\hat{G}/Z(\hat{G})$, we may as well temporarily simplify notation by assuming that $\hat{G}$ is adjoint (or, equivalently, that $G$ is semisimple and simply connected).

Since we are now assuming that $\hat{G}$ is adjoint, the group $Z(\hat{M})$ is a torus, and in fact is a subtorus of $\hat{A}$, so that $X_*(Z(\hat{M}))$ is a subgroup of $X_*(\hat{A})$. Inside the real vector space $X_*(\hat{A})_\mathbb{R}$ obtained from $X_*(\hat{A})$ by tensoring over $\mathbb{Z}$ with $\mathbb{R}$ we have the Weyl fan, and the subspace $X_*(Z(\hat{M}))_\mathbb{R}$ of $X_*(\hat{A})_\mathbb{R}$ is a union of cones in the Weyl fan; thus, the collection of cones in the Weyl fan that happen to lie inside the subspace $X_*(Z(\hat{M}))_{\mathbb{R}}$ gives us a fan in $X_*(Z(\hat{M}))_{\mathbb{R}}$, hence a toric variety $Y_M = Y_M^G$ for $Z(\hat{M})$, which is obviously complete and easily seen to be non-singular (again because $\hat{G}$ is adjoint) and projective.

The index set for the cones in this fan in $X_*(Z(\hat{M}))_{\mathbb{R}}$ is $\mathcal{F}(M)$, the set of parabolic subgroups $Q$ of $G$ such that $Q \supset M$. Thus the decomposition of $Y_M$ as a union of $Z(\hat{M})$-orbits is given by

$$Y_M = \bigcup_{Q \in \mathcal{F}(M)} Z(\hat{M})/Z(\hat{L}_Q),$$

where $L_Q$ denotes the unique Levi component of $Q$ that contains $M$.

We need to understand how $Y_M$ is related to $V$. Since the fan used to produce $Y_M$ can also be viewed as a fan in $X_*(\hat{A})_{\mathbb{R}}$ whose support is the subspace $X_*(Z(\hat{M}))_{\mathbb{R}}$, it also produces a toric variety $U_M$ for $\hat{A}$, sitting inside $V$ as an $\hat{A}$-stable open subvariety. The decomposition of $U_M$ as a union of $\hat{A}$-orbits is

$$U_M = \bigcup_{Q \in \mathcal{F}(M)} V_Q = \bigcup_{Q \in \mathcal{F}(M)} \hat{A}/Z(\hat{L}_Q).$$

Moreover $Y_M$ sits inside $V$ as a closed $Z(\hat{M})$-stable subspace, and the multiplication map $\hat{A} \times Y_M \to V$ has image $U_M$ and induces an isomorphism

$$\hat{A} \underset{Z(\hat{M})}{\times} Y_M \simeq U_M. \tag{23.2.1}$$

(The space on the left side of the identification is the quotient of $\hat{A} \times Y_M$ by the equivalence relation $(az, y) \sim (a, zy)$.)

Consider any $Q = LU \in \mathcal{F}(A)$ (with $L$ chosen so that $L \supset A$, as usual). If $Q \notin \mathcal{F}(M)$, then $\overline{V}_Q$ does not meet $Y_M$. On the other hand, if $Q \in \mathcal{F}(M)$, so that $L \supset M$, then

$$U_M \cap \overline{V}_Q = \bigcup_{\{Q' \in \mathcal{F}(M) : Q' \subset Q\}} V_{Q'},$$

and since $\{Q' \in \mathcal{F}(M) : Q' \subset Q\}$ can be identified with $\mathcal{F}^L(M)$, the set of parabolic subgroups of $L$ containing $M$, we see that

$$U_M \cap \overline{V}_Q = \hat{A} \underset{Z(\hat{M})}{\times} Y_M^L.$$

From these considerations we obtain the following result.

LEMMA 23.1. *Let $M$ be a Levi subgroup of $G$ containing $A$ and let $Q = LU$ be a parabolic subgroup of $G$ whose Levi component $L$ contains $A$. Recall that $\overline{V}_Q$ can be identified with the toric variety $V^L$. If $Q \notin \mathcal{F}(M)$, then $\overline{V}_Q$ does not meet $Y_M$. Otherwise $L$ contains $M$, and the non-singular closed subvarieties $\overline{V}_Q = V^L$ and $Y_M$ of $V$ intersect transversely, their intersection being the non-singular closed subvariety $Y_M^L$ in $V^L$.*

We have now completed our discussion of $Y_M$ in the case $\hat{G}$ is adjoint, so we return to the case of a general split group $G$ and Levi subgroup $M \in \mathcal{L}(A)$. We write $G_0$ for the simply connected cover of the derived group of $G$, and $M_0$ for the Levi subgroup in $G_0$ obtained as the inverse image of $M$ under $G_0 \to G$; thus $\hat{G}_0 = \hat{G}/Z(\hat{G})$, $\hat{M}_0 = \hat{M}/Z(\hat{G})$, and $Z(\hat{M}_0) = Z(\hat{M})/Z(\hat{G})$.

We have already defined the toric variety $Y_{M_0}^{G_0}$ for $Z(\hat{M}_0) = Z(\hat{M})/Z(\hat{G})$. Using the canonical surjection $Z(\hat{M}) \twoheadrightarrow Z(\hat{M}_0)$, we now view $Y_{M_0}^{G_0}$ as a space on which $Z(\hat{M})$ acts, and rename it $Y_M^G$. Thus, as a space, $Y_M^G$ depends only on $G_0$, but $Z(\hat{M})$ and its action on $Y_M^G$ reflect $G$ and $M$.

**23.3. Equivariant line bundles on $Y_M$.** The decomposition of $Y_M$ into $Z(\hat{M})$-orbits

$$Y_M = \bigcup_{Q\in\mathcal{F}(M)} Z(\hat{M})/Z(\hat{L}_Q),$$

proved in the case $\hat{G}$ is adjoint, obviously remains valid in the general case, since

$$Z(\hat{M})/Z(\hat{L}_Q) = (Z(\hat{M})/Z(\hat{G}))/(Z(\hat{L}_Q)/Z(\hat{G})).$$

Thus the fixed points of $Z(\hat{M})$ in $Y_M$ are indexed by $\mathcal{P}(M)$. Given $P = MU \in \mathcal{P}(M)$, the fixed point in $Y_M$ indexed by $P$ is the unique point in the 1-element set $\overline{V}_P \cap Y_M$. The character group of the diagonalizable group $Z(\hat{M})$ is $\Lambda_M$, as we noted before. A $Z(\hat{M})$-equivariant line bundle $\mathcal{M}$ on $Y_M$ gives us a $(G, M)$-orthogonal set of points $y_P \in \Lambda_M$, with $y_P$ defined as the character through which $Z(\hat{M})$ acts on the line (in $\mathcal{M}$) at the fixed point indexed by $P$, and in this way we get an isomorphism from the group of isomorphism classes of $Z(\hat{M})$-equivariant line bundles on $Y_M$ to the group of $(G, M)$-orthogonal sets in $\Lambda_M$. Note that all the points $y_P$ (with $P$ ranging through $\mathcal{P}(M)$) have the same image in $\Lambda_G$.

**23.4. Restriction of equivariant line bundles from $V$ to $Y_M$.** Now suppose that $\mathcal{L}$ is an $\hat{A}$-equivariant line bundle on $V$. From $\mathcal{L}$ we obtain a $(G, A)$-orthogonal set of points $x_B \in X_*(A)$. The restriction $\mathcal{L}|_{Y_M}$ of $\mathcal{L}$ to the subspace $Y_M$ is a $Z(\hat{M})$-equivariant line bundle on $Y_M$, hence yields a $(G, M)$-orthogonal set of points $y_P$ in $\Lambda_M$. Now $y_P$ is the character on which $Z(\hat{M})$ acts on the line in $\mathcal{L}$ at the unique point in $\overline{V}_P \cap Y_M$, but since $Z(\hat{M})$ acts trivially on $\overline{V}_P = V^M$, it acts by a single character on the lines in $\mathcal{L}$ at all points in $\overline{V}_P$ (as we have already discussed in 23.1), showing that $y_P$ is the common image $x_P$ of the points $x_B$ ($B \in \mathcal{B}(A)$ such that $B \subset P$) under $X_*(A) \twoheadrightarrow \Lambda_M$. In other words $(y_P) = (x_P)$, where $(x_P)$ is the $(G, M)$-orthogonal set in $\Lambda_M$ obtained from the $(G, A)$-orthogonal set $(x_B)$ by the procedure in 12.5.

**23.5. Euler characteristics of $Z(\hat{M})$-equivariant line bundles on $Y_M$.** Let $\mathcal{M}$ be a $Z(\hat{M})$-equivariant line bundle on $Y_M$ with associated $(G, M)$-orthogonal set $(y_P)_{P\in\mathcal{P}(M)}$ in $\Lambda_M$. We put

$$EP(\mathcal{M}) := \sum_i (-1)^i \dim H^i(Y_M, \mathcal{M}). \tag{23.5.1}$$

More generally, for $s \in Z(\hat{M})$ we put

$$EP(s, \mathcal{M}) := \sum_i (-1)^i \operatorname{tr}\big(s; H^i(Y_M, \mathcal{M})\big), \tag{23.5.2}$$

so that we recover $EP(\mathcal{M})$ when $s = 1$.

If the $(G, M)$-orthogonal set $(y_P)$ associated to $\mathcal{M}$ is positive, the higher cohomology groups $H^i(Y_M, \mathcal{M})$ ($i > 0$) vanish, and the representation of $Z(\hat{M})$ on $H^0(Y_M, \mathcal{M})$ is multiplicity free, with the character $y \in X^*(Z(\hat{M})) = \Lambda_M$ appearing in $H^0(Y_M, \mathcal{M})$ if and only if

(1) The image $\bar{y}$ of $y$ under $\Lambda_M \to \mathfrak{a}_M$ lies in the convex hull of the points $\bar{y}_P$ obtained as images under $\Lambda_M \to \mathfrak{a}_M$ of the points $y_P$ in our orthogonal set.
(2) The image of $y$ in $\Lambda_G$ coincides with the common image in $\Lambda_G$ of the points $y_P$.

Thus, when $(y_P)$ is positive, the number of points $y \in \Lambda_M$ satisfying the two conditions above is equal to $EP(\mathcal{M})$.

However the weight factor $u_M(g_1, g_2; \mu)$ (for large $\mu$) involves counting points in $X_*(A_M)$ rather than $\Lambda_M$, a circumstance which must be be taken into account. Recall the canonical injective homomorphism $X_*(A_M) \hookrightarrow \Lambda_M$, which we use to identify $X_*(A_M)$ with a subgroup of finite index in $\Lambda_M$. We write $\mathcal{Z}_M$ for the Pontryagin dual group

$$\mathcal{Z}_M := \mathrm{Hom}(\Lambda_M/X_*(A_M), \mathbb{C}^\times) \tag{23.5.3}$$

of $\Lambda_M/X_*(A_M)$. It is easy to see that $\mathcal{Z}_M$ can be identified with the subgroup of $Z(\hat{M}) = \mathrm{Hom}(\Lambda_M, \mathbb{C}^\times)$ obtained as the center of the derived group of $\hat{M}$.

When $(y_P)$ is positive, Fourier analysis on the finite abelian group $\mathcal{Z}_M$ shows that the number of points $y \in X_*(A_M)$ satisfying conditions (1) and (2) above is equal to

$$|\mathcal{Z}_M|^{-1} \sum_{s \in \mathcal{Z}_M} EP(s, \mathcal{M}). \tag{23.5.4}$$

**23.6. Definition of the weight factors $\tilde{u}_M$ and $\tilde{v}_M$.** Let $g_1, g_2 \in G$ and let $\mu$ be a dominant coweight. As in 23.1 we get an $\hat{A}$-equivariant line bundle $\mathcal{L}_{(g_1,g_2;\mu)}$ on $V$, which we restrict to the subspace $Y_M$, obtaining $\mathcal{L}_{(g_1,g_2;\mu)}|_{Y_M}$.

Define the weight factors $\tilde{u}_M$ and $\tilde{v}_M$ by

$$\tilde{u}_M(g_1, g_2; \mu) := |\mathcal{Z}_M|^{-1} \sum_{s \in \mathcal{Z}_M} EP(s, \mathcal{L}_{(g_1,g_2;\mu)}|_{Y_M}), \tag{23.6.1}$$

$$\tilde{v}_M(g_1, g_2; \mu) := |\mathcal{Z}_M|^{-1} EP(\mathcal{L}_{(g_1,g_2;\mu)}|_{Y_M}). \tag{23.6.2}$$

**23.7. Agreement of $u_M$ and $\tilde{u}_M$ when $\mu$ is big.** In this subsection we will check that

$$u_M(g_1, g_2; \mu) = \tilde{u}_M(g_1, g_2; \mu)$$

when $\mu$ is big enough relative to $g_1, g_2$.

Indeed, let $g_1, g_2 \in G$, let $x_1, x_2 \in X$ be the transforms of the basepoint $x_0 \in X = G/K$ under $g_1, g_2$ respectively, and let $\mu$ be a dominant coweight big enough that the conclusions of Theorem 22.3 hold for $x_1, x_2, \mu$. Put

$$x_B := \mu_B - H_B(x_2) + H_{\bar{B}}(x_1),$$

so that $(x_B)$ is a special $(G, A)$-orthogonal set in $X_*(A)$, and let $(x_P)_{P \in \mathcal{P}(M)}$ be the $(G, M)$-orthogonal set in $\Lambda_M$ obtained from $(x_B)$ as in 23.4 (and 12.5). From (8.4.5) and Theorem 22.3, we see that $u_M(g_1, g_2; \mu)$ is the number of points $x \in X_*(A_M)$ such that

(1) The point $x$ lies in the convex hull of the points $\{x_B : B \in \mathcal{B}(A)\}$.
(2) The image of $x$ in $\Lambda_G$ coincides with the common image in $\Lambda_G$ of the points $x_B$.

On the other hand, we have designed our definitions so that $\tilde{u}_M(g_1, g_2; \mu)$ is equal to the number of points $x \in X_*(A_M)$ such that

(1′) The image $\bar{x}$ of $x$ under $X_*(A_M) \hookrightarrow \Lambda_M \to \mathfrak{a}_M$ lies in the convex hull of the points $\bar{x}_P$ obtained as images under $\Lambda_M \to \mathfrak{a}_M$ of the points $x_P$.

(2′) The image of $x$ in $\Lambda_G$ coincides with the common image in $\Lambda_G$ of the points $x_P$.

Clearly conditions (2) and (2′) are equivalent. Moreover (1) and (1′) are equivalent by Proposition 12.1. Therefore

$$u_M(g_1, g_2; \mu) = \tilde{u}_M(g_1, g_2; \mu),$$

as desired.

**23.8. Qualitative behavior of $EP(s, \mathcal{M})$.** The group $E := \mathrm{Pic}_{Z(\hat{M})}(Y_M)$ of isomorphism classes of $Z(\hat{M})$-equivariant line bundles $\mathcal{M}$ on $Y_M$ is a finitely generated abelian group, isomorphic to the group of $(G, M)$-orthogonal sets $(y_P)_{P \in \mathcal{P}(M)}$ in $\Lambda_M$. There is an obvious embedding $\Lambda_M \hookrightarrow E$, obtained by using $y \in \Lambda_M = X^*(Z(\hat{M}))$ to define a $Z(\hat{M})$-equivariant line bundle on a point, and then pulling this back to $Y_M$; the corresponding orthogonal set is the one for which $y_P = y$ for all $P \in \mathcal{P}(M)$.

The quotient $E/\Lambda_M$ is isomorphic to $\mathrm{Pic}(Y_M) \simeq H^2(Y_M, \mathbb{Z})$, known to be a free abelian group (whose rank is easy to compute [**Ful93**]). It is also known that there exists (an obviously unique) polynomial $F$ of degree $\dim Y_M = \dim(A_M/A_G)$ on the $\mathbb{Q}$-vector space $E/\Lambda_M \otimes_{\mathbb{Z}} \mathbb{Q} = H^2(Y_M, \mathbb{Q})$ such that

$$EP(\mathcal{M}) = F(\mathcal{M}). \tag{23.8.1}$$

(In the right side of this equality we are using $E \twoheadrightarrow E/\Lambda_M$ to view $F$ as a function on $E$.)

Slightly more generally, now consider $s \in Z(\hat{G})$. Since $Z(\hat{G})$ acts trivially on $Y_M$, it acts on all lines in $\mathcal{M}$ by the same character, namely the character $z \in \Lambda_G = X^*(Z(\hat{G}))$ obtained as the common image in $\Lambda_G$ of the points $y_P$. Therefore

$$EP(s, \mathcal{M}) = \langle s, z \rangle F(\mathcal{M}) \tag{23.8.2}$$

for the same polynomial $F$ as before.

However, we need to understand the qualitative nature of the function $\mathcal{M} \mapsto EP(s, \mathcal{M})$ for any $s \in Z(\hat{M})$. For this it is convenient to use the localization theorem for equivariant $K$-theory [**Nie74, BFQ79**] (see also [**Bri88**]), which expresses $EP(s, \mathcal{M})$ in terms of contributions from the various connected components of the fixed point set $Y_M^s$ of $s$ on $Y_M$. Each of these connected components is a nonsingular projective toric variety for some quotient of $Z(\hat{M})$, and this leads to the conclusion that $EP(s, \mathcal{M})$ can be expressed as a finite sum of the form

$$\sum_{P \in \mathcal{P}(M)} \langle s, y_P \rangle F_{s,P}(\mathcal{M}) \tag{23.8.3}$$

where $F_{s,P}$ is some polynomial function on $(E/\Lambda_M) \otimes_{\mathbb{Z}} \mathbb{Q}$ of degree no bigger than $\dim(A_M/A_G)$. Note (although it causes no trouble) that the polynomials $F_{s,P}(\mathcal{M})$ are not unique (unless $s$ is generic in $Z(\hat{M})$), since the characters $(y_P) \mapsto \langle s, y_P \rangle$ on $E$ (one for each $P \in \mathcal{P}(M)$) need not be distinct.

Consequently (bearing in mind Lemma 20.4), we see that the weight factors $\tilde{u}_M$, $\tilde{v}_M$ both satisfy the same kind of estimate (see 20.5) as $u_M$, namely, there exists a positive constant $c$ such that

$$\tilde{u}_M(g_1, g_2; \mu) \leq c(1 + \log \|g_1\|_{M\backslash G} + \log \|g_2\|_{M\backslash G} + \|\mu\|_E)^{\dim(A_M)} \tag{23.8.4}$$

(and the same for $\tilde{v}_M$).

## 24. Local trace formula

**24.1. The goal.** In this section we will prove various versions of the local trace formula for our split group $G$. All versions will have the following shape. Recall that $\mathcal{L} = \mathcal{L}(A)$ denotes the set of Levi subgroups $M$ of $G$ such that $M \supset A$. For each $M \in \mathcal{L}$ we will have a weight factor $w_M(g_1, g_2)$ on $G \times G$, left invariant under $A_M \times A_M$ (and, in some cases, even under $M \times M$) and right invariant under $K \times K$. Given such a family $w = (w_M)_{M\in\mathcal{L}}$, we can define a distribution $T_w$ on $\mathfrak{g} \times \mathfrak{g}$ by

$$\begin{aligned} T_w(f_1, f_2) = \sum_{M\in\mathcal{L}} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \cdot \\ \int_{A_M\backslash G} \int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) w_M(h, g)\, d\dot{h}\, d\dot{g}\, dX, \end{aligned} \tag{24.1.1}$$

$f_1, f_2$ being two functions in $C_c^\infty(\mathfrak{g})$, so long as all these triple integrals make sense. Here the notation is the same as in our second form of the Weyl integration formula (see 7.11).

For each of the weight factors we will consider, the integrals will make sense, and we will have a version of the local trace formula, namely the equality

$$T_w(f_1, f_2) = T_w(\hat{f}_1, \check{f}_2). \tag{24.1.2}$$

**24.2. Weight factors $u_M$.** We already know (see 8.4) that for any dominant coweight $\mu$ the local trace formula holds for the weight factors $u_M(g_1, g_2, \mu)$. However, it seems that these weight factors are too complicated to be of much use.

**24.3. Weight factors $\tilde{u}_M$.** Again let $\mu$ be a dominant coweight. Let us now check that the local trace formula holds for the weight factors $\tilde{u}_M(g_1, g_2; \mu)$. To this end we need to choose an auxiliary dominant regular coweight $\mu_1$ (so that $\langle \alpha, \mu_1 \rangle > 0$ for every simple root $\alpha$). Replacing $\mu_1$ by $N\mu_1$ for some positive integer $N$, we may assume that

$$\langle s, \mu_1 \rangle = 1 \tag{24.3.1}$$

for all $M \in \mathcal{L}$ and all $s \in \mathcal{Z}_M$.

Fix $f_1, f_2 \in C_c^\infty(\mathfrak{g})$ and define complex valued functions $\varphi(d)$, $\tilde{\varphi}(d)$ on the set of non-negative integers $d$ by the following rules:

$$\varphi(d) := T_w(f_1, f_2) \text{ with } w_M(g, h) = u_M(g, h; \mu + d\mu_1), \tag{24.3.2}$$

$$\tilde{\varphi}(d) := T_w(f_1, f_2) \text{ with } w_M(g, h) = \tilde{u}_M(g, h; \mu + d\mu_1). \tag{24.3.3}$$

In view of the discussion in 23.8 our assumption (24.3.1) on $\mu_1$ guarantees that $\tilde{\varphi}$ is a polynomial function of $d$.

We claim that $\tilde{\varphi}(d) - \varphi(d) \to 0$ as $d \to +\infty$. Obviously $\tilde{\varphi}(d) - \varphi(d) = T_{w_d}(f_1, f_2)$ for the weight factors $w_d$ defined by

$$(w_d)_M(g, h) := \tilde{u}_M(g, h; \mu + d\mu_1) - u_M(g, h; \mu + d\mu_1),$$

and for fixed $g, h$ we know that this difference is 0 once $d$ is sufficiently big. Therefore the integrands in the integrals defining $T_{w_d}$ approach 0 pointwise, and to conclude that $\tilde{\varphi}(d)-\varphi(d) \to 0$, it is enough to justify the application of Lebesgue's dominated convergence theorem.

For this we need an estimate for $(w_d)_M$ that is independent of $d$, unlike the estimates we already have for the two terms we took the difference of to get $w_d$, which of course do depend on $d$.

However, our first step is to use the estimates we already have (see Lemma 20.7 and the inequality (23.8.4)) for $u_M$ and $\tilde{u}_M$ to conclude that there exists a positive constant $c$ such that

$$|(w_d)_M(g,h)| \leq c(1 + \log \|g\|_{M\backslash G} + \log \|h\|_{M\backslash G} + d)^{\dim(A_M)}. \tag{24.3.4}$$

Noting that for any simple root $\alpha$ there is an inequality $\langle \alpha, \mu + d\mu_1 \rangle \geq d$, we see from Theorem 22.3 and the discussion in 23.7 that there exists a positive constant $c'$ such that $(w_d)_M(g,h) = 0$ unless

$$d \leq c'(1 + d(g) + d(h))$$

(with $d(g) := d(gx_0, x_0) = \log \|g\|_G$). Since $(w_d)_M(g,h)$ is left $(A_M \times A_M)$-invariant, it follows that $(w_d)(g,h) = 0$ unless

$$d \leq c'(1 + d_{A_M\backslash G}(g) + d_{A_M\backslash G}(h)),$$

where

$$d_{A_M\backslash G}(g) := \inf\{d(a_M g) : a_M \in A_M\},$$

Also, since $\log \|g\|_{M\backslash G} = \inf\{d(mg) : m \in M\}$, we trivially have

$$\log \|g\|_{M\backslash G} \leq d_{A_M\backslash G}(g).$$

We conclude that for all $d \geq 0$ there is an inequality

$$|(w_d)_M(g,h)| \leq c\big(1+d_{A_M\backslash G}(g)+d_{A_M\backslash G}(h)+c'(1+d_{A_M\backslash G}(g)+d_{A_M\backslash G}(h))\big)^{\dim A_M},$$

which can be simplified to an inequality of the form

$$|(w_d)_M(g,h)| \leq c''\big(1 + d_{A_M\backslash G}(g) + d_{A_M\backslash G}(h)\big)^{\dim A_M}.$$

Consider the right side of this estimate as a family of weight factors $w_{est}$. The integrals appearing in $T_{w_{est}}(f_1, f_2)$ are all convergent by Proposition 20.1. Therefore we have justified the application of Lebesgue's dominated convergence theorem.

We can summarize what we have done so far as follows: $\tilde{\varphi}(d)$ is a polynomial function of $d$ such that $\tilde{\varphi}(d) - \varphi(d) \to 0$ as $d \to +\infty$. We used $f_1, f_2$ to define $\varphi$, $\tilde{\varphi}$; we now indicate this dependence by writing $\varphi_{f_1,f_2}$ and $\tilde{\varphi}_{f_1,f_2}$. The local trace formula for the weight factors $u_M$ tells us that $\varphi_{f_1,f_2} = \varphi_{\hat{f}_1,\check{f}_2}$. Therefore $\tilde{\varphi}_{f_1,f_2}(d) - \tilde{\varphi}_{\hat{f}_1,\check{f}_2}(d)$ is a polynomial function of $d$ that approaches 0 as $d \to +\infty$, which obviously implies that it is identically 0, or, in other words, that

$$\tilde{\varphi}_{f_1,f_2}(d) = \tilde{\varphi}_{\hat{f}_1,\check{f}_2}(d)$$

for all $d \geq 0$; taking $d = 0$ we conclude that the local trace formula holds for the weight factors $\tilde{u}_M(g_1, g_2; \mu)$, as desired.

**24.4. Weight factors $\tilde{v}_M$.** Recall that $\tilde{u}_M$ was defined by the equality

$$\tilde{u}_M(g_1, g_2; \mu) := |\mathcal{Z}_M|^{-1} \sum_{s \in \mathcal{Z}_M} EP(s, \mathcal{L}_{(g_1,g_2;\mu)}|_{Y_M}). \tag{24.4.1}$$

It follows from (23.8.3) that the function

$$\mu \mapsto EP(s, \mathcal{L}_{(g_1,g_2;\mu)}|_{Y_M}) \tag{24.4.2}$$

on the monoid of dominant coweights $\mu$ has the form

$$\sum_{w \in W} \langle w(s), \mu \rangle F_w \tag{24.4.3}$$

for some collection of polynomial functions $F_w$ of $\mu$. (As usual $W$ denotes the Weyl group of $A$.) Applying linear independence of characters on the monoid of dominant coweights, we conclude that the local trace formula holds for the weight factors $\tilde{v}_M(g_1, g_2; \mu)$ (since these were obtained from the weight factors $\tilde{u}_M$ by extracting the contribution of the trivial character on the monoid of dominant coweights).

**24.5. Weight factors $\tilde{\mathbf{v}}_M$.** So far all of our weight factors have been numbers. We now consider weight factors $\tilde{\mathbf{v}}_M$ (closely related to $\tilde{v}_M$) taking values in $K(V)_{\mathbb{C}}$, the complexification of the Grothendieck group $K(V)$ of vector bundles (in the sense of algebraic geometry) on our toric variety $V = V^G$. The Grothendieck groups $K(Y_M)$ will also play an auxiliary role. Since $V$, $Y_M$ are non-singular projective varieties, we may also view these $K$-groups as Grothendieck groups of coherent sheaves.

Consider the closed immersion $i_M : Y_M \hookrightarrow V$. Thinking of our $K$-groups in terms of vector bundles, we have a restriction (pull-back) map

$$i_M^* : K(V) \to K(Y_M), \tag{24.5.1}$$

and thinking of our $K$-groups in terms of coherent sheaves, we have a push-forward map

$$(i_M)_* : K(Y_M) \to K(V). \tag{24.5.2}$$

We now define our $K$-theoretic weight factor $\tilde{\mathbf{v}}_M$ as follows:

$$\tilde{\mathbf{v}}_M(g_1, g_2) := |\mathcal{Z}_M|^{-1} (i_M)_* i_M^* [\mathcal{L}_{(g_1,g_2)}] \in K(V)_{\mathbb{C}}, \tag{24.5.3}$$

where $[\mathcal{L}_{(g_1,g_2)}]$ denotes the class in $K$-theory of the line bundle $\mathcal{L}_{(g_1,g_2)}$ on $V$ obtained by taking $\mathcal{L}_{(g_1,g_2;\mu)}$ for $\mu = 0$.

Pushing forward from $V$ to a point, we get a homomorphism

$$EP : K(V) \to \mathbb{Z}, \tag{24.5.4}$$

whose value on the class of a coherent sheaf $\mathcal{F}$ is

$$\sum_i (-1)^i \dim H^i(V, \mathcal{F}).$$

Our definition of $\tilde{\mathbf{v}}_M$ was designed so that

$$EP(\tilde{\mathbf{v}}_M(g_1, g_2)) = \tilde{v}_M(g_1, g_2; 0). \tag{24.5.5}$$

Since the local trace formula holds for the weight factors $\tilde{v}_M(g_1, g_2; 0)$, it is reasonable to hope that it will also hold for the weight factors $\tilde{\mathbf{v}}_M(g_1, g_2)$ (as an equality between two elements in $K(V)_{\mathbb{C}}$), and we will now check that this really is the case.

For this we need to check that for all linear functionals $\lambda : K(V)_{\mathbb{C}} \to \mathbb{C}$, the local trace formula holds for the weight factors $v_M^\lambda$ defined by

$$v_M^\lambda(g_1, g_2) := \langle \lambda, \tilde{\mathbf{v}}_M(g_1, g_2) \rangle.$$

So far we know only that this is true for $\lambda = EP$.

Of course we only need to consider a collection of linear functionals $\lambda$ that spans the vector space dual to $K(V)_{\mathbb{C}}$. We now define such a collection of linear functionals $\lambda_P$, one for each parabolic subgroup $P = LU \in \mathcal{F}(A)$ (with $L$ being the unique Levi component of $P$ that contains $A$). Recall that inside $V$ we have the non-singular closed subvariety $\overline{V}_P = V^L$, the toric variety associated to $L$. We define $\lambda_P$ (as the complex linear extension of)

$$K(V) \to K(\overline{V}_P) \to \mathbb{Z},$$

where the first map is pull-back (restriction) from $V$ to $\overline{V}_P$ and the second is push-forward from $\overline{V}_P = V^L$ to a point.

It follows from Lemma 23.1 that

$$v_M^{\lambda_P}(g_1, g_2) = \begin{cases} \tilde{v}_M^L(l_1, l_2; 0) & \text{if } M \subset L \\ 0 & \text{otherwise} \end{cases} \tag{24.5.6}$$

where $\tilde{v}_M^L$ is the weight factor for the Levi subgroup $M$ of $L$, and where we have used the Iwasawa decomposition to decompose $g_i$ $(i = 1, 2)$ as $g_i = l_i u_i k_i$ for $l_i \in L$, $u_i \in U$, $k_i \in K$.

Using Lemma 13.3 and applying the local trace formula for $L$ (with weight factors $\tilde{v}_M^L(g_1, g_2; 0)$) to the functions $f_1^{(P)}$ and $f_2^{(P)}$ on $\mathrm{Lie}(L)$, we see that the local trace formula holds for the weight factors $v_M^{\lambda_P}$ and hence for the $K$-theoretic weight factors $\tilde{\mathbf{v}}_M$. Actually, for this we should check that the various measures we are using in our integrals on $L$ (and in the definition of $f \mapsto f^{(P)}$) are compatible with the ones we are using in our integrals on $G$, but we are going to omit this point.

**24.6. Weight factors $\mathbf{v}_M$ and $v_M$.** Finally we come to the weight factors $v_M$ we really want, those defined using volumes of convex hulls. These are related to our $K$-theoretic weight factors in the following way. The Chern character $ch$ induces an isomorphism (of $\mathbb{C}$-algebras)

$$ch : K(V)_{\mathbb{C}} \simeq H^\bullet(V, \mathbb{C}).$$

We write $\mathbf{v}_M(g_1, g_2) \in H^\bullet(V, \mathbb{C})$ for the image of the weight factor $\tilde{\mathbf{v}}_M(g_1, g_2) \in K(V)_{\mathbb{C}}$ under the Chern character isomorphism $ch$. Since the local trace formula holds for the weight factors $\tilde{\mathbf{v}}_M$, it also holds for the weight factors $\mathbf{v}_M$.

Consider the linear functional $\lambda$ on $H^\bullet(V, \mathbb{C})$ projecting $H^\bullet(V, \mathbb{C})$ onto its summand $H^{2r}(V, \mathbb{C}) = \mathbb{C}$ of top degree (with $r = \dim V = \dim(A/A_G)$). Define yet another weight factor by

$$v_M(g_1, g_2) := \langle \lambda, \mathbf{v}_M(g_1, g_2) \rangle. \tag{24.6.1}$$

Obviously the local trace formula holds for the weight factors $v_M$. Our next goal is to express $v_M$ in terms of volumes of convex hulls. We cannot do this without a better understanding of (24.6.1) when $M \neq A$, so we will rewrite (24.6.1) in terms of the cohomology ring $H^\bullet(Y_M, \mathbb{C})$ of $Y_M$. Consider the diagram

$$\begin{tikzcd} K(Y_M)_{\mathbb{C}} \arrow[r, "ch_M"] \arrow[d, "(i_M)_*"'] & H^\bullet(Y_M, \mathbb{C}) \arrow[d, "(i_M)_*"'] \\ K(V)_{\mathbb{C}} \arrow[r, "ch"] & H^\bullet(V, \mathbb{C}) \end{tikzcd} \tag{24.6.2}$$

in which the horizontal arrows are Chern character isomorphisms and the right vertical map is the usual Gysin push-forward map on cohomology, coming from the natural map $(i_M)_* : H_\bullet(Y_M, \mathbb{C}) \to H_\bullet(V, \mathbb{C})$ on homology after we use Poincaré duality to identify the cohomology of $V$ with its homology, and the same for $Y_M$. We claim that the diagram (24.6.2) is commutative. For this we need to consider the Riemann-Roch theorem for the morphism $i_M : Y_M \hookrightarrow V$. Let $N$ denote the normal bundle to $Y_M$ in $V$. To show that (24.6.2) commutes it is enough to show that the Todd class $td(N)$ is 1 (see the proof of [**Ful98**, Theorem 15.2]). In fact more is true: the normal bundle itself is trivial, as one sees from (23.2.1), which shows that the open neighborhood $U_M$ of $Y_M$ in $V$ is isomorphic to the product $S \times Y_M$ for any subtorus $S$ of $\hat{A}/Z(\hat{G})$ complementary to the subtorus $Z(\hat{M})/Z(\hat{G})$.

Consider the linear functional $\lambda_M$ on $H^\bullet(Y_M, \mathbb{C})$ projecting $H^\bullet(Y_M, \mathbb{C})$ onto its summand $H^{2s}(Y_M, \mathbb{C}) = \mathbb{C}$ of top degree (with $s = \dim Y_M = \dim(A_M/A_G)$). Then, as a consequence of the commutativity of the diagram (24.6.2), we have the equality

$$v_M(g_1, g_2) = |\mathcal{Z}_M|^{-1} \langle \lambda_M, ch_M[\mathcal{M}_{(g_1,g_2)}] \rangle, \tag{24.6.3}$$

where $\mathcal{M}_{(g_1,g_2)}$ is the restriction of $\mathcal{L}_{(g_1,g_2)}$ to $Y_M$, so that the corresponding $(G, M)$-orthogonal set in $\Lambda_M$ is $H_{\bar{P}}(g_1) - H_P(g_2)$.

**24.7. Volumes of positive orthogonal sets.** We need to specify the measures with respect to which we take our volumes. Consider a positive $(G, M)$-orthogonal set of points $(y_P)_{P \in \mathcal{P}(M)}$ in $\Lambda_M$. The points $y_P$ all have the same image in $\Lambda_G$. Pick $y \in \Lambda_M$ having the same image in $\Lambda_G$ as all the points $y_P$. Then the translated points $y_P - y$ all lie in the subgroup $\Lambda_M^0 := X^*(Z(\hat{M})/Z(\hat{G}))$ of $X^*(Z(\hat{M})) = \Lambda_M$. Since $\Lambda_M^0$ is a free abelian group, there is a canonical translation invariant measure on $\Lambda_M^0 \otimes_{\mathbb{Z}} \mathbb{R}$, namely the one that gives measure 1 to any fundamental domain for $\Lambda_M^0$. By definition, we take the volume of the convex hull of the points $y_P$ to be the volume in $\Lambda_M^0 \otimes_{\mathbb{Z}} \mathbb{R}$ (for the measure we just defined) of the translated points $y_P - y$. (Clearly this is independent of the choice of $y$.)

**24.8. Computation of $\langle \lambda_M, ch_M[\mathcal{M}] \rangle$ for certain line bundles $\mathcal{M}$.** Let $\mathcal{M}$ be a $Z(\hat{M})$-equivariant line bundle on $Y_M$ with associated $(G, M)$-orthogonal set $(y_P)_{P \in \mathcal{P}(M)}$ in $\Lambda_M$. It is known (see [**Ful93**]) that when the orthogonal set $y_P$ is positive, $\langle \lambda_M, ch_M[\mathcal{M}] \rangle$ is equal to the volume of the convex hull of the points $y_P$.

The map $\mathrm{Pic}(Y_M) \to K(Y_M)_{\mathbb{C}} \xrightarrow{ch_M} H^\bullet(Y_M, \mathbb{C}) \xrightarrow{\lambda_M} \mathbb{C}$ is a homogeneous polynomial function of degree $\dim Y_M = \dim(A_M/A_G)$. Therefore, if $y_P$ is a *negative* orthogonal set (in the sense that $-y_P$ is a positive orthogonal set), $\langle \lambda_M, ch_M[\mathcal{M}] \rangle$ is equal to $(-1)^{\dim(A_M/A_G)}$ times the volume of the convex hull of the points $y_P$. (There is no need to replace the points by their negatives since this does not affect the volume.)

**24.9. Computation of $v_M$ in terms of volumes of convex hulls.** Since $H_{\bar{P}}(g_1) - H_P(g_2)$ is a negative orthogonal set, we conclude that

$$v_M(g_1, g_2) = (-1)^{\dim(A_M/A_G)}|\mathcal{Z}_M|^{-1} \operatorname{vol}\big(\operatorname{Hull}\{H_{\bar{P}}(g_1) - H_P(g_2) : P \in \mathcal{P}(M)\}\big).$$

Thus, for these weight factors $v_M$ we have the final version of the local trace formula:

THEOREM 24.1 (Waldspurger [**Wal95**]). *Let $f_1, f_2 \in C_c^\infty(\mathfrak{g})$. Then $T(f_1, f_2) = T(\hat{f}_1, \check{f}_2)$, where*

$$\begin{aligned} T(f_1, f_2) = \sum_{M \in \mathcal{L}} \frac{|W_M|}{|W|} \sum_{T \in \mathcal{T}_M} \frac{1}{n_T^M} \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| \cdot \\ \int_{A_M\backslash G} \int_{A_M\backslash G} f_1(h^{-1}Xh) f_2(g^{-1}Xg) v_M(h, g)\, d\dot{h}\, d\dot{g}\, dX. \end{aligned} \tag{24.9.1}$$

**24.10. Splitting.** Recall that $\mathcal{M}_{(g_1,g_2)}$ is the $Z(\hat{M})$-equivariant line bundle on $Y_M$ associated to the negative $(G, M)$-orthogonal set $P \mapsto H_{\bar{P}}(g_1) - H_P(g_2)$. Thus it is natural to introduce (for $g \in G$) the $Z(\hat{M})$-equivariant line bundles $\mathcal{M}'_g$ and $\mathcal{M}_g$ associated to the negative $(G, M)$-orthogonal sets $B \mapsto H_{\bar{P}}(g)$ and $P \mapsto -H_P(g)$ respectively, as well as the elements

$$\mathbf{v}_M(g) := ch_M[\mathcal{M}_g] \in H^\bullet(Y_M, \mathbb{C}), \tag{24.10.1}$$

$$\mathbf{v}'_M(g) := ch_M[\mathcal{M}'_g] \in H^\bullet(Y_M, \mathbb{C}). \tag{24.10.2}$$

These definitions are set up so that

$$\mathbf{v}_M(g_1, g_2) = |\mathcal{Z}_M|^{-1}(i_M)_*\big(\mathbf{v}'_M(g_1) \cdot \mathbf{v}_M(g_2)\big), \tag{24.10.3}$$

the product on the right being taken in the cohomology ring $H^\bullet(Y_M, \mathbb{C})$.

Now let $T \in \mathcal{T}_M$ and let $X \in \mathfrak{t}_{\mathrm{reg}}$. Then for $f \in C_c^\infty(\mathfrak{g})$ we can define normalized weighted orbital integrals $\mathcal{J}_X(f) = \mathcal{J}_X^G(f)$ and $\mathcal{J}'_X(f) = (\mathcal{J}')_X^G(f)$ taking values in $H^\bullet(Y_M, \mathbb{C})$ by putting

$$\mathcal{J}_X(f) := |D(X)|^{1/2} \int_{A_M\backslash G} f(g^{-1}Xg)\mathbf{v}_M(g)\, d\dot{g}, \tag{24.10.4}$$

$$\mathcal{J}'_X(f) := |D(X)|^{1/2} \int_{A_M\backslash G} f(g^{-1}Xg)\mathbf{v}'_M(g)\, d\dot{g}. \tag{24.10.5}$$

These definitions are set up so that the expression

$$\mathcal{J}_X(f_1, f_2) := |D(X)| \int_{A_M\backslash G} \int_{A_M\backslash G} f_1(g_1^{-1}Xg_1) f_2(g_2^{-1}Xg_2) v_M(g_1, g_2)\, d\dot{g}_1\, d\dot{g}_2$$

occurring in Theorem 24.1 is given by

$$\mathcal{J}_X(f_1, f_2) = |\mathcal{Z}_M|^{-1}\langle \lambda_M, \mathcal{J}'_X(f_1) \cdot \mathcal{J}_X(f_2)\rangle, \tag{24.10.6}$$

the product $\mathcal{J}'_X(f_1) \cdot \mathcal{J}_X(f_2)$ being taken in $H^\bullet(Y_M, \mathbb{C})$.

By parabolic descent (see Lemma 13.3) for any parabolic subgroup $P = LU \in \mathcal{F}(M)$ (with $L \supset M$), the image of $\mathcal{J}_X(f) \in H^\bullet(Y_M, \mathbb{C})$ under the map

$$H^\bullet(Y_M, \mathbb{C}) \to H^\bullet(Y_M^L, \mathbb{C})$$

induced by $Y_M^L = \overline{V}_P \cap Y_M \hookrightarrow Y_M$ is equal to $\mathcal{J}_X^L(f^{(P)}) \in H^\bullet(Y_M^L, \mathbb{C})$, and the analogous statement holds for $\mathcal{J}'_X$.

Now assume that $f$ is a cusp form (see 13.2), so that $f^P = 0$ (and hence $f^{(P)} = 0$) for all $P \neq G$. Then, since the fundamental classes of $\overline{V}_P \cap Y_M$ ($P \in \mathcal{F}(M)$, $P \neq G$) in the homology groups $H_\bullet(Y_M, \mathbb{C})$ span (see [**Ful93**]) the subspace

$$\sum_{i=0}^{2\dim Y_M - 1} H_i(Y_M, \mathbb{C})$$

of $H_\bullet(Y_M, \mathbb{C})$, we see that for such $f$ the weighted orbital integrals $\mathcal{J}_X(f)$, $\mathcal{J}'_X(f)$ lie in the top degree subspace $H^{2\dim(Y_M)}(Y_M, \mathbb{C}) = \mathbb{C}$ of $H^\bullet(Y_M, \mathbb{C})$.

Therefore for any $f_1 \in C_c^\infty(\mathfrak{g})$ and any cusp form $f_2 \in C_c^\infty(\mathfrak{g})$ the product $\mathcal{J}'_X(f_1)\mathcal{J}_X(f_2)$ is equal to $\mathcal{J}_X(f_2)$ times the projection of $\mathcal{J}'_X(f_1)$ on $H^0(Y_M, \mathbb{C}) = \mathbb{C}$. Since the projection of $\mathbf{v}'_M(g_1)$ on $H^0(Y_M, \mathbb{C})$ is obviously $1 \in \mathbb{C} = H^0(Y_M, \mathbb{C})$, we conclude that

$$\mathcal{J}'_X(f_1)\mathcal{J}_X(f_2) = I_X(f_1)\mathcal{J}_X(f_2), \tag{24.10.7}$$

where $I_X(f)$ is the normalized orbital integral

$$I_X(f) := |D(X)|^{1/2} \int_{A_M\backslash G} f(g^{-1}Xg)\, d\dot{g}.$$

From (24.10.6) and (24.10.7) we conclude that when $f_2$ (and hence $\check{f}_2$) is a cusp form, the local trace formula (see Theorem 24.1) reduces to the statement that

$$T_{\text{cusp}}(f_1, f_2) = T_{\text{cusp}}(\hat{f}_1, \check{f}_2), \tag{24.10.8}$$

with $T_{\text{cusp}}$ defined by

$$\begin{aligned} T_{\text{cusp}}(f_1, f_2) := \sum_{M\in\mathcal{L}} (-1)^{\dim(A_M/A_G)} |\mathcal{Z}_M|^{-1} \frac{|W_M|}{|W|} \sum_{T\in\mathcal{T}_M} \frac{1}{n_T^M} \cdot \\ \int_{\mathfrak{t}_{\text{reg}}} |D(X)| O_X(f_1) WO_X(f_2)\, dX, \end{aligned} \tag{24.10.9}$$

where

$$O_X(f) = \int_{A_M\backslash G} f(g^{-1}Xg)\, d\dot{g}, \tag{24.10.10}$$

$$WO_X(f) = \int_{A_M\backslash G} f(g^{-1}Xg) v_M(g)\, d\dot{g}, \tag{24.10.11}$$

$v_M(g)$ being (as in earlier sections of this article) the volume of the convex hull of $\{H_P(g) : P \in \mathcal{P}(M)\}$.

## 25. An important application of the local trace formula

Following Waldspurger [**Wal95**], we are going to use the local trace formula on the Lie algebra to strengthen a result of Harish-Chandra [**HC78**] that is the first key step in his proof that the distribution characters of irreducible representations of $G$ are represented by locally integrable functions.

**25.1. Definition of support of a distribution.** Let's recall the notion of support of a distribution on an l.c.t.d space $X$. For this we need to remember that for open subsets $U \subset V$ of $X$ there is a restriction map

$$\mathcal{D}(V) \to \mathcal{D}(U)$$

on distributions that is dual to the inclusion

$$C_c^\infty(U) \to C_c^\infty(V).$$

With these restriction maps $U \mapsto \mathcal{D}(U)$ is a sheaf of vector spaces on $X$. In particular, given a distribution $D$ on $X$, there is a biggest open subset $U$ of $X$ such that the restriction of $D$ to $U$ is zero. The complement $Y$ of $U$ is called the *support* of $D$. Thus $Y$ is the smallest closed subset of $X$ for which $D$ is in the image of the embedding $\mathcal{D}(Y) \hookrightarrow \mathcal{D}(X)$.

**25.2. Definition of the invariant distribution $I_\phi$ on $\mathfrak{g}$.** Let $\phi \in C_c^\infty(\mathfrak{g})$ be a cusp form. For any $f \in C_c^\infty(\mathfrak{g})$ put

$$I_\phi(f) := T_{\mathrm{cusp}}(f, \phi), \tag{25.2.1}$$

with $T_{\mathrm{cusp}}$ as in (24.10.9) above. Thus $I_\phi$ is a well-defined distribution on $\mathfrak{g}$. It is clear from the definition that $I_\phi$ is invariant and supported on the closure of the set of $G$-conjugates of elements in the compact set $\mathrm{Supp}(\phi)$. In particular the support of $I_\phi$ is bounded modulo conjugation (see 15.2 for this notion).

Recall that the Fourier transform $\hat{T}$ of a distribution $T$ on $\mathfrak{g}$ is defined so that $\hat{T}(f) = T(\hat{f})$ for all test functions $f \in C_c^\infty(\mathfrak{g})$. Since the Fourier transform commutes with adjoint $G$-action, it takes invariant distributions to invariant distributions.

The next result makes use of the notion of nice conjugation invariant function on $\mathfrak{g}$ (discussed in 13.8).

THEOREM 25.1. *Let $\phi$ be a cusp form on $\mathfrak{g}$. Then $\hat{\phi}$ is also a cusp form, and there is an equality*

$$\hat{I}_\phi = I_{\hat{\phi}}. \tag{25.2.2}$$

*Moreover the distribution $I_\phi$ is represented by the nice conjugation invariant function $F_\phi$ on $\mathfrak{g}$ which is 0 off $\mathfrak{g}_{\mathrm{rs}}$ and whose value at any $X \in \mathfrak{t}_{\mathrm{reg}}$ for any $M \in \mathcal{L}$ and $T \in \mathcal{T}_M$ is given by*

$$F_\phi(X) = (-1)^{\dim(A_M/A_G)} |\mathcal{Z}_M|^{-1} WO_X(\phi). \tag{25.2.3}$$

PROOF. We observed long ago that the Fourier transform of a cusp form is a cusp form (see 13.2). The equality (25.2.2) follows from (24.10.8). The second statement of the theorem follows from (24.10.9), the Weyl integration formula, and the local constancy of $WO_X(\phi)$ as a function of $X \in \mathfrak{t}_{\mathrm{reg}}$ (see Theorem 17.11). $\square$

**25.3. Remarks.** This theorem can be regarded as a Lie algebra analog of a result of Arthur [**Art87**] which says that (up to a sign) the weighted orbital integrals of a matrix coefficient of a supercuspidal representation give the character values of that representation. In particular a cusp form $\phi$ on $\mathfrak{g}$ should be regarded as being analogous to a matrix coefficient for a supercuspidal representation of $G$.

**25.4. The special case in which $\phi \in C_c^\infty(\mathfrak{g}_e)$.** As before (see 10.4), we write $\mathfrak{g}_e$ for the open subset of $\mathfrak{g}$ consisting of elements whose centralizers are elliptic maximal tori in $G$. Let $\phi \in C_c^\infty(\mathfrak{g}_e)$. Clearly $\phi$ is a cusp form on $\mathfrak{g}$, so we can consider the invariant distribution $I_\phi$ of 25.2. As an immediate consequence of Theorem 25.1 we obtain the following result of Harish-Chandra and Waldspurger.

THEOREM 25.2 ([**HC78, Wal95**]). *The invariant distribution $\hat{I}_\phi$ is represented by a nice conjugation invariant function whose value at any $X \in \mathfrak{t}$ for any $M \in \mathcal{L}$, $T \in \mathcal{T}_M$ is given by*

$$(-1)^{\dim(A_M/A_G)}|\mathcal{Z}_M|^{-1} WO_X(\hat{\phi}). \tag{25.4.1}$$

Harish-Chandra proved that $\hat{I}_\phi$ is represented by a nice conjugation invariant function and proved the formula above for its value at elliptic elements $X \in \mathfrak{g}$, for which the weighted orbital integral reduces to an ordinary orbital integral. The formula (25.4.1) in the case of non-elliptic elements is due to Waldspurger.

We should note that because $\phi$ is supported on the elliptic set $\mathfrak{g}_e$, we get from (24.10.9) the following simple formula

$$I_\phi(f) = |\mathcal{Z}_G|^{-1} \sum_{T \in \mathcal{T}_G} \frac{1}{n_T^G} \cdot \int_{\mathfrak{t}_{\mathrm{reg}}} |D(X)| O_X(\phi) O_X(f)\, dX, \tag{25.4.2}$$

which can also be rewritten (using the Weyl integration formula) as

$$I_\phi(f) = |\mathcal{Z}_G|^{-1} \int_{A_G\backslash G} \int_{\mathfrak{g}} \phi(X) f(g^{-1}Xg)\, dX\, d\dot{g}. \tag{25.4.3}$$

## 26. Niceness of $\hat{O}_X$ for $X \in \mathfrak{g}_{\mathrm{rs}}$.

**26.1. Goal.** We see from (25.4.2) that the distribution $I_\phi$ is an integral of distributions of the form $O_X$ for $X \in \mathfrak{g}_e$. By varying $\phi$ we get many such integrals, and for each one we know that its Fourier transform $\hat{I}_\phi$ is represented by a nice conjugation invariant function. This suggests that for any $X \in \mathfrak{g}_e$ the Fourier transform $\hat{O}_X$ of $O_X$ is represented by a nice conjugation invariant function on $\mathfrak{g}$. In fact this is true, and is the main step in the proof of the following more general result of Harish-Chandra (which in turn is a special case of the yet more general result Theorem 27.8, again due to Harish-Chandra):

THEOREM 26.1 ([**HC78, HC99**]). *For every $X \in \mathfrak{g}_{\mathrm{rs}}$ the Fourier transform $\hat{O}_X$ of the orbital integral $O_X$ is represented by a nice conjugation invariant function on $\mathfrak{g}$.*

For the time being we remark only that it suffices to prove the theorem in case $X$ is elliptic. The general case will then follow from Lemma 13.2, Lemma 13.4, and equation (13.12.1). In 26.5 we will use Theorem 25.2 to treat the elliptic case. This will require Howe's finiteness theorem, to be discussed next.

**26.2. Howe's finiteness theorem.** Before stating Howe's finiteness theorem for $\mathfrak{g}$, we need a few preliminary remarks.

Let $V$ be a subset of $\mathfrak{g}$ that is conjugation invariant and bounded modulo conjugation (see 15.2 for this notion). We denote by $J(V)$ the space of invariant distributions on $\mathfrak{g}$ whose support is contained in $V$.

Now let $L$ be a lattice in $\mathfrak{g}$. Inside $C_c^\infty(\mathfrak{g})$ we have the subspace $C_c(\mathfrak{g}/L)$ consisting of functions that are compactly supported and translation invariant under $L$. There is of course a restriction map

$$\mathcal{D}(\mathfrak{g}) \to C_c(\mathfrak{g}/L)^* \tag{26.2.1}$$

where $C_c(\mathfrak{g}/L)^*$ denotes the vector space dual to $C_c(\mathfrak{g}/L)$.

Now we can state Howe's finiteness theorem, proved by Howe [**How73**] for $GL_n$ and by Harish-Chandra in the general case [**HC78**]. (There is an analogous result for $G$, known as Howe's conjecture, which was proved by Clozel [**Clo89**].)

THEOREM 26.2. *For any lattice $L$ in $\mathfrak{g}$ and any subset $V$ of $\mathfrak{g}$ that is conjugation invariant and bounded modulo conjugation, the image of $J(V)$ under the restriction map* (26.2.1) *is finite dimensional.*

We will use this theorem without proving it. For additional insight into why Howe's finiteness theorem is useful, see DeBacker's article in this volume.

We also need to understand what the theorem says in the Fourier transformed picture. The Fourier transform gives an isomorphism

$$C_c^\infty(\mathfrak{g}) \xrightarrow{FT} C_c^\infty(\mathfrak{g}),$$

and this isomorphism restricts to an isomorphism

$$C_c(\mathfrak{g}/L) \xrightarrow{FT} C_c^\infty(L^\perp),$$

where $L^\perp$ is the lattice in $\mathfrak{g}$ that is Pontryagin dual to $L$. (When we view elements of $\mathfrak{g}$ as characters on $\mathfrak{g}$, the lattice $L^\perp$ consists of all those characters that are trivial on $L$.) Since any lattice arises as the Pontryagin dual of some other lattice, Howe's theorem can be reformulated as follows.

THEOREM 26.3. *For any lattice $L$ in $\mathfrak{g}$ and any subset $V$ of $\mathfrak{g}$ that is conjugation invariant and bounded modulo conjugation, the image of $J(V)$ under the composed map*

$$\mathcal{D}(\mathfrak{g}) \xrightarrow{FT} \mathcal{D}(\mathfrak{g}) \xrightarrow{\text{res}} \mathcal{D}(L) \tag{26.2.2}$$

*is finite dimensional. The first map is the Fourier transform on distributions, and the second map is given by restriction of distributions from $\mathfrak{g}$ to its open subset $L$.*

**26.3. Topology on $V^*$.** When using Howe's finiteness theorem, it is useful to topologize $\mathcal{D}(\mathfrak{g})$. The topology we use is of the following type.

Let $V$ be any complex vector space (not assumed to be finite dimensional as the example we have in mind is $C_c^\infty(\mathfrak{g})$). For any subspace $U$ of $V$ we let $U^\perp$ denote the subspace of $V^*$ consisting of all linear forms that vanish on $U$. Similarly, for any subspace $W$ of $V^*$ we let $W^\perp$ denote the subspace of $V$ consisting of all vectors $v$ that vanish on $W$.

We write $V$ as the direct limit of its finite dimensional subspaces $U$. The dual space $V^*$ is then the projective limit of the dual spaces $U^* = V^*/U^\perp$. We give each dual space $U^*$ the discrete topology and then use the projective limit topology on $V^*$. In the topological vector space $V^*$ the subgroups $U^\perp$ are open and form a neighborhood base at 0. Thus two linear forms are close to each other if they agree on a large finite dimensional subspace of $V$. This topology is obviously Hausdorff. When $V$ is finite dimensional, $V^*$ has the discrete topology.

Taking $V = C_c^\infty(\mathfrak{g})$, we get the desired topology on $\mathcal{D}(\mathfrak{g}) = V^*$.

LEMMA 26.4. *The topology above has the following properties.*

(1) *For any linear map $f : V_1 \to V_2$ the dual map $f^* : V_2^* \to V_1^*$ is continuous, and its kernel and image are closed. If $f$ is surjective, then $f^*$ is a homeomorphism of $V_2^*$ onto a closed subspace of $V_1^*$.*
(2) *Let $W$ be any subspace of $V^*$. Then the closure of $W$ is $W^{\perp\perp}$.*
(3) *Any finite dimensional subspace $W$ of $V^*$ is closed, and the topology it inherits from $V^*$ is discrete.*

PROOF. We leave the first two items as exercises for the reader. To prove the last item, first note that the natural map $V \to W^*$ is surjective, then apply the second statement of the first item to this surjection. □

Combining the last statement of the lemma above with the Fourier transformed version of Howe's theorem (Theorem 26.3), we get the following useful result.

PROPOSITION 26.1. *Let $V$ be a conjugation invariant subset of $\mathfrak{g}$ that is bounded modulo conjugation, and let $U_1 \subset U_2$ be subspaces of $J(V)$ such that $U_2$ is contained in the closure of $U_1$. Let $L$ be any lattice in $\mathfrak{g}$. Then $U_1$ and $U_2$ have the same image under the composed map*

$$\mathcal{D}(\mathfrak{g}) \xrightarrow{FT} \mathcal{D}(\mathfrak{g}) \xrightarrow{\mathrm{res}} \mathcal{D}(L). \tag{26.3.1}$$

**26.4. Elliptic regular orbital integrals as limits of distributions $I_\phi$.** The next result illustrates how the topology on $\mathcal{D}(\mathfrak{g})$ works. It will be needed to complete the proof of Theorem 26.1.

LEMMA 26.5. *Let $T$ be an elliptic maximal torus in $G$, let $X \in \mathfrak{t}_{\mathrm{reg}}$, and let $\omega_T$ be a compact open neighborhood of $X$ in $\mathfrak{t}_{\mathrm{reg}}$. Then $O_X$ lies in the closure of the linear subspace*

$$\{I_\phi : \phi \in C_c^\infty(\mathrm{Ad}(G)(\omega_T))\}$$

*of $\mathcal{D}(\mathfrak{g})$.*

PROOF. We may shrink $\omega_T$ as needed. Recall the map

$$(T\backslash G) \times \mathfrak{t}_{\mathrm{reg}} \to \mathfrak{g}_{\mathrm{rs}} \tag{26.4.1}$$

(sending $(g, Y)$ to $g^{-1}Yg$) that we used when proving the Weyl integration formula. Its differential is an isomorphism at all points. Shrinking $\omega_T$, we can find a compact open neighborhood $\omega_{T\backslash G}$ of $1 \in T\backslash G$ and a compact open neighborhood $\omega_G$ of $X$ in $\mathfrak{g}_{\mathrm{rs}}$ such that the map (26.4.1) restricts to an isomorphism (of $p$-adic manifolds)

$$\omega_{T\backslash G} \times \omega_T \to \omega_G. \tag{26.4.2}$$

For $\phi \in C_c^\infty\big(\mathrm{Ad}(G)(\omega_T)\big)$ and any $f \in C_c^\infty(\mathfrak{g})$, we see from (25.4.2) that

$$I_\phi(f) = |\mathcal{Z}_G|^{-1} \int_{\omega_T} |D(Y)| O_Y(f) O_Y(\phi)\, dY. \tag{26.4.3}$$

Suppose that $f \in C_c^\infty(\mathfrak{g})$ is such that $I_\phi(f) = 0$ for all $\phi \in C_c^\infty\big(\mathrm{Ad}(G)(\omega_T)\big)$. In order to prove the lemma, we need to check that $O_X(f) = 0$. In fact we will check that $O_Y(f) = 0$ for all $Y \in \omega_T$. Indeed, since the integral in (26.4.3) vanishes for all $\phi \in C_c^\infty\big(\mathrm{Ad}(G)(\omega_T)\big)$, it is enough to convince ourselves that every locally constant function $\varphi$ on $\omega_T$ arises as $Y \mapsto O_Y(\phi)$ for some $\phi \in C_c^\infty\big(\mathrm{Ad}(G)(\omega_T)\big)$. But this is clear: pull $\varphi$ back to $\omega_{T\backslash G} \times \omega_T$ using the second projection, view this pullback as a function on $\omega_G$ using the isomorphism (26.4.2), and then divide by meas($\omega_{T\backslash G}$) to get a function $\phi \in C_c^\infty(\omega_G)$ that does the job. (Note that $\omega_G$ is an open subset of $\mathrm{Ad}(G)(\omega_T)$.) □

**26.5. Proof of Theorem 26.1.** As we already remarked, it suffices to prove that $\hat{O}_X$ is represented by a nice conjugation invariant function in the case when $X$ lies in $\mathfrak{t}_{\mathrm{reg}}$ for some elliptic maximal torus $T$. Let $\omega_T$ be a compact open neighborhood of $X$ in $\mathfrak{t}_{\mathrm{reg}}$. Put $\omega := \mathrm{Ad}(G)(\omega_T)$, a subset which is clearly bounded modulo conjugation.

By Lemma 26.5 $O_X$ lies in the closure of the space of distributions $I_\phi$ with $\phi \in C_c^\infty(\omega)$. Moreover the distributions $O_X$ and $I_\phi$ all lie in $J(\omega)$. Applying Proposition 26.1 to $J(\omega)$, we see that there exists $\phi \in C_c^\infty(\omega)$ such that $\hat{O}_X$ and $\hat{I}_\phi$ have the same restriction to $L$. From Theorem 25.2 we know that the restriction of $\hat{I}_\phi$ to $L$ is represented by an integrable function that is locally constant on $L \cap \mathfrak{g}_{\mathrm{rs}}$. Therefore the same is true of $\hat{O}_X$. This proves the theorem since the collection of all lattices covers $\mathfrak{g}$.

## 27. Deeper results on Shalika germs; Lie algebra analog of the local character expansion

Harish-Chandra's Theorem 26.1, together with Howe's finiteness theorem, will allow us to prove quite a number of deep results in harmonic analysis on $\mathfrak{g}$.

**27.1. Density of orbital integrals.** Our next main goal is to prove the linear independence of Shalika germs. This is closely related, as we will see, to the density of regular semisimple orbital integrals. The first step is the density of all orbital integrals. What do we mean by this? Recall that we have topologized the space of distributions on $\mathfrak{g}$. We consider the subspace $\mathcal{D}(\mathfrak{g})^G$ of invariant distributions with its inherited topology. Inside of $\mathcal{D}(\mathfrak{g})^G$ we have the linear subspace $\mathcal{D}(\mathfrak{g})_{\mathrm{orb}}$ spanned by all orbital integrals $O_X$ ($X \in \mathfrak{g}$). Now we can state the density result, but one should bear in mind that it is only of temporary interest, since we will soon prove a stronger (and more difficult) statement.

Proposition 27.1 ([**HC78**]). *The subspace $\mathcal{D}(\mathfrak{g})_{\mathrm{orb}}$ is dense in $\mathcal{D}(\mathfrak{g})^G$.*

Proof. As we saw when we discussed (in 26.3) the topology on duals of vector spaces, the statement we need to prove can be reformulated as follows. Let $f \in C_c^\infty(\mathfrak{g})$. If $O_X(f) = 0$ for all $X \in \mathfrak{g}$, then $I(f) = 0$ for every invariant distribution $I$ on $\mathfrak{g}$. In terms of coinvariants $C_c^\infty(\mathfrak{g})_G$, this can in turn be reformulated as the statement that if $O_X(f) = 0$ for all $X \in \mathfrak{g}$, then the image of $f$ in $C_c^\infty(\mathfrak{g})_G$ is 0.

So we need a better understanding of $C_c^\infty(\mathfrak{g})_G$. For this we will again use the map $\pi_G : \mathfrak{g} \to \mathbb{A}_G(F)$. For $x \in \mathbb{A}_G(F)$ we denote by $\mathfrak{g}_x$ the fiber $\pi_G^{-1}(x)$ over $x$. The conjugation action of $G$ preserves $\mathfrak{g}_x$, so we can also consider the coinvariants $C_c^\infty(\mathfrak{g}_x)_G$. Restriction of functions to the fiber induces a surjective map

$$C_c^\infty(\mathfrak{g})_G \to C_c^\infty(\mathfrak{g}_x)_G, \tag{27.1.1}$$

and these can be assembled to give a map

$$C_c^\infty(\mathfrak{g})_G \to \prod_x C_c^\infty(\mathfrak{g}_x)_G, \tag{27.1.2}$$

where $x$ runs over all points $x \in \mathbb{A}_G(F)$. It follows from Lemma 27.1 below that the map (27.1.2) is injective.

When $x$ is the origin in the affine space, the fiber $\mathfrak{g}_x$ is the nilpotent cone in $\mathfrak{g}$, and we already have a good understanding of $C_c^\infty(\mathfrak{g}_x)_G$: its dimension is the number of nilpotent orbits, and the integrals over the nilpotent orbits provide a basis for the dual of $C_c^\infty(\mathfrak{g}_x)_G$

The situation for arbitrary $x$ is quite similar. The fiber $\mathfrak{g}_x$ is a finite union of $G$-orbits, and the integrals over these orbits provide a basis for the dual of $C_c^\infty(\mathfrak{g}_x)_G$. This is proved the same way as for the nilpotent cone, so we will not discuss it any further.

Now return to our function $f$. Since all orbital integrals of $f$ vanish by hypothesis, the image of $f$ under (27.1.2) is 0. Since (27.1.2) is injective, $f$ is 0 in $C_c^\infty(\mathfrak{g})_G$, and we are done. □

The next lemma is similar to the material in section 2.36 of [**BZ76**].

Lemma 27.1. *Let $X$ and $Y$ be l.c.t.d spaces, and let $f : X \to Y$ be a continuous map. For $y \in Y$ we denote by $X_y$ the fiber $f^{-1}(y)$. Suppose that an abstract group $G$ acts on $X$, preserving the fibers of $f$. Restriction of functions from $X$ to $X_y$ induces a map*

$$C_c^\infty(X)_G \to C_c^\infty(X_y)_G, \tag{27.1.3}$$

*and these can be assembled to give a map*

$$C_c^\infty(X)_G \to \prod_{y\in Y} C_c^\infty(X_y)_G. \tag{27.1.4}$$

*The map (27.1.4) is injective.*

*Moreover, for any open neighborhood $U$ of $y \in Y$ there is a surjective restriction map*

$$C_c^\infty(f^{-1}U)_G \to C_c^\infty(X_y)_G,$$

*and these fit together to give an isomorphism*

$$\varinjlim_U C_c^\infty(f^{-1}U)_G \cong C_c^\infty(X_y)_G, \tag{27.1.5}$$

*where the colimit is taken over the set of open neighborhoods $U$ of $y$.*

Proof. Replacing $Y$ by its 1-point compactification (which is again a l.c.t.d space), we may assume without loss of generality that $Y$ is compact.

Suppose that we have a decomposition of $Y$ as a disjoint union of open (hence closed) subsets $Y_i$ $(i \in I)$. Then

$$C_c^\infty(X) = \bigoplus_{i\in I} C_c^\infty(Y_i)$$

and therefore

$$C_c^\infty(X)_G = \bigoplus_{i\in I} C_c^\infty(Y_i)_G. \tag{27.1.6}$$

For any open neighborhood $U$ of $y \in Y$ there is a surjective restriction map

$$C_c^\infty(f^{-1}U) \to C_c^\infty(X_y),$$

and these fit together to give an isomorphism

$$\varinjlim_U C_c^\infty(f^{-1}U) \cong C_c^\infty(X_y), \tag{27.1.7}$$

where the colimit is taken over the set of open neighborhoods $U$ of $y$. Surjectivity is clear, but let's check injectivity. So suppose that we have a function $\phi \in C_c^\infty(f^{-1}U)$ whose restriction to $X_y$ is 0. Then the support $S$ of $\phi$ is a compact set disjoint from $X_y$, so its image $f(S)$ does not contain $y$. Let $V$ be the open subset of $U$ obtained by removing all points in the compact set $f(S)$. Then $\phi$ becomes 0 in $C_c^\infty(f^{-1}V)$, hence in the colimit.

Taking coinvariants in (27.1.7), we get the isomorphism (27.1.5) mentioned in the last statement of the lemma. (Coinvariants commute with arbitrary colimits.)

Now we finish the proof. Let $\phi \in C_c^\infty(X)$ and suppose that the image of $\phi$ under (27.1.4) is 0. By (27.1.5) for every $y \in Y$ there exists a compact open neighborhood $U_y$ of $y$ such that $\phi$ is 0 in $C_c^\infty(f^{-1}U_y)_G$. Since $Y$ is compact, it can be covered by finitely many compact open subsets $U_1, \dots, U_n$ such that $\phi$ is 0 in $C_c^\infty(f^{-1}U_i)_G$ for all $i$. Now put

$$Y_1 = U_1,\, Y_2 = U_2 \setminus U_1, \dots, Y_n = U_n \setminus (U_1 \cup \cdots \cup U_{n-1}).$$

Thus we have written $Y$ as a disjoint union of open subsets $Y_i$ such that $\phi$ is 0 in $C_c^\infty(f^{-1}Y_i)_G$ for all $i$. It follows from (27.1.6) that $\phi$ is 0 in $C_c^\infty(X)_G$, as desired. □

**27.2. Preliminary remarks regarding linear independence of Shalika germs.** We will soon be proving that the Shalika germs $\Gamma_1, \dots, \Gamma_r$ (attached to the nilpotent orbits $\mathcal{O}_i$) are linearly independent functions on $\mathfrak{g}_{\mathrm{rs}}$.

LEMMA 27.2. *Assume that the Shalika germs $\Gamma_1, \dots, \Gamma_r$ are linearly independent functions on $\mathfrak{g}_{\mathrm{rs}}$. Then for any open neighborhood $U$ of $0$ in $\mathfrak{g}$, the restrictions of $\Gamma_1, \dots, \Gamma_r$ to $U \cap \mathfrak{g}_{\mathrm{rs}}$ remain linearly independent.*

PROOF. Without loss of generality we may assume that $U$ is a lattice in $\mathfrak{g}$. Now we use homogeneity of Shalika germs. The additive semigroup of non-negative integers acts on $U \cap \mathfrak{g}_{\mathrm{rs}}$, with $j$ acting by multiplication by the scalar $\pi^{2j} \in F^\times$, and therefore acts on the space of functions on $U \cap \mathfrak{g}_{\mathrm{rs}}$ (the action of $j$ transforming a function $F(X)$ into $F(\pi^{2j}X)$).

By homogeneity of Shalika germs (see (17.7.1)) the restriction of $\Gamma_i$ to $U \cap \mathfrak{g}_{\mathrm{rs}}$ transforms under the character

$$j \mapsto q^{j \dim \mathcal{O}_i}$$

on our semigroup. But in any representation of our semigroup, vectors transforming under distinct characters are linearly independent. Thus, in order to prove linear independence of the restrictions of Shalika germs to $U \cap \mathfrak{g}_{\mathrm{rs}}$, it is enough to fix a non-negative integer $d$ and prove linear independence of the restrictions of the Shalika germs for all nilpotent orbits of dimension $d$. But all these germs are homogeneous of the same degree, namely $d$, so it is clear that any dependence relation that holds on the subset $U \cap \mathfrak{g}_{\mathrm{rs}}$ will also hold on the whole set $\mathfrak{g}_{\mathrm{rs}}$. □

Next we relate linear independence of Shalika germs to the problem of writing nilpotent orbital integrals as limits (for our usual topology on $\mathcal{D}(\mathfrak{g})$, see 26.3) of linear combinations of regular semisimple orbital integrals.

LEMMA 27.3. *The functions $\Gamma_1, \dots, \Gamma_r$ on the set $\mathfrak{g}_{\mathrm{rs}}$ are linearly independent if and only if all nilpotent orbital integrals $\mu_i$ lie in the closure of the linear span of the subset*

$$\{O_X : X \in \mathfrak{g}_{\mathrm{rs}}\}$$

*of $\mathcal{D}(\mathfrak{g})$.*

PROOF. First we need to recall that the closure occurring in the statement of the lemma is equal to the set of all distributions $I$ such that $I(f) = 0$ for all $f \in C_c^\infty(\mathfrak{g})$ such that $O_X(f) = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}}$.

($\Rightarrow$) Suppose that the functions $\Gamma_i$ are linearly independent. Given $f \in C_c^\infty(\mathfrak{g})$ such that $O_X(f) = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}}$, we must show that $\mu_i(f) = 0$ for all $i$. By Shalika germ theory there exists an open neighborhood $U$ of 0 in $\mathfrak{g}$ such that

$$O_X(f) = \sum_{i=1}^{r} \mu_i(f)\Gamma_i(X) \quad \forall\ X \in U \cap \mathfrak{g}_{\mathrm{rs}}. \tag{27.2.1}$$

Since the function $X \mapsto O_X(f)$ on $\mathfrak{g}_{\mathrm{rs}}$ is identically zero, and since the restrictions of the functions $\Gamma_i$ to the subset $U \cap \mathfrak{g}_{\mathrm{rs}}$ remain linearly independent, we see that $\mu_i(f) = 0$ for all $i$, as desired.

($\Leftarrow$) Consider a dependence relation $a_1\Gamma_1 + \cdots + a_r\Gamma_r = 0$. By linear independence of the distributions $\mu_i$ there exists $f \in C_c^\infty(\mathfrak{g})$ such that $\mu_i(f) = a_i$ for all $i$. By Shalika germ theory there exists an open neighborhood $U$ of 0 in $\mathfrak{g}$ such that

$$O_X(f) = 0 \quad \forall\ X \in U \cap \mathfrak{g}_{\mathrm{rs}}. \tag{27.2.2}$$

It follows from Lemma 15.3 that $\mathrm{Ad}(G)(U)$ contains a $G$-invariant open and closed neighborhood $V$ of the nilpotent cone. Multiplying $f$ by the characteristic function of $V$, we obtain a function $f' \in C_c^\infty(\mathfrak{g})$ such that

$$O_X(f') = \begin{cases} O_X(f) & \text{if } X \in V \\ 0 & \text{if } X \notin V. \end{cases} \tag{27.2.3}$$

Combining (27.2.2) with (27.2.3), we see that $O_X(f') = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}}$. Therefore, since we are assuming that nilpotent orbital integrals are in the closure of the span of the regular semisimple orbital integrals, we conclude that $\mu_i(f') = 0$ for all $i$. But, again using (27.2.3), we find that $a_i = \mu_i(f) = \mu_i(f')$, and we are done. $\square$

Now let $S$ be a semisimple element of $\mathfrak{g}$, and let $H = G_S$, $\mathfrak{h}$, $\mathfrak{h}'$ be as in 17.10. For nilpotent $Y \in \mathfrak{h}$ we write $\mu_{S+Y}$ for integration over the $G$-orbit of $S + Y$ in $\mathfrak{g}$. We write $Y_1, \dots, Y_s$ for representatives of the $H$-orbits of nilpotent elements in $\mathfrak{h}$.

LEMMA 27.4. *Assume that the Shalika germs $\Gamma_1^H, \dots, \Gamma_s^H$ for $H$ are linearly independent functions on $\mathfrak{h}_{\mathrm{rs}}$. Then for every $X \in \mathfrak{g}$ whose semisimple part is $S$, the distribution $O_X$ lies in the closure of the linear span of the subset*

$$\{O_{X'} : X' \in \mathfrak{g}_{\mathrm{rs}}\}$$

*of $\mathcal{D}(\mathfrak{g})$.*

PROOF. This proof is almost the same as that of half of the previous lemma. Given $f \in C_c^\infty(\mathfrak{g})$ such that $O_X(f) = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}}$, we must show that $\mu_{S+Y_i}(f) = 0$ for all $i$. By Theorem 17.6 there exists an open neighborhood $U$ of $S$ in $\mathfrak{h}'$ such that

$$O_{X'}(f) = \sum_{i=1}^{s} \mu_{S+Y_i}(f) \cdot \Gamma_i^H(X') \tag{27.2.4}$$

for all $X' \in U \cap \mathfrak{h}_{\mathrm{rs}} = U \cap \mathfrak{g}_{\mathrm{rs}}$. Since the function $X' \mapsto O_{X'}(f)$ on $\mathfrak{g}_{\mathrm{rs}}$ is identically zero, and since the restrictions of the functions $\Gamma_i^H$ to the subset $U \cap \mathfrak{h}_{\mathrm{rs}}$ remain linearly independent, we see that $\mu_{S+Y_i}(f) = 0$ for all $i$, as desired. $\square$

**27.3. Linear independence of Shalika germs and density of linear combinations of regular semisimple orbital integrals.** Now we are ready to prove Harish-Chandra's theorem stating that Shalika germs are in fact linearly independent.

THEOREM 27.5 ([**HC78, HC99**]). *The Shalika germs $\Gamma_1, \dots, \Gamma_r$ are linearly independent functions on $\mathfrak{g}_{\mathrm{rs}}$. Indeed they remain linearly independent when restricted to $U \cap \mathfrak{g}_{\mathrm{rs}}$ for any open neighborhood $U$ of $0$ in $\mathfrak{g}$. Moreover every invariant distribution on $\mathfrak{g}$ lies in the closure of the linear span of the subset*

$$\{O_X : X \in \mathfrak{g}_{\mathrm{rs}}\}$$

*of $\mathcal{D}(\mathfrak{g})$.*

PROOF. We reproduce Harish-Chandra's beautiful proof, which uses just about everything we have done. By Lemma 27.2 the second statement of the theorem follows from the first. We prove the first and last statements of the theorem by induction on the dimension of $\mathfrak{g}$, the case when $\dim(G) = 0$ being trivial. Assuming the theorem is true for all connected reductive $H$ with $\dim(H) < \dim(G)$, we must show that it is true for $G$.

We claim that the first statement of the lemma holds for $G$ if and only if the last statement of the theorem holds for $G$. Indeed, if the last statement is true, then the first statement is true by Lemma 27.3. Now assume that the first statement is true. Then for every semisimple element $S$ in $\mathfrak{g}$, the Shalika germs for $G_S$ are linearly independent. By Lemma 27.4, for every $X \in \mathfrak{g}$ the distribution $O_X$ is in the closure of the linear span of the subset

$$\{O_{X'} : X' \in \mathfrak{g}_{\mathrm{rs}}\}$$

of $\mathcal{D}(\mathfrak{g})$. This, together with Proposition 27.1, shows that the last statement of the theorem is true.

Using Lemma 17.4, we reduce to the case in which the center of $\mathfrak{g}$ is trivial. It remains to verify the last statement of the theorem. For this we need to consider the two subspaces $C_2 \subset C_1$ of $C_c^\infty(\mathfrak{g})$ defined by

$$C_1 := \{f \in C_c^\infty(\mathfrak{g}) : O_X(f) = 0 \quad \forall X \in \mathfrak{g}_{\mathrm{rs}}\}$$
$$C_2 := \{f \in C_c^\infty(\mathfrak{g}) : O_X(f) = 0 \quad \forall X \in \mathfrak{g}\}.$$

Using our induction hypothesis and Lemma 27.4, we see that

$$C_1 = \{f \in C_c^\infty(\mathfrak{g}) : O_X(f) = 0 \quad \forall X \in \mathfrak{g} \text{ such that } X \text{ is not nilpotent}\},$$

from which we see that $C_1/C_2$ is finite dimensional and that the dual space $(C_1/C_2)^*$ is spanned by the images of the nilpotent orbital integrals $\mu_1, \dots, \mu_r$.

It follows from Proposition 27.1 that

$$C_2 = \{f \in C_c^\infty(\mathfrak{g}) : I(f) = 0 \text{ for every invariant distribution } I\},$$

from which it is clear that the Fourier transform takes $C_2$ isomorphically onto itself. We also see that in order to prove that the last statement of the theorem is true, we must show that $C_1/C_2 = 0$.

We claim that the Fourier transform also takes $C_1$ isomorphically onto itself. It is enough prove that the Fourier transform $f \mapsto \hat{f}$ carries $C_1$ into itself, since then the same will be true of the inverse Fourier transform $f \mapsto \check{f}$. So let $f \in C_1$. We must show that $O_X(\hat{f}) = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}}$. But $O_X(\hat{f}) = \hat{O}_X(f)$, and we know from Theorem 26.1 that $\hat{O}_X$ is represented by a nice conjugation invariant function on $\mathfrak{g}$. Since all regular semisimple orbital integrals of $f$ vanish, it is then clear that $\hat{O}_X(f)$ vanishes.

Thus the Fourier transform on functions induces an isomorphism

$$C_1/C_2 \xrightarrow{FT} C_1/C_2, \tag{27.3.1}$$

and the Fourier transform on distributions induces an isomorphism

$$(C_1/C_2)^* \xrightarrow{FT} (C_1/C_2)^*. \tag{27.3.2}$$

Thus the Fourier transforms $\hat{\mu}_1, \dots, \hat{\mu}_r$ of the nilpotent orbital integrals also span $(C_1/C_2)^*$.

Now we use homogeneity of nilpotent orbital integrals (see Lemma 17.2):

$$\mu_{\mathcal{O}}(f_{\alpha^2}) = |\alpha|^{-\dim \mathcal{O}} \mu_{\mathcal{O}}(f). \tag{27.3.3}$$

An easy calculation shows that $(f_\beta)\hat{} = |\beta|^{-\dim(G)}(\hat{f})_{\beta^{-1}}$; therefore the Fourier transform $\hat{\mu}_{\mathcal{O}}$ is also homogeneous:

$$\hat{\mu}_{\mathcal{O}}(f_{\alpha^2}) = |\alpha|^{\dim \mathcal{O} - 2\dim(G)}\, \hat{\mu}_{\mathcal{O}}(f). \tag{27.3.4}$$

Let $D$ be the set of integers that arise as the dimension of some nilpotent orbit for $G$. Assuming that $\dim(G) \neq 0$, as we may, then

$$d < \dim G \tag{27.3.5}$$

for all $d \in D$.

Therefore, there is one basis for $(C_1/C_2)^*$ in which each basis element scales by the factor $|\alpha|^{-d}$ for some $d \in D$, and another in which each scales by the factor $|\alpha|^{d-2\dim(G)}$ for some $d \in D$. By (27.3.5) $-d \neq d' - 2\dim G$ for all $d, d' \in D$. Therefore, by linear independence of characters, we have $C_1/C_2 = 0$, and this establishes that the last statement of the theorem holds for $G$. □

Corollary 27.6. *Let $C$ be an open and closed $G$-invariant subset of $\mathfrak{g}$. Then every invariant distribution on $\mathfrak{g}$ supported in $C$ lies in the closure of the linear span of the subset*

$$\{O_X : X \in \mathfrak{g}_{\mathrm{rs}} \cap C\}$$

*of $\mathcal{D}(\mathfrak{g})$.*

Proof. Let $f \in C_c^\infty(\mathfrak{g})$ and suppose that $O_X(f) = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}} \cap C$. We must show that $I(f) = 0$ for every invariant distribution supported on $C$. Put $f_0 := f 1_C$, where $1_C$ denotes the characteristic function of $C$. Clearly $I(f) = I(f_0)$. Moreover it is clear that $O_X(f_0) = 0$ for all $X \in \mathfrak{g}_{\mathrm{rs}}$. By the theorem above $I(f_0) = 0$. □

Now return to the situation in 13.6.

Corollary 27.7. *Parabolic induction $i_P^G$ is independent of the choice of parabolic subgroup $P$ having Levi component $M$.*

Proof. Let $P, P'$ be two parabolic subgroups with Levi component $M$. We want to show that

$$I(f^{(P)}) = I(f^{(P')}) \tag{27.3.6}$$

for every invariant distribution $I$ on $\mathfrak{m}$. By the theorem above (applied to $M$), it is enough to show that

$$O_X^M(f^{(P)}) = O_X^M(f^{(P')}) \tag{27.3.7}$$

for all $X \in \mathfrak{m}_{\mathrm{rs}}$. For $X \in \mathfrak{m} \cap \mathfrak{g}_{\mathrm{rs}}$ this follows from Lemma 13.3, and by continuity (use the local constancy statement in Theorem 17.10) it then follows for all elements $X \in \mathfrak{m}_{\mathrm{rs}}$. □

**27.4. Niceness of Fourier transforms of invariant distributions whose support is bounded modulo conjugation.** So far we know (see Theorem 26.1) that $\hat{O}_X$ is represented by a nice conjugation invariant function on $\mathfrak{g}$ for any $X \in \mathfrak{g}_{\mathrm{rs}}$. In fact the same is true for any $X \in \mathfrak{g}$ (the case of nilpotent $X$ being especially interesting). Our next theorem, also due to Harish-Chandra, says that an even stronger and more general result is true.

Let $\omega$ be a compact open subset of $\mathbb{A}_G(F)$ and put $C := \pi_G^{-1}(\omega)$. Let $J(C)$ be the space of invariant distributions supported in $C$. By Corollary 27.6 the linear span of the set

$$\{O_X : X \in \mathfrak{g}_{\mathrm{rs}} \cap C\} \tag{27.4.1}$$

is dense in $J(C)$.

Theorem 27.8 ([**HC78**]). *Let $I \in J(C)$, and let $L$ be any lattice in $\mathfrak{g}$. Then there exists a linear combination $I'$ of elements in the set* (27.4.1) *such that the distributions $\hat{I}$ and $\hat{I}'$ have the same restriction to $L$. Consequently, for any invariant distribution $I$ on $G$ whose support is bounded modulo conjugation, the Fourier transform $\hat{I}$ is represented by a nice conjugation invariant function on $\mathfrak{g}$.*

Proof. The first statement follows from Proposition 26.1. Now we derive the second statement from the first. By Lemma 15.2 $I$ is contained in $J(C)$ for suitably big $\omega$. Since the distributions $\hat{I}'$ appearing in the first statement of the theorem are nice by Theorem 26.1, we see that the restriction of $\hat{I}$ to any lattice $L$ is nice. This finishes the proof, since the collection of all lattices is an open cover of $\mathfrak{g}$. □

It follows from Theorem 27.8 that the Fourier transform $\hat{O}_X$ of any orbital integral $O_X$ is represented by a nice conjugation invariant function, which we will also denote by $\hat{O}_X$. Context will determine whether we are thinking about $\hat{O}_X$ as a distribution or as a nice function on $\mathfrak{g}$. The same goes for $\hat{\mu}_i$.

**27.5. Uniformity of Shalika germ expansions.** By the Shalika germ expansion, for any $f \in C_c^\infty(\mathfrak{g})$ there is a lattice $L'$ in $\mathfrak{g}$ such that

$$O_X(f) = \sum_{i=1}^{r} \mu_i(f) \cdot \Gamma_i(X) \tag{27.5.1}$$

for all $\in \mathfrak{g}_{\mathrm{rs}} \cap L'$. The lattice $L'$ depends on $f$. Of course, given finitely many functions $f$, we can find a single lattice that works for all of them at once, but there is no guarantee that we can do so for an infinite collection of functions. Nevertheless, we will now see that Howe's finiteness theorem implies that for any lattice $L$ in $\mathfrak{g}$, we can find a lattice $L'$ that works for all the functions in $C_c(\mathfrak{g}/L)$.

Proposition 27.2. *Let $L$ be a lattice in $\mathfrak{g}$. Then there exists a lattice $L'$ in $\mathfrak{g}$ such that*

$$O_X(f) = \sum_{i=1}^{r} \mu_i(f) \cdot \Gamma_i(X) \tag{27.5.2}$$

*for all $f \in C_c(\mathfrak{g}/L)$ and all $X \in \mathfrak{g}_{\mathrm{rs}} \cap L'$.*

PROOF. Pick a compact open neighborhood $\omega$ of 0 in $\mathbb{A}_G(F)$ and put $C := \pi_G^{-1}(\omega)$, an open neighborhood of the nilpotent cone. Note that $O_X \in J(C)$ for all $X \in C$. By Howe's theorem the image of $J(C)$ in $C_c(\mathfrak{g}/L)^*$ is finite dimensional. The subspace $W$ in $C_c(\mathfrak{g}/L)$ consisting of all functions annihilated by all distributions in $J(C)$ therefore has finite codimension, so that we can choose finitely many functions $f_1, \dots, f_m \in C_c(\mathfrak{g}/L)$ that together with $W$ span $C_c(\mathfrak{g}/L)$. For each $f_j$ there is a neighborhood $U_j$ of 0 in $\mathfrak{g}$ such that the Shalika germ expansion for $f_j$ works on $U_j$. The Shalika germ expansion for each $f \in W$ works on the open neighborhood $C$ of 0, since both sides of (27.5.2) vanish for such $f$. Therefore for any lattice $L'$ contained in $C \cap U_1 \cap \dots \cap U_m$ the Shalika germ expansion works on $L'$ for all $f \in C_c(\mathfrak{g}/L)$. □

PROPOSITION 27.3. *Let $L$ be a lattice in $\mathfrak{g}$. Then there exists a lattice $L'$ in $\mathfrak{g}$ such that for all $X \in \mathfrak{g}_{\mathrm{rs}} \cap L'$ and all $Y \in \mathfrak{g}_{\mathrm{rs}} \cap L$ there is an equality*

$$\hat{O}_X(Y) = \sum_{i=1}^{r} \Gamma_i(X) \cdot \hat{\mu}_i(Y). \tag{27.5.3}$$

PROOF. This proposition is the Fourier transform of the previous one. □

COROLLARY 27.9. *Let $L'$ be a lattice in $\mathfrak{g}$. Then there exists a lattice $L$ in $\mathfrak{g}$ such that for all $X \in \mathfrak{g}_{\mathrm{rs}} \cap L'$ and all $Y \in \mathfrak{g}_{\mathrm{rs}} \cap L$ there is an equality*

$$\hat{O}_X(Y) = \sum_{i=1}^{r} \Gamma_i(X) \cdot \hat{\mu}_i(Y). \tag{27.5.4}$$

PROOF. An easy calculation shows that $\hat{O}_X(Y) = \hat{O}_{\beta X}(\beta^{-1}Y)$ for all $\beta \in F^\times$ and all $X, Y \in \mathfrak{g}_{\mathrm{rs}}$. Moreover the right side of (27.5.4) does not change when $(X, Y)$ is replaced by $(\alpha^2 X, \alpha^{-2} Y)$ (for $\alpha \in F^\times$), because of the homogeneity properties of Shalika germs (17.7.1) and Fourier transforms of nilpotent orbital integrals (27.3.4). Therefore the equality (27.5.4) holds for all $X \in \mathfrak{g}_{\mathrm{rs}} \cap L'$ and all $Y \in \mathfrak{g}_{\mathrm{rs}} \cap L$ if and only if it holds for all $X \in \mathfrak{g}_{\mathrm{rs}} \cap \alpha^2 L'$ and all $Y \in \mathfrak{g}_{\mathrm{rs}} \cap \alpha^{-2} L$. From the previous proposition there exists some pair of lattices $L_0$, $L_0'$ on which the equality (27.5.4) holds. Pick $\alpha$ such that $L' \subset \alpha^2 L_0'$. Then the statement of the corollary holds for $L := \alpha^{-2} L_0$. □

**27.6. Linear independence of the restrictions of nilpotent orbital integrals to $C_c(\mathfrak{g}/L)$.** We have already observed that the nilpotent orbital integrals $\mu_1, \dots, \mu_r$ are linearly independent distributions. Now let $L$ be any lattice in $\mathfrak{g}$.

LEMMA 27.10. *The restrictions of $\mu_1, \dots, \mu_r$ to $C_c(\mathfrak{g}/L)$ are linearly independent.*

PROOF. Since $\mu_1, \dots, \mu_r$ are linearly independent, there exists some lattice $L'$ for which the lemma is true. There exists $\alpha \in F^\times$ such that $L \subset \alpha^2 L'$. The distributions $\mu_1', \dots, \mu_r'$ obtained from $\mu_1, \dots, \mu_r$ by scaling by $\alpha^2$ remain linearly independent on $C_c(\mathfrak{g}/\alpha^2 L')$ and hence on the bigger space $C_c(\mathfrak{g}/L)$ as well. But by homogeneity of nilpotent orbital integrals, $\mu_i'$ is a positive multiple of $\mu_i$. This proves the lemma. □

COROLLARY 27.11. *For any lattice $L$ in $\mathfrak{g}$ the restrictions to $L$ of the nice functions $\hat{\mu}_1, \dots, \hat{\mu}_r$ are linearly independent.*

Proof. This statement is the Fourier transform of the statement in the lemma. □

**27.7. Lie algebra analog of the local character expansion.** First we explain the statement of Harish-Chandra's local character expansion. Harish-Chandra proved [**HC78**] that the distribution character of any irreducible admissible representation of $G$ is represented by a locally constant function $\Theta$ on $G_{\mathrm{rs}}$ that is locally integrable on $G$. Use the exponential function to identify a suitable open neighborhood of 0 in $\mathfrak{g}$ with an open neighborhood of 1 in $G$. Then use the exponential function to transport the nice functions $\hat{\mu}_i$ to this neighborhood of 1. Harish-Chandra then proved that there are unique constants $c_i$ such that

$$\Theta(g) = \sum_{i=1}^{r} c_i \hat{\mu}_i(g) \tag{27.7.1}$$

for all regular semisimple $g$ in some suitably small neighborhood of 1 in $G$. How small the neighborhood has to be depends on the representation of $G$ that we started with.

His proof uses the Lie algebra analog of this statement. We have already made the point that Fourier transforms of orbital integrals are the Lie algebra analogs of irreducible characters on $G$. Therefore we would expect Fourier transforms of orbital integrals to appear on the left side of the Lie algebra analog of (27.7.1). Actually a more general statement is true: the Fourier transform of any invariant distribution whose support is bounded modulo conjugation has a local character expansion. In the case of Fourier transforms of regular semisimple orbital integrals, one even knows what the constants $c_i$ are: they are Shalika germs. Here is the precise statement of Harish-Chandra's Lie algebra analog of the local character expansion.

Theorem 27.12 ([**HC78**]). *Let $\omega$ be any compact open subset of $\mathbb{A}_G(F)$ and let $C := \pi_G^{-1}(\omega)$, a closed and open $G$-invariant subset of $\mathfrak{g}$ that is bounded modulo conjugation. There exists a lattice $L$ in $\mathfrak{g}$ such that the following two statements hold.*

(1) *For all $X \in C \cap \mathfrak{g}_{\mathrm{rs}}$ there is an equality*

$$\hat{O}_X = \sum_{i=1}^{r} \Gamma_i(X) \hat{\mu}_i$$

*of functions on $L \cap \mathfrak{g}_{\mathrm{rs}}$.*

(2) *For all $I \in J(C)$ there exist unique complex numbers $c_1, \dots, c_r$ such that*

$$\hat{I} = \sum_{i=1}^{r} c_i \hat{\mu}_i$$

*on $L \cap \mathfrak{g}_{\mathrm{rs}}$.*

Proof. By Lemma 15.2 there exists a lattice $L'$ such that $C \subset \mathrm{Ad}(G)(L')$. Thus the first statement follows from Corollary 27.9. In view of Theorem 27.8 the second statement follows from the first. □

## 28. Guide to notation

- See 4.1 for $F$, $\mathcal{O}$, $\pi$, $\mathrm{val}(x)$, $G$, $B = AN$, $W$, $K$, $G_{\mathrm{der}}$, $G_{\mathrm{sc}}$, $X_*(A)$.
- See 4.5 for $\Lambda_G$, $H_G$, $\mathfrak{a}$, $\mathfrak{a}_G$.
- See 7.8 for $\mathcal{L} = \mathcal{L}(A)$.
- See 7.12 for $\mathcal{F}(A)$, $\mathcal{P}(M)$.
- See 8.3 for $\mathcal{B}(A)$, $B_0 = AN_0$. See 14.2 for $\pi_G : \mathfrak{g} \to \mathbb{A}_G$.
- See 23.2 for $\mathcal{F}(M)$.

## References


[Art76] J. Arthur, *The characters of discrete series as orbital integrals*, Invent. Math. **32** (1976), 205–261.

[Art78] J. Arthur, *A trace formula for reductive groups. I. Terms associated to classes in $G(\mathbb{Q})$*, Duke Math. J. **45** (1978), 911–952.

[Art81] J. Arthur, *The trace formula in invariant form*, Ann. of Math. (2) **114** (1981), 1–74.

[Art87] J. Arthur, *The characters of supercuspidal representations as weighted orbital integrals*, Proc. Indian Acad. Sci. Math. Sci. **97** (1987), 3–19.

[Art89a] J. Arthur, *The $L^2$-Lefschetz numbers of Hecke operators*, Invent. Math. **97** (1989), 257–290.

[Art89b] J. Arthur, *Towards a local trace formula*, Algebraic analysis, geometry, and number theory (Baltimore, MD, 1988), Johns Hopkins Univ. Press, Baltimore, MD, 1989, pp. 1–23.

[Art91a] J. Arthur, *A local trace formula*, Inst. Hautes Études Sci. Publ. Math. **73** (1991), 5–96.

[Art91b] J. Arthur, *Some problems in local harmonic analysis*, Harmonic analysis on reductive groups (Brunswick, ME, 1989), Progr. Math., vol. 101, Birkhäuser Boston, Boston, MA, 1991, pp. 57–78.

[Art97] J. Arthur, *The problem of classifying automorphic representations of classical groups*, Advances in mathematical sciences: CRM's 25 years (Montreal, PQ, 1994), CRM Proc. Lecture Notes, vol. 11, Amer. Math. Soc., Providence, RI, 1997, pp. 1–12.

[BFQ79] P. Baum, W. Fulton, and G. Quart, *Lefschetz-Riemann-Roch for singular varieties*, Acta Math. **143** (1979), no. 3-4, 193–211.

[Bor91] A. Borel, *Linear algebraic groups*, Graduate Texts in Mathematics, vol. 126, Springer-Verlag, New York, 1991.

[Bou02] N. Bourbaki, *Lie groups and Lie algebras. Chapters 4–6*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 2002.

[Bri88] M. Brion, *Points entiers dans les polyèdres convexes*, Ann. Sci. École Norm. Sup. (4) **21** (1988), no. 4, 653–663.

[BT72] F. Bruhat and J. Tits, *Groupes réductifs sur un corps local. I*, Inst. Hautes Études Sci. Publ. Math. **41** (1972), 5–251.

[BZ76] I. N. Bernšteĭn and A. V. Zelevinskiĭ, *Representations of the group $GL(n, F)$, where $F$ is a local non-Archimedean field*, Uspehi Mat. Nauk **31** (1976), no. 3(189), 5–70.

[Car00] H. Carayol, *Preuve de la conjecture de Langlands locale pour* $\mathrm{GL}_n$*: travaux de Harris-Taylor et Henniart*, Astérisque (2000), no. 266, Exp. No. 857, 4, 191–243.

[Clo89] L. Clozel, *Orbital integrals on p-adic groups: a proof of the Howe conjecture*, Ann. of Math. (2) **129** (1989), no. 2, 237–251.

[Den84] J. Denef, *The rationality of the Poincaré series associated to the p-adic points on a variety*, Invent. Math. **77** (1984), no. 1, 1–23.

[Ful93] W. Fulton, *Introduction to toric varieties*, Ann. of Math. Studies, No. 131, Princeton University Press, 1993.

[Ful98] W. Fulton, *Intersection theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 2, Springer-Verlag, Berlin, 1998.

[HC70] Harish-Chandra, *Harmonic analysis on reductive p-adic groups. Notes by G. van Dijk*, Lecture Notes in Mathematics, No. 162, Springer-Verlag, 1970.

[HC78] Harish-Chandra, *Admissible invariant distributions on reductive p-adic groups*, Proceedings of the 1977 annual seminar of the Canadian Mathematical Congress (W. Rossmann, ed.), Queen's Papers in Pure and Applied Mathematics, no. 48, 1978, pp. 281–347.

[HC99] Harish-Chandra, *Admissible invariant distributions on reductive p-adic groups. Notes by S. DeBacker and P.J. Sally, Jr.*, University Lecture Series, vol. 16, AMS, 1999.

[How73] R. Howe, *Two conjectures about reductive p-adic groups*, Harmonic analysis on homogeneous spaces (Proc. Sympos. Pure Math., Vol. XXVI, Williams Coll., Williamstown, Mass., 1972), Amer. Math. Soc., Providence, R.I., 1973, pp. 377–380.

[Igu74] J.-I. Igusa, *Complex powers and asymptotic expansions. I. Functions of certain types*, J. Reine Angew. Math. **268/269** (1974), 110–130.

[Igu77] J.-I. Igusa, *Some observations on higher degree characters*, Amer. J. Math. **99** (1977), no. 2, 393–417.

[Kna86] A. Knapp, *Representation theory of semisimple groups*, Princeton University Press, Princeton, N.J., 1986.

[Lan66] R. P. Langlands, *Eisenstein series*, Algebraic Groups and Discontinuous Subgroups (Proc. Sympos. Pure Math., Boulder, Colo., 1965), Amer. Math. Soc., Providence, R.I., 1966, pp. 235–252.

[Lan76] R. P. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Mathematics, No. 544, Springer-Verlag, 1976.

[Lan80] R. P. Langlands, *Base change for $GL(2)$*, Ann. of Math. Studies 96, Princeton University Press, 1980.

[Lan02] S. Lang, *Algebra*, Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002.

[Nie74] H. Nielsen, *Diagonalizably linearized coherent sheaves*, Bull. Soc. Math. France **102** (1974), 85–97.

[Pan91] D. Panyushev, *Rationality of singularities and the Gorenstein property of nilpotent orbits*, Funct. Anal. Appl. **25** (1991), 76–78.

[Rao72] R. Rao, *Orbital integrals in reductive groups*, Ann. of Math. (2) **96** (1972), 505–510.

[Rog90] J. Rogawski, *Automorphic representations of unitary groups in three variables*, Annals of Mathematics Studies, vol. 123, Princeton University Press, Princeton, NJ, 1990.

[Ros61] M. Rosenlicht, *On quotient varieties and the affine embedding of certain homogeneous spaces*, Trans. Amer. Math. Soc. **101** (1961), 211–223.

[Ser92] J-P. Serre, *Lie algebras and Lie groups*, Lecture Notes in Mathematics, vol. 1500, Springer-Verlag, Berlin, 1992.

[Ser02] J-P. Serre, *Galois cohomology*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2002.

[Ser03] J-P. Serre, *Trees*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2003.

[Sha72] J. Shalika, *A theorem on semi-simple $\mathcal{P}$-adic groups*, Ann. of Math. (2) **95** (1972), 226–242.

[SS70] T. Springer and R. Steinberg, *Conjugacy classes*, Seminar in Algebraic Groups and Related Finite Groups, Lecture Notes in Math., 131, Springer Verlag, 1970, pp. 167–266.

[Ste75] R. Steinberg, *Torsion in reductive groups*, Advances in Math. **15** (1975), 63–92.

[Wal95] J.-L. Waldspurger, *Une formule des traces locale pour les algèbres de Lie p-adiques*, J. Reine Angew. Math. **465** (1995), 41–99.

[Wei82] A. Weil, *Adeles and algebraic groups*, Progress in Mathematics, vol. 23, Birkhäuser Boston, Mass., 1982.

Department of Mathematics, University of Chicago, 5734 University Ave., Chicago, IL 60637

*E-mail address*: `kottwitz@math.uchicago.edu`

Clay Mathematics Proceedings
Volume 4, 2005

# Homogeneity for Reductive $p$-adic Groups: An Introduction

Stephen DeBacker

Abstract. We discuss, in a fairly conversational manner, homogeneity results for reductive $p$-adic groups. We provide some motivation for why we expect such results to be true, and we discuss why they are important. We also discuss most of the mathematics required to prove homogeneity statements.

## 1. Introduction

The goal of these notes is to introduce the idea of homogeneity for reductive $p$-adic groups. Except in trivial cases, we are not in any position to verify homogeneity statements; rather, we shall try to motivate both why such results are important and why we should believe that they are true. To this end, we will also discuss many of the important mathematical ideas surrounding these statements. Finally, while I think that they are mathematically accurate, these notes are intended as an introduction, not as a reference.

I thank Joseph Rabinoff for producing the computer graphics for Figure 8. I learned nearly all that I know about harmonic analysis while under the excellent guidance of Bob Kottwitz and Paul Sally, Jr.. Although they are not directly referenced here, my understanding of Bruhat-Tits theory has been deeply influenced by the beautiful papers of Allen Moy and Gopal Prasad. Finally, I thank Jeff Adler for his excellent proofreading of these notes.

## 2. An introduction to homogeneity

We begin with some motivations for considering homogeneity questions and try to illustrate why their answers look the way that they do.

**2.1. The case $\mathrm{GL}_1$.** We begin with the completely trivial yet illuminating case of $G = k^\times = \mathrm{GL}_1(k)$ where $k$ is a $p$-adic field.

We first consider homogeneity statements on $k^\times$ and then turn our attention to its Lie algebra $k$. Let $C_c^\infty(k^\times)$ denote the space of compactly supported, locally constant functions on $k^\times$ (similar notation applies to $k$).

2000 *Mathematics Subject Classification.* Primary 22E50.

*Key words and phrases.* Bruhat-Tits theory, harmonic analysis, reductive $p$-adic group, stability.

Supported by the National Science Foundation Grant No. 0200542.

Suppose $\chi \in \widehat{k^\times}$, that is, $\chi$ is a complex-valued continuous multiplicative character of $k^\times$. We may define a distribution $\Theta_\chi \colon C_c^\infty(k^\times) \to \mathbb{C}$ by setting

$$\Theta_\chi(f) = \int_{k^\times} \chi(x) \cdot f(x)\, dx$$

for $f \in C_c^\infty(k^\times)$. Here $dx$ denotes a (fixed) Haar measure on $k^\times$.

Let $R$ denote the ring of integers of $k$ and let $\wp$ denote the prime ideal. Fix a uniformizer $\varpi$ (that is, $\wp = \varpi \cdot R$). To avoid complications, we suppose $\chi$ has depth $(m-1)$ with $m > 1$, that is, the restriction of $\chi$ to the filtration subgroup $1 + \wp^m$ is trivial and the restriction of $\chi$ to the filtration subgroup $1 + \wp^{(m-1)}$ is nontrivial.

Note that if the support of $f$ is contained in $1 + \wp^m$, then

$$\Theta_\chi(f) = \Theta_1(f)$$

where $\Theta_1$ denotes the distribution associated to the trivial character on $k^\times$. Therefore, we may write

$$\operatorname{res}_{C_c^\infty(1+\wp^m)} \Theta_\chi = \operatorname{res}_{C_c^\infty(1+\wp^m)} \Theta_1.$$

This is a homogeneity[1] statement: the distributions $\Theta_\chi$ and $\Theta_1$ agree on $C_c^\infty(1 + \wp^m)$.

We now focus on the Lie algebra $k$ of $k^\times$. We let $\mu_0$ denote the distribution on $C_c^\infty(k)$ which sends $f$ to $f(0)$. Suppose $T$ is a distribution on $k$, that is, a linear map from $C_c^\infty(k)$ to $\mathbb{C}$. Suppose $m$ is an integer such that $T$ belongs to $J(\wp^m)$, the space of distributions on $k$ having support in $\wp^m$. If $f$ belongs to $C_c(k/\wp^m)$, the space of compactly supported functions on $k$ which are translation invariant with respect to the lattice[2] $\wp^m$, then we can write

$$f = \sum_{\bar{X} \in k/\wp^m} f(X) \cdot [X + \wp^m]$$

where $[X + \wp^m]$ denotes the characteristic function of the coset $X + \wp^m$. For such a function we have

$$\begin{aligned} T(f) &= T\big(\sum_{\bar{X} \in k/\wp^m} f(X) \cdot [X + \wp^m]\big) = f(0) \cdot T([\wp^m]) \\ &= T([\wp^m]) \cdot \mu_0(f). \end{aligned}$$

That is, we have the homogeneity statement

$$\operatorname{res}_{C_c(k/\wp^m)} J(\wp^m) = \operatorname{res}_{C_c(k/\wp^m)} \mathbb{C} \cdot \mu_0.$$

Since $\mathrm{GL}_1(k)$ is abelian, we have not yet said anything nontrivial. The main idea you should keep in mind is that, by restricting to a subspace of a larger function space, we'd like to be able to express fairly arbitrary distributions in terms of well-understood distributions:

STATEMENT 2.1.1.

$$\operatorname{res}_{\substack{\textit{Function}\\\textit{space}}} \left\{ \begin{array}{c} \textit{Fairly arbitrary} \\ \textit{distributions} \end{array} \right\} = \operatorname{res}_{\substack{\textit{Function}\\\textit{space}}} \left\{ \begin{array}{c} \textit{Well-understood} \\ \textit{distributions} \end{array} \right\}.$$

[1] According to the Oxford English Dictionary [**11**], the word *homogeneity* means "identity of kind with something else," and according to Webster's Dictionary [**15**] it means "the state of having identical distribution functions or values."

[2] A compact, open $R$-submodule of a $p$-adic vector space is called a *lattice*.

Moreover, we'd like this statement to be optimal in some sense. For example, the following exercise shows that the homogeneity statements we made above are optimal.

EXERCISE 2.1.2. Suppose $\ell \leq m < n$. Show that

$$\operatorname{res}_{C_c(k/\wp^\ell)} J(\wp^m) = \operatorname{res}_{C_c(k/\wp^\ell)} \mathbb{C} \cdot \mu_0$$

and

$$\operatorname{res}_{C_c(k/\wp^n)} J(\wp^m) \neq \operatorname{res}_{C_c(k/\wp^n)} \mathbb{C} \cdot \mu_0.$$

Formulate and prove a similar statement for distributions on $k^\times$.

**2.2. Some history and an application.** If we do not wish to make an optimal homogeneity statement, then the type of results we seek have been known for a long time — we shall call these "prehomogeneity" results. However, it has become clear that a great many of the interesting problems in representation theory and harmonic analysis require more precision than these prehomogeneity results provide.

Let $G$ be a reductive $p$-adic group and let $\mathfrak{g}$ be the Lie algebra of $G$. So, for example, we could take $G$ to be $\mathrm{SL}_n(k)$ or $\mathrm{Sp}_{2n}(k)$ and then $\mathfrak{g}$ would be $\mathfrak{sl}_n(k)$ or $\mathfrak{sp}_{2n}(k)$. If $S \subset \mathfrak{g}$, then we set

$${}^G S := \{{}^g s := \mathrm{Ad}(g)s \mid g \in G \text{ and } s \in S\}.$$

The first result we discuss is a conjecture of Howe which was proved by Howe [**8**] for the general linear group and by Harish-Chandra [**7**] in a general context.

THEOREM 2.2.1 (Howe's conjecture for the Lie algebra). *If $L$ is a lattice in $\mathfrak{g}$ and $\omega \subset \mathfrak{g}$ is compact, then*

$$\dim_{\mathbb{C}}\big(\operatorname{res}_{C_c(\mathfrak{g}/L)} J(\omega)\big) < \infty.$$

In the statement of Howe's conjecture, the notation $J(\omega)$ denotes the space of invariant distributions[3] supported on the closure of the set ${}^G\omega$. So, for example, if $X \in \omega$, then the orbital integral $\mu_X$ belongs to $J(\omega)$. (Since $\mathrm{GL}_1(k)$ is abelian, this agrees with our earlier use of the notation $J$.) Note that since Howe's conjecture is not equating two sets of distributions, it is not really a homogeneity result — or even a prehomogeneity statement. However, for fixed $\omega$ and expanding $L$, the dimension of the left-hand side will stabilize. Thus, for sufficiently large $L$, we might expect to find a basis for the left hand side consisting of well-understood distributions on $\mathfrak{g}$ (see §2.3). In this section, we use the above result to prove a useful harmonic analysis result (which will later be improved using homogeneity results).

Suppose that $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}$. Let $\mathfrak{h}' = \mathfrak{h} \cap \mathfrak{g}^{\text{r.s.s}}$. (Here $\mathfrak{g}^{\text{r.s.s}}$ denotes the set of regular semisimple elements in $\mathfrak{g}$, that is, those elements of $\mathfrak{g}$ whose centralizer in $G$ is a torus.) We consider the map $\mathfrak{h}' \times C_c^\infty(\mathfrak{g}) \to \mathbb{C}$ defined by

$$(H, f) \mapsto \widehat{\mu}_H(f) := \mu_H(\hat{f}). \tag{*}$$

[3]A distribution $T$ is said to be *invariant* provided that $T(f^g) = T(f)$ for all $g \in G$ and $f \in C_c^\infty(\mathfrak{g})$. Here $f^g(X) = f({}^g X)$.

Here, we realize the Fourier transform as a map from $C_c^\infty(\mathfrak{g})$ to itself by setting

$$\hat{f}(X) = \int_{\mathfrak{g}} f(Y) \cdot \Lambda(B(Y,X))\, dY$$

where $dY$ is a Haar measure on $\mathfrak{g}$, $B$ is a nondegenerate, symmetric, invariant, bilinear form on $\mathfrak{g}$, and $\Lambda$ is a continuous additive character of $k$ that is trivial on the lattice $\wp$ and nontrivial on the lattice $R$.

There are two ways to think about the map defined by Equation (*):

(1) If we fix $H$ and vary $f$, then we are looking at a distribution on $\mathfrak{g}$. It is a result of Harish-Chandra that this distribution is represented by a locally integrable function on $\mathfrak{g}$ which we also call $\widehat{\mu}_H$. This means that for all $f \in C_c^\infty(\mathfrak{g})$ we have

$$\widehat{\mu}_H(f) = \int_{\mathfrak{g}} f(Y) \cdot \widehat{\mu}_H(Y)\, dY.$$

(2) If we fix $f$ and vary $H$, then we are looking at a locally constant function on $\mathfrak{h}'$.

We can combine these two ways of thinking about the map defined in Equation (*) by formulating a statement about the local constancy of the function $\widehat{\mu}_H$. Namely,

THEOREM 2.2.2 ([**7**]). *For all $H \in \mathfrak{h}'$ and for all compact open $\omega \subset \mathfrak{g}$, there exists a compact open $\omega_H \subset \mathfrak{h}'$ such that*

(1) *$H \in \omega_H$ and*
(2) *$\widehat{\mu}_{H'}(Y) = \widehat{\mu}_H(Y)$ for all $H' \in \omega_H$ and all $Y \in \omega$.*

To illustrate the usefulness of Howe's conjecture, we present here Harish-Chandra's proof of this result. In the proof, Howe's conjecture reduces a seemingly intractable problem to a simple linear algebra problem.

PROOF. Fix $H \in \mathfrak{h}'$ and $\omega \subset \mathfrak{g}$ compact and open. We begin by reformulating statement (2) of the theorem:

$$\widehat{\mu}_{H'}(Y) = \widehat{\mu}_H(Y) \text{ for all } H' \in \omega_H \text{ and all } Y \in \omega.$$

This statement is equivalent to the statement

$$\widehat{\mu}_{H'}(f) = \widehat{\mu}_H(f) \text{ for all } H' \in \omega_H \text{ and all } f \in C_c^\infty(\omega),$$

which, in turn, is equivalent to the statement

$$\mu_{H'}(\hat{f}) = \mu_H(\hat{f}) \text{ for all } H' \in \omega_H \text{ and all } f \in C_c^\infty(\omega).$$

By choosing a lattice $L$ in $\mathfrak{g}$ so that $f \in C_c^\infty(\omega)$ implies that $\hat{f} \in C_c(\mathfrak{g}/L)$, we see that this last formulation of the statement would be true if we knew that

$$\mu_{H'}(\varphi) = \mu_H(\varphi) \text{ for all } H' \in \omega_H \text{ and all } \varphi \in C_c(\mathfrak{g}/L).$$

We will establish this last statement (which, in itself, is a type of prehomogeneity statement).

Let $\omega'_H$ be any compact open neighborhood of $H$ in $\mathfrak{h}'$. Note that $\mu_{H'}$ belongs to $J(\omega'_H)$ for all $H' \in \omega'_H$. From Howe's conjecture for the Lie algebra, we have

$$\dim_{\mathbb{C}} \left(\operatorname{res}_{C_c(\mathfrak{g}/L)} J(\omega'_H)\right) < \infty.$$

Hence, we can choose $H_1, H_2, \ldots, H_m \in \omega'_H$ such that for every $H' \in \omega'_H$ the distribution $\mathrm{res}_{C_c(\mathfrak{g}/L)}\, \mu_{H'}$ belongs to the span of the linearly independent distributions

$$\mathrm{res}_{C_c(\mathfrak{g}/L)}\, \mu_{H_i}.$$

Fix $f_1,\, f_2,\, \ldots\, , f_m \in C_c(\mathfrak{g}/L)$ such that

$$\mu_{H_i}(f_j) = \delta_{ij}.$$

So, for all $H' \in \omega'_H$ we have

$$\mu_{H'}(f) = \sum_i \mu_{H'}(f_i) \cdot \mu_{H_i}(f)$$

for all $f \in C_c(\mathfrak{g}/L)$.

Fix a neighborhood $\omega_H$ of $H$ for which

(1) $\omega_H \subset \omega'_H$ and
(2) $\mu_{H'}(f_i) = \mu_H(f_i)$ for all $1 \leq i \leq m$ and for all $H' \in \omega_H$.

We then have that

$$\begin{aligned} \mu_{H'}(f) &= \sum_i \mu_{H'}(f_i) \cdot \mu_{H_i}(f) \\ &= \sum_i \mu_H(f_i) \cdot \mu_{H_i}(f) \\ &= \mu_H(f) \end{aligned}$$

for all $f \in C_c(\mathfrak{g}/L)$ and all $H' \in \omega_H$. □

**2.3. The nilpotent cone in** $\mathrm{SL}_2(\mathbb{R})$**.** In this section we look at the $\mathrm{SL}_2(\mathbb{R})$-orbits in $\mathfrak{sl}_2(\mathbb{R})$. We do this for two reasons: First, it gives us a way to visualize[4] the problems we are discussing. Second, it will help to clear up many of the common misunderstandings the reader may harbor about how things work over non-algebraically closed fields.

As vector spaces, we have $\mathbb{R}^3 \cong \mathfrak{sl}_2(\mathbb{R})$ via the map

$$(x, y, z) \mapsto \mathrm{M}(x, y, z) := \begin{pmatrix} x & y+z \\ y-z & -x \end{pmatrix}.$$

The characteristic polynomial of $\mathrm{M}(x, y, z)$ is

$$t^2 - \big((x^2 + y^2) - z^2\big),$$

and so we have three distinct types of elements depending on the eigenvalues of $\mathrm{M}(x, y, z)$ (see Table 1).

| Type of element | $(x^2 + y^2) - z^2$ |
|---|---|
| nilpotent | $0$ |
| split | $> 0$ |
| elliptic | $< 0$ |

Table 1. Types of elements in $\mathfrak{sl}_2(\mathbb{R})$

2.3.1. *Nilpotent elements.* In this case, we have $z^2 = x^2 + y^2$, and so $\mathcal{N}$, the nilpotent elements, is a cone in $\mathbb{R}^3$ (see Figure 1).

[4]It is hard to draw pictures of $p$-adic vector spaces; to paraphrase Paul Sally, Jr.: "We all have our own picture of the $p$-adics, but we dare not discuss it with others."

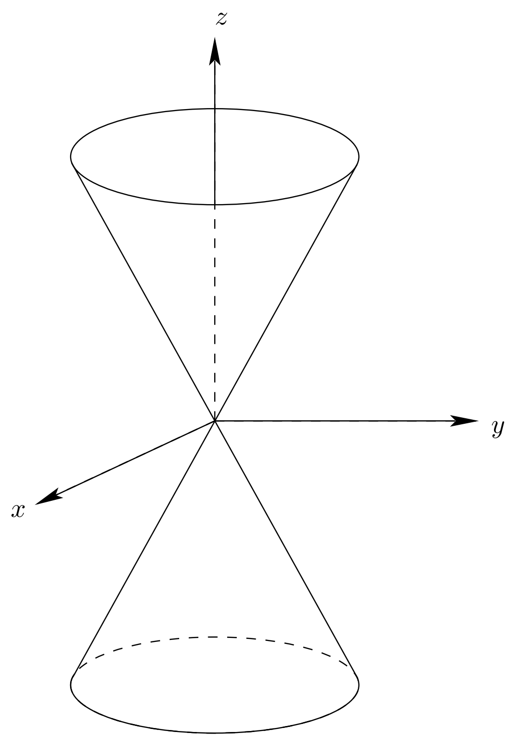



FIGURE 1. The nilpotent cone for $\mathrm{SL}_2(\mathbb{R})$

We let $\mathcal{O}(0)$ denote the set of nilpotent orbits. To decompose $\mathcal{N}$ into orbits, we notice that the unit circle $S^1$ embeds into $\mathrm{SL}_2(\mathbb{R})$ under the map

$$\theta \mapsto s(\theta) := \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

and

$$^{s(\theta)}M(x, y, z) = M(x \cdot \cos(2\theta) + y \cdot \sin(2\theta), y \cdot \cos(2\theta) - x \cdot \sin(2\theta), z).$$

Consequently, the set of nilpotent elements in $\mathfrak{sl}_2(\mathbb{R})$ having a fixed $z$ value are all conjugate. From the Jacobson-Morozov theorem [**3**, §5.3], for all $X \in \mathcal{N}$ we can produce a one-parameter subgroup

$$\lambda\colon \mathrm{GL}_1 \to \mathrm{SL}_2$$

such that

$$^{\lambda(t)}X = t^2 X$$

for all $t \in \mathbb{R}^\times$.

EXERCISE 2.3.1. Prove the above assertion.

Combining the action of $S^1$ with the above consequence of Jacobson-Morozov, we conclude that $\mathcal{O}(0)$ has at most three elements. In fact, there are exactly three nilpotent orbits.

REMARK 2.3.2. It is important to note that, except for the trivial orbit, it is not true that there is a single $g \in \mathrm{SL}_2(\mathbb{R})$ that acts by dilation on every element of a nilpotent orbit. More precisely, we know that if $\mathcal{O} \in \mathcal{O}(0)$, then

(1) $t^2\mathcal{O} = \mathcal{O}$ for all $t \in \mathbb{R}^\times$ and
(2) for each $X \in \mathcal{O}$, there is a $g_X \in \mathrm{SL}_2(\mathbb{R})$ such that ${}^{g_X}X = t^2X$.

Consequently, if $\mu_{\mathcal{O}}$ denotes an invariant measure (it is unique up to a constant) on $\mathcal{O}$ and $f$ is a nice function on $\mathcal{O}$, then

(1) for all $t \in \mathbb{R}^\times$
$$\mu_{\mathcal{O}}(f_{t^2}) = |t|^{-\dim(\mathcal{O})}\,\mu_{\mathcal{O}}(f)$$
where $f_{t^2}(Y) = f(t^2Y)$ for $Y \in \mathfrak{sl}_2(\mathbb{R})$ and
(2) for all $g \in \mathrm{SL}_2(\mathbb{R})$,
$$\mu_{\mathcal{O}}(f^g) = \mu_{\mathcal{O}}(f).$$

2.3.2. *Split and elliptic elements.* We now consider the two remaining cases. In both cases, the characteristic polynomial has distinct eigenvalues: real in the split case and complex in the elliptic case. Fix $\alpha > 0$.

We first consider the split case. The set of $\mathrm{M}(x,y,z)$ for which $\alpha^2 = z^2 - (x^2 + y^2)$ form a single orbit all of whose elements are conjugate to
$$\mathrm{M}(\alpha,0,0) = \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix}.$$
The orbit is a one sheeted hyperboloid which is asymptotic to (and outside of) the nilpotent cone.

For the elliptic case the elements $M(x,y,z)$ for which $-\alpha^2 = z^2 - (x^2 + y^2)$ form two orbits all of whose elements are conjugate to either
$$\mathrm{M}(0,0,\alpha) = \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix}$$
or
$$\mathrm{M}(0,-\alpha,0) = \begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix}.$$
Note that these two matrices are conjugate by an element of $\mathrm{SL}_2(\mathbb{C})$. These orbits form a two sheeted hyperboloid which is asymptotic to (and inside of) the nilpotent cone.

To complete our discussion of split and elliptic elements, we recall that a Cartan subalgebra (CSA) is a maximal subalgebra consisting of commuting semisimple elements. (If you prefer, you may think of a CSA as the Lie algebra of a maximal $\mathbb{R}$-torus of $\mathrm{SL}_2$.) For $\mathfrak{sl}_2(\mathbb{R})$, the CSAs are one-dimensional, given by lines through the origin of the form
$$\{M(\lambda a, \lambda b, \lambda c) \mid \lambda \in \mathbb{R}\}$$
with $a^2 + b^2 \neq c^2$. We therefore recover the "standard" split CSA
$$\{M(\lambda,0,0) \mid \lambda \in \mathbb{R}\} = \{\begin{pmatrix} x & 0 \\ 0 & -x \end{pmatrix} \mid x \in \mathbb{R}\}$$
and the "standard" elliptic CSA
$$\{M(0,0,\lambda) \mid \lambda \in \mathbb{R}\} = \{\begin{pmatrix} 0 & z \\ -z & 0 \end{pmatrix} \mid z \in \mathbb{R}\}.$$

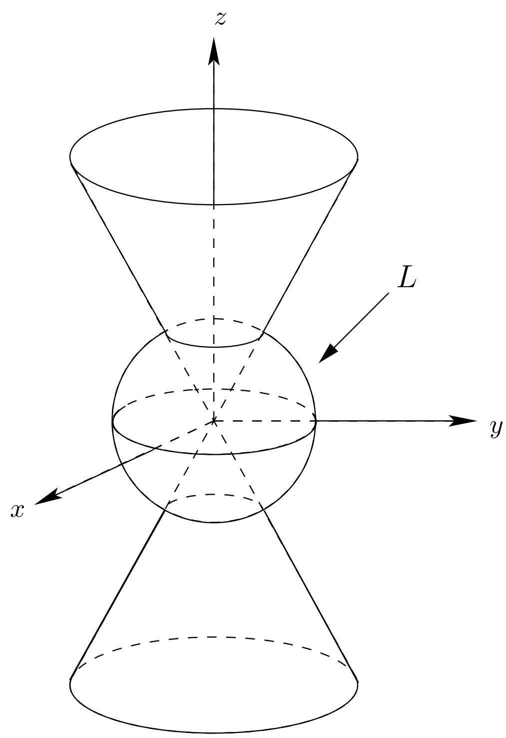



FIGURE 2. A "picture" of the lattice $L$

2.3.3. *A return to homogeneity.* We again consider Statement 2.1.1. From the preceding discussion, it is clear (at least for $\mathfrak{sl}_2(\mathbb{R})$) that every orbit is asymptotic to the nilpotent cone. Thus, it is believable that the right-hand side of Statement 2.1.1 should, ideally, consist of nilpotent orbital integrals.

If we pretend that we can draw pictures of what the nilpotent cone looks like $p$-adically, then we can even visualize Statement 2.1.1 . For simplicity, let us assume that we we are interested in invariant distributions supported on the closure of ${}^{\mathrm{SL}_2(k)}L$ for the lattice $L$ "drawn" in Figure 2

From our discussion above, we know that the closure of ${}^{\mathrm{SL}_2(k)}L$ is asymptotic to the nilpotent cone, and we "see" that, in fact,

$$^{\mathrm{SL}_2(k)}L \subset \mathcal{N} + L.$$

(Compare this with Lemma 5.1.1.) Consequently, it is not much of a stretch to think that our homogeneity statements should look like

$$\mathrm{res}_{C_c(\mathfrak{g}/L)}\, J(L) = \mathrm{res}_{C_c(\mathfrak{g}/L)}\, J(\mathcal{N})$$

where $J(\mathcal{N})$ denotes the space of invariant distributions spanned by the nilpotent orbital integrals.

## 3. An introduction to some aspects of Bruhat-Tits theory

We now have a guess as to what belongs on the right-hand side of Statement 2.1.1. The purpose of this section is to introduce, via examples, enough

Bruhat-Tits theory to help us refine our understanding of what to place on the left-hand side.

A good introduction to Bruhat-Tits theory may be found in Joe Rabinoff's Harvard senior thesis [**12**].

**3.1. Apartments.** Our immediate goal is to understand a bit of the mathematics behind the "Coxeter paper" that Bill Casselman has posted on his web page[5].

Recall that $G$ is a $p$-adic group, that is, $G$ is the group of $k$-rational points of a connected reductive linear algebraic $k$-group $\mathbf{G}$. For simplicity, we shall assume that $\mathbf{G}$ is a semisimple, $k$-split group which is defined over $\mathbb{Z}$. Thus, the notations $\mathbf{G}(R)$ and $\mathfrak{g}(R)$ make sense. So, for example, $\mathbf{G}$ could be $\mathrm{Sp}_{2n}$, realized in the usual way.

Following earlier lecturers, we fix a maximal $k$-split torus $\mathbf{A}$ in $\mathbf{G}$ which is defined over $\mathbb{Z}$. We let $A$ denote the group of $k$-rational points of $\mathbf{A}$. So, for example, $A$ could be the set of diagonal matrices in $\mathrm{Sp}_{2n}(k)$.

We let $\mathcal{A} = \mathbf{X}_*(\mathbf{A}) \otimes \mathbb{R}$ and call $\mathcal{A}$ the *apartment*[6] attached to $A$. For the group $\mathrm{Sp}_{2n}(k)$, the apartment is isomorphic to $\mathbb{R}^n$.

An apartment carries a natural polysimplicial decomposition; we now describe how this arises. We let $\Phi = \Phi(\mathbf{G}, \mathbf{A})$ denote the set of nontrivial eigencharacters for the action of $A$ on $\mathfrak{g}$. We assume that the valuation map $\nu\colon k^\times \to \mathbb{Z}$ is surjective, and we let $\Psi = \Psi(\mathbf{G}, \mathbf{A}, \nu)$ denote the corresponding set of affine roots, that is

$$\Psi = \{\gamma + n \,|\, \gamma \in \Phi\ ,\ n \in \mathbb{Z}\}.$$

Each $\psi = \gamma + n \in \Psi$ defines an affine function on $\mathcal{A}$ by

$$(\gamma + n)(\lambda \otimes r) := r \cdot \langle \lambda, \gamma \rangle + n$$

where $\langle\ ,\ \rangle$ denotes the natural perfect pairing $\mathbf{X}_*(\mathbf{A}) \times \mathbf{X}^*(\mathbf{A}) \to \mathbb{Z}$. (Here, $\mathbf{X}^*(\mathbf{A})$ denotes the group of characters of $\mathbf{A}$.) Consequently, for each $\psi \in \Psi$, we can define the hyperplane $H_\psi := \{x \in \mathcal{A} \,|\, \psi(x) = 0\} \subset \mathcal{A}$. These hyperplanes give us the familiar polysimplicial decomposition of $\mathcal{A}$. We usually call a polysimplex occurring in this decomposition a *facet* and the maximal facets are called *alcoves*.

Finally, just as the Weyl group $W = N_G(A)/A$ acts transitively on (spherical) chambers, the extended affine Weyl group $\tilde{W} = N_G(A)/\mathbf{A}(R)$ acts transitively on alcoves (but not, in general, simply transitively — think about the image of $\left(\begin{smallmatrix} 0 & 1 \\ \varpi & 0 \end{smallmatrix}\right)$ in $\mathrm{PGL}_2(k)$ and how it acts on the standard apartment of $\mathrm{PGL}_2(k)$).

3.1.1. $\mathrm{Sp}_4(k)$ *in detail.* For this subsection only, we let $G = \mathrm{Sp}_4(k)$ realized as the subgroup of the group of $4 \times 4$ matrices of nonzero determinant which preserve

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}.$$

We take $A$ to be the set of matrices $\{a(x,y) \,|\, x, y \in k^\times\}$ where

$$a(x,y) := \begin{pmatrix} x & 0 & 0 & 0 \\ 0 & y & 0 & 0 \\ 0 & 0 & y^{-1} & 0 \\ 0 & 0 & 0 & x^{-1} \end{pmatrix}.$$

[5]Look under Frivolities at `http://www.math.ubc.ca/people/faculty/cass/`

[6]Generally speaking, one does not want to fix (as we have) an origin.

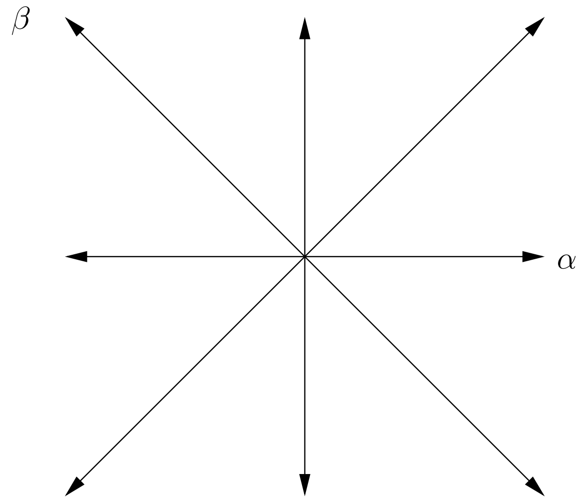



FIGURE 3. The $C_2$ root system

If we define $\alpha, \beta \in \mathbf{X}^*(\mathbf{A})$ by $\alpha(a(x,y)) = xy^{-1}$ and $\beta(a(x,y)) = y^2$, then

$$\Phi = \{\pm\alpha, \pm\beta, \pm(\beta+\alpha), \pm(\beta+2\alpha)\}$$

and the root system has the familiar diagram given in Figure 3.

The $\mathbb{Z}$-lattice of cocharacters $\mathbf{X}_*(\mathbf{A})$ is the $\mathbb{Z}$-linear span of $\lambda_1$ and $\lambda_2$ where $\lambda_1(t) = a(t,1)$ and $\lambda_2(t) = a(1,t)$ for $t \in k^\times$. In Figure 4 we have begun a sketch of the simplicial decomposition of $\mathcal{A}$ arising from the above data. The reader is encouraged to spend some time thinking about how we arrived at Figure 4.

REMARK 3.1.1. For those familiar with coroots, we note that $\check{\alpha} = \lambda_1 - \lambda_2$ while $\check{\beta} = \lambda_2$.

**3.2. Objects associated to facets.** To each facet in $\mathcal{A}$ we can attach many types of objects. Some of these live in $G$, others in $\mathfrak{g}$, and still others are properly thought of as objects over $\mathfrak{f} := R/\wp$, the residue field of $k$. In this section, we introduce these items.

For each $\gamma \in \Phi$ we have a root group, denoted $U_\gamma$, in $G$ and a root space, denoted $\mathfrak{g}_\gamma$, in $\mathfrak{g}$. In each case, these groups are isomorphic to $k$.

EXAMPLE 3.2.1. In the example of $\mathrm{Sp}_4(k)$ introduced above, we have that $U_\alpha$ consists of matrices of the form

$$\begin{pmatrix} 1 & a & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -a \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and $\mathfrak{g}_\alpha$ consists of $4 \times 4$ matrices of the form

$$\begin{pmatrix} 0 & a & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -a \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The field $k$ carries a natural filtration, indexed by $\mathbb{Z}$, consisting of compact open subgroups:

$$k \supset \cdots \supset \wp^{-2} \supset \wp^{-1} \supset R \supset \wp \supset \wp^2 \supset \cdots \supset \{0\}.$$

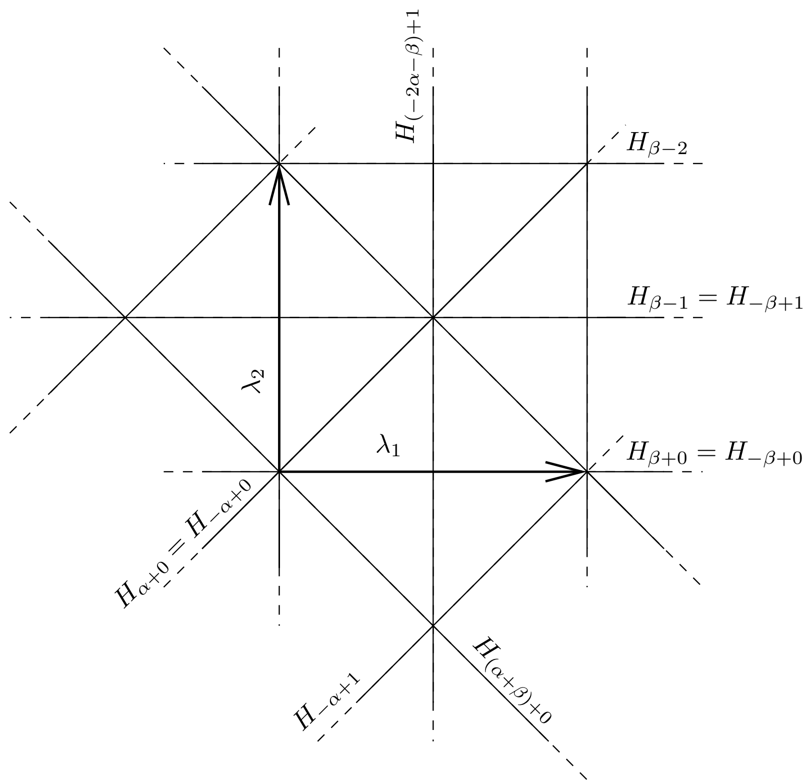



FIGURE 4. A sketch of an apartment for $\mathrm{Sp}_4(k)$

We'd like to use the set $\{\gamma+n \mid n \in \mathbb{Z}\}$ to index the corresponding natural filtration in $U_\gamma$ (resp. $\mathfrak{g}_\gamma$). To fix this indexing, we make the following choices:

$$U_{\gamma+1} \subsetneq U_{\gamma+0} := \mathbf{G}(R) \cap U_\gamma$$

and

$$\mathfrak{g}_{\gamma+1} \subsetneq \mathfrak{g}_{\gamma+0} := \mathfrak{g}(R) \cap \mathfrak{g}_\gamma.$$

We can now define some of the objects we are interested in. For $x \in \mathcal{A}$, we define $G_x$, the parahoric subgroup attached to $x$, by

$$G_x := \langle \mathbf{A}(R), U_\psi \rangle_{\psi \in \Psi; \psi(x) \geq 0}.$$

That is, $G_x$ is the group generated by $\mathbf{A}(R)$ and the subgroups $U_\psi$ for $\psi \in \Psi$ with $\psi(x) \geq 0$. Since a facet $F$ in $\mathcal{A}$ is determined by the intersection of hyperplanes, we have $G_x = G_y$ for $x, y \in F$. Consequently, the notation $G_F$ makes sense. If $o$ is the origin in $\mathcal{A}$, then $G_o = \mathbf{G}(R)$.

EXAMPLE 3.2.2. We consider the case of $\mathrm{SL}_2(k)$ with $A$ realized as the set of diagonal matrices. In Figure 5 we have sketched and labeled part of the corresponding apartment. After fixing an orientation, the parahoric subgroups associated to each facet are given[7] in the second column of Table 2.

[7]For example, the notation $\begin{pmatrix} R & \wp \\ \wp^{-1} & R \end{pmatrix}$ means the group of matrices in $\mathrm{SL}_2(k)$ having entries in the indicated rings.

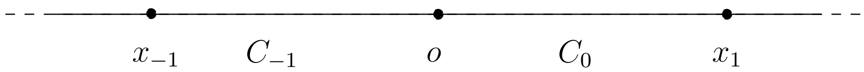



FIGURE 5. A sketch of an apartment for $\mathrm{SL}_2(k)$

| $F$ | $G_F$ | $G_F^+$ | $G_F/G_F^+$ |
|---|---|---|---|
| $x_{-1}$ | $\begin{pmatrix} R & \wp \\ \wp^{-1} & R \end{pmatrix}$ | $1 + \begin{pmatrix} \wp & \wp^2 \\ R & \wp \end{pmatrix}$ | $\mathrm{SL}_2(\mathfrak{f})$ |
| $C_{-1}$ | $\begin{pmatrix} R & \wp \\ R & R \end{pmatrix}$ | $1 + \begin{pmatrix} \wp & \wp \\ R & \wp \end{pmatrix}$ | $\mathrm{GL}_1(\mathfrak{f})$ |
| $o$ | $\mathrm{SL}_2(R)$ | $1 + \begin{pmatrix} \wp & \wp \\ \wp & \wp \end{pmatrix}$ | $\mathrm{SL}_2(\mathfrak{f})$ |
| $C_0$ | $\begin{pmatrix} R & R \\ \wp & R \end{pmatrix}$ | $1 + \begin{pmatrix} \wp & R \\ \wp & \wp \end{pmatrix}$ | $\mathrm{GL}_1(\mathfrak{f})$ |
| $x_1$ | $\begin{pmatrix} R & \wp^{-1} \\ \wp & R \end{pmatrix}$ | $1 + \begin{pmatrix} \wp & R \\ \wp^2 & \wp \end{pmatrix}$ | $\mathrm{SL}_2(\mathfrak{f})$ |

TABLE 2. Various groups associated to facets in $\mathrm{SL}_2(k)$

The parahoric $G_F$ always has a normal subgroup $G_F^+$, called the pro-unipotent radical, with the property that the quotient $G_F/G_F^+$ is the group of $\mathfrak{f}$-rational points of a connected reductive $\mathfrak{f}$-group $\mathsf{G}_F$. To define $G_F^+$, we must first consider the torus $A$. We set

$$\mathbf{A}(R)^+ := \{a \in \mathbf{A}(R) \mid \nu(\chi(a) - 1) > 0 \text{ for all } \chi \in \mathbf{X}^*(\mathbf{A})\}.$$

EXAMPLE 3.2.3. In $\mathrm{SL}_2(k)$, $\mathbf{A}(R)^+$ consists of the matrices

$$\begin{pmatrix} 1+\wp & 0 \\ 0 & 1+\wp \end{pmatrix}.$$

and in $\mathrm{Sp}_4(k)$, we have $\mathbf{A}(R)^+ := \{a(x,y) \mid x, y \in 1 + \wp\}$.

For $x \in \mathcal{A}$ we define $G_x^+$ by

$$G_x^+ := \langle \mathbf{A}(R)^+, U_\psi \rangle_{\psi \in \Psi; \psi(x) > 0}.$$

As before, for a facet $F$ in $\mathcal{A}$, the notation $G_F^+$ makes sense. The various subgroups associated to each facet in $\mathcal{A}$ for $\mathrm{SL}_2(k)$ are given in Table 2.

It is a general fact, which is clearly exhibited in the example of $\mathrm{SL}_2(k)$, that if $F_1$ and $F_2$ are two facets for which $F_1$ belongs to the closure of $F_2$, then

$$G_{F_1}^+ < G_{F_2}^+ < G_{F_2} < G_{F_1}$$

and $G_{F_2}/G_{F_1}^+$ is a parabolic subgroup of $\mathsf{G}_{F_1}(\mathfrak{f}) = G_{F_1}/G_{F_1}^+$ with unipotent radical isomorphic to $G_{F_2}^+/G_{F_1}^+$ and Levi factor isomorphic to $\mathsf{G}_{F_2}(\mathfrak{f})$. In particular, if $F_2$ is an alcove, then $G_{F_2}/G_{F_1}^+$ may be identified with a Borel subgroup of $\mathsf{G}_{F_1}(\mathfrak{f})$.

We end this section with a few examples.

EXAMPLE 3.2.4. In Figure 6 we label each of the facets in a fixed alcove of an apartment for $\mathrm{SL}_3(k)$ with the name of the corresponding $\mathfrak{f}$-group.

EXAMPLE 3.2.5. In Figure 7 we label each of the facets in a fixed alcove of an apartment for $\mathrm{Sp}_4(k)$ with the name of the corresponding $\mathfrak{f}$-group.

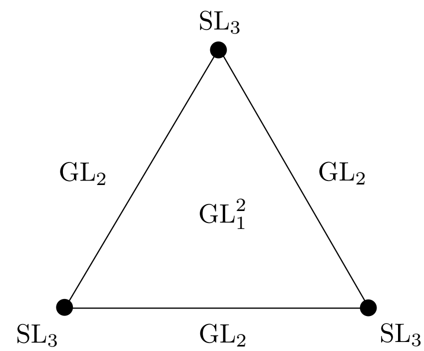



FIGURE 6. An alcove for $\mathrm{SL}_3(k)$

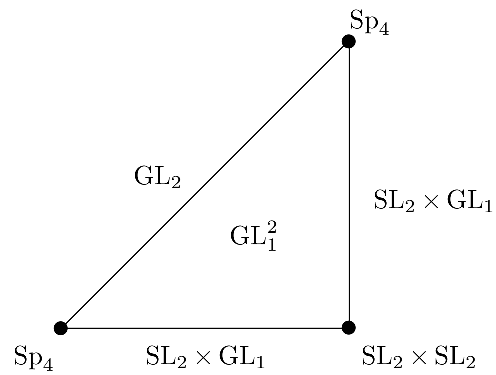



FIGURE 7. An alcove for $\mathrm{Sp}_4(k)$

EXAMPLE 3.2.6. In Figure 8 there is a model, produced by Joseph Rabinoff, for an alcove of $\mathrm{Sp}_6(k)$. Each of the facets has been labeled with the name of the corresponding $\mathfrak{f}$-group. This model can be quite instructive. For example, after assembling the model, one sees that it can be realized as that part of a cube cut out by placing vertices at a vertex of the cube, the midpoint of an adjacent edge, the center of an adjacent face, and the center of the cube. The cube decomposes into forty-eight such solids, and the Weyl group of $\mathrm{Sp}_6(k)$ acts simply transitively on them (take the origin of $\mathcal{A}$ as the center of the cube).

All of the above can be carried out for the Lie algebra. In particular, for a facet $F$ there is a lattice $\mathfrak{g}_F^+$ so that $\mathfrak{g}_F/\mathfrak{g}_F^+$ is $\mathsf{L}_F(\mathfrak{f}) := \mathrm{Lie}(\mathsf{G}_F)(\mathfrak{f})$, the Lie algebra of $\mathsf{G}_F(\mathfrak{f})$.

## 4. Parameterizations via Bruhat-Tits theory: nilpotent orbits

The main idea of this section is to relate certain aspects of the structure theory of $G$ to the structure theory of the various finite groups of Lie type that arise naturally via Bruhat-Tits theory. We shall treat the structure theory of finite groups of Lie type as a black box. These results will play a key role in our understanding and use of the homogeneity statements to come.

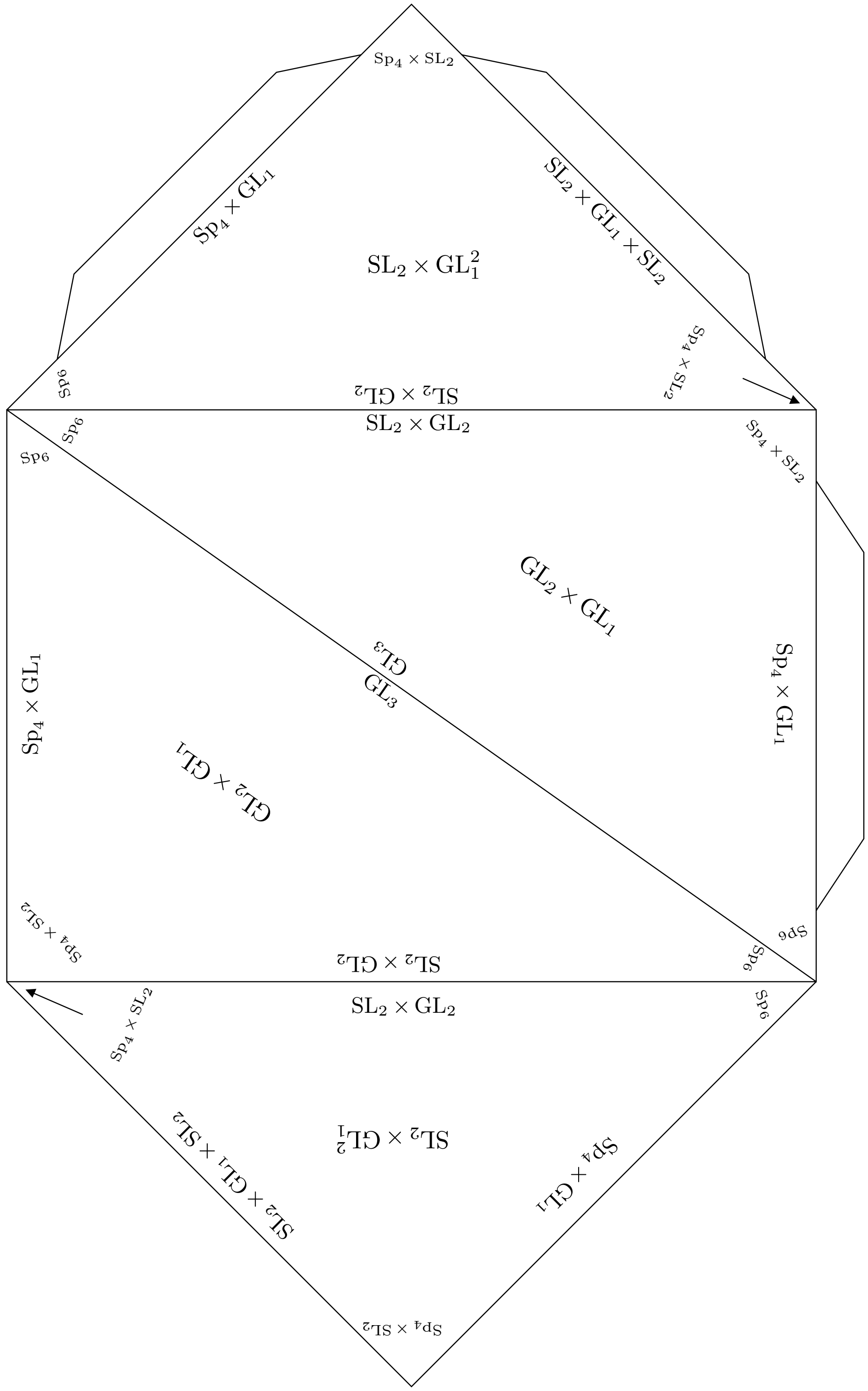



Figure 8. An alcove for $\mathrm{Sp}_6(k)$

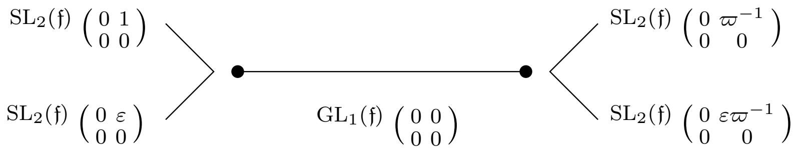



FIGURE 9. Distinguished nilpotent orbits associated to facets for $\mathrm{SL}_2(k)$

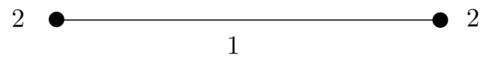



FIGURE 10. Enumeration of distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits for $\mathrm{SL}_2(k)$

**4.1. A parameterization of nilpotent orbits: examples.** None of the material in this section works unless $p$, the residual characteristic of $k$, is sufficiently large (as a function of the root datum of $\mathbf{G}$). We begin with an example.

EXAMPLE 4.1.1. When $p \neq 2$ the group $\mathrm{SL}_2(k)$ has five nilpotent orbits. These are represented by the elements of the set

$$\left\{ \begin{pmatrix} 0 & \theta \\ 0 & 0 \end{pmatrix} \mid \theta \in \{0, 1, \varepsilon, \varpi^{-1}, \varpi^{-1}\varepsilon\} \right\}$$

where $\varepsilon \in R^\times \setminus (R^\times)^2$. On the other hand, we have that the group $\mathrm{SL}_2(\mathfrak{f})$ has two distinguished[8] orbits

$$^{\mathrm{SL}_2(\mathfrak{f})}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad ^{\mathrm{SL}_2(\mathfrak{f})}\begin{pmatrix} 0 & \varepsilon \\ 0 & 0 \end{pmatrix}$$

where $\varepsilon \in \mathfrak{f}^\times \setminus (\mathfrak{f}^\times)^2$, and $\mathrm{GL}_1(\mathfrak{f})$ has one distinguished orbit — the trivial orbit. When we encode this information in our preferred chamber, we produce a picture like Figure 9.

Note that in the diagram we've included the factor $\varpi^{-1}$ to emphasize the obvious $p$-adic lift. For consistency with later examples, in Figure 10 we enumerate the distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits attached to each facet in an alcove for $\mathrm{SL}_2(k)$. Note that there are five orbits enumerated in Figure 10.

The example of $\mathrm{SL}_2(k)$ indicates that there is a simple connection between $\mathcal{O}(0)$, the set of nilpotent orbits for a $p$-adic group, and the nilpotent orbits for Lie groups of finite type. We have the following result due to D. Barbasch and A. Moy [**2**].

FACT 4.1.2. If $F$ is a facet and $\bar{\mathcal{O}} \subset \mathsf{L}_F(\mathfrak{f}) = \mathfrak{g}_F/\mathfrak{g}_F^+$ is a nilpotent orbit, then there exists a unique nilpotent orbit in $\mathfrak{g}$ of minimal dimension which intersects the preimage of $\bar{\mathcal{O}}$ nontrivially.

REMARK 4.1.3. The reader is urged to verify this fact for the group $\mathrm{SL}_2(k)$.

EXAMPLE 4.1.4. Since the heuristics of $\mathrm{SL}_2(k)$ worked so well, let us now turn our attention to $\mathrm{SL}_3(k)$ with $p > 3$. It is easy to see that $\mathcal{O}(0)$ has $2 + 3 \cdot \left|\mathfrak{f}^\times/(\mathfrak{f}^\times)^3\right|$ elements. On the other hand, in Figure 11 we have enumerated the number of distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits in $\mathsf{L}_F(\mathfrak{f})$ for each facet in an alcove of $\mathrm{SL}_3(k)$. When we proceed without thinking (that is, we sum), we find that our indexing set has $4 + 3 \cdot \left|\mathfrak{f}^\times/(\mathfrak{f}^\times)^3\right|$ elements — two too many! However, whenever two line segments

[8]A nilpotent orbit which does not intersect a proper Levi subalgebra is called *distinguished*.

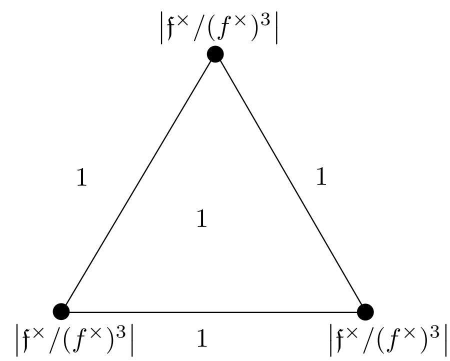



FIGURE 11. Enumeration of distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits for $\mathrm{SL}_3(k)$

in the closure of an alcove are incident (see Figure 6), the associated general linear groups are conjugate in $\mathrm{SL}_3(\mathfrak{f})$. That is, in some real sense we are summing two too many things.

**4.2. An equivalence relation on $\mathcal{A}$.** We now introduce an equivalence relation on the set of facets of $\mathcal{A}$ that will account for the over counting encountered in the $\mathrm{SL}_3(k)$ example above.

DEFINITION 4.2.1. If $F$ is a facet in $\mathcal{A}$, then we let $A(F)$ denote the smallest affine subspace of $\mathcal{A}$ containing $F$.

EXAMPLE 4.2.2. If $F$ is a vertex, then $A(F)$ is the vertex itself. At the opposite extreme, if $F$ is an alcove, then $A(F)$ is $\mathcal{A}$.

Recall that $\tilde{W} = N_G(A)/\mathbf{A}(R)$ acts transitively on the set of alcoves in $\mathcal{A}$

DEFINITION 4.2.3. Suppose $F_1$ and $F_2$ are two facets in $\mathcal{A}$. If there is a $w \in \tilde{W}$ such that

$$A(F_1) = A(wF_2),$$

then we write $F_1 \sim F_2$.

One easily verifies that the rule $\sim$ defines an equivalence relation on the set of facets in $\mathcal{A}$. Moreover, since $\tilde{W}$ acts transitively on alcoves, a set of representatives for the equivalence classes under $\sim$ can always be found among the facets occurring in the closure of a fixed alcove.

EXAMPLE 4.2.4. Here are some examples that the reader is encouraged to verify.

- Two vertices are equivalent if and only if they belong to the same $\tilde{W}$-orbit.
- If $C_1$ and $C_2$ are two alcoves in $\mathcal{A}$, then $C_1 \sim C_2$.
- For $\mathrm{SL}_2(k)$ and $\mathrm{Sp}_4(k)$, the set of facets occurring in the closure of a fixed alcove forms a complete set of representatives for the relation $\sim$.
- The only equivalent facets occurring in the closure of an alcove for $\mathrm{Sp}_6(k)$ are the two faces for which $\mathsf{G}_F$ is $\mathrm{GL}_2 \times \mathrm{GL}_1$.
- The only equivalent facets occurring in the closure of an alcove for $\mathrm{SL}_3(k)$ are the three edges. That is, the facets with hatch marks in Figure 12.

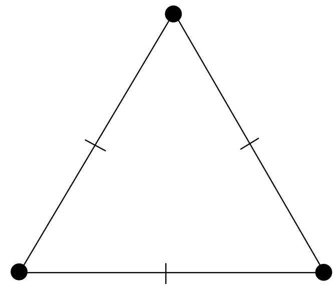


FIGURE 12. Equivalent edges in an alcove for $\mathrm{SL}_3(k)$

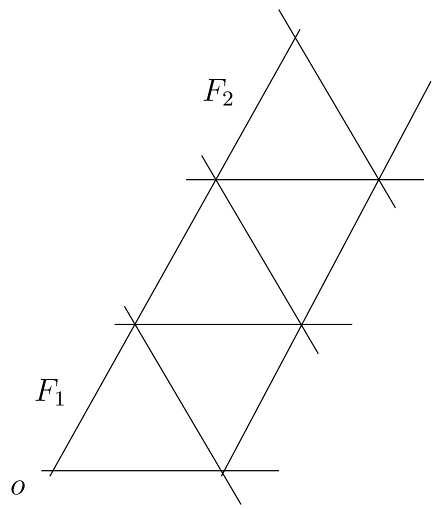



FIGURE 13. Part of the apartment for $\mathrm{SL}_3(k)$

**4.3. The key idea.** We now present the key ingredient that makes everything work. If $F_1$ and $F_2$ are two facets in $\mathcal{A}$ such that $A(F_1) = A(F_2)$, then the natural map

$$G_{F_1} \cap G_{F_2} \to \mathsf{G}_{F_i}(\mathfrak{f})$$

is surjective with kernel $G^+_{F_1} \cap G^+_{F_2}$. In fact, this leads to an $\mathfrak{f}$-isomorphism between $\mathsf{G}_{F_1}$ and $\mathsf{G}_{F_2}$ which we write as $\mathsf{G}_{F_1} \stackrel{i}{=} \mathsf{G}_{F_2}$ (or, for the Lie algebra, as $\mathsf{L}_{F_1} \stackrel{i}{=} \mathsf{L}_{F_2}$).

If you recall how the facets were created, then the above observation becomes less surprising. We now present an example to reinforce the idea.

EXAMPLE 4.3.1. Consider the facets $F_1$ and $F_2$ in the standard apartment for $\mathrm{SL}_3(k)$ as in Figure 13. In Table 3, we list the parahoric subgroup, its pro-unipotent radical, and the $\mathfrak{f}$-group associated to each of these facets. The reader may verify that this example works as advertised.

**4.4. A parameterization of nilpotent orbits: the general case.** We now present some definitions which allow us to extend the examples presented in §4.1.

DEFINITION 4.4.1. Let $I^d$ denote the set of pairs $(F, \bar{\mathcal{O}})$ where $F$ is a facet in $\mathcal{A}$ and $\bar{\mathcal{O}}$ is a distinguished $\mathsf{G}_F(\mathfrak{f})$-orbit in $\mathsf{L}_F(\mathfrak{f})$.

| $F$ | $G_F$ | $G_F^+$ | $\mathsf{G}_F$ |
|---|---|---|---|
| $F_1$ | $\begin{pmatrix} R & R & R \\ R & R & R \\ \wp & \wp & R \end{pmatrix}$ | $1+\begin{pmatrix} \wp & \wp & R \\ \wp & \wp & R \\ \wp & \wp & R \end{pmatrix}$ | $\mathrm{GL}_2$ |
| $F_2$ | $\begin{pmatrix} R & R & \wp^{-2} \\ R & R & \wp^{-2} \\ \wp^3 & \wp^3 & R \end{pmatrix}$ | $1+\begin{pmatrix} \wp & \wp & \wp^{-2} \\ \wp & \wp & \wp^{-2} \\ \wp^3 & \wp^3 & \wp \end{pmatrix}$ | $\mathrm{GL}_2$ |

TABLE 3. Various groups associated to some facets for $\mathrm{SL}_3(k)$

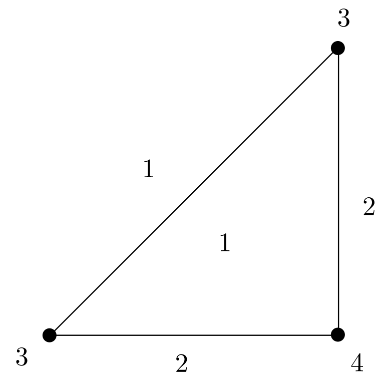



FIGURE 14. An enumeration of the distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits for $\mathrm{Sp}_4(k)$.

DEFINITION 4.4.2. Suppose $(F_1, \bar{\mathcal{O}}_1)$ and $(F_2, \bar{\mathcal{O}}_2)$ are two elements of $I^d$. We write $(F_1, \bar{\mathcal{O}}_1) \sim (F_2, \bar{\mathcal{O}}_2)$ provided that there exists $n \in N_G(A)$ such that

(1) $A(F_1) = A(nF_2)$ and

(2) $\bar{\mathcal{O}}_1 \stackrel{i}{=} {}^n\bar{\mathcal{O}}_2$ in $\mathsf{L}_{F_1}(\mathfrak{f}) \stackrel{i}{=} \mathsf{L}_{nF_2}(\mathfrak{f})$.

We can now state the main result for this section.

THEOREM 4.4.3 ([**6**]). *Suppose $p$ is sufficiently large. The map that sends $(F, \bar{\mathcal{O}}) \in I^d$ to the unique nilpotent $G$-orbit of minimal dimension which intersects the preimage of $\bar{\mathcal{O}}$ nontrivially induces a bijective correspondence*

$$I^d/\!\sim \;\longleftrightarrow\; \mathcal{O}(0).$$

We remark that the theorem is false if $p$ is not large enough. Consider, for example, $\mathrm{SL}_2(\mathbb{Q}_2)$.

We finish our discussion with some examples.

EXAMPLE 4.4.4. It is known that for $\mathrm{Sp}_4(k)$ and $p \neq 2$ the cardinality of $\mathcal{O}(0)$ is sixteen. We have already discussed the fact that none of the facets in the closure of a fixed alcove for $\mathrm{Sp}_4(k)$ are equivalent under $\sim$. In Figure 14 we enumerate the number of distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits in $\mathsf{L}_F(\mathfrak{f})$ for each facet $F$ in the closure of an alcove of $\mathrm{Sp}_4(k)$. As a warning to those who might wish to think further about these matters, we note that the three distinguished orbits found at each of the $\mathrm{Sp}_4$ vertices arise in a somewhat surprising way: Over the algebraic closure, there is one regular nilpotent orbit and one subregular nilpotent orbit (which intersects the

Lie algebra of the $\mathrm{GL}_2$-Levi of $\mathrm{Sp}_4$). Upon descent to the field $\mathfrak{f}$, the regular orbit breaks into two distinguished $\mathrm{Sp}_4(\mathfrak{f})$-orbits and the subregular orbit breaks into two $\mathrm{Sp}_4(\mathfrak{f})$-orbits. One of these orbits intersects the $\mathfrak{f}$-rational points of the Lie algebra of the $\mathrm{GL}_2$-Levi; the other is distinguished.

EXAMPLE 4.4.5. It is known that for $\mathrm{Sp}_6(k)$ and $p \neq 2$ the cardinality of $\mathcal{O}(0)$ is forty-five. We have already discussed the fact that exactly two of the facets in the closure of a fixed alcove for $\mathrm{Sp}_6(k)$ are equivalent under $\sim$. In Table 4 we enumerate

| $\mathsf{G}$ | number of distinguished $\mathsf{G}(\mathfrak{f})$ -orbits |
|---|---|
| $\mathrm{Sp}_6$ | six |
| $\mathrm{Sp}_4 \times \mathrm{SL}_2$ | six |
| $\mathrm{Sp}_4 \times \mathrm{GL}_1$ | three |
| $\mathrm{SL}_2 \times \mathrm{GL}_1^2$ | two |
| $\mathrm{SL}_2 \times \mathrm{GL}_1 \times \mathrm{SL}_2$ | four |
| $\mathrm{SL}_2 \times \mathrm{GL}_2$ | two |
| $\mathrm{GL}_2 \times \mathrm{GL}_1$ | one |
| $\mathrm{GL}_3$ | one |
| $\mathrm{GL}_1^3$ | one |

TABLE 4. An enumeration of distinguished $\mathsf{G}(\mathfrak{f})$-orbits

the number of distinguished $\mathsf{G}_F(\mathfrak{f})$-orbits in $\mathsf{L}_F(\mathfrak{f})$ for each facet $F$ in the closure of an alcove of $\mathrm{Sp}_6(k)$. The subsequent counting exercise is left to the reader.

Finally, we note that there does not exist a complete description of the distinguished orbits in the Lie algebra of a finite group of Lie type. But, although it seems that we have reduced one problem about which we know very little to another problem about which we also know very little, this reduction will be quite useful.

## 5. A precise homogeneity statement

Recall that our goal is to make Statement 2.1.1 into something reasonable and provable. In §2.3.3 we discussed the fact that the "$G$-orbit" of every compact set was asymptotic to the nilpotent cone. This motivated the idea that perhaps $J(\mathcal{N})$, the span of the nilpotent orbital integrals, was a reasonable candidate for the right-hand side of Statement 2.1.1. We are still searching for a candidate for the left-hand side; we begin with a very precise asymptotic result.

### 5.1. An asymptotic result.

LEMMA 5.1.1 ([**1**]). *For facets $F_1, F_2$ in $\mathcal{A}$ we have $\mathfrak{g}_{F_1} \subset \mathfrak{g}_{F_2} + \mathcal{N}$.*

EXAMPLE 5.1.2. In Figure 15 we have described the lattices $\mathfrak{g}_F$ for the standard apartment in $\mathrm{SL}_2(k)$. We observe that if $F_2$ lies to the left of $F_1$, then $\mathfrak{g}_{F_1} \subset \mathfrak{g}_{F_2} + \mathfrak{u}$ where $\mathfrak{u}$ is the set of strictly upper triangular two-by-two matrices.

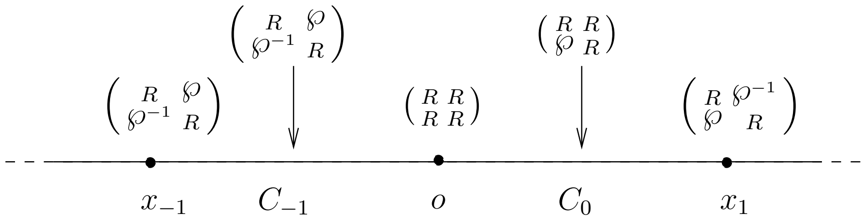



FIGURE 15. Some lattices in $\mathfrak{sl}_2(k)$

PROOF. Choose $x \in F_1$ and $y \in F_2$. Let $\vec{v} = y - x$. Let $\Phi^+$ denote the set of roots that pair nonnegatively against $\vec{v}$ and let $\Phi^- = \Phi \setminus \Phi^+$. We have

$$\sum_{\alpha \in \Phi^-} \mathfrak{g}_\alpha \subset \mathcal{N}$$

and

$$\begin{aligned}
\mathfrak{g}_{F_1} &= \mathrm{Lie}(\mathbf{A})(R) \oplus \sum_{\alpha \in \Phi;\, n \in \mathbb{Z};\, (\alpha+n)(x) > 0} \mathfrak{g}_{\alpha+n} \\
&= \mathrm{Lie}(\mathbf{A})(R) \oplus \sum_{\alpha \in \Phi^+;\, n \in \mathbb{Z};\, (\alpha+n)(x) > 0} \mathfrak{g}_{\alpha+n} \oplus \sum_{\alpha \in \Phi^-;\, n \in \mathbb{Z};\, (\alpha+n)(x) > 0} \mathfrak{g}_{\alpha+n} \\
&\subset \mathfrak{g}_{F_2} + \sum_{\alpha \in \Phi^-;\, n \in \mathbb{Z};\, (\alpha+n)(x) > 0} \mathfrak{g}_{\alpha+n} \\
&\subset \mathfrak{g}_{F_2} + \mathcal{N}.
\end{aligned}$$

The second to last line is true because if $\alpha \in \Phi^+$, then

$$\begin{aligned}
(\alpha+n)(y) = (\alpha+n)(x+\vec{v}) &= (\alpha+n)(x) + \langle \vec{v}, \alpha \rangle \\
&\geq (\alpha+n)(x).
\end{aligned}$$

□

To facilitate our discussion, we fix an alcove $C$ in $\mathcal{A}$.

DEFINITION 5.1.3. We set

$$\mathfrak{g}_0 := \bigcup_{F \subset \bar{C}} {}^G(\mathfrak{g}_F)$$

where the union is over the facets occurring in the closure of a fixed alcove $C$.

The set $\mathfrak{g}_0$ is usually referred to as the set of compact elements in $\mathfrak{g}$; for $\mathrm{GL}_n(k)$ it is exactly the set of elements in $\mathrm{M}_n(k)$ for which each eigenvalue has nonnegative valuation.

COROLLARY 5.1.4. *We have* $\mathfrak{g}_0 \subset \mathfrak{g}_C + \mathcal{N}$.

PROOF. From Bruhat-Tits theory we can write

$$G = G_C \tilde{W} G_C.$$

The result follows. □

**5.2. A homogeneity statement.** The above asymptotic results, along with our previous discussions should, I hope, make the following homogeneity statement both natural and plausible.

THEOREM 5.2.1 ([**14**], [**4**]). *Suppose $p$ is sufficiently large.*

(1)
$$\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathfrak{g}_0) = \operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathcal{N}).$$

(2) *For $T \in J(\mathfrak{g}_0)$ we have*
$$\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} T = 0$$
*if and only if*
$$\operatorname{res}_{\sum_{F\subset \bar{C}} C(\mathfrak{g}_F/\mathfrak{g}_C)} T = 0.$$

The first proof of this result, for "unramified classical" groups, is due to Waldspurger [**14**]. We shall not attempt to prove this theorem, which is a special case of a much more general result. However, we do have enough tools on hand to sketch how statement (2) implies statement (1): We have
$$\operatorname{res}_{\sum_{F\subset \bar{C}} C(\mathfrak{g}_F/\mathfrak{g}_C)} T = 0$$
if and only if
$$\operatorname{res}_{\sum_{F\subset \bar{C}} C(\mathfrak{g}_C^+/\mathfrak{g}_F^+)} \widehat{T} = 0.$$
(Note, we are assuming in this statement that the form $B$ introduced in §2.2 has certain properties — for example, that it descends to a nondegenerate, symmetric, nondegenerate, bilinear form on $\mathsf{L}_F(\mathfrak{f})$.) However, as discussed previously, $\mathfrak{g}_C^+/\mathfrak{g}_F^+$ is the nilradical of a Borel subgroup of $\mathsf{G}_F(\mathfrak{f})$. Thus
$$\operatorname{res}_{\sum_{F\subset \bar{C}} C(\mathfrak{g}_F/\mathfrak{g}_C)} T = 0$$
if and only if
$$\widehat{T}([(F,\bar{\mathcal{O}})]) = 0$$
for all $(F,\bar{\mathcal{O}}) \in I^d$ where $[(F,\bar{\mathcal{O}})]$ denotes the characteristic function of the preimage of $\bar{\mathcal{O}}$. It is then not difficult to see that this is equivalent to the statement
$$\widehat{T}([(F,\bar{\mathcal{O}})]) = 0$$
where $(F,\bar{\mathcal{O}}) \in I^d$ runs over a set of representatives for $I^d/\sim$. But from Theorem 4.4.3, this implies that the dimension of $\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathfrak{g}_0)$ is less than or equal to the cardinality of $\mathcal{O}(0)$. On the other hand, $J(\mathcal{N}) \subset J(\mathfrak{g}_0)$ and from Harish-Chandra [**7**] we know that the dimension of
$$\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathcal{N})$$
is equal to the number of nilpotent orbits. So (1) follows from (2).

**5.3. Some applications.** We present here two quick applications that are related to material presented elsewhere in this workshop. The final section of these notes is dedicated to giving a more thorough (yet still incomplete) treatment of an application.

First, the above homogeneity statement gives us a sharpened version of the Harish-Chandra–Howe local character expansion. Suppose, as usual, that $p$ is large. Let $(\pi, V)$ be an irreducible admissible representation of $G$. If there exists a facet

$F$ in $\mathcal{A}$ for which $V^{G_F^+} \neq \{0\}$ (that is, $(\pi, V)$ has depth zero), then there exist complex constants $c_{\mathcal{O}}(\pi)$ for which

$$\Theta_\pi(\exp(X)) = \sum_{\mathcal{O}\in\mathcal{O}(0)} c_{\mathcal{O}}(\pi)\cdot \widehat{\mu}_{\mathcal{O}}(X)$$

for all regular semisimple $X \in \mathfrak{g}_{0^+}$. Here $\mathfrak{g}_{0^+}$ denotes the set of topologically nilpotent elements, or, more precisely,

$$\mathfrak{g}_{0^+} := \bigcup_{F\subset\bar{C}} {}^G\mathfrak{g}_F^+.$$

For $\mathrm{GL}_n(k)$ the set of topologically nilpotent elements is exactly the set of elements in $\mathrm{M}_n(k)$ for which each eigenvalue has positive valuation. Note that we are also assuming that $\exp\colon \mathfrak{g}_{0^+} \to G_{0^+}$ is bijective.

Second, again assuming that $p$ is sufficiently large, we can derive a sharpened Shalika-germ expansion. Namely, for all regular semisimple $X \in \mathfrak{g}_0$ we have

$$\widehat{\mu}_X(Y) = \sum_{\mathcal{O}\in\mathcal{O}(0)} \Gamma_{\mathcal{O}}(X)\cdot \widehat{\mu}_{\mathcal{O}}(Y)$$

for all regular semisimple $Y \in \mathfrak{g}_{0^+}$.

## 6. An application: stable distributions supported on the nilpotent cone

In this section, we sketch a final application of the homogeneity result stated above. This section should be thought of as an introduction to the techniques found in Waldspurger's tome [**13**].

**6.1. Stability.** For some purposes, the concept of stable invariance is more natural than the concept of invariance; however, the definition of stable invariance is far less natural. In order to motivate the definition of stability, we begin by recalling a result of Harish-Chandra.

We define $\mathcal{D}^{\mathrm{ann}}$ to be the space of functions that vanish on every regular semisimple orbital integral. That is

Definition 6.1.1.

$$\mathcal{D}^{\mathrm{ann}} = \{f \in C_c^\infty(\mathfrak{g}) \,|\, \mu_X(f) = 0 \text{ for all regular semisimple } X \in \mathfrak{g}\}.$$

We then have

Theorem 6.1.2 ([**7**]). *Suppose $T \in C_c^\infty(\mathfrak{g})^*$, that is, $T$ is a distribution on $\mathfrak{g}$ (not necessarily invariant). We have*

$$T \text{ is invariant if and only if } \mathrm{res}_{\mathcal{D}^{ann}} T = 0.$$

In other words, regular semisimple orbital integrals are dense in the space of invariant distributions. We remark that a key step in the proof is to show that $\mathrm{res}_{\mathcal{D}^{\mathrm{ann}}} \mu_{\mathcal{O}} = 0$ for each $\mathcal{O} \in \mathcal{O}(0)$.

Motivated by this result of Harish-Chandra, we can now define $J^{\mathrm{st}}(\mathfrak{g})$, the space of stably invariant distributions on $\mathfrak{g}$. We begin by introducing the idea of a stable orbital integral. Suppose $X \in \mathfrak{g}$ is regular semisimple. There is a finite set $\{X_\ell \,|\, 1 \leq \ell \leq n\}$ of regular semisimple elements in $\mathfrak{g}$ so that ${}^{\mathbf{G}(\bar{k})}X \cap \mathfrak{g}$ can be written as a disjoint union

$${}^{\mathbf{G}(\bar{k})}X \cap \mathfrak{g} = {}^G X_1 \sqcup {}^G X_2 \sqcup \cdots \sqcup {}^G X_n.$$

After suitably normalizing measures, we set

$$S\mu_X = \sum_{\ell=1}^{n} \mu_{X_\ell}$$

and we call $S\mu_X$ a *stable orbital integral*.

The analogue of $\mathcal{D}^{\mathrm{ann}}$ becomes the space of functions that vanish on every stable orbital integral. That is,

DEFINITION 6.1.3.

$$\mathcal{D}^{\mathrm{stann}} := \{f \in C_c^\infty(\mathfrak{g}) \mid S\mu_X(f) = 0 \text{ for all regular semisimple } X \in \mathfrak{g}\}.$$

We then define

DEFINITION 6.1.4.

$$J^{\mathrm{st}}(\mathfrak{g}) := \{T \in C_c^\infty(\mathfrak{g})^* \mid \mathrm{res}_{D^{\mathrm{stann}}} T = 0\}.$$

Note that since $\mathcal{D}^{\mathrm{ann}} \subset \mathcal{D}^{\mathrm{stann}}$, every element of $J^{\mathrm{st}}(\mathfrak{g})$ is an invariant distribution on $\mathfrak{g}$.

EXAMPLE 6.1.5. Here are some examples of elements of $J^{\mathrm{st}}(\mathfrak{g})$.

- For all regular semisimple $X \in \mathfrak{g}$, the distribution $S\mu_X$ is stable.
- The distribution $\mu_{\{0\}}$ is stable.
- The distribution which sends $f \in C_c^\infty(\mathfrak{g})$ to $\int_{\mathfrak{g}} f(X)\,dX$ is stable.

Herein lies the basic problem: beyond the examples listed above, we have essentially no general understanding of $J^{\mathrm{st}}(\mathfrak{g})$. A natural first question to ask is: can we understand $J^{\mathrm{st}}(\mathcal{N}) := J(\mathcal{N}) \cap J^{\mathrm{st}}(\mathfrak{g})$? For certain unramified classical groups, Waldspurger has provided an affirmative answer to this question.

**6.2. A first step towards understanding $J^{\mathrm{st}}(\mathcal{N})$.** The following result, due to Waldspurger [**13**], gives us a way to tackle the problem of describing $J^{\mathrm{st}}(\mathcal{N})$. The argument is very similar to one that Harish-Chandra used to prove Theorem 6.1.2.

LEMMA 6.2.1 ([**13**]). *Suppose $T \in J(\mathfrak{g}_0)$. Let*

$$D = \sum_{\mathcal{O} \in \mathcal{O}(0)} c_{\mathcal{O}}(T) \cdot \mu_{\mathcal{O}}$$

*(with $c_{\mathcal{O}}(T) \in \mathbb{C}$) denote the unique element in $J(\mathcal{N})$ for which*

$$\mathrm{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} T = \mathrm{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} D.$$

*If $T \in J^{st}(\mathfrak{g})$, then $D \in J^{st}(\mathcal{N})$.*

PROOF. Fix $f \in \mathcal{D}^{\mathrm{stann}}$. We need to show that $D(f) = 0$.

We note that if $t \in k^\times$, then $f_{t^2} \in \mathcal{D}^{\mathrm{stann}}$ . Choose $t \in k^\times \smallsetminus R^\times$ such that $f_{t^{2n}} \in C_c(\mathfrak{g}/\mathfrak{g}_C)$ for all $n \geq 1$. For all $n \geq 1$ we have

$$\begin{aligned} 0 = T(f_{t^2 n}) &= D(f_{t^2 n}) \\ &= \sum_{\mathcal{O} \in \mathcal{O}(0)} c_{\mathcal{O}}(T) \cdot \mu_{\mathcal{O}}(f_{t^{2n}}) \\ &= \sum_{i=0} |t|^{-in} \sum_{\mathcal{O} \in \mathcal{O}(0); \dim(\mathcal{O}) = i} c_{\mathcal{O}}(T) \cdot \mu_{\mathcal{O}}(f). \end{aligned}$$

Since the characters $n \mapsto |t|^{-in}$ are linearly independent, each of the terms

$$\sum_{\mathcal{O}\in\mathcal{O}(0);\dim(\mathcal{O})=i} c_{\mathcal{O}}(T)\cdot\mu_{\mathcal{O}}(f)$$

must be zero. Consequently $D(f) = 0$. $\square$

Thus, one way to find a basis for $J^{\mathrm{st}}(\mathcal{N})$ is to first produce a basis for $\mathrm{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathfrak{g}_0)$ with the properties

- the elements of the basis are of the form $\mathrm{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)}\,\mu_X$ with $X \in \mathfrak{g}_0$ regular semisimple, and
- we can easily describe which combinations of the $\mu_X$ are stable.

**6.3. A dual basis.** Fix a set of representatives $\{(F_i,\bar{\mathcal{O}}_i)\in I^d \mid 1\le i\le |\mathcal{O}(0)|\}$ for $I^d/\sim$. Recall that for $T\in J(\mathfrak{g}_0)$ we have $\mathrm{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} T = 0$ if and only if $\widehat{T}([(F_i,\bar{\mathcal{O}}_i)]) = 0$ for $1\le i\le|\mathcal{O}(0)|$. Thus the Fourier transforms of the functions $[(F_i,\bar{\mathcal{O}}_i)]$ form a dual basis for $\mathrm{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathfrak{g}_0)$. (Note that the Fourier transform of the function $[(F,\bar{\mathcal{O}})]$ does not belong to $C_c(\mathfrak{g}/\mathfrak{g}_C)$, but, rather, it belongs to $\sum_{g\in G} C_c(\mathfrak{g}/{}^g\mathfrak{g}_C)$. However, since $T$ is an invariant distribution, this will not cause us any difficulties.) So, the idea is to produce well-understood functions on $\mathsf{L}_F(\mathfrak{f})$ that separate distinguished nilpotent orbits and (might) have something to do with regular semisimple orbital integrals. Thanks to work of Deligne, Kazhdan, Lusztig, and others, such functions exist:

FACT 6.3.1 ([**10**]). There exist class functions on $\mathsf{L}_F(\mathfrak{f})$, called *generalized Green functions*, such that

- the functions span the set of class functions supported on the nilpotent elements in $\mathsf{L}_F(\mathfrak{f})$,
- the cuspidal[9] generalized Green functions separate distinguished orbits, and
- the functions are well understood.

EXAMPLE 6.3.2. If $\mathsf{T}\le\mathsf{G}_F$ is an $\mathfrak{f}$-minisotropic torus[10], then the usual Green function

$$Q_{\mathsf{T}}^{\mathsf{G}_F}(\bar{X}) = \begin{cases} 0 & \bar{X} \text{ is not nilpotent} \\ R_{\mathsf{T}}^{\mathsf{G}_F}(1)(\exp(\bar{X})) & \text{otherwise.}\end{cases}$$

is a cuspidal generalized Green function. Note that exp makes sense in this context because $\bar{X}$ is nilpotent, and we are assuming that $p$ is not too small.

Note that not all cuspidal generalized Green functions occur as in this example; this is already the case for $\mathrm{SL}_2(\mathfrak{f})$.

We define $I^G$ to be the set of pairs $(F,\mathcal{G})$ where $F$ is a facet in $\mathcal{A}$ and $\mathcal{G}$ is a cuspidal generalized Green function on $\mathsf{L}_F(\mathfrak{f})$. As in the case of $I^d$, the set $I^G$ carries a natural equivalence relation, which we also denote by $\sim$. Given the above discussion, it is not hard to believe that the following lemma is valid.

---

[9]A function is called *cuspidal* provided that summing against the nilradical of any proper parabolic yields zero.

[10]An $\mathfrak{f}$-torus is called $\mathfrak{f}$-*minisotropic* in $\mathsf{G}_F$ provided that its maximal $\mathfrak{f}$-split torus lies in the center of $\mathsf{G}_F$.

LEMMA 6.3.3 ([**13**]). *Suppose $T \in J(\mathfrak{g}_0)$. We have*

$$\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} T = 0 \quad \textit{if and only if } T(\hat{\mathcal{G}}_F) = 0$$

*for all $(F, \mathcal{G}) \in I^G/\sim$. Here $\hat{\mathcal{G}}_F$ denotes the inflation of $\hat{\mathcal{G}}$ to a function on $\mathfrak{g}$.*

**6.4. A well-chosen basis for $\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathfrak{g}_0)$.** As discussed before, we want to find a basis for $\operatorname{res}_{C_c(\mathfrak{g}/\mathfrak{g}_C)} J(\mathfrak{g}_0)$ with several good properties. It would be even better if this basis were dual to $I^G/\sim$. As evidenced by the size of [**13**], this is quite a difficult problem. However, it is not too difficult to sketch how to carry out this program for the generalized Green functions of the form $Q_{\mathsf{T}}^{\mathsf{G}_F}$.

Fix an element of $I^G$ of the form $(F, Q_{\mathsf{T}}^{\mathsf{G}_F})$. Choose absolutely any $X_{\mathsf{T}} \in \mathfrak{g}_F$ for which the centralizer in $\mathsf{G}_F$ of the image of $X_{\mathsf{T}}$ in $\mathsf{L}_F$ is $\mathsf{T}$. Note that such an $X_{\mathsf{T}}$ is necessarily regular semisimple and $\mu_{X_{\mathsf{T}}} \in J(\mathfrak{g}_0)$. Using results of Kazhdan [**9**], Waldspurger proves

LEMMA 6.4.1 ([**13**]). *For $X_{\mathsf{T}}$ as above and $(F', \mathcal{G}') \in I^G$ we have*

$$\mu_{X_{\mathsf{T}}}(\hat{\mathcal{G}}_{F'}) = \begin{cases} 0 & (F', \mathcal{G}') \not\sim (F, Q_{\mathsf{T}}^{\mathsf{G}_F}) \\ N & \textit{otherwise.} \end{cases}$$

*where $N$ is a nice nonzero number which is independent of the choice of $X_{\mathsf{T}}$.*

As a consequence of this lemma, we have that $\operatorname{res}_{C_c^\infty(\mathfrak{g}_{0+})} \hat{\mu}_{X_{\mathsf{T}}}$ is independent of how $X_{\mathsf{T}}$ was chosen. This is a much stronger version of Lemma 2.2.2. To see why the elements $\operatorname{res}_{C_c^\infty(\mathfrak{g}_{0+})} \mu_{X_{\mathsf{T}}}$ are particularly nice to deal with, we must return to Bruhat-Tits theory.

**6.5. Parameterizing maximal unramified tori.** A subgroup $T \leq G$ is called an *unramified torus* provided that it is the group of $k$-rational points of a torus which splits over an unramified extension of $k$.

EXAMPLE 6.5.1. We begin by considering some examples.

- The group $A$ is always a maximal unramified torus.
- If $p \neq 2$ and $\varepsilon \in R^\times \setminus (R^\times)^2$, then

  $$\{ \left(\begin{smallmatrix} a & b \\ b\varepsilon & a \end{smallmatrix}\right) \mid a^2 - b^2\varepsilon = 1\}$$

  is a maximal unramified torus in $\mathrm{SL}_2(k)$, but the torus

  $$\{ \left(\begin{smallmatrix} a & b \\ b\varpi & a \end{smallmatrix}\right) \mid a^2 - b^2\varpi = 1\}$$

  is not.

Just as we parameterized the elements of $\mathcal{O}(0)$ in terms of similar objects over the finite field, we would like to do the same for conjugacy classes of maximal unramified tori. This time, the objects over the finite field will be conjugacy classes of maximal $\mathfrak{f}$-minisotropic tori.

Suppose $\mathsf{G}$ is a connected $\mathfrak{f}$-split reductive group. From Carter [**3**] the $\mathsf{G}(\mathfrak{f})$-conjugacy classes of maximal $\mathfrak{f}$-tori in $\mathsf{G}$ are parameterized by the conjugacy classes in the Weyl group of $\mathsf{G}$. We sketch how this parameterization works: Let $\mathsf{S}$ be a maximal $\mathfrak{f}$-split torus in $\mathsf{G}$ and let $\sigma$ denote a topological generator for $\operatorname{Gal}(\bar{\mathfrak{f}}/\mathfrak{f})$. If $\mathsf{T}$ is any $\mathfrak{f}$-torus, then there is a $g \in \mathsf{G}(\bar{\mathfrak{f}})$ such that $\mathsf{T} = {}^g\mathsf{S}$. Since $\mathsf{T}$ and $\mathsf{S}$ are $\sigma$-stable, the element $\sigma(g)^{-1}g$ belongs to the normalizer of $\mathsf{S}$ in $\mathsf{G}$ and so determines a conjugacy class in the Weyl group.

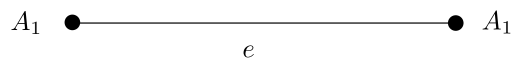



FIGURE 16. An enumeration of classes of $\mathfrak{f}$-minisotropic tori for $\mathrm{SL}_2$

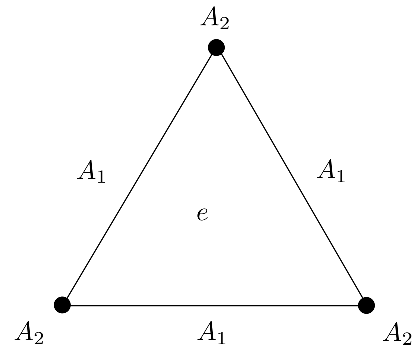



FIGURE 17. An enumeration of classes of $\mathfrak{f}$-minisotropic tori for $\mathrm{SL}_3$

The maximal $\mathfrak{f}$-minisotropic tori in $\mathsf{G}$ are parameterized by the anisotropic[11] conjugacy classes of the Weyl group. We shall use Carter's notation for the conjugacy classes in the Weyl group.

EXAMPLE 6.5.2. The group $\mathrm{SL}_2$ has two $\mathrm{SL}_2(\mathfrak{f})$-conjugacy classes of maximal $\mathfrak{f}$-tori. One is $\mathfrak{f}$-minisotropic and corresponds to $A_1$ (see Figure 16), the nontrivial conjugacy class in the Weyl group, while the other is $\mathfrak{f}$-split and corresponds to the trivial conjugacy class in the Weyl group. The group $\mathrm{GL}_1$ has a single $\mathrm{GL}_1(\mathfrak{f})$-conjugacy class of maximal $\mathfrak{f}$-tori, namely $\mathrm{GL}_1(\mathfrak{f})$ itself. In Figure 16 we enumerate the number of $\mathsf{G}_F(\mathfrak{f})$-conjugacy classes of $\mathfrak{f}$-minisotropic tori for each facet $F$ in an alcove for $\mathrm{SL}_2(k)$. The sum of the enumerated classes is three, and the number of $\mathrm{SL}_2(k)$-conjugacy classes of maximal unramified tori is three. (Can you produce a representative for the third class?)

The map from tori over $\mathfrak{f}$ to tori over $k$ is not as easy to describe as in the nilpotent case, but it has the advantage of working independent of the residual characteristic.

In general, we want to consider the set of pairs

$$I^m := \{(F, {}^{\mathsf{G}_F(\mathfrak{f})}\mathsf{T})\}$$

where $F$ is a facet in $\mathcal{A}$ and ${}^{\mathsf{G}_F(\mathfrak{f})}\mathsf{T}$ is short-hand for the set of $\mathfrak{f}$-tori which are $\mathsf{G}_F(\mathfrak{f})$-conjugate to the $\mathfrak{f}$-minisotropic torus $\mathsf{T}$. As with the sets $I^d$ and $I^G$, the set $I^m$ carries a natural equivalence relation, which we again denote by $\sim$.

THEOREM 6.5.3 ([**5**]). *We have a natural bijective correspondence between* $I^m/\!\sim$ *and the set of* $G$*-conjugacy classes of maximal unramified tori.*

EXAMPLE 6.5.4. The group $\mathrm{SL}_3(k)$ has five conjugacy classes of maximal unramified tori. In Figure 17 we use Carter's labeling for the conjugacy classes in the Weyl group to enumerate the $\mathsf{G}_F(\mathfrak{f})$-conjugacy classes of $\mathfrak{f}$-minisotropic tori for

[11]A conjugacy class in a Weyl group is called *anisotropic* provided that it does not intersect a proper parabolic subgroup of the Weyl group

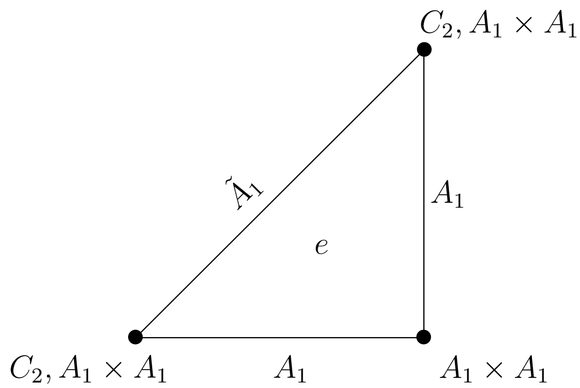



FIGURE 18. An enumeration of classes of $\mathfrak{f}$-minisotropic tori for $\mathrm{Sp}_4$

each facet $F$ in an alcove for $\mathrm{SL}_3(k)$. (Recall that the line segments in the closure of the alcove are equivalent.)

EXAMPLE 6.5.5. The group $\mathrm{Sp}_4(k)$ has nine conjugacy classes of maximal unramified tori. In Figure 18 we again list the anisotropic Weyl group conjugacy classes to enumerate the $\mathsf{G}_F(\mathfrak{f})$-conjugacy classes of $\mathfrak{f}$-minisotropic tori for each facet $F$ in an alcove for $\mathrm{Sp}_4(k)$.

**6.6. The finish.** To complete these notes, we remark that it is now nearly trivial to describe the number of distributions in $J^{\mathrm{st}}(\mathcal{N})$ arising from pairs of the form $(F, Q_{\mathsf{T}}^{\mathsf{G}_F})$.

In the preceding sections, we have discussed how to associate to the pair $(F, Q_{\mathsf{T}}^{\mathsf{G}_F}) \in I^G$ a regular semisimple orbital integral $\mu_{X_{\mathsf{T}}}$. On the other hand, $(F, Q_{\mathsf{T}}^{\mathsf{G}_F})$ is naturally associated to the pair $(F, \mathsf{T})$ which is associated to a conjugacy class in the Weyl group of $\mathsf{G}_F$. We can lift this conjugacy class to a $\tilde{W}$-conjugacy class in the extended affine Weyl group $\tilde{W}$ and then quotient by $A$ to arrive at a conjugacy class, call it $w_{\mathsf{T}}$, in $W$.

Suppose $(F', Q_{\mathsf{T}'}^{\mathsf{G}_{F'}})$ is another element of $I^G$ with associated regular semisimple orbital integral $\mu_{X_{\mathsf{T}'}}$. From [**5**] the elements $X_{\mathsf{T}}$ and $X_{\mathsf{T}'}$ can be chosen to be stably conjugate if and only if $w_{\mathsf{T}} = w_{\mathsf{T}'}$. Consequently, to each $W$-conjugacy class in $W$ we can associate one distribution in $J^{\mathrm{st}}(\mathcal{N})$. Thus, the dimension of $J^{\mathrm{st}}(\mathcal{N})$ is at least equal to the number of $W$-conjugacy classes in $W$.

EXAMPLE 6.6.1. From the above discussion, we can conclude the following. For $\mathrm{SL}_2(k)$, the dimension of $J^{\mathrm{st}}(\mathcal{N})$ is at least two (in fact, it is two). For $\mathrm{SL}_3(k)$, the dimension of $J^{\mathrm{st}}(\mathcal{N})$ is at least three (in fact, it is three). For $\mathrm{Sp}_4(k)$, the dimension of $J^{\mathrm{st}}(\mathcal{N})$ is at least five (in fact, it is six).

To describe the elements of $J^{\mathrm{st}}(\mathcal{N})$ is an entirely different and much more demanding problem. Such a description will rely on all that we have discussed here and more.

## References

[1] J. Adler and S. DeBacker *Some applications of Bruhat-Tits theory to harmonic analysis on the Lie algebra of a reductive p-adic group*, Michigan Math. J. **50** (2002), no. 2, pp. 263–286.

[2] D. Barbasch and A. Moy, *Local character expansions*, Ann. Sci. École Norm. Sup. (4) **30** (1997), no. 5, pp. 553–567.

[3] R. Carter, *Finite groups of Lie type. Conjugacy classes and complex characters*, Reprint of the 1985 original, Wiley Classics Library, John Wiley & Sons, Ltd., Chichester, 1993.

[4] S. DeBacker, *Homogeneity results for invariant distributions of a reductive p-adic group*, Ann. Sci. École Norm. Sup., **35** (2002), no. 3, pp. 391-422.

[5] ———, *Parameterizing conjugacy classes of maximal unramified tori via Bruhat-Tits theory*, preprint, 2001.

[6] ———, *Parameterizing nilpotent orbits via Bruhat-Tits theory*, Ann. of Math., **156** (2002), no. 1, pp. 295-332.

[7] Harish-Chandra, *Admissible invariant distributions on reductive p-adic groups*, Preface and notes by Stephen DeBacker and Paul J. Sally, Jr., University Lecture Series, **16**, American Mathematical Society, Providence, RI, 1999.

[8] R. Howe, *The Fourier transform and germs of characters (case of $Gl_n$ over a p-adic field)*, Math. Ann. **208** (1974), pp. 305–322.

[9] D. Kazhdan, *Proof of Springer's hypohesis*, Israel J. Math. **28** (1977), pp. 272–286.

[10] G. Lusztig, *Green functions and character sheaves*, Ann. Math., **131** (1990), pp. 355–408.

[11] *Oxford English Dictionary*, second edition, Oxford University Press, Oxford, 1989.

[12] J. Rabinoff, *The Bruhat-Tits builidng of a p-adic Chevalley group and an application to representation theory*, Harvard Senior Thesis, 2003, http://shadowfax.homelinux.net:8080/building/

[13] J.-L. Waldspurger, *Intégrales orbitales nilpotentes et endoscopie pour les groupes classiques non ramifiés*, Astérisque No. 269, 2001.

[14] ——— *Quelques resultats de finitude concernant les distributions invariantes sur les algèbres de Lie p-adiques*, preprint, 1993.

[15] *Webster's Ninth New Collegiate Dictionary*, Merriam-Webster, Springfield, MA, 1985.

The University of Michigan, Ann Arbor, MI 48109
*E-mail address*: smdbackr@umich.edu

Clay Mathematics Proceedings
Volume 4, 2005

# Compactifications and Cohomology of Modular Varieties

Mark Goresky

## 1. Overview

Let $\mathbf{G}$ be a connected reductive linear algebraic group defined over $\mathbb{Q}$. Denote by $\mathbf{G}(\mathbb{Q})$ (resp. $\mathbf{G}(\mathbb{R})$) the group of points in $\mathbf{G}$ with entries in $\mathbb{Q}$ (resp. $\mathbb{R}$). It is common to write $G = \mathbf{G}(\mathbb{R})$. Fix a maximal compact subgroup $K \subset G$ and let $A_G = \mathbf{A_G}(\mathbb{R})^+$ (see §4.1) be the (topologically) connected identity component of the group of real points of the greatest $\mathbb{Q}$-split torus $\mathbf{A_G}$ in the center of $\mathbf{G}$. (If $\mathbf{G}$ is semisimple then $A_G = \{1\}$.) We refer to $D = G/KA_G$ as the "symmetric space" for $\mathbf{G}$. We assume it is Hermitian, that is, it carries a $G$-invariant complex structure. Fix an arithmetic subgroup $\Gamma \subset \mathbf{G}(\mathbb{Q})$ and let $X = \Gamma\backslash D$. We refer to $X$ as a locally symmetric space. In general, $X$ is a rational homology manifold: at worst, it has finite quotient singularities. If $\Gamma$ is torsion-free then $X$ is a smooth manifold. It is usually noncompact. (If $\mathbb{A}$ denotes the adèles of $\mathbb{Q}$ and $\mathbb{A}_f$ denotes the finite adèles, and if $K_f \subset \mathbf{G}(\mathbb{A}_f)$ is a compact open subgroup, then the topological space $Y = \mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})/A_GK \cdot K_f$ is a disjoint union of finitely many locally symmetric spaces for $\mathbf{G}$. To compactify $Y$ it suffices to compactify each of these locally symmetric spaces.)

There are (at least) four important compactifications of $X$ : the Borel-Serre compactification $\overline{X}^{\mathrm{BS}}$ (which is a manifold with corners), the reductive Borel-Serre compactification, $\overline{X}^{\mathrm{RBS}}$ (which is a stratified singular space), the Baily-Borel (Satake) compactification $\overline{X}^{\mathrm{BB}}$ (which is a complex projective algebraic variety, usually singular), and the toroidal compactification $\overline{X}^{\mathrm{tor}}_{\Sigma}$ (which is a resolution of singularities of $\overline{X}^{\mathrm{BB}}$). (Actually there is a whole family of toroidal compactifications, depending on certain choices $\Sigma$.) The identity mapping $X \to X$ extends to unique continuous mappings

$$\overline{X}^{\mathrm{BS}} \longrightarrow \overline{X}^{\mathrm{RBS}} \xrightarrow{\tau} \overline{X}^{\mathrm{BB}} \longleftarrow \overline{X}^{\mathrm{tor}}_{\Sigma}.$$

The first three of these compactifications are obtained as the quotient under $\Gamma$ of corresponding "partial compactifications"

$$\overline{D}^{\mathrm{BS}} \longrightarrow \overline{D}^{\mathrm{RBS}} \longrightarrow \overline{D}^{\mathrm{BB}}$$

of the symmetric space $D$.

Partially supported by National Science Foundation grant no. 0139986.

Besides its ordinary (singular) cohomology, the two singular compactifications $\overline{X}^{\mathrm{BB}}$ and $\overline{X}^{\mathrm{RBS}}$ also support various exotic sorts of cohomology, defined in terms of a complex of sheaves of differential forms with various sorts of restrictions near the singular strata. The $L^2$ cohomology of $X$ may be realized as the cohomology of the sheaf of $L^2$ differential forms on $\overline{X}^{\mathrm{BB}}$. The (middle) intersection cohomology of $\overline{X}^{\mathrm{BB}}$ is obtained from differential forms which satisfy a condition (see §6.6) near each singular stratum, defined in terms of the dimension of the stratum. The Zucker conjecture [**Z1**], proven by E. Looijenga [**Lo**] and L. Saper and M. Stern [**SS**], says that the $L^2$ cohomology of $\overline{X}^{\mathrm{BB}}$ coincides with its intersection cohomology, and that the same is true of any open subset $U \subset \overline{X}^{\mathrm{BB}}$.

The (middle) weighted cohomology complex on $\overline{X}^{\mathrm{RBS}}$ is defined in a manner similar to that of the intersection cohomology, however the restrictions on the chains (or on differential forms) are defined in terms of the weights of a certain torus action which exists near each singular stratum. Although the weighted cohomology and the intersection cohomology do not agree on every open subset of $\overline{X}^{\mathrm{RBS}}$, it has recently been shown ([**S1**], [**S2**]) that they do agree on subsets of the form $\tau^{-1}(U)$ for any open set $U \subset \overline{X}^{\mathrm{BB}}$.

This article is in some sense complementary to the survey articles [**Sch**] and [**B4**].

*Notation.* Throughout this article, algebraic groups over $\mathbb{Q}$ will be indicated in bold, and the corresponding group of real points in Roman, so $G = \mathbf{G}(\mathbb{R})$. The group of $n \times r$ matrices over a field $k$ is denoted $M_{n\times r}(k)$. The rank $n$ identity matrix is denoted $I_n$ and the zero matrix is $0_n$.

*Acknowledgments.* This article was prepared during the Clay Mathematics Institute Summer School which was held at the Fields Institute in Toronto, in June 2003. The author is grateful to the organizers of the Summer School, the Clay Mathematics Institute, and the Fields Institute for their hospitality and support. This work was partially supported by the National Science Foundation, grant no. DMS-0139986.

## 2. The Baily Borel (Satake) compactification

**2.1. The case $\mathbf{G} = \mathbf{SL}(2)$.** Recall the fundamental domains for the action of $\Gamma = \mathbf{SL}(2,\mathbb{Z})$ on the upper half plane $\mathfrak{h}$.

The quotient $X = \Gamma\backslash\mathfrak{h}$ may be compactified, $\overline{X} = X \cup \{\infty\}$ by adding a single cusp[1] at infinity. If we wish to realize this as the quotient under $\Gamma$ of a partial compactification $\overline{\mathfrak{h}}^{\mathrm{BB}}$ of the upper half plane, then we must add to $\mathfrak{h}$ all the $\Gamma$-translates of $\{\infty\}$. This consists of all the rational points $x \in \mathbb{Q}$ on the real line (which also coincides with the $\mathbf{SL}(2,\mathbb{Q})$ orbit of the point at infinity). With this candidate for $\overline{\mathfrak{h}}^{\mathrm{BB}}$, the quotient under $\Gamma$ will fail to be Hausdorff. The solution is to re-topologise this union so as to “separate” the added points $x \in \mathbb{Q}$.

A neighborhood basis for the point at infinity may be chosen to consist of the open sets $U_\tau = \{z \in \mathfrak{h} | \operatorname{Im}(z) > \tau\}$ for $\tau \geq 2$ (say). If we also throw in all the $\mathbf{SL}(2,\mathbb{Q})$-translates of these sets $U_\tau$ then we obtain a new topology, the *Satake topology*, on $\overline{\mathfrak{h}} = \mathfrak{h} \cup \mathbb{Q} \cup \{\infty\}$, in which each point $x \in \mathbb{Q}$ has a neighborhood

[1] Although they are called cusps, the points which are added to compactify a modular curve are in fact nonsingular points of the resulting compactifications.

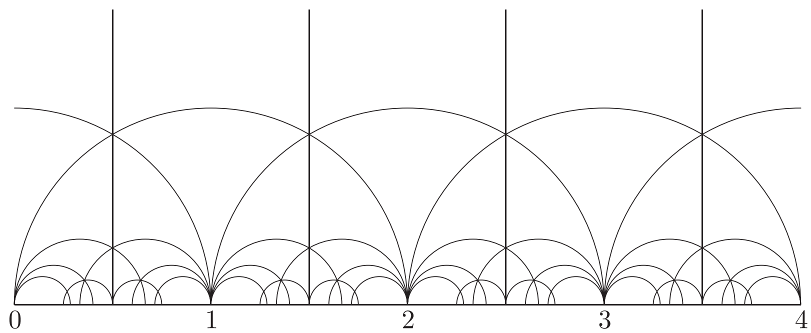



FIGURE 1. Fundamental domains for $\mathbf{SL}(2,\mathbb{Z})$

homeomorphic to the neighborhood of the point at infinity. The group $\Gamma$ still fails to act "properly" on $\overline{\mathfrak{h}}$ because for each boundary point $x \in \mathbb{Q} \cup \infty$ there are infinitely many elements $\gamma \in \Gamma$ which fix $x$. However it does satisfy (cf. [**AMRT**] p. 258) the following conditions:

(S1) If $x, x' \in \overline{\mathfrak{h}}$ are not equivalent under $\Gamma$ then there exist neighborhoods $U, U'$ of $x, x'$ respectively, such that $(\Gamma \cdot U) \cap U' = \phi$.

(S2) For every $x \in \overline{\mathfrak{h}}$ there exists a fundamental system of neighborhoods $\{U\}$, each of which is preserved by the stabilizer $\Gamma_x$, such that if $\gamma \notin \Gamma_x$ then $(\gamma \cdot U) \cap U = \phi$.

These properties guarantee that the quotient $\overline{X} = \Gamma \backslash \overline{\mathfrak{h}}$ is Hausdorff, and in fact it is compact. The same partial compactification $\overline{\mathfrak{h}}$ may be used for any arithmetic subgroup $\Gamma' \subset \mathbf{SL}(2,\mathbb{Q})$, giving a uniform method for compactifying all arithmetic quotients $\Gamma' \backslash \mathfrak{h}$. One would like to do the same sort of thing for symmetric spaces of higher rank.

**2.2. A warmup problem.** The following example, although not Hermitian, illustrates the phenomena which are encountered in the Baily Borel compactification of higher rank locally symmetric spaces. See [**AMRT**] Chapt. II for more details. The group $\mathbf{GL}(n,\mathbb{R})$ acts on the vector space $S_n(\mathbb{R})$ of real symmetric $n \times n$ matrices (through change of basis) by

$$g \cdot A = gA\,{}^t g. \tag{2.2.1}$$

The orbit of the identity matrix $I_n$ is the open (homogeneous self adjoint) convex cone $\mathcal{P}_n$ of positive definite symmetric matrices. The stabilizer of $I$ is the maximal compact subgroup $\mathbf{O}(n)$. The center $A_G$ of $\mathbf{GL}(n,\mathbb{R})$ (which consists of the scalar matrices) acts by homotheties. The action of $\mathbf{GL}(n,\mathbb{R})$ preserves the closure $\overline{\mathcal{P}}_n$ of $\mathcal{P}_n$ in $S_n(\mathbb{R})$, whose boundary $\partial\mathcal{P}_n = \overline{\mathcal{P}_n} - \mathcal{P}_n$ decomposes into a disjoint union of (uncountably many) *boundary components* as follows. A *supporting* hyperplane $H \subset S_n(\mathbb{R})$ is a hyperplane such that $H \cap \mathcal{P}_n = \phi$ and $H \cap \partial\mathcal{P}_n$ contains nonzero elements. Let $\overline{F} = \partial\mathcal{P}_n \cap H$ where $H$ is a supporting hyperplane. Then there is a unique smallest linear subspace $L \subset S_n(\mathbb{R})$ containing $\overline{F}$. The interior $F$ of $\overline{F}$ in $L$ is called a boundary component of $\mathcal{P}_n$ (much in the same way that the closure of each face of a convex polyhedron $P \subset \mathbb{R}^m$ is the intersection $P \cap H$ of $P$ with a supporting affine hyperplane $H \subset \mathbb{R}^m$). Distinct boundary components do not intersect.

Let $\mathbf{B} \subset \mathbf{GL}(n)$ be the (rational) Borel subgroup of upper triangular matrices. Parabolic subgroups containing $\mathbf{B}$ will referred to as *standard*. Each boundary component is a $\mathbf{GL}(n,\mathbb{R})$ translate of exactly one of the following *standard boundary components* $F_r$ $(1 \le r \le n-1)$ consisting of matrices $\left(\begin{smallmatrix} E & 0 \\ 0 & 0 \end{smallmatrix}\right)$ such that $E \in \mathcal{P}_r$ is positive definite. The normalizer $P \subset \mathbf{GL}(n,\mathbb{R})$ of this boundary component (meaning the set of elements which preserve $F_r$) is the standard maximal parabolic subgroup consisting of matrices $g = \left(\begin{smallmatrix} A & B \\ 0 & D \end{smallmatrix}\right)$ where $A \in \mathbf{GL}(r,\mathbb{R})$, $B \in \mathbf{Hom}(\mathbb{R}^{n-r},\mathbb{R}^r)$, and $D \in \mathbf{GL}(n-r,\mathbb{R})$. (It is the group of real points $P = \mathbf{P}(\mathbb{R})$ of the obvious maximal parabolic subgroup $\mathbf{P} \subset \mathbf{G}$.) The supporting subspace $L$ of $F_r$ is the set of all symmetric matrices $t = \left(\begin{smallmatrix} T & 0 \\ 0 & 0 \end{smallmatrix}\right)$ where $T \in S_r(\mathbb{R})$. The action of such an element $g \in P$ on the element $t$ is given by $g \cdot t = gt\,{}^t g$, that is,

$$\begin{pmatrix} A & B \\ 0 & D \end{pmatrix} \cdot \begin{pmatrix} T & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} AT\,{}^tA & 0 \\ 0 & 0 \end{pmatrix} \tag{2.2.2}$$

for any $T \in S_r(\mathbb{R})$. In particular, the Levi component of $P$ decomposes as a product $\mathbf{GL}(r,\mathbb{R}) \times \mathbf{GL}(n-r,\mathbb{R})$ where the first factor, $A$, acts transitively on the boundary component $F_r$ and the second factor, $D$, acts trivially. The standard parabolic subgroups correspond to subsets of the Dynkin diagram of $\mathbf{G}$, and the maximal parabolic subgroup $\mathbf{P}$ corresponds to the deletion of a single node, $\alpha$.

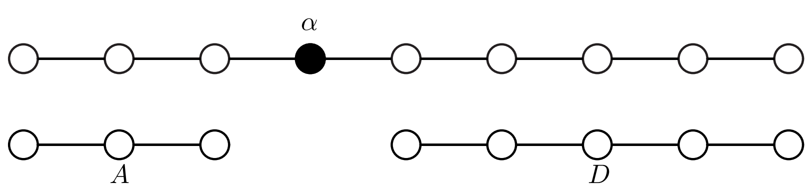



FIGURE 2. Dynkin diagrams for $\mathbf{G}$, $\mathbf{P}$, and its Levi factor.

A boundary component $F$ is *rational* if the subspace $L$ containing it is rational, or equivalently, if the normalizer $P$ is a rational parabolic subgroup. Define the *standard* partial compactification, $\overline{\mathcal{P}}_n^{\mathrm{std}}$ to be the union of $\mathcal{P}_n$ with all its rational boundary components, with the Satake topology[2]. Then $\mathbf{GL}(n,\mathbb{Q})$ acts on $\overline{\mathcal{P}}_n^{\mathrm{std}}$. For any arithmetic group $\Gamma \subset \mathbf{GL}(n,\mathbb{Q})$ the quotient

$$\overline{X}^{\mathrm{std}} = \Gamma\backslash\overline{\mathcal{P}}_n^{\mathrm{std}}/A_G = \Gamma\backslash\overline{\mathcal{P}}_n^{\mathrm{std}}/\text{homotheties}$$

is a compact singular space which is stratified with finitely many strata of the form $X_F = \Gamma_F\backslash F/\text{homotheties}$ (where $F$ is a rational boundary component and

[2] The Satake topology ([**AMRT**] p. 258, [**BB**] Thm. 4.9): is uniquely determined by requiring that conditions (S1) and (S2) (above) hold for any arithmetic group $\Gamma$, as well as the following: for any Siegel set $S \subset D$ its closure in $\overline{D}$ and its closure in $\overline{D}^{\mathrm{BB}}$ have the same topology.

$\Gamma_F$ is an appropriate arithmetic group). The closure $\overline{X}_F$ in $\overline{X}^{\text{std}}$ is the standard compactification of $F$/homotheties.

A similar construction holds for any rational self adjoint homogeneous cone. These are very interesting spaces. Although they are not algebraic varieties, they have a certain rigid structure. For example, each $\Gamma$-invariant rational polyhedral simplicial cone decomposition of $\overline{\mathcal{P}}_n^{\text{std}}$ (in the sense of [**AMRT**]) passes to a "flat, rational" triangulation of $\overline{X}^{\text{std}}$. In some cases ([**GT2**]) there is an associated real algebraic variety.

**2.3. Hermitian symmetric domains.** (Standard references for this section include [**AMRT**] Chapt. III and [**Sa1**] Chapt. II.) Assume that $\mathbf{G}$ is semisimple, defined over $\mathbb{Q}$, that $K \subset \mathbf{G}(\mathbb{R})$ is a maximal compact subgroup and that $D = G/K$ is Hermitian. The symmetric space $D$ may be holomorphically embedded in Euclidean space $\mathbb{C}^m$ as a bounded (open) domain, by the Harish Chandra embedding ([**AMRT**] p. 170, [**Sa1**] §II.4). The action of $G$ extends to the closure $\overline{D}$. The boundary $\partial D = \overline{D} - D$ is a smooth manifold which decomposes into a (continuous) union of *boundary components*. Let us say that a real affine hyperplane $H \subset \mathbb{C}^m$ is a *supporting* hyperplane if $H \cap \overline{D}$ is nonempty but $H \cap D$ is empty. Let $H$ be a supporting hyperplane and let $\overline{F} = H \cap \overline{D} = H \cap \partial D$. Let $L$ be the smallest affine subspace of $\mathbb{C}^m$ which contains $\overline{F}$. Then $\overline{F}$ is the closure of a nonempty open subset $F \subset L$ which is then a single boundary component of $D$ ([**Sa1**], III.8.11). The boundary component $F$ turns out to be a bounded symmetric domain in $L$. Distinct boundary components have nonempty intersection, and the collection of boundary components decomposes $\partial D$. Alternatively, it is possible ([**Sa1**] III.8.13) to characterize each boundary component as a single holomorphic path component of $\partial D$ : two points $x, y, \in \partial D$ lie in a single boundary component $F$ iff they are both in the image of a holomorphic "path" $\alpha : \Delta \to \partial D$ (where $\Delta$ denotes the open unit disk). In this case $\alpha(\Delta)$ is completely contained in $F$.

Fix a boundary component $F$. The normalizer $N_G(F)$ (consisting of those group elements which preserve the boundary component $F$) turns out to be a (proper) parabolic subgroup of $G$. The boundary component $F$ is *rational* if this subgroup is rationally defined in $\mathbf{G}$. There are countably many rational boundary components. If we decompose $\mathbf{G}$ into its $\mathbb{Q}$ simple factors, $\mathbf{G} = \mathbf{G_1} \times \ldots \times \mathbf{G_k}$ then the symmetric space $D$ decomposes similarly, $D = D_1 \times \ldots \times D_k$. Each (rational) boundary component $F$ of $D$ is then the product $F = F_1 \times \ldots \times F_k$ where either $F_i = D_i$ or $F_i$ is a proper (rational) boundary component of $D_i$. The normalizer of $F$ is the product $N_G(F) = N_{G_1}(F_1) \times \ldots \times N_{G_k}(F_k)$ (writing $N_{G_i}(D_i) = G_i$ whenever necessary). If $\mathbf{G}$ is $\mathbb{Q}$-simple then the normalizer $N_G(F)$ is a *maximal* (rational) proper parabolic subgroup of $\mathbf{G}$.

2.1. Definition. The Baily-Borel-Satake partial compactification $\overline{D}^{\text{BB}}$ is the union of $D$ together with all its rational boundary components, with the Satake topology.

2.1. Theorem. ([**BB**]) *The closure $\overline{F}$ of each rational boundary component $F \subset \overline{D}^{BB}$ is the Baily-Borel-Satake partial compactification $\overline{F}^{BB}$ of $F$. The group $\mathbf{G}(\mathbb{Q})$ acts continuously, by homeomorphisms on the partial compactification $\overline{D}^{BB}$. The action of any arithmetic group $\Gamma \subset \mathbf{G}(\mathbb{Q})$ on $\overline{D}^{BB}$ satisfies conditions (S1)* *and (S2) of §2.1 and the quotient* $\overline{X}^{BB} = \Gamma \backslash \overline{D}^{BB}$ *is compact. Moreover, it admits the structure of a complex projective algebraic variety.*

**2.4. Remarks.** Dividing by $\Gamma$ has two effects: it identifies (rational) boundary components whose normalizers are $\Gamma$-conjugate, and it makes identifications within each (rational) boundary component. The locally symmetric space $X$ is open and dense in $\overline{X}^{\mathrm{BB}}$. If $\kappa : \overline{D}^{\mathrm{BB}} \to \overline{X}^{\mathrm{BB}}$ denotes the quotient mapping, and if $F$ is a rational boundary component then its image $X_F = \kappa(F)$ is the quotient $\Gamma_F \backslash F$ under the subgroup $\Gamma_F = \Gamma \cap N_G(F)$ which preserves $F$, and it is referred to as a *boundary stratum*. If $\Gamma$ is *neat* (see §4.1) then the stratum $X_F$ is a complex manifold.

**2.5. Symplectic group.** In this section we illustrate these concepts for the case of the symplectic group $G = \mathbf{G}(\mathbb{R}) = \mathbf{Sp}(2n, \mathbb{R})$, which may be realized as the group of $2n$ by $2n$ real matrices $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ such that ${}^tAD - {}^tCB = I$; ${}^tAC$ and ${}^tBD$ are symmetric. These are the linear transformations which preserve the symplectic form $J = \left(\begin{smallmatrix} 0 & I_n \\ -I_n & 0 \end{smallmatrix}\right)$ on $\mathbb{R}^{2n}$. The symplectic group acts on the Siegel upper half space

$$\mathfrak{h}_n = \left\{ Z = X + iY \in M_{n \times n}(\mathbb{C}) \mid {}^tZ = Z, Y > 0 \right\}$$

(meaning that $Y$ is positive definite) by fractional linear transformations:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot Z = (AZ + B)(CZ + D)^{-1}.$$

The stabilizer of the basepoint $iI_n$ is the unitary group $K = \mathbf{U}(n)$, embedded in the symplectic group by $A + iB \mapsto \left(\begin{smallmatrix} A & B \\ -B & A \end{smallmatrix}\right)$. It is a maximal compact subgroup, so $\mathfrak{h}_n = G/K$ is a Hermitian locally symmetric space.

The Harish-Chandra embedding $\phi : \mathfrak{h}_n \to D_n$ is given by the Cayley transformation. Here,

$$D_n = \left\{ w \in M_{n \times n}(\mathbb{C}) \mid {}^tw = w \text{ and } I_n - w\overline{w} > 0 \right\}$$

is a bounded domain, and

$$\phi(z) = (z - iI_n)(z + iI_n)^{-1}.$$

The closure $\overline{D}_n$ is given by relaxing the positive definite condition to positive semi-definite: $I_n - w\overline{w} \geq 0$. Each boundary component (resp. rational boundary component) is a $\mathbf{G}(\mathbb{R})$-translate (resp. $\mathbf{G}(\mathbb{Q})$-translate) of one of the $n$ different *standard* boundary components $D_{n,r}$ (with $0 \leq r \leq n-1$) consisting of all complex $n \times n$ matrices of the form $\left(\begin{smallmatrix} w & 0 \\ 0 & I_{n-r} \end{smallmatrix}\right)$ such that $w \in D_r$. The normalizer $P_{n,r}$ in $G$ of the boundary component $D_{n,r}$ is the maximal parabolic subgroup consisting of matrices $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ such that (cf. [**Kl**] §5)

$$A = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}, \; C = \begin{pmatrix} * & 0 \\ 0 & 0 \end{pmatrix}, \; D = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}.$$

(The upper left block has size $r \times r$ in each of these.)

Each maximal parabolic subgroup $P$ of $\mathbf{Sp}(2n, \mathbb{R})$ is the normalizer of an isotropic subspace $E \subset \mathbb{R}^n$ (meaning that the symplectic form vanishes on $E$). If a symplectic group element preserves $E$ then it also preserves the symplectic orthogonal subspace $E^{\perp} \supseteq E$ (which is co-isotropic, meaning that the induced symplectic form vanishes on $\mathbb{R}^{2n}/E^{\perp}$). So $P$ may also be described as the normalizer of the isotropic-co-isotropic flag $E \subset E^{\perp}$. In the case of $P_{n,r}$,

$$E = (\mathbb{R}^{n-r} \times 0_r) \times 0_n \subset \mathbb{R}^n \times \mathbb{R}^n, \quad \text{and } E^{\perp} = \mathbb{R}^n \times (\mathbb{R}^r \times 0_{n-r}) \subset \mathbb{R}^n \times \mathbb{R}^n.$$

To make these matrices look more familiar, reverse the numbering of the coordinates in the first copy of $\mathbb{R}^n$. Then the symplectic form becomes

$$J' = \begin{pmatrix} 0 & \alpha \\ -\alpha & 0 \end{pmatrix}$$

where $\alpha$ is the anti-diagonal matrix of ones. In these coordinates, the parabolic subgroup $P_{n,r}$ consists of all matrices preserving $J'$ of the following form:

$$\left(\begin{array}{c|cc|c} *_{n-r} & * & * & * \\ \hline 0 & *_r & *_r & * \\ 0 & *_r & *_r & * \\ \hline 0 & 0 & 0 & *_{n-r} \end{array}\right)$$

(where $*_t$ denotes a square $t \times t$ matrix). As such, it has a Levi decomposition $P_{n,r} = LU$ with

$$L = \left(\begin{array}{c|c|c} A & 0 & 0 \\ \hline 0 & B & 0 \\ \hline 0 & 0 & A' \end{array}\right), \quad U = \left(\begin{array}{c|c|c} I_{n-r} & * & * \\ \hline 0 & I_r & * \\ \hline 0 & 0 & I_{n-r} \end{array}\right)$$

where $A \in \mathbf{GL}(n-r, \mathbb{R})$, $A' = \alpha\, {}^tA^{-1}\alpha^{-1}$, where $B \in \mathbf{Sp}(2r, \mathbb{R})$. The center, $Z_U$ of $U$ is

$$Z_U = \left(\begin{array}{c|c|c} I_{n-r} & 0 & C \\ \hline 0 & I_r & 0 \\ \hline 0 & 0 & I_{n-r} \end{array}\right),$$

which is easily seen to be isomorphic to the vector space $S_{n-r}(\mathbb{R})$ of symmetric matrices. So $L$ splits as a direct product of a Hermitian factor $L_{Ph} = \mathbf{Sp}(2r, \mathbb{R})$ and a "linear" factor $L_{P\ell} = \mathbf{GL}(n-r, \mathbb{R})$. The Dynkin diagrams for these factors are obtained from the Dynkin diagram for $\mathbf{G}$ by deleting the node $\alpha$ corresponding to the maximal parabolic subgroup $P$ as illustrated in Figure 2.

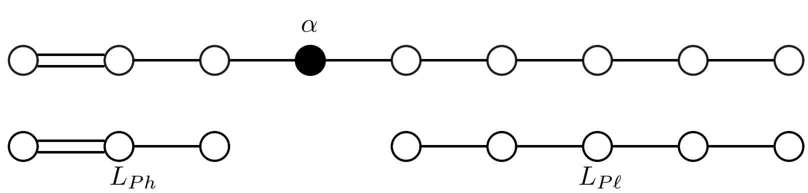



FIGURE 3. Dynkin diagrams for $L_{Ph}$ and $L_{P\ell}$

It is not too difficult to verify that the standard parabolic group $P$ acts on the boundary component $D_{n,r}$ via the first factor $\mathbf{Sp}(2r, \mathbb{R})$, in analogy with the situation in equation (2.2.2). Observe also that the second factor $\mathbf{GL}(n-r, \mathbb{R})$ acts linearly on the Lie algebra $\mathfrak{z} \cong S_{n-r}(\mathbb{R})$ of the center $Z(U_P)$ of the unipotent radical of $P$, by the action (2.2.1). This action preserves the self-adjoint homogeneous cone $\mathcal{P}_{n-r} \subset \mathfrak{z}$ described in §2.2. If $r' < r$ then $D_{n,r'} \subset \overline{D}_{n,r}$ and $\mathcal{P}_{n-r} \subset \overline{\mathcal{P}}_{n-r'}$.

**2.6. Stratifications.** A subset $S$ of a locally compact Hausdorff space $Y$ is locally closed iff it is the intersection of an open set and a closed set. A *manifold decomposition* of a locally compact Hausdorff space $Y$ is a decomposition $Y = \coprod_\alpha S_\alpha$ of $Y$ into locally finitely many locally closed smooth manifolds $S_\alpha$ (called strata), which satisfies the axiom of the frontier: the closure of each stratum is a union of strata. In this case the open cone

$$c^0(Y) = Y \times [0,1)/(y,0) \sim (y',0) \text{ for all } y, y' \in Y$$

may be decomposed with strata $S_\alpha \times (0,1)$ and the cone point $*$. Stratified sets are defined inductively. Every smooth manifold is stratified with a single stratum. Let $B^s$ denote the open unit ball in $\mathbb{R}^s$.

2.2. Definition. A manifold decomposition $Y = \coprod_\alpha S_\alpha$ of a locally compact Hausdorff space $Y$ is a *stratification* if for each stratum $S_\alpha$ there exists a compact stratified space $L_\alpha$, and for each point $x \in S_\alpha$ there exists an open neighborhood $V_x \subset Y$ of $x$ and a stratum preserving homeomorphism

$$V_x \cong B^s \times c^0(L_\alpha) \tag{2.6.1}$$

(where $s = \dim(S_\alpha)$ which is smooth on each stratum, which takes $x$ to $0 \times \{*\}$ and which takes $V_x \cap S_\alpha$ to $B^s \times \{*\}$. Such a neighborhood $V_x$ is a *distinguished* neighborhood of $x$.

The space $L_\alpha$ is called the *link* of the stratum $S_\alpha$. A stratification $Y = \coprod_\alpha S_\alpha$ is *regular* if the local trivializations (2.6.1) fit together to make a bundle over $S_\alpha$. (We omit the few paragraphs that it takes in order to make this precise since we will not have to make use of regularity.) There are many other possible “regularity” conditions on stratified sets, but all the useful ones (such as the Whitney conditions) imply the local triviality (2.6.1) of the stratification.

**2.7. Singularities of the Baily-Borel compactification.** Returning to the general case, suppose $\mathbf{G}$ is a semi-simple algebraic group defined over $\mathbb{Q}$, of Hermitian type, meaning that the symmetric space $D = G/K$ is Hermitian. Let $F$ be a rational boundary component with normalizing parabolic subgroup $P$. Let $U_P$ be the unipotent radical of $P$ and $L_P = P/U_P$ the Levi quotient. There is ([**BS**]) a unique lift $L_P \to P$ of the Levi quotient which is stable under the Cartan involution corresponding to $K$. The group $L_P$ decomposes as an almost direct product (meaning a commuting product with finite intersection), $L_P = L_{Ph}L_{P\ell}$ into factors of Hermitian and “linear” type[3] with $A_P \subset L_{P\ell}$. Here, “linear” means that the symmetric space $L_{P\ell}/K_\ell$ for $L_{P\ell}$ is a self-adjoint homogeneous cone $C_P$, which is open in some real vector space $V$ (in this case, $V = \mathrm{Lie}(Z(U_P))$) on which $L_{P\ell}$ acts by linear transformations which preserve $C_P$. The group $P$ acts on $F$ through $L_{Ph}$, identifying $F$ with the symmetric space for $L_{Ph}$. There is a diffeomorphism $D = P/K_P \cong \mathcal{U}_P \times F \times C_P$.

2.1. Lemma. ([**AMRT**] *§4.4*) *Let* $\mathbf{P} \neq \mathbf{P}'$ *be standard rational parabolic subgroups, normalizing the standard boundary components* $F \neq F'$ *respectively. Then the following statements are equivalent, in which case we write* $\mathbf{P}' \prec \mathbf{P}$ :

(1) $L_{P'h} \subset L_{Ph}$
(2) $L_{P\ell} \subset L_{P'\ell}$

[3] It is possible to absorb the compact factors of $L_P$, if there are any, into $L_{Ph}$ and $L_{P\ell}$ in such a way that both $L_{Ph}, L_{P\ell}$ are defined over $\mathbb{Q}$.

(3) $Z(U_P) \subset Z(U_{P'})$
(4) *$F'$ is a rational boundary component of $F$*
(5) *The cone $C_P$ is a rational boundary component of $C_{P'}$.*

Suppose $\mathbf{P}$ is a rational parabolic subgroup of $\mathbf{G}$ such that $P = \mathbf{P}(\mathbb{R})$ normalizes a rational boundary component $F$. Let $L_P = L_{Ph}L_{P\ell}$ be the almost direct product decomposition of its Levi component as discussed above. So we obtain identifications $D = P/K_P$, $F = L_{Ph}/K_h$, and $C_P = L_{P\ell}/K_\ell$ for appropriate maximal compact subgroups $K_P = K \cap P \subset L_P \subset P$, $K_h \subset L_{Ph}$, and $K_\ell \subset L_{P\ell}$.

Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be a neat ([**B1**]) arithmetic subgroup. Set $\Gamma_P = \Gamma \cap P$ and $\Gamma_U = \Gamma \cap U_P$. Then

$$N_P = \Gamma_U \backslash U_P$$

is a compact "nilmanifold" whose fundamental group is the nilpotent group $\Gamma_U$. Let $\Gamma_L = \Gamma_P/\Gamma_U \subset L_P$ and set $\Gamma_\ell = \Gamma_L \cap L_{P\ell}$. Let $\Gamma_h \subset L_{Ph}$ be the projection of $\Gamma_L$ to the Hermitian factor $L_{Ph}$. We obtain an identification between the boundary stratum $X_F$ and the quotient $\Gamma_h \backslash F = \Gamma_h \backslash L_{Ph}/K_h$ (cf. §2.4).

It follows that $\Gamma_P \backslash D = \Gamma_P \backslash P/K_P$ fibers over the locally symmetric space $X_P = \Gamma_L \backslash L_P/K_P$ (cf. equation (5.1.2)) with fiber $N_P$; and that $X_P$ in turn fibers over the boundary stratum $X_F$ with fiber $\Gamma_\ell \backslash C_P$.

2.2. Theorem. ([**BB**]) *The boundary strata of the Baily-Borel compactification form a regular stratification of $\overline{X}^{BB}$. Let $X_F$ be such a stratum, corresponding to the $\Gamma$-conjugacy class of the rational parabolic subgroup $\mathbf{P} \subset \mathbf{G}$. Then there exists a parabolic neighborhood (see §4.5) $V_F$ of $X_F$ whose intersection $V_F \cap X$ with $X$ is diffeomorphic to the quotient $\Gamma_P \backslash D$. Hence the geodesic projection $\pi_F : V_F \cap X \to X_F$ is a smooth fiber bundle with a fiber $W$, which is itself a fiber bundle, $W \to \Gamma_\ell \backslash C_P$ with fiber diffeomorphic to the compact nilmanifold $N_P = \Gamma_U \backslash U_P$. If $x \in X_F$ and if $B_x \subset X_F$ is a sufficiently small ball in $X_F$, containing $x$, then the pre-image $\pi_F^{-1}(B_x) \subset V_F$ is a distinguished neighborhood of $x$ in $\overline{X}^{BB}$, whose intersection with $X$ is therefore homeomorphic to the product $B_x \times W$. The closure of the stratum $X_F$ is the Baily-Borel compactification of $X_F$. It consists of the union of all strata $X_{F'}$ such that the normalizing parabolic subgroup $\mathbf{P}'$ is $\Gamma$-conjugate to some $\mathbf{Q} \prec \mathbf{P}$.*

Despite its precision, this result does not fully describe the topology of the neighborhood $V_F$; only that of $V_F \cap X$. Moreover it does not describe the manner in which such neighborhoods for different strata are glued together. A complete (but cumbersome) description of the local structure of the Baily-Borel compactification exists, but it is sometimes more useful to describe various sorts of "resolutions" of $\overline{X}^{\mathrm{BB}}$.

## 3. Toroidal compactifications and automorphic vector bundles

**3.1.** The toroidal compactification is quite complicated and we will not attempt to provide a complete description here. (The standard reference is [**AMRT**]. An excellent introduction appears in the book [**Nk**], but it takes many pages. Brief summaries are described in [**GT2**] §7.5 and [**GP**] §14.5.) Instead we will list some of its main features. As in the preceding section we suppose that $X = \Gamma \backslash G/K$ is a Hermitian locally symmetric space arising from a semisimple algebraic group $\mathbf{G}$ defined over $\mathbb{Q}$. There are many toroidal compactifications. Each depends on a certain choice $\Sigma$ of combinatorial data, and we sometimes indicate this by writing $\overline{X}_\Sigma^{\text{tor}}$ for this compactification. Each $\overline{X}_\Sigma^{\text{tor}}$ admits the structure of a complex algebraic space. For certain "good" choices of $\Sigma$ the resulting variety is projective and nonsingular, and the complement $\overline{X}_\Sigma^{\text{tor}} - X$ is a divisor with normal crossings, which is therefore stratified by the multi-intersections of the divisors. The identity extends to a unique continuous mapping $\kappa_\Sigma : \overline{X}^{\text{tor}} \to \overline{X}^{\text{BB}}$. It is a holomorphic morphism which takes strata to strata. If $\Sigma$ is "good" in the above sense then $\kappa_\Sigma$ is a resolution of singularities.

**3.2. The cone again.** In this section we describe the combinatorial data $\Sigma$ which determines the choice of toroidal compactification $\overline{X}_\Sigma^{\text{tor}}$ in the case that $\mathbf{G}$ is $\mathbb{Q}$-simple. In this case there is a natural ordering among the standard proper rational parabolic subgroups, with $\mathbf{P} \prec \mathbf{Q}$ if $L_{Ph} \subset L_{Qh}$, or equivalently (see Lemma 2.1), if $L_{Q\ell} \subset L_{P\ell}$. Let $\mathbf{P}$ be the standard maximal rational parabolic subgroup which comes first in this total ordering and let $C_P \cong L_{P\ell}/K_{P\ell}$ be the corresponding self adjoint homogeneous cone, with its partial compactification $\overline{C}_P^{\text{std}}$, as in §2.2. It is contained in $Z(U_P)$ which we identify with $\mathfrak{z} = \text{Lie}(Z(U_P))$ by the exponential map, and it is rational with respect to the lattice $\Lambda = \Gamma \cap Z(U_P)$. Choose a $\Gamma_{P\ell}$-invariant rational simplicial cone decomposition of $\overline{C}_P^{\text{std}}$, or equivalently, a rational flat triangulation of the compact (singular) space $\Gamma_{P\ell}\backslash\overline{C}_P^{\text{std}}$/homotheties, which is subordinate to the stratification by boundary strata. (This means that the closure of each stratum should be a subcomplex.)

Up to $\Gamma$ conjugacy, there are finitely many maximal rational parabolic subgroups $P$ that are minimal with respect to the ordering $\prec$ . The data $\Sigma$ refers to a choice of cone decomposition of $\overline{C}_P^{\text{std}}$ for each of these, which are compatible in the sense that if $C_Q \subset \overline{C}_P^{\text{std}} \cap \overline{C}_{P'}^{\text{std}}$ then the two resulting cone decompositions of $C_Q$ coincide.

By the theory of torus embeddings, such a cone decomposition of $C_P$ determines a $\Gamma_{P\ell}$ equivariant partial compactification of the algebraic torus $(\mathfrak{z} \otimes \mathbb{C})/\Lambda$, which is one of the key ingredients in the construction of the toroidal compactification. The rest of the construction, which is rather complicated, consists of "attaching" the resulting torus embedding to $X$.

Actually, this data only determines a resolution $\overline{X}_\Sigma^{\text{tor}}$ which is "rationally nonsingular" (has finite quotient singularities). A truly nonsingular compactification is obtained when we place a further integrality condition on the cone decomposition of $\overline{C}_P^{\text{std}}$, namely that the shortest vectors in the 1-dimensional cones in any top dimensional simplicial cone should form an integral basis of the lattice $\Lambda$. There is a further (convexity) criterion on the cone decompositions to guarantee that the resulting $\overline{X}_\Sigma^{\text{tor}}$ is projective. Cone decompositions satisfying these additional conditions exist, although the literature is a little sketchy on this point. A more difficult problem is to find (canonical) models for $\overline{X}_\Sigma^{\text{tor}}$ defined over a number field, or possibly over the reflex field, when $X$ is a Shimura variety. See, for example [**FC**].

**3.3. Automorphic vector bundles.** Let $\lambda : K \to \mathbf{GL}(E)$ be a representation of $K$ on some complex vector space $E$. Then we obtain a homogeneous vector bundle $\mathbf{E} = G \times_K E$ on $D$, meaning that we identify $(h, e)$ with $(hk, \lambda(k)^{-1}e)$ whenever $k \in K$ and $h \in G$. Denote the equivalence class of such a pair by $[h, e]$. The action of $G$ on $D$ is covered by an action of $G$ on $\mathbf{E}$ which is given by $g \cdot [h, e] = [gh, e]$. So dividing by $\Gamma$ we obtain a *automorphic vector bundle* $\mathbf{E}_\Gamma = \Gamma \backslash \mathbf{E} \to X$, which may also be described as $\mathbf{E}_\Gamma = (\Gamma \backslash G) \times_K E$. Such a vector bundle carries a canonical connection. If the representation $\lambda$ is the restriction to $K$ of a representation of $G$, then $E_\Gamma$ also carries a (different) flat connection (cf. [**GP**] §5).

Smooth sections of $\mathbf{E}_\Gamma$ may be identified (see also §3.5 below) with smooth mappings $f : G \to E$ such that $f(\gamma g k) = \lambda(k^{-1}) f(g)$ (for all $k \in K, \gamma \in \Gamma$). The *holomorphic* sections of $\mathbf{E}_\Gamma$ correspond to those functions that are killed by certain differential operators, as observed in [**B3**]. The complexified Lie algebra of $G$ decomposes under the Cartan involution into $+1, +i$, and $-i$ eigenspaces,

$$\mathfrak{g}(\mathbb{C}) = \mathfrak{g} \otimes \mathbb{C} = \mathfrak{k}(\mathbb{C}) \oplus \mathfrak{p}^+ \oplus \mathfrak{p}^-$$

respectively (where $\mathfrak{k}(\mathbb{C})$ is the complexification of $\mathrm{Lie}(K)$). Each $V = X + iY \in \mathfrak{g}(\mathbb{C})$ acts on functions $f : G \to \mathbb{C}$ by $V(f) = X(f) + iY(f)$. Then a smooth section of $\mathbf{E}_\Gamma$ is holomorphic if and only if the corresponding function $f : G \to \mathbb{E}$ satisfies the Cauchy-Riemann equations: $V(f) = 0$ for all $V \in \mathfrak{p}^-$. Let us further say ([**B3**]) that such a holomorphic section $f$ is a *holomorphic automorphic form* if it has *polynomial growth*, that is, if there exists $C > 0$ and $n \geq 1$ such that $|f(g)| \leq C ||g||_G^n$. (Here, $||g||_G$ is the norm $||g||_G = tr(Ad(\theta(g^{-1})) \cdot Ad(g))$, where $\theta$ is the Cartan involution.)

One might hope to interpret this condition in terms of the Baily Borel compactification $\overline{X}^{\mathrm{BB}}$. However the automorphic vector bundle $\mathbf{E}_\Gamma \to X$ does not necessarily extend to the Baily-Borel compactification. It does extend to $\overline{X}^{\mathrm{RBS}}$, but only as a topological vector bundle. However, in [**M**], Mumford constructs a canonical extension $\mathbf{E}_\Sigma \to \overline{X}_\Sigma^{\mathrm{tor}}$ as a holomorphic vector bundle, and shows that the (global holomorphic) sections of $\mathbf{E}_\Sigma$ are precisely the holomorphic sections of $\mathbf{E}_\Gamma \to X$ with polynomial growth, that is, they are holomorphic automorphic forms.

**3.4. Proportionality theorem.** In [**M**], Mumford proved that the Chern classes $c^i(\mathbf{E}_\Sigma) \in H^{2i}(\overline{X}_\Sigma^{\mathrm{tor}})$ of the bundle $\mathbf{E}_\Sigma$ satisfy Hirzebruch's proportionality theorem: there exists a single rational number $v(\Gamma)$ so that for any automorphic vector bundle $\mathbf{E}_\Gamma$ on $X$, for any toroidal compactification $\overline{X}_\Sigma^{\mathrm{tor}}$, and for any partition $I : n_1 + n_2 + \cdots + n_k = 2n$ where $n = \dim_{\mathbb{C}}(X)$, the corresponding Chern number of the canonical extension $\mathbf{E}_\Sigma$

$$c^I(\mathbf{E}_\Sigma) = c^{n_1}(\mathbf{E}_\Sigma) \cup c^{n_2}(\mathbf{E}_\Sigma) \cup \cdots \cup c^{n_k}(\mathbf{E}_\Sigma) \cap [\overline{X}_\Sigma^{\mathrm{tor}}] \in \mathbb{Q}$$

(where $[\overline{X}_\Sigma^{\mathrm{tor}}] \in H_{2n}(\overline{X}_\Sigma^{\mathrm{tor}})$ is the fundamental class) satisfies

$$c^I(\mathbf{E}_\Sigma) = v(\Gamma) c^I(\check{\mathbf{E}})$$

where $\check{\mathbf{E}}$ is the corresponding vector bundle on the compact dual symmetric space, $\check{D}$. The fact that these Chern numbers are independent of the resolution $\overline{X}_\Sigma^{\mathrm{tor}}$ suggests that they might be related to the topology of $\overline{X}^{\mathrm{BB}}$. This possibility was realized in [**GP**] where it was shown that for any automorphic vector bundle $\mathbf{E}_\Gamma \to X$, each Chern class $c^k(\mathbf{E}_\Gamma) \in H^{2k}(X; \mathbb{C})$ has a particular lift to cohomology $\overline{c}^k(\mathbf{E}_\Gamma) \in H^{2k}(\overline{X}^{\mathrm{BB}}; \mathbb{C})$ such that for any toroidal resolution $\kappa_\Sigma : \overline{X}_\Sigma^{\mathrm{tor}} \to \overline{X}^{\mathrm{BB}}$ the lift satisfies $\kappa_\Sigma^*(\overline{c}^k(\mathbf{E}_\Gamma)) = c^k(\mathbf{E}_\Sigma)$. Therefore the proportionality formula holds for these lifts $\overline{c}^k(\mathbf{E}_\Sigma)$ as well. In many cases this accounts for sufficiently many cohomology classes to prove that the cohomology $H^*(\check{D},\mathbb{C})$ of the compact dual symmetric space is contained in the cohomology $H^*(\overline{X}^{\mathrm{BB}},\mathbb{C})$ of the Baily-Borel compactification.

**3.5. Automorphy factors.** There is a further (and more classical) description of the sections of an automorphic vector bundle $\mathbf{E}_{\mathbf{\Gamma}} = \Gamma\backslash(G \times_K E)$ corresponding to a representation $\lambda : K \to \mathbf{GL}(E)$. A (smooth) *automorphy factor* $J : G \times D \to \mathbf{GL}(E)$ for $\mathbf{E}$ is a (smooth) mapping such that

(1) $J(gg', x) = J(g, g'x)J(g', x)$ for all $g, g' \in G$ and $x \in D$
(2) $J(k, x_0) = \lambda(k)$ for all $k \in K$.

It follows (by taking $g = 1$) that $J(1, x) = I$. The automorphy factor $J$ is determined by its values $J(g, x_0)$ at the basepoint: any smooth mapping $j : G \to \mathbf{GL}(E)$ such that $j(gk) = j(g)\lambda(k)$ (for all $k \in K$ and $g \in G$) extends in a unique way to an automorphy factor $J : G \times D \to \mathbf{GL}(E)$ by setting $J(g, hx_0) = j(gh)j(h)^{-1}$.

An automorphy factor $J$ determines a (smooth) trivialization

$$\Phi_J : G \times_K E \to (G/K) \times E$$

by $[g, v] \mapsto (gK, J(g, x_0)v)$. With respect to this trivialization the action of $\gamma \in G$ is given by

$$\gamma \cdot (x, v) = (\gamma x, J(\gamma, x)v). \tag{3.5.1}$$

Conversely any smooth trivialization $\Phi : \mathbf{E} \cong (G/K) \times E$ of $\mathbf{E}$ determines a unique automorphy factor $J$ such that $\Phi = \Phi_J$. Such a trivialization allows one to identify smooth sections $s$ of $\mathbf{E}$ with smooth mappings $r : D \to E$. If the section $s$ is given by a smooth mapping $s : G \to E$ such that $s(gk) = \lambda(k^{-1})s(g)$ then the corresponding mapping $r$ is $r(gK) = J(g, x_0)s(g)$ (which is easily seen to be well defined). By (3.5.1), sections $s$ which are invariant under $\gamma \in \Gamma \subset \mathbf{G}(\mathbb{Q})$ then correspond to functions $r : D \to E$ which satisfy the familiar relation

$$r(\gamma x) = J(\gamma, x)r(x) \tag{3.5.2}$$

for all $x \in D$. Moreover, there exists a *canonical automorphy factor* ([**Sa1**] II §5),

$$J_0 : G \times D \to \mathbf{K}(\mathbb{C})$$

which determines an automorphy factor $J = \lambda_{\mathbb{C}} \circ J_0$ for every homogeneous vector bundle $\mathbf{E} = G \times_K E$, where $\lambda_{\mathbb{C}} : \mathbf{K}(\mathbb{C}) \to \mathbf{GL}(E)$ denotes the complexification of $\lambda$. With this choice for $J$, *holomorphic* sections $s$ of $\mathbf{E}_{\mathbf{\Gamma}}$ correspond to *holomorphic* functions $r : D \to E$ which satisfy (3.5.2).

## 4. Borel-Serre compactification

**4.1. About the center, and other messy issues.** In this section and in the remainder of this article, $\mathbf{G}$ will be a connected reductive algebraic group defined over $\mathbb{Q}$; $K \subset G$ will be a chosen maximal compact subgroup and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ will be an arithmetic group.

The identity component (in the sense of algebraic groups) of the center of $\mathbf{G}$ is an algebraic torus defined over $\mathbb{Q}$. It has three parts: a greatest $\mathbb{Q}$-split subtorus $\mathbf{A_G}$, an $\mathbb{R}$-split but $\mathbb{Q}$-anisotropic part $\mathbf{A^1_G}$, and an $\mathbb{R}$-anisotropic (i.e. compact) part, $\mathbf{A^2_G}$. Unfortunately it is not simply the direct product of these three parts, however we can at least isolate the group $A_G = \mathbf{A_G}(\mathbb{R})^+$, the topologically connected identity component of the group of real points of $\mathbf{A_G}$. Define ([**BS**] §1.1)

$$^0\mathbf{G} = \bigcap_{\chi} \ker(\chi^2)$$

to be the intersection of the kernels of the squares of the rationally defined characters $\chi : \mathbf{G} \to \mathbf{G_m}$. It is a connected reductive linear algebraic group defined over $\mathbb{Q}$ which contains every compact subgroup and every arithmetic subgroup of $G$ ([**BS**] §1.2). The group of real points $G = \mathbf{G}(\mathbb{R})$ decomposes as a direct product, $G = {}^0\mathbf{G}(\mathbb{R}) \times A_G$. Then $D = G/KA_G = {}^0G/K$. So, to study the topology and geometry of $D$ (and its arithmetic quotients) one may assume that the group $\mathbf{G}$ contains no nontrivial $\mathbb{Q}$-split torus in its center. This is not a good assumption to make from the point of view of representation theory or from the point of view of Shimura varieties since in these cases the center $\mathbf{A_G}$ plays an important role. Nevertheless we will occasionally make this assumption when it simplifies the exposition.

The part $A^1_G$ of the center contributes a Euclidean factor to the symmetric space $D$. However, after dividing by $\Gamma$ this Euclidean space will get rolled up into circles, which explains why it does not interfere with our efforts to compactify $\Gamma\backslash D$. Even if $\mathbf{A_G}$ and $\mathbf{A^1_G}$ are trivial, the group $G$ may still contain a compact torus in its center, but this will be contained in any maximal compact subgroup $K \subset G$ so it will not appear in the symmetric space $G/K$.

We will also assume, for simplicity, that $\Gamma$ is torsion free, which implies that $\Gamma$ acts freely on $D$ and that the quotient $X$ is a smooth manifold. It is often convenient to make the slightly stronger assumption that $\Gamma$ is *neat* ([**B1**]), which implies ([**AMRT**] p. 276) that $(\Gamma \cap H_2(\mathbb{C}))/(\Gamma \cap H_1(\mathbb{C}))$ is torsion-free whenever $\mathbf{H_1} \triangleleft \mathbf{H_2} \subset \mathbf{G}$ are rationally defined algebraic subgroups. This guarantees that all the boundary strata are smooth manifolds also. Every arithmetic group contains neat arithmetic subgroups of finite index, however much of what follows will continue to hold even when $\Gamma$ has torsion.

**4.2.** The Borel-Serre compactification $\overline{X}^{\mathrm{BS}}$ is (topologically) a smooth manifold (of some dimension $m$) with boundary. However the boundary has the differentiable structure of "corners": it is decomposed into a collection of smooth manifolds of various dimensions, and a point on one of these boundary manifolds of dimension $d$ has a neighborhood which is diffeomorphic to the product $B^d \times [0, 1)^{m-d}$ where $B^d$ is the open unit ball in $\mathbb{R}^d$. This compactification is obtained as the quotient under $\Gamma$ of a "partial" compactification $\overline{D}^{\mathrm{BS}}$ which is obtained from $D$ by attaching a "boundary component" for each proper rational parabolic subgroup $\mathbf{P} \subset \mathbf{G}$.

**4.3. Geodesic action.** Let $P$ be the group of real points of a rational parabolic subgroup $\mathbf{P}$. Let $U_P$ be its unipotent radical and $\nu : P \to L_P$ be the projection to the Levi quotient. Then $L_P$ is the group of real points of a rationally defined reductive group $\mathbf{L_P}$ and as such, we have $L_P = M_P A_P$ where $M_P = {}^0\mathbf{L_P}(\mathbb{R})$ as in §4.1. The choice of $K \subset G$ corresponds to a Cartan involution $\theta : G \to G$ and there is a unique $\theta$ stable lift ([**BS**]) of $L_P$ to $P$. So we obtain the *Langlands decomposition*

$$P = U_P A_P M_P. \tag{4.3.1}$$

The intersection $K_P = K \cap P$ is completely contained in $M_P$. It follows from the Iwasawa decomposition that $P$ acts transitively on $D$. Define the *right* action of

$A_P$ on $D = P/K_P$ by $(gK_P) \cdot a = gaK_P$ for $g \in P$ and $a \in A_P$. This action is well defined since $A_P$ commutes with $K_P \subset M_P$, but it also turns out to be independent of the choice of basepoint. Moreover each orbit of this $A_P$ action is a totally geodesic submanifold of $D$ (with respect to any invariant Riemannian metric). Define the (Borel-Serre) boundary component $e_P = D/A_P$.

Intuitively, we want to "attach" $e_P$ to $D$ as the set of limit points of each of these geodesic orbits. For the upper half plane $\mathfrak{h}_1$ and the standard Borel subgroup $B \subset \mathbf{SL}(2, \mathbb{R})$, if $a = \left(\begin{smallmatrix} t & 0 \\ 0 & t^{-1} \end{smallmatrix}\right)$ then the geodesic action is $(x + iy) \cdot a = x + it^2 y$ for any $t \neq 0$, so the geodesic orbits are "vertical" half lines. Then $e_P$ is a line at infinity, parallel to the real axis, which is glued onto the upper half plane so as to make a strip $\mathbb{R} \times (0, \infty]$. (See figure 3.)

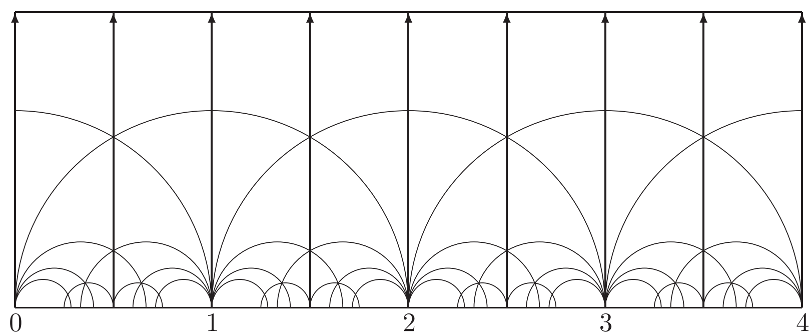



Figure 4. $\overline{\mathfrak{h}}^{\mathrm{BS}}$ and geodesic action

When $\Gamma$ "acts" on this union $D \cup e_P$, only the translations $\Gamma_P = \Gamma \cap P$ act nontrivially on the boundary component $e_P$ so the resulting circle $\Gamma_P \backslash e_P$ becomes glued to $X$ where, previously in the Baily-Borel compactification, we had placed a cusp. Unfortunately the group $\Gamma$ does not actually act on $D \cup e_P$, a difficulty which may be rectified by attaching additional boundary components $e_Q$ for every rational parabolic subgroup $\mathbf{Q}$, using a "Satake topology" in which each $e_Q$ has a neighborhood isomorphic to that of $e_P$. Although it is difficult to visualize the resulting space $\overline{D}^{\mathrm{BS}}$, it is nevertheless a (real) two dimensional manifold with boundary, whose boundary consists of countably many disjoint copies of $\mathbb{R}$.

**4.4.** In this section we return to the general case but we assume for simplicity that $\mathbf{A_G}$ is trivial. As a set, $\overline{D}^{\mathrm{BS}}$ is defined to be the disjoint union of $D$ and all the Borel-Serre boundary components $e_P$ corresponding to rational proper parabolic subgroups $\mathbf{P}$. Let $\mathbf{P_0} \subset \mathbf{G}$ be a fixed minimal rational parabolic subgroup. The parabolic subgroups containing $\mathbf{P_0}$ are the *standard* parabolics. Denote by $\mathbf{A_0}$ the greatest $\mathbb{Q}$ split torus in the center of the (canonical lift of the) Levi component $L_0 = L_{P_0}$ and let $\Phi = \Phi(\mathbf{G}, \mathbf{A_0})$ be the corresponding system of rational roots, with simple rational roots $\Delta$. For any standard parabolic subgroup $\mathbf{P}$ let $\Delta_P$ be the set of restrictions of the roots in $\Delta$ to $A_P \subset A_0$. If $\Delta_P = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ the isomorphism

$$A_P \cong (0, \infty)^r \quad \text{given by } t \mapsto (\alpha_1(t), \alpha_2(t), \ldots, \alpha_r(t))$$

determines a partial compactification $\overline{A}_P \cong (0, \infty]^r$. Let

$$D(P) = D \times_{A_P} \overline{A}_P$$

where $A_P$ acts on $D$ by the geodesic action. Then $D(P)$ contains $D$ and it also contains $e_P$ as the set of points (or, rather, equivalence classes) $[x, (\infty, \infty, \ldots, \infty)]$. The projection $D \to e_P$ extends continuously to $\pi_P : D(P) \to e_P$ which we refer to as the *geodesic projection*.

It is a bit easier to picture this construction in terms of coordinates. For each $\alpha \in \Delta_P$ define the *root function* $f_\alpha^P : D \to (0, \infty)$ by

$$f_\alpha^P(x) = f_\alpha^P(gK_P) = f_\alpha^P(uamK_P) = \alpha(a)$$

where $x = gK_P$ and $g = uam \in P$ has been decomposed according to the Langlands' decomposition (4.3.1). The root function is well defined because the mapping $P \to A_P$ given by $uam \mapsto a$ is a group homomorphism. If $g' = u'a'm' \in P$ and if $b \in A_P$ then

$$f_\alpha^P(g'x \cdot b) = \alpha(a'b) f_\alpha^P(x). \tag{4.4.1}$$

The root functions clearly extend to $D(P)$ and together with $\pi_P$ they determine a diffeomorphism

$$D(P) \cong e_P \times (0, \infty]^r. \tag{4.4.2}$$

If $\mathbf{P} \subset \mathbf{Q}$ then $A_Q \subset A_P$ and the restriction of the geodesic action for $A_P$ to $A_Q$ coincides with the geodesic action for $A_Q$. Therefore there is a natural inclusion $\overline{A_Q} \subset \overline{A_P}$ as a coordinate subspace, and we see that $D(P)$ is the disjoint union of boundary components $e_Q$ for $\mathbf{Q} \supseteq \mathbf{P}$ (including $e_G = D$.) We wish to declare this set $D(P)$ to be an open neighborhood of $e_P$ in $\overline{D}^{\mathrm{BS}}$.

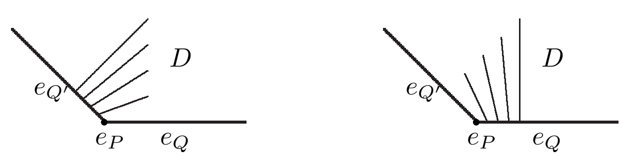



FIGURE 5. $D(P)$ and level curves of $f_\alpha^P$ for $\alpha_1, \alpha_2 \in \Delta_P$.

The following theorem says that it is possible to similarly attach boundary components $e_P$ for any rational parabolic subgroup $\mathbf{P}$, so as to obtain a partial compactification $\overline{D}^{\mathrm{BS}}$ of $D$.

4.1. THEOREM. ([**BS**] §*7.1*) *There is a unique topology* (*the Satake topology*) *on the union* $\overline{D}^{BS}$ *of* $D$ *with all its rational boundary components, so that the action of* $\mathbf{G}(\mathbb{Q})$ *on* $D$ *extends continuously to an action by homeomorphisms on* $\overline{D}^{BS}$, *and* *so that each $D(P) \subset \overline{D}^{BS}$ is open. The parabolic subgroup $P$ is the normalizer of the boundary component $e_P$. The closure $\overline{e}_P$ of $e_P$ in $\overline{D}^{BS}$ is the Borel-Serre partial compactification of $e_P$.*

(An annoying problem arises because $e_P$ is not a symmetric space, and in fact it is a homogeneous space under the non-reductive group $P$. In order to apply inductive arguments, Borel and Serre found it necessary to work within a wider class of groups and homogeneous spaces which include $P$ and $e_P$. Fortunately the current context provides the author with a poetic license to ignore these further complications.)

**4.5. Quotient under $\Gamma$.** Fix a neat arithmetic group $\Gamma \subset \mathbf{G}(\mathbb{Q})$. Let $\kappa : \overline{D}^{\mathrm{BS}} \to \overline{X}^{\mathrm{BS}} = \Gamma\backslash\overline{D}^{\mathrm{BS}}$ be the quotient. The action of $\Gamma$ on $\overline{D}^{\mathrm{BS}}$ will identify some boundary components and it will also make identifications within a single boundary component. There is a risk that this will completely destroy the local picture $D(P)$ of $e_P$ which was developed above. It is a remarkable fact that this risk never materializes. To be precise, there is a neighborhood $V$ of $e_P$ in $\overline{D}^{\mathrm{BS}}$ so that

(P1) Two points in $V$ are identified under $\Gamma$ if and only if they are identified under $\Gamma_P = \Gamma \cap P$.

(P2) The neighborhood $V$ is preserved by the geodesic action of $A_P(\geq 1)$.

Here, $A_P(\geq 1) = \{a \in A_P | \alpha(a) \geq 1 \text{ for all } \alpha \in \Delta_P\}$ is the part of $A_P$ that moves points in $D$ "towards the boundary." Such a neighborhood $V$ is called a $\Gamma$-*parabolic neighborhood* and we will also refer to its image $\kappa(V) \subset \overline{X}^{\mathrm{BS}}$ as a parabolic neighborhood. The intersection $\kappa(V) \cap X$ is diffeomorphic to the quotient $\Gamma_P\backslash D$.

There is another way to say this. Let $V'$ be the image of $V$ in $\Gamma_P\backslash\overline{D}^{\mathrm{BS}}$. Since $\Gamma_P \subset \Gamma$ we have a covering $\beta : \Gamma_P\backslash\overline{D}^{\mathrm{BS}} \to \Gamma\backslash\overline{D}^{\mathrm{BS}}$. For points far away from $e_P$ this is a nontrivial covering. However for points in $V' \subset \overline{D}^{\mathrm{BS}}$ the covering $V' \to \beta(V')$ is actually one to one. This fact (a consequence of reduction theory) allows us to study a neighborhood of $e_P$ using the structure of the parabolic subgroup $P$. (In the case of the upper half plane, it is easy to see from Figure 1 that the set of points $z \in \mathbb{C}$ with $\mathrm{Re}(z) > 2$ forms such a parabolic neighborhood of the point at infinity.)

Define the *Borel-Serre stratum* $Y_P = \kappa(e_P)$ to be the image of the boundary component $e_P$. A $\Gamma$-parabolic neighborhood of $Y_P \subset \overline{X}^{\mathrm{BS}}$ is diffeomorphic to a neighborhood of $Y_P$ in $\Gamma_P\backslash D(P)$. If $\gamma \in \Gamma_P$ then for any $\alpha \in \Delta_P$ and any $x \in D(P)$ we have: $f_\alpha^P(\gamma x) = f_\alpha^P(x)$. This follows from (4.4.1) and the fact that the projection $P \to A_P$ kills $\Gamma_P$. Therefore the diffeomorphism (4.4.2) passes to a diffeomorphism

$$\Gamma_P\backslash D(P) \cong \Gamma_P\backslash e_P \times (0, \infty]^r$$

which says that the stratum $Y_P$ has a neighborhood in $\overline{X}^{BS}$ which is a manifold with corners. As described above, these corners fit together: if $\mathbf{P} \subset \mathbf{Q}$ then the inclusion $e_Q \subset D(P)$ induces an mapping $Y_P \times (0, \infty)^s \to Y_Q$ (for an appropriate coordinate subspace $(0, \infty)^s$), which is one to one near $Y_P$. (Once we leave the parabolic neighborhood of $Y_P$ this mapping is no longer one to one.) With a bit more work one concludes the following

4.2. Theorem. ([**BS**]) *The quotient $\overline{X}^{BS} = \Gamma\backslash\overline{D}^{BS}$ is compact. It is stratified with finitely many strata $Y_P = \Gamma_P\backslash e_P$, one for each $\Gamma$-conjugacy class of rational* *parabolic subgroups* $\mathbf{P} \subset \mathbf{G}$. *Each stratum* $Y_P$ *has a parabolic neighborhood* $V$ *diffeomorphic to* $Y_P \times (0, \infty]^r$ *(where* $r$ *is the rank of* $A_P$*) whose faces* $Y_P \times (0, \infty)^s$ *are the intersections* $Y_Q \cap V$ *for appropriate* $\mathbf{Q} \supset \mathbf{P}$.

Two important applications are given in §6.1. For many purposes the Borel-Serre compactification is too big. For example, each stratum $Y_P$ is the quotient of a non-reductive group $P$ by an arithmetic subgroup $\Gamma_P$. The reductive Borel-Serre compactification (first studied in [**Z1**] §4.2, p. 190; see also [**GHM**] §8) is better behaved. It is obtained by replacing this stratum by an appropriate arithmetic quotient of the Levi component of $P$.

## 5. Reductive Borel-Serre Compactification

**5.1.** As in the previous section we suppose $\mathbf{G}$ is a reductive algebraic group defined over $\mathbb{Q}$ with associated symmetric space $D = G/KA_G$. Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup and set $X = \Gamma \backslash D$. Let $\mathbf{P}$ be a proper rational parabolic subgroup with Langlands' decomposition (4.3.1), let $\Gamma_P = \Gamma \cap P$, let $e_P = D/A_P$ be the Borel-Serre boundary component, and let $Y_P = \Gamma_P \backslash e_P$ be the Borel-Serre stratum. Let us first examine the structure of $e_P \to Y_P$.

Using the Levi decomposition $P = U_P L_P$ we may write $D = U_P L_P / K_P A_G$. The group $K_P$ and the geodesic action of the group $A_P$ act (from the right) only on the factor $L_P$. So we obtain a diffeomorphism

$$e_P \cong U_P \times (L_P/K_P A_P) = U_P \times D_P \tag{5.1.1}$$

where $D_P$ is the *reductive Borel-Serre boundary component* $L_P/K_P A_P$. In these coordinates, the action of $g \in P$ is given by

$$g.(u, zK_P A_P) = (gui\nu_P(g)^{-1}, \nu_P(g) z K_P A_P)$$

where $\nu_P : P \to L_P$ is the projection to the Levi quotient, and where $i : L_P \to P$ is its canonical splitting from §2.7. So the unipotent radical of $P$ acts only on the $U_P$ factor, while $P$ acts on the $D_P$ factor through its Levi quotient.

Define the *reductive Borel-Serre stratum*

$$X_P = \Gamma_L \backslash D_P = \Gamma_L \backslash L_P / K_P A_P \tag{5.1.2}$$

where $\Gamma_L = \nu_P(\Gamma_P) \subset L_P = M_P A_P$. Then the Borel-Serre stratum $Y_P$ is a fiber bundle over the reductive Borel-Serre stratum $X_P$,

$$Y_P = \Gamma_P \backslash e_P = \Gamma_P \backslash P / K_P A_P \to X_P = \Gamma_L \backslash D_P$$

whose fiber is the compact nilmanifold $N_P = \Gamma_U \backslash U_P$.

5.1. Definition. The reductive Borel-Serre partial compactification $\overline{D}^{\mathrm{RBS}}$ (resp. $\overline{X}^{\mathrm{RBS}}$) is the quotient of $\overline{D}^{\mathrm{BS}}$ (resp. $\overline{X}^{\mathrm{BS}}$) which is obtained by collapsing each $e_P$ to $D_P$ (resp. $Y_P$ to $X_P$).

5.1. Theorem. ([**Z1**], [**GHM**] §*8.10*) *The group* $\Gamma$ *acts on* $\overline{D}^{RBS}$ *with compact quotient* $\Gamma \backslash \overline{D}^{RBS} = \overline{X}^{RBS}$. *The boundary strata form a regular stratification of* $\overline{X}^{RBS}$ *and the stratum*

$$X_P = \Gamma_L \backslash M_P / K_P = \Gamma_L \backslash L_P / K_P A_P$$

*is a locally symmetric space corresponding to the reductive group* $L_P$. *Its closure* $\overline{X}_P$ *in* $\overline{X}^{RBS}$ *is the reductive Borel-Serre compactification of* $X_P$. *The geodesic* *projection $\pi_P : D \to e_P \to D_P$ passes to a geodesic projection, $\pi_P : V \to X_P$ defined on any parabolic neighborhood $V \subset \overline{X}^{RBS}$ of $X_P$. The pre-image $\pi_P^{-1}(B^r)$ of an open ball $B^r \subset X_P$ is a distinguished neighborhood of any $x \in B^r$.*

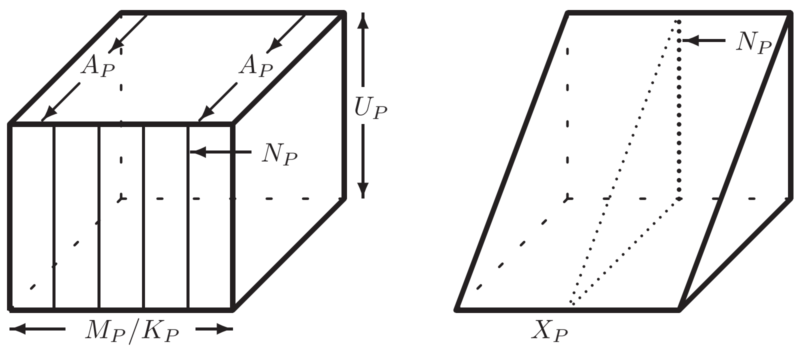



FIGURE 6. Borel-Serre and reductive Borel-Serre compactifications

The diagram on the left of Figure 5 represents the Borel-Serre compactification. This may be thought of as a "local" picture, but one may also imagine a "global" picture by identifying the top and bottom of the box, and identifying the left and right sides of the box. The box is a manifold with boundary: the front face is the boundary stratum $Y_P$. It is foliated by nilmanifolds isomorphic to $N_P$, and in general the vertical lines are the (images of) orbits of $U_P$. The geodesic action of $A_P$ moves points towards the front face. The (images of) orbits of $M_P$ are horizontal. On the right hand side, the nilmanifold fibers in $Y_P$ have been collapsed to points, leaving the stratum $X_P$. Nothing else has changed. However we now see that a normal slice through $X_P$ (indicated by a dotted triangle) is diffeomorphic to the cone over $N_P$, that is, the nilmanifold $N_P$ is the *link* of $X_P$ (see next section).

**5.2. Singularities of $\overline{X}^{\mathbf{RBS}}$.** The reductive Borel-Serre compactification of $X$ is a highly singular, non-algebraic space. Although the singularities are complicated, they can be precisely described, and as a consequence it is possible to compute the stalk cohomology of various sheaves on $\overline{X}^{\mathrm{RBS}}$. Here is a description of the link (cf. §2.6) of the stratum $X_P$ in $\overline{X}^{\mathrm{RBS}}$.

If $\mathbf{P}$ is a (proper) maximal rational parabolic subgroup of $\mathbf{G}$ then the link of the stratum $X_P$ is the compact nilmanifold $N_P = \Gamma_{U_P}\backslash U_P$ where $U_P$ is the unipotent radical of $P$ and $\Gamma_{U_P} = \Gamma \cap U_P$.

If $\mathbf{P} \subset \mathbf{Q}$ then $\mathbf{P}$ determines a parabolic subgroup $P/U_Q \subset L_Q$ with unipotent radical $U_P^Q = U_P/U_Q$ and discrete group $\Gamma_P^Q = \Gamma_{U_P}/\Gamma_{U_Q}$. Let $N_P^Q = \Gamma_P^Q\backslash U_P^Q$ be the associated nilmanifold. It is the quotient of $N_P$ under the action of $U_Q$ so there is a surjection $T_{PQ} : N_P \to N_P^Q$. Similarly, if $\mathbf{P} \subset \mathbf{R} \subset \mathbf{Q}$ we obtain a canonical surjection

$$N_P^Q \to N_P^R. \tag{5.2.1}$$

To make the notation more symmetric, let us also write $N_P = N_P^G$.

If $\mathbf{P} = \mathbf{Q}_1 \cap \mathbf{Q}_2$ is the intersection of two maximal rational (proper) parabolic subgroups then the link of the stratum $X_P$ in $\overline{X}^{\mathrm{RBS}}$ is the double mapping cylinder of the diagram

$$N_P^{Q_1} \xleftarrow[T_{PQ_1}]{} N_P^G \xrightarrow[T_{PQ_2}]{} N_P^{Q_2}.$$

In other words, it is the disjoint union $N_P^{Q_1} \cup (N_P \times [-1,1]) \cup N_P^{Q_2}$ modulo relations $(x,-1) \sim T_{PQ_1}(x)$ and $(x,1) \sim T_{PQ_2}(x)$ for all $x \in N_P$.

In the general case, suppose $\mathbf{P}$ is a rational parabolic subgroup of $\mathbf{G}$ with $\dim(A_P) = r$. The rational parabolic subgroups containing $\mathbf{P}$ (including $\mathbf{G}$) are in one to one correspondence with the faces of the $r-1$ dimensional simplex $\Delta^{r-1}$, in an inclusion-preserving manner, with the interior face corresponding to $\mathbf{G}$.

5.2. Theorem. ([**GHM**] §8) *The link of the stratum $X_P$ in $\overline{X}^{RBS}$ is homeomorphic (by a stratum preserving homeomorphism which is smooth on each stratum) to the geometric realization of the contravariant functor $N : \Delta^{r-1} \to \{manifolds\}$ defined on the category whose objects are faces of the $r-1$ simplex (and whose morphisms are inclusions of faces), which associates to each face $\mathbf{Q}$ the nilmanifold $N_P^Q$ and to each inclusion of faces $\mathbf{R} \subset \mathbf{Q}$ the morphism (5.2.1).*

5.3. Theorem. ([**Z2**]) *Suppose the symmetric space $D = G/K$ is Hermitian and let $\overline{X}^{BB}$ be the Baily-Borel compactification of $X = \Gamma\backslash D$. Then there exist unique continuous mappings*

$$\overline{X}^{BS} \longrightarrow \overline{X}^{RBS} \xrightarrow{\ \tau\ } \overline{X}^{BB}.$$

*which restrict to the identity on $X$.*

**5.3.** The first map is part of the definition of the reductive Borel-Serre compactification. The mapping $\tau$, if it exists, is determined by the fact that it is the identity on $X$. However at first glance it appears unlikely to exist since, when $\mathbf{G}$ is $\mathbb{Q}$ simple, the strata of $\overline{X}^{\mathrm{BB}}$ are indexed by ($\Gamma$ conjugacy classes of) *maximal* rational parabolic subgroups, while the strata of $\overline{X}^{\mathrm{RBS}}$ are indexed by ($\Gamma$ conjugacy classes of) all rational parabolic subgroups. Suppose for the moment that $\mathbf{G}$ is $\mathbb{Q}$ simple. (The general case follows from this.) Then the rational Dynkin diagram for $G$ is of type $C_n$ or $BC_n$, as in Figure 2. A rational parabolic subgroup corresponds to a subset of the Dynkin diagram, so its Levi quotient decomposes as an almost direct product (commuting product with finite intersections):

$$L_P = L_{Ph} \times L_{\ell 1} \times L_{\ell 2} \times \ldots \times L_{\ell m} \times H \tag{5.3.1}$$

of a (semisimple) Hermitian factor $L_{Ph}$ with a number of "linear factors" $L_{\ell i}$, (each of which acts as a group of automorphisms of a self adjoint homogeneous cone in some real vector space) and a compact group $H$. (In what follows we will assume the compact factor $H$, if it exists, has been absorbed into the other factors. It is possible to arrange this so that each of the resulting factors is defined over the rational numbers.)

So there is a projection $D_P \to F$ from the reductive Borel-Serre boundary component $D_P = L_P/K_PA_P$ to the Borel-Serre boundary component $F = L_{Ph}/K_{Ph}$ (for appropriate maximal compact subgroup $K_{Ph}$). This boundary component $F$ was associated (in §2.3) to the maximal (proper, rational) parabolic subgroup $\mathbf{Q}$ whose Levi component $L_Q$ decomposes as $L_Q = L_{Qh} \times L_{Q\ell}$ with $L_{Qh} = L_{Ph}$. In other words, $L_P$ and $L_Q$ have the same Hermitian factor.

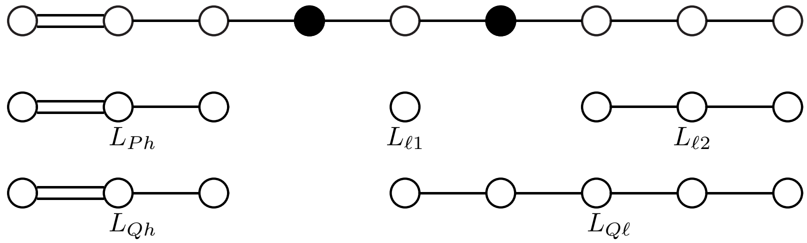



FIGURE 7. Dynkin diagrams for $G$, $L_P$, and $L_Q$

Moreover, the closure $\overline{D_P} = \overline{D_P}^{\mathrm{RBS}}$ decomposes as the product of reductive Borel-Serre partial compactifications of the locally symmetric spaces corresponding to the factors in (5.3.1). The symmetric spaces for the linear factors $L_{\ell 1} \times \ldots \times L_{\ell m}$ show up as boundary components of the symmetric space for $L_{Q\ell}$. With a bit more work it can be shown that

5.4. THEOREM. *The mapping $\tau : \overline{X}^{RBS} \to \overline{X}^{BB}$ takes strata to strata and it is a submersion on each stratum. If $X_F \subset \overline{X}^{BB}$ is the stratum corresponding to a maximal parabolic subgroup $\mathbf{Q} \subset \mathbf{G}$ then $\tau^{-1}(X_F) \to X_F$ is a fiber bundle whose fiber is isomorphic, by a stratum preserving isomorphism, to the reductive Borel-Serre compactification of the arithmetic quotient $\Gamma_{Q\ell} \backslash L_{Q\ell}/K_{Q\ell}$ of the self adjoint homogeneous cone $L_{Q\ell}/K_{Q\ell}$.*

(Here, $\Gamma_{Q\ell} \subset L_{Q\ell}(\mathbb{Q})$ is the arithmetic group which is obtained by first projecting $\Gamma \cap Q$ to $L_Q$ and then intersecting with $L_{Q\ell}$.)

In summary we have a diagram of partial compactifications and compactifications,

$$\begin{array}{ccccc} \overline{D}^{\mathrm{BS}} & \longrightarrow & \overline{D}^{\mathrm{RBS}} & \longrightarrow & \overline{D}^{\mathrm{BB}} \\ \downarrow & & \downarrow & & \downarrow \\ \overline{X}^{\mathrm{BS}} & \longrightarrow & \overline{X}^{\mathrm{RBS}} & \longrightarrow & \overline{X}^{\mathrm{BB}} \end{array}$$

with corresponding boundary components and boundary strata

$$\begin{array}{ccccc} e_P = P/K_PA_P & \longrightarrow & D_P = L_P/K_PA_P & \longrightarrow & F = L_{Ph}/K_{Ph} \\ \downarrow & & \downarrow & & \downarrow \\ Y_P = \Gamma_P \backslash e_P & \longrightarrow & X_P = \Gamma_L \backslash M_P/K_P & \longrightarrow & X_F = \Gamma_{Ph} \backslash F \end{array}$$

**5.4.** A very similar picture applies when $X = \Gamma \backslash \mathcal{P}_n$ is an arithmetic quotient of the symmetric cone of positive definite real matrices (or, more generally, when $X$ is an arithmetic quotient of any rationally defined $\mathbb{Q}$ irreducible self adjoint homogeneous cone). The identity mapping $X \to X$ has unique continuous extensions

$$\overline{X}^{\mathrm{BS}} \longrightarrow \overline{X}^{\mathrm{RBS}} \xrightarrow{\ \tau\ } \overline{X}^{\mathrm{std}}$$

which take strata to strata. A stratum $X_F$ of $\overline{X}^{\mathrm{std}}$ corresponds to a maximal rational parabolic subgroup $\mathbf{Q}$ whose Levi component factors, $L_Q = L_1 L_2$ as a product of two "linear" factors. The stratum $X_F$ is an arithmetic quotient of the self adjoint homogeneous cone for $L_1$. The pre-image $\tau^{-1}(X_F)$ is a fiber bundle over $X_F$ whose fiber over a point $x \in X_F$ is isomorphic to the reductive Borel-Serre compactification of an arithmetic quotient of the self adjoint homogeneous cone for $L_2$.

## 6. Cohomology

**6.1. Group cohomology.** (see [**Bro**] Chapt. I or [**W**] Chapt. 6) As in the preceding section we assume that $\mathbf{G}$ is reductive, defined over $\mathbb{Q}$, that $\mathbf{A_G}$ is trivial, that $K \subset \mathbf{G}(\mathbb{R})$ is a maximal compact subgroup and that the symmetric space $D = G/K$ is Hermitian with basepoint $x_0$. Fix a neat arithmetic subgroup $\Gamma \subset \mathbf{G}(\mathbb{Q})$ and set $X = \Gamma\backslash D$. Let $\lambda : G \to \mathbf{GL}(E)$ be a finite dimensional representation of $G$ on some complex vector space, and let $\mathbf{E} = E \times_\Gamma D$ be the resulting local system (flat vector bundle) on $X$. Since $\Gamma$ acts freely on the contractible manifold $D$ we see that the cohomology $H^i(X, \mathbf{E})$ is naturally isomorphic to the group cohomology $H^i(\Gamma, E)$ of the representation $\lambda|\Gamma$.

6.1. Theorem. *The cohomology $H^*(\Gamma, E)$ is finite dimensional. The group $\Gamma$ is finitely presented.*

If $\Gamma$ is neat, the proof follows just from the existence of the Borel-Serre compactification: the inclusion $X \to \overline{X}^{\mathrm{BS}}$ is a homotopy equivalence. Since $\Gamma$ is the fundamental group of $X$, it is finitely presented. Moreover any compact manifold with boundary (or compact manifold with corners) may be triangulated using finitely many simplices, so its cohomology is finite dimensional (and vanishes in dimensions greater than $\dim(X)$). In fact, these two consequences of the existence of the Borel-Serre compactification were first proven by M. S. Raghunathan [**R**], who showed that $X$ was diffeomorphic to the interior of a smooth compact manifold with boundary.

It can also be shown, using the Borel-Serre compactification, that the Euler characteristic $\chi(X)$ and the Euler characteristic with compact supports $\chi_c(X) = \sum_{i\geq 0}(-1)^i \dim(H^i_c(X))$ are equal. This follows from the fact that their difference is the Euler characteristic of the boundary $\partial\overline{X} = \overline{X} - X$ for any compactification $\overline{X}$ of $X$. One checks by induction that the Euler characteristic of the Borel-Serre boundary vanishes, since each "corner" $Y_P$ is fibered over $X_P$ with fiber a compact nilmanifold $N_P$, whose Euler characteristic $\chi(N_P) = 0$ vanishes.

**6.2. $L^2$ cohomology.** A choice of $K$-invariant inner product on the tangent space $T_{x_0}D$ determines a complete $G$-invariant Riemannian metric on $D$ which then passes to a complete Riemannian metric (with negative curvature) on $X$. Let $\Omega^i(X)$ be the vector space of smooth complex valued differential $i$-forms and let

$$\Omega^i_{(2)}(X) = \left\{ \omega \in \Omega^i(X) \,\middle|\, \int \omega \wedge *\omega < \infty, \int d\omega \wedge *d\omega < \infty \right\}$$

be the vector space of $L^2$ differential $i$-forms on $X$. These form a complex whose cohomology

$$H^i_{(2)}(X) = \ker(d)/\mathrm{im(d)}$$

is called the $L^2$ cohomology of $X$. It is finite dimensional (when $D$ is Hermitian symmetric, which we are currently assuming). We may similarly define the $L^2$

cohomology $H^i(X, \mathbf{E})$ with coefficients in a local system $\mathbf{E}$ arising from a finite dimensional irreducible representation $\lambda : \mathbf{G}(\mathbb{R}) \to \mathbf{GL}(E)$ on some complex vector space $E$.

The $L^2$ cohomology is a representation-theoretic object, and it may be identified ([**BW**]) with the relative Lie algebra cohomology

$$H^i_{(2)}(X, \mathbf{E}) \cong H^i(\mathfrak{g}, K; L^2(\Gamma\backslash G, E)) \tag{6.2.1}$$

of the module of $L^2$ functions on $\Gamma\backslash G$ with values in $E$. One would like to understand the decomposition of this module under the regular representation of $G$. (See lectures of J. Arthur in this volume.) This decomposition of $L^2(\Gamma\backslash G)$ is reflected in the resulting decomposition of its cohomology (6.2.1), which is somewhat easier to understand, but the information flows in both directions. For example, it is known (when $\mathbf{E}$ is trivial) that the trivial representation occurs exactly once in $L^2(\Gamma\backslash G)$, and that its $(\mathfrak{g}, K)$ cohomology coincides with the ordinary cohomology of the compact dual symmetric space $\check{D}$. Hence, $H^*(\check{D})$ occurs in $H^*_{(2)}(X, \mathbb{C})$.

**6.3. Zucker conjecture.** In [**Z1**], S. Zucker conjectured there is an isomorphism

$$H^i_{(2)}(X, \mathbf{E}) \cong IH^i(\overline{X}^{\mathrm{BB}}, \mathbf{E})$$

between the $L^2$ cohomology and the intersection cohomology of the Baily-Borel compactification. This beautiful conjecture relates an analytic and representation theoretic object, the $L^2$ cohomology, with a topological invariant, the intersection cohomology. (An analogous result, for "metrically conical" singular spaces, had been previously discovered [**Ch0, Ch1, Ch2**] by J. Cheeger. A relatively simple, piecewise linear analog is developed in [**BGM**].) Moreover, if $X$ is (replaced by) a Shimura variety, then it has a canonical model defined over a number field and there is an associated variety defined over various finite fields. In this case the intersection cohomology of $\overline{X}^{\mathrm{BB}}$ has an étale version, on which a certain Galois group acts. So the Zucker conjecture provides a "path" from automorphic representations to Galois representations, the understanding of which constitutes one of the goals of Langlands' program.

Zucker proved the conjecture in the case of $\mathbb{Q}$-rank one. Further special cases were proven by Borel, Casselman, and Zucker ([**B2**], [**BC1**], [**BC2**], [**Z3**]). Finally, in [**Lo**] and [**SS**] the conjecture was proven in full generality. Looijenga's proof uses the decomposition theorem ([**BBD**] thm. 6.2.5) and the toroidal compactification, while the proof of Saper and Stern uses analysis (essentially on the reductive Borel-Serre compactification). Among the many survey articles on this material we mention [**B2**], [**CGM**], [**Go**], and [**S4**].

Both the $L^2$ cohomology and the intersection cohomology are the (hyper) cohomology groups of complexes of sheaves, $\mathbf{\Omega}^{\bullet}_{(2)}(\overline{X}^{\mathrm{BB}}, \mathbf{E})$ and $\mathbf{I\Omega}^{\bullet}(\overline{X}^{\mathrm{BB}}, \mathbf{E})$ respectively. The proofs of Looijenga and Saper and Stern construct a quasi-isomorphism between these complexes of sheaves. This implies, for example, the existence of an isomorphism between the $L^2$ cohomology and the intersection cohomology of any open set $V \subset \overline{X}^{\mathrm{BB}}$, and these isomorphisms are compatible with the maps induced by inclusion of open sets, exact sequences of pairs, and Mayer Vietoris sequences.

**6.4. Review of sheaf theory.** Let $Z$ be a stratified space with a regular stratification and let $\mathbf{S}$ be a sheaf (of finite dimensional vector spaces over some field) on $Z$. The stalk of the sheaf $\mathbf{S}$ at the point $x \in Z$ is denoted $\mathbf{S}_x$. A *local system* on $Z$ is a locally trivial sheaf. Denote by $\Gamma(U, \mathbf{S})$ the sections of $\mathbf{S}$ over an open set $U \subset Z$. The sheaf $\mathbf{S}$ is *fine* if it admits partitions of unity. (That is, for any locally finite cover $\{U_\alpha\}$ of $Z$, and for any open $V \subset Z$, every section $\omega \in \Gamma(V, \mathbf{S})$ can be written as a sum of sections $\omega_\alpha$ supported in $U_\alpha \cap V$.) If $f : Z \to W$ is a continuous mapping and if $\mathbf{S}$ is a sheaf on $Z$ then its push forward $f_*(\mathbf{S})$ is the sheaf on $W$ whose sections over an open set are

$$\Gamma(U, f_*(\mathbf{S})) = \Gamma(f^{-1}(U), \mathbf{S}).$$

Let

$$\mathbf{S}^0 \xrightarrow{d_0} \mathbf{S}^1 \xrightarrow{d_1} \cdots$$

be a complex of sheaves (of vector spaces) on $Z$ which is bounded from below. Such a complex is denoted $\mathbf{S}^\bullet$ rather than $\mathbf{S}^*$ to indicate that it comes with a differential. It is common to write $\mathbf{S}[k]^\bullet$ or $\mathbf{S}^\bullet[k]$ for the shifted complex, $\mathbf{S}[k]^i = \mathbf{S}^{k+i}$.

If $\mathbf{S}^\bullet$ is a complex of sheaves on $Z$ its *stalk cohomology* $H^i_x(\mathbf{S}^\bullet)$ and *stalk cohomology with compact supports* $H^i_{c,x}(\mathbf{S}^\bullet) = H^i_{\{x\}}(\mathbf{S}^\bullet)$ at a point $x \in X_F \subset \overline{X}^{\mathrm{BB}}$ are the limits

$$\lim_{\to} H^i(U_x, \mathbf{S}^\bullet) \text{ and } \lim_{\leftarrow} H^i_c(U_x, \mathbf{S}^\bullet) \tag{6.4.1}$$

respectively, over a basis of neighborhoods $U_x \subset \overline{X}^{\mathrm{BB}}$ of $x$ (ordered with respect to containment $U_x \supset U'_x \supset \cdots$). Sheaves form an abelian category so $\ker(d_i)$ and $\mathrm{Im}(d_{i-1})$ are sheaves, and we may form the *cohomology sheaf* $\mathbf{H}^i(\mathbf{S}) = \ker(d_i)/\mathrm{Im}(d_{i-1})$. Its stalk at a point $x \in Z$ is the stalk cohomology $H^i_x(\mathbf{S}^\bullet)$. The complex of sheaves $\mathbf{S}^\bullet$ is *cohomologically constructible* with respect to the stratification if, for each $i$, the restriction of the cohomology sheaf $\mathbf{H}^i(\mathbf{S}^\bullet)$ to each stratum is finite dimensional and locally trivial. This implies that $H^i(Z, \mathbf{S}^\bullet)$ is finite dimensional provided $Z$ is compact.

A morphism $f : \mathbf{S}^\bullet \to \mathbf{T}^\bullet$ of complexes of sheaves is a *quasi-isomorphism* if it induces isomorphisms $H^i_x(\mathbf{S}^\bullet) \cong H^i_x(\mathbf{T}^\bullet)$ for every $i$ and for every $x \in Z$. Such a quasi-isomorphism $\mathbf{S}^\bullet \to \mathbf{T}^\bullet$ induces isomorphisms $H^i(U, \mathbf{S}^\bullet) \cong H^i(U, \mathbf{T}^\bullet)$ for any open set $U \subseteq Z$ and these isomorphisms are compatible with the maps induced by inclusions and with Mayer Vietoris sequences.

If $\mathbf{S}^\bullet$ is a complex of fine sheaves, then for any open set $U \subseteq Z$ the cohomology $H^i(U, \mathbf{S}^\bullet)$ is the cohomology of the complex of sections over $U$,

$$\to \Gamma(U, \mathbf{S}^{i-1}) \to \Gamma(U, \mathbf{S}^i) \to \Gamma(U, \mathbf{S}^{i+1}) \to$$

However if $\mathbf{S}^\bullet$ is not fine, then this procedure gives the wrong answer. (Take, for example, the constant sheaf on a smooth manifold.) A *fine resolution* of $\mathbf{S}^\bullet$ is a quasi-isomorphism $\mathbf{S}^\bullet \to \mathbf{T}^\bullet$ where $\mathbf{T}^\bullet$ is fine. Then, in general, the cohomology $H^i(U, \mathbf{S}^\bullet)$ is defined to be the cohomology $H^i(U, \mathbf{T}^\bullet)$ for any fine (or flabby, or injective) resolution $\mathbf{T}^\bullet$ of $\mathbf{S}^\bullet$.

A similar problem arises when $f : Z \to W$ is a continuous mapping: if $\mathbf{S}^\bullet$ is a complex of fine sheaves on $Z$ then the push forward $f_*(\mathbf{S}^\bullet)$ will satisfy

$$H^i(U, f_*(\mathbf{S}^\bullet)) \cong H^i(f^{-1}(U), \mathbf{S}^\bullet) \tag{6.4.2}$$

for any open set $U \subseteq W$. However if $\mathbf{S}^\bullet$ is not fine then (6.4.2) may fail, and $\mathbf{S}^\bullet$ should first be replaced by a fine (or flabby or injective) resolution before pushing forward. The resulting complex of sheaves (or rather, its quasi-isomorphism class) is denoted $Rf_*(\mathbf{S}^\bullet)$.

These apparently awkward constructions have their most natural expression in terms of the *derived category* of sheaves on $Z$, for which many excellent references exist. (See [**I**], [**GeM**], [**GeM2**]). Brief summaries are given in [**GM**], [**B5**].) However, the sheaves to be studied in the following sections will be fine, so no further resolutions are required.

Originally it was felt that the "dual" of a sheaf (or of a complex of sheaves) should be a co-sheaf (an object similar to a sheaf, but for which the restriction arrows are reversed). However, in [**BM**], Borel and Moore constructed the *dual sheaf* $\mathbf{T}^\bullet$ of a complex of sheaves $\mathbf{S}^\bullet$ on $Z$. They showed, for any open set $U \subset Z$, that $H^i_c(U, \mathbf{T}^\bullet)$ is the vector space dual of $H^i(U, \mathbf{S}^\bullet)$. In [**V**], Verdier showed there was a sheaf $\mathbf{D}^\bullet$ (called the dualizing sheaf) such that the Borel-Moore dual $\mathbf{T}^\bullet$ was quasi-isomorphic to the sheaf $\mathbf{Hom}^\bullet(\mathbf{S}^\bullet, \mathbf{D}^\bullet)$. In particular, the dual of the dual of $\mathbf{S}^\bullet$ is not equal to $\mathbf{S}^\bullet$, however it is quasi-isomorphic to $\mathbf{S}^\bullet$. There are many quasi-isomorphic models for the dualizing sheaf. Possible models include the sheaf of (singular) chains on $Z$ (or piecewise linear chains, or subanalytic chains, if $Z$ has a piecewise linear or subanalytic structure). If $Z$ is compact, orientable (meaning that the top stratum of $Z$ is orientable), and purely $n$-dimensional, then $H^i(Z, \mathbf{D}^\bullet)$ is the homology $H_{n-i}(Z)$.

**6.5. The $L^2$ sheaf.** Return to the situation of §6, with $X = \Gamma\backslash G/K$ a Hermitian locally symmetric space. The sheaf $\mathbf{\Omega}^i_{(2)}(\overline{X}^{\mathrm{BB}}, \mathbf{E})$ of (smooth) $L^2$ differential forms on $\overline{X}^{\mathrm{BB}}$ is defined to be the sheafification of the presheaf whose sections over an open set $U \subset \overline{X}^{\mathrm{BB}}$ are

$$\left\{ \omega \in \Omega^i(U \cap X, E) \,\middle|\, \int_{U\cap X} \omega \wedge *\omega < \infty \text{ and } \int_{U\cap X} d\omega \wedge *d\omega < \infty \right\}$$

A common mistake is to confuse this with the direct image

$$j_*\mathbf{\Omega}^i_{(2)}(X, \mathbf{E})$$

of the sheaf of (smooth) $L^2$ $\mathbf{E}$-valued differential forms on $X$, where $j : X \to \overline{X}^{\mathrm{BB}}$ is the inclusion. In fact the sheafification of the presheaf of smooth $L^2$ ($\mathbf{E}$-valued) differential forms on $X$ is the sheaf of all smooth ($\mathbf{E}$-valued) differential forms on $X$. Its cohomology is the ordinary cohomology $H^*(X, \mathbf{E})$ and so the same is true of $j_*\mathbf{\Omega}^i_{(2)}(X, \mathbf{E})$.

6.2. Theorem ([**Z1**]). *The sheaf $\mathbf{\Omega}^\bullet_{(2)}(\mathbf{E}) = \mathbf{\Omega}^\bullet_{(2)}(\overline{X}^{BB}, \mathbf{E})$ of smooth $L^2$ differential forms on $\overline{X}^{BB}$ is fine.*

This implies that one may calculate the (hyper) cohomology of this complex of sheaves simply by taking the cohomology of the global sections (that is, globally defined $L^2$ differential forms), so we do indeed get the $L^2$ cohomology, that is,

$$H^i_{(2)}(X, \mathbf{E}) \cong H^i(\overline{X}^{\mathrm{BB}}, \mathbf{\Omega}^\bullet_{(2)}(\mathbf{E})).$$

**6.6. Middle Intersection cohomology.** There is a construction of intersection cohomology using differential forms, which R. MacPherson and I worked out some years ago (see [**Bry**] and [**P**]). Let $\pi_F : V_F \to F$ be the geodesic projection of a parabolic neighborhood of a boundary stratum $X_F \subset \overline{X}^{\mathrm{BB}}$ of (complex) codimension $c$. Let us say that a smooth differential form $\omega \in \Omega^i(X, \mathbf{E})$ is *allowable near* $X_F$ if there exists a neighborhood $V_\omega \subset V_F$ of $X_F$ in $\overline{X}^{\mathrm{BB}}$ such that for any choice of $c$ smooth vector fields $A_1, A_2, \ldots, A_c$ in $V_\omega \cap X$, each tangent to the fibers of $\pi$, the contractions

$$i(A_1)i(A_2)\cdots i(A_c)(\omega) = 0 \quad \text{and} \quad i(A_1)i(A_2)\cdots i(A_c)(d\omega) = 0$$

vanish in $V_\omega \cap X$. We say a smooth differential form $\omega \in \Omega^i(X, E)$ is *allowable* if it is allowable near $X_F$, for every stratum $X_F$ of $\overline{X}^{\mathrm{BB}}$.

6.1. Definition. The sheaf $\mathbf{I\Omega^i}(\mathbf{E})$ on $\overline{X}^{\mathrm{BB}}$ is the sheafification of the presheaf whose sections over an open set $U \subset \overline{X}^{\mathrm{BB}}$ are

$$\left\{\omega \in \Omega^i(U \cap X, \mathbf{E}) \mid \omega \text{ is the restriction of an allowable form on } X\right\}.$$

This sheaf is fine, so its cohomology $IH^*(\overline{X}^{\mathrm{BB}}, \mathbf{E})$ coincides with the cohomology of the complex of allowable differential forms on $X$. Moreover its stalk cohomology and stalk cohomology with compact supports (6.4.1) are achieved in any *distinguished* neighborhood $V_x \subset \overline{X}^{\mathrm{BB}}$ (see §2.7 and §4.5), that is,

$$IH^i_x(\mathbf{E}) \cong IH^i(V_x, \mathbf{E}) \text{ and } IH^i_{\{x\}}(\mathbf{E}) \cong IH^i_c(V_x, \mathbf{E}).$$

The (stalk) cohomology is even the cohomology of the complex of allowable differential forms in $V_x$ which satisfy the allowability condition with respect to $X_F$ throughout the neighborhood $V_x$. (The corresponding statement for the stalk cohomology with compact supports is false.)

The complex of sheaves $\mathbf{I\Omega^\bullet}(\mathbf{E})$ has the following properties.

(1) It is constructible: its stalk cohomology (at any point) is finite dimensional, and its cohomology sheaves are locally trivial when restricted to any stratum.
(2) The restriction $\mathbf{I\Omega^\bullet}(\mathbf{E})|X$ is a fine resolution of the sheaf (of sections of) $\mathbf{E}$.
(3) If $F$ is a stratum of complex codimension $c$ then for any $x \in F$,

$$H^i_x(\mathbf{I\Omega^\bullet}(\mathbf{E})) = 0 \text{ for all } i \geq c$$
$$H^i_{c,x}(\mathbf{I\Omega^\bullet}(\mathbf{E})) = 0 \text{ for all } i \leq c.$$

Condition (3) says that the sheaf of differential forms has been *truncated by degree* at the stratum $X_F$, that is, the allowability condition has killed all the stalk cohomology of degree $\geq c$. In [**GM**] it is shown that any complex of sheaves $\mathbf{S}^\bullet$ satisfying these three conditions is quasi-isomorphic to the intersection complex, meaning that in the appropriate bounded constructible derived category $D^b_c(\overline{X}^{\mathrm{BB}})$ there is an isomorphism $\mathbf{S}^\bullet \cong \mathbf{I\Omega^\bullet}(\mathbf{E})$. So the proof of the Zucker conjecture amounts to checking that the sheaf of $L^2$ differential forms satisfies these conditions. Conditions (1) and (2) are easy, however checking condition (3), which is local in $\overline{X}^{\mathrm{BB}}$, involves a detailed understanding both of the local topology of $\overline{X}^{\mathrm{BB}}$ and of its metric structure.

The intersection cohomology sheaf is (Borel-Moore-Verdier) self dual. In particular, if $E_1$ and $E_2$ are dual finite dimensional representations of $G$ then for each open set $U \subset \overline{X}^{\text{BB}}$ the intersection cohomology vector spaces

$$IH^i(U, \mathbf{E_1}) \text{ and } IH_c^{2n-i}(U, \mathbf{E_2})$$

are dual, where $n = \dim_{\mathbb{C}}(X)$.

**6.7. Remark.** Condition (3) above says that $\mathbf{I\Omega}^\bullet$ is a perverse sheaf ([**BBD**]) on $\overline{X}^{\text{BB}}$. In fact the simple objects in the category $\text{Perv}_c(\overline{X}^{\text{BB}})$ of (constructible) perverse sheaves are the just the intersection complexes $j_*(\mathbf{I\Omega}^\bullet(\overline{X}_F, \mathbf{E_F}))[c_F]$ of closures of strata, where $j : \overline{X}_F \to \overline{X}^{\text{BB}}$ is the inclusion of the closure of a stratum $X_F$ of codimension $c_F$ and where $\mathbf{E_F}$ is a local coefficient system on $X_F$.

**6.8. Weighted cohomology.** If $f : Y \to Z$ is a morphism and if $\mathbf{S}^\bullet$ is a complex of fine sheaves on $Y$ then $f_*(\mathbf{S}^\bullet)$ is a complex of fine sheaves on $Z$ whose cohomology is the same: $H^i(Z, f_*(\mathbf{S}^\bullet)) \cong H^i(Y, \mathbf{S}^\bullet)$. So we can study the cohomology of $\mathbf{S}^\bullet$ locally on $Z$. However the converse is not always true: if $\mathbf{T}^\bullet$ is a complex of sheaves on $Z$, there does not necessarily exist a complex of sheaves $\mathbf{S}^\bullet$ on $Y$ so that $f_*(\mathbf{S}^\bullet) \cong \mathbf{T}^\bullet$.

One would like to study the intersection cohomology $IH^*(\overline{X}^{\text{BB}}, \mathbf{E})$ locally on the reductive Borel-Serre compactification, which is in many ways a simpler space than $\overline{X}^{\text{BB}}$. One might hope to use the sheaf of $L^2$ differential forms on $\overline{X}^{\text{RBS}}$, which again makes sense on the reductive Borel-Serre compactification. It is again a fine sheaf [**Z1**], and its cohomology is $H^*_{(2)}(X, \mathbf{E})$. However, the $L^2$ sheaf on $\overline{X}^{\text{RBS}}$ is not constructible: its stalk cohomology at a boundary point $x$ may be infinite dimensional. The weighted cohomology sheaf $\mathbf{WC}^\bullet(\overline{X}^{\text{RBS}}, \mathbf{E})$ is designed to be a good replacement; see Theorem 6.4 below. The idea is the following. For any stratum $X_P \subset \overline{X}^{\text{RBS}}$ the torus $A_P$ acts (by geodesic action) on any parabolic neighborhood $V_P$. This action should give rise to a decomposition of the stalk cohomology (of the sheaf $\mathbf{\Omega}^\bullet(\overline{X}^{\text{RBS}}, \mathbf{E})$ of all smooth differential forms) at any point $x \in X_P$ into weight spaces. We would like to kill all the cohomology with weights greater than or equal to some fixed value, that is, we would like a *weight truncation* of the sheaf of smooth differential forms. Unfortunately, the complex of smooth differential forms is infinite dimensional, and the torus $A_P$ does not act semi-simply (near $X_P$) on this complex. So it is first necessary to find an appropriate collection of differential forms with the same cohomology, which decomposes under the action of $A_P$. In [**GHM**] a subsheaf $\mathbf{\Omega}^\bullet_{\text{sp}}(\overline{X}^{\text{RBS}}, \mathbf{E})$ of smooth "special" differential forms is constructed with this property.

Assume $\mathbf{G}$ is reductive and $\mathbf{A_G}$ is trivial. Fix a standard minimal rational parabolic subgroup $\mathbf{P_0} \subset \mathbf{G}$. Let $\mathbf{A_0}$ be the greatest $\mathbb{Q}$ split torus in the center of $\mathbf{P_0}$. Fix a "weight profile" $\nu \in X^*_{\mathbb{Q}}(\mathbf{A_0})$, that is, a rational character of $\mathbf{A_0}$. This will be used to determine weight cutoffs for each stratum. Suppose $P$ is a standard rational parabolic subgroup. The choice of basepoint $x_0 \in D$ determines a lift $L_P \subset P$ (see §2.7), so the action of $P$ on its unipotent radical restricts to an action of $L_P \subset P$ on the complex

$$C^\bullet(\mathfrak{N}_P, E) = \text{Hom}_{\mathbb{R}}(\wedge^\bullet(\mathfrak{N}_P), E)$$

(where $\mathfrak{N}_P = \text{Lie}(U_P)$) and hence determines a local system

$$\mathbf{C}^\bullet(\mathfrak{N}_P, E) = C^\bullet(\mathfrak{N}_P, E) \times_{\Gamma_L} (L_P \backslash K_P A_P)$$

over the reductive Borel-Serre stratum $X_P = \Gamma_L \backslash L_P / K_P A_P$, cf. (5.1.2). The torus $A_P$ acts on $C^\bullet(\mathfrak{N}_P, E)$ so we obtain a decomposition into weight submodules

$$C^\bullet(\mathfrak{N}_P, E) \cong \bigoplus_{\mu \in X(A_P)} C^\bullet(\mathfrak{N}_P, E)_\mu.$$

Using the weight profile $\nu$, define the submodule

$$C^\bullet(\mathfrak{N}_P, E)_{\geq \nu} = \bigoplus_{\mu \geq \nu} C^\bullet(\mathfrak{N}_P, E)_\nu$$

where $\mu \geq \nu$ means that $\mu - (\nu | A_P)$ lies in the positive cone spanned by the simple rational roots $\alpha \in \Delta_P$. This definition also makes sense when $\mathbf{P}$ is an arbitrary rational parabolic subgroup, by conjugation.

Suppose $V \subset \overline{X}^{\mathrm{RBS}}$ is a parabolic neighborhood of $X_P$. Then it turns out that the complex of differential forms which are special throughout $V$ may be identified with the complex $\Omega^\bullet_{sp}(X_P, \mathbf{C}^\bullet(\mathfrak{N}_P, E))$ of special differential forms on $X_P$ with coefficients in the (finite dimensional) local system $\mathbf{C}^\bullet(\mathfrak{N}_P, E)$. Define $\Omega^\bullet_{sp}(V)_{\geq \nu}$ to be the subcomplex of special differential forms on $X_P$ with coefficients in the subbundle $\mathbf{C}^\bullet(\mathfrak{N}_P, E)_{\geq \nu}$. The subcomplex of $\Omega^\bullet_{sp}(V)$ is independent of the choice of basepoint.

6.2. Definition. The weighted cohomology $\mathbf{W}^{\geq \nu}\mathbf{C}^\bullet(\overline{X}^{\mathrm{RBS}}, \mathbf{E})$ is the sheafification of the complex of sheaves whose sections over an open set $U \subset \overline{X}^{\mathrm{RBS}}$ consist of smooth differential forms $\omega$ on $U \cap X$ such that for each stratum $X_P$ there exists a parabolic neighborhood $V = V(\omega, X_P) \subset \overline{X}^{\mathrm{RBS}}$ with $\omega | V \in \Omega^\bullet_{sp}(V)_{\geq \nu}$.

It is possible to similarly define $\mathbf{W}^{> \nu}\mathbf{C}^\bullet(\overline{X}^{\mathrm{RBS}}, \mathbf{E})$. It will coincide with the sheaf $\mathbf{W}^{\geq \nu}\mathbf{C}^\bullet(\overline{X}^{\mathrm{RBS}}, E)$ if, for each rational parabolic subgroup $\mathbf{P}$, the weight $\nu | A_P$ does not occur in any of the cohomology groups $H^i(\mathfrak{N}_P, E)$.

6.3. Theorem. *The complex $\mathbf{W}^{\geq \nu}\mathbf{C}^\bullet$ is constructible with respect to the canonical stratification of $\overline{X}^{RBS}$, so its cohomology is finite dimensional. Its restriction to $X$ is a fine resolution of the sheaf (of sections of) $\mathbf{E}$. The stalk cohomology, and compactly supported stalk cohomology at a point $x \in X_P$ are given by*

$$WH^j_x \cong H^j(\mathfrak{N}_P, E)_{\geq \nu}$$
$$WH^j_{c,x} \cong H^{j-d-s}(\mathfrak{N}_P, E)_{< \nu}$$

*where $s = \dim(A_P)$ and $d = \dim(X_P)$.*

It is possible that these conditions uniquely determine the weighted cohomology sheaf in the bounded constructible derived category of $\overline{X}^{\mathrm{RBS}}$. In any case this theorem is considerably more complete than the corresponding result in §6.6 for intersection cohomology (which only specifies the region in which these stalk cohomology groups vanish). This illustrates the fact that the reductive Borel-Serre compactification is easier to understand than the Baily-Borel compactification.

There are many possible weight truncations. The two extreme truncations ($\nu = -\infty$ and $\nu = \infty$) give rise to a weighted cohomology sheaf on $\overline{X}^{\mathrm{RBS}}$ whose cohomology is the ordinary cohomology $H^*(X, \mathbf{E})$ and the ordinary cohomology with compact supports $H^*_c(X, \mathbf{E})$ of $X$, respectively. Another weight truncation $\nu = 0$ (and $\mathbf{E}$ trivial) gives the ordinary cohomology $H^*(\overline{X}^{\mathrm{RBS}}, \mathbb{C})$.

If $E_1$ and $E_2$ are dual (finite dimensional) irreducible representations of $G$ and if $\mu+\nu=-2\rho$ then the weighted cohomology complexes $\mathbf{W}^{\geq\nu}\mathbf{C}^\bullet(\overline{X}^{\mathrm{RBS}},\mathbf{E_1})$ and $\mathbf{W}^{>\mu}\mathbf{C}^\bullet(\overline{X}^{\mathrm{RBS}},\mathbf{E_2})$ are (Verdier) dual sheaves. (Here $\rho$ is one-half the sum of the positive roots.) In particular, for any open set $U\subseteq\overline{X}^{\mathrm{RBS}}$ the cohomology groups

$$W^{\geq\nu}H^i(U,\mathbf{E_1}) \text{ and } W^{>\mu}H_c^{n-i}(U,\mathbf{E_2})$$

are dual vector spaces (where $n=\dim(X)$). Thus, taking $m=-\rho$ there are two "middle" weighted cohomology sheaves (which may coincide),

$$\mathbf{W}^{\geq m}\mathbf{C}^\bullet(\overline{E}^{\mathrm{RBS}}) \quad\text{and}\quad \mathbf{W}^{>m}\mathbf{C}^\bullet(\overline{X}^{\mathrm{RBS}},\mathbf{E}).$$

The weighted cohomology construction makes sense whether or not $D$ is Hermitian. But in the Hermitian case we also have the mapping $\tau:\overline{X}^{\mathrm{RBS}}\to\overline{X}^{\mathrm{BB}}$ of §5.3. Let $E$ be an irreducible finite dimensional representation of $G$.

6.4. Theorem. ([**GHM**] *Theorem 23.2*) *The above mapping $\tau$ induces quasi-isomorphisms*

$$\tau_*(\mathbf{W}^{\geq m}\mathbf{C}^\bullet(\overline{X}^{RBS},\mathbf{E}))\cong\tau_*(\mathbf{W}^{>m}\mathbf{C}^\bullet(\overline{X}^{RBS},\mathbf{E}))\cong\mathbf{I\Omega}^\bullet(\overline{X}^{BB},\mathbf{E})$$

*and in particular the weighted cohomology of $\overline{X}^{RBS}$ is canonically isomorphic to the intersection cohomology of $\overline{X}^{BB}$.*

**6.9. Hecke correspondences.** Any $g\in\mathbf{G}(\mathbb{Q})$ gives rise to a *Hecke correspondence* $X'\rightrightarrows X$, meaning that we have two finite surjective mappings $c_1,c_2$ from $X'$ to $X$. It is defined as follows. Let $\Gamma'=\Gamma\cap g^{-1}\Gamma g$, $X'=\Gamma'\backslash D$. The two mappings are: $\Gamma' hK\mapsto(\Gamma hK,\Gamma ghK)$. They give an immersion $X'\to X\times X$ whose image may be thought of as a multi-valued mapping $X\to X$. The Hecke correspondence defined by any $g'\in\Gamma g\Gamma$ is the same as that defined by $g$ (cf. [**GM2**] §6.6). The composition of Hecke correspondences defined by $g,g'\in\mathbf{G}(\mathbb{Q})$ is not the Hecke correspondence defined by $gg'$, but rather, it is a finite linear combination of Hecke correspondences (cf. [**Sh**] §3.1). So the set of finite formal linear combinations of Hecke correspondences form a ring, the *Hecke ring* or Hecke algebra of $\Gamma$.

Fix a Hecke correspondence $(c_1,c_2):X'\rightrightarrows X$. Differential forms on $X$ may be pulled back by $c_2$ then pushed forward by $c_1$, and $L^2$ forms are taken to $L^2$ forms by this procedure. The induced mapping on $H^i_{(2)}(X,\mathbf{E})$ is called a Hecke operator. Using the trace formula, J. Arthur ([**A**]) gave an expression for the Lefschetz number of this operator, that is, the alternating sum of the traces of the induced mapping on the $L^2$ cohomology.

Both mappings $(c_1,c_2):X'\rightrightarrows X$ extend to finite mappings

$$\overline{X'}^{\mathrm{RBS}}\rightrightarrows\overline{X}^{\mathrm{RBS}} \quad\text{and}\quad \overline{X'}^{\mathrm{BB}}\rightrightarrows\overline{X}^{\mathrm{BB}}.$$

A *fixed point* $x\in\overline{X'}^{\mathrm{RBS}}$ is a point such that $c_1(x)=c_2(x)$. In [**GM2**] the Lefschetz fixed point formula for the action of this Hecke correspondence on the weighted cohomology of $\overline{X}^{\mathrm{RBS}}$ was computed. In [**GKM**] it was shown how the contributions from individual fixed point components in $\overline{X}^{\mathrm{RBS}}$ may be grouped together so as to make the Lefschetz formula (for the middle weighted cohomology) agree, term by term, with the $L^2$ Lefschetz formula of Arthur. This gives a purely topological interpretation (and re-proof) of Arthur's formula, as well as similar formulas for other weighted cohomology groups.

## 7. A selection of further developments

Throughout this section we assume that $X = \Gamma\backslash G/K$ is a Hermitian locally symmetric space, with $\mathbf{G}$ semi-simple, as in the preceding section.

**7.1.** Let $\widehat{X}$ be the closure of $X$ in $\overline{X}_\Sigma^{\text{tor}} \times \overline{X}^{\text{RBS}}$. In [**GT1**] it is shown that the fibers of the mapping $\pi_1 : \widehat{X} \to \overline{X}_\Sigma^{\text{tor}}$ are contractible, so there exists a homotopy inverse $\mu : \overline{X}_\Sigma^{\text{tor}} \to \widehat{X}$, that is, $\pi_1\mu$ and $\mu\pi_1$ are homotopic to the identity. The composition $\overline{X}_\Sigma^{\text{tor}} \to \widehat{X} \to \overline{X}^{\text{RBS}}$ allows one to compare the cohomology of $\overline{X}^{\text{RBS}}$ and $\overline{X}_\Sigma^{\text{tor}}$.

**7.2.** In [**GHMN**] it is shown that the restriction of the weighted cohomology sheaf to the closure of any stratum of $\overline{X}^{\text{RBS}}$ decomposes as a direct sum of weighted cohomology sheaves for that stratum. (The analogous statement for intersection cohomology is false.)

**7.3.** In [**Z4**], S. Zucker showed that for large $p$, the $L^p$ cohomology of $X = \Gamma\backslash G/K$ is naturally isomorphic to the ordinary cohomology $H^*(\overline{X}^{\text{RBS}})$ of the reductive Borel-Serre compactification. Although this result is much easier to prove than the original Zucker conjecture, it went surprisingly unnoticed for twenty years. In [**Nr**], A. Nair showed that the weighted cohomology $WH^*(\overline{X}^{\text{RBS}}, \mathbf{E})$ is canonically isomorphic to the weighted $L^2$ cohomology of J. Franke [**Fr**]. In [**S1**], [**S2**], L. Saper showed that the push forward $\tau_*(\mathbf{I\Omega}^\bullet(\overline{X}^{\text{RBS}}, \mathbf{E}))$ is canonically isomorphic to $\mathbf{I\Omega}^\bullet(\overline{X}^{\text{BB}}, \mathbf{E})$. This gives the surprising result that

$$IH^i(\overline{X}^{\text{RBS}}, \mathbf{E}) \cong IH^i(\overline{X}^{\text{BB}}, \mathbf{E}) \cong W^{\geq m}H^i(\overline{X}^{\text{RBS}}, \mathbf{E}).$$

However, on $\overline{X}^{\text{RBS}}$, the weighted cohomology sheaf and sheaf of intersection forms are definitely not quasi-isomorphic: Saper's theorem depends on delicate global vanishing results for the weighted cohomology groups of various boundary strata.

**7.4.** Many other compactifications of $\Gamma\backslash G/K$ were constructed by Satake ([**Sa2**], [**Z2**]). Each Satake compactifications depend on a choice of (what is now called) a "geometrically rational" representation of $G$. If the representation is rational, then it is geometrically rational, however the Baily-Borel compactification arises from a geometrically rational representation that is not necessarily rational. So the issue of determining which representations are geometrically rational is quite subtle. See [**Ca**] and [**S3**] for more details.

**7.5.** The most successful method for computing the $L^2$ cohomology involves understanding relative Lie algebra cohomology and automorphic representations. See, for example, [**Ko, LS, Sch, BW**].

## References

[A] J. Arthur, The $L^2$ Lefschetz numbers of Hecke operators, Invent. Math. **97** (1989), pp. 257–290.

[AMRT] A. Ash, D. Mumford, M. Rapoport, Y. -S. Tai, **Smooth Compactification of Locally Symmetric Varieties**, Math. Sci. Press, Brookline MA, 1975.

[BB] W. Baily and A. Borel, Compactification of arithmetic quotients of bounded symmetric domains. Ann. Math. **84** (1966), pp. 442-528.

[BBD] A. Beilinson, J. Bernstein and P. Deligne, **Faisceaux Pervers**, Astérisque **100**, Soc. Mat. de France, Paris, 1982.

[B1] A. Borel, **Introduction aux Groupes Arithmétiques**, Hermann, Paris, 1969.

[B2] A. Borel, $L^2$ cohomology and intersection cohomology of certain arithmetic varieties, in **Emmy Noether in Bryn Mawr**, Springer Verlag, New York, 1983, pp. 119–131.

[B3] A. Borel, Introduction to automorphic forms, in **Arithmetic Groups**, Proc. Symp. Pure Math. **IX**, Amer. Math. Soc., Providence R.I., 1966.

[B4] A. Borel, Cohomology and spectrum of an arithmetic group, **Proceedings of a Conference on Operator Algebras and Group Representations, Neptun, Romania**, Pitman, New York, 1983, p. 28–45.

[B5] A. Borel et al, **Intersection Cohomology**, Progress in Mathematics **50**, Birkhäuser Boston, Boston MA, 1984.

[BC1] A. Borel and W. Casselman, $L^2$ cohomology of locally symmetric manifolds of finite volume, Duke Math. J. **50** (1983), pp. 625–647.

[BC2] A. Borel and W. Casselman, Cohomologie d'intersection et $L_2$-cohomologie de variétés arithmétiques de rang rationnnel 2, C. R. Acad. Sci. Paris **301** Ser. 1 no. 7 (1985), pp. 369–373.

[BM] A. Borel and J. C. Moore, Homology theory for locally compact spaces, Mich. Math. J. **7** (1960), pp. 137–159.

[BS] A. Borel and J. P. Serre, Corners and arithmetic groups, Comment. Math. Helv. **48** (1973), pp. 436-491.

[BW] A. Borel and N. Wallach, **Continuous Cohomology, Discrete Subgroups, and Representations of Reductive Groups**, Princeton Univ. Press, Princeton N.J., 1980.

[BGM] J. P. Brasselet, M. Goresky, and R. MacPherson, Simplicial differential forms with poles, Amer. J. Math., **113** (1991), pp. 1019–1052.

[Bro] K. Brown, **Cohomology of Groups**, Graduate Texts in Mathematics **87**, Springer Verlag, New York, 1982.

[Bry] J. -L. Brylinski, Equivariant intersection cohomology, in **Kazhdan-Lusztig Theory and Related Topics**, Contemp. Math. **139** (1992), Amer. Math. Soc., Providence R.I., 1992.

[Ca] W. Casselman, Geometric rationality of Satake compactifications, in **Algebraic Groups and Lie Groups**, G. Lehrer ed., Cambridge University Press, Cambridge, UK, 1997.

[Ch0] J. Cheeger, On the spectral geometry of spaces with cone-like singularities, Proc. Nat. Acad. Sci. **76** (1979), pp. 2103–2106.

[Ch1] J. Cheeger, On the Hodge theory of Riemannian pseudomanifolds, in **Geometry of the Laplace Operator**, Proc. Symp. Pure Math **36**, Amer. Math. Soc., Providence R.I., 1980, pp. 91–146.

[Ch2] J. Cheeger, Spectral geometry of singular Riemannian spaces, J. Diff. Geom. **18** (1983), pp. 575–657.

[CGM] J. Cheeger, M. Goresky, and R. MacPherson, $L^2$ cohomology and intersection homology for singular varieties, in **Seminar in Differential Geometry** (S. T. Yau, ed.), Princeton University Press, Princeton N.J., 1982.

[FC] G. Faltings and C.-L. Chai, **Degeneration of Abelian Varieties**, Ergeb. Math. **22**, Springer Verlag, Berlin, 1990.

[Fr] J. Franke, Harmonic analysis in weighted $L_2$ spaces, Ann. Sci. École Norm. Sup. **31** (1998), pp. 181–279.

[GeM] S. Gelfand and Y. Manin, **Algebra V: Homological Algebra**, Encyclopedia of Mathematical Sciences **38**, Springer Verlag, Berlin, 1994.

[GeM2] S. Gelfand and Y. Manin **Methods of Homological Algebra**, Springer Verlag, Berlin, 2002.

[Go] M. Goresky, $L^2$ cohomology is intersection cohomology, in **The Zeta Functions of Picard Modular Surfaces**, R. P. Langlands and D. Ramakrishnan, ed., Centre de Recherches Mathématiques, Montréal, 1992.

[GHM] M. Goresky, G. Harder, and R. MacPherson, Weighted cohomology, Invent. Math. **116** (1994), pp. 139–213.

[GHMN] M. Goresky, G. Harder, R. MacPherson, and A. Nair, Local intersection cohomology of Baily Borel compactifications, Comp. Math. **134** (2002), pp. 243–268.

[GKM] M. Goresky, R. Kottwitz, and R. MacPherson, Discrete series characters and the Lefschetz formula for Hecke operators, Duke Math. J. **89** (1997), pp.477–554.

[GM] M. Goresky and R. MacPherson, Intersection homology II, Invent. Math. **72** (1983), pp. 77–129.

[GM2] M. Goresky and R. MacPherson, The topological trace formula, Crelle's Journal, **560** (2003), 77-150. .

[GP] M. Goresky and W. Pardon, Chern classes of automorphic vector bundles, Invent. Math. **147** (2002), pp. 561–612.

[GT1] M. Goresky and Y.-S. Tai, Toroidal and reductive Borel-Serre compactifications of locally symmetric spaces, Amer. J. Math. **121** (1999), pp. 1095–1151.

[GT2] M. Goresky and Y.-S. Tai, The moduli space of real abelian varieties with level structure, Comp. Math. (to appear).

[I] B. Iverson, **Cohomology of Sheaves**, Universitext, Springer Verlag, Berlin, 1986.

[Kl] H. Klingen, **Introductory Lectures on Siegel Modular Forms**, Cambridge University Press, Cambridge UK, 1990.

[Ko] R. Kottwitz, Shimura varieties and $\lambda$-adic representations, in **Automorphic Forms, Shimura varieties, and L-functions**, vol I, Academic Press, New York, 1990, p. 161–209.

[LS] J. -S. Li and J. Schwermer, On the Eisenstein cohomology of arithmetic groups. Preprint, Univ. Wien, Austria;

[Lo] E. Looijenga, $L^2$ cohomology of locally symmetric varieties. Comp. Math. **67** (1988), pp. 3-20.

[M] D. Mumford, Hirzebruch's proportionality theorem in the non-compact case, Inv. Math. **42** (1977), pp. 239–272.

[Nr] A. Nair, Weighted cohomology of arithmetic groups, Ann. Math. **150** (1999), pp. 1–31.

[Nk] Y. Namikawa, **Toroidal Compactification of Siegel Spaces**, Lecture Notes in Mathematics **812**, Springer Verlag, Berlin, 1980.

[P] G. Pollini, Modello di de Rham per la coomologia d'intersezione, Tesi di Laurea, University of Pisa, 2003.

[R] M. S. Raghunathan, A note on quotients of real algebraic groups by arithmetic subgroups, Invent. Math. **4** (1968), pp. 318–355.

[S1] L. Saper, $L$-modules and the conjecture of Rapoport and Goresky-MacPherson, `[arxiv.org/math.RT/0211138]`

[S2] L. Saper, $L$-modules and micro-support, `[arxiv.org/math.RT/0112251]`.

[S3] L. Saper, Geometric rationality of equal-rank Satake compactifications, preprint, Duke Univ., 2003.

[S4] L. Saper, On the Cohomology of Locally Symmetric Spaces and of their Compactifications, to appear in the proceedings of the Harvard/M.I.T. conference **Current Developments in Mathematics, 2002**, held in November 2002 [arxiv.org/math.RT/0306403] .

[SS] L. Saper and M. Stern, $L_2$ cohomology of arithmetic varieties, Ann. of Math. **132** (1990), pp. 1-69.

[Sa1] I. Satake, **Algebraic Structures Of Symmetric Domains**, Princeton University Press, Princeton N.J., 1980.

[Sa2] I. Satake, On compactifications of the quotient spaces for arithmetically defined discontinuous groups, Ann. Math. **72** (1960), pp. 555-580.

[Sch] J. Schwermer, Cohomology of arithmetic groups, automorphic forms and L-functions, in **Cohomology of Arithmetic Groups and Automorphic Forms**, Lecture Notes in Mathematics **1447**, Springer Verlag, Berlin, 1990, pp. 1–30.

[Sh] G. Shimura, **Introduction to the Arithmetic Theory of Automorphic Functions**, Princeton University Press, Princeton N.J., 1971.

[V] J. L. Verdier, Dualité dans la cohomologie des espaces localment compactes. Sém. Bourbaki **300** (1965).

[W] C. Weibel, **An Introduction to Homological Algebra**, Cambridge studies in advanced Mathematics **38**, Cambridge University Press, Cambridge UK, 1994.

[Z1] S. Zucker, $L_2$ cohomology of warped products and arithmetic groups. Invent. Math. **70** (1982), pp. 169–218.

[Z2] S. Zucker, Satake compactifications, Comment. Math. Helv. **58** (1983), pp. 312-343.

[Z3] S. Zucker, $L_2$ cohomology and intersection homology of locally symmetric varieties II, Comp. Math. **59** (1986), pp. 339–398.

[Z4] S. Zucker, On the reductive Borel-Serre compactification: $L^p$ cohomology of arithmetic groups (for large $p$), Amer. J. Math. **123** (2001), pp. 951–984.

School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540
*E-mail address:* goresky@ias.edu

Clay Mathematics Proceedings
Volume 4, 2005

# Introduction to Shimura Varieties with Bad Reduction of Parahoric Type

Thomas J. Haines

Contents



## 1. Introduction

This survey article is intended to introduce the reader to several important concepts relating to Shimura varieties with parahoric level structure at $p$. The main tool is the Rapoport-Zink local model [**RZ**], which plays an important role in several aspects of the theory. We discuss local models attached to general linear and symplectic groups, and we illustrate their relation to Shimura varieties in two examples: the *simple* or *"fake" unitary* Shimura varieties with parahoric level structure, and the Siegel modular varieties with $\Gamma_0(p)$-level structure. In addition, we present some applications of local models to questions of flatness, stratifications

Research partially supported by NSF grant DMS 0303605.

of the special fiber, and the determination of the semi-simple local zeta functions for simple Shimura varieties.

There are several good references for material of this sort that already exist in the literature. This survey has a great deal of overlap with two articles of Rapoport: [**R1**] and [**R2**]. A main goal of this paper is simply to make more explicit some of the ideas expressed very abstractly in those papers. Hopefully it will shed some new light on the earlier seminal works of Rapoport-Zink [**RZ82**], and Zink [**Z**]. This article is also closely related to some recent work of H. Reimann [**Re1**], [**Re2**], [**Re3**].

Good general introductions to aspects of the Langlands program which might be consulted while reading parts of this report are those of Blasius-Rogawski [**BR**], and T. Wedhorn [**W2**].

Several very important developments have taken place in the theory of Shimura varieties with bad reduction, which are completely ignored in this report. In particular, we mention the book of Harris-Taylor [**HT**] which uses bad reduction of Shimura varieties to prove the local Langlands conjecture for $\mathrm{GL}_n(\mathbb{Q}_p)$, and the recent work of L. Fargues and E. Mantovan [**FM**].

Most of the results stated here are well-known by now (although scattered around the literature, with differing systems of conventions). However, the author took this opportunity to present a few new results (and some new proofs of old results). For example, there is the proof of the non-emptiness of the Kottwitz-Rapoport strata in any connected component of the Siegel modular and "fake" unitary Shimura varieties with Iwahori-level structure (Lemmas 13.1, 13.2), some foundational relations between Newton strata, Kottwitz-Rapoport strata, and affine Deligne-Lusztig varieties (Prop. 12.6), and the verification of the conjectural non-emptiness of the basic locus in the "fake" unitary case (Cor. 12.12). The main new proofs relate to topological flatness of local models attached to Iwahori subgroups of unramified groups (see §7) and to the description of the nonsingular locus of Shimura varieties with Iwahori-level structure (see §8.4). Finally, some of the results explained here (especially in §11) are background material necessary for the author's as yet unpublished joint work with B.C. Ngô [**HN3**].

I am grateful to U. Görtz, R. Kottwitz, B.C. Ngô, G. Pappas, and M. Rapoport for all they have taught me about this subject over the years. I thank them for their various comments and suggestions on an early version of this article. Also, I heartily thank U. Görtz for providing the figures. Finally, I thank the Clay Mathematics Institute for sponsoring the June 2003 *Summer School on Harmonic Analysis, the Trace Formula, and Shimura Varieties*, which provided the opportunity for me to write this survey article.

## 2. Notation

**2.1. Some field-theoretic notation.** Fix a rational prime $p$. We let $F$ denote a non-Archimedean local field of residual characteristic $p$, with ring of integers $\mathcal{O}$. Let $\mathfrak{p} \subset \mathcal{O}$ denote the maximal ideal, and fix a uniformizer $\pi \in \mathfrak{p}$. The residue field $\mathcal{O}/\mathfrak{p}$ has cardinality $q$, a power of $p$.

We will fix an algebraic closure $k$ of the finite field $\mathbb{F}_p$. The Galois group $\mathrm{Gal}(k/\mathbb{F}_p)$ has a canonical generator (the Frobenius automorphism), given by $\sigma(x) = x^p$. For each positive integer $r$, we denote by $k_r$ the fixed field of $\sigma^r$. Let $W(k_r)$ (resp. $W(k)$) denote the ring of Witt vectors of $k_r$ (resp. $k$), with fraction field $L_r$

(resp. $L$). We also use the symbol $\sigma$ to denote the Frobenius automorphism of $L$ induced by that on $k$.

We fix throughout a rational prime $\ell \neq p$, and a choice of algebraic closure $\mathbb{Q}_\ell \subset \overline{\mathbb{Q}}_\ell$.

**2.2. Some group-theoretic notation.** The symbol $\mathbf{G}$ will always denote a connected reductive group over $\mathbb{Q}$ (sometimes defined over $\mathbb{Z}$). Unless otherwise indicated, $G$ will denote the base-change $\mathbf{G}_F$, where $F$ is a suitable local field (usually, $G = \mathbf{G}_{\mathbb{Q}_p}$).

Now let $G$ denote a connected reductive $F$-group. Fix once and for all a Borel subgroup $B$ and a maximal torus $T$ contained in $B$. We will usually assume $G$ is split over $F$, in which case we can even assume $G, B$ and $T$ are defined and split over the ring $\mathcal{O}$. For $\mathrm{GL}_n$ or $\mathrm{GSp}_{2n}$, $T$ will denote the usual "diagonal" torus, and $B$ will denote the "upper triangular" Borel subgroup.

We will often refer to "standard" parahoric and "standard" Iwahori subgroups. For the group $G = \mathrm{GL}_n$ (resp. $G = \mathrm{GSp}_{2n}$), the "standard" hyperspecial maximal compact subgroup of $G(F)$ will be the subgroup $G(\mathcal{O})$. The "standard" Iwahori subgroup will be inverse image of $B(\mathcal{O}/\mathfrak{p})$ under the reduction modulo $\mathfrak{p}$ homomorphism

$$G(\mathcal{O}) \to G(\mathcal{O}/\mathfrak{p}).$$

A "standard" parahoric will be defined similarly as the inverse image of a standard (= upper triangular) parabolic subgroup.

For $\mathrm{GL}_n(F)$, the standard Iwahori is the subgroup stabilizing the standard lattice chain (defined in §3). The standard parahorics are stabilizers of standard partial lattice chains. Similar remarks apply to the group $\mathrm{GSp}_{2n}(F)$. The symbols $I$ or $I_r$ or $K_p^{\mathbf{a}}$ will always denote a standard Iwahori subgroup of $G(F)$ defined in terms of our fixed choices of $B \supset T$, and $G(\mathcal{O})$ as above (for some local field $F$). Often (but not always) $K$ or $K_r$ or $K_p^0$ will denote our fixed hyperspecial maximal compact subgroup $G(\mathcal{O})$.

We have the associated spherical Hecke algebra $\mathcal{H}_K := C_c(K\backslash G(F)/K)$, a convolution algebra of $\mathbb{C}$-valued (or $\overline{\mathbb{Q}}_\ell$-valued) functions on $G(F)$ where convolution is defined using the measure giving $K$ volume 1. Similarly, $\mathcal{H}_I := C_c(I\backslash G(F)/I)$ is a convolution algebra using the measure giving $I$ volume 1. For a compact open subset $U \subset G(F)$, $\mathbb{I}_U$ denotes the characteristic function of the set $U$.

The *extended affine Weyl group* of $G(F)$ is defined as the group $\widetilde{W} = N_{G(F)}T/T(\mathcal{O})$. The map $X_*(T) \to T(F)/T(\mathcal{O})$ given by $\lambda \mapsto \lambda(\pi)$ is an isomorphism of abelian groups. The finite Weyl group $W_0 := N_{G(F)}T/T(F)$ can be identified with $N_{G(\mathcal{O})}T/T(\mathcal{O})$, by choosing representatives of $N_{G(F)}T/T(F)$ in $G(\mathcal{O})$. Thus we have a canonical isomorphism

$$\widetilde{W} = X_*(T) \rtimes W_0.$$

We will denote elements in this group typically by the notation $t_\nu w$ (for $\nu \in X_*(T)$ and $w \in W_0$).

Our choice of $B \supset T$ determines a unique *opposite* Borel subgroup $\bar{B}$ such that $B \cap \bar{B} = T$. We have a notion of $B$-positive (resp. $\bar{B}$-positive) root $\alpha$ and coroot $\alpha^\vee$. Also, the group $W_0$ is a Coxeter group generated by the simple reflections $s_\alpha$ in the vector space $X_*(T) \otimes \mathbb{R}$ through the walls fixed by the $B$-positive (or $\bar{B}$-positive) simple roots $\alpha$. Let $w_0$ denote the unique element of $W_0$ having greatest length with respect to this Coxeter system.

We will often need to consider $\widetilde{W}$ as a subset of $G(F)$. We choose the following conventions. For each $w \in W_0$, we fix once and for all a lift in the group $N_{G(\mathcal{O})}T$. We identify each $\nu \in X_*(T)$ with the element $\nu(\pi) \in T(F) \subset G(F)$.

Let $\mathbf{a}$ denote the alcove in the building of $G(F)$ which is fixed by the Iwahori $I$, or equivalently, the unique alcove in the apartment corresponding to $T$ which is contained in the $\bar{B}$-positive (i.e., the $B$-*negative*) Weyl chamber, whose closure contains the origin (the vertex fixed by the maximal compact subgroup $G(\mathcal{O})$) [1].

The group $\widetilde{W}$ permutes the set of affine roots $\alpha + k$ ($\alpha$ a root, $k \in \mathbb{Z}$) (viewed as affine linear functions on $X_*(T) \otimes \mathbb{R}$), and hence permutes (transitively) the set of alcoves. Let $\Omega$ denote the subgroup which stabilizes the base alcove $\mathbf{a}$. Then we have a semi-direct product

$$\widetilde{W} = W_{\mathrm{aff}} \rtimes \Omega,$$

where $W_{\mathrm{aff}}$ (the *affine Weyl group*) is the Coxeter group generated by the reflections $S_{\mathrm{aff}}$ through the walls of $\mathbf{a}$. In the case where $G$ is an almost simple group of rank $l$, with simple $B$-positive roots $\alpha_1, \dots, \alpha_l$, then $S_{\mathrm{aff}}$ consists of the $l$ simple reflections $s_i = s_{\alpha_i}$ generating $W_0$, along with one more simple affine reflection $s_0 = t_{-\widetilde{\alpha}^\vee} s_{\widetilde{\alpha}}$, where $\widetilde{\alpha}$ is the highest $B$-positive root.

The Coxeter system $(W_{\mathrm{aff}}, S_{\mathrm{aff}})$ determines a length function $\ell$ and a Bruhat order $\leq$ on $W_{\mathrm{aff}}$, which extend naturally to $\widetilde{W}$: for $x_i \in W_{\mathrm{aff}}$ and $\sigma_i \in \Omega$ ($i = 1, 2$), we define $x_1\sigma_1 \leq x_2\sigma_2$ in $\widetilde{W}$ if and only if $\sigma_1 = \sigma_2$ and $x_1 \leq x_2$ in $W_{\mathrm{aff}}$. Similarly, we set $\ell(x_1\sigma_1) = \ell(x_1)$.

By the Bruhat-Tits decomposition, the inclusion $\widetilde{W} \hookrightarrow G(F)$ induces a bijection

$$\widetilde{W} = I \backslash G(F)/I.$$

In the function-field case (e.g., $F = \mathbb{F}_p((t))$), the affine flag variety $\mathcal{F}l = G(F)/I$ is naturally an ind-scheme, and the closures of the $I$-orbits $\mathcal{F}l_w := IwI/I$ are determined by the Bruhat order on $\widetilde{W}$:

$$\mathcal{F}l_x \subset \overline{\mathcal{F}l_y} \Longleftrightarrow x \leq y.$$

Similar statements hold for the affine Grassmannian, $\mathrm{Grass} = G(F)/G(\mathcal{O})$. Now the $G(\mathcal{O})$-orbits $\mathcal{Q}_\lambda := G(\mathcal{O})\lambda G(\mathcal{O})/G(\mathcal{O})$ are given (using the Cartan decomposition) by the $B$-dominant coweights $X_+(T)$:

$$X_+(T) \leftrightarrow G(\mathcal{O}) \backslash G(F)/G(\mathcal{O}).$$

By definition, $\lambda$ is $B$-dominant if $\langle \alpha, \lambda \rangle \geq 0$ for all $B$-positive roots $\alpha$. Here $\langle \cdot, \cdot \rangle : X^*(T) \times X_*(T) \to \mathbb{Z}$ is the canonical duality pairing.

The closure relations in Grass are given by the partial order on $B$-dominant coweights $\lambda$ and $\mu$:

$$\mathcal{Q}_\lambda \subset \overline{\mathcal{Q}_\mu} \Longleftrightarrow \lambda \preceq \mu,$$

where $\lambda \prec \mu$ means that $\mu - \lambda$ is a sum of $B$-positive coroots.

Unless otherwise stated, a *dominant* coweight $\lambda \in X_*(T)$ will always mean one that is $B$-dominant.

For the group $G = \mathrm{GL}_n$ or $\mathrm{GSp}_{2n}$, there is a $\mathbb{Z}_p$-ind-scheme $M$ which is a deformation of the affine Grassmannian $\mathrm{Grass}_{\mathbb{Q}_p}$ to the affine flag variety $\mathcal{F}l_{\mathbb{F}_p}$ for the underlying $p$-adic group $G$ (see [**HN1**], and Remark 4.1). A very similar

[1] This choice of base alcove results from our convention of embedding $X_*(T) \hookrightarrow G(F)$ by the rule $\lambda \mapsto \lambda(\pi)$; to see this, consider how vectors spanning the standard periodic lattice chain in §3 are identified with vectors in $X_*(T) \otimes \mathbb{R}$.

deformation $\mathrm{Fl}_X$ over a smooth curve $X$ (due to Beilinson) exists for any group $G$ in the function field setting, and has been extensively studied by Gaitsgory [**Ga**]. For any dominant coweight $\lambda \in X_+(T)$, the symbol $M_\lambda$ will always denote the $\mathbb{Z}_p$-scheme which is the scheme-theoretic closure in $M$ of $\mathcal{Q}_\lambda \subset \mathrm{Grass}_{\mathbb{Q}_p}$.

**2.3. Duality notation.** If $A$ is any abelian scheme over a scheme $S$, we denote the dual abelian scheme by $\widehat{A}$. The existence of $\widehat{A}$ over an arbitrary base is a delicate matter; see [**CF**], §I.1.

If $M$ is a module over a ring $R$, we denote the dual module by $M^\vee = \mathrm{Hom}_R(M,R)$. Similar notation applies to quasi-coherent $\mathcal{O}_S$-modules over a scheme $S$.

If $G$ is a connected reductive group over a local field $F$, then the Langlands dual (over $\mathbb{C}$ or $\overline{\mathbb{Q}}_\ell$) will be denoted $\widehat{G}$. The Langlands $L$-group is the semi-direct product ${}^LG = \widehat{G} \rtimes W_F$, where $W_F$ is the Weil-group of $F$.

**2.4. Miscellaneous notation.** We will use the following abbreviation for elements of $R^n$ (here $R$ can be any set): let $a_1, \dots, a_r$ be a sequence of positive integers whose sum is $n$. A vector of the form $(x_1, \dots, x_1, x_2, \dots, x_2, \dots, x_r, \dots, x_r)$, where for $i = 1, \dots, r$, the element $x_i$ is repeated $a_i$ times, will be denoted by

$$(x_1^{a_1}, x_2^{a_2}, \dots, x_r^{a_r}).$$

We will denote by $\mathbb{A}$ the adeles of $\mathbb{Q}$, by $\mathbb{A}_f$ the finite adeles, and by $\mathbb{A}_f^p$ the finite adeles away from $p$ (with the exception of two instances, where $\mathbb{A}$ denotes affine space!).

## 3. Iwahori and parahoric subgroups

**3.1. Stabilizers of periodic lattice chains.** We discuss the definitions for the groups $\mathrm{GL}_n$ and $\mathrm{GSp}_{2n}$.

3.1.1. *The linear case.* For each $i \in \{1, \dots, n\}$, let $e_i$ denote the $i$-th standard vector $(0^{i-1}, 1, 0^{n-i})$ in $F^n$, and let $\Lambda_i \subset F^n$ denote the free $\mathcal{O}$-module with basis $\pi^{-1}e_1, \dots, \pi^{-1}e_i, e_{i+1}, \dots, e_n$. We consider the diagram

$$\Lambda_\bullet : \ \Lambda_0 \longrightarrow \Lambda_1 \longrightarrow \cdots \longrightarrow \Lambda_{n-1} \longrightarrow \pi^{-1}\Lambda_0,$$

where the morphisms are inclusions. The lattice chains $\pi^n\Lambda_\bullet$ $(n \in \mathbb{Z})$ fit together to form an infinite complete lattice chain $\Lambda_i$, $(i \in \mathbb{Z})$. If we identify each $\Lambda_i$ with $\mathcal{O}^n$, then the diagram above becomes

$$\mathcal{O}^n \xrightarrow{m_1} \mathcal{O}^n \xrightarrow{m_2} \cdots \xrightarrow{m_{n-1}} \mathcal{O}^n \xrightarrow{m_n} \mathcal{O}^n,$$

where $m_i$ is the morphism given by the diagonal matrix

$$m_i = \mathrm{diag}(1, \dots, \pi, \dots, 1),$$

the element $\pi$ appearing in the $i$th place. We define the *(standard) Iwahori subgroup* $I$ of $\mathrm{GL}_n(F)$ by

$$I = \bigcap_i \ \mathrm{Stab}_{\mathrm{GL}_n(F)}(\Lambda_i).$$

Similarly, for any non-empty subset $J \subset \{0, 1, \dots, n-1\}$, we define the *parahoric subgroup* of $\mathrm{GL}_n(F)$ corresponding to the subset $J$ by

$$\mathcal{P}_J = \bigcap_{i \in J} \ \mathrm{Stab}_{\mathrm{GL}_n(F)}(\Lambda_i).$$

Note that $\mathcal{P}_J$ is a compact open subgroup of $\mathrm{GL}_n(F)$, and that $\mathcal{P}_{\{0\}} = \mathrm{GL}_n(\mathcal{O})$ is a hyperspecial maximal compact subgroup, in the terminology of Bruhat-Tits, cf. [**T**].

3.1.2. *The symplectic case.* The definitions for the group of symplectic similitudes $\mathrm{GSp}_{2n}$ are similar. We define this group using the alternating matrix

$$\tilde{I} = \begin{bmatrix} 0 & \tilde{I}_n \\ -\tilde{I}_n & 0 \end{bmatrix},$$

where $\tilde{I}_n$ denotes the $n \times n$ matrix with 1 along the anti-diagonal and 0 elsewhere. Let $(x, y) := x^t \tilde{I} y$ denote the corresponding alternating pairing on $F^{2n}$. For an $\mathcal{O}$-lattice $\Lambda \subset F^{2n}$, we define $\Lambda^\perp := \{x \in F^{2n} \mid (x, y) \in \mathcal{O} \text{ for all } y \in \Lambda\}$. The lattice $\Lambda_0$ is self-dual (i.e., $\Lambda_0^\perp = \Lambda_0$). Consider the infinite lattice chain in $F^{2n}$

$$\cdots \longrightarrow \Lambda_{-2n} = \pi\Lambda_0 \longrightarrow \cdots \longrightarrow \Lambda_{-1} \longrightarrow \Lambda_0 \longrightarrow \cdots \longrightarrow \Lambda_{2n} = \pi^{-1}\Lambda_0 \longrightarrow \cdots$$

We have $\Lambda_i^\perp = \Lambda_{-i}$ for all $i \in \mathbb{Z}$. Now we define the *(standard) Iwahori subgroup* $I$ of $\mathrm{GSp}_{2n}(F)$ by

$$I = \bigcap_i \mathrm{Stab}_{\mathrm{GSp}_{2n}(F)}(\Lambda_i).$$

For any non-empty subset $J \subset \{-n, \ldots, -1, 0, 1, \ldots, n\}$ such that $i \in J \Leftrightarrow -i \in J$, we define the *parahoric subgroup* corresponding to $J$ by

$$\mathcal{P}_J = \bigcap_{i \in J} \mathrm{Stab}_{\mathrm{GSp}_{2n}(F)}(\Lambda_i).$$

**3.2. Bruhat-Tits group schemes.** In Bruhat-Tits theory, parahoric groups are defined as the groups $\mathfrak{G}^0_{\Delta_J}(\mathcal{O})$, where $\mathfrak{G}^0_{\Delta_J}$ is the neutral component of a group scheme $\mathfrak{G}_{\Delta_J}$, defined and smooth over $\mathcal{O}$, which has generic fiber the $F$-group $G$, and whose $\mathcal{O}$-points are the subgroup of $G(F)$ fixing the facet $\Delta_J$ of the Bruhat-Tits building corresponding to the set $J$; see [**BT2**], p. 356. By [**T**], 3.4.1 (see also [**BT2**], 1.7) we can characterize the group scheme $\mathfrak{G}_{\Delta_J}$ as follows: it is the unique $\mathcal{O}$-group scheme $\mathcal{P}$ satisfying the following three properties:

(1) $\mathcal{P}$ is smooth and affine over $\mathcal{O}$;
(2) The generic fiber $\mathcal{P}_F$ is $G_F$;
(3) For any unramified extension $F'$ of $F$, letting $\mathcal{O}_{F'} \subset F'$ denote ring of integers, the group $\mathcal{P}(\mathcal{O}_{F'}) \subset G(F')$ is the subgroup of elements which fix the facet $\Delta_J$ in the Bruhat-Tits building of $G_{F'}$.

Let us show how automorphism groups of periodic lattice chains $\Lambda_\bullet$ give a concrete realization of the Bruhat-Tits parahoric group schemes, in the $\mathrm{GL}_n$ and $\mathrm{GSp}_{2n}$ cases. We will discuss the Iwahori subgroups of $\mathrm{GL}_n$ in some detail, leaving for the reader the obvious generalizations to parahoric subgroups of $\mathrm{GL}_n$ (and $\mathrm{GSp}_{2n}$).

For any $\mathcal{O}$-algebra $R$, we may consider the diagram $\Lambda_{\bullet,R} = \Lambda_\bullet \otimes_{\mathcal{O}} R$, and we may define the $\mathcal{O}$-group scheme Aut whose $R$-points are the isomorphisms of the diagram $\Lambda_{\bullet,R}$. More precisely, an element of $\mathrm{Aut}(R)$ is an $n$-tuple of $R$-linear automorphisms

$$(g_0, \ldots, g_{n-1}) \in \mathrm{Aut}(\Lambda_{0,R}) \times \cdots \times \mathrm{Aut}(\Lambda_{n-1,R})$$

such that the following diagram commutes

$$\begin{array}{ccccccc}
\Lambda_{0,R} & \longrightarrow & \cdots & \longrightarrow & \Lambda_{n-1,R} & \longrightarrow & \Lambda_{n,R} \\
\downarrow g_0 & & & & \downarrow g_{n-1} & & \downarrow g_0 \\
\Lambda_{0,R} & \longrightarrow & \cdots & \longrightarrow & \Lambda_{n-1,R} & \longrightarrow & \Lambda_{n,R}.
\end{array}$$

The group functor Aut is obviously represented by an affine group scheme, also denoted Aut. Further, it is not hard to see that Aut is formally smooth, hence smooth, over $\mathcal{O}$. To show this, one has to check the lifting criterion for formal smoothness: if $R$ is an $\mathcal{O}$-algebra containing a nilpotent ideal $\mathcal{J} \subset R$, then any automorphism of $\Lambda_\bullet \otimes_{\mathcal{O}} R/\mathcal{J}$ can be lifted to an automorphism of $\Lambda_\bullet \otimes_{\mathcal{O}} R$. This is proved on page 135 of [**RZ**]. Thus Aut satisfies condition (1) above.

Alcoves in the Bruhat-Tits building for $\mathrm{GL}_n$ over $F$ can be described as complete periodic $\mathcal{O}$-lattice chains in $F^n$

$$\cdots \longrightarrow \mathcal{L}_0 \longrightarrow \mathcal{L}_1 \longrightarrow \cdots \longrightarrow \mathcal{L}_{n-1} \longrightarrow \mathcal{L}_n = \pi^{-1}\mathcal{L}_0 \longrightarrow \cdots$$

where the arrows are inclusions. We can regard $\Lambda_\bullet$ as the base alcove in this building. It is clear that since $\pi$ is invertible over the generic fiber $F$, the automorphism $g_0$ determines the other $g_i$'s over $F$ and so $\mathrm{Aut}_F = \mathrm{GL}_n$. By construction we have $\mathrm{Aut}(\mathcal{O}) = \mathrm{Stab}_{\mathrm{GL}_n(F)}(\Lambda_\bullet)$. This is unchanged if we replace $F$ by an unramified extension $F'$. It follows that Aut satisfies conditions (2) and (3) above. Thus, by uniqueness, $\mathrm{Aut} = \mathfrak{G}_{\Delta_J}$ for $J = \{0, \dots, n-1\}$.

Further, one can check that the special fiber $\mathrm{Aut}_k$ is an extension of the Borel subgroup $B_k$ by a connected unipotent group over $k$; hence the special fiber is connected. It follows that Aut is a connected group scheme (cf. [**BT2**], 1.2.12). So in this case $\mathrm{Aut} = \mathfrak{G}_{\Delta_J} = \mathfrak{G}^0_{\Delta_J}$. We conclude that $\mathrm{Aut}(\mathcal{O})$ is the Bruhat-Tits Iwahori subgroup fixing the base alcove $\Lambda_\bullet$.

Exercise 3.1. 1) Check the lifting criterion which shows Aut is formally smooth directly for the case $n = 2$, by explicit calculations with $2 \times 2$ matrices.
2) By identifying $\mathrm{Aut}(\mathcal{O})$ with its image in $\mathrm{GL}_n(F)$ under the inclusion $g_\bullet \mapsto g_0$, show that the Iwahori subgroup is the preimage of $B(k)$ under the canonical surjection $\mathrm{GL}_n(\mathcal{O}) \to \mathrm{GL}_n(k)$.
3) Prove that $\mathrm{Aut}_k \to \mathrm{Aut}(\Lambda_{0,k})$, $g_\bullet \mapsto g_0$, has image $B_k$, and kernel a connected unipotent group.

## 4. Local models

Given a certain triple $(G, \mu, K_p)$ consisting of a $\mathbb{Z}_p$-group $G$, a minuscule coweight $\mu$ for $G$, and a parahoric subgroup $K_p \subset G(\mathbb{Q}_p)$, one may construct a projective $\mathbb{Z}_p$-scheme $\mathbf{M}^{\mathrm{loc}}$ which (étale locally) models the singularities found in the special fiber of a certain Shimura variety with parahoric-level structure at $p$. The advantage of $\mathbf{M}^{\mathrm{loc}}$ is that it is defined in terms of linear algebra and is therefore easier to study than the Shimura variety itself. These schemes are called "local models", or sometimes "Rapoport-Zink local models"; the most general treatment is given in [**RZ**], but in special cases they were also investigated in [**DP**] and [**deJ**].

In this section we recall the definitions of local models associated to $\mathrm{GL}_n$ and $\mathrm{GSp}_{2n}$. For simplicity, we limit the discussion to models for Iwahori-level structure. In each case, the local model is naturally associated to a dominant minuscule coweight $\mu$, which we shall always mention. In fact, it turns out that if the Shimura data give rise to $(G,\mu)$, then the Rapoport-Zink local model $\mathbf{M}^{\mathrm{loc}}$ (for Iwahori-level structure) can be identified with the space $M_{-w_0\mu}$ mentioned in §2. See Remark 4.1 below.

**4.1. Linear case.** We use the notation $\Lambda_\bullet$ from section 3 to denote the "standard" lattice chain over $\mathcal{O} = \mathbb{Z}_p$ here.

Fix an integer $d$ with $1 \le d \le n-1$. We define the scheme $\mathbf{M}^{\mathrm{loc}}$ by defining its $R$-points for any $\mathbb{Z}_p$-algebra $R$ as the set of isomorphism classes of commutative diagrams

$$\begin{tikzcd}
\Lambda_{0,R} \arrow[r] & \cdots \arrow[r] & \Lambda_{n-1,R} \arrow[r] & \Lambda_{n,R} \arrow[r, "p"] & \Lambda_{0,R} \\
\mathcal{F}_0 \arrow[r] \arrow[u] & \cdots \arrow[r] & \mathcal{F}_{n-1} \arrow[r] \arrow[u] & \mathcal{F}_n \arrow[r, "p"] \arrow[u] & \mathcal{F}_0 \arrow[u]
\end{tikzcd}$$

where the vertical arrows are inclusions, and each $\mathcal{F}_i$ is an $R$-submodule of $\Lambda_{i,R}$ which is Zariski-locally on $R$ a direct factor of corank $d$. It turns out that this is identical to the space $M_{-w_0\mu}$ of §2, where $\mu = (0^{n-d}, (-1)^d)$. It is clear that $\mathbf{M}^{\mathrm{loc}}$ is represented by a closed subscheme of a product of Grassmannians over $\mathbb{Z}_p$, hence it is a projective $\mathbb{Z}_p$-scheme. One can also formulate the moduli problem using quotients of rank $d$ instead of submodules of corank $d$.

Another way to formulate the same moduli problem which is sometimes useful (see [**HN1**]) is given by adding an indeterminate $t$ to the story (following a suggestion of G. Laumon). We replace the "standard" lattice chain term $\Lambda_{i,R}$ with $\mathcal{V}_{i,R} := \alpha^{-i}R[t]^n$, where $\alpha$ is the $n \times n$ matrix

$$\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ t+p & & & 0 \end{pmatrix} \in \mathrm{GL}_n(R[t, t^{-1}, (t+p)^{-1}]).$$

One can identify $\mathbf{M}^{\mathrm{loc}}(R)$ with the set of chains

$$\mathcal{L}_\bullet = (\mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_n = (t+p)^{-1}\mathcal{L}_0)$$

of $R[t]$-submodules of $R[t,t^{-1},(t+p)^{-1}]^n$ satisfying the following properties

(1) for all $i = 0, \ldots, n-1$, we have $t\mathcal{V}_{i,R} \subset \mathcal{L}_i \subset \mathcal{V}_{i,R}$;
(2) as an $R$-module, $\mathcal{L}_i/t\mathcal{V}_{i,R}$ is locally a direct factor of $\mathcal{V}_{i,R}/t\mathcal{V}_{i,R}$ of corank $d$.

Remark 4.1. With the second definition, it is easy to see that the geometric generic fiber $\mathbf{M}^{\mathrm{loc}}_{\overline{\mathbb{Q}}_p}$ is contained in the affine Grassmannian $\mathrm{GL}_n(\overline{\mathbb{Q}}_p((t)))/\mathrm{GL}_n(\overline{\mathbb{Q}}_p[[t]])$, and the geometric special fiber $\mathbf{M}^{\mathrm{loc}}_{\overline{\mathbb{F}}_p}$ is contained in the affine flag variety $\mathrm{GL}_n(\overline{\mathbb{F}}_p((t)))/I_{\overline{\mathbb{F}}_p}$, where $I_{\overline{\mathbb{F}}_p} := \mathrm{Aut}(\overline{\mathbb{F}}_p[[t]])$ is the Iwahori subgroup of $\mathrm{GL}_n(\overline{\mathbb{F}}_p[[t]])$ corresponding to the upper triangular Borel subgroup $B \subset \mathrm{GL}_n$. Moreover, it is possible to view $\mathbf{M}^{\mathrm{loc}}$ as a piece of a $\mathbb{Z}_p$-ind-scheme $M$ which forms a deformation of the affine Grassmannian to the affine flag variety over the base $\mathrm{Spec}(\mathbb{Z}_p)$, in analogy with Beilinson's deformation $\mathrm{Fl}_X$ over a smooth $\mathbb{F}_p$-curve $X$ in the function field case (cf. [**Ga**], [**HN1**]). In fact, letting $e_0$ denote the base point in the affine Grassmannian $\mathrm{GL}_n(\mathbb{Q}_p((t)))/\mathrm{GL}_n(\mathbb{Q}_p[[t]])$, and for $\lambda \in X_+(T)$, letting $\mathcal{Q}_\lambda$ denote the $\mathrm{GL}_n(\mathbb{Q}_p[[t]])$-orbit of the point $\lambda(t)e_0$, it turns out that $\mathbf{M}^{\mathrm{loc}}$ coincides with the scheme-theoretic closure $M_{-w_0\mu}$ of $\mathcal{Q}_{-w_0\mu}$ taken in the model $M$. A similar statement holds in the symplectic case.

The identification $\mathbf{M}^{\rm loc} = M_{-w_0\mu}$ is explained in §8. It is a consequence of the determinant condition and the "homology" definition of our local models; also the flatness of $\mathbf{M}^{\rm loc}$ (see §7) plays a role.

**4.2. The symplectic case.** Recall that for our group $\mathrm{GSp}_{2n} = \mathrm{GSp}(V, (\cdot,\cdot))$, we have an identification of $X_*(T)$ with the lattice $\{(a_1, \dots, a_n, b_n, \dots, b_1) \in \mathbb{Z}^{2n} \mid \exists\, c \in \mathbb{Z},\ a_i + b_i = c, \forall i\}$. The Shimura coweight that arises here has the form $\mu = (0^n, (-1)^n)$.

For the group $\mathrm{GSp}_{2n}$, the symbol $\Lambda_\bullet$ now denotes the self-dual $\mathbb{Z}_p$-lattice chain in $\mathbb{Q}_p^{2n}$, discussed in section 3 in the context of $\mathrm{GSp}_{2n}$. Let $(\cdot,\cdot)$ denote the alternating pairing on $\mathbb{Z}_p^{2n}$ discussed in that section, and let the dual $\Lambda^\perp$ of a lattice $\Lambda$ be defined using $(\cdot,\cdot)$. As above, there are (at least) two equivalent ways to define $\mathbf{M}^{\rm loc}(R)$ for a $\mathbb{Z}_p$-algebra $R$. We define $\mathbf{M}^{\rm loc}(R)$ to be the set of isomorphism classes of diagrams

$$\begin{array}{ccccccc}
\Lambda_{0,R} & \longrightarrow & \cdots & \longrightarrow & \Lambda_{n-1,R} & \longrightarrow & \Lambda_{n,R} \\
\uparrow & & & & \uparrow & & \uparrow \\
\mathcal{F}_0 & \longrightarrow & \cdots & \longrightarrow & \mathcal{F}_{n-1} & \longrightarrow & \mathcal{F}_n
\end{array}$$

where the vertical arrows are inclusions of $R$-submodules with the following properties:

(1) for $i = 0, \dots, n$, Zariski-locally on $R$ the submodule $\mathcal{F}_i$ is a direct factor of $\Lambda_{i,R}$ of corank $n$;
(2) $\mathcal{F}_0$ is isotropic with respect to $(\cdot,\cdot)$ and $\mathcal{F}_n$ is isotropic with respect to $p(\cdot,\cdot)$.

As in the linear case, this can be described in a way more transparently connected to affine flag varieties. In this case, $\mathcal{V}_{i,R}$ has the same meaning as in the linear case, except that the ambient space is now $R[t, t^{-1}, (t+p)^{-1}]^{2n}$. We may describe $\mathbf{M}^{\rm loc}(R)$ as the set of chains

$$\mathcal{L}_\bullet = (\mathcal{L}_0 \subset \mathcal{L}_1 \subset \cdots \subset \mathcal{L}_n)$$

of $R[t]$-submodules of $R[t, t^{-1}, (t+p)^{-1}]^{2n}$ satisfying the following properties

(1) for $i = 0, 1, \dots, n$, $t\mathcal{V}_{i,R} \subset \mathcal{L}_i \subset \mathcal{V}_{i,R}$;
(2) as $R$-modules, $\mathcal{L}_i / t\mathcal{V}_{i,R}$ is locally a direct factor of $\mathcal{V}_{i,R}/t\mathcal{V}_{i,R}$ of corank $n$;
(3) $\mathcal{L}_0$ is self-dual with respect to $t^{-1}(\cdot,\cdot)$, and $\mathcal{L}_n$ is self-dual with respect to $t^{-1}(t+p)(\cdot,\cdot)$.

**4.3. Generic and special fibers.** In the linear case with $\mu = (0^{n-d}, (-1)^d)$, the generic fiber of $M_{-w_0\mu}$ is the Grassmannian $\mathrm{Gr}(d,n)$ of $d$-planes in $\mathbb{Q}_p^n$. In the symplectic case with $\mu = (0^n, (-1)^n)$, the generic fiber of $M_{-w_0\mu}$ is the Grassmannian of isotropic $n$-planes in $\mathbb{Q}_p^{2n}$ with respect to the alternating pairing $(\cdot,\cdot)$.

In each case, the special fiber is a union of finitely many Iwahori-orbits $IwI/I$ in the affine flag variety, indexed by elements $w$ in the extended affine Weyl group $\widetilde{W}$ for $\mathrm{GL}_n$ (resp. $\mathrm{GSp}_{2n}$) ranging over the so-called $-w_0\mu$-*admissible* subset $\mathrm{Adm}(-w_0\mu)$, [**KR**]. Let $\lambda \in X_+(T)$. Then by definition

$$\mathrm{Adm}(\lambda) = \{w \in \widetilde{W} \mid \exists\, \nu \in W\lambda, \text{ such that } w \le t_\nu\}.$$

Here, $W\lambda$ is the set of conjugates of $\lambda$ under the action of the finite Weyl group $W_0$, and $t_\nu \in \widetilde{W}$ is the translation element corresponding to $\nu$, and $\leq$ denotes the Bruhat order on $\widetilde{W}$. Actually (see §8.1), the set that arises naturally from the moduli problem is the $-w_0\mu$*-permissible* subset $\mathrm{Perm}(-w_0\mu) \subset \widetilde{W}$ from [**KR**]. Let us recall the definition of this set, following loc. cit. Let $\lambda \in X_+(T)$ and suppose $t_\lambda \in W_{\mathrm{aff}}\tau$, for $\tau \in \Omega$. Then $\mathrm{Perm}(\lambda)$ consists of the elements $x \in W_{\mathrm{aff}}\tau$ such that $x(a) - a \in \mathrm{Conv}(\lambda)$ for every vertex $a \in \mathbf{a}$. Here $\mathrm{Conv}(\lambda)$ denotes the convex hull of $W\lambda$ in $X_*(T) \otimes \mathbb{R}$.

The strata in the special fiber of $\mathbf{M}^{\mathrm{loc}} = M_{-w_0\mu}$ are naturally indexed by the set $\mathrm{Perm}(-w_0\mu)$, which agrees with $\mathrm{Adm}(-w_0\mu)$ by the following non-trivial combinatorial theorem due to Kottwitz and Rapoport.

THEOREM 4.2 ([**KR**]; see also [**HN2**]). *For every minuscule coweight $\lambda$ of either $\mathrm{GL}_n$ or $\mathrm{GSp}_{2n}$, we have the equality*

$$\mathrm{Perm}(\lambda) = \mathrm{Adm}(\lambda).$$

Using the well-known correspondence between elements in the affine Weyl group and the set of alcoves in the standard apartment of the Bruhat-Tits building, one can "draw" pictures of $\mathrm{Adm}(\mu)$ for low-rank groups. Figures 1 and 2 depict this set for $G = \mathrm{GL}_3$, $\mu = (-1,0,0)$, and $G = \mathrm{GSp}_4$, $\mu = (-1,-1,0,0)$ [2]. Actually, we draw the image of $\mathrm{Adm}(\mu)$ in the apartment for $\mathrm{PGL}_3$ (resp. $\mathrm{PGSp}_4$); the base alcove is labeled by $\tau$.

**4.4. Computing the singularities in the special fiber of $\mathbf{M}^{\mathrm{loc}}$.** In certain cases, the singularities in $\mathbf{M}^{\mathrm{loc}}_{\mathbb{F}_p}$ can be analyzed directly by writing down equations. As the simplest example of how this is done, we analyze the local model for $\mathrm{GL}_2$, $\mu = (0,-1)$. For a $\mathbb{Z}_p$-algebra $R$, we are looking at the set of pairs $(\mathcal{F}_0, \mathcal{F}_1)$ of locally free rank 1 $R$-submodules of $R^2$ such that the following diagram commutes

$$\begin{tikzcd}
R \oplus R \arrow[r, "\begin{bmatrix} p & 0 \\ 0 & 1 \end{bmatrix}"] & R \oplus R \arrow[r, "\begin{bmatrix} 1 & 0 \\ 0 & p \end{bmatrix}"] & R \oplus R \\
\mathcal{F}_0 \arrow[u] \arrow[r] & \mathcal{F}_1 \arrow[u] \arrow[r] & \mathcal{F}_0 \arrow[u]
\end{tikzcd}$$

Obviously this functor is represented by a certain closed subscheme of $\mathbb{P}^1_{\mathbb{Z}_p} \times \mathbb{P}^1_{\mathbb{Z}_p}$. Locally around a fixed point $(\mathcal{F}_0, \mathcal{F}_1) \in \mathbb{P}^1(R) \times \mathbb{P}^1(R)$ we choose coordinates such that $\mathcal{F}_0$ is represented by the homogeneous column vector $[1 : x]^t$ and $\mathcal{F}_1$ by the vector $[y : 1]^t$, for $x, y \in R$. We see that $(\mathcal{F}_0, \mathcal{F}_1) \in \mathbf{M}^{\mathrm{loc}}(R)$ if and only if

$$xy = p,$$

so $\mathbf{M}^{\mathrm{loc}}$ is locally the same as $\mathrm{Spec}(\mathbb{Z}_p[X,Y]/(XY - p)$, the usual deformation of $\mathbb{A}^1_{\mathbb{Q}_p}$ to a union of two $\mathbb{A}^1_{\mathbb{F}_p}$'s which intersect transversally at a point. Indeed, $\mathbf{M}^{\mathrm{loc}}$ is globally this kind of deformation:

- In the generic fiber, the matrices are invertible and so $\mathcal{F}_0$ uniquely determines $\mathcal{F}_1$; thus $\mathbf{M}^{\mathrm{loc}}_{\mathbb{Q}_p} \cong \mathbb{P}^1_{\mathbb{Q}_p}$;

[2] Note that in Figure 2 there is an alcove of length one contained in the Bruhat-closure of all four distinct translations. This already tells us something about the singularities: the special fiber of the Siegel variety for $\mathrm{GSp}_4$ is not a union of divisors with normal crossings; see §8.

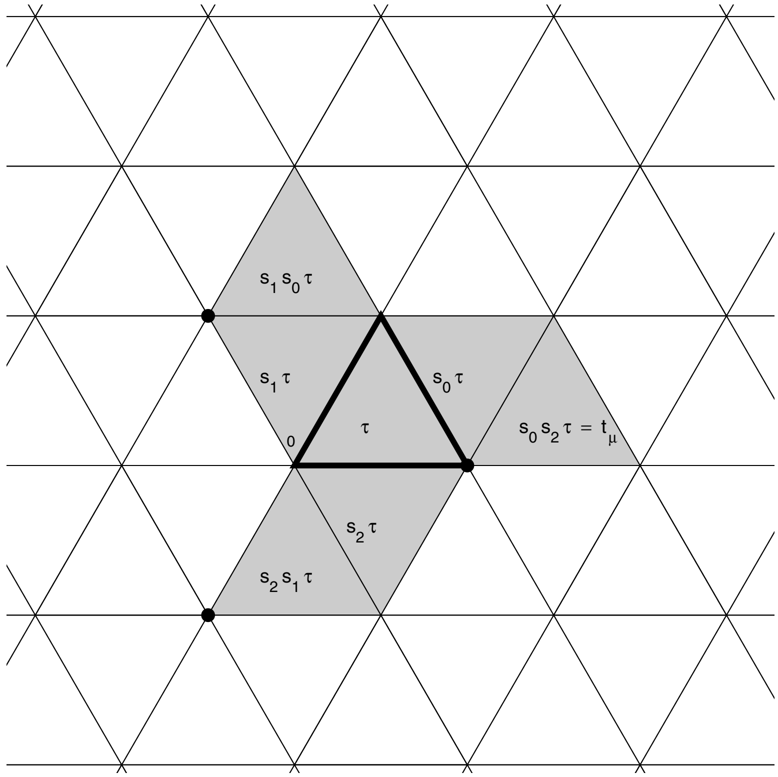



FIGURE 1. The admissible alcoves $\mathrm{Adm}(\mu)$ for $\mathrm{GL}_3$, $\mu = (-1,0,0)$. The base alcove is labeled by $\tau$.

- In the special fiber, $p = 0$ and one can check that $\mathbf{M}^{\mathrm{loc}}_{\mathbb{F}_p}$ is the union of the closures of two Iwahori-orbits in the affine flag variety $\mathrm{GL}_2(\mathbb{F}_p((t)))/I_{\mathbb{F}_p}$, each of dimension 1, which meet in a point. Thus $\mathbf{M}^{\mathrm{loc}}_{\mathbb{F}_p}$ is the union of two $\mathbb{P}^1_{\mathbb{F}_p}$'s meeting in a point.

We refer to the work of U. Görtz for many more complicated calculations of this kind: [**Go1**], [**Go2**], [**Go3**], [**Go4**].

## 5. Some PEL Shimura varieties with parahoric level structure at $p$

**5.1. PEL-type data.** Given a Shimura datum $(\mathbf{G}, \{h\}, \mathbf{K})$ one can construct a Shimura variety $Sh(\mathbf{G}, h)_{\mathbf{K}}$ which has a canonical model over the reflex field $\mathbf{E}$, a number field determined by the datum. We write $G$ for the $p$-adic group $\mathbf{G}_{\mathbb{Q}_p}$. Let us assume that the compact open subgroup $\mathbf{K} \subset \mathbf{G}(\mathbb{A}_f)$ is of the form $\mathbf{K} = K^p K_p$, where $K^p \subset \mathbf{G}(\mathbb{A}_f^p)$ is a sufficiently small compact open subgroup, and $K_p \subset G(\mathbb{Q}_p)$ is a parahoric subgroup.

Let us fix once and for all embeddings $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$, and $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$. We denote by $\mathfrak{p}$ the corresponding place of $\mathbf{E}$ over $p$ and by $E = \mathbf{E}_{\mathfrak{p}}$ the completion of $\mathbf{E}$ at $\mathfrak{p}$.

If the Shimura datum comes from PEL-type data, then it is possible to define a moduli problem (in terms of chains of abelian varieties with additional structure)

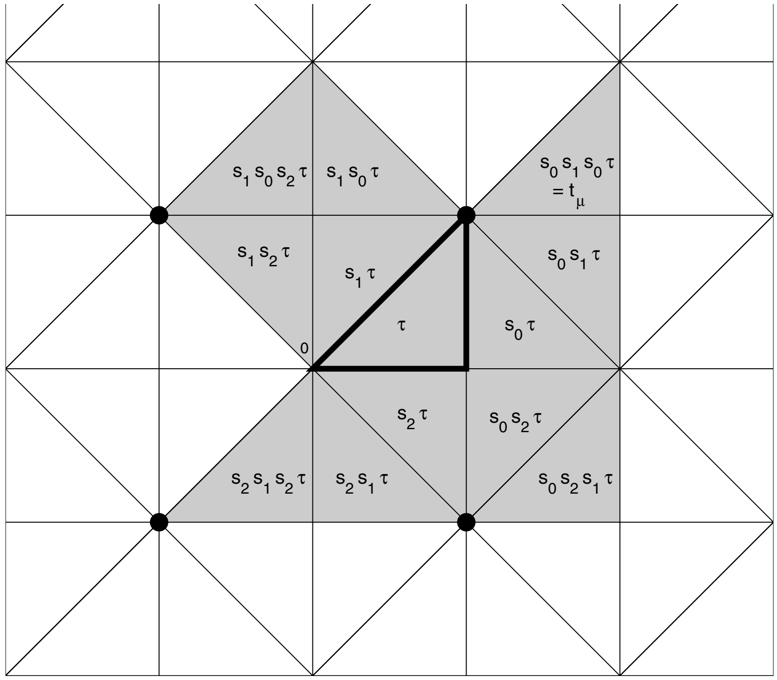



FIGURE 2. The admissible alcoves $\mathrm{Adm}(\mu)$ for $\mathrm{GSp}_4$, $\mu = (-1, -1, 0, 0)$.

over the ring $\mathcal{O}_E$. This moduli problem is representable by a quasi-projective $\mathcal{O}_E$-scheme whose generic fiber is the base-change to $E$ of the initial Shimura variety $Sh(\mathbf{G}, h)_{\mathbf{K}}$ (or at least a finite union of Shimura varieties, one of which is the canonical model $Sh(\mathbf{G}, h)_{\mathbf{K}}$). This is done in great generality in Chapter 6 of [**RZ**]. Our aim in this section is only to make somewhat more explicit the definitions in loc. cit., in two very special cases attached to the linear and symplectic groups.

First, let us recall briefly PEL-type data. Let $B$ denote a finite-dimensional semi-simple $\mathbb{Q}$-algebra with positive involution $\iota$. Let $V \neq 0$ be a finitely-generated left $B$-module, and let $(\cdot, \cdot)$ be a non-degenerate alternating form $V \times V \to \mathbb{Q}$ on the underlying $\mathbb{Q}$-vector space, such that $(bv, w) = (v, b^\iota w)$, for $b \in B$, $v, w \in V$. The form $(\cdot, \cdot)$ determines a "transpose" involution on $\mathrm{End}(V)$, denoted by $*$ (so viewing the left-action of $b$ as an element of $\mathrm{End}(V)$, we have $b^\iota = b^*$). We denote by $\mathbf{G}$ the $\mathbb{Q}$-group whose points in a $\mathbb{Q}$-algebra $R$ are

$$\{g \in \mathrm{GL}_{B\otimes R}(V \otimes R) \mid g^* g = c(g) \in R^\times\}.$$

We assume $\mathbf{G}$ is a *connected* reductive group; this means we are excluding the orthogonal case. Consider the $\mathbb{R}$-algebra $C := \mathrm{End}_B(V) \otimes \mathbb{R}$. We let $h_0 : \mathbb{C} \to C$ denote an $\mathbb{R}$-algebra homomorphism satisfying $h_0(\overline{z}) = h_0(z)^*$, for $z \in \mathbb{C}$. We fix a choice of $i = \sqrt{-1}$ in $\mathbb{C}$ once and for all, and we assume the symmetric bilinear form $(\cdot\,, h_0(i)\,\cdot) : V_{\mathbb{R}} \times V_{\mathbb{R}} \to \mathbb{R}$ is positive definite. Let $h$ denote the inverse of the restriction of $h_0$ to $\mathbb{C}^\times$. Then $h$ induces an algebraic homomorphism

$$h : \mathbb{C}^\times \to \mathbf{G}(\mathbb{R})$$

of real groups which defines on $V_{\mathbb{R}}$ a Hodge structure of type $(1,0)+(0,1)$ (in the terminology of [**Del2**], section 1) and which satisfies the usual Riemann conditions with respect to $(\cdot,\cdot)$ (see [**Ko92**], Lemma 4.1). For any choice of (sufficiently small) compact open subgroup $\mathbf{K}$, the data $(\mathbf{G}, h, \mathbf{K})$ determine a (smooth) Shimura variety over a reflex field $\mathbf{E}$; cf. [**Del**].

We recall that $h$ gives rise to a minuscule coweight

$$\mu := \mu_h : \mathbb{G}_{m,\mathbb{C}} \to \mathbf{G}_{\mathbb{C}}$$

as follows: the complexification of the real group $\mathbb{C}^\times$ is the torus $\mathbb{C}^\times \times \mathbb{C}^\times$, the factors being indexed by the two $\mathbb{R}$-algebra automorphisms of $\mathbb{C}$; we assume the first factor corresponds to the identity and the second to complex conjugation. Then we define

$$\mu(z) := h_{\mathbb{C}}(z,1).$$

By definition of Shimura data, the homomorphism $h : \mathbb{C}^\times \to \mathbf{G}_{\mathbb{R}}$ is only specified up to $\mathbf{G}(\mathbb{R})$-conjugation, and therefore $\mu$ is only well-defined up to $\mathbf{G}(\mathbb{C})$-conjugation. However, this conjugacy class is at least defined over the reflex field $\mathbf{E}$ (in fact we *define* $\mathbf{E}$ as the field of definition of the conjugacy class of $\mu$). Via our choice of field embeddings $\overline{\mathbb{C}} \hookleftarrow \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$, we get a well-defined conjugacy class of minuscule coweights

$$\mu : \mathbb{G}_{m,\overline{\mathbb{Q}}_p} \to G_{\overline{\mathbb{Q}}_p},$$

which is defined over $E$.

The argument of [**Ko84**], Lemma (1.1.3) shows that $E$ is contained in any subfield of $\overline{\mathbb{Q}}_p$ which splits $G$. Therefore, when $G$ is split over $\mathbb{Q}_p$ (the case of interest in this report), it follows that $E = \mathbb{Q}_p$ and the conjugacy class of $\mu$ contains a $\mathbb{Q}_p$-rational and $B$-dominant element, usually denoted also by the symbol $\mu$. It is this same $\mu$ which was mentioned in the definitions of local models in section 4.

For use in the definition to follow, we decompose the $B_{\mathbb{C}}$-module $V_{\mathbb{C}}$ as $V_{\mathbb{C}} = V_1 \oplus V_2$, where $h_0(z)$ acts by $z$ on $V_1$ and by $\overline{z}$ on $V_2$, for $z \in \mathbb{C}$. Our conventions imply that $\mu(z)$ acts by $z^{-1}$ on $V_1$ and by 1 on $V_2$ ($z \in \mathbb{C}^\times$). We choose $E' \subset \overline{\mathbb{Q}}_p$ a finite extension field $E' \supset E$ over which this decomposition is defined:

$$V_{E'} = V_1 \oplus V_2.$$

(We are implicitly using the diagram $\mathbb{C} \hookleftarrow \overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ to make sense of this.)

Recall that we are interested in defining an $\mathcal{O}_E$-integral model for $Sh(\mathbf{G}, h)_{\mathbf{K}}$ in the case where $K_p \subset \mathbf{G}(\mathbb{Q}_p)$ is a parahoric (more specifically, an Iwahori) subgroup.

To define an integral model over $\mathcal{O}_E$, we need to specify certain additional data. We suppose $\mathcal{O}_B$ is a $\mathbb{Z}_{(p)}$-order in $B$ whose $p$-adic completion $\mathcal{O}_B \otimes \mathbb{Z}_p$ is a maximal order in $B_{\mathbb{Q}_p}$, stable under the involution $\iota$. Using the terminology of [**RZ**], 6.2, we assume we are given a self-dual multichain $\mathcal{L}$ of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices in $V_{\mathbb{Q}_p}$ (the notion of multichain $\mathcal{L}$ is a generalization of the lattice chain $\Lambda_\bullet$ appearing in section 4; specifying $\mathcal{L}$ is equivalent to specifying a parahoric subgroup, namely $K_p := \mathrm{Aut}(\mathcal{L})$, of $\mathbf{G}(\mathbb{Q}_p)$). We can then give the definition of a model $Sh_{K^p}$ that depends on the above data and the choice of a small compact open subgroup $K^p$ [3].

[3]In the sequel, we sometimes drop the subscript $K^p$ on $Sh_{K^p}$, or replace it with the subscript $K_p$, depending on whether $K^p$, or $K_p$ (or both) is understood.

DEFINITION 5.1. A point of the functor $Sh_{K^p}$ with values in the $\mathcal{O}_E$-scheme $S$ is given by the following set of data up to isomorphism [4].

(1) An $\mathcal{L}$-set of abelian $S$-schemes $A = \{A_\Lambda\}, \ \Lambda \in \mathcal{L}$, compatibly endowed with an action of $\mathcal{O}_B$:

$$i : \mathcal{O}_B \otimes \mathbb{Z}_{(p)} \to \mathrm{End}(A) \otimes \mathbb{Z}_{(p)};$$

(2) A $\mathbb{Q}$-homogeneous principal polarization $\overline{\lambda}$ of the $\mathcal{L}$-set $A$;
(3) A $K^p$-level structure

$$\bar{\eta} : V \otimes \mathbb{A}_f^p \cong H_1(A, \mathbb{A}_f^p) \mod K^p$$

that respects the bilinear forms on both sides up to a scalar in $(\mathbb{A}_f^p)^\times$, and commutes with the $B = \mathcal{O}_B \otimes \mathbb{Q}$-actions.

We impose the condition that under

$$i : \mathcal{O}_B \otimes \mathbb{Z}_{(p)} \to \mathrm{End}(A) \otimes \mathbb{Z}_{(p)},$$

we have $i(b^\iota) = \lambda^{-1} \circ (i(b))^\vee \circ \lambda$; in other words, $i$ intertwines $\iota$ and the Rosati involution on $\mathrm{End}(A) \otimes \mathbb{Z}_{(p)}$ determined by $\overline{\lambda}$. In addition, we impose the following determinant condition: for each $b \in \mathcal{O}_B$ and $\Lambda \in \mathcal{L}$:

$$\det_{\mathcal{O}_S}(b, \mathrm{Lie}(A_\Lambda)) = \det_{E'}(b, V_1).$$

We will not explain all the notions entering this definition; we refer to loc. cit., Chapter 6 as well as [**Ko92**], section 5, for complete details. However, in the simple examples we make explicit below, these notions will be made concrete and their importance will be highlighted. For example, an $\mathcal{L}$-set of abelian varieties $\{A_\Lambda\}$ comes with a family of "periodicity isomorphisms"

$$\theta_a : A_\Lambda^a \to A_{a\Lambda},$$

see [**RZ**], Def. 6.5, and we will describe these explicitly in the examples to follow.

Note that one can see from this definition why some of the conditions on PEL data are imposed. For example, since the Rosati involution is always positive (see [**Mu**], section 21), we see that the involution $\iota$ on $B$ must be positive for the moduli problem to be non-empty.

**5.2. Some "fake" unitary Shimura varieties.** This section concerns the so-called "simple" or "fake unitary" Shimura varieties investigated by Kottwitz in [**Ko92b**]. They are indeed "simple" in the sense that they are compact Shimura varieties for which there are no problems due to endoscopy (see loc. cit.).

Kottwitz made assumptions ensuring that the local group $\mathbf{G}_{\mathbb{Q}_p}$ be unramified, and that the level structure at $p$ be given by a hyperspecial maximal compact subgroup. Here we will work in a situation where $\mathbf{G}_{\mathbb{Q}_p}$ is *split*, but we only impose parahoric-level structure at $p$. For simplicity, we explain only the case $F_0 = \mathbb{Q}$ (notation of loc. cit.).

[4] We say $\{A_\Lambda\}$ is isomorphic to $\{A'_\Lambda\}$ if there is a compatible family of prime-to-$p$ isogenies $A_\Lambda \to A'_\Lambda$ which preserve all the structures.

5.2.1. *The group-theoretic set-up.* Let $F$ be an imaginary quadratic extension of $\mathbb{Q}$, and let $(D, *)$ be a division algebra with center $F$, of dimension $n^2$ over $F$, together with an involution $*$ which induces on $F$ the non-trivial element of $\mathrm{Gal}(F/\mathbb{Q})$. Let $\mathbf{G}$ be the $\mathbb{Q}$-group whose points in a commutative $\mathbb{Q}$-algebra $R$ are

$$\{x \in D \otimes_{\mathbb{Q}} R \mid x^* x \in R^\times\}.$$

The map $x \mapsto x^* x$ is a homomorphism of $\mathbb{Q}$-groups $\mathbf{G} \to \mathbb{G}_m$ whose kernel $\mathbf{G}_0$ is an inner form of a unitary group over $\mathbb{Q}$ associated to $F/\mathbb{Q}$. Let us suppose we are given an $\mathbb{R}$-algebra homomorphism

$$h_0 : \mathbb{C} \to D \otimes_{\mathbb{Q}} \mathbb{R}$$

such that $h_0(z)^* = h_0(\overline{z})$ and the involution $x \mapsto h_0(i)^{-1} x^* h_0(i)$ is positive.

Given the data $(D, *, h_0)$ above, we want to explain how to find the PEL-data $(B, \iota, V, (\cdot,\cdot), h_0)$ used in the definition of the scheme $Sh_{K^p}$.

Let $B = D^{opp}$ and let $V = D$ be viewed as a left $B$-module, free of rank 1, using right multiplications. Thus we can identify $C := \mathrm{End}_B(V)$ with $D$ (left multiplications). For $h_0 : \mathbb{C} \to C_{\mathbb{R}}$ we use the homomorphism $h_0 : \mathbb{C} \to D \otimes_{\mathbb{Q}} \mathbb{R}$ we are given.

Next, one can show that there exist elements $\xi \in D^\times$ such that $\xi^* = -\xi$ and the involution $x \mapsto \xi x^* \xi^{-1}$ is positive. To see this, note that the Skolem-Noether theorem implies that the involutions of the second type on $D$ are precisely the maps of the form $x \mapsto b x^* b^{-1}$, for $b \in D^\times$ such that $b(b^*)^{-1}$ lies in the center $F$. Since positive involutions of the second kind exist (see [**Mu**], p. 201-2), for some such $b$ the involution $x \mapsto b x^* b^{-1}$ is positive. We have $N_{F/\mathbb{Q}}(b(b^*)^{-1}) = 1$, so by Hilbert's Theorem 90, we may alter any such $b$ by an element in $F^\times$ so that $b^* = b$. There exists $\epsilon \in F^\times$ such that $\epsilon^* = -\epsilon$. We then put $\xi = \epsilon b$.

We define the positive involution $\iota$ by $x^\iota := \xi x^* \xi^{-1}$, for $x \in B = D^{opp}$.

Now we define the non-degenerate alternating pairing $(\cdot,\cdot) : D \times D \to \mathbb{Q}$ by

$$(x, y) = tr_{D/\mathbb{Q}}(x \xi y^*).$$

It is clear that $(bx, y) = (x, b^\iota y)$ for any $b \in B = D^{opp}$, remembering that the left action of $b$ is right multiplication by $b$. We also have $(h_0(z)\, x, y) = (x, h_0(\overline{z})\, y)$, since $h_0(z) \in D$ acts by left multiplication on $D$.

Finally, we claim that $(\cdot\,,\, h_0(i)\,\cdot)$ is always positive or negative definite; thus we can always arrange for it to be positive definite by replacing $\xi$ with $-\xi$ if necessary. To prove the definiteness, choose an isomorphism

$$D \otimes_{\mathbb{Q}} \mathbb{R} \xrightarrow{\sim} M_n(\mathbb{C})$$

such that the positive involution $x \mapsto x^\iota$ goes over to the standard positive involution $X \mapsto \overline{X}^t$ on $M_n(\mathbb{C})$. Let $H \in M_n(\mathbb{C})$ be the image of $\xi h_0(i)^{-1}$ under this isomorphism, so that the symmetric pairing $\langle x, y\rangle = (x\,,\, h_0(i) y)$ goes over to the pairing

$$\langle X, Y\rangle = tr_{M_n(\mathbb{C})/\mathbb{R}}(X\, \overline{Y}^t\, H).$$

We conclude by invoking the following exercise for the reader.

Exercise 5.2. The matrix $H$ is Hermitian and either positive or negative definite. If positive definite, we then have $tr(X\, \overline{X}^t\, H) > 0$ whenever $X \neq 0$.

(Hint: use the argument of [**Mu**], p. 200.)

5.2.2. *The minuscule coweight* $\mu$. How is the minuscule coweight $\mu$ described in terms of the above data? Recall our decomposition

$$D_{\mathbb{C}} = V_1 \oplus V_2.$$

The homomorphism $h_0$ makes $D_{\mathbb{C}}$ into a $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}$-module. Of course

$$\begin{aligned} \mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} &\xrightarrow{\sim} \mathbb{C} \times \mathbb{C} \\ z_1 \otimes z_2 &\mapsto (z_1 z_2, \overline{z}_1 z_2), \end{aligned}$$

which induces the above decomposition of $D_{\mathbb{C}}$ ($h_0(z_1 \otimes 1)$ acts by $z_1$ on $V_1$ and by $\overline{z}_1$ on $V_2$).

The factors $V_1$, $V_2$ are stable under right multiplications of

$$D_{\mathbb{C}} = D \otimes_{F,\nu} \mathbb{C} \times D \otimes_{F,\nu^*} \mathbb{C},$$

where $\nu, \nu^* : F \hookrightarrow \mathbb{C}$ are the two embeddings. (We may assume our fixed choice $\overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$ extends $\nu$.) Also $h_0(z \otimes 1)$ is the endomorphism given by left multiplication by a certain element of $D_{\mathbb{C}}$. We can choose an isomorphism

$$D \otimes_{F,\nu} \mathbb{C} \times D \otimes_{F,\nu^*} \mathbb{C} \cong M_n(\mathbb{C}) \times M_n(\mathbb{C})$$

such that $h_0(z \otimes 1)$ can be written explicitly as

$$h_0(z \otimes 1) = \mathrm{diag}(\overline{z}^{n-d}, z^d) \times \mathrm{diag}(z^{n-d}, \overline{z}^d),$$

for some integer $d$, $0 \le d \le n$. (One can then identify $V_1$ resp. $V_2$ as the span of certain columns of the two matrices.) We know that $\mu(z) = h_{\mathbb{C}}(z, 1)$ acts by $z^{-1}$ on $V_1$ and by 1 on $V_2$. Hence we can identify $\mu(z)$ as

$$\mu(z) = \mathrm{diag}(1^{n-d}, (z^{-1})^d) \times \mathrm{diag}((z^{-1})^{n-d}, 1^d).$$

We may label this by $(0^{n-d}, (-1)^d) \in \mathbb{Z}^n$, via the usual identification applied to the first factor.

Here is another way to interpret the number $d$. Let $W$ (resp. $W^*$) be the (unique up to isomorphism, $n$-dimensional) simple right-module for $D \otimes_{F,\nu} \mathbb{C}$ (resp. $D \otimes_{F,\nu^*} \mathbb{C}$). Then as right $D_{\mathbb{C}}$-modules we have

$$V_1 = W^d \oplus (W^*)^{n-d}, \;\; \text{resp. } V_2 = W^{n-d} \oplus (W^*)^d.$$

Finally, let us remark that if we choose the identification of $D \otimes_{\mathbb{Q}} \mathbb{R} = D \otimes_{F,\nu} \mathbb{C}$ with $M_n(\mathbb{C})$ in such a way that the positive involution $x \mapsto h_0(i)^{-1} x^* h_0(i)$ goes over to $X \mapsto \overline{X}^t$, then we get an isomorphism

$$G(\mathbb{R}) \;\cong\; \mathrm{GU}(d, n-d).$$

(See also [**Ko92b**], section 1.)

In applications, it is sometimes necessary to prescribe the value of $d$ ahead of time (with the additional constraint that $1 \le d \le n-1$). However, it can be a delicate matter to arrange things so that a prescribed value of $d$ is achieved. To see how this is done for the case of $d = 1$, see [**HT**], Lemma I.7.1.

5.2.3. *Assumptions on $p$ and integral data.* We first make some assumptions on the prime $p$ [5], and then we specify the integral data at $p$.

*First assumption on $p$*: The prime $p$ splits in $F$ as a product of distinct prime ideals

$$(p) = \mathfrak{p}\mathfrak{p}^*,$$

where $\mathfrak{p}$ is the prime determined by our fixed choice of embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$, and $\mathfrak{p}^*$ is its image under the non-trivial element of $\mathrm{Gal}(F/\mathbb{Q})$.

Under this assumption $F_{\mathfrak{p}} = F_{\mathfrak{p}^*} = \mathbb{Q}_p$. Further the algebra $D_{\mathbb{Q}_p}$ is a product

$$D \otimes \mathbb{Q}_p = D_{\mathfrak{p}} \times D_{\mathfrak{p}^*}$$

where each factor is a central simple $\mathbb{Q}_p$-algebra. We have $D_{\mathfrak{p}} \xrightarrow{\sim} D_{\mathfrak{p}^*}^{opp}$ via $*$. Therefore for any $\mathbb{Q}_p$-algebra $R$ we can identify the group $G(R)$ with the group

$$\{(x_1, x_2) \in (D_{\mathfrak{p}} \otimes R)^\times \times (D_{\mathfrak{p}^*} \otimes R)^\times \mid x_1 = c(x_2^*)^{-1}, \text{ for some } c \in R^\times\}.$$

Therefore there is an isomorphism of $\mathbb{Q}_p$-groups $G \cong D_{\mathfrak{p}}^\times \times \mathbb{G}_m$ given by $(x_1, x_2) \mapsto (x_1, c)$.

*Second assumption on $p$*: The algebra $D_{\mathbb{Q}_p}$ *splits*: $D_{\mathfrak{p}} \cong M_n(\mathbb{Q}_p)$.

In this case the involution $*$ becomes isomorphic to the involution on $M_n(\mathbb{Q}_p) \times M_n(\mathbb{Q}_p)$ given by

$$* : (X, Y) \mapsto (Y^t, X^t).$$

Our assumptions imply that $G = \mathrm{GL}_n \times \mathbb{G}_m$, a split $p$-adic group (and thus $E := \mathbf{E}_{\mathfrak{p}} = \mathbb{Q}_p$). Why is this helpful? As we shall see, this allows us to use the local models for $\mathrm{GL}_n$ described in section 4 to describe the reduction modulo $p$ of the Shimura variety $Sh_{K^p}$, see §6.3.3. Also, we can use the description in [**HN1**] of nearby cycles on such models to compute the semi-simple local zeta function at $p$ of $Sh_{K^p}$, see [**HN3**] and Theorem 11.7. One expects that this is still possible in the general case (where $D_{\mathfrak{p}}^\times$ is not a split group), but there the crucial facts about nearby cycles on the corresponding local models are not yet established.

*Integral data.* We need to specify a $\mathbb{Z}_{(p)}$-order $\mathcal{O}_B \subset B$ and a self-dual multichain $\mathcal{L} = \{\Lambda\}$ of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices. To give a multichain we need to specify first a (partial) $\mathbb{Z}_p$-lattice chain in $V_{\mathbb{Q}_p} = D_{\mathfrak{p}} \times D_{\mathfrak{p}^*}$. We do this one factor at a time. First, we may fix an isomorphism

$$D_{\mathbb{Q}_p} = D_{\mathfrak{p}} \times D_{\mathfrak{p}^*} \cong M_n(\mathbb{Q}_p) \times M_n(\mathbb{Q}_p) \tag{5.2.1}$$

such that the involution $x \mapsto x^\iota = \xi x^* \xi^{-1}$ goes over to $(X, Y) \mapsto (Y^t, X^t)$. So $\xi$ gets identified with an element of the form $(\chi^t, -\chi)$, for $\chi \in \mathrm{GL}_n(\mathbb{Q}_p)$ [6], and our pairing $(x, y) = tr_{D/\mathbb{Q}}(x\xi y^*) = tr_{D/\mathbb{Q}}(xy^\iota \xi)$ goes over to

$$\langle (X_1, X_2), (Y_1, Y_2) \rangle = tr_{D_{\mathbb{Q}_p}/\mathbb{Q}_p}(X_1 Y_2^t \chi^t, -X_2 Y_1^t \chi).$$

Next we define a (partial) $\mathbb{Z}_p$-lattice chain $\Lambda_{-n}^* \subset \cdots \subset \Lambda_0^* = p^{-1}\Lambda_{-n}^*$ in $D_{\mathfrak{p}^*}$ by setting

$$\Lambda_{-i}^* = \chi^{-1} \, \mathrm{diag}(p^i, 1^{n-i}) \, M_n(\mathbb{Z}_p),$$

---

[5] It would make sense to include in our discussion another case, where $p$ remains inert in $F$, and where the group $\mathbf{G}_{\mathbb{Q}_p}$ is a quasi-split unitary group associated to the extension $F_p/\mathbb{Q}_p$. However, we shall postpone discussion of this case to a future occasion.

[6] For use in §11, note that if we multiply $\xi$ by any integral power of $p$, we change neither its properties nor the isomorphism class of the symplectic space $V, (\cdot, \cdot)$. Hence we may assume $\chi^{-1} \in \mathrm{GL}_n(\mathbb{Q}_p) \cap M_n(\mathbb{Z}_p)$.

for $i = 0, 1, \ldots, n$. We can then extend this by periodicity to define $\Lambda_i^*$ for all $i \in \mathbb{Z}$. Similarly, we define the $\mathbb{Z}_p$-lattice chain $\Lambda_0 \subset \cdots \subset \Lambda_n = p^{-1}\Lambda_0$ in $D_{\mathfrak{p}}$ by setting

$$\Lambda_i = \mathrm{diag}((p^{-1})^i, 1^{n-i})\, M_n(\mathbb{Z}_p),$$

for $i = 0, 1, \cdots, n$ (and then extending by periodicity to define for all $i$). We note that

$$(\Lambda_i \oplus \Lambda_i^*)^\perp = \Lambda_{-i} \oplus \Lambda_{-i}^*,$$

where $\perp$ is defined in the usual way using the pairing $(\cdot, \cdot)$. Setting $\mathcal{O}_B \subset B$ to be the unique maximal $\mathbb{Z}_{(p)}$-order such that under our fixed identification $D_{\mathbb{Q}_p} \cong M_n(\mathbb{Q}_p) \times M_n(\mathbb{Q}_p)$, we have

$$\mathcal{O}_B \otimes \mathbb{Z}_p \xrightarrow{\sim} M_n^{opp}(\mathbb{Z}_p) \times M_n^{opp}(\mathbb{Z}_p),$$

one can now check that $\mathcal{L} := \{\Lambda \oplus \Lambda^*\}$ is a self-dual multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices. It is clear that $(\mathcal{O}_B \otimes \mathbb{Z}_p)^\iota = \mathcal{O}_B \otimes \mathbb{Z}_p$.

5.2.4. *The moduli problem.* We have now constructed all the data that enters into the definition of $Sh_{K^p}$. By the determinant condition, the abelian varieties have (relative) dimension $\dim(V_1) = n^2$. An $S$-point in our moduli space is a chain of abelian schemes over $S$ of relative dimension $n^2$, equipped with $\mathcal{O}_B \otimes \mathbb{Z}_{(p)}$-actions, indexed by $\mathcal{L}$ (we set $A_i = A_{\Lambda_i \oplus \Lambda_i^*}$ for all $i \in \mathbb{Z}$)

$$\cdots \xrightarrow{\alpha} A_0 \xrightarrow{\alpha} A_1 \xrightarrow{\alpha} \cdots \xrightarrow{\alpha} A_n \xrightarrow{\alpha} \cdots$$

such that

- each $\alpha$ is an isogeny of height $2n$ (i.e., of degree $p^{2n}$);
- there is a "periodicity isomorphism" $\theta_p : A_{i+n} \to A_i$ such that for each $i$ the composition

$$A_i \xrightarrow{\alpha} A_{i+1} \xrightarrow{\alpha} \cdots \xrightarrow{\alpha} A_{i+n} \xrightarrow{\theta_p} A_i$$

  is multiplication by $p : A_i \to A_i$;
- the morphisms $\alpha$ commute with the $\mathcal{O}_B \otimes \mathbb{Z}_{(p)}$-actions;
- the determinant condition holds: for every $i$ and $b \in \mathcal{O}_B$,

$$\det_{\mathcal{O}_S}(b, \mathrm{Lie}(A_i)) = \det_{E'}(b, V_1).$$

(See [**RZ**], Def. 6.5.)

In addition, we have a principal polarization and a $K^p$-level structure (see [**RZ**], Def. 6.9). Giving a polarization is equivalent to giving a commutative diagram whose vertical arrows are isogenies

$$\begin{tikzcd}
\cdots \arrow[r, "\alpha"] & A_{-1} \arrow[r, "\alpha"] \arrow[d] & A_0 \arrow[r, "\alpha"] \arrow[d] & A_1 \arrow[r, "\alpha"] \arrow[d] & \cdots \arrow[r, "\alpha"] & A_n \arrow[r, "\alpha"] \arrow[d] & \cdots \\
\cdots \arrow[r, "\alpha^\vee"] & \widehat{A}_1 \arrow[r, "\alpha^\vee"] & \widehat{A}_0 \arrow[r, "\alpha^\vee"] & \widehat{A}_{-1} \arrow[r, "\alpha^\vee"] & \cdots \arrow[r, "\alpha^\vee"] & \widehat{A}_{-n} \arrow[r, "\alpha^\vee"] & \cdots
\end{tikzcd}$$

such that for each $i$ the quasi-isogeny

$$A_i \to \widehat{A}_{-i} \to \widehat{A}_i$$

is a rational multiple of a polarization of $A_i$. If up to a $\mathbb{Q}$-multiple the vertical arrows are all isomorphisms, we say the polarization is principal.

The fact that $\mathrm{End}_B(V)$ is a division algebra implies that the moduli space $Sh_{K^p}$ is proper over $\mathcal{O}_E$ (Kottwitz verified the valuative criterion of properness in the case of maximal hyperspecial level structure in [**Ko92**] p. 392, using the theory of Néron models; the same proof applies here.)

**5.3. Siegel modular varieties with $\Gamma_0(p)$-level structure.** The set-up is much simpler here. The group $\mathbf{G}$ is $\mathrm{GSp}(V)$ where $V$ is the standard symplectic space $\mathbb{Q}^{2n}$ with the alternating pairing $(\cdot,\cdot)$ given by the matrix $\tilde{I}$ in 3.1.2. We have $B = \mathbb{Q}$ with involution $\iota = \mathrm{id}$, and $h_0 : \mathbb{C} \to \mathrm{End}(V_{\mathbb{R}})$ is defined as the unique $\mathbb{R}$-algebra homomorphism such that

$$h_0(i) = \tilde{I}.$$

For the multichain $\mathcal{L}$ we use the standard complete self-dual lattice chain $\Lambda_\bullet$ in $\mathbb{Q}_p^{2n}$ that appeared in section 4. We take $\mathcal{O}_B = \mathbb{Z}_{(p)}$.

The group $G = \mathrm{GSp}_{2n,\mathbb{Q}_p}$ is split, so again we have $E = \mathbb{Q}_p$, so $\mathcal{O}_E = \mathbb{Z}_p$. It turns out that the minuscule coweight $\mu$ is

$$\mu = (0^n, (-1)^n),$$

in other words, the same that appeared in the definition of local models in the symplectic case in section 4.

The moduli problem over $\mathbb{Z}_p$ can be expressed as follows. For a $\mathbb{Z}_p$-scheme $S$, an $S$-point is an element of the set of 4-tuples (taken up to isomorphism)

$$\mathcal{A}_{K^p}(S) = \{(A_\bullet, \lambda_0, \lambda_n, \bar{\eta})\}$$

consisting of

- a chain $A_\bullet$ of (relative) $n$-dimensional abelian varieties $A_0 \xrightarrow{\alpha} A_1 \xrightarrow{\alpha} \cdots \xrightarrow{\alpha} A_n$ such that each morphism $\alpha : A_i \to A_{i+1}$ is an isogeny of degree $p$ over $S$;
- principal polarizations $\lambda_0 : A_0 \xrightarrow{\sim} \widehat{A}_0$ and $\lambda_n : A_n \xrightarrow{\sim} \widehat{A}_n$ such that the composition of

$$\begin{tikzcd} A_0 \arrow[r, "\alpha"] & \cdots \arrow[r, "\alpha"] & A_n \arrow[d, "\lambda_n"] \\ \widehat{A}_0 \arrow[u, "\lambda_0^{-1}"] & \cdots \arrow[l, "\alpha^\vee"'] & \widehat{A}_n \arrow[l, "\alpha^\vee"'] \end{tikzcd}$$

  starting and ending at any $A_i$ or $\widehat{A}_i$ is multiplication by $p$;
- a level $K^p$-structure $\bar{\eta}$ on $A_0$.

Exercise 5.3. Show that the above description of $\mathcal{A}_{K^p}$ is equivalent to the definition given in Definition 6.9 of [**RZ**] (cf. our Def. 5.1) for the group-theoretic data $(B, \iota, V, ...)$ we described above.

Note that the only information imparted by the determinant condition in this case is that $\dim(A_i) = n$ for every $i$.

There is another convenient description of the same moduli problem, used by de Jong [**deJ**]. We define another moduli problem $\mathcal{A}'_{K^p}$ whose $S$-points is the set of 4-tuples

$$\mathcal{A}'_{K^p}(S) = \{(A_0, \lambda_0, \bar{\eta}, H_\bullet)\}$$

consisting of

- an $n$-dimensional abelian variety $A_0$ with principal polarization $\lambda_0$ and $K^p$-level structure $\bar{\eta}$;

- a chain $H_\bullet$ of finite flat group subschemes of $A_0[p] := \ker(p : A_0 \to A_0)$ over $S$
$$(0) = H_0 \subset H_1 \subset \cdots \subset H_n \subset A_0[p]$$
such that $H_i$ has rank $p^i$ over $S$ and $H_n$ is totally isotropic with respect to the Riemann form $e_{\lambda_0}$, defined by the diagram
$$\begin{tikzcd} A_0[p] \times A_0[p] \arrow[r, "e_{\lambda_0}"] \arrow[d, "\mathrm{id}\times\lambda_0"] & \mu_p \\ A_0[p] \times \widehat{A}_0[p] \arrow[r, "\cong"] & A_0[p] \times \widehat{A_0[p]}. \arrow[u, "can"'] \end{tikzcd}$$

Here $\widehat{A_0[p]} = \mathrm{Hom}(A_0[p], \mathbb{G}_m)$ denotes the Cartier dual of the finite group scheme $A_0[p]$ and *can* denotes the canonical pairing (which takes values in the $p$-th roots of unity group subscheme $\mu_p \subset \mathbb{G}_m$). (See [**Mu**], section 20, or [**Mi**], section 16.)

The isomorphism $\mathcal{A} \xrightarrow{\sim} \mathcal{A}'$ is given by
$$(A_\bullet, \lambda_0, \lambda_n, \bar{\eta}) \mapsto (A_0, \lambda_0, \bar{\eta}, H_\bullet); \quad H_i := \ker[\alpha^i : A_0 \to A_i].$$
The inverse map is given by setting $A_i = A_0/H_i$ (the condition on $H_n$ allows us to define a principal polarization $\lambda_n : A_0/H_n \xrightarrow{\sim} (\widehat{A_0/H_n})$ using $\lambda_0$).

In [**deJ**], de Jong analyzed the singularities of $\mathcal{A}$ in the case $n = 2$, and deduced that the model $\mathcal{A}$ is flat in that case (by passing from $\mathcal{A}$ to a local model $\mathbf{M}^{\mathrm{loc}}$ according to the procedure of section 6 and then by writing down equations for $\mathbf{M}^{\mathrm{loc}}$).

In the sequel, we will denote the model $\mathcal{A}$ (and $\mathcal{A}'$) by the symbol $Sh$, sometimes adding the subscript $K_p$ when the level-structure at $p$ is not already understood.

## 6. Relating Shimura varieties and their local models

**6.1. Local model diagrams.** Here we describe the desiderata for local models of Shimura varieties. Quite generally, consider a diagram of finite-type $\mathcal{O}_E$-schemes
$$\mathcal{M} \xleftarrow{\varphi} \widetilde{\mathcal{M}} \xrightarrow{\psi} \mathcal{M}^{\mathrm{loc}}.$$

Definition 6.1. We call such a diagram a *local model diagram* provided the following conditions are satisfied:

(1) the morphisms $\varphi$ and $\psi$ are smooth and $\varphi$ is surjective;
(2) étale locally $\mathcal{M} \cong \mathcal{M}^{\mathrm{loc}}$: there exists an étale covering $V \to \mathcal{M}$ and a section $s : V \to \widetilde{\mathcal{M}}$ of $\varphi$ over $V$ such that $\psi \circ s : V \to \mathcal{M}^{\mathrm{loc}}$ is étale.

In practice $\mathcal{M}$ is the scheme we are interested in, and $\mathcal{M}^{\mathrm{loc}}$ is somehow simpler to study; $\widetilde{\mathcal{M}}$ is just some intermediate scheme used to link the other two. Every property that is local for the étale topology is shared by $\mathcal{M}$ and $\mathcal{M}^{\mathrm{loc}}$. For example, if $\mathcal{M}^{\mathrm{loc}}$ is *flat* over $\mathrm{Spec}(\mathcal{O}_E)$, then so is $\mathcal{M}$. The singularities in $\mathcal{M}$ and $\mathcal{M}^{\mathrm{loc}}$ are the same.

**6.2. The general definition of local models.** We briefly recall the general definition of local models, following [**RZ**], Def. 3.27. We suppose we have data $G, \mu, V, V_1, \ldots$ coming from a PEL-type data as in section 5.1. We assume $\mu$ and $V_1$ are defined over a finite extension $E' \supset E$. We suppose we are given a self-dual multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices $\mathcal{L} = \{\Lambda\}$.

Definition 6.2 ([**RZ**], 3.27). A point of $\mathbf{M}^{\mathrm{loc}}$ with values in an $\mathcal{O}_E$-scheme $S$ is given by the following data.

(1) A functor from the category $\mathcal{L}$ to the category of $\mathcal{O}_B \otimes_{\mathbb{Z}_p} \mathcal{O}_S$-modules on $S$

$$\Lambda \mapsto t_\Lambda, \quad \Lambda \in \mathcal{L};$$

(2) A morphism of functors $\psi_\Lambda : \Lambda \otimes_{\mathbb{Z}_p} \mathcal{O}_S \to t_\Lambda$.

We require the following conditions are satisfied:

(i) $t_\Lambda$ is a locally free $\mathcal{O}_S$-module of finite rank. For the action of $\mathcal{O}_B$ on $t_\Lambda$ we have the determinant condition

$$\det_{\mathcal{O}_S}(a; t_\Lambda) = \det_{E'}(a; V_1), \qquad a \in \mathcal{O}_B;$$

(ii) the morphisms $\psi_\Lambda$ are surjective;

(iii) for each $\Lambda$ the composition of the following map is zero:

$$t_\Lambda^\vee \xrightarrow{\psi_\Lambda^\vee} (\Lambda \otimes \mathcal{O}_S)^\vee \xrightarrow[\cong]{(\cdot,\cdot)} \Lambda^\perp \otimes \mathcal{O}_S \xrightarrow{\psi_{\Lambda^\perp}} t_{\Lambda^\perp}.$$

It is clear that one can associate to any PEL-type Shimura variety $Sh = Sh_{K_p}$ a scheme $\mathbf{M}^{\mathrm{loc}}$ (just use the same PEL-type data and multichain $\mathcal{L}$ used to define $Sh_{K_p}$). It is less clear why the resulting scheme $\mathbf{M}^{\mathrm{loc}}$ really is a local model for $Sh_{K_p}$, in the sense described above. We shall see this below, thus justifying the terminology "local model". Then we will show that in our two examples – the "fake" unitary and the Siegel cases – this definition agrees with the concrete ones defined for $\mathrm{GL}_n$ and $\mathrm{GSp}_{2n}$ in section 4.

### 6.3. Constructing local model diagrams for Shimura varieties.

6.3.1. *The abstract construction.* For one of our models $Sh = Sh_{K_p}$ from §5, we want to construct a local model diagram

$$Sh \xleftarrow{\varphi} \widetilde{Sh} \xrightarrow{\psi} \mathbf{M}^{\mathrm{loc}}.$$

In the following we use freely the notation of the appendix, §14. For an abelian scheme $a : A \to S$, let $M(A)$ be the locally free $\mathcal{O}_S$-module dual to the de Rham cohomology

$$M^\vee(A) = H^1_{DR}(A/S) := R^1 a_*(\Omega^\bullet_{A/S}).$$

This is a locally free $\mathcal{O}_S$-module of rank $2\dim(A/S)$. We have the Hodge filtration

$$0 \to \mathrm{Lie}(\widehat{A})^\vee \to M(A) \to \mathrm{Lie}(A) \to 0.$$

This is dual to the usual Hodge filtration on de Rham cohomology

$$0 \subset \omega_{A/S} := a_*\Omega^1_{A/S} \subset H^1_{DR}(A/S).$$

We shall call $M(A)$ the *crystal associated to* $A/S$ (this is perhaps non-standard terminology). If $A$ carries an action of $\mathcal{O}_B$, then by functoriality so does $M(A)$. Note also that $M(A)$ is covariant as a functor of $A$. So if $\mathcal{L}$ denotes a self-dual multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices, and $\{A_\Lambda\}$ denotes an $\mathcal{L}$-set of abelian schemes over $S$ with $\mathcal{O}_B$-action and polarization (as in Definition 5.1), then applying the functor $M(\cdot)$ gives us a polarized multichain $\{M(A_\Lambda)\}$ of $\mathcal{O}_B \otimes \mathcal{O}_S$-modules of type $(\mathcal{L})$, in the sense of [**RZ**], Def. 3.6, 3.10, 3.14.

One key consequence of the conditions imposed in loc. cit., Def. 3.6, is that locally on $S$ there is an isomorphism of polarized multichains of $\mathcal{O}_B \otimes \mathcal{O}_S$-modules

$$\gamma_\Lambda : M(A_\Lambda) \xrightarrow{\sim} \Lambda \otimes_{\mathbb{Z}_p} \mathcal{O}_S.$$

In fact we have the following result which guarantees this.

THEOREM 6.3 ([**RZ**], Theorems 3.11, 3.16). *Let $\mathcal{L} = \{\Lambda\}$ be a (self-dual) multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices in $V$. Let $S$ be any $\mathbb{Z}_p$-scheme where $p$ is locally nilpotent. Then any (polarized) multichain $\{M_\Lambda\}$ of $\mathcal{O}_B \otimes_{\mathbb{Z}_p} \mathcal{O}_S$-modules of type $(\mathcal{L})$ is locally (for the étale topology on $S$) isomorphic to the (polarized) multichain $\{\Lambda \otimes_{\mathbb{Z}_p} \mathcal{O}_S\}$.*

*Moreover, the functor* Isom

$$T \mapsto \mathrm{Isom}(\{M_\Lambda \otimes \mathcal{O}_T\}, \{\Lambda \otimes \mathcal{O}_T\}),$$

*is represented by a smooth affine scheme over $S$.*

The analogous statements hold for any $\mathbb{Z}_p$-scheme $S$, see [**P**]. In particular for such $S$ we have a smooth affine group scheme $\mathcal{G}$ over $S$ given by

$$\mathcal{G}(T) = \mathrm{Aut}(\{\Lambda \otimes \mathcal{O}_T\}),$$

and the functor Isom is obviously a left-torsor under $\mathcal{G}$. This generalizes the smoothness of the groups Aut in section 3.2. Moreover, by the same arguments as in section 3.2, for $S = \mathrm{Spec}(\mathbb{Z}_p)$ the group $\mathcal{G}_{\mathbb{Z}_p}$ is a Bruhat-Tits parahoric group scheme corresponding to the parahoric subgroup of $G(\mathbb{Q}_p) = \mathcal{G}(\mathbb{Q}_p)$ which stabilizes the multichain $\mathcal{L}$ [7].

In the special case of lattice chains for $\mathrm{GSp}_{2n}$, the theorem was proved by de Jong [**deJ**] (he calls what are "polarized (multi)chains" here by the name "systems of $\mathcal{O}_S$-modules of type II").

Now we define the local model diagram for $Sh$. We assume $\mathcal{O}_E = \mathbb{Z}_p$ for simplicity. Let us define $\widetilde{Sh}$ to be the $\mathbb{Z}_p$-scheme representing the functor whose points in a $\mathbb{Z}_p$-scheme $S$ is the set of pairs

$$(\{A_\Lambda\}, \bar{\lambda}, \bar{\eta}) \in Sh(S); \quad \gamma_\Lambda : M(A_\Lambda) \xrightarrow{\sim} \Lambda \otimes_{\mathbb{Z}_p} \mathcal{O}_S,$$

where $\gamma_\Lambda$ is an isomorphism of polarized multichains of $\mathcal{O}_B \otimes \mathcal{O}_S$-modules. The morphism

$$\varphi : \widetilde{Sh} \to Sh$$

is the obvious morphism which forgets $\gamma_\Lambda$. By Theorem 6.3, $\varphi$ is smooth (being a torsor for a smooth group scheme) and surjective. Now we want to define

$$\psi : \widetilde{Sh}(S) \to \mathbf{M}^{\mathrm{loc}}(S).$$

We define it to send an $S$-point $(\{A_\Lambda\}, \bar{\lambda}, \bar{\eta}, \gamma_\Lambda)$ to the morphism of functors

$$\Lambda \otimes_{\mathbb{Z}_p} \mathcal{O}_S \to \mathrm{Lie}(A_\Lambda)$$

induced by the composition $\gamma_\Lambda^{-1} : \Lambda \otimes_{\mathbb{Z}_p} \mathcal{O}_S \cong M(A_\Lambda)$ with the canonical surjective morphism

$$M(A_\Lambda) \to \mathrm{Lie}(A_\Lambda).$$

It is not completely obvious that the morphisms $\Lambda \otimes \mathcal{O}_S \to \mathrm{Lie}(A_\Lambda)$ satisfy the condition (iii) of Definition 6.2. We will explain it in the Siegel case below, as a

[7] More precisely, the *connected component* of $\mathcal{G}$ is the Bruhat-Tits group scheme. As G. Pappas points out, in some cases (e.g. the unitary group for ramified quadratic extensions), the stabilizer group $\mathcal{G}$ is not connected.

consequence of Proposition 5.1.10 of [**BBM**] (our Prop. 14.1). We omit discussion of this point in other cases.

The theory of Grothendieck-Messing ([**Me**]) shows that the morphism $\psi$ is formally smooth. Since both schemes are of finite type over $\mathbb{Z}_p$, it is smooth. In summary:

THEOREM 6.4 ([**RZ**], §3). *The diagram*

$$Sh \xleftarrow{\varphi} \widetilde{Sh} \xrightarrow{\psi} \mathbf{M}^{\mathrm{loc}}$$

*is a local model diagram. The morphism $\varphi$ is a torsor for the smooth affine group scheme $\mathcal{G}$.*

PROOF. We have indicated why condition (1) of Definition 6.1 is satisfied. Condition (2) is proved in [**RZ**], 3.30-3.35; see also [**deJ**], Cor. 4.6. □

We will describe the local model diagrams more explicitly for each of our two main examples next. Our goal is to show that their local models are none other than the ones defined in section 4.

6.3.2. *Symplectic case.* Following [**deJ**] and [**GN**] we change conventions slightly and replace the de Rham homology functor with the cohomology functor

$$A/S \mapsto H^1_{DR}(A/S).$$

What kind of data do we get by applying the de Rham cohomology functor to a point in our moduli problem $Sh$ from section 5.3? For notational convenience, let us now number the chains of abelian varieties in the opposite order:

$$\{A_{\Lambda_\bullet}\} = A_n \to A_{n-1} \to \cdots \to A_0.$$

LEMMA 6.5. *The result of applying $H^1_{DR}$ to a point $(\{A_{\Lambda_\bullet}\}, \lambda_0, \lambda_n)$ in $Sh(S)$ is a datum of form $(M_0 \xrightarrow{\alpha} M_1 \xrightarrow{\alpha} \cdots \xrightarrow{\alpha} M_n, q_0, q_n)$ satisfying*

- *$M_i$ is a locally free $\mathcal{O}_S$-module of rank $2n$;*
- *$\mathrm{Coker}(M_{i-1} \to M_i)$ is a locally free $\mathcal{O}_S/p\mathcal{O}_S$-module of rank 1;*
- *for $i = 0, n$, $q_i : M_i \otimes M_i \to \mathcal{O}_S$ is a non-degenerate symplectic pairing;*
- *for any $i$, the composition of*

$$\begin{tikzcd} M_0 \arrow[r, "\alpha"] & \cdots \arrow[r, "\alpha"] & M_n \arrow[d, "q_n"] \\ M_0^\vee \arrow[u, "q_0"] & \cdots \arrow[l, "\alpha^\vee"'] & M_n^\vee \arrow[l, "\alpha^\vee"'] \end{tikzcd}$$

*starting and ending at $M_i$ or $M_i^\vee := \mathrm{Hom}(M_i, \mathcal{O}_S)$ is multiplication by $p$.*

PROOF. The pairings $q_0, q_n$ come from the polarizations $\lambda_0, \lambda_n$. The various properties are easy to check, using the canonical natural isomorphism $H^1_{DR}(\widehat{A}/S) = (H^1_{DR}(A/S))^\vee$; cf. Prop. 14.1. □

Our next goal is to rephrase Definition 6.2 in terms of data similar to that in Lemma 6.5, which will take us closer to the definition of $\mathbf{M}^{\mathrm{loc}}$ in §4.

Let $\mathcal{L} = \Lambda_\bullet$ be the standard self-dual lattice chain in $V = \mathbb{Q}_p^{2n}$, with respect to the usual pairing $(x, y) = x^t \tilde{I} y$. Clearly we may rephrase Definition 6.2 using the sub-objects $\omega'_\Lambda := \ker(\psi_\Lambda)$ of $\Lambda \otimes \mathcal{O}_S$ rather than the quotients $t_\Lambda$. Then condition (iii) becomes (iii') $(\omega'_\Lambda)^{perp} \subset \omega'_{\Lambda^\perp}$, which is equivalent to $(\omega'_\Lambda)^{perp} = \omega'_{\Lambda^\perp}$,

in other words, under the canonical pairing $(\cdot,\cdot) : \Lambda \otimes \mathcal{O}_S \times \Lambda^\perp \otimes \mathcal{O}_S \to \mathcal{O}_S$, the submodules $\omega'_\Lambda$ and $\omega'_{\Lambda^\perp}$ are perpendicular. If $\Lambda = \Lambda^\perp$, this means

$$(\cdot,\cdot)|_{\omega'_\Lambda \times \omega'_\Lambda} \equiv 0,$$

and if $\Lambda^\perp = p\Lambda$, this means

$$p(\cdot,\cdot)|_{\omega'_\Lambda \times \omega'_\Lambda} \equiv 0,$$

since the pairing on $\Lambda \otimes \mathcal{O}_S \times \Lambda \otimes \mathcal{O}_S$ is defined by composing the standard pairing on $\Lambda \otimes \mathcal{O}_S \times \Lambda^\perp \otimes \mathcal{O}_S$ with the periodicity isomorphism

$$p : \Lambda \xrightarrow{\sim} \Lambda^\perp$$

in the second variable. For the "standard system" $(\Lambda_0 \to \Lambda_1 \to \cdots \to \Lambda_n, q_0, q_n)$ as in Lemma 6.5, the (perfect) pairings are given by

$$q_0 = (\cdot,\cdot) : \Lambda_0 \times \Lambda_0 \to \mathbb{Z}_p$$
$$q_n = p(\cdot,\cdot) : \Lambda_n \times \Lambda_n \to \mathbb{Z}_p.$$

Note that if $\omega'_i := \omega'_{\Lambda_i}$, the identity $(\omega'_i)^{perp} = \omega'_{-i}$ ((iii) of Def. 6.2) means that $\omega'_\bullet$ is uniquely determined by the elements $\omega'_0, \dots, \omega'_n$. Conversely, suppose we are given $\omega'_0, \dots, \omega'_n$ such that $(\omega'_0)^{perp} = \omega'_0$ and $(\omega'_n)^{perp} = p\,\omega'_n =: \omega'_{-n}$. Then we can *define* $\omega'_{-i} = (\omega'_i)^{perp}$ for $i = 0, \dots, n$, and then extend by periodicity to get an infinite chain $\omega'_\bullet$ as in Definition 6.2 (condition (iii) being satisfied by fiat).

We thus have the following reformulation of Definition 6.2, which shows that that definition agrees with the one in section 4 for $\mathrm{GSp}_{2n}$.

LEMMA 6.6. *In the Siegel case, an $S$-point of $\mathbf{M}^{\mathrm{loc}}$ (in the sense of Definition 6.2) is a commutative diagram*

$$\begin{tikzcd}
\Lambda_0 \otimes \mathcal{O}_S \arrow[r] & \Lambda_1 \otimes \mathcal{O}_S \arrow[r] & \cdots \arrow[r] & \Lambda_n \otimes \mathcal{O}_S \\
\omega'_0 \arrow[r] \arrow[u] & \omega'_1 \arrow[r] \arrow[u] & \cdots \arrow[r] & \omega'_n, \arrow[u]
\end{tikzcd}$$

*such that*

- *for each $i$, $\omega'_i$ is a locally free $\mathcal{O}_S$-submodule of $\Lambda_i \otimes \mathcal{O}_S$ of rank $n$;*
- *$\omega'_0$ is totally isotropic for $(\cdot,\cdot)$ and $\omega'_n$ is totally isotropic for $p(\cdot,\cdot)$.*

Finally, we promised to explain why the morphism $\psi : \widetilde{Sh} \to \mathbf{M}^{\mathrm{loc}}$ really takes values in $\mathbf{M}^{\mathrm{loc}}$. We must also redefine it in terms of cohomology. Recall we now have the Hodge filtration

$$\omega_{A_\Lambda/S} \subset H^1_{DR}(A_\Lambda/S).$$

We define $\psi$ to send $(\{A_0 \leftarrow \cdots \leftarrow A_n\}, \lambda_0, \lambda_n, \bar{\eta}, \gamma_\Lambda)$ to the locally free, rank $n$, $\mathcal{O}_S$-submodules

$$\gamma_\Lambda(\omega_{A_\Lambda}) \subset \Lambda \otimes \mathcal{O}_S,$$

where now $\gamma_\Lambda$ is an isomorphism of polarized multichains of $\mathcal{O}_S$-modules

$$\gamma_\Lambda : H^1_{DR}(A_\Lambda/S) \xrightarrow{\sim} \Lambda \otimes \mathcal{O}_S.$$

The following result ensures that this map really takes values in $\mathbf{M}^{\mathrm{loc}}$.

LEMMA 6.7. *The morphism $\psi$ takes values in $\mathbf{M}^{\mathrm{loc}}$, i.e., condition (iii') holds.*

PROOF. Setting $\omega_{A_{\Lambda_i}} = \omega_i$, we need to see that the Hodge filtration $\omega_0$ resp. $\omega_n$ is totally isotropic with respect to the pairing $q_0$ resp. $q_n$ induced by the polarization $\lambda_0$ resp. $\lambda_n$. But this is Proposition 5.1.10 of [**BBM**] (our Prop. 14.1). See also [**deJ**], Cor. 2.2. □

*Comparison of homology and cohomology local models.* One further remark is in order. Let us consider a point

$$A = (A_0 \to \cdots \to A_n, \lambda_0, \lambda_n, \bar{\eta})$$

in our moduli problem $Sh$. Note that this data gives us another point in $Sh$, namely

$$\widehat{A} = (\widehat{A}_n \to \cdots \to \widehat{A}_0, \lambda_n^{-1}, \lambda_0^{-1}, \bar{\eta}).$$

(We need to use the assumption that the polarizations $\lambda_i$ are required to be *symmetric* isogenies $A_i \to \widehat{A}_i$, in the sense that $\widehat{\lambda}_i = \lambda_i$.)

The moduli problem $Sh$ is thus equipped with an automorphism of order 2, given by $A \mapsto \widehat{A}$.

This comes in handy in comparing the "homology" and "cohomology" constructions of the local model diagram. Namely, since $M(A_i) = H^1_{DR}(\widehat{A}_i)$ (Prop. 14.1), an isomorphism $\gamma_\bullet : M(A_\bullet) \xrightarrow{\sim} \Lambda_\bullet \otimes \mathcal{O}_S$ is simultaneously an isomorphism $\gamma_\bullet : H^1_{DR}(\widehat{A}_\bullet) \xrightarrow{\sim} \Lambda_\bullet \otimes \mathcal{O}_S$. In the "homology" version, $\psi$ sends $(A, \gamma_\bullet)$ to the quotient chain

$$\Lambda_\bullet \otimes \mathcal{O}_S \to \mathrm{Lie}(A_\bullet),$$

defined using $\gamma_\bullet^{-1}$. On the other hand, in the "cohomology" version, $\psi$ sends $(\widehat{A}, \gamma_\bullet)$ to the sub-object chain

$$\omega_{\widehat{A}_\bullet} \subset \Lambda_\bullet \otimes \mathcal{O}_S$$

(identifying $\omega_\bullet$ with $\gamma_\bullet(\omega_\bullet)$). But the exact sequence

$$0 \to \omega_{\widehat{A}} \to M(A) \to \mathrm{Lie}(A) \to 0$$

(Prop. 14.1) means that the two chains correspond: they give exactly the same element of $\mathbf{M}^{\mathrm{loc}}$. In summary, we have the following result relating the "homology" and "cohomology" constructions of the local model diagram.

PROPOSITION 6.8. *There is a commutative diagram*

$$\begin{tikzcd} \widetilde{Sh}^{hom} \arrow[r, "\psi^{hom}"] \arrow[d] & \mathbf{M}^{\mathrm{loc}} \arrow[d, "="] \\ \widetilde{Sh}^{coh} \arrow[r, "\psi^{coh}"] & \mathbf{M}^{\mathrm{loc}}, \end{tikzcd}$$

*where the left vertical arrow is the automorphism* $(A, \gamma_\bullet) \mapsto (\widehat{A}, \gamma_\bullet)$.

6.3.3. *"Fake" unitary case.* Here the "standard" polarized multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices is given by $\{\Lambda_i \oplus \Lambda_i^*\}$, in the notation of section 5.2.3. Recall that

$$\mathcal{O}_B \otimes \mathbb{Z}_p \cong M_n^{opp}(\mathbb{Z}_p) \times M_n^{opp}(\mathbb{Z}_p),$$

according to the decomposition of $B^{opp}_{\mathbb{Q}_p} = D_{\mathbb{Q}_p}$

$$D_{\mathbb{Q}_p} = D_{\mathfrak{p}} \times D_{\mathfrak{p}^*} \cong M_n(\mathbb{Q}_p) \times M_n(\mathbb{Q}_p).$$

Let $W$ (resp. $W^*$) be $\mathbb{Z}_p^n$ viewed as a left $\mathcal{O}_B \otimes \mathbb{Z}_p$-module, via right multiplications by elements of the first (resp. second) factor of $M_n(\mathbb{Z}_p) \times M_n(\mathbb{Z}_p)$. The ring $B_{\mathbb{Q}_p}$ has two simple left modules: $W_{\mathbb{Q}_p}$ and $W^*_{\mathbb{Q}_p}$. We may write

$$V_{E'} = V_1 \oplus V_2$$

as before. The determinant condition now implies (at least over $E'$) that

$$V_1 = W^d_{E'} \oplus (W^*_{E'})^{n-d};$$

(comp. section 5.2.2). Using the "sub-object" variant of Definition 6.2, it follows that an $S$-point of $\mathbf{M}^{\rm loc}$ is a commutative diagram (here $\Lambda_i$ being understood as $\Lambda_i \otimes \mathcal{O}_S$)

$$\begin{tikzcd} \Lambda_0 \oplus \Lambda_0^* \arrow[r] & \Lambda_1 \oplus \Lambda_1^* \arrow[r] & \cdots \arrow[r] & \Lambda_n \oplus \Lambda_n^* \\ \mathcal{F}_0 \oplus \mathcal{F}_0^* \arrow[r] \arrow[u] & \mathcal{F}_1 \oplus \mathcal{F}_1^* \arrow[r] \arrow[u] & \cdots \arrow[r] & \mathcal{F}_n \oplus \mathcal{F}_n^* \arrow[u] \end{tikzcd}$$

where $\mathcal{F}_i \oplus \mathcal{F}_i^*$ is an $\mathcal{O}_B \otimes \mathcal{O}_S$-submodule of $\Lambda_i \oplus \Lambda_i^*$ which, locally on $S$, is a direct factor isomorphic to $W^{n-d}_{\mathcal{O}_S} \oplus (W^*_{\mathcal{O}_S})^d$.

The analogue of condition (iii'), which is imposed in Definition 6.2, is

$$(\mathcal{F}_i \oplus \mathcal{F}_i^*)^{perp} = \mathcal{F}_{-i} \oplus \mathcal{F}^*_{-i}.$$

On the other hand, from the definition of $\langle \cdot , \cdot \rangle$ in section 5.2.3 it is immediate that

$$(\mathcal{F}_i \oplus \mathcal{F}_i^*)^{perp} = \mathcal{F}_i^{*,perp} \oplus \mathcal{F}_i^{perp}.$$

We see thus that the first factor $\mathcal{F}_\bullet$ uniquely determines the second factor $\mathcal{F}^*_\bullet$ (and vice-versa). Thus $\mathbf{M}^{\rm loc}$ is given by chains of right $M_n(\mathcal{O}_S) = M_n(\mathbb{Z}_p) \otimes \mathcal{O}_S$-modules

$$\mathcal{F}_0 \to \mathcal{F}_1 \to \cdots \to \mathcal{F}_n$$

which are locally direct factors in

$$M_n(\mathcal{O}_S) \to \mathrm{diag}(p^{-1}, 1^{n-1}) M_n(\mathcal{O}_S) \to \cdots \to p^{-1} M_n(\mathcal{O}_S),$$

each term locally isomorphic to $(\mathcal{O}_S^n)^{n-d}$. By Morita equivalence, $\mathbf{M}^{\rm loc}$ is just given by the definition in section 4 (for the integer $d$).

## 7. Flatness

Because of the local model diagram, the flatness of the moduli problem $Sh$ can be investigated by considering its local model. The following fundamental result is due to U. Görtz. It applies to all parahoric subgroups.

THEOREM 7.1 ([**Go1**], [**Go2**]). *Suppose $\mathbf{M}^{\rm loc}$ is a local model attached to a group $\mathrm{Res}_{F/\mathbb{Q}_p}(\mathrm{GL}_n)$ or $\mathrm{Res}_{F/\mathbb{Q}_p}(\mathrm{GSp}_{2n})$, where $F/\mathbb{Q}_p$ is an unramified extension. Then $\mathbf{M}^{\rm loc}$ is flat over $\mathcal{O}_E$. Moreover, its special fiber is reduced, and has rational singularities.*

We give the idea for the proof. One reduces to the case where $F = \mathbb{Q}_p$. In order to prove flatness over $\mathcal{O}_E = \mathbb{Z}_p$ it is enough to prove the following facts (comp. [**Ha**], III.9.8):
1) The special fiber is reduced, as a scheme over $\mathbb{F}_p$;
2) The model is *topologically flat*: every closed point in the special fiber is contained in the the scheme-theoretic closure of the generic fiber.

The innovation behind the proof of 1) is to embed the special fiber into the affine flag variety and then to make systematic use of the theory of Frobenius-splitting to prove affine Schubert varieties are compatibly Frobenius-split. See [**Go1**].

To prove 2), suppose $\mu$ is such that $\mathbf{M}^{\mathrm{loc}} = M_{-w_0\mu}$. It is enough by a result of Kottwitz-Rapoport (Theorem 4.2) to prove that the generic element in a stratum of the special fiber indexed by a translation element in $\mathrm{Adm}(-w_0\mu)$ can be lifted to characteristic zero. This statement is checked by hand in [**Go1**].

We will provide an alternative, calculation-free, proof by making use of nearby cycles [8]. We freely make use of material on nearby cycles from §10, 11.

We fix an element $\lambda \in W(-w_0\mu)$ and consider the stratum of $\mathbf{M}^{\mathrm{loc}}$ indexed by $t_\lambda$. We want to show this stratum is in the closure of the generic fiber.

The nearby cycles sheaf $R\Psi^{\mathbf{M}^{\mathrm{loc}}}(\overline{\mathbb{Q}}_\ell)$ is supported only on this closure (Theorem 10.1), and so it is enough to show that

$$\mathrm{Tr}^{ss}(\Phi_p^r, R\Psi^{\mathbf{M}^{\mathrm{loc}}}_{t_\lambda}(\overline{\mathbb{Q}}_\ell)) \neq 0.$$

But it is clear that

$$z_{-w_0\mu,r}(t_\lambda) \neq 0$$

from the definition of Bernstein functions (see [**Lu**] or [**HKP**]), and we are done by Theorem 11.3 (which also holds for the group $\mathrm{GSp}_{2n}$, see [**HN1**]).

As G. Pappas has observed [**P**], the local models attached (by [**RZ**]) to the groups above can fail to be flat if $F/\mathbb{Q}_p$ is *ramified*. In their joint works [**PR1**], [**PR2**], Pappas and Rapoport provide alternative definitions of local models in that case (in fact they treat nearly all the groups considered in [**RZ**]), and these new models *are* flat. However, these new models cannot always be described as the scheme representing a "concrete" moduli problem.

## 8. The Kottwitz-Rapoport stratification

Let us assume $Sh$ is the model over $\mathcal{O}_E$ for one of the Shimura varieties $Sh(\mathbf{G}, h)_{\mathbf{K}}$ discussed in §5, i.e. a "fake" unitary or a Siegel modular variety. We assume (for simplicity of statements) that $K_p$ is an Iwahori subgroup of $G(\mathbb{Q}_p)$. Let us summarize what we know so far.

The group $G_{\mathbb{Q}_p}$ is either isomorphic to $\mathrm{GL}_{n,\mathbb{Q}_p} \times \mathbb{G}_{m,\mathbb{Q}_p}$ or $\mathrm{GSp}_{2n,\mathbb{Q}_p}$. These groups being split over $\mathbb{Q}_p$, we have $E = \mathbb{Q}_p$ and $\mathcal{O}_E = \mathbb{Z}_p$.

The Shimura datum $h$ gives rise to a dominant minuscule cocharacter $\mu$ of $\mathrm{GL}_{n,\mathbb{Q}_p}$ or $\mathrm{GSp}_{2n,\mathbb{Q}_p}$, respectively. The functorial description of the local model $\mathbf{M}^{\mathrm{loc}}$ shows that it can be embedded into the deformation $M$ from the affine Grassmannian $\mathrm{Grass}_{\mathbb{Q}_p}$ to the affine flag variety $\mathcal{F}l_{\mathbb{F}_p}$ associated to $G$, and has generic fibre $\mathcal{Q}_{-w_0\mu}$. Since the local model is flat with reduced special fiber [**Go1**], [**Go2**] (see §7) and is closed in $M$, it coincides with the scheme-theoretic closure $M_{-w_0\mu}$ of $\mathcal{Q}_{-w_0\mu}$ in this deformation. We thus identify $\mathbf{M}^{\mathrm{loc}} = M_{-w_0\mu}$. (For all this, keep in mind we use the "homology" definition of the local model diagram.)

The relation between the model of the Shimura variety over $\mathbb{Z}_p$ and its local model is given by a diagram

$$Sh \xleftarrow{\varphi} \widetilde{Sh} \xrightarrow{\psi} \mathbf{M}^{\mathrm{loc}}$$

[8]We emphasize that this proof is much less elementary than the original proof of Görtz [**Go1**], relying as it does on the full strength of [**HN1**].

of $\mathbb{Z}_p$-schemes, where $\varphi$ is a torsor under the smooth affine group scheme $\mathcal{G}$ of §6.3 (also termed Aut in §4), and $\psi$ is smooth. The fibres of $\varphi$ are geometrically connected (more precisely, this holds for the restriction of $\varphi$ to any geometric connected component of $\widetilde{Sh}$). One can show that étale-locally around each point of the special fiber of $Sh$, the schemes $Sh$ and $\mathbf{M}^{\mathrm{loc}}$ are isomorphic.

The stratification of the special fibre of $\mathbf{M}^{\mathrm{loc}}$ (by Iwahori-orbits) induces stratifications of the special fibers of $\widetilde{Sh}$ and $Sh$ (see below). The resulting stratification of $Sh_{\mathbb{F}_p}$ is called the *Kottwitz-Rapoport (or KR-) stratification.*

**8.1. Construction of the KR-stratification.** Essentially following [**GN**], we will recall the construction and basic properties of the KR-stratification. The difference between their treatment and ours is that they construct local models in terms of de Rham *cohomology*, whereas here they are constructed in terms of de Rham *homology*. This is done for compatibility with the computations in §11.

For later use in §11, we give a detailed treatment here for the case of "fake" unitary Shimura varieties.

Let $k$ denote the algebraic closure of the residue field of $\mathbb{Z}_p$, and let $\widetilde{\Lambda}_\bullet = \Lambda_\bullet \oplus \Lambda_\bullet^*$ denote the self-dual multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices from §5.2.3. Recall that a point in $\mathbf{M}^{\mathrm{loc}}(k)$ is a "quotient chain" of $k$-vector spaces

$$\widetilde{\Lambda}_\bullet \otimes k \to t_{\widetilde{\Lambda}_\bullet},$$

self-dual in the sense of Definition 6.2 (iii), and such that each $t_{\widetilde{\Lambda}_i}$ satisfies the determinant condition, that is,

$$t_{\widetilde{\Lambda}_i} = W_k^d \oplus (W^*)_k^{n-d}$$

as $\mathcal{O}_B \otimes k$-modules. We can identify this object with a lattice chain in the affine flag variety for $\mathrm{GL}_n(k((t)))$ as follows. Let $\mathcal{V}_{\bullet,k}$ denote the "standard" complete lattice chain from §4.1. Using duality and Morita equivalence (see §6.3.3), the quotient $t_{\widetilde{\Lambda}_\bullet}$ can be identified with a quotient $t_{\Lambda_\bullet}$ of the standard lattice chain $\mathcal{V}_{\bullet,k} \subset k(\!(t)\!)^n$. Then we may write

$$t_{\Lambda_i} = \mathcal{V}_{i,k}/\mathcal{L}_i$$

for a unique lattice chain $\mathcal{L}_\bullet = (\mathcal{L}_0 \subset \cdots \subset \mathcal{L}_n = t^{-1}\mathcal{L}_0)$ consisting of $k[[t]]$-submodules of $k((t))^n$ which satisfy for each $i = 0, \ldots, n$,

- $t\mathcal{V}_{i,k} \subset \mathcal{L}_i \subset \mathcal{V}_{i,k}$;
- the $k$-vector space $\mathcal{V}_{i,k}/\mathcal{L}_i$ has dimension $d$ (determinant condition).

The set of such lattice chains $\mathcal{L}_\bullet$ is the special fiber of the model $M_{-w_0\mu}$ attached to the dominant coweight $-w_0\mu = (1^d, 0^{n-d})$ of $\mathrm{GL}_n$. Indeed, the two conditions above mean that for each $i$,

$$\mathrm{inv}_K(\mathcal{L}_i, \mathcal{V}_{i,k}) = \mu$$

and thus

$$\mathrm{inv}_K(\mathcal{V}_{i,k}, \mathcal{L}_i) = -w_0\mu.$$

Here $\mathrm{inv}_K$ is the standard notion of relative position of $k[\![t]\!]$-lattices in $k(\!(t)\!)^n$, relative to the base point $\mathcal{V}_{0,k} = k[\![t]\!]^n$: we say $\mathrm{inv}_K(g\mathcal{V}_{0,k}, g'\mathcal{V}_{0,k}) = \lambda \in X_+(T)$ if $g^{-1}g' \in K\lambda K$, where $K = \mathrm{GL}_n(k[\![t]\!])$. Recall we have identified $\mu$ with $(0^{n-d}, (-1)^d)$ and have embedded coweights into the loop group by the rule $\lambda \mapsto \lambda(t)$.

If $\mathcal{L}_\bullet = x(\mathcal{V}_{\bullet,k})$ for $x \in \widetilde{W}(\mathrm{GL}_n)$, this means that $x \in \mathrm{Perm}(-w_0\mu)$, which is also the set $\mathrm{Adm}(-w_0\mu)$, see §4.

Recall that the Iwahori subgroup $I = I_{k((t))}$ fixing $\mathcal{V}_{\bullet,k}$ preserves the subset $M_{-w_0\mu,k} \subset \mathcal{F}l_k$ and so via the identification $\mathbf{M}^{\rm loc} = M_{-w_0\mu}$, it also acts on the local model. The Iwahori-orbits give a cellular decomposition

$$\mathbf{M}^{\rm loc}_k = \coprod_{w \in \mathrm{Adm}(\mu)} \mathbf{M}^{\rm loc}_w.$$

Here we define $\mathbf{M}^{\rm loc}_w$ to be the set of $\mathcal{L}_\bullet$ above such that

$$\mathrm{inv}_I(\mathcal{V}_{\bullet,k}, \mathcal{L}_\bullet) = w^{-1},$$

or equivalently

$$\mathrm{inv}_I(\mathcal{L}_\bullet, \mathcal{V}_{\bullet,k}) = w,$$

for $w^{-1} \in \mathrm{Adm}(-w_0\mu)$ (which happens if and only if $w \in \mathrm{Adm}(\mu)$). Here we define $\mathrm{inv}_I(g\mathcal{V}_{\bullet,k}, g'\mathcal{V}_{\bullet,k}) = w$ if $g^{-1}g' \in IwI$.

Each stratum is smooth (in fact $\mathbf{M}^{\rm loc}_w = \mathbb{A}^{\ell(w)}$), and the closure relations are determined by the Bruhat order on $\widetilde{W}$; that is, $\mathbf{M}^{\rm loc}_w \subset \overline{\mathbf{M}^{\rm loc}_{w'}}$ if and only if $w \leq w'$.

There is a surjective homomorphism $I_{k((t))} \to \mathrm{Aut}_k$, where Aut is the group scheme $\mathcal{G}$ of Theorem 6.3 which acts on the whole local model diagram. The action of $I_{k((t))}$ on $\mathbf{M}^{\rm loc}_k$ factors through $\mathrm{Aut}_k$, so that the strata above can also be thought of as $\mathrm{Aut}_k$-orbits. The morphism $\psi$ is clearly equivariant for $\mathrm{Aut}_k$, hence we have a stratification of $\widetilde{Sh}_k$

$$\widetilde{Sh}_k = \coprod_{w \in \mathrm{Adm}(\mu)} \psi^{-1}(\mathbf{M}^{\rm loc}_w),$$

whose strata are non-empty (Lemma 13.1), smooth, and stable under the action of $\mathrm{Aut}_k$. Since $\varphi_k$ is a torsor for the smooth group scheme $\mathrm{Aut}_k$, the stratification descends to $Sh_k$:

$$Sh_k = \coprod_{w \in \mathrm{Adm}(\mu)} Sh_w$$

such that $\varphi^{-1}(Sh_w) = \psi^{-1}(\mathbf{M}^{\rm loc}_w)$. These strata are still smooth, non-empty, and satisfy closure relations determined by the Bruhat order.

All statements above remain true over the base field $\mathbb{F}_p$ instead of its algebraic closure $k$.

**8.2. Relating nearby cycles on local models and Shimura varieties.** We will need in §11 the following result relating the nearby cycles on $Sh$, $\widetilde{Sh}$, and $\mathbf{M}^{\rm loc}$, which follows immediately from the above remarks and Theorem 10.1 below (cf. [**GN**]):

LEMMA 8.1. *There are canonical isomorphisms*

$$\varphi^* R\Psi^{Sh}(\overline{\mathbb{Q}}_\ell) = R\Psi^{\widetilde{Sh}}(\overline{\mathbb{Q}}_\ell) = \psi^* R\Psi^{\mathbf{M}^{\rm loc}}(\overline{\mathbb{Q}}_\ell).$$

*Moreover, $R\Psi^{Sh}(\overline{\mathbb{Q}}_\ell)$ is constant on each stratum $Sh_w$, and if $\Phi_\mathfrak{p} \in \mathrm{Gal}(\overline{\mathbb{Q}}_p/E)$ is a geometric Frobenius element, then for any elements $x \in Sh_w(k_r)$ and $x_0 \in \mathbf{M}^{\rm loc}_w(k_r)$, we have*

$$\mathrm{Tr}^{ss}(\Phi^r_\mathfrak{p}, R\Psi^{Sh}_x(\overline{\mathbb{Q}}_\ell)) = \mathrm{Tr}^{ss}(\Phi^r_\mathfrak{p}, R\Psi^{\mathbf{M}^{\rm loc}}_{x_0}(\overline{\mathbb{Q}}_\ell)).$$

Here we use the notion of semi-simple trace, which is explained below in §9.3. The above lemma plays a key role in the determination of the semi-simple local zeta function in §11.

**8.3. The Genestier-Ngô comparison with $p$-rank.** We assume in this section that $Sh$ is the Siegel modular variety with Iwahori-level structure from §5.3.

Recall that any $n$-dimensional abelian variety $A$ over an algebraically closed field $k$ of characteristic $p$ has

$$\#A[p](k) = p^j,$$

for some integer $0 \leq j \leq n$. The integer $j$ is called the *$p$-rank* of $A$. Ordinary abelian varieties are those whose $p$-rank is $n$, the largest possible. The $p$-rank is constant on isogeny classes, and therefore it determines a well-defined function on the set of geometric points $Sh_{\bar{\mathbb{F}}_p}$. The level sets determine the *stratification by $p$-rank*.

It is natural to ask how this stratification relates to the KR-stratification. In [**GN**], Genestier and Ngô have very elegantly derived the relationship using local models and work of de Jong [**deJ**]. As they point out, their theorem yields interesting results even in the case of Siegel modular varieties having *good reduction* at $p$: they derive a short and beautiful proof that the ordinary locus in such Shimura varieties is open and dense in the special fiber (comp. [**W1**]).

To state their result, we define for $w \in \widetilde{W}(\mathrm{GSp}_{2n})$ an integer $r(w)$ as follows. Its image $\overline{w}$ in the finite Weyl group $W(\mathrm{GSp}_{2n})$ is a permutation of the set $\{1, \cdots, 2n\}$ commuting with the involution $i \mapsto 2n+1-i$. The set of fixed points of $\overline{w}$ is stable under the involution, and therefore has even cardinality (the involution is without fixed-points). Define

$$2r(w) = \#\{\text{fixed points of } \overline{w}\}.$$

THEOREM 8.2 (Genestier-Ngô [**GN**]). *The $p$-rank is constant on each KR-stratum $Sh_w$. More precisely, the $p$-rank of a point in $Sh_w$ is the integer $r(w)$.*

COROLLARY 8.3 ([**GN**]). *The ordinary locus in $Sh_{\mathbb{F}_p}$ is precisely the union of the KR-strata indexed by the translation elements in $\mathrm{Adm}(\mu)$, that is, the elements $t_\lambda$, for $\lambda \in W\mu$. Moreover, the ordinary locus is dense and open in $Sh_{\mathbb{F}_p}$.*

PROOF. By the theorem, the $p$-rank is $n$ on $Sh_w$ if and only $r(w) = n$; writing $w = t_\lambda \overline{w}$, this is equivalent to $\overline{w} = 1$. We conclude the first statement by noting that $w \in \mathrm{Adm}(\mu) \Rightarrow \lambda \in W\mu$.

Finally, the union of the strata $\mathbf{M}^{\mathrm{loc}}_{t_\lambda}$ for $\lambda \in W\mu$ is clearly dense and open in $\mathbf{M}^{\mathrm{loc}}_{\mathbb{F}_p}$. □

REMARK 8.4. It should be noted that in [**GN**] the local model, and thus the KR-stratification, is defined in terms of the "cohomology" local model diagram, whereas here everything is stated using the "homology" version. Furthermore, in [**GN**] the "standard" lattice chain is "opposite" from ours, so that an element $w \in \widetilde{W}(\mathrm{GSp}_{2n})$ is used to index a double coset $\bar{I}w\bar{I}/\bar{I}$, where $\bar{I}$ is an "opposite" Iwahori subgroup. Nevertheless, our conventions and those of [**GN**] yield *the same answer*, that is, the $p$-rank on $Sh_w$ is given by $r(w)$ is both cases. This may be seen by using the comparison between "homology" and "cohomology" local model diagrams in Prop. 6.8, and by imitating the proof of [**GN**] with our conventions in force.

**8.4. The smooth locus of $Sh_{\mathbb{F}_p}$.** Also, in [**GN**] one finds the proof of the following related fact.

PROPOSITION 8.5 (Genestier-Ngô [**GN**]). *The smooth locus of $Sh_{\mathbb{F}_p}$ is the union of the KR-strata indexed by $t_\lambda$, for $\lambda \in W\mu$ (in particular the smooth locus agrees with the ordinary locus).*

8.4.1. *The geometric proof of* [**GN**]. The crucial observation is that any stratum $Sh_w$, where $w$ is *not* a translation element, is contained in the singular locus. Genestier and Ngô deduce this by showing that for such $w$,

$$\mathrm{Tr}^{ss}(\Phi_{\mathfrak{p}}^r, R\Psi_w^{Sh}(\overline{\mathbb{Q}}_\ell)) \neq 1, \tag{8.4.1}$$

which by the general geometric principle explained below, shows that $w$ is singular. Now (8.4.1) itself is proved by combining the main theorems of [**HN1**] and [**H2**], and by taking into account Lemma 8.1.

Here is the geometric principle implicit in [**GN**] and a sketch of the proof from [**GN**]. We will use freely the material from sections 9.3 and 10 below.

LEMMA 8.6. *Let $S = (S, s, \eta)$ be a trait, with $k(s) = \mathbb{F}_q$ a finite field. Suppose $M \to S$ is a finite type flat model with $M_\eta$ smooth. Then $x \in M(\mathbb{F}_{q^r})$ is a smooth point of $M_{\bar{s}}$ only if $\mathrm{Tr}^{ss}(\Phi_q^r, R\Psi_x^M(\overline{\mathbb{Q}}_\ell)) = 1$.*

PROOF. Let $M' \subset M$ be the open subscheme obtained by removing the singular locus of the special fiber of $M$. We see that $M' \to S$ is smooth (since $M' \to S$ is flat of finite-type, it suffices to check the smoothness fiber by fiber, and by construction $M'_\eta = M_\eta$ and $M'_s$ are both smooth). Now we invoke the general fact (Theorem 10.1) that for smooth models $M'$, $R\Psi^{M'}(\overline{\mathbb{Q}}_\ell) \cong \overline{\mathbb{Q}}_\ell$, the constant sheaf on the special fiber. This implies that the semi-simple trace of nearby cycles at $x \in M'_s$ is 1. $\square$

*Proof of Proposition 8.5.* We consider the stratum $Sh_w$, or equivalently, $\mathbf{M}_w^{\mathrm{loc}}$, for $w \in \mathrm{Adm}(\mu)$. We recall that $\mathbf{M}^{\mathrm{loc}} = M_{-w_0\mu}$ and the stratum $\mathbf{M}_w^{\mathrm{loc}}$ is the Iwahori-orbit indexed by $x := w^{-1}$, contained in $M_{-w_0\mu}$. For such an $x$, we have from [**HN1**], [**H2**], [**HP**] an explicit formula for the semi-simple trace of Frobenius on nearby cycles at $x$

$$\mathrm{Tr}^{ss}(\Phi_q, R\Psi_x^{\mathbf{M}^{\mathrm{loc}}}(\overline{\mathbb{Q}}_\ell)) = (-1)^{\ell(t_\mu)+\ell(x)}\, R_{x,t_{\lambda(x)}}(q),$$

where $\lambda(x)$ is the translation part of $x$ ($x = t_{\lambda(x)}w$, for $\lambda(x) \in X_*$, and $w \in W_0$), and $R_{x,y}(q)$ is the Kazhdan-Lusztig $R$-polynomial. This polynomial can be computed explicitly, but we need only the fact that it is always a polynomial in $q$ of degree $\ell(y) - \ell(x)$. It follows from this and the above lemma that whenever $x$ corresponds to a stratum of codimension $\geq 1$, every point of that stratum is singular.

8.4.2. *A combinatorial proof.* There is however a more elementary way to proceed: we prove below that every codim $\geq 1$ stratum in $\mathbf{M}_{\mathbb{F}_p}^{\mathrm{loc}}$ is contained at least two irreducible components. The same goes for $Sh_{\mathbb{F}_p}$, proving the proposition. (There is even a third proof of the proposition, given in [**GH**].)

PROPOSITION 8.7. *Let $\mu$ be minuscule. For any $x \in \mathrm{Adm}(\mu)$ of codimension 1, there exist exactly two distinct translation elements $\lambda_1, \lambda_2$ in $W\mu$ such that $x \leq t_{\lambda_i}$ (for $i = 1, 2$). Thus, any codimension 1 KR-stratum in the special fiber of a Shimura variety $Sh$ with Iwahori-level structure at $p$ is contained in exactly two irreducible components.*

PROOF. We give a purely combinatorial proof. Suppose $x \in \mathrm{Adm}(\mu)$ has codimension 1. We have $x < t_\nu$, for some $\nu \in W\mu$. By properties of the Bruhat order, there exists an affine reflection $s_{\beta+k}$, where $\beta$ is a $B$-positive root, such that $x = t_\nu s_{\beta+k}$. Since $s_{\beta+k} = t_{-k\beta^\vee} s_\beta$, this means $x = t_{\nu - k\beta^\vee} s_\beta$. The translation part must lie in $\mathrm{Perm}(\mu) \cap X_* = W\mu$, hence we have $x = t_\lambda s_\beta$, for $\lambda \in W\mu$. By comparing lengths, we have $t_\lambda s_\beta < t_\lambda$ and $t_\lambda s_\beta < s_\beta t_\lambda s_\beta = t_{s_\beta \lambda}$. We claim that $\langle \beta, \lambda \rangle < 0$. Indeed, writing $\epsilon_\beta \in [-1, 0)$ for the infimum of the set $\beta(\mathbf{a})$ (recalling that our base alcove $\mathbf{a}$ is contained in the $\bar{B}$-positive chamber) , we have

$$
\begin{aligned}
t_\lambda s_\beta < t_\lambda &\Leftrightarrow \mathbf{a} \text{ and } t_{-\lambda}\mathbf{a} \text{ are on opposite sides of the hyperplane } \beta = 0 \\
&\Leftrightarrow \beta(-\lambda + \mathbf{a}) \subset [0, \infty) \\
&\Leftrightarrow -\langle \beta, \lambda \rangle + (\epsilon_\beta, 0) \subset [0, \infty) \\
&\Leftrightarrow \langle \beta, \lambda \rangle < 0.
\end{aligned}
$$

We see that $s_\beta \lambda \neq \lambda$, and so $x$ precedes at least the two distinct translation elements $t_\lambda$ and $t_{s_\beta \lambda}$ in $\mathrm{Adm}(\mu)$. It remains to prove that these are the only such translation elements. So suppose now that $t_\lambda s_\beta < t_{\lambda'}$, where $\lambda' \in W\mu$; we will show that $\lambda' \in \{\lambda, s_\beta \lambda\}$. As above, there is an affine reflection $s_{\alpha+n}$ such that $t_\lambda s_\beta = t_{\lambda'} s_{\alpha+n} = t_{\lambda' - n\alpha^\vee} s_\alpha$, where $\alpha$ is $B$-positive. We see that $\alpha = \beta$, and $\lambda' - n\beta^\vee = \lambda$. Thus, $\lambda'$, $\lambda$, and $s_\beta \lambda$ all lie on the line $\lambda + \mathbb{R}\beta^\vee$. Since all elements in $W\mu$ are vectors with the same Euclidean length, this can only occur if $\lambda' \in \{\lambda, s_\beta \lambda\}$. □

## 9. Langlands' strategy for computing local $L$-factors

The well-known general strategy for computing the local $L$-factor at $p$ of a Shimura variety in terms of automorphic $L$-functions is due to the efforts of many people, beginning with Eichler, Shimura, Kuga, Sato, and Ihara, and reaching its final conjectural form with Langlands, Rapoport, and Kottwitz.

Let us fix a rational prime $p$, and a compact open subgroup $K_p \subset G(\mathbb{Q}_p)$ at $p$; we consider the Shimura variety $Sh(\mathbf{G}, h)_{\mathbf{K}}$ as in §5.1.

Roughly, the method of Langlands is to start with a cohomological definition of the local factor of the Hasse-Weil zeta function for $Sh(\mathbf{G}, h)_{\mathbf{K}}$, and express its logarithm via the Grothendieck-Lefschetz trace formula as a certain sum of orbital integrals for the group $\mathbf{G}(\mathbb{A})$. This involves both a process of counting points (with "multiplicity" – the trace of the correspondence on the stalk of an appropriate sheaf), and then a "pseudo-stabilization" like that done to stabilize the geometric side of the Arthur-Selberg trace formula (we are ignoring the appearance of endoscopic groups other than $\mathbf{G}$ itself in this stage). At this point, we can apply the Arthur-Selberg stable trace formula and express the sum as a trace of a function on automorphic representations appearing in the discrete part of $L^2(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A}))$. This equality of traces implies a relation like that in Theorem 11.7 below.

More details on the general strategy as well as the precise conjectural description of $IH^i(Sh \times_E \overline{\mathbb{Q}}_p, \overline{\mathbb{Q}}_\ell)$ in terms of automorphic representations of $\mathbf{G}$ can be found in [**Ko90**], [**Ko92b**], and [**BR**]. These sources discuss the case of good reduction at $\mathfrak{p}$.

Some details of the analogous strategy in case of bad reduction will be given below in §11. Let us explain more carefully the relevant definitions.

**9.1. Definition of local factors of the Hasse-Weil Zeta function.** Let $\mathfrak{p}$ denote a prime of a number field $\mathbf{E}$, lying over $p$. Let $X$ be a smooth $d$-dimensional variety over $\mathbf{E}$. We define the Hasse-Weil zeta function (of a complex variable $s$) as an Euler product

$$Z(s,X) = \prod_{\mathfrak{p}} Z_{\mathfrak{p}}(s,X),$$

where the local factors are defined as follows.

DEFINITION 9.1. The factor $Z_{\mathfrak{p}}(s,X)$ is defined to be

$$\prod_{i=0}^{2d} \det(1 - N\mathfrak{p}^{-s}\Phi_{\mathfrak{p}}\,;\, H^i_c(X \times_E \overline{\mathbb{Q}}_p,\, \overline{\mathbb{Q}}_\ell)^{\Gamma^0_{\mathfrak{p}}})^{(-1)^{i+1}},$$

where

- $\ell$ is an auxiliary prime, $\ell \neq p$;
- $\Phi_{\mathfrak{p}}$ is the inverse of an arithemetic Frobenius element for the extension $\mathbf{E}_{\mathfrak{p}}/\mathbb{Q}_p$;
- $N\mathfrak{p} = \mathrm{Norm}_{\mathbf{E}_{\mathfrak{p}}/\mathbb{Q}_p}\mathfrak{p}$;
- $\Gamma^0_{\mathfrak{p}} \subset \Gamma_{\mathfrak{p}} := \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbf{E}_{\mathfrak{p}})$ is the inertia subgroup.

It is believed that $Z(s,X)$ has good analytical properties (it should satisfy a functional equation and have a meromorphic analytic continuation on $\mathbb{C}$) and that analytical invariants (e.g. residues, special values, orders of zeros and poles) carry important arithmetic information about $X$. Moreover, it is believed that the local factor is indeed independent of the choice of the auxiliary prime $\ell$. At present these remain only guiding principles, as very little has been actually proved. As Taniyama originally proposed, a promising strategy for establishing the functional equation is to express the zeta function as a product of automorphic $L$-functions, whose analytic properties are easier to approach. The hope that this can be done is at the heart of the Langlands program.

**9.2. Definition of local factors of automorphic $L$-functions.** Let $\pi_p$ be an irreducible admissible representation of $G(\mathbb{Q}_p)$. Let us assume that the local Langlands conjecture holds for the group $G(\mathbb{Q}_p)$. Then associated to $\pi_p$ is a local Langlands parameter, that is, a homomorphism

$$\varphi_{\pi_p} : W_{\mathbb{Q}_p} \times \mathrm{SL}_2(\mathbb{C}) \to {}^L G,$$

where $W_{\mathbb{Q}_p}$ is the Weil group for $\mathbb{Q}_p$ and, letting $\widehat{G}$ denote the Langlands dual group over $\mathbb{C}$ associated to $G_{\mathbb{Q}_p}$, the $L$-group is defined to be ${}^L G = W_{\mathbb{Q}_p} \ltimes \widehat{G}$. Let $r = (r, V)$ be a rational representation of ${}^L G$. The local $L$-function attached to $\pi_p$ is defined using $\varphi_{\pi_p}$ and $r$ as follows.

DEFINITION 9.2. We define $L(s, \pi_p, r)$ to be

$$\det(1 - p^{-s}\, r\, \varphi_{\pi_p}(\Phi_p \times \begin{bmatrix} p^{-1/2} & 0 \\ 0 & p^{1/2} \end{bmatrix});\, (\ker N)^{\Gamma^0_p})^{-1},$$

where

- $\Phi_p \in W_{\mathbb{Q}_p}$ is the inverse of the arithmetic Frobenius for $\mathbb{Q}_p$;
- $N$ is a nilpotent endomorphism on $V$ coming from the action of $\mathfrak{sl}_2$ on the representation $r\varphi_{\pi_p}$, namely $N := d(r\varphi_{\pi_p})(1 \times \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix})$; equivalently, $N$ is determined by $r\varphi_{\pi_p}(1 \times \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}) = \exp(N)$;

- the action of inertia $\Gamma_p^0 \subset W_{\mathbb{Q}_p}$ on $\ker(N)$ is the restriction of $r\varphi_{\pi_p}$ to $\Gamma_p^0 \times \mathrm{id} \subset W_{\mathbb{Q}_p} \times \mathrm{SL}_2(\mathbb{C})$.

REMARK 9.3. If $\pi_p$ is a spherical representation, then $(\ker N)^{\Gamma_p^0} = V$ and therefore in that case the local factor takes the more familiar form

$$\det(1 - p^{-s} r(\mathrm{Sat}(\pi_p));\, V)^{-1},$$

where $\mathrm{Sat}(\pi_p)$ is the Satake parameter of $\pi_p$; see [**Bo**], [**Ca**].

Any irreducible admissible representation $\pi$ of $\mathbf{G}(\mathbb{A})$ has a tensor factorization $\pi = \otimes_v \pi_v$ ($v$ ranges over all places of $\mathbb{Q}$) where $\pi_v$ is an admissible representation of the local group $\mathbf{G}(\mathbb{Q}_v)$. If $v = \infty$, there is a suitable definition of $L(s, \pi_\infty, r)$ (see [**Ta**]). We then define the automorphic $L$-function

$$L(s, \pi, r) = \prod_v L(s, \pi_v, r).$$

**9.3. Problems in case of bad reduction, and definition of semi-simple local factors.**

9.3.1. *Semi-simple zeta function.* Let us fix $\mathfrak{p}$ and set $E = \mathbf{E}_\mathfrak{p}$. In the case where $X$ possesses an integral model over $\mathcal{O}_E$, one can study the local factor by reduction modulo $\mathfrak{p}$. In the case of good reduction (meaning this model is smooth over $\mathcal{O}_E$), the inertia group acts trivially and the cohomology of $X \times_E \overline{\mathbb{Q}}_p$ can be identified with that of the special fiber $X \times_{k_E} \overline{\mathbb{F}}_p$ ($k_E$ being the residue field of $E$). By the Grothendieck-Lefschetz trace formula, the local zeta function then satisfies the familiar identity

$$\log(Z_\mathfrak{p}(s, X)) = \sum_{r=1}^{\infty} \#X(k_{E,r})\, \frac{N\mathfrak{p}^{-rs}}{r},$$

where $k_{E,r}$ is the degree $r$ extension of $k_E$ in its algebraic closure $k$. In the case of bad reduction, inertia can act non-trivially and the smooth base-change theorems of $\ell$-adic cohomology no longer apply in such a simple way. Following Rapoport [**R1**], we bypass the first difficulty by working with the *semi-simple* local zeta function, defined below. The second difficulty forces us to work with nearby cycles (see §10): if $X$ is a proper scheme over $\mathcal{O}_E$, then there is a $\Gamma_\mathfrak{p}$-equivariant isomorphism

$$H^i(X \times_E \overline{\mathbb{Q}}_p,\, \overline{\mathbb{Q}}_\ell) = H^i(X \times_{k_E} \overline{k}_E,\, R\Psi(\overline{\mathbb{Q}}_\ell)),$$

so that the Grothendieck-Lefschetz trace formula gives rise to the problem of "counting counts $x \in X(k_{E,r})$ with multiplicity", i.e., to computing the semi-simple trace on the stalks of the complex of nearby cycles:

$$\mathrm{Tr}^{ss}(\Phi_\mathfrak{p}^r, R\Psi(\overline{\mathbb{Q}}_\ell)_x),$$

in order to understand the semi-simple zeta function.

How do we define semi-simple trace and semi-simple zeta functions? We recall the discussion from [**HN1**]. Suppose $V$ is a finite-dimensional continuous $\ell$-adic representation of the Galois group $\Gamma_\mathfrak{p}$. Grothendieck's local monodromy theorem states that this representation is necessarily *quasi-unipotent*: there exists a finite index subgroup $\Gamma_1 \subset \Gamma_\mathfrak{p}^0$ such that $\Gamma_1$ acts unipotently on $V$. Hence there is a finite increasing $\Gamma_\mathfrak{p}$-invariant filtration $V_\bullet = (0 \subset V_1 \subset \cdots \subset V_m = V)$ such that $\Gamma_\mathfrak{p}^0$ acts on $\oplus_k gr_k V_\bullet$ through a finite quotient. Such a filtration is called *admissible*. Then we define

$$\mathrm{Tr}^{ss}(\Phi_{\mathfrak{p}}; V) = \sum_k \mathrm{Tr}(\Phi_{\mathfrak{p}}; gr_k(V_\bullet)^{\Gamma^0_{\mathfrak{p}}}).$$

This is independent of the choice of admissible filtration. Moreover, the function $V \mapsto \mathrm{Tr}^{ss}(\Phi_{\mathfrak{p}}, V)$ factors through the Grothendieck group of the category of $\ell$-adic representations $V$, and using this one proves that it extends naturally to give a "sheaf-function dictionary" à la Grothendieck: a complex $\mathcal{F}$ in the "derived" category $D^b_c(X \times_\eta s, \overline{\mathbb{Q}}_\ell)$, [9] gives rise to the $\overline{\mathbb{Q}}_\ell$-valued function

$$x \mapsto \mathrm{Tr}^{ss}(\Phi^r_{\mathfrak{p}}, \mathcal{F}_x)$$

on $X(k_{E,r})$. Furthermore, the formation of this function is compatible with the pull-back and proper-push-forward operations on the derived catgegory, and a Grothendieck-Leftschetz trace formula holds. (For details, see [**HN1**].) We can then *define* the semi-simple local zeta function $Z^{ss}_{\mathfrak{p}}(s, X)$ by the identity

DEFINITION 9.4.

$$\log(Z^{ss}_{\mathfrak{p}}(s, X)) = \sum_{r=1}^{\infty} \big( \sum_i (-1)^i \,\mathrm{Tr}^{ss}(\Phi^r_{\mathfrak{p}};\, H^i_c(X \times_E \overline{\mathbb{Q}}_p,\, \overline{\mathbb{Q}}_\ell))\big)\ \frac{N\mathfrak{p}^{-rs}}{r}.$$

REMARK 9.5. Note that in the case where $\Gamma^0_{\mathfrak{p}}$ acts unipotently (not just quasi-unipotently) on the cohomology of the generic fiber, then we have

$$\mathrm{Tr}^{ss}(\Phi^r_{\mathfrak{p}};\, H^i_c(X \times_E \overline{\mathbb{Q}}_p,\, \overline{\mathbb{Q}}_\ell)) = \mathrm{Tr}(\Phi^r_{\mathfrak{p}};\, H^i_c(X \times_E \overline{\mathbb{Q}}_p,\, \overline{\mathbb{Q}}_\ell)).$$

As we shall see below in §10.2, $\Gamma^0_{\mathfrak{p}}$ does indeed act unipotently on the cohomology of a proper Shimura variety with Iwahori-level structure at $p$. The definition of $Z^{ss}$ thus simplifies in that case.

As before, the global semi-simple zeta function is defined to be an Euler product over all finite places of the local functions: $Z^{ss}(s, X) = \prod_{\mathfrak{p}} Z^{ss}_{\mathfrak{p}}(s, X)$.

9.3.2. *Semi-simple local L-functions.* Retain the notation of §9.2. The Langlands parameters $\varphi = \varphi_{\pi_p}$ being used here have the property that the representation $r\varphi$ on $V$ arises from a representation $(\rho, N)$ of the Weil-Deligne group $W'_{\mathbb{Q}_p}$ on $V$, see [**Ta**]. In that case we have for $w \in W_{\mathbb{Q}_p}$

$$\rho(w) = r\varphi(w \times \begin{bmatrix} \|w\|^{1/2} & 0 \\ 0 & \|w\|^{-1/2} \end{bmatrix})$$
$$\exp(N) = r\varphi(1 \times \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}),$$

where $\|w\|$ is the power to which $w$ raises elements in the residue field of $\mathbb{Q}_p$.

Now suppose that via some choice of isomorphism $\overline{\mathbb{Q}}_\ell \cong \mathbb{C}$, the pair $(\rho, N)$ comes from an $\ell$-adic representation $(\rho_\lambda, V_\lambda)$ of $W_{\mathbb{Q}_p}$, by the rule

$$\rho_\lambda(\Phi^n\sigma) = \rho(\Phi^n\sigma)\exp(N\, t_\ell(\sigma)),$$

where $n \in \mathbb{Z}$, $\sigma \in \Gamma^0_p$, and $t_\ell : \Gamma^0_p \to \mathbb{Q}_\ell$ is a nonzero homomorphism (cf. [**Ta**], Thm 4.2.1).

[9] This is the "derived category of $\overline{\mathbb{Q}}_\ell$-sheaves" – although this is somewhat misleading terminology (see [**KW**] for a detailed account) – on $X \times_{k_E} \overline{k_E}$ equipped with a continuous action of $\Gamma_{\mathfrak{p}}$ compatible with the action of its quotient $\mathrm{Gal}(\overline{k_E}/k_E)$ on $X \times \overline{k_E}$. See §10.

For each $\sigma \in \Gamma_p^0$, $\rho_\lambda(\sigma) = \rho(\sigma)\exp(Nt_\ell(\sigma))$ is the multiplicative Jordan decomposition of $\rho_\lambda(\sigma)$ into its semi-simple and unipotent parts, respectively. Therefore a vector $v \in V$ is fixed by $\rho_\lambda(\sigma)$ if and only if it is fixed by both $\rho(\sigma)$ and $\exp(Nt_\ell(\sigma))$. Thus we have the following result.

LEMMA 9.6. *If $(\rho, N, V)$ and $(\rho_\lambda, V_\lambda)$ are the two avatars above of a representation of the Weil-Deligne group $W'_{\mathbb{Q}_p}$, then*

$$(\ker N)^{\Gamma_p^0} = V_\lambda^{\Gamma_p^0}.$$

*Furthermore $\rho$ is trivial on $\Gamma_p^0$ if and only if $\rho_\lambda(\Gamma_p^0)$ acts unipotently on $V_\lambda$, and in that case*

$$(\ker N)^{\Gamma_p^0} = \ker N = V_\lambda^{\Gamma_p^0}.$$

COROLLARY 9.7. *The local $L$-function can also be expressed as*

$$L(s, \pi_p, r) = \det(1 - p^{-s}\rho_\lambda(\Phi_p);\, V_\lambda^{\Gamma_p^0})^{-1}.$$

Note the similarity with Definition 9.1. The representation $\rho_\lambda(\Gamma_p^0)$ being quasi-unipotent, we can define the semi-simple $L$-function as in Definition 9.4.

DEFINITION 9.8.

$$\log(L^{ss}(s, \pi_p, r)) = \sum_{r=1}^{\infty} \mathrm{Tr}^{ss}(\rho_\lambda(\Phi_p^r);\, V_\lambda)\ \frac{p^{-rs}}{r}.$$

(The symbol $r$ occurs with two different meanings here, but this should not cause confusion.)

REMARK 9.9. In analogy with Remark 9.5, in the case that $(\rho, N, V)$ has $\rho(\Gamma_p^0) = 1$, in view of Lemma 9.6, we have

$$\mathrm{Tr}^{ss}(\rho_\lambda(\Phi_p^r);\, V) = \mathrm{Tr}(\rho_\lambda(\Phi_p^r);\, V),$$

and the definition of $L^{ss}$ simplifies. The dictionary set-up by the local Langlands correspondence asserts that if $\pi_p$ has an Iwahori-fixed vector, then $\rho(\Gamma_p^0)$ is trivial (cf. [**W2**]). Moreover, in the situation of parahoric level structure at $p$, the only representations $\pi_p$ which will arise necessarily have Iwahori-fixed vectors. Hence the simplified definition of $L^{ss}$ will apply in our situation.

Let $\pi = \otimes_v \pi_v$ be an irreducible admissible representation of $\mathbf{G}(\mathbb{A})$. It is to be hoped that a reasonable definition of $L^{ss}(s, \pi_v, r)$ exists for Archimedean places $v$, and if so we can then define

$$L^{ss}(s, \pi, r) = \prod_v L^{ss}(s, \pi_v, r).$$

## 10. Nearby cycles

**10.1. Definitions and general facts.** Let $X$ be a scheme of finite type over a finite (or algebraically closed) field $k$. (The following also works if we assume that $k$ is the fraction field of a discrete valuation ring $R$ with finite residue field, and that $X$ is finite-type over $R$, cf. [**Ma**].) Denote by $\overline{k}$ an algebraic closure of $k$, and by $X_{\overline{k}}$ the base change $X \times_k \overline{k}$.

We denote by $D_c^b(X, \overline{\mathbb{Q}}_\ell)$ the 'derived' category of $\overline{\mathbb{Q}}_\ell$-sheaves on $X$. Note that this is not actually the derived category of the category of $\overline{\mathbb{Q}}_\ell$-sheaves, but is defined via a limit process. See [**BBD**] 2.2.14 or [**Weil2**] 1.1.2, or [**KW**] for more details. Nevertheless, $D^b_c(X, \overline{\mathbb{Q}}_\ell)$ is a triangulated category which admits the usual functorial formalism, and which can be equipped with a 'natural' $t$-structure having as its core the category of $\overline{\mathbb{Q}}_\ell$-sheaves. If $f : X \longrightarrow Y$ is a morphism of schemes of finite type over $k$, we have the derived functors $f_*,\ f_! : D^b_c(X, \overline{\mathbb{Q}}_\ell) \to D^b_c(Y, \overline{\mathbb{Q}}_\ell)$ and $f^*,\ f^! : D^b_c(Y, \overline{\mathbb{Q}}_\ell) \longrightarrow D^b_c(X, \overline{\mathbb{Q}}_\ell)$. Ocassionally we denote these same functors using the symbols $Rf_*$, etc.

Let $(S, s, \eta)$ denote an Henselian trait: $S$ is the spectrum of a complete discrete valuation ring, with special point $s$ and generic point $\eta$. The key examples for us are $S = \mathrm{Spec}(\mathbb{Z}_p)$ (the $p$-adic setting) and $S = \mathrm{Spec}(\mathbb{F}_p[[t]])$ (the function-field setting). Let $k(s)$ resp. $k(\eta)$ denote the residue fields of $s$ resp. $\eta$.

We choose a separable closure $\bar{\eta}$ of $\eta$ and define the Galois group $\Gamma = \mathrm{Gal}(\bar{\eta}/\eta)$ and the inertia subgroup $\Gamma_0 = \ker[\mathrm{Gal}(\bar{\eta}/\eta) \to \mathrm{Gal}(\bar{s}/s)]$, where $\bar{s}$ is the residue field of the normalization $\bar{S}$ of $S$ in $\bar{\eta}$.

Now let $X$ denote a finite-type scheme over $S$. The category $D^b_c(X \times_s \eta, \overline{\mathbb{Q}}_\ell)$ is the category of sheaves $\mathcal{F} \in D^b_c(X_{\bar{s}}, \overline{\mathbb{Q}}_\ell)$ together with a continuous action of $\mathrm{Gal}(\overline{\eta}/\eta)$ which is compatible with the action on $X_{\bar{s}}$. (Continuity is tested on cohomology sheaves.)

For $\mathcal{F} \in D^b_c(X_\eta, \overline{\mathbb{Q}}_\ell)$, we define the *nearby cycles sheaf* to be the object in $D^b_c(X \times_s \eta, \overline{\mathbb{Q}}_\ell)$ given by

$$R\Psi^X(\mathcal{F}) = \bar{i}^* R\bar{j}_*(\mathcal{F}_{\bar{\eta}}),$$

where $\bar{i} : X_{\bar{s}} \hookrightarrow X_{\bar{S}}$ and $\bar{j} : X_{\bar{\eta}} \hookrightarrow X_{\bar{S}}$ are the closed and open immersions of the geometric special and generic fibers of $X/S$, and $\mathcal{F}_{\bar{\eta}}$ is the pull-back of $\mathcal{F}$ to $X_{\bar{\eta}}$.

Here we list the basic properties of $R\Psi^X$, extracted from the standard references: [**BBD**], [**Il**], [**SGA7 I**], [**SGA7 XIII**]. The final listed property has been proved (in this generality) only recently, and is due to the author and U. Görtz [**GH**].

THEOREM 10.1. *The following properties hold for the functors*

$$R\Psi : D^b_c(X_\eta, \overline{\mathbb{Q}}_\ell) \to D^b_c(X \times_s \eta, \overline{\mathbb{Q}}_\ell) :$$

(a) $R\Psi$ *commutes with proper-push-forward: if* $f : X \to Y$ *is a proper* $S$*-morphism, then the canonical base-change morphism of functors to* $D^b_c(Y \times_s \eta, \overline{\mathbb{Q}}_\ell)$ *is an isomorphism:*

$$R\Psi\, f_* \xrightarrow{\sim} f_*\, R\Psi;$$

*In particular, if* $X \to S$ *is proper there is a* $\mathrm{Gal}(\bar{\eta}/\eta)$*-equivariant isomorphism*

$$H^i(X_{\bar{\eta}}, \overline{\mathbb{Q}}_\ell) = H^i(X_{\bar{s}}, R\Psi(\overline{\mathbb{Q}}_\ell)).$$

(b) *Suppose* $f : X \to S$ *is finite-type but not proper. Suppose that there is a compactification* $j : X \hookrightarrow \overline{X}$ *over* $S$ *such that the boundary* $\overline{X}\backslash X$ *is a relative normal crossings divisor over* $S$. *Then there is a* $\mathrm{Gal}(\bar{\eta}/\eta)$*-equivariant isomorphism*

$$H^i_c(X_{\bar{\eta}}, \overline{\mathbb{Q}}_\ell) = H^i_c(X_{\bar{s}}, R\Psi(\overline{\mathbb{Q}}_\ell)).$$

(c) $R\Psi$ *commutes with smooth pull-back: if* $p : X \to Y$ *is a smooth* $S$*-morphism, then the base-change morphism is an isomorphism:*

$$p^* R\Psi \xrightarrow{\sim} R\Psi p^*.$$

(d) *If $\mathcal{F} \in D^b_c(X, \overline{\mathbb{Q}}_\ell)$, we define $R\Phi(\mathcal{F})$ (the* vanishing cycles*) to be the cone of the canonical morphism*

$$\mathcal{F}_{\bar{s}} \to R\Psi(\mathcal{F}_\eta);$$

*there is a distinguished triangle*

$$\mathcal{F}_{\bar{s}} \to R\Psi(\mathcal{F}_\eta) \to R\Phi(\mathcal{F}) \to \mathcal{F}_{\bar{s}}[1].$$

*If $X \to S$ is smooth, then $R\Phi(\overline{\mathbb{Q}}_\ell) = 0$; in particular, $\overline{\mathbb{Q}}_\ell \cong R\Psi(\overline{\mathbb{Q}}_\ell)$ in this case;*

(e) *$R\Psi$ commutes with Verdier duality, and preserves perversity of sheaves (for the middle perversity);*

(f) *For $x \in X_{\bar{s}}$, $R^i\Psi(\overline{\mathbb{Q}}_\ell)_x = H^i(X_{(\bar{x})\bar{\eta}}, \overline{\mathbb{Q}}_\ell)$, where $X_{(\bar{x})\bar{\eta}}$ is the fiber over $\bar{\eta}$ of the strict henselization of $X$ in a geometric point $\bar{x}$ with center $x$. In particular, the support of $R\Psi(\overline{\mathbb{Q}}_\ell)$ is contained in the scheme-theoretic closure of $X_\eta$ in $X_{\bar{s}}$;*

(g) *If the generic fiber $X_\eta$ is non-singular, then the complex $R\Psi^X(\overline{\mathbb{Q}}_\ell)$ is mixed, in the sense of* [**Weil2**].

Remark 10.2. The "fake" unitary Shimura varieties $Sh(\mathbf{G}, h)_\mathbf{K}$ discussed in §5.2 are proper over $\mathcal{O}_E$ and hence by (a) we can use nearby cycles to study their semi-simple local zeta functions. The Siegel modular schemes of §5.3 are not proper over $\mathbb{Z}_p$, so (a) does not apply directly; in fact it is not a priori clear that there is a Galois equivariant isomorphism on cohomology with compact supports as in (b). In order to apply the method of nearby cycles to study the semi-simple local zeta function, we need such an isomorphism.

Conjecture 10.3. *Let $Sh_{K_p}$ denote a model over $\mathcal{O}_E$ for a PEL Shimura variety with parahoric level structure $K_p$, as in §5. Then the natural morphism*

$$H^i_c(Sh_{K_p} \times_{\mathcal{O}_E} \overline{k_E},\, R\Psi(\overline{\mathbb{Q}}_\ell)) \to H^i_c(Sh_{K_p} \times_{\mathcal{O}_E} \overline{\mathbb{Q}}_p,\, \overline{\mathbb{Q}}_\ell)$$

*is an isomorphism.*

In the Siegel case one should be able to prove this by finding a suitably nice compactification, perhaps by adapting the methods of [**CF**].

Such an isomorphism would allow us to study the semi-simple local zeta function in the Siegel case by the same approach applied to the "fake" unitary case in §11. The local geometric problems involving nearby cycles have already been resolved in [**HN1**], and Conjecture 10.3 encapsulates the remaining geometric difficulty (which is global in nature). There will be additional group-theoretic problems in applying the Arthur-Selberg trace formula, however, due to endoscopy.

**10.2. Concerning the inertia action on certain nearby cycles.** Let $Sh(\mathbf{G}, h)_\mathbf{K}$ denote a "fake" unitary Shimura variety as in §5.2, where $K_p$ is an Iwahori subgroup. Suppose $Sh$ is its integral model over $\mathcal{O}_E = \mathbb{Z}_p$ defined by the moduli problem in §5.2.4. Our goal in §11 is to explain how to identify its $Z^{ss}_\mathfrak{p}$ with a product of semi-simple local $L$-functions (see Theorem 11.7). We first want to justify our earlier claim that the simplified definitions of $Z^{ss}$ and $L^{ss}$ apply to this case. The key ingredient is a theorem of D. Gaitsgory showing that the inertia action on certain nearby cycles is unipotent.

We first recall the theorem of Gaitsgory, which is contained in [**Ga**]. Let $\lambda$ denote a dominant coweight of $G_{\overline{\mathbb{Q}}_p}$ and consider the corresponding $G(\mathbb{Q}_p[[t]])$-orbit $\mathcal{Q}_\lambda$ in the affine Grassmannian $\mathrm{Grass}_{\overline{\mathbb{Q}}_p} = G(\overline{\mathbb{Q}}_p((t)))/G(\overline{\mathbb{Q}}_p[\![t]\!])$. Let $M_\lambda$ denote the scheme-theoretic closure of $\mathcal{Q}_\lambda$ in the deformation $M$ of $\mathrm{Grass}_{\overline{\mathbb{Q}}_p}$ to $\mathcal{F}l_{\overline{\mathbb{F}}_p}$, from Remark 4.1. Let $IC_\lambda$ denote the intersection complex of the closure $\overline{\mathcal{Q}}_\lambda$.

THEOREM 10.4 (Gaitsgory [**Ga**]). *The inertia group $\Gamma_p^0$ acts unipotently on $R\Psi^{M_\lambda}(IC_\lambda)$.*

See also [**GH**], §5, for a detailed proof of this theorem. We remark that Gaitsgory proves this statement for nearby cycles taken with respect to Beilinson's deformation of the affine Grassmannian of $G(\mathbb{F}_p[\![t]\!])$ to its affine flag variety, but the same proof applies in the present $p$-adic setting; see loc. cit.

We can apply this to the local model $\mathbf{M}^{\mathrm{loc}} = M_{-w_0\mu}$. Because the morphisms in the local model diagram are smooth and surjective (Lemma 13.1), by taking [**GH**], Lemma 5.6 into account, we see that unipotence of nearby cycles on $\mathbf{M}^{\mathrm{loc}}$ implies unipotence of nearby cycles on $Sh$:

COROLLARY 10.5. *The inertia group $\Gamma_p^0$ acts unipotently on $R\Psi^{Sh}(\overline{\mathbb{Q}}_\ell)$. Consequently, by Theorem 10.1 (a), it also acts unipotently on $H^i(Sh \times_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p, \overline{\mathbb{Q}}_\ell)$.*

This justifies the assertion made in Remark 9.5. Together with Remark 9.9, we thus see that the simplified definitions of $Z^{ss}$ and $L^{ss}$ apply in this case, which will be helpful in §11 below.

## 11. The semi-simple local zeta function for "fake" unitary Shimura varieties

We assume in this section that $Sh$ is the model from §5.2: a "fake" unitary Shimura variety. We also assume that $K_p$ is the "standard" Iwahori subgroup of $\mathbf{G}(\mathbb{Q}_p)$, i.e., the subgroup stabilizing the "standard" self-dual multichain of $\mathcal{O}_B \otimes \mathbb{Z}_p$-lattices

$$\widetilde{\Lambda}_\bullet = \Lambda_\bullet \oplus \Lambda_\bullet^*$$

from §5.2.3.

Following the strategy of Kottwitz [**Ko92**], [**Ko92b**], we will explain how to express $Z_{\mathfrak{p}}^{ss}(s, Sh)$ in terms of the functions $L^{ss}(s, \pi_p, r)$.

There are two equations to be proved. The first equation is an expression for the semi-simple Lefschetz number

$$\sum_{x \in Sh(k_r)} \mathrm{Tr}^{ss}(\Phi_{\mathfrak{p}}^r, R\Psi_x(\overline{\mathbb{Q}}_\ell)) = \sum_{\gamma_0} \sum_{(\gamma,\delta)} c(\gamma_0; \gamma, \delta)\ O_\gamma(f^p)\ TO_{\delta\sigma}(\phi_r). \tag{11.0.1}$$

The left hand side is termed the semi-simple Lefschetz number $\mathrm{Lef}^{ss}(\Phi_{\mathfrak{p}}^r)$. The right hand side has exactly the same form (with essentially the same notation, see below) as [**Ko92**], p. 442 [10]. Recall that the twisted orbital integral is defined as

$$TO_{\delta\sigma}(\phi_r) = \int_{\mathbf{G}_{\delta\sigma} \backslash \mathbf{G}(L_r)} \phi_r(x^{-1}\delta\sigma(x)),$$

[10]Note that the factor $|\ker^1(\mathbb{Q}, \mathbf{G})| = |\ker^1(\mathbb{Q}, Z(\mathbf{G}))|$ appearing in loc. cit. does not appear here since our assumptions on $\mathbf{G}$ guarantee that this number is 1; see loc. cit. §7.

with an appropriate choice of measures (and $O_\gamma(f^p)$ is similarly defined). However in contrast to loc. cit., here the Haar measure on $\mathbf{G}(L_r)$ is the one giving the standard Iwahori subgroup $I_r \subset \mathbf{G}(L_r)$ volume 1.

The second equation relates the sum of (twisted) orbital integrals on the right to the spectral side of the Arthur-Selberg trace formula for $\mathbf{G}$:

$$\sum_{\gamma_0} \sum_{(\gamma,\delta)} c(\gamma_0;\gamma,\delta)\ O_\gamma(f^p)\ TO_{\delta\sigma}(\phi_r) = \sum_\pi m(\pi)\ \mathrm{Tr}\,\pi(f^p f_p^{(r)} f_\infty). \tag{11.0.2}$$

All notation on the right hand side is as in [**Ko92b**] (cf. §4) where the "fake" unitary Shimura varieties were analyzed in the case that $K_p$ is a *hyperspecial maximal compact* subgroup rather than an Iwahori. In particular, $\pi$ ranges over irreducible admissible representations of $\mathbf{G}(\mathbb{A})$ which occur in the discrete part of $L^2(\mathbf{G}(\mathbb{Q})A_{\mathbf{G}}(\mathbb{R})^0\backslash\mathbf{G}(\mathbb{A}))$, with multiplicity $m(\pi)$. The function $f^p \in C_c^\infty(K^p\backslash\mathbf{G}(\mathbb{A}_f^p)/K^p)$ is just the characteristic function $\mathbb{I}_{K^p}$ of $K^p$, whereas $f_\infty$ is the much more mysterious function from [**Ko92b**], §1 [11]. The function $f_p^{(r)}$ is a kind of "base-change" of $\phi_r$ and will be further explained below.

The equality (11.0.2) comes from the "pseudo-stabilization" of its left hand side, similar to that done in [**Ko92b**], §4. One important ingredient in that is the "base-change fundamental lemma" between the test function $\phi_r$ and its "base-change" $f_p^{(r)}$; the novel feature here is that the function $\phi_r$ is more complicated than when $K_p$ is maximal compact: it is not a spherical function, but rather an element in the center of an Iwahori-Hecke algebra (see below). The "base-change fundamental lemma" for such functions is proved in [**HN3**], and further discussion will be omitted here. Finally, after pseudo-stablilization and the fundamental lemma, we apply a simple form of the Arthur-Selberg trace formula, which produces the right hand side of (11.0.2). See [**HN3**] for details.

Our object here is to explain (11.0.1), following the strategy of [**Ko92**] which handles the case where $K_p$ is maximal compact (the case of good reduction). The main difficulty is to identify the test function $\phi_r$ that appears in the right hand side.

**11.1. Finding the test function $\phi_r$ via the Kottwitz conjecture.** To understand the function $\phi_r$, we will use the full strength of our description of local models in §6.3, in particular §6.3.3, and also the material in the appendix §14.

In (11.0.1), the index $\gamma_0$ roughly parametrizes polarized $n^2$-dimensional $B$-abelian varieties over $k_r$, up to $\overline{\mathbb{Q}}$-isogeny. The index $(\gamma,\delta)$ roughly parametrizes those polarized $n^2$-dimensional $B$-abelian varieties, up to $\mathbb{Q}$-isogeny, which belong to the $\overline{\mathbb{Q}}$-isogeny class indexed by $\gamma_0$. (For precise statements, see [**Ko92**].) Therefore, the summand roughly counts (with "multiplicity") the elements in $Sh(k_r)$ which belong to a fixed $\mathbb{Q}$-isogeny class. We will make this last statement precise, and also explain the crucial meaning of "multiplicity" here.

Let us fix a polarized $n^2$-dimensional $B$-abelian variety over $k_r$, up to isomorphism: $(A', \lambda', i')$. We assume it possesses some $K^p$-level structure $\bar\eta'$. We fix once and for all an isomorphism of skew-Hermitian $\mathcal{O}_B \otimes \mathbb{A}_f^p$-modules

$$V \otimes \mathbb{A}_f^p = H_1(A', \mathbb{A}_f^p), \tag{11.1.1}$$

[11] Up to the sign $(-1)^{\dim(Sh_E)}$, a pseudo-coefficient of a representation $\pi_\infty^0$ in the packet of discrete series $\mathbf{G}(\mathbb{R})$-representations having trivial central and infinitesimal characters.

(in the terminology of [**Ko92**], §4). Since it comes from a level-structure $\bar{\eta}'$, this isomorphism is Galois-equivariant (with the trivial action on $V \otimes \mathbb{A}_f^p$).

Associated to $(A', \lambda', i')$ is also an $L_r$-isocrystal $(H'_{L_r}, \Phi)$ as in §14. In brief, $H' = H(A')$ is the the $W(k_r)$-dual of $H^1_{\mathrm{crys}}(A'/W(k_r))$, and $\Phi$ is the $\sigma$-linear bijection on $H'_{L_r}$ such that $p^{-1}H' \supset \Phi H' \supset H'$ (i.e., $\Phi$ is $V^{-1}$, where $V$ is the Verschiebung from Cor. 14.4). Because of (11.1.1) and the determinant condition, there is also an isomorphism of skew-Hermitian $\mathcal{O}_B \otimes L_r$-modules

$$V \otimes_{\mathbb{Q}_p} L_r = H'_{L_r} \tag{11.1.2}$$

which we also fix once and for all. (See [**Ko92**], p. 430.)

From these isomorphisms we construct the elements $(\gamma_0; \gamma, \delta)$ that appear in (11.0.1). Namely, the absolute Frobenius $\pi_{A'}$ for $A'/k_r$ acting on $H_1(A', \mathbb{A}_f^p)$ induces the automorphism $\gamma^{-1} \in \mathbf{G}(\mathbb{A}_f^p)$, and the $\sigma$-linear bijection $\Phi$ acting on $H'_{L_r}$ induces the element $\delta\sigma$, for $\delta \in \mathbf{G}(L_r)$. The element $\gamma_0 \in \mathbf{G}(\mathbb{Q})$ is constructed from $(\gamma, \delta)$ as in [**Ko92**], §14. The existence of $\gamma_0$ is proved roughly as follows. By Cor. 14.4, we have $\Phi^r = \pi_{A'}^{-1}$ acting on $H'_{L_r}$. Hence the elements $\gamma_l \in \mathbf{G}(\mathbb{Q}_l)$ (for $l \neq p$) and $N\delta \in \mathbf{G}(L_r)$ come from $l$-adic (resp. $p$-adic) realizations of the endomorphism $\pi_{A'}^{-1}$ of $A'$. Using Honda-Tate theory (and a bit more), one can view $\pi_{A'}^{-1}$ as a semi-simple element $\gamma_0 \in \mathbf{G}(\mathbb{Q})$, which is well-defined up to stable conjugacy. By its very construction, $\gamma_0$ is stably conjugate to $\gamma$, resp. $N\delta$.

Let $(A_\bullet, \lambda, i, \bar{\eta}) \in Sh(k_r)$ be a point in the moduli problem (§5.2.4). We want to classify those such that $(A_0, \lambda, i)$ is $\mathbb{Q}$-isogenous to $(A', \lambda', i')$. Let us consider the category

$$\{(A_\bullet, \lambda, i, \xi)\}$$

consisting of chains of polarized $\mathcal{O}_B$-abelian varieties over $k_r$ (up to $\mathbb{Z}_{(p)}$-isogeny), equipped with a $\mathbb{Q}$-isogeny of polarized $\mathcal{O}_B$-abelian varieties $\xi : A_0 \to A'$, defined over $k_r$. Of course the integral "isocrystal" functor $A \mapsto H(A)$ of §14 is a covariant functor from this category to the category of $W(k_r)$-free $\mathcal{O}_B \otimes W(k_r)$-modules in $H'_{L_r}$ equipped with Frobenius and Verschiebung endomorphisms. In fact, $(A_\bullet, \lambda, i, \xi) \mapsto \xi(H(A_\bullet))$ gives an isomorphism

$$\{(A_\bullet, \lambda, i, \xi)\} \xrightarrow{\sim} Y_p,$$

where $Y_p$ is the category consisting of all type $(\widetilde{\Lambda}_\bullet)$ multichains of $\mathcal{O}_B \otimes W(k_r)$-lattices $H_\bullet$ in $H'_{L_r} = V \otimes L_r$, self-dual up to a scalar in $\mathbb{Q}^\times$ [12], such that for each $i$, $p^{-1}H_i \supset \Phi H_i \supset H_i$, and $\sigma^{-1}(\Phi H_i/H_i)$ satisfies the determinant condition.

11.1.1. *Interpreting the determinant condition.* The final condition on $H_i$ comes from the determinant condition on $\mathrm{Lie}(A_i)$, and Cor. 14.4. Let us see what this means more concretely. By Morita equivalence (see §6.3.3), a type $(\widetilde{\Lambda}_\bullet)$ multichain of $\mathcal{O}_B \otimes W(k_r)$-lattices $H_\bullet$, self-dual up to a scalar in $\mathbb{Q}^\times$, inside $H'_{L_r} = V \otimes_{\mathbb{Q}_p} L_r$ can be regarded as a complete $W(k_r)$-lattice chain $H^0_\bullet$ in $L_r^n$. By working with $H^0_i$ instead of $H_i$, we can work in $L_r^n$ instead of $V \otimes_{\mathbb{Q}_p} L_r$ (which has $\dim_{L_r} = 2n^2$). Recall our minuscule coweight $\mu = (0^{n-d}, (-1)^d)$ of $\mathrm{GL}_n$ (§5.2.2),and write $\Phi$ for the Morita equivalent $\sigma$-linear bijection of $L_r^n$. The determinant condition now reads

$$\Phi H^0_i / H^0_i \cong \sigma(k_r)^d,$$

[12] In particular, these multichains are *polarized* in the sense of [**RZ**], Def. 3.14.

that is, the relative position of the $W(k_r)$-lattices $H_i^0$ and $\Phi H_i^0$ in $L_r^n$ is given by

$$\mathrm{inv}_K(H_i^0, \Phi H_i^0) = \sigma(\mu(p)) = \mu(p).$$

(We write $\mu(p)$ in place of $\mu$ to emphasize our convention that coweights $\lambda$ are embedded in $\mathrm{GL}_n(\mathbb{Q}_p)$ by the rule $\lambda \mapsto \lambda(p)$.) The same identity holds for $H_i$ replacing $H_i^0$, when we interpret $\mu$ as a coweight for the group $\mathbf{G}(L_r) \subset \mathrm{Aut}_B(V \otimes_{\mathbb{Q}_p} L_r)$.

By Theorem 6.3 (and the proof of [**Ko92**], Lemma 7.2), we may find $x \in \mathbf{G}(L_r)$ such that

$$H_\bullet = x\widetilde{\Lambda}_{\bullet, W(k_r)},$$

where $\widetilde{\Lambda}_{\bullet, W(k_r)} = \widetilde{\Lambda}_\bullet \otimes_{\mathbb{Z}_p} W(k_r)$ is the "standard" self-dual multichain of $\mathcal{O}_B \otimes W(k_r)$-lattices in $V \otimes_{\mathbb{Q}_p} L_r$.

The determinant condition now reads: for every index $i$ in the chain $\widetilde{\Lambda}_\bullet$,

$$\mathrm{inv}_K(\widetilde{\Lambda}_{i, W(k_r)},\, x^{-1}\delta\sigma(x)\widetilde{\Lambda}_{i, W(k_r)}) = \mu(p). \tag{11.1.3}$$

Letting $I_r \subset \mathbf{G}(L_r)$ denote the stabilizer of $\widetilde{\Lambda}_{\bullet, W(k_r)}$, we have the Bruhat-Tits decomposition

$$\widetilde{W}(\mathrm{GL}_n \times \mathbb{G}_m) \cong \widetilde{W}(\mathbf{G}) = I_r \backslash \mathbf{G}(L_r)/I_r.$$

Equation (11.1.3) recalls the definition of the $\mu$-permissible set (§4.3). The determinant condition can now be interpreted as:

$$x^{-1}\delta\sigma(x) \in I_r w I_r$$

for some $w \in \mathrm{Perm}^{\mathbf{G}}(\mu)$. The equality $\mathrm{Adm}^{\mathbf{G}}(\mu) = \mathrm{Perm}^{\mathbf{G}}(\mu)$ holds (this translates under Morita equivalence to the analogous statement for $\mathrm{GL}_n$), and therefore we have proved that the determinant condition can now be interpreted as:

$$x^{-1}\delta\sigma(x)\widetilde{\Lambda}_{\bullet, W(k_r)} \in \mathbf{M}_\mu(k_r),$$

where by definition $\mathbf{M}_\mu(k_r)$ is the set of type $(\widetilde{\Lambda}_\bullet)$ multichains of $\mathcal{O}_B \otimes W(k_r)$-lattices in $V \otimes L_r$, self-dual up to a scalar in $\mathbb{Q}^\times$, of form

$$g\widetilde{\Lambda}_{\bullet, W(k_r)}$$

for some $g \in \mathbf{G}(L_r)$ such that $I_r g I_r = I_r w I_r$ for an element $w \in \mathrm{Perm}^{\mathbf{G}}(\mu) = \mathrm{Adm}^{\mathbf{G}}(\mu)$.

Now let $I$ denote the $\mathbb{Q}$-group of self-$\mathbb{Q}$-isogenies of $(A', \lambda', i')$. Our above remarks and the discussion in [**Ko92**], §16 show that there is a bijection from the set of points $(A_\bullet, \lambda, i, \bar{\eta}) \in Sh(k_r)$ such that $(A_0, \lambda, i)$ is $\mathbb{Q}$-isogenous to $(A', \lambda', i')$, to the set $I(\mathbb{Q})\backslash(Y^p \times Y_p)$, where

$$Y^p = \{y \in \mathbf{G}(\mathbb{A}_f^p)/K^p \mid y^{-1}\gamma y \in K^p\},$$
$$Y_p = \{x \in \mathbf{G}(L_r)/I_r \mid I_r x^{-1}\delta\sigma(x)\widetilde{\Lambda}_{\bullet, W(k_r)} \subset \mathbf{M}_\mu(k_r)\}.$$

11.1.2. *Compatibility of "de Rham" and "crystalline" maps.*

Lemma 11.1. *Fix $w \in \mathrm{Adm}(\mu)$. Let $A = (A_\bullet, \lambda, i, \bar{\eta}) \in Sh(k_r)$. Suppose that $A$ is $\mathbb{Q}$-isogenous to $(A', \lambda', i')$ and that via a choice of $\xi : A_0 \to A'$ we have $\xi H(A_\bullet) = x\widetilde{\Lambda}_{\bullet, W(k_r)}$ as above, for $x \in \mathbf{G}(L_r)/I_r$* [13].

[13] Note that $x \in \mathbf{G}(L_r)/I_r$ depends on $\xi$, but its image in $I(\mathbb{Q})\backslash\mathbf{G}(L_r)/I_r$ is independent of $\xi$, and hence the double coset $I_r x^{-1}\delta\sigma(x) I_r$ is well-defined.

*Then $A$ belongs to the KR-stratum $Sh_w$ if and only if*

$$I_r x^{-1}\delta\sigma(x)I_r = I_r w I_r.$$

Proof. Recall that the local model $\mathbf{M}^{\mathrm{loc}}$ is naturally identified with the model $M_{-w_0\mu}$ and as such its special fiber carries an action of the standard Iwahori-subgroup of $\mathrm{GL_n}(k_r[\![t]\!])$; further the KR-stratum $Sh_w$ is the set of points which give rise to a point in the Iwahori-orbit indexed by $w^{-1}$ under the "de Rham" map $\psi : \widetilde{Sh} \to \mathbf{M}^{\mathrm{loc}}$. Let us recall the definition of $\psi$. We choose any isomorphism

$$\gamma_\bullet : M(A_\bullet) \xrightarrow{\sim} \widetilde{\Lambda}_{\bullet,k_r}$$

of polarized $\mathcal{O}_B \otimes k_r$-multichains. The quotient $M(A_\bullet) \to \mathrm{Lie}(A_\bullet)$ then determines via Morita equivalence a quotient

$$\frac{\mathcal{V}_{\bullet,k_r}}{t\mathcal{V}_{\bullet,k_r}} \to \frac{\mathcal{V}_{\bullet,k_r}}{\mathcal{L}_\bullet}$$

for a uniquely determined $k_r[\![t]\!]$-lattice chain $\mathcal{L}_\bullet$ satisfying $t\mathcal{V}_{\bullet,k_r} \subset \mathcal{L}_\bullet \subset \mathcal{V}_{\bullet,k_r}$ (see §8.1). The "de Rham" map $\psi$ sends the point $(A, \gamma_\bullet) \in \widetilde{Sh}(k_r)$ to $\mathcal{L}_\bullet$. By definition $A \in Sh_w$ if and only if

$$\mathrm{inv}_I(\mathcal{L}_\bullet, \mathcal{V}_{\bullet,k_r}) = w,$$

where the invariant measures the relative position of complete $k_r[\![t]\!]$-lattice chains in $k_r(\!(t)\!)^n$. (Note that $w$ is independent of the choice of $\gamma_\bullet$.)

In the previous paragraph the ambient group is $\mathrm{GL}_n(k_r(\!(t)\!))$, the function-field analogue of $\mathrm{GL}_n(L_r)$. But it is more natural to work directly with $\mathbf{G}(L_r)$. Given $\mathcal{L}_\bullet$ as above, there exists a unique type $(\widetilde{\Lambda}_\bullet)$ polarized multichain of $\mathcal{O}_B \otimes W(k_r)$-lattices $\widetilde{\mathcal{L}}_\bullet$ in $V \otimes_{\mathbb{Q}_p} L_r$ such that $p\widetilde{\Lambda}_{\bullet,W(k_r)} \subset \widetilde{\mathcal{L}}_\bullet \subset \widetilde{\Lambda}_{\bullet,W(k_r)}$ and the polarized $\mathcal{O}_B \otimes k_r$-multichain $\widetilde{\Lambda}_{\bullet,W(k_r)}/\widetilde{\mathcal{L}}_\bullet$ is Morita equivalent to the $k_r$-lattice chain $\mathcal{V}_{\bullet,k_r}/\mathcal{L}_\bullet$. Thus we can think of $\psi$ as the map

$$(A, \gamma_\bullet) \mapsto \widetilde{\mathcal{L}}_\bullet.$$

Thus, $A \in Sh_w$ if and only if

$$\mathrm{inv}_{I_r}(\widetilde{\mathcal{L}}_\bullet\, ,\, \widetilde{\Lambda}_{\bullet,W(k_r)}) = w.$$

We have an isomorphism of polarized multichains of $\mathcal{O}_B \otimes k_r$-lattices

$$\frac{\widetilde{\Lambda}_{\bullet,W(k_r)}}{\widetilde{\mathcal{L}}_\bullet} = \mathrm{Lie}(A_\bullet) = \sigma^{-1}\left(\frac{x^{-1}\delta\sigma(x)\widetilde{\Lambda}_{\bullet,w(k_r)}}{\widetilde{\Lambda}_{\bullet,W(k_r)}}\right).$$

The second equality comes from Cor. 14.4; the map sending $(A_\bullet, \xi)$ to the multichain $x\widetilde{\Lambda}_{\bullet,W(k_r)}$ may be termed "crystalline" – it is defined using crystalline homology. This equality shows that the "de Rham" and "crystalline" maps are compatible (and ultimately rests on Theorem 14.2 of Oda).

Putting these remarks together, we see that $A \in Sh_w$ if and only if

$$\mathrm{inv}_{I_r}(\widetilde{\Lambda}_{\bullet,W(k_r)}\, ,\ x^{-1}\delta\sigma(x)\widetilde{\Lambda}_{\bullet,W(k_r)}) = w,$$

which completes the proof.

□

Remark 11.2. The crux of the above proof is the aforementioned compatibility between the "de Rham" and "crystalline" maps. This compatibility can be rephrased as the commutativity of the diagram at the end of §7 in [**R2**], when the morphisms there are suitably interpreted.

11.1.3. *Identifying* $\phi_r$. To find $\phi_r$ we need to count the points in the set $I(\mathbb{Q})\backslash(Y^p \times Y_p)$ with the correct "multiplicity". The test function $\phi_r$ in the twisted orbital integral must be such that

$$\sum_A \mathrm{Tr}^{ss}(\Phi_{\mathfrak{p}}^r; R\Psi_A^{Sh}(\overline{\mathbb{Q}}_\ell)) = \int_{I(\mathbb{Q})\backslash(\mathbf{G}(\mathbb{A}_f^p)\times\mathbf{G}(L_r))} f^p(y^{-1}\gamma y)\,\phi_r(x^{-1}\delta\sigma(x)), \tag{11.1.4}$$

where $A$ ranges over points $(A_\bullet, \lambda, i, \bar{\eta}) \in Sh(k_r)$ such that $(A_0, \lambda, i)$ is $\mathbb{Q}$-isogenous to $(A', \lambda', i')$ (other notation and measures as in [**Ko92**], §16, with the exception that here the Haar measure on $G(L_r)$ gives $I_r$ volume 1).

Now by Lemma 11.1, equation (11.1.4) will hold if $\phi_r$ is a function in the Iwahori-Hecke algebra of $\overline{\mathbb{Q}}_\ell$-valued functions

$$\mathcal{H}_{I_r} = C_c(I_r\backslash\mathbf{G}(L_r)/I_r)$$

such that

$$\phi_r(I_r w I_r) = \mathrm{Tr}^{ss}(\Phi_p^r, R\Psi_{w^{-1}}^{M_{-w_0\mu}}(\overline{\mathbb{Q}}_\ell)),$$

for elements $w \in \mathrm{Adm}(\mu)$, and zero elsewhere. Note that on the right hand side, $w^{-1}$ really represents an Iwahori-orbit in the affine flag variety $\mathcal{F}l_{\mathbb{F}_p}$ over the *function field*. The nearby cycles are equivariant for the Iwahori-action in a suitable sense, so that the semi-simple trace function is constant on these orbits. Hence the right hand side is a well-defined element of $\overline{\mathbb{Q}}_\ell$.

We can simply *define* the function $\phi_r$ by this equality. But such a description of $\phi_r$ will not be useful unless we can identify it with an explicit function in the Iwahori-Hecke algebra (we need to know its traces on representations with Iwahori-fixed vectors, at least, if we want to make the spectral side of (11.0.2) explicit). This however is possible, due to the following theorem. This result was conjectured by Kottwitz in a more general form, which inspired Beilinson to conjecture that nearby cycles can be used to give a geometric construction of the center of the affine Hecke algebra for *any* reductive group in the function-field setting. This latter conjecture was proved by Gaitsgory [**Ga**], whose ideas were adapted to prove the p-adic analogue in [**HN1**].

THEOREM 11.3 (The Kottwitz Conjecture;[**Ga**],[**HN1**]). *Let* $G = \mathrm{GL}_n$ *or* $\mathrm{GSp}_{2n}$. *Let $\lambda$ be a minuscule dominant coweight of $G$, with corresponding $\mathbb{Z}_p$-model $M_\lambda$ (cf. Remark 4.1). Let $\mathcal{H}_{k_r((t))}$ denote the Iwahori-Hecke algebra $C_c(I_r\backslash G(k_r((t)))/I_r)$. Let $z_{\lambda,r} \in Z(\mathcal{H}_{k_r((t))})$ denote the Bernstein function in the center of the Iwahori-Hecke algebra which is associated to $\lambda$. Then*

$$\mathrm{Tr}(\Phi_p^r, R\Psi^{M_\lambda}(\overline{\mathbb{Q}}_\ell)) = p^{r\dim(M_\lambda)/2}\, z_{\lambda,r}.$$

Let $K_r$ be the stabilizer in $\mathbf{G}(L_r)$ of $\widetilde{\Lambda}_{0,W(k_r)}$; this is a hyperspecial maximal compact subgroup of $\mathbf{G}(L_r)$ containing $I_r$. The Bernstein function $z_{\lambda,r}$ is characterized as the unique element in the center of $\mathcal{H}_{I_r}$ such that the image of $p^{r\langle\rho,\lambda\rangle}z_{\lambda,r}$ under the Bernstein isomorphism

$$- * \mathbb{I}_{K_r} : Z(\mathcal{H}_{I_r}) \xrightarrow{\sim} \mathcal{H}_{K_r} := C_c(K_r\backslash\mathbf{G}(L_r)/K_r)$$

is the spherical function $f_{\lambda,r} := \mathbb{I}_{K_r\lambda K_r}$. Here, we have used the fact that $\lambda$ is minuscule. Recall that $\rho$ is the half-sum of the $B$-positive roots of $G$, and that $\dim(M_\lambda)/2 = \langle\rho,\lambda\rangle$.

The above theorem for $\lambda = -w_0\mu$ implies that

$$\phi_r(w) = p^{r\dim(Sh)/2}\, z_{-w_0\mu}(w^{-1}).$$

Now invoking the identity $z_\mu(w) = z_{-w_0\mu}(w^{-1})$ (see [**HKP**], §3.2), we have proved the following result.

PROPOSITION 11.4. *Let $z_{\mu,r}$ denote the Bernstein function in the center of the Iwahori-Hecke algebra $\mathcal{H}_{L_r} = C_c(I_r\backslash\mathbf{G}(L_r)/I_r)$ corresponding to $\mu$. Then the test function is given by*

$$\phi_r = p^{r\dim(Sh)/2}\, z_{\mu,r}.$$

*Remarks:* 1) Because $\phi_r$ is central we can define its "base-change" function $b(\phi_r) =: f_p^{(r)}$, an element in the center of the Iwahori-Hecke algebra for $\mathbf{G}(\mathbb{Q}_p)$. We define the base-change homomorphism for centers of Iwahori-Hecke algebras as the unique homomorphism $b : Z(\mathcal{H}_{I_r}) \to Z(\mathcal{H}_I)$ which induces, via the Bernstein isomorphism, the usual base-change homomorphism for spherical Hecke algebras $b : \mathcal{H}_{K_r} \to \mathcal{H}_K$ for any special maximal compact $K_r$ containing $I_r$ (setting $I = I_r \cap \mathbf{G}(\mathbb{Q}_p)$ and $K = K_r \cap \mathbf{G}(\mathbb{Q}_p)$). This gives a well-defined homomorphism, independent of the choice of $K_r$. Moreover, the pair of functions $\phi_r$, $f_p^{(r)}$ have matching (twisted) orbital integrals; see [**HN3**].
2) We know how $z_{\mu,r}$ (and hence how its base-change $b(z_{\mu,r})$) acts on unramified principal series. This plays a key role in Theorem 11.7 below. In fact, we have the following helpful lemma.

LEMMA 11.5. *Let $I$ denote our Iwahori subgroup $K_p^{\mathbf{a}} \subset G(\mathbb{Q}_p)$, whose reduction modulo $p$ is $B(\mathbb{F}_p)$. Suppose $\pi_p$ is an irreducible admissible representation of $G(\mathbb{Q}_p)$ with $\pi_p^I \neq 0$. Let $r_\mu$ be the irreducible representation of ${}^L\mathbf{G}_{\mathbb{Q}_p}$ having extreme weight $\mu$. Let $d = \dim(Sh_E)$. Then*

$$\mathrm{Tr}\,\pi_p\,(p^{rd/2}\,b(z_{\mu,r})) = \dim(\pi_p^I)\; p^{rd/2}\,\mathrm{Tr}\big(r_\mu\varphi_{\pi_p}\big(\Phi^r \times \begin{bmatrix} p^{-r/2} & 0 \\ 0 & p^{r/2}\end{bmatrix}\big)\big).$$

PROOF. Write $G = G(\mathbb{Q}_p)$ and $B = B(\mathbb{Q}_p)$ and suppose that $\pi_p$ is an irreducible subquotient of the normalized unramified principal series representation $i_B^G(\chi)$, for an unramified quasi-character $\chi : T(\mathbb{Q}_p) \to \mathbb{C}^\times$. Suppose $K_r \supset I_r$ (thus also $K \supset I$) is a hyperspecial maximal compact, and suppose that $\pi_\chi$ is the unique $K$-spherical subquotient of $i_B^G(\chi)$. Suppose $\varphi : W_{\mathbb{Q}_p} \to {}^LG$ is the unramified parameter associated to $\pi_\chi$. Our normalization of the correspondence $\pi_p \mapsto \varphi_{\pi_p}$ is such that for all $r \geq 1$, the element $\varphi(\Phi^r) \in \widehat{G} \rtimes W_{\mathbb{Q}_p}$ can be described as

$$\varphi(\Phi^r) = \varphi_{\pi_p}\big(\Phi^r \times \begin{bmatrix} p^{-r/2} & 0 \\ 0 & p^{r/2}\end{bmatrix}\big) = (\chi \rtimes \Phi)^r \tag{11.1.5}$$

where $\Phi$ is a *geometric* Frobenius element in $W_{\mathbb{Q}_p}$, and where we identify $\chi$ with an element in the dual torus $\widehat{T}(\mathbb{C}) \subset \widehat{G}(\mathbb{C})$ and take the product on the right in the group ${}^LG$. Note that our normalization of the local Langlands correspondence is the one compatible with Deligne's normalization of the reciprocity map in local class field theory, where a uniformizer is sent to a geometric Frobenius element (see [**W2**]).

Let $f_{\mu,r} = \mathbb{I}_{K_r\mu K_r}$ in the spherical Hecke algebra $\mathcal{H}_{K_r}$; it is the image of $p^{rd/2}z_{\mu,r}$ under the Bernstein isomorphism

$$- * \mathbb{I}_{K_r} \;:\; Z(\mathcal{H}_{I_r}) \xrightarrow{\sim} \mathcal{H}_{K_r}.$$

Further, $b(p^{rd/2}z_{\mu,r}) * \mathbb{I}_K = b(f_{\mu,r})$.

Now by [**Ko84**], Thm (2.1.3), we know that under the usual left action of $\mathcal{H}_K$, $b(f_{\mu,r})$ acts on $\pi_\chi^K$ by the scalar

$$p^{rd/2}\,\mathrm{Tr}(r_\mu\varphi(\Phi^r)).$$

Taking (11.1.5) into account along with the well-known fact that elements in $Z(\mathcal{H}_I)$ act on the entire space $i_B^G(\chi)^I$ by a scalar (see e.g. [**HKP**]), we are done. $\square$

REMARK 11.6. For application in Theorem 11.7, we need the "$\ell$-adic" analogue of this lemma, i.e., we need to work with the dual group $\widehat{G}(\overline{\mathbb{Q}}_\ell)$ instead of $\widehat{G}(\mathbb{C})$. For a discussion of how to do this, see [**Ko92b**], §1.

**11.2. The semi-simple local zeta function in terms of semi-simple $L$-functions.** The foregoing discussion culminates in the following result from [**HN3**], to which we refer for details of the proof.

THEOREM 11.7 ([**HN3**]). *Suppose $Sh$ is a simple ("fake" unitary) Shimura variety. Suppose $K_p$ is an Iwahori subgroup. Suppose $r_\mu$ is the irreducible representation of ${}^L(\mathbf{G}_{\mathbf{E}_\mathfrak{p}})$ with extreme weight $\mu$, where $\mu$ is the minuscule coweight determined by the Shimura data. Then we have*

$$Z_\mathfrak{p}^{ss}(s, Sh) = \prod_{\pi_f} L^{ss}(s-\frac{d}{2}, \pi_p, r_\mu)^{a(\pi_f)\dim(\pi_f^{\mathbf{K}})}$$

*where the product runs over all admissible representations $\pi_f$ of $\mathbf{G}(\mathbb{A}_f)$, and the integer number $a(\pi_f)$ is given by*

$$a(\pi_f) = \sum_{\pi_\infty\in\Pi_\infty} m(\pi_f\otimes\pi_\infty)\mathrm{Tr}\,\pi_\infty(f_\infty),$$

*where $m(\pi_f\otimes\pi_\infty)$ is the multiplicity of $\pi_f\otimes\pi_\infty$ in $L^2(\mathbf{G}(\mathbb{Q})A_\mathbf{G}(\mathbb{R})^0\backslash\mathbf{G}(\mathbb{A}))$. Here $\Pi_\infty$ is the set of admissible representations of $G(\mathbb{R})$ whose central and infinitesimal characters are trivial. Also, $d$ denotes the dimension of $Sh_E$.*

Let us remark that it is our firm belief that this result continues to hold when $K_p$ is a general parahoric subgroup of $\mathbf{G}(\mathbb{Q}_p)$, but some details are more difficult than the Iwahori case treated in [**HN3**], and remain to be worked out.

Finally, recall our assumptions on $\mathfrak{p}$ implied that $\mathbf{E}_\mathfrak{p} = \mathbb{Q}_p$ and $N\mathfrak{p} = (p)$. In more general circumstances one or both of these statements will fail to hold, and the result will be a slightly more complicated expression for $Z_\mathfrak{p}^{ss}(s, Sh)$.

## 12. The Newton stratification on Shimura varieties over finite fields

**12.1. Review of the Kottwitz and Newton maps.** As usual $L$ denotes the fraction field of the Witt vectors $W(\overline{\mathbb{F}}_p)$, with Frobenius automorphism $\sigma$. Let $G$ denote a connected reductive group over $\mathbb{Q}_p$. Then we have the pointed set $B(G) = B(G)_{\mathbb{Q}_p}$ consisting of $\sigma$-conjugacy classes in $G(L)$. Let us also assume (only for simplicity) that the connected reductive $\mathbb{Q}_p$-group $G$ is *unramified*. We have the Kottwitz map

$$\kappa_G : B(G) \to X^*(Z(\widehat{G})^{\Gamma_p}),$$

where $\Gamma_p = \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ (see [**Ko97**], §7) . We also have the Newton map

$$\overline{\nu} : B(G) \to \mathfrak{A}^+$$

where the notation is as follows. We choose a $\mathbb{Q}_p$-rational Borel subgroup $B \subset G$ and a maximal $\mathbb{Q}_p$-torus $T$ which is contained in $B$. Then $\mathfrak{U} = X_*(T)_{\mathbb{R}}^{\Gamma_p}$, and $\mathfrak{U}^+$ denotes the intersection of $\mathfrak{U}$ with the cone of $B$-dominant elements in $X_*(T)_{\mathbb{R}}$. We call $b \in B(G)$ *basic* if $\overline{\nu}_b$ is central, i.e., $\overline{\nu}_b \in X_*(Z)_{\mathbb{R}}$.

Suppose $\{\lambda\}$ is a conjugacy class of one-parameter subgroups of $G$, defined over $\mathbb{Q}_p$. We may represent the class by a unique cocharacter $\overline{\lambda} \in X_*(T) = X^*(\widehat{T})$ lying in the $B$-positive Weyl chamber of $X_*(T)_{\mathbb{R}}$. The Weyl-orbit of $\overline{\lambda}$ is stabilized by $\Gamma_p$. The notion of $B$-dominant being preserved by $\Gamma_p$ (since $B$ is $\mathbb{Q}_p$-rational), we see that $\overline{\lambda}$ is fixed by $\Gamma_p$, hence it belongs to $\mathfrak{U}^+$. Also, restricting $\{\lambda\}$ to $Z(\widehat{G})$ determines a well-defined element $\lambda^\natural \in X^*(Z(\widehat{G})^{\Gamma_p})$.

We can now define the subset $B(G, \lambda) \subset B(G)$ to be the set of classes $[b] \in B(G)$ such that

$$\kappa_G(b) = \lambda^\natural$$
$$\overline{\nu}_b \preceq \overline{\lambda}.$$

Here $\preceq$ denotes the usual partial order on $\mathfrak{U}^+$ for which $\nu \preceq \nu'$ if $\nu' - \nu$ is a nonnegative linear combination of simple relative coroots.

**12.2. Definition of the Newton stratification.** Suppose that the Shimura variety $Sh_{K_p} = Sh(\mathbf{G}, h)_{K^pK_p}$ is given by a moduli problem of abelian varieties, and that $K_p \subset \mathbf{G}(\mathbb{Q}_p)$ is a parahoric subgroup. Also, assume for simplicity that $E := \mathbf{E}_{\mathfrak{p}} = \mathbb{Q}_p$. Let $G = \mathbf{G}_{\mathbb{Q}_p}$, which again for simplicity we assume is unramified. Let $k$ denote as usual an algebraic closure of the residue field $\mathcal{O}_E/\mathfrak{p} = \mathbb{F}_p$, and let $k_r$ denote the unique subfield of $k$ having cardinality $p^r$. Let $L_r$ denote the fraction field of the Witt vectors $W(k_r)$.

We denote by $\mathbf{a}$ the base alcove and suppose $0 \in \overline{\mathbf{a}}$ is a hyperspecial vertex. Let $K_p^{\mathbf{a}}$ (resp. $K_p^0$) denote the corresponding Iwahori (resp. hyperspecial maximal compact) subgroup of $G(\mathbb{Q}_p)$. We will let $K_p$ denote a "standard" parahoric subgroup, i.e., one such that $K_p^{\mathbf{a}} \subset K_p \subset K_p^0$.

We can define a map

$$\delta_{K_p} : Sh_{K_p}(k) \to B(G)$$

as follows. A point $A_\bullet = (A_\bullet, \lambda, i, \overline{\eta}) \in Sh_{K_p}(k_r)$ gives rise to a $c$-polarized virtual $B$-abelian variety over $k_r$ up to prime-to-$p$ isogeny (cf. [**Ko92**] §10), which we denote by $(A, \lambda, i)$. That in turn determines an $L$-isocystal $(H(A)_L, \Phi)$ as in §14, cf. loc. cit. It is not hard to see that $\delta_{K_p}$ takes values in the subset $B(G, \mu) \subset B(G)$. In fact, if we choose an isomorphism $(H(A)_L, \Phi) = (V \otimes L, \delta\sigma)$ of isocrystals for the group $G(L)$, then $\delta \in G(L)$ satisfies

$$\kappa_G(\delta) = \mu^\natural,$$

as follows from the determinant condition $\sigma(\mathrm{Lie}(A_0)) = \Phi H(A_0)/H(A_0)$, for any $A_0$ in the chain $A_\bullet$ (use the argument of [**Ko92**], p. 431). Furthermore, the Mazur inequality

$$\overline{\nu}_\delta \preceq \overline{\mu}$$

can be proved by reducing to the case where $K_p = K_p^0$ (which has been treated by Rapoport-Richartz [**RR**] – see also [**Ko03**]) [14].

[14]Note that our assumption that $K_p$ be *standard*, i.e., $K_p \subset K_p^0$, was used in the last step, because we want to invoke [**RR**]. The results of loc. cit. probably hold for nonspecial maximal compact subgroups, so this assumption is probably unnecessary.

DEFINITION 12.1. We call the fibers of $\delta_{K_p}$ the *Newton strata* of $Sh_{K_p}$. The inverse image of the basic set

$$\delta_{K_p}^{-1}\big(B(G,\mu) \cap B(G)_{basic}\big).$$

is called the *basic locus* of $Sh_{K_p}$.

We denote the Newton stratum $\delta_{K_p}^{-1}([b])$ by $\mathcal{S}_{K_p,[b]}$, or if $K_p$ is understood, simply by $\mathcal{S}_{[b]}$.

The following conjecture is fundamental to the subject. It asserts that all the Newton strata are nonempty.

CONJECTURE 12.2 (Rapoport, Conj. 7.1 [**R2**]). *The map*

$$\delta_{K_p} : Sh_{K_p}(k) \to B(G,\mu)$$

*is surjective. In particular, the basic locus is nonempty.*

REMARK 12.3. Note that $\mathrm{Im}(\delta_{K_p^0})$ can be interpreted purely in terms of group theory: $[b] \in B(G,\mu)$ lies in the image of $\delta_{K_p^0}$ if and only if for one (equivalently, for all sufficiently divisible) $r \geq 1$, $[b]$ contains an element $\delta \in G(L_r)$ belonging to a triple $(\gamma_0;\gamma,\delta)$, which satisfies the conditions in [**Ko90**] §2 (except for the following correction, noted at the end of [**Ko92**]: under the canonical map $B(G)_{\mathbb{Q}_p} \to X^*(Z(\widehat{G})^{\Gamma_p})$, the $\sigma$-conjugacy class of $\delta$ goes to $\mu|_{X^*(Z(\widehat{G})^{\Gamma_p})}$ and not its negative), and for which the following four conditions also hold:

(a) $\gamma_0\gamma_0^* = c_0^{-1}p^{-r}$, where $c_0 \in \mathbb{Q}^\times$ is a $p$-adic unit;
(b) the Kottwitz invariant $\alpha(\gamma_0;\gamma,\delta)$ is trivial;
(c) there exists a lattice $\Lambda$ in $V_{L_r}$ such that $\delta\sigma\Lambda \supset \Lambda$;
(d) $O_\gamma(\mathbb{I}_{K^p})\, TO_{\delta\sigma}(\mathbb{I}_{K_r\mu K_r}) \neq 0$.

Here $K_r \subset \mathbf{G}(L_r)$ is the hyperspecial maximal compact subgroup such that $K_r \cap \mathbf{G}(\mathbb{Q}_p) = K_p^0$.

To see this, use [**Ko92**], Lemmas 15.1, 18.1 to show that the first three conditions ensure the existence of a $c_0p^r$-polarized virtual $B$-abelian variety $(A',\lambda',i')$ over $k_r$ up to prime-to-$p$ isogeny, giving rise to $(\gamma_0;\gamma,\delta)$. In the presence of the first three, the last condition shows that there exists a $k_r$-rational point $(A,\lambda,i,\bar\eta) \in Sh_{K_p^0}$ such that $(A,\lambda,i)$ is $\mathbb{Q}$-isogenous to $(A',\lambda',i')$. Hence $\delta_{K_p^0}((A,\lambda,i,\bar\eta)) = [\delta]$.

Note that by counting fixed points of $\Phi_{\mathfrak{p}}^r \circ f$ for any Hecke operator $f$ away from $p$ as in [**Ko92**], §16, we may replace condition (d) with

(d') For some $g \in \mathbf{G}(\mathbb{A}_f^p)$, we have $O_\gamma(\mathbb{I}_{K^pg^{-1}K^p})\, TO_{\delta\sigma}(\mathbb{I}_{K_r\mu K_r}) \neq 0$.

Since we may always choose $g = \gamma^{-1}$, we may also replace (d) or (d') with

(d") $TO_{\delta\sigma}(\mathbb{I}_{K_r\mu K_r}) \neq 0$.

In the Siegel case with $K_p$ hyperspecial, we have the following result of Oort [**Oo**] (comp. [**R2**], Thm. 7.4) which in particular proves Conjecture 12.2 in that case.

THEOREM 12.4 (Oort, [**Oo**]). *Suppose $Sh_{K_p^0}$ is a Siegel modular variety with maximal hyperspecial level structure $K_p^0$ at $p$. Then the Newton strata are all nonempty and equidimensional (and the dimension is given by a simple formula in terms of the partially ordered set $B(\mathrm{GSp}_{2n},\mu)$).*

In Corollary 12.8 below, we shall generalize the part of Theorem 12.4 that asserts the nonemptiness of the Newton strata: in the Siegel case, all Newton strata of $Sh_{K_p}$ are nonempty, when $K_p$ is an arbitrary standard parahoric subgroup. In the "fake" unitary case, we shall later prove in Cor. 12.12 only that the basic locus is nonempty (again for standard parahorics $K_p$).

**12.3. The relation between Newton strata, KR-strata, and affine Deligne-Lusztig varieties.** Fix a $\sigma$-conjugacy class $[b] \in B(G)$, and fix an element $w \in \widetilde{W}$. Let $\widetilde{K}_p^{\mathbf{a}} \subset G(L)$ denote the Iwahori subgroup such that $\widetilde{K}_p^{\mathbf{a}} \cap G(\mathbb{Q}_p) = K_p^{\mathbf{a}}$.

DEFINITION 12.5. We define the *affine Deligne-Lusztig variety* [15]

$$X_w(b)_{\widetilde{K}_p^{\mathbf{a}}} = \{x \in G(L)/\widetilde{K}_p^{\mathbf{a}} \mid x^{-1}b\sigma(x) \in \widetilde{K}_p^{\mathbf{a}}\, w\, \widetilde{K}_p^{\mathbf{a}}\},$$

and for any dominant coweight $\lambda$,

$$X(\lambda, b)_{\widetilde{K}_p^{\mathbf{a}}} = \bigcup_{w \in \mathrm{Adm}(\lambda)} X_w(b)_{\widetilde{K}_p^{\mathbf{a}}}.$$

A similar definition can be made for a parahoric subgroup replacing $K_p^{\mathbf{a}}$ (cf. [**R2**]).

A fundamental problem is to determine the relations between the Kottwitz-Rapoport and Newton stratifications. The following shows how this problem is related to the nonemptiness of certain affine Deligne-Lusztig varieties.

PROPOSITION 12.6. *Let $\mu$ be the minuscule coweight attached to the Shimura data for $Sh = Sh_{K_p^{\mathbf{a}}}$. Suppose $w \in \mathrm{Adm}(\mu)$. Then for every $[b] \in \mathrm{Im}(\delta_{K_p^0})$, we have*

$$X_w(b)_{\widetilde{K}_p^{\mathbf{a}}} \neq \emptyset \Leftrightarrow Sh_w \cap \mathcal{S}_{[b]} \neq \emptyset.$$

PROOF. By Remark 12.3, for all sufficiently divisible $r \geq 1$, $[b]$ contains an element $\delta \in G(L_r)$ which is part of a Kottwitz triple $(\gamma_0; \gamma, \delta)$ satisfying the conditions (a-d). We consider such a triple, up to equivalence (we say $(\gamma_0'; \gamma', \delta')$ is equivalent to $(\gamma_0; \gamma, \delta)$ if $\gamma_0'$ and $\gamma_0$ are stably-conjugate, $\gamma'$ and $\gamma$ are conjugate, and $\delta'$ and $\delta$ are $\sigma$-conjugate). Then by the arguments in §11 together with [**Ko92**], §18, 19, we have the equality

$$\#\{A_\bullet \in Sh_w(k_r) \mid A_\bullet \rightsquigarrow (\gamma_0; \gamma, \delta)\} = (\mathrm{vol})\, O_\gamma(\mathbb{I}_{K^p})\ \#X_w(\delta)_{I_r}. \tag{12.3.1}$$

Let us explain the notation. The notation $A_\bullet \rightsquigarrow (\gamma_0; \gamma, \delta)$ means that $A_\bullet$ gives rise to the equivalence class of $(\gamma_0; \gamma, \delta)$; cf. [**Ko92**], §18, 19. The term vol denotes the nonzero rational number $|\ker^1(\mathbb{Q}, \mathbf{G})| c(\gamma_0; \gamma, \delta)$, where the second term is the number defined in loc. cit. Also $I_r = \widetilde{K}_p^{\mathbf{a}} \cap G(L_r)$.

This equality would imply the proposition, if we knew that $O_\gamma(\mathbb{I}_{K^p}) \neq 0$. But this follows from condition (d) in Remark 12.3. □

The following result of Wintenberger [**Wi**] proves a conjecture of Kottwitz and Rapoport in a suitably unramified case (cf. [**R2**], Conj. 5.2, and the notes at the end).

[15] Strictly speaking, this is only a set, not a variety. The sets are the affine analogues of the usual Deligne-Lusztig varieties in the theory of finite groups of Lie type.

THEOREM 12.7 (Wintenberger). *Let $G$ be any connected reductive group, defined and quasi-split over $L$. Suppose $\{\lambda\}$ is a conjugacy class of 1-parameter subgroups, defined over $L$. Suppose $[b] \in B(G)$ and let $K$ be any standard parahoric subgroup (that is, one contained in a special maximal parahoric subgroup). Then*

$$X(\lambda, b)_K \neq \emptyset \Leftrightarrow [b] \in B(G, \lambda).$$

COROLLARY 12.8. *Let $Sh$ be either a "fake" unitary or a Siegel modular variety, as in* §5.

(a) *In the "fake" unitary case, for any two standard parahoric sugroups $K_p' \subset K_p''$, we have* $\mathrm{Im}(\delta_{K_p'}) = \mathrm{Im}(\delta_{K_p''})$.

(b) *In the Siegel case, we have* $\mathrm{Im}(\delta_{K_p}) = B(G, \mu)$ *for every standard parahoric subgroup $K_p$.*

PROOF. Consider first (a). We need to prove $\mathrm{Im}(\delta_{K_p'}) \supset \mathrm{Im}(\delta_{K_p''})$. Clearly it is enough to consider the case $K_p' = K_p^{\mathbf{a}}$ and $K_p'' = K_p^0$. The natural morphism $Sh_{K_p^{\mathbf{a}}} \to Sh_{K_p^0}$ is proper, surjective on generic fibers (look at $\mathbb{C}$-points), and the target is flat (even smooth). Therefore in the special fiber the morphism is surjective. This completes the proof of (a).

Consider now part (b). In the Siegel case the morphism $Sh_{K_p^{\mathbf{a}}} \to Sh_{K_p^0}$ is still projective with flat image, so the same argument combined with Theorem 12.4 yields the stronger result of (b). Here is another argument using Theorem 12.4, Proposition 12.6 and Theorem 12.7. Let $G = \mathrm{GSp}_{2n}$ and $\mu = (0^n, (-1)^n)$. By Oort's theorem, $\mathrm{Im}(\delta_{K_p^0}) = B(G, \mu)$, so it is enough to prove $B(G, \mu) \subset \mathrm{Im}(\delta_{K_p^{\mathbf{a}}})$. Let $[b] \in B(G, \mu)$. By Wintenberger's theorem, there exists $w \in \mathrm{Adm}(\mu)$ such that $X_w(b)_{\widetilde{K}_p^{\mathbf{a}}} \neq \emptyset$, which implies the result by Proposition 12.6.

Note that a similar argument provides an alternative proof for part (a). □

REMARK 12.9. Let $G = \mathrm{GSp}_{2n}$, and suppose $\mu$ is minuscule. We can give a proof of Theorem 12.7 in this case using Oort's theorem, as follows. Let $[b] \in B(G, \mu)$. By Corollary 12.8, there is a point $A = (A_\bullet, \lambda, i, \bar{\eta})$ such that $\delta_{\widetilde{K}_p^{\mathbf{a}}}(A) = [b]$. Now $A$ belongs to *some* KR-stratum $Sh_w$, so $Sh_w \cap \mathcal{S}_{[b]} \neq \emptyset$. Now Proposition 12.6 implies that $X_w([b])_{\widetilde{K}^{\mathbf{a}}} \neq \emptyset$.

**12.4. The basic locus is nonempty in the "fake" unitary case.** [16]

First consider a "fake" unitary variety $Sh$ with $\mu = (0^{n-d}, (-1)^d)$. We shall consider both hyperspecial and Iwahori level structures.

Recall that the subset $\mathrm{Adm}(\mu) \subset \widetilde{W}(\mathrm{GL}_n)$ contains a unique minimal element. To find this element, we consider first the element

$$\tau_1 = t_{(-1, 0^{n-1})}(12 \cdots n) \in \widetilde{W}(\mathrm{GL}_n),$$

where the cycle $(12 \cdots n)$ acts on a vector $(x_1, x_2 \dots, x_n) \in X_*(T) \otimes \mathbb{R}$ by sending it to $(x_n, x_1, \dots, x_{n-1})$. Note that this element preserves our base alcove

$$\mathbf{a} = \{(x_1, \dots, x_n) \in X_*(T) \otimes \mathbb{R} \mid x_n - 1 < x_1 < \cdots < x_{n-1} < x_n\}.$$

Hence its $d$-th power

$$\tau = t_{((-1)^d, 0^{n-d})}(12 \cdots n)^d \in \widetilde{W}(\mathrm{GL}_n),$$

[16] This non-emptiness is implicit in both articles of [**FM**], and can be justified (indirectly) from their main theorems. Our object here is only to give a simple direct proof.

is the unique element of $\Omega$ which is congruent modulo $W_{\mathrm{aff}}$ to $t_\mu$, so is the desired element. This element is identified with an element in $\mathrm{GL}_n(L)$ using our usual convention (the vector part is sent to $\mathrm{diag}((p^{-1})^d, 1^{n-d})$). Via Morita equivalence, we can view it as an element of $\mathbf{G}(L)$. In fact, via (5.2.1), $\tau$ becomes the element

$$\tau = (X, p^{-1}\chi^{-1}(X^t)^{-1}\chi),$$

where by definition $X = \mathrm{diag}((p^{-1})^d, 1^{n-d})(12\cdots n)^d$.

Next consider the Siegel case, where $\mu = (0^n, (-1)^n)$. The unique element $\tau \in \Omega$ which is congruent to $t_\mu$ modulo $W_{\mathrm{aff}}$ is given by

$$\tau = t_{((-1)^n, 0^n)}(12\cdots 2n)^n \in \widetilde{W}(\mathrm{GSp}_{2n}).$$

We will show that either case, $\tau \in \mathbf{G}(L)$ is *basic*. We will also show that $[\tau]$ belongs to the image of $\delta_{K_p^0}$. By virtue of Corollary 12.8, this will show that the basic locus of $Sh_{K_p}$ is nonempty for every standard parahoric $K_p$.

Let us handle the second statement first.

LEMMA 12.10. *Let $\delta = \tau \in \mathbf{G}(L)$. Then there exists a Kottwitz triple $(\gamma_0; \gamma, \delta)$ satisfying the conditions in Remark 12.3. Hence $[\tau] \in \mathrm{Im}(\delta_{K_p^0})$.*

PROOF. Consider the "fake" unitary case. Note that $\delta \in \mathbf{G}(L_n)$ is clearly fixed by $\sigma$, so that

$$N\delta = \delta^n = \mathrm{diag}((p^{-d})^n) \times \mathrm{diag}((p^{d-n})^n), \tag{12.4.1}$$

where the first factor is the diagonal matrix with the entry $p^{-d}$ repeated $n$ times (similarly for the second factor). This is the image of a unique element $\gamma_0 \in \mathbf{G}(\mathbb{Q})$ under the composition of the inclusion $\mathbf{G}(\mathbb{Q}) \hookrightarrow \mathbf{G}(\mathbb{Q}_p)$ and the isomorphism (5.2.1). In fact $\gamma_0$ belongs to the center of $\mathbf{G}$, and thus is clearly an elliptic element in $\mathbf{G}(\mathbb{R})$. Moreover, $\gamma_0\gamma_0^* = p^{-n}$. For all primes $l \neq p$, we define $\gamma_l$ to the image of $\gamma_0$ under the inclusion $\mathbf{G}(\mathbb{Q}) \hookrightarrow \mathbf{G}(\mathbb{Q}_l)$. We set $\gamma = (\gamma_l)_l \in \mathbf{G}(\mathbb{A}_f^p)$. The resulting triple $(\gamma_0; \gamma, \delta)$ is clearly a Kottwitz triple satisfying the conditions (a-c,d") of Remark 12.3 (with $r = n$). One can check that $\alpha(\gamma_0; \gamma, \delta) = 1$ from the definitions, but in fact this is not necessary, as the group to which $\alpha(\gamma_0; \gamma, \delta)$ belongs is itself trivial in the "fake" unitary case (see [**Ko92b**], Lemma 2). Hence by Remark 12.3, $\delta$ arises from a point in $Sh_{K_p^0}(k_n)$, i.e., $[\tau]$ is in the image of $\delta_{K_p^0}$.

In the Siegel case, the same argument works, if we let $\delta = \tau \in \mathrm{GL}_n(L_2)$ and note

$$N\delta = \delta^2 = \mathrm{diag}((p^{-1})^{2n}). \tag{12.4.2}$$

□

LEMMA 12.11. *The element $\tau \in G(L)$ is basic.*

PROOF. We want to use the following special case of the characterization of $\bar{\nu}_b$, for certain $b \in G(L)$: suppose we are given an element $b \in G(L)$ such that for sufficiently divisible $s \in \mathbb{N}$, we may write in the semidirect product $G(L) \rtimes \langle\sigma\rangle$ the identity

$$(b\sigma)^s = (s\nu)(p)\ \sigma^s$$

for a rational $B$-dominant cocharacter $s\nu : \mathbf{G}_m \to Z(G)$ defined over $\mathbb{Q}_p$. Then in that case, $b$ is basic and

$$\overline{\nu}_b = \frac{1}{s}(s\nu) \in X_*(T)_{\mathbb{Q}}^{\Gamma_p}.$$

This follows immediately from the general characterization of $\bar{\nu}_b$ given in [**Ko85**], §4.3.

This characterization applies to the element $\tau$ because of the identities (12.4.1) and (12.4.2). In fact we see that, in $\mathrm{GL}_n$, the Newton point of $\tau$ is

$$\overline{\nu}_\tau = ((\frac{-d}{n})^n) \in X_*(T(\mathrm{GL}_n))_{\mathbb{Q}}^{\Gamma_p},$$

which clearly factors through the center of $G$. In the Siegel case,

$$\overline{\nu}_\tau = ((\frac{-1}{2})^{2n}) \in X_*(T(\mathrm{GSp}_{2n}))_{\mathbb{Q}}^{\Gamma_p}.$$

Thus $\tau$ is basic in each case. □

We get the following, which of course we already knew in the Siegel case, by Theorem 12.4.

Corollary 12.12. *Let $Sh$ be a "fake" unitary or Siegel modular Shimura variety with level structure given by a standard parahoric subgroup $K_p$. Then the basic locus of $Sh$ is nonempty.*

For (more detailed) information concerning the Newton stratification on some other Shimura varieties, the reader might consult the work of Andreatta-Goren [**AG**] and Goren-Oort [**GOo**].

## 13. The number of irreducible components in $Sh_{\overline{\mathbb{F}}_p}$

Let $Sh$ be a "fake" unitary or a Siegel modular variety with Iwahori-level structure as in §5. Recall that $\mathbf{M}^{\mathrm{loc}}_{\mathbb{F}_p}$ is a connected variety with an Iwahori-orbit stratification indexed by the finite subset $\mathrm{Adm}(\mu) \subset \widetilde{W}$, where $\mu = (0^{n-d}, (-1)^d)$, resp. $(0^n, (-1)^n)$. Its irreducible components are indexed by the maximal elements in this set, namely by the translation elements $t_\lambda$ for $\lambda$ belonging to the Weyl-orbit $W\mu$ of the coweight $\mu$.

It is natural to hope that similar statements apply in some sense to $Sh$. The varieties $Sh_{\mathbb{F}_p}$ and $\widetilde{Sh}_{\mathbb{F}_p}$ are not geometrically connected (the number of connected components of $Sh_{\overline{\mathbb{F}}_p}$ depends on the choice of subgroup $K^p \subset \mathbf{G}(\mathbb{A}_f^p)$; see below). Nevertheless the following two Lemmas (13.1 and 13.2) show that in every connected component of $Sh$, all the KR-strata are nonempty.

Lemma 13.1 ([**Ge**], Prop. 1.3.2, in the Siegel case). *If $Sh$ is either a "fake" unitary or a Siegel modular Shimura variety, and $K_p$ is any standard parahoric subgroup, then the morphism $\psi : \widetilde{Sh} \to \mathbf{M}^{\mathrm{loc}}_{K_p}$ is surjective.*

Proof. First we claim that it suffices to prove the lemma for $K_p^{\mathbf{a}}$, our standard Iwahori subgroup. Indeed, it is enough to observe that the following diagram commutes

$$\begin{tikzcd} \widetilde{Sh}_{K_p^{\mathbf{a}}} \arrow[r, "\psi"] \arrow[d] & \mathbf{M}^{\mathrm{loc}} \arrow[d, "p"] \\ \widetilde{Sh}_{K_p} \arrow[r, "\psi"] & \mathbf{M}^{\mathrm{loc}}_{K_p}, \end{tikzcd}$$

and that the right vertical arrow $p$ is surjective on the level of KR-strata: every $\mathrm{Aut}_{K_p}$-orbit in $\mathbf{M}^{\mathrm{loc}}_{K_p}$ contains an element in the image of $p$; this follows from [**KR**], Prop. 9.3, 10.6. See also [**Go3**].

Next consider the Siegel case with Iwahori level structure, where this result is due to Genestier, loc. cit. We briefly recall his argument. It is easy to see that $\psi$ is surjective on generic fibers, because it is Aut-equivariant, and the generic fiber of $\mathbf{M}^{\mathrm{loc}}$ is a single orbit under Aut. Because $\psi$ is smooth, the complement of its image is an Aut-invariant, Zariski-closed subset of the special fiber of $\mathbf{M}^{\mathrm{loc}}$. On the other hand, there is a unique closed (zero-dimensional) Aut-orbit (denoted here by $\tau^{-1}$) in that fiber which belongs to the Zariski-closure of every other Aut-orbit, and one can show (by writing down an explicit chain of supersingular abelian varieties) that the point $\tau^{-1}$ belongs to the image of $\psi$. It follows that $\psi$ is surjective.

In the "fake" unitary case with Iwahori-level structure, consider the element $\tau$ from §12.4 above. Then the element $\tau^{-1}$ indexes the unique zero-dimensional Aut-orbit in $\mathbf{M}^{\mathrm{loc}} = M_{-w_0\mu}$. By Genestier's argument, it is enough to prove that $\tau^{-1}$ belongs to the image of $\psi$, or equivalently, the stratum $Sh_\tau$ is nonempty. But we have proved in §12.4 that $[\delta] = [\tau]$ belongs to the image of $\delta_{K^0_p}$. Furthermore, it is obvious that $X_\tau(\tau)_{\widetilde{K}^{\mathbf{a}}_p} \neq \emptyset$, and hence by Proposition 12.6, we conclude that $Sh_\tau \cap \mathcal{S}_{[\tau]} \neq \emptyset$. (*Note:* Using the element $\tau \in \widetilde{W}(\mathrm{GSp}_{2n})$, this argument applies just as well to the Siegel case, thus providing an alternative to Genestier's step of finding an explicit chain of abelian varieties in the moduli problem which maps to $\tau^{-1}$.) □

Recall that $Sh$ is not a connected scheme. In fact, for our two examples, the set of connected components of the geometric generic fiber $Sh_{\overline{\mathbb{Q}}_p}$ carries a simply transitive action by the finite abelian group

$$\pi_0 = \mathbb{Z}^+_{(p)}\backslash(\mathbb{A}^p_f)^\times/c(K^p) = \mathbb{Q}^+\backslash\mathbb{A}^\times_f/c(K^p)\mathbb{Z}^\times_p = \mathbb{Q}^\times\backslash\mathbb{A}^\times/c(\mathbf{K}K_\infty),$$

where $c : \mathbf{G} \to \mathbb{G}_m$ is the similitude homomorphism, $K_\infty$ denotes the centralizer of $h_0$ in $\mathbf{G}(\mathbb{R})$, and the superscript $+$ designates the positive elements of the set; cf. [**Del**]. To see the groups above are actually isomorphic, use the fact that $c(K_\infty) \supset \mathbb{R}^+$ and $c(K_p) = \mathbb{Z}^\times_p$, as one can easily check for each of our two examples. Fixing isomorphisms $\mathbb{A}^p_f(1) = \mathbb{A}^p_f$ and $\overline{\mathbb{Q}}_p = \mathbb{C}$ once and for all, an element $(A_\bullet, \lambda, i, \bar\eta) \in Sh(\overline{\mathbb{Q}}_p)$ belongs to the connected component indexed by $a \in \pi_0$ if and only if the Weil-pairing $(\cdot,\cdot)_\lambda$ on $H_1(A, \mathbb{A}^p_f)$ pulls-back via $\bar\eta$ to the pairing $a(\cdot,\cdot)$ on $V \otimes \mathbb{A}^p_f$; see [**H3**] §2.

Let $\overline{\mathbb{Z}}_p \subset \overline{\mathbb{Q}}_p$ denote the subring of elements integral over $\mathbb{Z}_p$, and fix $a \in \pi_0$. Let $Sh^0$ denote the moduli space (over e.g. $\overline{\mathbb{Z}}_p$) of points $(A_\bullet, \lambda, i, \bar\eta) \in Sh$ such that

$$\bar\eta^*(\cdot,\cdot)_\lambda = a(\cdot,\cdot).$$

Lemma 13.2. *Suppose $K_p$ is any standard parahoric subgroup. Then the fibers of $Sh^0 \to \mathrm{Spec}(\overline{\mathbb{Z}}_p)$ are connected. Furthermore, the morphism $\psi : \widetilde{Sh^0} \to \mathbf{M}^{\mathrm{loc}}_{K_p}$ is surjective.*

Proof. By [**Del**], the generic fiber $Sh^0_{\overline{\mathbb{Q}}_p}$ is connected. In the "fake" unitary case, $Sh^0 \to \mathrm{Spec}(\overline{\mathbb{Z}}_p)$ is proper and flat, and hence by the Zariski connectedness principle (comp. [**Ha**], Ex. III.11.4), the special fiber is also connected. In the

Siegel case $Sh^0 \to \mathrm{Spec}(\overline{\mathbb{Z}}_p)$ is flat but not proper, so the same argument does not apply (but see [**CF**] IV.5.10 when $K_p$ is maximal hyperspecial). However, the connectedness of the special fiber still holds and can be proved in an indirect way from the $p$-adic monodromy theorem of [**CF**]; for details see [**Yu**].

The statement regarding surjectivity follows from the proof of Lemma 13.1: the local model diagram does not "see" $\bar{\eta}$, and so we can arrange matters so that all constructions occur within $Sh^0$ and $\widetilde{Sh}^0$. □

From now on we assume $K_p = K_p^{\mathbf{a}}$. The above lemma proves that in any connected component $Sh^0$, all KR-strata $Sh_w^0$ are nonempty.

In fact, because the KR-stratum $Sh_\tau$ is zero-dimensional, the nonemptiness of $Sh_\tau^0 \cap \mathcal{S}_{[\tau]}$ proves slightly more. The following statement is in some sense the opposite extreme of the result of Genestier-Ngô in Corollary 8.3.

COROLLARY 13.3. *In the "fake" unitary or the Siegel case with $K_p = K_p^{\mathbf{a}}$, let $Sh_\tau^0$ denote the the zero-dimensional KR-stratum in a connected component $Sh^0$. Then $Sh_\tau^0$ is nonempty and is contained in the basic locus of $Sh_{K_p^{\mathbf{a}}}$.*

How can we describe the irreducible components in $Sh^0_{\overline{\mathbb{F}}_p}$? These are clearly just the closures of the irreducible components of the KR-strata $Sh^0_{t_\lambda}$, as $\lambda$ ranges over the Weyl-orbit $W\mu$. A priori, each of these maximal KR-strata might be the (disjoint) union of several irreducible components, all having the same dimension.

COROLLARY 13.4. *In the Siegel or "fake" unitary case, $Sh^0_{\overline{\mathbb{F}}_p}$ is equidimensional, and the number of irreducible components is at least $\#W\mu$.*

In the Siegel case, a much more precise statement has been established by C.-F. Yu [**Yu**], answering in the affirmative a question raised in [**deJ**].

THEOREM 13.5 ([**Yu**]). *In the Siegel case with $K_p = K_p^{\mathbf{a}}$, each maximal KR-stratum $Sh^0_{t_\lambda}$ is irreducible. Hence $Sh^0_{\overline{\mathbb{F}}_p}$ has exactly $2^n$ irreducible components. An analogous statement holds for any standard parahoric subgroup $K_p$.*

It is reasonable to expect that similar methods will apply to the "fake" unitary case to prove that the number of irreducible components in $Sh^0_{\overline{\mathbb{F}}_p}$ is exactly $\#W\mu$. In fact, it would be interesting to determine whether this last statement remains true for any PEL Shimura variety attached to a group whose $p$-adic completion is unramified.

## 14. Appendix: Summary of Dieudonné theory and de Rham and crystalline cohomology for abelian varieties

This summary is extracted from some standard references – [**BBM**], [**Dem**], [**Fon**], [**Il**], [**MaMe**], [**Me**], and [**O**] – as well as from [**deJ**].

**14.1. de Rham cohomology and the Hodge filtration.** To an abelian scheme $a : A \to S$ of relative dimension $g$ is associated the de Rham complex $\Omega^\bullet_{A/S}$ of $\mathcal{O}_A$-modules. We define the de Rham cohomology sheaves

$$R^i a_*(\Omega^\bullet_{A/S}).$$

The first de Rham cohomology sheaf

$$R^1 a_*(\Omega^\bullet_{A/S})$$

is a locally free $\mathcal{O}_S$-module of rank $2g$. If $S$ is the spectrum of a Noetherian ring $R$, then

$$H^1_{DR}(A/S) := H^1(A, \Omega^\bullet_{A/S}) = \Gamma(S, R^1 a_*(\Omega^\bullet_{A/S}))$$

is a locally free $R$-module of rank $2g$.

The Hodge-de Rham spectral sequence degenerates at $E_1$ ([**BBM**], §2.5), yielding the exact sequence

$$0 \to R^0 a_*(\Omega^1_{A/S}) \to R^1 a_*(\Omega^\bullet_{A/S}) \to R^1 a_*(\mathcal{O}_A) \to 0.$$

We define $\omega_A := R^0 a_*(\Omega^1_{A/S})$, a locally free sub-$\mathcal{O}_S$-module of rank $g$. The term $R^1 a_*(\mathcal{O}_A)$ may be identified with $\mathrm{Lie}(\widehat{A})$, the Lie algebra of the dual abelian scheme $\widehat{A}/S$. It is also locally free of rank $g$. Thus we have the *Hodge filtration* on de Rham cohomology

$$0 \to \omega_A \to R^1 a_*(\Omega^\bullet_{A/S}) \to \mathrm{Lie}(\widehat{A}) \to 0.$$

Recall that in our formulation of the moduli problem defining $\mathrm{Sh}(\mathbf{G}, h)_{K_p}$, the important determinant condition refers to the Lie algebra $\mathrm{Lie}(A)$, and not to $\mathrm{Lie}(\widehat{A})$. Because of this it is convenient (although not, strictly-speaking, necessary) to work with the *covariant* analogue $M(A)$ of $R^1 a_*(\Omega^\bullet_{A/S})$. To define it, recall that for any $\mathcal{O}_S$-module $N$, we define the dual $\mathcal{O}_S$-module $N^\vee$ by

$$N^\vee := \mathcal{H}om_{\mathcal{O}_S}(N, \mathcal{O}_S).$$

Let $M(A) := (R^1 a_*(\Omega^\bullet_{A/S}))^\vee$ be the dual of de Rham cohomology. This is a locally free $\mathcal{O}_S$-module of rank $2g$. By the proposition below, we can identify $\omega_A^\vee = \mathrm{Lie}(A)$ and so the *Hodge filtration* on $M(A)$ takes the form

$$0 \to \mathrm{Lie}(\widehat{A})^\vee \to M(A) \to \mathrm{Lie}(A) \to 0.$$

It is sometimes convenient to denote $\mathbb{D}(A)_S := R^1 a_*(\Omega_{A/S})$ (this notation refers to crystalline cohomology, see [**BBM**], [**Il**]).

Proposition 14.1 ([**BBM**], Prop. 5.1.10). *There is a commutative diagram whose vertical arrows are isomorphisms*

$$\begin{tikzcd}
0 \arrow[r] & \mathrm{Lie}(\widehat{A})^\vee \arrow[r] \arrow[d, "\cong"'] & \mathbb{D}(A)_S^\vee \arrow[r] \arrow[d, "\cong"'] & \omega_A^\vee \arrow[r] \arrow[d, "\cong"'] & 0 \\
0 \arrow[r] & \omega_{\widehat{A}} \arrow[r] & \mathbb{D}(\widehat{A})_S \arrow[r] & \mathrm{Lie}(A) \arrow[r] & 0.
\end{tikzcd}$$

**14.2. Crystalline cohomology.** Let $k_r = \mathbb{F}_{p^r}$ be a finite field with ring of Witt vectors $W(k_r)$. The fraction field $L_r$ of $W(k_r)$ is an unramified extension of $\mathbb{Q}_p$ and its Galois group is the cyclic group of order $r$ generated by the Frobenius element $\sigma : x \mapsto x^p$; note also that $\sigma$ acts on Witt vectors by the rule $\sigma(a_0, a_1, \dots) = (a_0^p, a_1^p, \dots)$.

Let $A$ be an abelian variety over $k_r$ of dimension $g$. We have the *integral isocrystal* associated to $A/k_r$, given by the data

$$\mathbb{D}(A) = (H^1_{\mathrm{crys}}(A/W(k_r)), F, V).$$

Here the crystalline cohomology group $H^1_{\mathrm{crys}}(A/W(k_r))$ is a free $W(k_r)$-module of rank $2g$, equipped with a $\sigma$-linear endomorphism $F$ ("Frobenius") and the $\sigma^{-1}$-linear endomorphism $V$ ("Verschiebung") which induce bijections on

$H^1_{\rm crys}(A/W(k_r)) \otimes_{W(k_r)} L_r$. We have the identity $FV = VF = p$ (by *definition* of $V$), hence the inclusions of $W(k_r)$-lattices

$$H^1_{\rm crys}(A/W(k_r)) \supset FH^1_{\rm crys}(A/W(k_r)) \supset pH^1_{\rm crys}(A/W(k_r))$$

(as well as the analogous inclusions for $V$ replacing $F$).

The endomorphism $F$ has the property that $F^r = \pi_A$ on $H^1_{\rm crys}$, where $\pi_A$ denotes the absolute Frobenius morphism of $A$ relative to the field of definition $k_r$ (on projective coordinates $x_i$ for $A$, $\pi_A$ induces the map $x_i \mapsto x_i^{p^r}$).

**14.3. Relation with Dieudonné theory.** The crystalline cohomology of $A/k_r$ is intimately connected to the (contravariant) Dieudonné module of the $p$-divisible group $A[p^\infty] := \varinjlim A[p^n]$, the union of the sub-groupschemes $A[p^n] = \ker(p^n : A \to A)$. Recall that the classical contravariant Dieudonné functor $G \mapsto D(G)$ establishes an exact anti-equivalence between the categories

$$\Big\{p\text{-divisible groups } G = \varinjlim G_n \text{ over } k_r\Big\}$$

and

$$\{\text{free } W(k_r)\text{-modules } M = \varprojlim M/p^nM, \text{ equipped with operators } F, V\};$$

see [**Dem**], [**Fon**]. Here $F$ and $V$ are $\sigma$- resp. $\sigma^{-1}$-linear, inducing bijections on $M \otimes_{W(k_r)} L_r$.

The crystalline cohomology of $A/k_r$, together with the operators $F$ and $V$, is the same as the Dieudonné module of the $p$-divisible group $A[p^\infty]$, in the sense that there is a canonical isomorphism

$$H^1_{\rm crys}(A/W(k_r)) \cong D(A[p^\infty]) \tag{14.3.1}$$

which respects the endomorphisms $F$ and $V$ on both sides, cf. [**BBM**]. Moreover, we have the following identifications

$$\mathbb{D}(A)_{k_r} := H^1_{\rm crys}(A/W(k_r)) \otimes_{W(k_r)} k_r \cong H^1_{DR}(A/k_r) \cong D(A[p]). \tag{14.3.2}$$

The second isomorphism is due to Oda [**O**]; see below. The first isomorphism is a standard fact ([**BBM**]), but can also be deduced via Oda's theorem by reducing equation (14.3.1) modulo $p$: the exactness of the functor $D$ implies that $D(A[p]) = D(A[p^\infty])/pD(A[p^\infty]) = D(A[p^\infty]) \otimes_{W(k_r)} k_r$. In particular, the $k_r$-vector space $H^1_{DR}(A/k_r)$ inherits endomorphisms $F$ and $V$ ($\sigma$- resp. $\sigma^{-1}$-linear).

The theorem of Oda [**O**] includes as well the relation between the Hodge filtration on the de Rham cohomology of $A$ and a suitable filtration on the isocrystal $\mathbb{D}(A)$.

THEOREM 14.2 ([**O**], Cor. 5.11). *There is a natural isomorphism $\psi : \mathbb{D}(A)_{k_r} \tilde{\to} H^1_{DR}(A/k_r)$, and under this isomorphism, $V\mathbb{D}(A)_{k_r}$ is taken to $\omega_A$. In particular there is an exact sequence*

$$0 \to V\mathbb{D}(A) \to \mathbb{D}(A) \to \mathrm{Lie}(\widehat{A}) \to 0.$$

**14.4. Remarks on duality.** We actually make use of this in a dual formulation. Define $H = H(A)$ to be the $W(k_r)$-linear dual of the isocrystal $\mathbb{D}(A)$

$$H(A) = \mathrm{Hom}_{W(k_r)}(\mathbb{D}(A), W(k_r)), \tag{14.4.1}$$

and define $H_{L_r} = H(A)_{L_r} = H(A) \otimes_{W(k_r)} L_r$. Letting $\langle\ ,\ \rangle : H \times \mathbb{D}(A) \to W(k_r)$ denote the canonical pairing, we define $\sigma$- resp. $\sigma^{-1}$-linear injections $F$ resp. $V$ on $H$ (they are bijective on $H_{L_r}$) by the formulae

$$\langle Fu, a\rangle = \sigma\langle u, Va\rangle \tag{14.4.2}$$

$$\langle Vu, a\rangle = \sigma^{-1}\langle u, Fa\rangle \tag{14.4.3}$$

for $u \in H$ and $a \in \mathbb{D}(A) = H^1_{\mathrm{crys}}(A/W(k_r))$.

Of course $H(A)_{k_r} := H \otimes_{W(k_r)} k_r$ is the $k_r$-linear dual of $\mathbb{D}(A)_{k_r}$, hence

$$H(A)_{k_r} = M(A/k_r).$$

Lemma 14.3. *Let $S = \mathrm{Spec}(k_r)$. Equip $H = \mathbb{D}(A)^\vee$ with operators $F, V$ as in (14.4.2). Then the isomorphism*

$$\mathbb{D}(A)^\vee_{k_r} \xrightarrow{\sim} \mathbb{D}(\widehat{A})_{k_r}$$

*of Proposition 14.1 is an isomorphism of $W(k_r)[F,V]$-modules.*

Proof. From [**Dem**], Theorem 8 (p. 71), for a p-divisible group $G$ with Serre dual $G'$ (loc. cit., p. 46) there is a duality pairing in the category of $W(k_r)[F,V]$-modules

$$\langle\ ,\ \rangle : D(G') \times D(G) \to W(k_r)(-1)$$

where $W(k_r)(-1)$ is the isocrystal with underlying space $W(k_r)$ and $\sigma$-linear endomorphism $p\sigma$. That is, we have the identity

$$\langle F_{G'}x, F_G y\rangle = p\sigma\langle x, y\rangle,$$

or

$$\langle F_{G'}x, y\rangle = \sigma\langle x, pF_G^{-1}y\rangle = \sigma\langle x, V_G y\rangle.$$

There is a canonical identification

$$(A[p^\infty])' = \widehat{A}[p^\infty].$$

Now from the above pairing with $G = A[p^\infty]$ and (14.3.1) we deduce a duality pairing of $W(k_r)[F,V]$-modules

$$\mathbb{D}(\widehat{A}) \times \mathbb{D}(A) \to W(k_r)(-1),$$

which induces the isomorphism of Proposition 14.1. The lemma follows from these remarks. □

Applying Oda's theorem (14.2) to $\widehat{A}$ and invoking the above lemma gives $\mathrm{Lie}(\widehat{A})^\vee = VM(A/k_r)$, thus there is an exact sequence

$$0 \to V\,H(A) \to H(A) \to \mathrm{Lie}(A) \to 0.$$

Thus, we have

Corollary 14.4. *Let $F, V$ be the $\sigma$- resp. $\sigma^{-1}$-linear endomorphisms of $H = H(A)$ defined in (14.4.2). There is a natural isomorphism*

$$\frac{V^{-1}\,H}{H} = \sigma(\mathrm{Lie}(A)).$$

*Moreover, on $H_{L_r}$ we have the identity $V^{-r} = \pi_A^{-1}$. (Comp.* [**Ko92**]*, §10, 16.)*

## References

[AG] F. Andreatta, E.Z. Goren, *Geometry of Hilbert modular varieties over totally ramified primes*, IMRN **33**, (2003), 1785-1835.

[BBD] A. Beilinson, I.N. Bernstein, P. Deligne, *Faisceaux Pervers*, Astérisque **100**, (1981).

[BBM] P. Berthelot, L. Breen, W. Messing, *Théorie de Dieudonné cristalline*, vol. 2, LNM **930**, Springer 1982.

[BR] D. Blasius, J.D. Rogawski, *Zeta functions of Shimura varieties*, in: Motives, vol. 2, Proc. Symp. Pure Math. **55.2** (1994), 525-571.

[Bo] A. Borel, *Automorphic L-functions*, In: *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math., vol. **33**, part 2, Amer. Math. Soc., Providence, RI, 1979, pp. 27–61.

[BT1] F. Bruhat and J. Tits, *Groupes réductifs sur un corps local. I*, Inst. Hautes Études Sci. Publ. Math. **41** (1972), 5–251.

[BT2] F. Bruhat and J. Tits, *Groupes réductifs sur un corps local. II*, Inst. Hautes Études Sci. Publ. Math. **60** (1984), 5-184.

[Ca] P. Cartier, *Representations of $\mathfrak{p}$-adic groups: a survey* In: *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math., vol. **33**, part 1, Amer. Math. Soc., Providence, RI, 1979, pp. 111–155.

[CF] C.-L. Chai, G. Faltings, *Degeneration of abelian varieties*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Band 22, Springer-Verlag (1990).

[deJ] A.J. de Jong, *The moduli spaces of principally polarized abelian varieties with $\Gamma_0(p)$-level structure*, J. Alg. Geom. **2** (1993), 667-688.

[SGA7 XIII] P. Deligne, *Le formalisme des cycles évanescents*, SGA7, Exp. XIII.

[Del] P. Deligne, *Travaux de Shimura*, Sem. Bourbaki 389. LNM **244**, Springer 1971.

[Del2] P. Deligne, *Variétés de Shimura: Interprétation modulaire, et techniques de construction de modèles canoniques*, Proc. Symp. Pure Math. **33** (1979), part 2, 247-290.

[Weil2] P. Deligne, *La conjecture de Weil II*, Publ. Math. IHES **52**, 137–252 (1980)

[Dem] M. Demazure, *Lectures on p-divisible groups*, Springer LNM **302** (1972), 98 pp.

[DP] P. Deligne, G. Pappas, *Singularities des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant*, Composition Math. **90** (1994), 59-79.

[FM] L. Fargues, E. Mantovan *Variétés de Shimura, espaces de Rapoport-Zink et correspondances de Langlands locales*, Astérisque **291** (2004), 331 pp.

[Fon] J.-M. Fontaine, *Groupes p-divisibles sur les corps locaux*, Astérisque **47-48**, (1977), 262 pp.

[Ga] D. Gaitsgory. *Construction of central elements in the affine Hecke algebra via nearby cycles*, Invent. Math. **144** (2001), 253-280.

[Ge] A. Genestier, *Un modèle semi-stable de la variété de Siegel de genre 3 avec structures de niveau de type* $\Gamma_0(p)$, Compositio Math. **123** (2000), no. 3, 303-328.

[GN] A. Genestier, B.C. Ngô, *Alcoves et p-rang des variétés abéliennes*, Ann. Inst. Fourier **52** (2002), 1665-1680.

[GOo] E.Z. Goren, F. Oort, *Stratifications of Hilbert modular varieties*, J. Alg. Geom. **9**, (2000), no. 1, 111-154.

[Go1] U. Görtz, *On the flatness of models of certain Shimura varieties of PEL-type*, Math. Ann. **321** (2001), no. 3, 689-727.

[Go2] U. Görtz, *On the flatness of local models for the symplectic group*, Adv. Math. **176**, (2003), no.1, 89-115.

[Go3] U. Görtz, *Topological flatness of local models in the ramified case*, math.AG/0205050.

[Go4] U. Görtz, *Computing the alternating trace of Frobenius on the sheaves of nearby cycles on local models for $GL_4$ and $GL_5$*, J. of Algebra **278** (2004), no.1 , 148-172.

[GH] U. Görtz, T. Haines, *The Jordan-Hölder series for nearby cycles on some Shimura varieties and affine flag varieties*, preprint (2003). math.AG/0402143.

[SGA7 I] A. Grothendieck, *Groupes de monodromie en géométrie algébrique*, Exp. I, (rédigé par P. Deligne), Springer LNM **288** (1972), 1-24.

[H1] T. Haines, *The combinatorics of Bernstein functions*, Trans. Amer. Math. Soc. **353** (2001), 1251–1278.

[H2] T. Haines, *Test functions for Shimura varieties: the Drinfeld case*, Duke Math. J. **106** (2001), no. 1, 19–40.

[H3] T. Haines, *On connected components of Shimura varieties*, Canad. J. Math., vol. **54** (2), 2002, pp. 352-395.

[HKP] T. Haines, R. Kottwitz, and A. Prasad, *Iwahori-Hecke Algebras*, math.RT/0309168.

[HN1] T. Haines, B.C. Ngô, *Nearby cycles for local models of some Shimura varieties*, Compositio Math. **133** (2002), no. 2, 117–150.

[HN2] T. Haines, B.C. Ngô, *Alcoves associated to special fibers of local models*, Amer. J. Math. **124** (2002), no. 6, 1125–1152.

[HN3] T. Haines, B.C. Ngô, *On the semi-simple local L-factors for some simple Shimura varieties*, in preparation.

[HP] T. Haines, A. Pettet, *Formulae relating the Bernstein and Iwahori-Matsumoto presentations of an affine Hecke algebra*, J. Algebra **252** (2002), no. 1, 127–149.

[HT] M. Harris, R. Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Math. Studies, Princeton Univ. Press, 276 pp + vi, 2001.

[Ha] R. Hartshorne, *Algebraic Geometry*, GTM **52**, Springer-Verlag, 1977.

[Il] L. Illusie, *Report on crystalline cohomology*, Proc. Symp. Pure Math. **29** (1975), 459-478.

[Ko84] R. Kottwitz, *Shimura varieties and twisted orbital integrals*, Math. Ann. **269** (1984), 287-300.

[Ko85] R. Kottwitz, *Isocrystals with additional structure*, Compositio Math. **56** (1985), 201-220.

[Ko90] R. Kottwitz, *Shimura varieties and $\lambda$-adic representations*, In: Automorphic Forms, Shimura Varieties and $L$-functions, part 1. Perspectives in Mathematics 10. Academic Press, 1990, pp. 161-209.

[Ko92] R. Kottwitz, *Points on some Shimura varieties over finite fields*, J. Amer. Math. Soc. **5** (1992), 373-444.

[Ko92b] R. Kottwitz, *On the $\lambda$-adic representations associated to some simple Shimura varieties*, Invent. Math. **108** (1992), 653-665.

[Ko97] R. Kottwitz, *Isocrystals with additional structure. II*, Compositio Math. **109**, (1997), 255-339.

[Ko03] R. Kottwitz, *On the Hodge-Newton decomposition for split groups*, IMRN **26** (2003), 1433-1447.

[KR] R. Kottwitz, M. Rapoport, *Minuscule alcoves for $GL_n$ and $GSp_{2n}$*, Manuscripta Math. **102** (2000), no. 4, 403–428.

[KW] R. Kiehl, R. Weissauer, *Weil conjectures, Perverse sheaves, and $\ell$-adic Fourier transform*, Springer Erg. Math. 3 Folg., **42** (2001).

[Lu] G. Lusztig, *Affine Hecke algebras and their graded versions*, J. Amer. Math. Soc.**2** (1989), 599–635.

[Ma] B. Mazur, *Notes on étale cohomology of number fields*, Ann. Sci. Ecole Norm. Sup. **6** (1973), 521-556.

[MaMe] B. Mazur, W. Messing, *Universal extensions and one dimensional crystalline cohomology*, Springer LNM **370** (1974), 134 pp.

[Me] W. Messing, *The crystals associated to Barsotti-Tate groups*, LNM **264**, Springer 1962.

[Mi] J.S. Milne, *Abelian Varieties*, p. 103-150, In: Arithmetic Geometry, ed. G. Cornell, J.H. Silverman, Springer-Verlag, 1986.

[Mu] D. Mumford, *Abelian varieties*, Oxford University Press, 1974.

[O] T. Oda, *The first de Rham cohomology group and Dieudonné modules*, Ann. Scient. École Norm. Sup., 4 ser., t. 2, (1969), 63-135.

[Oo] F. Oort, *Newton polygon strata in the moduli space of abelian varieties*, In: Moduli of abelian varieties (Texel Island, 1999), 417-440, Progr. Math., **195**, Birkhäuser, Basel, 2001.

[P] G. Pappas, *On the arithmetic moduli schemes of PEL Shimura varieties*, J. Alg. Geom. **9** (2000), 577-605.

[PR1] G. Pappas, M. Rapoport, *Local models in the ramified case I. The EL-case*, J. Alg. Geom. **12** (2003), 107-145.

[PR2] G. Pappas, M. Rapoport, *Local models in the ramified case II. Splitting models*, Duke Math. J. **127** (2005), no. 2, 193-250.

[R1] M. Rapoport, *On the bad reduction of Shimura varieties.* Automorphic Forms, Shimura Varieties, and $L$-functions, Vol. II, (Ann Arbor, MI, 1988), 253-321, Perspec. Math **11**, Academic Press, Boston, MA, 1990.

[R2] M. Rapoport, *A Guide to the reduction modulo p of Shimura varieties*, preprint (2002). math.AG/0205022.

[RR] M. Rapoport, M. Richartz, *On the classification and specialization of $F$-isocrystals with additional structure*, Compositio Math. **103** (1996), 153-181.

[RZ82] M. Rapoport, Th. Zink, *Über die lokale Zetafunktion von Shimuravarietäten. Monodromiefiltration und verschwindende Zyklen in ungleicher Charakteristik*, Invent. Math. **68** (1982), no.1, 21-101.

[RZ] M. Rapoport, T. Zink, *Period domains for p-divisible groups*, Ann. of Math. Stud. **141**, Princeton University Press (1996).

[Re1] H. Reimann, *The semi-simple zeta function of quaternionic Shimura varieties*, Lecture Notes in Mathematics **1657**, Springer-Verlag, Berlin, 1997, viii + 143 pp.

[Re2] H. Reimann, *Reduction of quaternionic Shimura varieties at parahoric levels*, Manuscripta Math. **107**, (2002), no. 3, 355-390.

[Re3] H. Reimann, *Cohomology of quaternionic Shimura varieties at parahoric levels*, Manuscripta Math. **107**, (2002), no. 3, 391-407.

[Ta] J. Tate, *Number theoretic background*, In: *Automorphic Forms, Representations and L-functions*, Proc. Sympos. Pure Math., vol. **33**, part 2, Amer. Math. Soc., Providence, RI, 1979, pp. 3 –26.

[T] J. Tits, *Reductive groups over local fields*, In: Automorphic Forms, Representations, and $L$-finctions, ed. A. Borel, W. Casselman, Proc. Symp. Pure Math. **33** (1979), Part I, 29-69.

[W1] T. Wedhorn, *Ordinariness in good reductions of Shimura varieties of PEL-type*, Ann. Sci. École Norm. Sup. (IV), **32**, (1999), no. 5, 575-618.

[W2] T. Wedhorn, *The local Langlands correspondence for* $\mathrm{GL}(n)$ *over p-adic fields*, math.AG/0011210 v2.

[Wi] J.-P. Wintenberger, *Existence de $F$-cristaux avec structures supplementaires*, Adv. Math. **190** (2005), no. 1, 196-224.

[Yu] C.-F. Yu, *Irreducibility of the Siegel moduli spaces with parahoric level structure*, IMRN, **48** (2004), 2593-2597.

[Z] Th. Zink, *Über die schlechte Reducktion einiger Shimuramannigfaltigkeiten*, Compositio Math. **45** (1981), 15-107.

Mathematics Department, University of Maryland, College Park, MD 20742-4015
*E-mail address*: `tjh@math.umd.edu`

Clay Mathematics Proceedings
Volume 4, 2005

# A Statement of the Fundamental Lemma

Thomas C. Hales

ABSTRACT. These notes give a statement of the fundamental lemma, which is a conjectural identity between $p$-adic integrals.

## 1. Introduction

**Notation.** Let $F$ be a $p$-adic field, given either as a finite field extension of $\mathbb{Q}_p$, or as the field $F = \mathbb{F}_q((t))$. Let $\mathbb{F}_q$ (a finite field with $q$ elements and characteristic $p$) be the residue field of $F$. Let $\bar{F}$ be a fixed algebraic closure of $F$. Let $F^{un}$ be the maximal unramified extension of $F$ in $\bar{F}$. For simplicity, we also assume that the characteristic of $F$ is not 2.

The fundamental lemma pertains to groups that satisfy a series of hypotheses. Here is the first.

ASSUMPTION 1.1. *$G$ is a connected reductive linear algebraic group that is defined over $F$.*

The following examples give the $F$-points of three different families of connected reductive linear algebraic groups: orthogonal, symplectic, and unitary groups.

EXAMPLE 1.2. Let $M(n, F)$ be the algebra of $n$ by $n$ matrices with coefficients in $F$. Let $J \in M(n, F)$ be a symmetric matrix with nonzero determinant. The special orthogonal group with respect to the matrix $J$ is

$$\mathrm{SO}(n, J, F) = \{X \in M(n, F) \mid {}^tXJX = J, \quad \det(X) = 1\}.$$

EXAMPLE 1.3. Let $J \in M(n, F)$, with $n = 2k$, be a skew-symmetric matrix ${}^tJ = -J$ with nonzero determinant. The symplectic group with respect to $J$ is defined in a similar manner:

$$\mathrm{Sp}(2k, J, F) = \{X \in M(2k, F) \mid {}^tXJX = J\}.$$

EXAMPLE 1.4. Let $E/F$ be a separable quadratic extension. Let $\bar{x}$ be the Galois conjugate of $x \in E$ with respect to the nontrivial automorphism of $E$ fixing $F$. For any $A \in M(n, E)$, let $\bar{A}$ be the matrix obtained by taking the Galois

I thank R. Kottwitz and S. DeBacker for many helpful comments.
This work was supported in part by the NSF.

conjugate of each coefficient of $A$. Let $J \in M(n,E)$ satisfy ${}^t\bar{J} = J$ and have a nonzero determinant. The unitary group with respect to $J$ and $E/F$ is

$$U(n,J,F) = \{X \in M(n,E) \mid {}^t\bar{X}JX = J\}.$$

The algebraic groups $SO(n,J)$, $Sp(2k,J)$, and $U(n,J)$ satisfy Assumption 1.1.

Assumption 1.5. *$G$ splits over an unramified field extension.*

That is, there is an unramified extension $F_1/F$ such that $G \times_F F_1$ is split.

Example 1.6. In the first two examples above (orthogonal and symplectic), if we take $J$ to have the special form

$$J = \begin{pmatrix} 0 & 0 & * \\ 0 & * & 0 \\ * & 0 & 0 \end{pmatrix} \tag{1.6.1}$$

(that is, nonzero entries from $F$ along the cross-diagonal and zeros elsewhere), then $G$ splits over $F$. In the third example (unitary), if $J$ has this same form and if $E/F$ is unramified, then the unitary group splits over the unramified extension $E$ of $F$.

Assumption 1.7. *$G$ is quasi-split.*

This means that there is an $F$-subgroup $B \subset G$ such that $B \times_F \bar{F}$ is a Borel subgroup of $G \times_F \bar{F}$.

Example 1.8. In all three cases (orthogonal, symplectic, and unitary), if $J$ has the cross-diagonal form 1.6.1, then $G$ is quasi-split. In fact, we can take the points of $B$ to be the set of upper triangular matrices in $G(F)$.

Assumption 1.9. *$K$ is a hyperspecial maximal compact subgroup of $G(F)$, in the sense of Definition 1.11.*

Example 1.10. Let $O_F$ be the ring of integers of $F$ and let $K = GL(n, O_F)$. This is a hyperspecial maximal compact subgroup of $GL(n,F)$.

Definition 1.11. $K$ is *hyperspecial* if there exists $\mathcal{G}$ such that the following conditions are satisfied.

- $\mathcal{G}$ is a smooth group scheme over $O_F$,
- $G = \mathcal{G} \times_{O_F} F$,
- $\mathcal{G} \times_{O_F} \mathbb{F}_q$ is reductive,
- $K = \mathcal{G}(O_F)$.

Example 1.12. In all three examples (orthogonal, symplectic, and unitary), take $G$ to have the form of Example 1.6. Assume that each cross-diagonal entry is a *unit* in the ring of integers. Assume further that the residual characteristic is not 2. Then the equations

$${}^tXJX = J \quad \text{(or in the unitary case } {}^t\bar{X}JX = J\text{)}$$

define a group scheme $\mathcal{G}$ over $O_F$, and $\mathcal{G}(O_F)$ is hyperspecial.

## 2. Classification of Unramified Reductive Groups

Definition 2.1. If $G$ is quasi-split and splits over an unramified extension (that is, if $G$ satisfies Assumptions 1.5 and 1.7), then $G$ is said to be an *unramified reductive group.*

Let $G$ be an unramified reductive group. It is classified by data (called root data)

$$(X^*, X_*, \Phi, \Phi^\vee, \sigma).$$

The data are as follows:

- $X^*$ is the character group of a Cartan subgroup of $G$.
- $X_*$ is the cocharacter group of the Cartan subgroup.
- $\Phi \subset X^*$ is the set of roots.
- $\Phi^\vee \subset X_*$ is the set of coroots.
- $\sigma$ is an automorphism of finite order of $X^*$ sending a set of simple roots in $\Phi$ to itself.

  $\sigma$ is obtained from the action on the character group induced from the Frobenius automorphism of $\mathrm{Gal}(F^{un}/F)$ on the maximally split Cartan subgroup in $G$.

The first four elements $(X^*, X_*, \Phi, \Phi^\vee)$ classify split reductive groups $G$ over $F$. For such groups $\sigma = 1$.

## 3. Endoscopic Groups

$H$ is an unramified endoscopic group of $G$ if it is an unramified reductive group over $F$ whose classifying data has the form

$$(X^*, X_*, \Phi_H, \Phi_H^\vee, \sigma_H).$$

The first two entries are the same for $G$ as for $H$. To distinguish the data for $H$ from that for $G$, we add subscripts $H$ or $G$, as needed. The data for $H$ are subject to the constraints that there exists an element $s \in \mathrm{Hom}(X_*, \mathbb{C}^\times)$ and a Weyl group element $w \in W(\Phi_G)$ such that

- $\Phi_H^\vee = \{\alpha \in \Phi_G^\vee \mid s(\alpha) = 1\}$,
- $\sigma_H = w \circ \sigma_G$, and
- $\sigma_H(s) = s$.

**3.1. Endoscopic groups for $SL(2)$.** As an example, we determine the unramified endoscopic groups of $G = SL(2)$. The character group $X^*$ can be identified with $\mathbb{Z}$, where $n \in \mathbb{Z}$ is identified with the character on the diagonal torus given by

$$\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \mapsto t^n.$$

The set $\Phi$ can be identified with the subset $\{\pm 2\}$ of $\mathbb{Z}$:

$$\begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} \mapsto t^{\pm 2}.$$

The cocharacter group $X_*$ is also identified with $\mathbb{Z}$, where $n \in \mathbb{Z}$ is identified with

$$t \mapsto \begin{pmatrix} t^n & 0 \\ 0 & t^{-n} \end{pmatrix}.$$

Under this identification $\Phi^\vee = \{\pm 1\}$. Since the group is split, $\sigma = 1$.

We get an unramified endoscopic group by selecting $s \in \mathrm{Hom}(X_*, \mathbb{C}^\times) \cong \mathbb{C}^\times$ and $w \in W(\Phi)$.

$$
\begin{aligned}
\Phi_H^\vee = \{\alpha \mid s(\alpha) = 1\} \quad &= \{n \in \{\pm 1\} \mid s^n = 1\} \\
&= \text{if } (s = 1) \text{ then } \Phi_G^\vee, \text{ else } \emptyset.
\end{aligned}
\tag{3.0.1}
$$

We consider two cases, according as $w$ is nontrivial or trivial. If $w$ is the nontrivial reflection, then $\sigma_H = w$ acts by negation on $\mathbb{Z}$. Thus,

$$(\sigma_H(s) = s) \implies (s^{-1} = s) \implies (s = \pm 1).$$

If $s = 1$, then $\sigma_H$ does not fix a set of simple roots as required. So $s = -1$ and $\Phi_H^\vee = \emptyset$. Thus, $H$ has root data

$$(\mathbb{Z}, \mathbb{Z}, \emptyset, \emptyset, w)$$

This determines $H$ up to isomorphism as $H = U_E(1)$, a 1-dimensional torus split by an unramified quadratic extension $E/F$.

If $w$ is trivial, then there are two further cases, according as $\Phi_H$ is empty or not:

- The endoscopic group $\mathbb{G}_m$ has root data
$$(\mathbb{Z}, \mathbb{Z}, \emptyset, \emptyset, 1).$$
- The endoscopic group $H = SL(2)$ has root data
$$(\mathbb{Z}, \mathbb{Z}, \{\pm 2\}, \{\pm 1\}, 1).$$

In summary, the three unramified endoscopic groups of $SL(2)$ are $U_E(1)$, $\mathbb{G}_m$, and $SL(2)$ itself.

**3.2. Endoscopic groups for $PGL(2)$.** As a second complete example, we determine the endoscopic groups of $PGL(2)$. The group $PGL(2)$ is dual to $SL(2)$ in the sense that the coroots of one group can be identified with the roots of the other group. $PGL(2)$ has root data

$$(\mathbb{Z}, \mathbb{Z}, \{\pm 1\}, \{\pm 2\}, 1).$$

When the Weyl group element is trivial, then the calculation is almost identical to the calculation for $SL(2)$. We find that there are again two cases, according as $\Phi_H$ is empty or not:

- The endoscopic group $\mathbb{G}_m$ has root data
$$(\mathbb{Z}, \mathbb{Z}, \emptyset, \emptyset, 1).$$
- The endoscopic group $H = PGL(2)$ has root data
$$(\mathbb{Z}, \mathbb{Z}, \{\pm 1\}, \{\pm 2\}, 1).$$

When the Weyl group element $w$ is nontrivial, then $s \in \{\pm 1\}$, as in the $SL(2)$ calculation.

$$\Phi_H^\vee = \{\alpha \mid s(\alpha) = 1\} = \{n \in \{\pm 2\} \mid s^n = 1\} = \Phi_G^\vee. \tag{3.0.2}$$

From this, we see that picking $w$ to be nontrivial is incompatible with the requirement that $\sigma_H = w$ must fix a set of simple roots. Thus, there are no endoscopic groups with $w$ nontrivial.

In summary, the two endoscopic groups of $PGL(2)$ are $\mathbb{G}_m$ and $PGL(2)$ itself.

### 3.3. Elliptic Endoscopic groups.

DEFINITION 3.1. An unramified endoscopic group $H$ is said to be *elliptic* if

$$(\mathbb{R}\Phi_G)^{W(\Phi_H)\rtimes\langle\sigma_H\rangle} = (0).$$

That is, the span of the set of roots of $G$ has no invariant vectors under the Weyl group of $H$ and the automorphism $\sigma_H$.

The origin of the term *elliptic* is the following. We will see below that each Cartan subgroup of $H$ is isomorphic to a Cartan subgroup of $G$. (Here and elsewhere, when we speak of an isomorphism between algebraic groups defined over $F$, we mean an isomorphism over $F$.) The condition on $H$ for it to be elliptic is precisely the condition that is needed for some Cartan subgroup of $H$ to be isomorphic to an *elliptic* Cartan subgroup of $G$.

EXAMPLE 3.2. We calculate the elliptic unramified endoscopic subgroups of $SL(2)$. We may identify $\mathbb{R}\Phi$ with $\mathbb{R}\{\pm 2\}$ and hence with $\mathbb{R}$. An unramified endoscopic group is elliptic precisely when $W(\Phi_H)$ or $\langle\sigma_H\rangle$ contains the nontrivial reflection $x \mapsto -x$. When $H = SL(2)$, the Weyl group contains the nontrivial reflection. When $H = U_E(1)$, the element $\sigma_H$ is the nontrivial reflection. But when $H = \mathbb{G}_m$, both $W(\Phi_H)$ and $\langle\sigma_H\rangle$ are trivial. Thus, $H = SL(2)$ and $H = U_E(1)$ are elliptic, but $H = \mathbb{G}_m$ is not.

**3.4. An exercise: elliptic endoscopic groups of unitary groups.** This exercise is a calculation of the elliptic unramified endoscopic groups of $U(n,J)$. We assume that $J$ is a cross-diagonal matrix with units along the cross-diagonal as in Section 1.6.1. We give a few facts about the endoscopic groups of $U(n,J)$ and leave it as an exercise to fill in the details.

Let $T = \{\mathrm{diag}(t_1,\ldots,t_n)\}$ be the group of diagonal $n$ by $n$ matrices. The character group $X^*$ can be identified with $\mathbb{Z}^n$ in such a way that the character

$$\mathrm{diag}(t_1,\ldots,t_n) \mapsto t_1^{k_1}\cdots t_n^{k_n}$$

is identified with $(k_1,\ldots,k_n)$.

The cocharacter group can be identified with $\mathbb{Z}^n$ in such a way that the cocharacter

$$t \mapsto \mathrm{diag}(t^{k_1},\ldots,t^{k_n})$$

is identified with $(k_1,\ldots,k_n)$.

Let $e_i$ be the basis vector of $\mathbb{Z}^n$ whose $j$-th entry is Kronecker $\delta_{ij}$. The set of roots can be identified with

$$\Phi = \{e_i - e_j \mid i \neq j\}.$$

The set of coroots $\Phi^\vee$ can be identified with the set of roots $\Phi$ under the isomorphism $X_* \cong \mathbb{Z}^n \cong X^*$.

We may identify $\mathrm{Hom}(X_*,\mathbb{C}^\times)$ with $\mathrm{Hom}(\mathbb{Z}^n,\mathbb{C}^\times) = (\mathbb{C}^\times)^n$. Thus, we take the element $s$ in the definition of endoscopic group to have the form $s = (s_1,\ldots,s_n) \in (\mathbb{C}^\times)^n$. The element $\sigma = \sigma_G$ acts on characters and cocharacters by

$$\sigma(k_1,\ldots,k_n) = (-k_n,\ldots,-k_1).$$

Let $I = \{1,\ldots,n\}$. Show that if $H$ is an elliptic unramified endoscopic group, then there is a partition

$$I = I_1 \coprod I_2$$

with $s_i = 1$ for $i \in I_1$ and $s_i = -1$ otherwise. The elliptic endoscopic group is a product of two smaller unitary groups $H = U(n_1) \times U(n_2)$, where $n_i = \#I_i$, for $i = 1, 2$.

## 4. Cartan subgroups

All unramified reductive groups are classified by their root data. This includes the classification of unramified tori $T$ as a special case (in this case, the set of roots and the set of coroots are empty):

$$(X^*(T), X_*(T), \emptyset, \emptyset, \sigma).$$

We can extend this classification to ramified tori. If $T$ is any torus over $F$, it is classified by

$$(X^*(T), X_*(T), \rho),$$

where $\rho$ is now allowed to be any homomorphism

$$\rho : \mathrm{Gal}(\bar{F}/F) \to \mathrm{Aut}(X^*(T))$$

with finite image.

A basic fact is that $T$ embeds over $F$ as a Cartan subgroup in a given unramified reductive group $G$ if and only if the following two conditions hold.

- The image of $\rho$ in $\mathrm{Aut}(X^*(T))$ is contained in $W(\Phi_G) \rtimes \langle \sigma_G \rangle$.
- There is a commutative diagram:

$$\begin{tikzcd} \mathrm{Gal}(\bar{F}/F) \arrow[r] \arrow[d, "\rho"'] & \mathrm{Gal}(F^{un}/F) \arrow[d, "\mathrm{Frob} \mapsto \sigma_G"] \\ W(\Phi_G) \rtimes \langle \sigma_G \rangle \arrow[r, "w \rtimes \tau \mapsto \tau"'] & \langle \sigma_G \rangle. \end{tikzcd}$$

It follows that every Cartan subgroup $T_H$ of $H$ is isomorphic over $F$ to a Cartan subgroup $T_G$ of $G$. (To check this, simply observe that these two conditions are more restrictive for $H$ than the corresponding conditions for $G$.) The isomorphism can be chosen to induce an isomorphism of Galois modules between the character group (and cocharacter group) of $T_H$ and that of $T_G$.

We say that a semisimple element in a reductive group is *strongly regular*, if its centralizer is a Cartan subgroup. If $\gamma \in H(F)$ is strongly regular semisimple, then its centralizer $T_H$ is isomorphic to some $T_G \subset G$. Let $\gamma_0 \in T_G(F) \subset G(F)$ be the element in $G(F)$ corresponding to $\gamma \in T_H(F) \subset H(F)$, under this isomorphism.

REMARK 4.1. The element $\gamma_0$ is not uniquely determined by $\gamma$. The Cartan subgroup $T_G$ can always be replaced with a conjugate $g^{-1}\, T_G\, g$, $g \in G(F)$, without altering the root data. However, the non-uniqueness runs deeper than this. An example will be worked in Section 8.1 to show how to deal with the problem of non-uniqueness. Non-uniqueness of $\gamma_0$ is related to stable conjugacy, which is our next topic.

## 5. Stable Conjugacy

DEFINITION 5.1. Let $\delta$ and $\delta'$ be strongly regular semisimple elements in $G(F)$. They are *conjugate* if $g^{-1}\delta g = \delta'$ for some $g \in G(F)$. They are *stably conjugate* if $g^{-1}\delta g = \delta'$ for some $g \in G(\bar{F})$.

EXAMPLE 5.2. Let $G = SL(2)$ and $F = \mathbb{Q}_p$. Assume that $p \neq 2$ and that $u$ is a unit that is not a square in $\mathbb{Q}_p$. Let $\epsilon = \sqrt{u}$ in an unramified quadratic extension of $\mathbb{Q}_p$. We have the matrix calculation

$$\begin{pmatrix} 1+p & 1 \\ 2p+p^2 & 1+p \end{pmatrix} \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix} = \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix} \begin{pmatrix} 1+p & u^{-1} \\ (2p+p^2)u & 1+p \end{pmatrix}.$$

This matrix calculation shows that the matrices

$$\begin{pmatrix} 1+p & 1 \\ 2p+p^2 & 1+p \end{pmatrix} \text{ and } \begin{pmatrix} 1+p & u^{-1} \\ (2p+p^2)u & 1+p \end{pmatrix} \tag{5.2.1}$$

of $SL(2, \mathbb{Q}_p)$ are stably conjugate. The diagonal matrix that conjugates one to the other has coefficients that lie in a quadratic extension. A short calculation shows that the matrices 5.2.1 are not conjugate by a matrix of $SL(2, \mathbb{Q}_p)$.

**5.1. Cocycles.** Let $\gamma_0$ and $\gamma'$ be stably conjugate strongly regular semisimple elements of $G(F)$. We view $\gamma_0$ as a fixed base point and $\gamma'$ as variable. If $\tau \in \mathrm{Gal}(\bar{F}/F)$, then

$$\begin{aligned} &g^{-1}\gamma_0 g &&= \gamma', (\text{with } g \in G(\bar{F}), \gamma_0, \gamma' \in G(F)) \\ &\tau(g)^{-1}\tau(\gamma_0)\tau(g) &&= \tau(\gamma'), \\ &\tau(g)^{-1}\gamma_0\tau(g) &&= g^{-1}\gamma_0 g, \\ &\gamma_0\left(\tau(g)g^{-1}\right) &&= \left(\tau(g)g^{-1}\right)\gamma_0, \\ &\gamma_0 a_\tau &&= a_\tau\gamma_0, \text{ with } a_\tau = \tau(g)g^{-1}. \end{aligned} \tag{5.2.2}$$

The element $a_\tau$ centralizes $\gamma_0$ and hence gives an element of the centralizer $T$. Viewed as a function of $\tau \in \mathrm{Gal}(\bar{F}/F)$, $a_\tau$ satisfies the cocycle relation

$$\tau_1(a_{\tau_2})a_{\tau_1} = a_{\tau_1\tau_2}.$$

It is continuous in the sense that there exists a field extension $F_1/F$ for which $a_\tau = 1$, for all $\tau \in \mathrm{Gal}(\bar{F}/F_1)$. Thus, $a_\tau$ gives a class in

$$H^1(\mathrm{Gal}(\bar{F}/F), T(\bar{F})),$$

which is defined to be the group of all continuous cocycles with values in $T$, modulo the subgroup of all continuous cocycles of the form

$$b_\tau = \tau(t)t^{-1},$$

for some $t \in T(\bar{F})$.

A general calculation of the group $H^1(\mathrm{Gal}(\bar{F}/F), T)$ is achieved by the *Tate-Nakayama* isomorphism. Let $F_1/F$ be a Galois extension that splits the Cartan subgroup $T$.

THEOREM 5.3 (Tate-Nakayama isomorphism [**27**]). *The cohomology group*

$$H^1(\mathrm{Gal}(\bar{F}/F), T)$$

*is isomorphic to the quotient of the group*

$$\{u \in X_* \mid \sum_{\tau \in \mathrm{Gal}(F_1/F)} \tau u = 0\}$$

*by the subgroup generated by the set*

$$\{u \in X_* \mid \exists \tau \in \mathrm{Gal}(F_1/F)\ \exists v \in X_*.\ u = \tau v - v\}.$$

EXAMPLE 5.4. Let $T = U_E(1)$ (the torus that made an appearance earlier as an endoscopic group of $SL(2)$). As was shown above, the group of cocharacters can be identified with $\mathbb{Z}$. The splitting field of $T$ is the quadratic extension field $E$. The nontrivial element $\tau \in \mathrm{Gal}(E/F)$ acts by reflection on $X_* \cong \mathbb{Z}$: $\tau(u) = -u$. By the Tate-Nakayama isomorphism, the group $H^1(\mathrm{Gal}(\bar{F}/F), U_E(1))$ is isomorphic to

$$\{u \in \mathbb{Z} \mid u + \tau u = 0\}/\{u \in \mathbb{Z} \mid \exists v.\ u = \tau v - v\} = \mathbb{Z}/2\mathbb{Z}.$$

Let $H$ be an unramified endoscopic group of $G$. Suppose that $T_H$ is a Cartan subgroup of $H$. Let $T_G$ be an isomorphic Cartan subgroup in $G$. The data defining $H$ includes the existence of an element $s \in \mathrm{Hom}(X_*, \mathbb{C}^\times)$; that is, a character of the abelian group $X_*$. Fix one such character $s$. We can restrict this character to get a character of

$$\{u \in X_* \mid \sum_{\tau \in \mathrm{Gal}(F_1/F)} \tau u = 0\}.$$

It can be shown that the character $s$ is trivial on

$$\{u \in X_* \mid \exists \tau \in \mathrm{Gal}(F_1/F)\ \exists v \in X_*.\ u = \tau v - v\}.$$

Thus, by the Tate-Nakayama isomorphism, the character $s$ determines a character $\kappa$ of the cohomology group

$$H^1(\mathrm{Gal}(\bar{F}/F), T).$$

In this way, each cocycle $a_\tau$ gives a complex constant $\kappa(a_\tau) \in \mathbb{C}^\times$.

EXAMPLE 5.5. The element $s \in \mathbb{C}^\times$ giving the endoscopic group $H = U_E(1)$ of $SL(2)$ is $s = -1$, which may be identified with the character $n \mapsto (-1)^n$ of $\mathbb{Z}$. This gives the nontrivial character $\kappa$ of

$$H^1(\mathrm{Gal}(\bar{F}/F), U_E(1)) \cong \mathbb{Z}/2\mathbb{Z}.$$

## 6. Statement of the Fundamental Lemma

**6.1. Context.** Let $G$ be an unramified connected reductive group over $F$. Let $H$ be an unramified endoscopic group of $G$. Let $\gamma \in H(F)$ be a strongly regular semisimple element. Let $T_H = C_H(\gamma)$, and let $T_G$ be a Cartan subgroup of $G$ that is isomorphic to it. More details will be given below about how to choose $T_G$. The choice of $T_G$ matters! Let $\gamma \in T_H(F)$ map to $\gamma_0 \in T_G(F)$ under this isomorphism.

By construction, $\gamma_0$ is semisimple. However, as $G$ may have more roots than $H$, it is possible for $\gamma_0$ to be singular, even when $\gamma$ is strongly regular. If $\gamma \in H(F)$ is a strongly regular semisimple element with the property that $\gamma_0$ is also strongly regular, then we will call $\gamma$ a *strongly $G$-regular* element of $H(F)$.

If $\gamma'$ is stably conjugate to $\gamma_0$ with cocycle $a_\tau$, then $s \in \mathrm{Hom}(X_*, \mathbb{C}^\times)$ gives $\kappa(a_\tau) \in \mathbb{C}^\times$.

Let $K_G$ and $K_H$ be hyperspecial maximal compact subgroups of $G$ and $H$. Let $\chi_{G,K}$ and $\chi_{H,K}$ be the characteristic functions of these hyperspecial subgroups. Set

$$\Lambda_{G,H}(\gamma) = \tag{6.0.1}$$
$$\prod_{\alpha \in \Phi_G} |\alpha(\gamma_0) - 1|^{1/2} \left[\frac{\mathrm{vol}(K_T, dt)}{\mathrm{vol}(K, dg)}\right] \sum_{\gamma' \sim \gamma_0} \kappa(a_\tau) \int_{C_G(\gamma', F)\backslash G(F)} \chi_{G,K}(g^{-1}\gamma' g) \frac{dg}{dt'}.$$

The set of roots $\Phi_G$ are taken to be those relative to $T_G$. The sum runs over all stable conjugates $\gamma'$ of $\gamma_0$, up to conjugacy. This is a finite sum. The group $K_T$ is defined to be the maximal compact subgroup of $T_G$. Equation 6.0.2 is a finite linear combination of orbital integrals (that is, integrals over conjugacy classes in the group with respect to an invariant measure). The Haar measures $dt'$ on $C_G(\gamma', F)$ and $dt$ on $T_G(F)$ are chosen so that stable conjugacy between the two groups is measure preserving. This particular linear combination of integrals is called a $\kappa$-orbital integral because of the term $\kappa(a_\tau)$ that gives the coefficients of the linear combination. Note that the integration takes place in the group $G$, and yet the parameter $\gamma$ is an element of $H(F)$.

The volume terms $\mathrm{vol}(K, dg)$ and $\mathrm{vol}(K_T, dt)$ serve no purpose other than to make the entire expression independent of the choice of Haar measures $dg$ and $dt$, which are only defined up to a scalar multiple.

We can form an analogous linear combination of orbital integrals on the group $H$. Set

$$\text{(6.0.2)} \qquad \Lambda_H^{st}(\gamma) = \left(\prod_{\alpha \in \Phi_H} |\alpha(\gamma) - 1|^{1/2}\right) \left[\frac{\mathrm{vol}(K_T, dt)}{\mathrm{vol}(K_H, dh)}\right] \sum_{\gamma' \sim \gamma} \int_{C_H(\gamma', F)\backslash H(F)} \chi_{H,K}(h^{-1}\gamma' h) \frac{dh}{dt'}.$$

This linear combination of integrals is like $\Lambda_{G,H}(\gamma)$, except that $H$ replaces $G$, $K_H$ replaces $K_G$, $\Phi_H$ (taken relative to $T_H$) replaces $\Phi_G$, and so forth. Also, the factor $\kappa(a_\tau)$ has been dropped. The linear combination of Equation 6.0.3 is called a stable orbital integral, because it extends over all stable conjugates of the element $\gamma$ without the factor $\kappa$. The superscript *st* in the notation is for 'stable.'

CONJECTURE 6.1. *(The fundamental lemma) For every $\gamma \in H(F)$ that is strongly $G$-regular semisimple,*

$$\Lambda_{G,H}(\gamma) = \Lambda_H^{st}(\gamma).$$

REMARK 6.2. There have been serious efforts over the past twenty years to prove the fundamental lemma. These efforts have not yet led to a proof. Thus, the fundamental lemma is not a lemma; it is a conjecture with a misleading name. Its name leads one to speculate that the authors of the conjecture may have severely underestimated the difficulty of the conjecture.

REMARK 6.3. Special cases of the fundamental lemma have been proved. The case $G = SL(n)$ was proved by Waldspurger [**28**]. Building on the work of [**5**], Laumon has proved that the fundamental lemma for $G = U(n)$ follows from a purity conjecture [**21**]. The fundamental lemma has not been proved for any other general families of groups. The fundamental lemma has been proved for some groups $G$ of small rank, such as $SU(3)$ and $Sp(4)$. See [**2**], [**7**], [**10**].

**6.2. The significance of the fundamental lemma.** The Langlands program predicts correspondences $\pi \leftrightarrow \pi'$ between the representation theory of different reductive groups. There is a local program for the representation theory of reductive groups over locally compact fields, and a global program for *automorphic* representations of reductive groups over the adele rings of global fields.

The Arthur-Selberg trace formula has emerged as a powerful tool in the Langlands program. In crude terms, one side of the trace formula contains terms related to the characters of automorphic representations. The other side contains terms such as orbital integrals. *Thanks to the trace formula, identities between orbital integrals on different groups imply identities between the representations of the two groups.*

It is possible to work backwards: from an analysis of the terms in the trace formula and a precise conjecture in representation theory, it is possible to make precise conjectures about identities of orbital integrals. The most basic identity that appears in this way is the fundamental lemma, articulated above.

The proofs of many major theorems in automorphic representation theory depend in one way or another on the proof of a fundamental lemma. For example, the proof of Fermat's Last Theorem depends on Base Change for $GL(2)$, which in turn depends on the fundamental lemma for cyclic base change [**17**]. The proof of the local Langlands conjecture for $GL(n)$ depends on automorphic induction, which in turn depends on the fundamental lemma for $SL(n)$ [**11**], [**12**], [**28**]. Properties of the zeta function of Picard modular varieties depend on the fundamental lemma for $U(3)$ [**26**], [**2**]. Normally, the dependence of a major theorem on a particular lemma would not be noteworthy. It is only because the fundamental lemma has not been proved in general, and because the lack of proof has become a serious impediment to progress in the field, that the conjecture has become the subject of increased scrutiny.

## 7. Reductions

To give a trivial example of the fundamental lemma, if $\gamma$ and $\gamma_0$ and their stable conjugates are not in any compact subgroup, then

$$\chi_{G,K}(g^{-1}\gamma' g) = 0 \text{ and } \chi_{H,K}(h^{-1}\gamma' h) = 0$$

so that both $\Lambda_{G,H}(\gamma)$ and $\Lambda_H^{st}(\gamma)$ are zero. Thus, the fundamental lemma holds for trivial reasons for such $\gamma$.

**7.1. Topological Jordan decomposition.** A somewhat less trivial reduction of the problem is provided by the topological Jordan decomposition. Suppose that $\gamma$ lies in a compact subgroup. It can be written uniquely as a product

$$\gamma = \gamma_s\gamma_u = \gamma_u\gamma_s,$$

where $\gamma_s$ has finite order, of order prime to the residue field characteristic $p$, and $\gamma_u$ is topologically unipotent. That is,

$$\lim_{n\to\infty} \gamma_u^{p^n} = 1.$$

The limit is with respect to the $p$-adic topology. A special case of the topological Jordan decomposition $\gamma \in O_F^\times \subset \mathbb{G}_m(F)$ is treated in [**13**, p20]. In that case, $\gamma_s$ is defined by the formula

$$\gamma_s = \lim_{n\to\infty} \gamma^{q^n}.$$

Let $\gamma$, $\gamma_0$, and $\gamma'$ be chosen as in Section 6.1. Each of these elements has a topological Jordan decomposition. Let $G_s = C_G(\gamma_{0s})$ and $H_s = C_H(\gamma_s)$. It turns out that $G_s$ is an unramified reductive group with unramified endoscopic group $H_s$. Descent for orbital integrals gives the formulas [**20**] [**8**]

$$\begin{aligned} \Lambda_{G,H}(\gamma) &= \Lambda_{G_s,H_s}(\gamma_u) \\ \Lambda_H^{st}(\gamma) &= \Lambda_{H_s}^{st}(\gamma_u). \end{aligned}$$

This reduces the fundamental lemma to the case that $\gamma$ is a topologically unipotent element.

**7.2. Lie algebras.** It is known (at least when the $p$-adic field $F$ has characteristic zero), that the fundamental lemma holds for fields of arbitrary residual characteristic provided that it holds when the $p$-adic field has sufficiently large residual characteristic [**9**]. Thus, if we are willing to restrict our attention to fields of characteristic zero, we may assume that the residual characteristic of $F$ is large. In fact, in our discussion of a reduction to Lie algebras in this section, we simply assume that the characteristic of $F$ is zero.

A second reduction is based on Waldspurger's homogeneity results for classical groups. (Homogeneity results have since been reworked and extended to arbitrary reductive groups by DeBacker, again assuming mild restrictions on $G$ and $F$.)

When the residual characteristic is sufficiently large, there is an exponential map from the Lie algebra to the group that has every topologically unipotent element in its image. Write

$$\gamma_u = \exp(X),$$

for some element $X$ in the Lie algebra. We may then consider the behavior of orbital integrals along the curve $\exp(\lambda^2 X)$. A difficult result of Waldspurger for classical groups states that if $|\lambda| \leq 1$, then

$$\begin{aligned} \Lambda_{G,H}(\exp(\lambda^2 X)) &= \sum a_i |\lambda|^i \\ \Lambda_H^{st}(\exp(\lambda^2 X)) &= \sum b_i |\lambda|^i; \end{aligned}$$

that is, both sides of the fundamental lemma identity are polynomials in $|\lambda|$. If a polynomial identity holds when $|\lambda| < \epsilon$ for some $\epsilon > 0$, then it holds for all $|\lambda| \leq 1$. In particular, it holds at $\gamma_u$ for $\lambda = 1$. The polynomial growth of orbital integrals makes it possible to prove the fundamental lemma in a small neighborhood of the identity element, and then conclude that it holds in general. In this manner, the fundamental lemma can be reduced to a conjectural identity in the Lie algebra.

## 8. The problem of base points

The fundamental lemma was formulated above with one omission: we never made precise how to fix an isomorphism $T_H \leftrightarrow T_G$ between Cartan subgroups in $H$ and $G$. Such isomorphisms exist, because the two Cartan subgroups have the same root data. But the statement of the fundamental lemma is sensitive to how an isomorphism is selected between $T_H$ and a Cartan subgroup of $G$. If we change the isomorphism, we change the $\kappa$-orbital integral by a root of unity $\zeta \in \mathbb{C}^\times$. The correctly chosen isomorphism will depend on the element $\gamma \in H(F)$.

The ambiguity of the isomorphism was removed by Langlands and Shelstad in [**19**]. They define a *transfer factor* $\Delta(\gamma_H, \gamma_G)$, which is a complex valued function on $H(F) \times G(F)$. The transfer factor can be defined to have the property that it is zero unless $\gamma_H \in H(F)$ is strongly regular semisimple, $\gamma_G \in G(F)$ is strongly regular semisimple, and there exists an isomorphism (preserving character groups) from the centralizer of $\gamma_H$ to the centralizer of $\gamma_G$. There exists $\gamma_0 \in G(F)$ such that

$$\Delta(\gamma_H, \gamma_0) = 1. \tag{8.0.1}$$

The correct formulation of the fundamental lemma is to pick the base point $\gamma_0 \in G(F)$ so that Condition 8.0.1 holds.

For classical groups, Waldspurger gives a simplified formula for the transfer factor $\Delta$ in [**31**]. Furthermore, because of the reduction of the fundamental lemma to the Lie algebra (Section 7.2), the transfer factor may be expressed as a function on the Lie algebras of $G$ and $H$, rather than as a function on the group.

**8.1. Base points for unitary groups.** More recently, Laumon (while working on the fundamental lemma for unitary groups) observed a similarity between Waldspurger's simplified formula for the transfer factor and the explicit formula for differents that is found in [**27**]. In this way, Laumon found a simple description of the matching condition $\gamma \leftrightarrow \gamma_0$ implicit in the statement of the fundamental lemma.

## 9. Geometric Reformulations of the Fundamental Lemma

From early on, those trying to prove the fundamental lemma have sought geometric interpretations of the identities of orbital integrals. Initially these geometric interpretations were rather crude. In the hands of Goresky, Kottwitz, MacPherson, and Laumon these geometric interpretations have become increasingly sophisticated [**5**], [**6**], [**21**], [**22**].

This paper is intended to give an introduction to the fundamental lemma, and the papers giving a geometric interpretation of the fundamental lemma do not qualify as introductory material. In this section, we will be content to describe the geometric interpretation in broad terms.

**9.1. Old-style geometric interpretations: buildings.** We begin with a geometric interpretation of the fundamental lemma that was popular in the late seventies and early eighties. It was eventually discarded in favor of other approaches when the combinatorial difficulties became too great.

This approach is to use the geometry of the Bruhat-Tits building to understand orbital integrals. We illustrate the approach with the group $G = SL(2)$. The term $\chi_{G,K}(g^{-1}\gamma' g)$ that appears in the fundamental lemma can be manipulated as follows:

$$\begin{aligned} \chi_{G,K}(g^{-1}\gamma' g) \neq 0 \quad & \Leftrightarrow g^{-1}\gamma' g \in K \\ & \Leftrightarrow \gamma' g \in gK \\ & \Leftrightarrow \gamma'(gK) = (gK) \\ & \Leftrightarrow gK \text{ is a fixed point of } \gamma' \text{ on } G(F)/K. \end{aligned}$$

The set $G(F)/K$ is in bijective correspondence with a set of vertices in the Bruhat-Tits building of $SL(2)$. Thus, we may interpret the orbital integral geometrically as the number of fixed points of $\gamma'$ in the building that are vertices of a given type.

Under this interpretation, it is possible to use counting arguments to obtain explicit formulas for orbital integrals as a function of $\gamma'$. In this way, the fundamental lemma was directly verified for a few groups of small rank such as $SL(2)$ and $U(3)$.

**9.2. Affine Grassmannians.** Until the end of Section 9, let $F = k((t))$, a field of formal Laurent series. Except for the discussion of the results of Kazhdan and Lusztig, the field $k$ will be taken to be a finite field: $k = \mathbb{F}_q$.

In 1988, Kazhdan and Lusztig showed that if $F = \mathbb{C}((t))$, then $G(F)/K$ can be identified with the points of an ind-scheme (that is, an inductive limit of schemes) [**15**]. This ind-scheme is called the affine Grassmannian. The set of fixed points of an element $\gamma$ can be identified with the set of points of a scheme over $\mathbb{C}$, known as the affine Springer fiber. The corresponding construction over $\mathbb{F}_q((t))$ is mentioned briefly in the final paragraphs of their paper. Rather than counting fixed points in the building, orbital integrals can be computed by counting the number of points on a scheme over $\mathbb{F}_q$.

Based on a description of orbital integrals as the number of points on schemes over finite fields, Kottwitz, Goresky, and MacPherson give a geometrical formulation of the fundamental lemma. Furthermore, by making a thorough investigation of the equivariant cohomology of these schemes, they prove the geometrical conjecture when $\gamma$ comes from an unramified Cartan subgroup [**5**].

**9.3. Geometric interpretations.** Each of the terms in the fundamental lemma has a nice geometric interpretation. Let us give a brief description of the geometrical counterpart of each term in the fundamental lemma. We work with the unitary group, so that we may include various insights of Laumon.

The geometrical counterparts of cosets $gK$ are self-dual lattices in a vector space $V$ over $F$.

The counterpart of the support set, $\mathrm{SUP} = \{g \mid g^{-1}\gamma g \in K\}$, is the affine Springer fiber $X_\gamma$.

The counterpart of the integral of the support set SUP over $G$ is counting points on the scheme $X_\gamma$. The integral over all of $G$ diverges and the number of fixed points on the scheme is infinite. For that reason the orbital integral is an integral over $T\backslash G$, where $T$ is the centralizer of $\gamma$, rather than over all of $G$.

The counterpart of the integral over $T\backslash G$ is counting points on a quotient space $Z_\gamma = X_\gamma/\mathbb{Z}^\ell$. (There is a free action of a group $\mathbb{Z}^\ell$ on $X_\gamma$, and $Z_\gamma$ is the quotient.)

The geometric counterpart of $\kappa(a_\tau)$ is somewhat more involved. For elliptic endoscopic groups of unitary groups $\kappa$ has order 2. The character $\kappa$ has the form

$$\kappa : H^1(\mathrm{Gal}(\bar{F}/F), T) \cong (\mathbb{Z}/2\mathbb{Z})^\ell \to \{\pm 1\}.$$

The character $\kappa$ pulls back to a character of $\mathbb{Z}^\ell$. The rational points of $X_\gamma$ are identified with self-dual lattices: $A^\perp = A$. The points of the quotient space $Z_\gamma$ are lattices that are self-dual modulo the group action: $A^\perp = \lambda \cdot A$, for some $\lambda \in \mathbb{Z}^\ell$. The character $\kappa$ then partitions the points of $Z_\gamma$ into two sets, depending on the sign of $\kappa(\lambda)$:

$$Z_\gamma^\pm = \{A \mid A^\perp = \lambda A; \quad \kappa(\lambda) = \pm 1\}.$$

(In a more sophisticated treatment of $\kappa(a_\tau)$, it gives rise to a local system on $Z_\gamma$; and counting points on varieties gives way to Grothendieck's trace formula.)

The counterpart of the $\kappa$-orbital integral $\Lambda_{G,H}(\gamma)$ is the number

$$\#Z_\gamma^+ - \#Z_\gamma^-.$$

The counterpart of the stable-orbital integral $\Lambda_H^{st}(\gamma)$ is the number

$$\#Z_\gamma^{H,st}$$

for a corresponding variety constructed from the endoscopic group.

The factors $\prod_\Phi |\alpha(\gamma) - 1|^{1/2}$ that appear on the two sides of the fundamental lemma can be combined into a single term

$$\prod_{\alpha \in \Phi_G \setminus \Phi_H} |\alpha(\gamma) - 1|^{1/2}.$$

This has the form $q^{-d}$ for some value $d = d(\gamma)$. The factor $q^{-d}$ has been interpreted in various ways. We mention that [**24**] interprets $q^d$ as the points on an affine space of dimension $d$. That paper expresses the hope that it might be possible to find an embedding $Z_\gamma^- \to Z_\gamma^+$ such that the complement of the embedded $Z_\gamma^-$ in $Z_\gamma^+$ is a rank $d$ fiber bundle over $Z_\gamma^{H,st}$. The realization of this hope would give an entirely geometric interpretation of the fundamental lemma. Laumon and Rapoport found that this construction works over $\mathbb{F}_{q^2}((t))$, but not over $\mathbb{F}_q((t))$. In more recent work of Laumon, the constant $d$ is interpreted geometrically as the intersection multiplicity of two singular curves.

**9.4. Compactified Jacobians.** Laumon, in the case of unitary groups, has made the splendid discovery that the orbital integrals – as they appear in the fundamental lemma – count points on the compactification of the Jacobians of a singular curve associated with the semisimple element $\gamma$. (In fact, $Z_\gamma$ is homeomorphic to and can be replaced with the compactification of a Jacobian.) Thus, the fundamental lemma may be reformulated as a relation between the compactified Jacobians of these curves. By showing that the singular curve for the endoscopic group $H$ is a perturbation of the singular curve for the group $G$, he is able to relate the compactified Jacobians of the two curves, and prove the fundamental lemma for unitary groups (assuming a purity hypothesis related to the cohomology of the schemes).

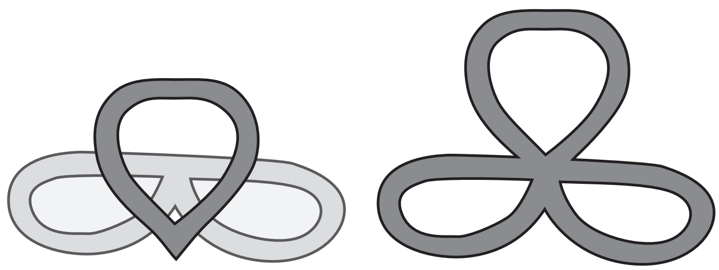


FIGURE 1. The singular curve on the left can be deformed into the singular curve on the right by pulling up on the center ring. The curve on the left controls $\Lambda_H^{st}(\gamma)$, and the curve on the right controls $\Lambda_{G,H}(\gamma)$. This deformation relating the two curves is a key part of Laumon's work on the fundamental lemma for unitary groups.

The origin of the curve $C$ is the following. The ring $O_F[\gamma]$ is the completion at a point of the local ring of a curve $C$. In the interpretation in terms of Jacobians, the self-dual lattices $A^\perp = A$ that appear in the geometric interpretation above are replaced with $O_C$-modules, where $O_C$ is the structure sheaf of $C$.

The audio recording of Laumon's lecture at the Fields Institute on this research is highly recommended [**23**].

**9.5. Final remarks.**

REMARK 9.1. The fundamental lemma is an open-ended problem, in the sense that as researchers develop new trace formulas (the symmetric space trace formula [**14**], the twisted trace formula [**16**], and so forth) and as they compare trace formulas for different groups, it will be necessary to formulate and prove generalized versions of the fundamental lemma. The version of the fundamental lemma stated in this paper should be viewed as a template that should be adapted according to an evolving context.

Remark 9.2. The methods of Goresky, Kottwitz, MacPherson, and Laumon are limited to fields of positive characteristic. This may at first seem to be a limitation of their method. However, there are ideas about how to use motivic integration to lift their results from positive characteristic to characteristic zero (see [**3**]). Waldspurger also has results about lifting to characteristic zero that were presented at the Labesse conference, but I have not seen a preprint [**32**].

Remark 9.3. In some cases, it is now known how to deduce stronger forms of the fundamental lemma from weaker versions. For example, it is known how to go from the characteristic function of the hyperspecial maximal compact groups to the full Hecke algebra [**9**]. A descent argument replaces twisted orbital integrals by ordinary orbital integrals. However, relations between weighted orbital integrals remain a serious challenge.

Remark 9.4. There has been much research on the fundamental lemma that has not been discussed in detail in this paper, including other forms of the fundamental lemma. For just one example, see [**25**] for the fundamental lemma of Jacquet and Ye. Other helpful references include [**18**] and [**30**].

## References

1. J. Arthur, A stable trace formula. I. General expansions. J. Inst. Math. Jussieu 1 (2002), no. 2, 175–277.
2. D. Blasius and J. D. Rogawski, Fundamental lemmas for U(3) and related groups. The zeta functions of Picard modular surfaces, 363–394, Univ. Montr?al, Montreal, QC, 1992.
3. C. Cunningham, T. C. Hales, Good Orbital Integrals., math.RT/0311353, preprint.
4. S. DeBacker, Homogeneity results for invariant distributions of a reductive $p$-adic group. Ann. Sci. École Norm. Sup. (4) 35 (2002), no. 3, 391–422.
5. M. Goresky, R. Kottwitz, R. MacPherson, Homology of affine Springer fibers in the unramified case, math.RT/0305141
6. M. Goresky, R. Kottwitz, R. MacPherson, Purity of Equivalued Affine Springer Fibers, math.RT/0305144
7. T. C. Hales, Orbital integrals on $U(3)$, The Zeta Function of Picard Modular Surfaces, Les Publications CRM, (R.P. Langlands and D. Ramakrishnan, eds.), 1992. MR 93d:22020
8. T.C. Hales, A Simple Definition of Transfer Factors for Unramified Groups, Contemporary Math., 145 (1993) 109-143
9. T.C. Hales, On the Fundamental Lemma for Standard Endoscopy: Reduction to Unit Elements, Canad. J. Math. 47, 974-994, 1995.
10. T. C. Hales, The fundamental lemma for Sp(4). Proc. Amer. Math. Soc. 125 (1997), no. 1, 301–308.
11. M. Harris and R. Taylor, The geometry and cohomology of some simple Shimura varieties, Annals of Math. Studies 151, PUP 2001.
12. G. Henniart and R. Herb, Automorphic induction for $GL(n)$ (over local nonarchimedean fields, Duke Math. J. 78 (1995), 131-192.
13. K. Iwasawa, Local class field theory. Oxford Science Publications. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1986. viii+155 pp. ISBN: 0-19-504030-9
14. H. Jacquet, K. F. Lai, S. Rallis, A trace formula for symmetric spaces. Duke Math. J. 70 (1993), no. 2, 305–372.
15. D. Kazhdan and G. Lusztig, Fixed point varieties on affine flag manifolds. Israel J. Math. 62 (1988), no. 2, 129–168.
16. R. Kottwitz, D. Shelstad, Foundations of twisted endoscopy. Astérisque No. 255 (1999), vi+190 pp.

17. R. P. Langlands, Base change for GL(2). Annals of Mathematics Studies, 96. Princeton University Press, Princeton, 1980.
18. R. P. Langlands, Les débuts d'une formule des traces stable. Pub. Math. de l'Univ. Paris VII, 13. Paris, 1983.
19. R. P. Langlands and D. Shelstad, On the definition of transfer factors, Math. Ann. 278 (1987), 219-271.
20. R. P. Langlands and D. Shelstad, Descent for transfer factors. The Grothendieck Festschrift, Vol. II, 485–563, Progr. Math., 87, Birkhauser Boston, 1990.
21. G. Laumon, Sur le lemme fondamental pour les groupes unitaires, math.AG/0212245G.
22. G. Laumon, Fibres de Springer et jacobiennes compactifiées., math.AG/0204109
23. G. Laumon, On the fundamental lemma for unitary groups, lecture The Fields Institute, March 4, 2003.
`http://www.fields.utoronto.ca/audio/02-03/shimura/laumon/`
24. G. Laumon and M. Rapoport, A geometric approach to the fundamental lemma for unitary groups, alg-geom/9711021.
25. B. C. Ngô, Faisceaux Pervers, homomorphisme de changement de base et lemme fondamental de Jacquet et Ye, Ann. scient. Éc. Norm. Sup. 4e série, 32, 1999, 619–679.
26. J. Rogawski, Automorphic representations of unitary groups in three variables. Annals of Mathematics Studies, 123. Princeton University Press, Princeton, NJ, 1990.
27. J.-P. Serre, Local Fields, Springer-Verlag, 1979 (GTM 67).
28. J.-L. Waldspurger, Sur les intégrales orbitales tordues pour les groupes linéaires: un lemme fondamental. Canad. J. Math. 43 (1991), no. 4, 852–896.
29. J.-L. Waldspurger, Homogénéité de certaines distributions sur les groupes $p$-adiques. Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 25–72.
30. J.-L. Waldspurger, Comparaison d'intégrales orbitales pour des groupes $p$-adiques. Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 807–816, Birkhauser, Basel, 1995.
31. J.-L. Waldspurger, Intégrales orbitales nilpotentes et endoscopie pour les groupes classiques non ramifiés, Astérisque, vol. 269, SMF, 2001.
32. J.-L. Waldspurger, *Endoscopie et changement de caractéristique*, lecture at Institut Henri Poincaré, Conference in honor of Jean-Pierre Labesse, Sept 26, 2003, http://www.math.jussieu.fr/congres-labesse/programme.html.

Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260
*E-mail address*: `hales@pitt.edu`

Clay Mathematics Proceedings
Volume 4, 2005

# Notes on the Generalized Ramanujan Conjectures

Peter Sarnak

Contents



## 1. $GL_n$

Ramanujan's original conjecture is concerned with the estimation of Fourier coefficients of the weight 12 holomorphic cusp form $\Delta$ for $SL(2,\mathbb{Z})$ on the upper half plane $\mathbb{H}$. The conjecture may be reformulated in terms of the size of the eigenvalues of the corresponding Hecke operators or equivalently in terms of the local representations which are components of the automorphic representation associated with $\Delta$. This spectral reformulation of the problem of estimation of Fourier coefficients (or more generally periods of automorphic forms) is not a general feature. For example, the Fourier coefficients of Siegel modular forms in several variables carry more information than just the eigenvalues of the Hecke operators. Another example is that of half integral weight cusp forms on $\mathbb{H}$ where the issue of the size of the Fourier coefficients is equivalent to special instances of the Lindelöf Hypothesis for automorphic $L$-functions (see [Wal], [I-S]). As such, the general problem of estimation of Fourier coefficients appears to lie deeper (or rather farther out of reach at the present time). In these notes we discuss the spectral or representation theoretic generalizations of the Ramanujan Conjectures ($GRC$ for short). While we are still far from being able to establish the full Conjectures in general, the approximations to the conjectures that have been proven suffice for a number of the intended applications.

We begin with some general comments. In view of Langlands Functoriality Conjectures (see [A1]) all automorphic forms should be encoded in the $GL_n$ automorphic spectrum. Moreover, Arthur's recent conjectural description of the discrete spectrum for the decomposition of a general group [A2],[A3] has the effect of reducing the study of the spectrum of a classical group, for example, to that of $GL_n$.

From this as well as the point of view of $L$-functions, $GL_n$ plays a special role. Let $F$ be a number field, $\mathbb{A}_F$ its ring of adèles, $v$ a place of $F$ (archimedian or finite) and $F_v$ the corresponding local field. Let $GL_n$ be the group of $n \times n$ invertible matrices and $GL_n(\mathbb{A}_F)$, $GL_n(F)$, $GL_n(F_v) \cdots$ be the corresponding group with entries in the indicated ring. The abelian case $GL_1$, is well understood and is a guide (though it is way too simplistic) to the general case. The constituents of the decomposition of functions on $GL_1(F)\backslash GL_1(\mathbb{A}_F)$ or what is the same, the characters of $F^*\backslash\mathbb{A}_F^*$, can be described in terms of class field theory. More precisely, if $W_F$ is Weil's extension of the Galois group $\mathrm{Gal}(\bar{F}/F)$ then the 1-dimensional representations of $F^*\backslash\mathbb{A}_F^*$ correspond naturally to the 1-dimensional representations of $W_F$ (see [Ta]). As Langlands has pointed out [Lang1] it would be very, nice for many reasons, to have an extended group $L_F$ whose $n$-dimensional representations would correspond naturally to the automorphic forms on $GL_n$. The basic such forms are constituents of the decomposition of the regular representation of $GL_n(\mathbb{A}_F)$ on $L^2(Z(\mathbb{A}_F)GL_n(F)\backslash GL_n(\mathbb{A}_F), w)$. Here $Z$ is the center of $GL_n$ and $w$ is a unitary character of $Z(\mathbb{A}_F)/Z(F)$. In more detail, the $L^2$ space consists of functions $f : GL_n(\mathbb{A}_F) \longrightarrow \mathbb{C}$ satisfying $f(\gamma z g) = w(z)f(g)$ for $\gamma \in GL_n(F)$, $z \in Z(\mathbb{A}_F)$ and

$$(1) \qquad \int_{Z(\mathbb{A}_F)GL_n(F)\backslash GL_n(\mathbb{A}_F)} |f(g)|^2\, dg \,<\, \infty\,.$$

Notwithstanding the success by Harris-Taylor and Henniart [H-T] giving a description in the local case of the representations of $GL_n(F_v)$ in terms of $n$-dimensional representations of the Deligne-Weil group $W_F'$, or the work of Lafforgue in the case of $GL_n(F)$ where $F$ is a function field over a finite field, it is difficult to imagine a direct definition of $L_F$ in the number field case. My reason for saying this is that $L_F$ would have to give, through its finite dimensional representations, an independent description of the general Maass cusp form for say $GL_2(\mathbb{A}_{\mathbb{Q}})$ (see [Sa] for a recent discussion of these). These are eigenfunctions of elliptic operators on infinite dimensional spaces with presumably highly transcendental eigenvalues. Arthur in his definition [A3] of $L_F$ gets around this difficulty by using among other things the $GL_n$ cusp forms as building blocks for the construction of the group. With this done, he goes on to describe a much more precise form of the general functoriality conjectures.

We turn to $GL_n$ and a description of the generalized Ramanujan Conjectures. According to the general theory of Eisenstein series $L^2(Z(\mathbb{A}_F)GL_n(F)\backslash GL_n(\mathbb{A}_F), w)$ decomposes into a discrete part and a continuous part. The discrete part coming from residues of Eisenstein series, as well as the continuous part coming from Eisenstein series, are described explicitly in [M-W]. They are given in terms of the discrete spectrum on $GL_m$, $m < n$. This leaves the cuspidal spectrum as the fundamental part. It is defined as follows:

$$(2) \qquad L^2_{\mathrm{cusp}}(Z(\mathbb{A}_F)GL_n(F)\backslash GL_n(\mathbb{A}_F), w)$$

$$= \{f \text{ satisfying (1) and } \int_{N(F)\backslash N(\mathbb{A}_F)} f(ng)dn \,=\, 0$$

$$\text{for all unipotent radicals } N \text{ of proper parabolic subgroups } P \text{ of } G(F)\}$$

The decomposition into irreducibles of $GL_n(\mathbb{A}_F)$ on $L^2_{\text{cusp}}$ is discrete and any irreducible constituent $\pi$ thereof is called an automorphic cusp form (or representation). Now such a $\pi$ is a tensor product $\pi = \underset{v}{\otimes}\pi_v$, where $\pi_v$ is an irreducible unitary representation of the local group $GL_n(F_v)$. The problem is to describe or understand which $\pi_v$'s come up in this way. For almost all $v$, $\pi_v$ is unramified, that is, $\pi_v$ has a nonzero $K_v$ invariant vector, where $K_v$ is a maximal compact subgroup of $GL_n(F_v)$. If $v$ is finite then $K_v = GL_n(O(F_v))$, $O(F_v)$ being the ring of integers at $v$. Such "spherical" $\pi_v$ can be described using the theory of spherical functions (Harish-Chandra, Satake) or better still in terms of the Langlands dual group ${}^LG$. For $G = GL_n$, ${}^LG = GL(n,\mathbb{C})$ (or rather the connected component of ${}^LG$ is $GL(n,\mathbb{C})$ but for our purposes here this will suffice) and an unramified representation $\pi_v$ is parameterized by a semi-simple conjugacy class

$$\alpha(\pi_v) = \begin{bmatrix} \alpha_1(\pi_v) & & & & \\ & \ddots & & 0 & \\ & & \ddots & & \\ & 0 & & \ddots & \\ & & & & \alpha_n(\pi_v) \end{bmatrix} \in {}^LG \tag{3}$$

as follows: Let $B$ be the subgroup of upper triangular matrices in $GL_n$. For $b \in B(F_v)$ and $\mu_1(v), \dots, \mu_n(v)$ in $\mathbb{C}$ let $\chi_\mu$ be the character of $B(F_v)$,

$$\chi_\mu(b) = |b_{11}|_v^{\mu_1} |b_{22}|_v^{\mu_2} \cdots |b_{nn}|_v^{\mu_n} . \tag{4}$$

Then $\psi_\mu = \operatorname{Ind}_{B(F_v)}^{G(F_v)} \chi_\mu$ yields a spherical representation of $G(F_v)$ (the induction is normalized unitarily and at $\mu$'s for which it is reducible we take the spherical constituent). $\psi_\mu$ is equivalent to $\psi_{\mu'}$ with $\mu$ and $\mu'$ considered $\mod \mathbb{Z}2\pi i\big/\log N(v)$ iff $\mu = \sigma\mu'$, where $\sigma$ is a permutation. In this notation $\alpha(\pi_v)$ corresponds to $\psi_{\mu(v)}$ by setting $\alpha_j(\pi_v) = N(v)^{\mu_j(v)}$ for $j = 1, 2, \dots, n$. The trivial representation of $G(F_v)$, or constant spherical function, corresponds to

$$\mu = \left(\frac{n-1}{2}, \frac{n-3}{2} \cdots \frac{1-n}{2}\right) \tag{5}$$

In terms of these parameters the local $L$-function $L(s,\pi_v)$ corresponding to such an unramified $\pi_v$ has a simple definition:

$$\begin{aligned} L(s,\pi_v) &= \det(I - \alpha(\pi_v)\, N(v)^{-s})^{-1} \\ &= \Pi_{j=1}^n (1 - \alpha_j(\pi_v)\, N(v)^{-s})^{-1} \end{aligned} \tag{6}$$

if $v$ is finite, and

$$L(s,\pi_v) = \Pi_{j=1}^n \Gamma_v(s - \mu_j(\pi_v)) , \tag{7}$$

with $\Gamma_v(s) = \pi^{-s/2}\,\Gamma\left(\frac{s}{2}\right)$ if $F_v \cong \mathbb{R}$ and $\Gamma_v(s) = (2\pi)^{-s}\,\Gamma(s)$ if $F_v \cong \mathbb{C}$.

More generally if $\rho : {}^LG \longrightarrow GL(\nu)$ is a representation of ${}^LG$ then the local $L$-function is defined by

$$L(s,\pi_v,\rho) = \det(I - \rho(\alpha(\pi_v))N(v)^{-s})^{-1} . \tag{8}$$

We digress and discuss some local harmonic analysis for more general groups. Let $G(F_v)$ be a reductive group defined over a local field $F_v$. We denote by $\widehat{G(F_v)}$ the unitary dual of $G(F_v)$, that is, the set of irreducible unitary representations of $G(F_v)$ up to equivalence. $G(F_v)$ has a natural topology, the Fell topology, coming from convergence of matrix coefficients on compact subsets of $G(F_v)$. Of particular interest is the tempered subset of $\widehat{G(F_v)}$, which we denote by $\widehat{G(F_v)}_{\text{temp}}$. These are the representations which occur weakly (see [Di]) in the decomposition of the regular representation of $G(F_v)$ on $L^2(G(F_v))$ or what is the same thing $\operatorname{Ind}_H^{G(F_v)} 1$ where $H = \{e\}$. If $G(F_v)$ is semi-simple then the tempered spectrum can be described in terms of decay of matrix coefficients of the representation. For $\psi$ a unitary representation of $G(F_v)$ on a Hilbert space $H$, these are the functions $F_w(g)$ on $G(F_v)$ given by $F_w(g) = \langle \psi(g)w, w\rangle_H$ for $w \in H$. Clearly, such a function is bounded on $G(F_v)$ and if $\psi$ is the trivial representation (or possibly finite dimensional) then $F_w(g)$ does not go to zero as $g \to \infty$ (we assume $w \neq 0$). However, for other $\psi$'s these matrix coefficients do decay (Howe-Moore [H-M])* and the rate of decay is closely related to the "temperedness" of $\psi$ and is important in applications. In particular, $\psi$ is tempered iff its $K_v$-finite matrix coefficients are in $L^{2+\epsilon}(G(F_v))$ for all $\epsilon > 0$.

For spherical representations (and in fact for the general ones too) one can use the asymptotics at infinity of spherical functions (that is, $K_v$ bi-invariant eigenfunctions of the Hecke algebra) to determine which are tempered. For $GL_n(F_v)$ this analysis shows that the $\pi_v$ defined in (3) and (4) is tempered iff

$$|\alpha_j(\pi_v)| = 1 \text{ for } j = 1, 2, \ldots, n \text{ if } v \text{ finite} \tag{9}$$

and

$$\Re(\mu_j(\pi_v)) = 0 \text{ if } v \text{ archimedean}\,. \tag{10}$$

For ramified representations of $G(F_v)$ one can give a similar description of the tempered representations in terms of Langlands parameters (see Knapp-Zuckerman [K-Z] for $v$ archimedean).

To complete our digression into more general local groups $G(F_v)$ we recall property $T$. Recall that $G(F_v)$ has property $T$ if the trivial representation is isolated in $\widehat{G(F_v)}$. Kazhdan in introducing this property showed that if $G(F_v)$ is simple and has rank at least two then it satisfies property $T$. One can quantify this property in these cases (as well as in the rank one groups which satisfy property $T$) by giving uniform estimates for the exponential decay rates of any non-trivial unitary representation of $G(F_v)$. In [Oh], Oh gives such bounds which are in fact sharp in many cases (such as for $SL_n, (F_v)$, $n \geq 3$ and $Sp_{2n}(F_v)$). In the case that $F_v \cong \mathbb{R}$, Li [Li1] determines the largest $p = p(G(F_v))$ for which every non-trivial representation of $G(F_v)$ is in $L^{p+\epsilon}(G(F_v))$ for all $\epsilon > 0$. Besides the isolation of the trivial representation in $\widehat{G(F_v)}$ it is also very useful to know which other representations are isolated. For $v$ archimedean and $\pi_v$ cohomological (in the sense of Borel and Wallach [B-W]) Vogan [Vo1] gives a complete description of the isolated points.

We return now to the global setting with $G = GL_n$ and formulate the main Conjecture.

---

*if say $G(F_v)$ is simple

Generalized Ramanujan Conjecture for $GL_n$:

Let $\pi = \bigotimes_v \pi_v$ be an automorphic cuspidal representation of $GL_n(\mathbb{A}_F)$ with a unitary central character, then for each place $v$, $\pi_v$ is tempered.

**Remarks**

(1) At the (almost all) places at which $\pi_v$ is unramified the Conjecture is equivalent to the explicit description of the local parameters satisfying (9) and (10).

(2) For analytic applications the more tempered (i.e. the faster the decay of the matrix coefficients) the better. It can be shown (compare with (28) of §2) that the $\pi_v$'s which occur cuspidally and automorphically are dense in the tempered spectrum, hence $GRC$ if true is sharp.

(3) Satake [Sat] appears to have been the first to observe that the classical Ramanujan Conjecture concerning the Fourier coefficients of $\Delta(z)$ can be formulated in the above manner. The $GRC$ above generalizes both these classical Ramanujan-Petersson Conjectures for holomorphic forms of even integral weight as well as Selberg's 1/4 eigenvalue conjecture for the Laplace spectrum of congruence quotients of the upper half plane [Sel]. In this representation theoretic language the latter is concerned with $\pi_\infty$ which are unramified and for which $\pi = \bigotimes_v \pi_v$ is an automorphic cuspidal representation of $GL_2(\mathbb{A}_\mathbb{Q})$.

There are some special but important cases of $\pi$'s for which the full $GRC$ is known. These are contained in cases where $\pi_v$ for $v$ archimedian is of special type. For $GL_2(\mathbb{A}_\mathbb{Q})$ and $\pi_\infty$ being holomorphic discrete series (that is, the case of classical holomorphic cusp forms of even integral weight) $RC$ was established by Deligne. For a recent treatment see the book by Conrad [Con]. The proof depends on $\Gamma_0(n)\backslash\mathbb{H}$ being a moduli space for elliptic curves (with level structure) and this leads eventually to an identification of $\alpha_j(\pi_p)$, $j = 1, 2$ in terms of arithmetic algebraic geometric data, specifically as eigenvalues of Frobenius acting on $\ell$-adic cohomology groups associated with a related moduli problem. The $RC$, i.e. $|\alpha_1(\pi_p)| = |\alpha_2(\pi_p)| = 1$, then follows from the purity theorem (the Weil Conjectures) for eigenvalues of Frobenius, which was established by Deligne.

Recently, Harris and Taylor [H-T], following earlier work of Clozel have established $GRC$ for an automorphic cusp form $\pi$ on $GL_n(\mathbb{A}_F)$ for which the following are satisfied:

(i) $F$ is a $CM$ field.

(ii) $\widetilde{\Pi} \cong \Pi^c$ (i.e. a contragredient - Galois conjugate condition.)

(iii) $\Pi_\infty$ ($\infty$ here being the product over all archimedian places of $F$) has the same infinitesimal character over $\mathbb{C}$ as the restriction of scalars from $F$ to $\mathbb{Q}$ of an algebraic representation of $GL_n$ over $\mathbb{C}$. In particular, $\pi_\infty$ is a special type of cohomological representation.

(iv) For some finite place $v$ of $F$, $\pi_v$ is square-integrable (that is, its matrix coefficients are square integrable).

The proof of the above is quite a tour-de-force. It combines the trace formula (see Arthur's Lectures) and Shimura varieties and eventually appeals to the purity theorem. To appreciate some of the issues involved consider for example $F$ an imaginary quadratic extension of $\mathbb{Q}$. In this case $F$ has one infinite place $v_\infty$ for which $F_{v_\infty} \simeq \mathbb{C}$. Hence, automorphic forms for $GL_n(F)$ live on quotients of the symmetric space $SL_n(\mathbb{C})\big/SU(n)$, which is not Hermitian. So there is no apparent algebro-geometric moduli interpretation for these quotient spaces. The basic idea is to transfer the given $\pi$ on $GL_n(\mathbb{A}_F)$ to a $\pi'$ on a Shimura variety (see Milne's lectures for definitions of the latter). The Shimura varieties used above are arithmetic quotients of unitary groups (see example 3 of Section 2). The transfer of $\pi$ to $\pi'$ is achieved by the trace formula. While the complete functorial transfers are not known for the general automorphic form, enough is known and developed by Harris, Taylor, Kottwitz and Clozel to deal with the $\pi$ in question. Conditions (i), (ii) and (iv) are used to ensure that $\pi$ corresponds to a $\pi'$ on an appropriate unitary group, while condition (iii) ensures that at the archimedian place, $\pi'$ is cohomological. The latter is essential in identifying the eigenvalues of $\pi'_v$ ($v$ finite) in terms of Frobenius eigenvalues.

In most analytic applications of $GRC$ all $\pi$'s enter and so knowing that the Conjecture is valid for special $\pi$'s is not particularly useful. It is similar to the situation with zeros of the Riemann Zeta Function and $L$-functions where it is not information about zeros on $\Re(s) = \frac{1}{2}$ that is so useful, but rather limiting the locations of zeros that are to the right of $\Re(s) = \frac{1}{2}$. We describe what is known towards $GRC$ beginning with the local bounds. If $\pi = \underset{v}{\otimes}\, \pi_v$ is automorphic and cuspidal on $GL_n(\mathbb{A}_F)$ then $\pi$ and hence $\pi_v$ is firstly unitary and secondly generic. The latter asserts that $\pi_v$ has a Whittaker model (see Cogdell's lectures [Co]). That $\pi_v$ is generic for $\pi$ cuspidal follows from the Fourier Expansions on $GL_n(\mathbb{A}_F)$ of Jacquet, Piatetski-Shapiro and Shalika [J-PS-S]. Now, Jacquet and Shalika [J-S] show that for $\pi_v$ generic the local Rankin-Selberg $L$-function of $\pi_v$ with its contragredient $\tilde{\pi}_v$,

$$L(s, \pi_v \times \tilde{\pi}_v) \,=\, \det\left(I - \alpha(\pi_v) \,\otimes\, \alpha\left(\tilde{\pi}_v\right) N(v)^{-s}\right)^{-1} \tag{11}$$

is analytic in $\Re(s) > 1$.
This leads directly to bounds towards $GRC$. Specifically, in the most important case when $\pi_v$ is unramified, (11) implies that

$$\left|\,\log_{N(v)} |\,\alpha_j(\pi_v)|\,\right| \,<\, \frac{1}{2} \quad \text{for } j = 1, \ldots n, \text{ and } v \text{ finite} \tag{12}$$

and

$$|\Re(\mu_j(\pi_v))\,| \,<\, \frac{1}{2} \quad \text{for } j = 1, 2, \ldots n, \text{ and } v \text{ archimedian}\,. \tag{13}$$

Within the context of generic unitary representations of $GL_n(F_v)$, (12) and (13) are sharp. Recall that the trivial representation corresponds to $\mu$ as given in (5), so that for $n = 2$ (12) and (13) recover the trivial bound. However, for $n \geq 3$ these bounds are non-trivial. For $n = 3$, (12) and (13) correspond to the sharp decay rates for matrix coefficients of non-trivial representations of $SL_3(F_v)$ mentioned earlier. For $n > 3$, the bounds (12) and (13) are much stronger (the trivial bound being $\frac{n-1}{2}$).

For many applications these local bounds fall just short of what is needed (this is clear in the case $n = 2$). One must therefore bring in further global information. The global Rankin-Selberg $L$-function is the key tool. In fact, it was already used by Rankin and Selberg in the case $n = 2$, $F = \mathbb{Q}$ and $v$ finite, for such a purpose. The extension of their analysis to general $n$ and $F$ was observed by Serre [Ser]. However, this method which uses twisting by quasi-characters $\alpha^s$ (a technique which we now call deformation in a family (see [I-S]) of $L$-functions, in this case the parameter being $s$) and a theorem of Landau [La], has the drawback of only working for $v$ finite and also it deteriorates in quality as the extension degree of $F$ over $\mathbb{Q}$ increases, the latter being a result of the increasing number of Gamma factors in the complete $L$-function (see [I-S]). In [L-R-S1], the use of the Rankin-Selberg $L$-functions in a different way and via deformation in another family was developed. It has the advantage of applying to the archimedian places as well as being uniform in its applicability. It leads to the following bounds towards $GRC$. Let $\pi = \underset{v}{\otimes} \pi_v$ be an automorphic cuspidal representation of $GL_n(\mathbb{A}_F)$.

For $v$ finite and $\pi_v$ unramified and $j = 1, \dots, n$

$$|\log_{N(v)}\ \alpha_j(\pi_v)| \le \frac{1}{2} - \frac{1}{n^2+1}\,. \tag{14}$$

For $v$ archimedian and $\pi_v$ unramified and $j = 1, 2, \dots, n$

$$|\Re(\mu_j(\pi_v))\,| \le \frac{1}{2} - \frac{1}{n^2+1}\,. \tag{15}$$

In [M-S] these bounds are extended to include analogous bounds for places $v$ at which $\pi_v$ is ramified.

We describe briefly this use of the global Rankin-Selberg $L$-function. Let $\pi$ be as above and $v_0$ a place at which $\pi_{v_0}$ is unramified. For $\chi$ a ray class character of $F^*\backslash\mathbb{A}_F^*$ which is trivial at $v_0$, we consider the global Rankin-Selberg $L$-functions

$$\begin{aligned}\Lambda\,(s, \pi \times (\tilde{\pi} \times \chi)) &:= \prod_v L(s, \pi_v \times (\tilde{\pi}_v \times \chi_v)) \\ &= L(s, \pi_{v_0} \times \tilde{\pi}_{v_0})\, L_{S_0}(s, \pi \times (\tilde{\pi} \times \chi))\,,\end{aligned} \tag{16}$$

where $L_S(s)$ denotes the partial $L$-function obtained as the product over all places except those in $S$ and $S_0 = \{v_0\}$. Now, according to the theory of the Rankin-Selberg $L$-function ([J-PS-S], [Sh], [M-W]) the left-hand side of (16) is analytic for $0 < \Re(s) < 1$. In particular, if $0 < \sigma_0 < 1$ is a pole of $L(s, \pi_{v_0} \times \tilde{\pi}_{v_0})$ (which will be present according to (9), (10) and (11) if $GRC$ fails for $\pi_{v_0}$) then

$$L_{S_0}(\sigma_0, \pi \times \tilde{\pi} \times \chi) = 0 \text{ for all } \chi \text{ with } \chi \text{ trivial at } v_0\,. \tag{17}$$

Thus, we are led to showing that $L_{S_0}(\sigma_0, \pi \times \tilde{\pi} \times \chi) \neq 0$ for some $\chi$ in this family. To see this one averages these $L$-functions over the set of all such $\chi$'s of a large conductor $q$. The construction of $\chi$'s satisfying the condition at $v_0$ is quite delicate (see [Roh]). In any event, using techniques from analytic number theory for averaging over families of $L$- functions, together with the positivity of the coefficients of $L(s, \pi \times \tilde{\pi})$, one shows that these averages are not zero if $N(q)$ is large enough and $\sigma_0$ is not too small. Combined with (17) this leads to (14) and (15).

The bounds (14) and (15) are the best available for $n \ge 3$. For $n = 2$ much better bounds are known and these come from the theory of higher tensor power $L$-functions. Recall that for $G = GL_n$, ${}^LG^0 = GL(n, \mathbb{C})$. In the case of $n = 2$ and $k \ge 1$ let $\mathrm{sym}^k : {}^LG^0 \longrightarrow GL(k+1, \mathbb{C})$ be the representation of $GL(2, \mathbb{C})$ on symmetric $k$-tensors (i.e. the action on homogeneous polynomials of degree $k$ in $x_1, x_2$ by linear substitutions). The corresponding local $L$-function associated to an automorphic cusp form $\pi$ on $GL_2(\mathbb{A}_F)$ and the representation $\mathrm{sym}^k$ of ${}^LG^0$ is given in (8). The global $L$-function with appropriate definitions at ramified places is given as usual by

$$\Lambda(s, \pi, \mathrm{sym}^k) = \prod_v L(s, \pi_v, \mathrm{sym}^k)\,. \tag{18}$$

Langlands [Lang2] made an important observation that if $\Lambda(s,\pi,\mathrm{sym}^k)$ is analytic in $\Re(s) > 1$ for all $k \geq 1$ (as he conjectured it to be) then a simple positivity argument yields $GRC$ for $\pi$.† Moreover, his general functoriality conjectures assert that $\Lambda(s,\pi,\mathrm{sym}^k)$ should be the global $L$-function of an automorphic form $\Pi_k$ on $GL_{k+1}(\mathbb{A}_F)$. Hence, the functoriality conjectures imply $GRC$. There have been some striking advances recently in this direction. The functorial lift $\pi \to \Pi_k$ of $GL_2$ to $GL_{k+1}$ is now known for $k = 2, 3$ and 4. The method of establishing these lifts is based on the converse theorem (see Cogdell's lectures [Cog]). This asserts that $\Pi$ is automorphic on $GL_n(\mathbb{A}_F)$ as long as the $L$-functions $\Lambda(s,\pi \times \pi_1)$ are entire and satisfy appropriate functional equations for all automorphic forms $\pi_1$ on $GL_m(\mathbb{A}_F)$ for $m \leq n-1$ (one can even allow $m < n-1$ if $n \geq 3$). In this way automorphy is reduced to establishing these analytic properties. This might appear to beg the question; however for $k = 2$ (and $\pi = \Pi_2$ on $GL_3$ as above) the theory of theta functions and half integral weight modular forms, combined with the Rankin-Selberg method, yields the desired analytic properties of $\Lambda(s,\pi,\mathrm{sym}^2)$ (Shimura [Shi]). For $k = 3, 4$ the analytic properties were established by Kim and Shahidi [K-S] , [K]. They achieve this using the Langlands-Shahidi method which appeals to the analytic properties of Eisenstein series on exceptional groups (up to and including $E_8$, so that this method is limited) to realize the functions $\Lambda(s,\Pi_k \times \pi')$ above in terms of the coefficients of Eisenstein series along parabolic subgroups. The general theory of Eisenstein series and their meromorphic continuation (Langlands) yields in this way the meromorphic continuation and functional equations for these $\Lambda(s,\Pi_k \times \pi')$. The proof that they are entire requires further ingenious arguments. Their work is precise enough to determine exactly when $\Pi_k$ is cuspidal (which is the case unless $\pi$ is very special, that being it corresponds to a two-dimensional representation of the Weil group $W_F$ in which case $GRC$ for $\pi$ is immediate). Now, using that $\Pi_k$ for $1 \leq k \leq 4$ is cuspidal on $GL_{k+1}(\mathbb{A}_F)$ and forming the Rankin-Selberg $L$-functions of pairs of these leads to $\Lambda(s,\pi,\mathrm{sym}^k)$ being analytic for $\Re(s) > 1$ and $k \leq 9$. From this one deduces that for $\pi$ as above, cuspidal on $GL_2(\mathbb{A}_F)$ and $\pi_v$ unramified (if $\pi_v$ is ramified on $GL_2(F_v)$ then it is tempered) that

$$|\log_{N(v)} |\alpha_j(\pi_v)|| \leq \frac{1}{9} \qquad \text{for } j = 1, 2 \text{ and } v \text{ finite} \tag{19}$$

and

$$|\Re(\mu_j(\pi_v))| \leq \frac{1}{9} \qquad \text{for } j = 1, 2 \text{ and } v \text{ archimedian}\,. \tag{20}$$

There is a further small improvement of (19) and (20) that has been established in the case $F = \mathbb{Q}$ [Ki-Sa]. One can use the symmetric square $L$-function in place of the Rankin-Selberg $L$-function in (16). This has the effect of reducing the "analytic conductor" (see [I-S] for the definition and properties of the latter). Applying the technique of Duke and Iwaniec [D-I] at the finite places and [L-R-S 2] at the archimedian place, one obtains the following refined estimates. For $n \leq 4$ and $\pi$

---

†This approach to the local statements involved in $GRC$ via the analytic properties of the global $L$-functions associated with large irreducible representations of ${}^LG$ has been influential. In Deligne's proof of the Weil Conjectures mentioned earlier, this procedure was followed. In that case, ${}^LG$ is replaced by the monodromy representation of the fundamental group of the parameter space for a family of zeta functions for whose members the Weil Conjectures are to be established. The analytic properties of the corresponding global $L$-functions follows from Grothendieck's cohomology theory.

an automorphic cusp form on $GL_n(\mathbb{A}_\mathbb{Q})$, or if $n=5$ and $\pi = \mathrm{sym}^4\psi$ with $\psi$ a cusp form an $GL_2(\mathbb{A}_\mathbb{Q})$, we have

$$|\log_p|\alpha_j(\pi_p)|| \le \frac{1}{2} - \frac{1}{1+\frac{n(n+1)}{2}}, p \text{ finite} \tag{21}$$

and

$$|\Re(\mu_j(\pi_\infty))| \le \frac{1}{2} - \frac{1}{1+\frac{n(n+1)}{2}}, p = \infty\,. \tag{22}$$

In particular, if we apply this to a cusp from $\psi$ on $GL_2(\mathbb{A}_\mathbb{Q})$ we get

$$|\log_p \alpha_j(\psi_p)| \le \frac{7}{64}, \text{ for } j=1,2 \text{ and } p<\infty \tag{23}$$

and

$$|\Re\mu_j(\psi_\infty)| \le \frac{7}{64}, \text{ for } j=1,2\,. \tag{24}$$

(24) is equivalent to the following useful bound towards Selberg's 1/4 conjecture concerning the first eigenvalue of the Laplacian $\lambda_1(\Gamma(N)\backslash\mathbb{H})$ for a congruence quotient of the upper half plane $\mathbb{H}$.

$$\lambda_1(\Gamma(N)\backslash\mathbb{H}) \ge \frac{975}{4096} = 0.238\ldots\,. \tag{25}$$

## 2. General $G$

Let $G$ be a reductive linear algebraic group defined over $F$. The principle of functoriality gives relations between the spectra of $G(F)\backslash G(\mathbb{A}_F)$ for different $G$'s and $F$'s. In particular, in cases where versions of this principle are known or better yet where versions of the more precise conjectures of Arthur are known, one can transfer information towards the Ramanujan Conjectures from one group to another. For example, if $D$ is a quaternion algebra over $F$, then the Jacquet-Langlands correspondence [Ge] from $D^*(F)\backslash D^*(\mathbb{A}_F)$ into $GL_2(F)\backslash GL_2(\mathbb{A}_F)$ allows one to formulate a precise *GRC* for $D$ as well as to establish bounds towards it using (19) and (20). In fact, if $D/\mathbb{Q}$ is such that $D\otimes\mathbb{R}\cong H(\mathbb{R})$, the Hamilton quaternions, then the transfer to $GL_2(\mathbb{A}_\mathbb{Q})$ yields only $\pi$'s for which $\pi_\infty$ is a holomorphic representation of $GL_2(\mathbb{R})$. Hence for such $D$'s the full *GRC* is known by Deligne's results mentioned in Section 1. Our main interest however is in $G$'s for which $G(F_v)$ is non-compact for at least one archimedean place $v$ of $F$. The remarks above about quaternion algebras apply to division algebras of degree $n$ over $F$ using the correspondence to $GL_n(\mathbb{A}_F)$ established by Arthur and Clozel [A-C]. Another example is that of unitary groups $G$ over $F$ in 3 variables and the transfer established by Rogawski [Ro] of the non-lifted forms (from $U(2)\times U(1)$) on $G(\mathbb{A}_F)$ to $GL_3(\mathbb{A}_E)$ where $E$ is a quadratic extension of $F$ (we discuss this example further in example 3 below). In all of the above examples the forms are lifted to $GL_n$ and after examining for cuspidality (14) and (15) yield the best approximations to *GRC* for the corresponding $G$. We note that, in the cases above, functoriality is established using the trace formula.

For a general semi-simple $G$ (for the rest of this Section we will assume that $G$ is semi-simple) defined over $F$, the Ramanujan Conjecture can be very complicated. It has been known for some time, at least since Kurokawa [Ku], that there are non-tempered automorphic cuspidal representations for groups such as $GSp(4)$. So the naive generalization of the $GL_n$ $GRC$ is not valid. Today the general belief is that such non-tempered representations are accounted for by functorial lifts from smaller groups.

One approach to $GRC$ for more general $G$, and which is along the lines of the original Ramanujan Conjecture, is to formulate the problem in a cruder form which is well-suited for analytic applications of the spectrum. For the latter, one wants to know the extent to which the local representations appearing as components of a global automorphic representation are limited. Put another way, which local representations in $\widehat{G(F_v)}$ can be excited arithmetically? Let $\pi = \underset{v}{\otimes} \pi_v$ be an automorphic representation appearing in $L^2(G(F)\backslash G(\mathbb{A}_F))$. That is, $\pi$ occurs cuspidally or as a residues of Eisenstein series or as part of a unitary integral of Eisenstein series. We will not distinguish the part of the spectrum in which these occur. This is one sense in which we seek cruder information. Now fix a place $w$ of $F$ and define the subset $\widehat{G(F_w)}_{\mathrm{AUT}}$ of $\widehat{G(F_w)}$ to be the closure in the Fell topology of the set of $\pi_w$'s for which $\pi = \underset{v}{\otimes} \pi_v$ occurs in $L^2(G(F)\backslash G(\mathbb{A}_F))$. This closure process is the second sense in which we seek cruder information. We call $\widehat{G(F_w)}_{\mathrm{AUT}}$ the automorphic dual of $G$ at $w$. More generally, if $S$ is a finite set of places of $F$ we define $\widehat{G(S)}_{\mathrm{AUT}}$ to be the closure in $\widehat{G(S)}$ of $\underset{w\in S}{\otimes} \pi_w$ as $\pi$ varies over all $\pi$ in $L^2(G(F)\backslash G(\mathbb{A}_F))$ and $G(S) = \Pi_{w\in S} G(F_w)$. Similarly, one can define $\widehat{G_{\mathrm{AUT}}}$ to be the corresponding closure in $\Pi_v \widehat{G(F_v)}$. By approximation theorems for adèle groups we can describe these sets in terms of congruence subgroups as follows.[‡] Let $S_\infty$ be the set of archimedian places of $F$. Then $\widehat{G(S_\infty)}_{\mathrm{AUT}}$ is the closure of all $\underset{w\in S_\infty}{\otimes} \Pi_w$ in $\widehat{G(S_\infty)}$ which occur in $L^2(\Gamma\backslash G(S_\infty))$ where $\Gamma$ varies over all congruence subgroups of $G(O_F)$, $O_F$ being the ring of integers of $F$. Similarly, if $S$ is a finite set of places containing $S_\infty$ then $\widehat{G(S)}_{\mathrm{AUT}}$ is the closure in $\underset{v\in S}{\Pi} \widehat{G(F_v)}$ of all $\underset{w\in S}{\otimes} \pi_w$ which occur in $L^2(\Gamma\backslash G(S))$, as $\Gamma$ varies over all congruence subgroups of the $S$-arithmetic group $G(O_S)$, with $O_S$ being the ring of $S$-integers of $F$. We can now state the basic problem for $G$.

Generalized Ramanujan Problem ($GRP$):

To determine, for a given $G$ defined over $F$, the sets $\widehat{G(F_v)}_{\mathrm{AUT}}$ and more generally $\widehat{G}_{\mathrm{AUT}}$.

We emphasize that the local data $\widehat{G(F_v)}_{\mathrm{AUT}}$ is determined by the global group $G$. Also, while the set of $\pi_w$'s in $\widehat{G(F_w)}$ that arise as the $w$ component of an automorphic $\pi$ in $L^2(G(F)\backslash G(\mathbb{A}_F))$ is typically very difficult to describe, the closure process in the definition of the automorphic dual makes this task much simpler. Moreover, the above formulation allows one to measure progress towards $GRP$ by giving set theoretic upper and lower bounds for $\widehat{G}_{\mathrm{AUT}}$. Non-trivial upper bounds are what are most useful in applications while various methods for constructing

[‡]At least if $G$ is simply connected and $F$ simple, otherwise the description is more complicated.

automorphic forms (some of which are discussed in the examples below) produce lower bounds. We denote by $\widehat{G(F_v)}^{\text{sph}}_{\text{AUT}}$ the spherical part of $\widehat{G(F_v)}_{\text{AUT}}$.

Let $G$ be defined over $F$ and let $H$ be a semi-simple subgroup of $G$ also defined over $F$. Then $\widehat{H(F_v)}_{\text{AUT}}$ and $\widehat{G(F_v)}_{\text{AUT}}$ and more generally, $\widehat{H(S)}_{\text{AUT}}$ and $\widehat{G(S)}_{\text{AUT}}$, satisfy some simple functorial properties.

If $\sigma \in \widehat{H(F_v)}_{\text{AUT}}$ then

$$\operatorname{Ind}_{H(F_v)}^{G(F_v)} \sigma \subset \widehat{G(F_v)}_{\text{AUT}} . \tag{26}$$

If $\beta \in \widehat{G(F_v)}_{\text{AUT}}$ then

$$\operatorname{Res}_{H(F_v)}^{G(F_v)} \beta \subset \widehat{H(F_v))}_{\text{AUT}} . \tag{27}$$

The induction and restriction computations involved in (26) and (27) are purely local. Their precise meaning is that any irreducible $\psi$ which is contained (weakly) on the left is contained on the right-hand side of the inclusions. These inclusions were proven in [B-S] and [B-L-S1] for $F_v = \mathbb{R}$ and in general (that is, for finitely many places at a time) in [C-U]. The proof of (26) depends on realizing the congruence subgroups of $H(F)$ as geometric limits (specifically as infinite intersections) of congruence subgroups of $G(F)$ and applying the spectral theory of such infinite volume quotients. In [Ven] a characterization of such intersections of congruence subgroups of $G(F)$ is given. (27) is established by approximating diagonal matrix coefficients of $\operatorname{Res}_{H(F_v)}^{G(F_v)} \beta$ by matrix coefficients of elements in $\widehat{H(F_v)}_{\text{AUT}}$. This is done by constructing suitable sequences of $H$ cycles, in a given congruence quotient of $G$, which become equidistributed in the limit. The latter can be done either using Hecke operators or using ergodic theoretic techniques associated with unipotent flows.

(26) and (27) may be used to give upper and lower bounds for $\widehat{G}_{\text{AUT}}$. For example, if $H = \{e\}$ and $\sigma = 1$ then (26) applies and since $\overline{\operatorname{Ind}_{\{e\}}^{G(F_v)} 1} = \widehat{G(F_v)}_{\text{temp}}$, we obtain the general lower bound

$$\widehat{G(F_v)}_{\text{AUT}} \supset \widehat{G(F_v)}_{\text{temp}} . \tag{28}$$

Next, we illustrate by way of examples, some bounds towards $GRP$ that have been established using current techniques.

Example 1. $SL$

Let $G = SL_2$ over $\mathbb{Q}$. The local components of the unitary Eisenstein integrals involved in the spectral decomposition of $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}_{\mathbb{Q}})$ satisfy $GRC$ at all places. Moreover, the only residue of the Eisenstein series is the trivial representation. Hence, the Ramanujan and Selberg Conjectures for the cuspidal spectrum of $SL_2(\mathbb{A}_{\mathbb{Q}})$ are equivalent to

$$\widehat{SL_2(\mathbb{Q}_v)}_{\text{AUT}} = \{1\} \cup \widehat{SL_2(\mathbb{Q}_v)}_{\text{temp}}, \text{ for all places } v \text{ of } \mathbb{Q} . \tag{29}$$

For this case (23) and (24) give the best known upper bounds towards (29).

Let $G = SL_3$ over $\mathbb{Q}$. Again, there are no poles of the Eisenstein series yielding residual spectrum other than the trivial representation. However, there is an integral of non-tempered unitary Eisenstein series contributing to $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}_{\mathbb{Q}}))$. These correspond to the Eisenstein series on the maximal $(2,1)$ parabolic subgroup of $G$ taken with the trivial representation on its Levi. In particular for any place $v$ of $\mathbb{Q}$, $\widehat{G(\mathbb{Q}_v)}_{\mathrm{AUT}}$ contains the following non-tempered spherical principal series (we use the parameters in (4) above):

$$\mathrm{Cont}(v) = \left\{\mu_t = \left(\frac{1}{2} + it,\, -2it,\, -\frac{1}{2} + it\right) \Big|\, t \in \mathbb{R}\right\} \subset \widehat{G(\mathbb{Q}_v)}^{\mathrm{sph}}. \tag{29'}$$

If (29) is true then the rest of the Eisenstein series contribution to $SL_3$, consists of tempered spectrum. Hence using (12) and (13) we see that the cuspidal $GRC$ for $SL_3(\mathbb{A}_{\mathbb{Q}})$ is equivalent to:
For any place $v$ of $\mathbb{Q}$

$$\widehat{G(\mathbb{Q}_v)}_{\mathrm{AUT}} = \{1\} \cup \mathrm{Cont}(v) \cup \widehat{G(\mathbb{Q}_v)}_{\mathrm{temp}}. \tag{29''}$$

The best upper bound on $\widehat{G}_{\mathrm{AUT}}$ in this case is given by (21) and (22) which assert that for any place $v$ of $\mathbb{Q}$

$$\begin{aligned} \widehat{G(\mathbb{Q}_v)}^{\mathrm{sph}}_{\mathrm{AUT}} \subset \Big\{\mu \in \widehat{G(\mathbb{Q}_v)}^{\mathrm{sph}} \big| \mu = (1,0,-1);\ \mu = \left(\frac{1}{2} + it,\, -2it,\, -\frac{1}{2} + it\right) \\ t \in \mathbb{R};\ \mu \text{ such that } |\Re(\mu_j)| \le \frac{5}{14}.\Big\} \end{aligned} \tag{29'''}$$

Using [M-W] one can make a similar analysis for $SL_n$, $n \ge 4$.

Example 2. Orthogonal Group

Let $f$ be the quadratic form over $\mathbb{Q}$ in $n+1$ variables given by

$$f(x_1, x_2, \dots x_{n+1}) = 2x_1x_{n+1} + x_2^2 + \cdots + x_n^2. \tag{30}$$

Let $G = SO_f$ be the special orthogonal group of $(n+1)\times(n+1)$ matrices preserving $f$. $G$ is defined over $\mathbb{Q}$ and is given explicitly by

$$G = \left\{ g \in SL_{n+1} \Big| g^t \begin{pmatrix} & & 1 \\ & I_{n-1} & \\ 1 & & \end{pmatrix} g = \begin{pmatrix} & & 1 \\ & I_{n-1} & \\ 1 & & \end{pmatrix} \right\}. \tag{31}$$

Thus $G(\mathbb{Q}_\infty) = G(\mathbb{R}) \cong SO_{\mathbb{R}}(n,1)$, which has real rank 1. The corresponding symmetric space $G(\mathbb{R})/K$ with $K \cong SO_{\mathbb{R}}(n)$ is hyperbolic $n$-space. Let $M(\mathbb{R}), N(\mathbb{R})$ and $A(\mathbb{R})$ be the subgroups of $G(\mathbb{R})$

$$A(\mathbb{R}) = \left\{ \begin{pmatrix} a & & \\ & I_{n-1} & \\ & & a^{-1} \end{pmatrix} \Big| a \in \mathbb{R}^* \right\} \tag{32}$$

$$N(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & -u^t & -\frac{1}{2}\langle u, u\rangle \\ & I_{n-1} & u \\ & & 1 \end{pmatrix} \Big| u \in \mathbb{R}^{n-1} \right\} \tag{33}$$

$$M(\mathbb{R}) = \left\{ \begin{pmatrix} 1 & & \\ & h & \\ & & 1 \end{pmatrix} \Big| h^t h = I_{n+1} \right\} \tag{34}$$

Then $P(\mathbb{R}) = M(\mathbb{R})A(\mathbb{R})N(\mathbb{R})$ is a parabolic subgroup of $G$ with Levi factor $MA$ and unipotent radical $N$. The spherical unitary dual of $G(\mathbb{R})$ may be described in terms of the principal series. For $s \in \mathbb{C}$ let

$$\pi_s = \mathrm{Ind}_{M(\mathbb{R})A(\mathbb{R})N(\mathbb{R})}^{G(\mathbb{R})} 1_M \otimes |a|^s . \tag{35}$$

(For $s \in \mathbb{C}$ for which $\pi_s$ is reducible we take the spherical constitutent for $\pi_s$.) In this normalization $s = \frac{n-1}{2} := \rho_n$ corresponds to the trivial representation and the tempered spherical representations consist of $\pi_s$ with $s \in i\mathbb{R}$. For $-\rho_n \leq s \leq \rho_n$, $\pi_s$ is unitarizable and these constitute the complementary series. Moreover, $\pi_s$ is equivalent to $\pi_{-s}$. These yield the entire spherical unitary dual, that is,

$$\widehat{G(\mathbb{R})}^{\mathrm{sph}} = \left\{\pi_s \mod \pm 1 \,\middle|\, s \in i\mathbb{R} \cup [-\rho_n, \rho_n] \right\} . \tag{36}$$

Here $i\mathbb{R}$ is identified with $\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{temp}}$ and $(0, \rho_n]$ is identified with the non-tempered part of $\widehat{G(\mathbb{R})}^{\mathrm{sph}}$. Towards the $GRP$ for $G_f$ we have the following inclusions ($n \geq 3$), see [B-S]:

$$i\mathbb{R} \cup \{\rho_n, \rho_n - 1, \ldots, \rho_n - [\rho_n]\} \subset \widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}} \subset i\mathbb{R} \cup \{\rho_n\} \cup \left[0, \rho_n - \frac{7}{9}\right] . \tag{37}$$

In particular, for $n \geq 4$, $\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}}$ contains non-tempered points besides the trivial representation. (37) is deduced from (26) and (27) as follows. Let $H$ be the subgroup of $G$ stabilizing $x_2$. $H$ together with $\sigma = 1$ satisfies the assumptions in (26). Hence

$$\widehat{G(\mathbb{R})}_{\mathrm{AUT}} \supset \mathrm{Ind}_{H(\mathbb{R})}^{G(\mathbb{R})} \mathbf{1} . \tag{38}$$

The space $G(\mathbb{R})/H(\mathbb{R})$ is an affine symmetric space and for general such spaces the induction on the right-hand side of (38) has been computed explicitly by Oshima (see [O-M] and [Vo2]). For the case at hand, one has

$$\mathrm{Ind}_{H(\mathbb{R})}^{G(\mathbb{R})} \supset \{\rho_n, \rho_n - 1 \ldots, \rho_n - [\rho_n]\} \cup i\mathbb{R} . \tag{39}$$

This gives the lower bound in (37).

To see the upper bound, first note that for $n = 3$ we have

$$\widehat{G(\mathbb{R})}^{\mathrm{sph}} \subset i\mathbb{R} \cup \{1\} \cup \left[0, \frac{2}{9}\right] . \tag{40}$$

This follows by passing from this $SO_f$ to its spin double cover which at $\mathbb{Q}_\infty$ is $SL_2(\mathbb{C})$ and then invoking the bound (20) for $GL_2(\mathbb{A}_E)$ where $E$ is an imaginary quadratic extension of $\mathbb{Q}$. If $n > 3$ we let $H$ be the subgroup of $G$ which stabilizes $x_2, \ldots, x_{n-2}$. Then $H = G_{f'}$ with $f'$ a form in 4 variables of signature $(3, 1)$. Thus according to (40)

$$\widehat{H(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}} \subset i\mathbb{R} \cup \{1\} \cup \left[0, \frac{2}{9}\right] . \tag{41}$$

Now apply the restriction principle (27) with the pair $G$ and $H$ as above and with $\beta$ a potential non-tempered element in $\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}}$. Computing the local restriction $\mathrm{Res}_{H(\mathbb{R})}^{G(\mathbb{R})} \beta$ and applying (41) leads to the upper bound in (37).

One is led to a precise $GRC$ for $G$ at $\mathbb{Q}_\infty$:

Conjecture: Let $G$ be as in (31); then

$$\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}} = i\mathbb{R} \cup \{\rho_n, \rho_n - 1, \ldots, \rho_n - [\rho_n]\}\,. \tag{42}$$

Example 3. Unitary Group

Let $SU(2,1)$ be the special unitary group of $3 \times 3$ complex matrices of determinant equal to one, that is, such matrices preserving the Hermitian form $|z_1|^2 + |z_2|^2 - |z_3|^2$. If $g \in SU(2,1)$ and $g = \begin{bmatrix} A & b \\ c^* & d \end{bmatrix}$ with $A$ $2 \times 2$, $b$ and $c$ $2 \times 1$ and $d$ $1 \times 1$ complex matrices, then $g$ acts projectively on

$$B^2 = \left\{(z_1, z_2) \,|\, |z_1|^2 + |z_2|^2 < 1\right\} \text{ by}$$

$$gz = (\langle z, c\rangle + d)^{-1}(Az + b)\,. \tag{43}$$

In this way $B^2 \simeq SU(2,1)/K$, with $K = S(U(2) \times U(1))$, is the corresponding Hermitian symmetric space. The biholomorphic action (43) extends to the closed ball $\overline{B^2}$. If $e_1 = (1,0) \in \overline{B^2} \setminus B^2$ then its stabilizer $P = \{g \in SU(2,1) | ge_1 = e_1\}$ is a parabolic subgroup of $SU(2,1)$. Let $\Gamma$ be a co-compact lattice in $SU(2,1)$. It acts discontinuously on $B^2$ and we form the compact quotient $X_\Gamma = \Gamma\backslash B^2$ which is a compact, complex Kahler surface. We examine the Betti numbers $b_j(\chi_\Gamma)$ for $j = 0, 1, 2, 3$ and 4. According to the Gauss-Bonnet-Chern formula

$$\chi(X_\Gamma) = b_0 - b_1 + b_2 - b_3 + b_4 = \mathrm{Vol}(\Gamma\backslash SU(2,1)) \tag{44}$$

with $dg$ being a suitable fixed normalized Haar measure on $SU(2,1)$ (this is a special case of the "Euler-Poincaré measure" in [Ser2]). By duality this yields

$$\mathrm{Vol}(\Gamma\backslash SU(2,1)) = b_2 - 2b_1 + 2\,. \tag{45}$$

It follows that if $\mathrm{Vol}(\Gamma\backslash SU(2,1))$ goes to infinity then so does $b_2(X_\Gamma)$. Thus for large volume $X_\Gamma$ will have cohomology in the middle dimension. The behavior of $b_1(X_\Gamma)$ is subtle and in algebraic surface theory this number is known as the irregularity of $X_\Gamma$. It can be calculated from the decomposition of the regular representation of $SU(2,1)$ on $L^2(\Gamma\backslash SU(2,1))$, for a discussion see Wallach [Wa]. We indicate briefly how this is done. The representation $\mathrm{Ind}_P^{SU(2,1)} 1$ (nonunitary induction) of $SU(2,1)$ is reducible. Besides containing the trivial representation as a subrepresentation it also contains two irreducible subquotients $\pi_0^+$ and $\pi_0^-$ (see [J-W]). $\pi_0^\pm$ are non-tempered unitary representations of $SU(2,1)$, in fact their $K$-finite matrix coefficients lie in $L^p(SU(2,1))$ for $p > 4$, but not in $L^4$. Let $m_\Gamma(\pi_0^+)$ and $m_\Gamma(\pi_0^-)$ be the multiplicities with which $\pi_0^+$(respectively $\pi_0^-$) occur in the decomposition of $L^2(\Gamma\backslash SU(2,1))$. For the example at hand, these multiplicities are equal (which is a reflection of $X_\Gamma$ being Kahler). The following is a particular case of Matsushima's formula (see Borel-Wallach [B-W]) which gives the dimensions of various cohomology groups of a general locally symmetric space $\Gamma\backslash G/K$ in terms of the multiplicities with which certain $\pi$'s in $\widehat{G}$ occur in $L^2(\Gamma\backslash G)$.

$$b_1(X_\Gamma) = m_\Gamma(\pi_0^+) + m_\Gamma(\pi_0^-) = 2m_\Gamma(\pi_0^+)\,. \tag{46}$$

We examine the above in the case that $\Gamma$ is a special arithmetic lattice. Let $E$ be an imaginary quadratic extension of $\mathbb{Q}$ and let $D$ be a degree 3 division algebra central over $E$ and which carries an involution $\alpha$ of the second kind, that is, the restriction of $\alpha$ to $E$ is Galois conjugation $E/\mathbb{Q}$. Let $G$ be the $\mathbb{Q}$-algebraic group whose $\mathbb{Q}$ points $G(\mathbb{Q})$ equals $\{g \in D^* | \alpha(g)g = 1 \text{ and } Nrd(g) = 1\}$. Here

$Nrd$ is the reduced norm on $D$. $G$ is the special unitary group $SU(D,f)$ where $f$ is the 1-dimensional (over $D$) Hermitian form $f(x,y) = \alpha(x)y$. Localizing $G$ at $\mathbb{Q}_\infty = \mathbb{R}$ we obtain $G(\mathbb{R})$ which is a special unitary group in 3 variables and which we assume has signature $(2,1)$, that is, $G(\mathbb{R}) \simeq SU(2,1)$. In this case $G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q})$ is compact and we consider its automorphic dual and specifically $\widehat{G(\mathbb{R})}_{\mathrm{AUT}}$. The key to obtaining information about $\widehat{G}_{\mathrm{AUT}}$ is the explicit description by Rogawski [Ro] of the spectrum of $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q}))$ in terms of certain automorphic forms on $GL_3(\mathbb{A}_E)$ (see his Chapter 14 which discusses inner forms). Not surprisingly the $\Pi$'s on $GL_3(\mathbb{A}_E)$ arising this way satisfy conditions similar to (i), (ii) and (iii) on page 6. If $\pi = \bigotimes_v \pi_v$ is an automorphic representation of $G(\mathbb{A}_\mathbb{Q})$ and is not 1-dimensional then the lifted form $\Pi = \bigotimes_w \Pi_w$ is cuspidal on $GL_3(\mathbb{A}_E)$. The relation between $\Pi_w$ for $w|v$ and $\pi_v$ is given explicitly. Thus the $GRC$ for $G$ takes the simplest form: If $\pi$ is not 1-dimensional then $\pi_v$ is tempered for all places $v$ of $\mathbb{Q}$. Moreover (14) and (15) yield corresponding non-trivial bounds on $\widehat{G}_{\mathrm{AUT}}$.

We fixate on the representations $\pi_0^\pm$ in $\widehat{G(\mathbb{R})}$. (15) implies that

$$\pi_0^\pm \notin \widehat{G(\mathbb{R})}_{\mathrm{AUT}} \tag{47}$$

(see [B-C]).
This upper bound on $GRC$ for this $G$ implies a fortiori that $m_\Gamma(\pi_0^-)$ are zero for any congruence subgroup $\Gamma$ of $G(\mathbb{Z})$. This combined with (46) has the following quite striking vanishing theorem as a consequence (and was proved in this way by Rogawski)

$$b_1(X_\Gamma) = 0, \text{ for } \Gamma \text{ any congruence subgroup of } G(\mathbb{Z}). \tag{48}$$

In particular, these arithmetic surfaces $X_\Gamma$ have no irregularities and all their non-trivial cohomology is in the middle degree and its dimension is given by the index (45).

The vanishing theorem (48) is of an arithmetic nature. It is a direct consequence of restrictions imposed by the Ramanujan Conjectures. It should be compared with vanishing theorems which are consequences of Matsushima's formula, by which we mean the vanishing of certain cohomology groups of general locally symmetric spaces $X_\Gamma = \Gamma\backslash G/K$, independent of $\Gamma$. The vanishing results from the fact that none of the potential $\pi$'s which contribute to Matsushima's formula are unitary. A complete table of the cohomological unitary representations and the corresponding vanishing degrees for general real groups $G$ is given in [V-Z].

Example 4. Exceptional groups

The theory of theta functions and its extension to general dual pairs provides a powerful method for constructing "lifted" automorphic forms and in particular non-tempered elements in $\widehat{G}_{\mathrm{AUT}}$. Briefly, a reductive dual pair is a triple of reductive algebraic groups $H, H'$ and $G$ with $H$ and $H'$ being subgroups of $G$ which centralize each other. If $\pi$ is a representation of $G$ then the analysis of the restriction $\pi\big|_{H\times H'}$ (here $\pi((h,h')) = \pi(hh'))$ can lead to a transfer of representations on $H$ to $H'$ (or vice-versa). The classical case of theta functions is concerned with $G$ being the symplectic group and $\pi = w$, the oscillator representation. That $w$ is automorphic was shown in Weil [We] while the general theory in this setting is due to Howe [Ho]. Recent works ([Ka-Sav], [R-S2], [G-G-J]) for example show that this rich theory can be extended to other groups $G$ such as exceptional groups with $\pi$ being the minimal representation. For such suitably split $G$ the minimal representation is shown to be automorphic by realizing it as a residue of Eisenstein series [G-R-S]. For an account of the general theory of dual pairs and the minimal representation see [Li2].

For example, the dual pair $O(n,1) \times SL_2$ in a suitable symplectic group may be used to give another proof of the lower bound in (37). Restricting the oscillator representation to this dual pair one finds that holomorphic discrete series of weight $k$ on $SL_2$ correspond to the point $\rho_n - k$ in (37); see Rallis-Schiffmann [R-S1] and [B-L-S2].

We illustrate these methods with a couple of examples of exceptional groups. Let $G$ be the automorphism group of the split Cayley algebra over $\mathbb{Q}$ (see [R-S] for explicit descriptions of the group as well as various data associated with it). $G$ is a linear algebraic group defined over $\mathbb{Q}$ and is split of type $G_2$. It is semi-simple, it has rank 2 and as a root system for a maximal split torus we can take $\Delta = \{\pm(e_1 - e_2),\ \pm(e_1 - e_3),\ \pm(e_2 - e_3),\ \pm(2e_1 - e_2 - e_3),\ \pm(2e_2 - e_1 - e_3),\ \pm(2e_3 - e_1 - e_2)\}$ in $V = \{(a,b,c) | a+b+c = 0\}$ and with the standard pairing $\langle\, ,\rangle$. Here $e_1, e_2, e_3$ are the standard basis vectors. The corresponding Weyl group $W$ is of order 12. It is generated by reflections along the roots and preserves $\langle\, ,\rangle$. The long root $\beta_1 = 2e_1 - e_2 - e_3$ together with the short root $\beta_6 = -e_1 + e_2$ form a basis and determine corresponding positive roots $\beta_1, \beta_2, \dots, \beta_6$, see Figure 1. Up to conjugacy $G$ has 3 parabolic subgroups; $P_0$ the minimal parabolic subgroup, $P_1$ the maximal parabolic corresponding to $\beta_1$ and $P_2$ the maximal parabolic corresponding to $\beta_6$. The parabolic subgroup $P_j$ factorizes as $L_j N_j$ with $L_j$ the Levi factor and $N_j$ the unipotent factor. Here $L_1$ and $L_2$ are isomorphic to $GL_2$. We examine the automorphic dual $\widehat{G}_{\mathrm{AUT}}$ associated with the spectrum of $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}_{\mathbb{Q}}))$ and specifically $\widehat{G(\mathbb{R})}_{\mathrm{AUT}}$.

We recall the classification by Vogan [Vo3] of the unitary spherical dual $\widehat{G(\mathbb{R})}^{\mathrm{sph}}$. The maximal compact subgroup $K$ of $G(\mathbb{R})$ is $SU(2) \times SU(2)$. The corresponding Riemannian symmetric space $G(\mathbb{R})/K$ is 8-dimensional. For $j = 0,1,2$ let $P_j(\mathbb{R}) = M_j(\mathbb{R}) A_j(\mathbb{R}) N_j(\mathbb{R})$ be the Langlands decomposition of the parabolic subgroup $P_j(\mathbb{R})$. $M_0(\mathbb{R})A_0(\mathbb{R})$ is a split Cartan subgroup of $G(\mathbb{R})$ and we identify the dual Lie algebra of $A_0$, denoted $\mathfrak{a}_{\mathbb{R}}^*$, with $V = \{(a_1, a_2, a_3) \in \mathbb{R}^3 | a_1 + a_2 + a_3 = 0\}$. The corresponding root system $\Delta(\mathfrak{g}, \mathfrak{a})$ is $\Delta$. Here $M_0(\mathbb{R})$ is the dihedral group $D_4$ while $M_1(\mathbb{R})$ and $M_2(\mathbb{R})$ are isomorphic to $SL_2(\mathbb{R})$. For $\chi$ a unitary character of $\mathbb{A}_0(\mathbb{R})$ let $I_{P_0(\chi)}$ be the spherical constituent of $\mathrm{Ind}_{P_0(\mathbb{R})}^{G(\mathbb{R})}(1_{M_0(\mathbb{R})} \otimes \chi)$. For $j = 1$ or 2 and $\chi_j$ a unitary character of $A_j(\mathbb{R})$ and $0 < \sigma \leq \frac{1}{2}$ a complementary series representation of $M_j(\mathbb{R})$, let $I_{P_j}(\sigma, \chi)$ be the spherical constituent of $\mathrm{Ind}_{P_j(\mathbb{R})}^{G(\mathbb{R})}(\sigma \otimes \chi_j)$. The representations $I_{P_0}(\chi)$ are tempered and as we vary over all unitary $\chi$ these exhaust $\widehat{G(\mathbb{R})}_{\mathrm{temp}}^{\mathrm{sph}}$. The representations $I_{P_j}(\sigma, \chi)$ are nontempered and unitary, they together with the tempered representations exhaust all the nonreal part of $\widehat{G(\mathbb{R})}^{\mathrm{sph}}$ (i.e. the spherical representations with nonreal infinitesimal characters). The rest of $\widehat{G(\mathbb{R})}^{\mathrm{sph}}$ may be described as a subset of $\mathfrak{a}_{\mathbb{R}}^*$ with $\alpha \in \mathfrak{a}_{\mathbb{R}}^*$ corresponding to $\mathrm{Ind}_{P_0(\mathbb{R})}^{G(\mathbb{R})}(1_{M_0(\mathbb{R})} \otimes \exp(\alpha(\cdot)))$. According to Vogan [Vo3] the set of such $\alpha$'s which are unitary is the brown and red shaded region in Figure 1.

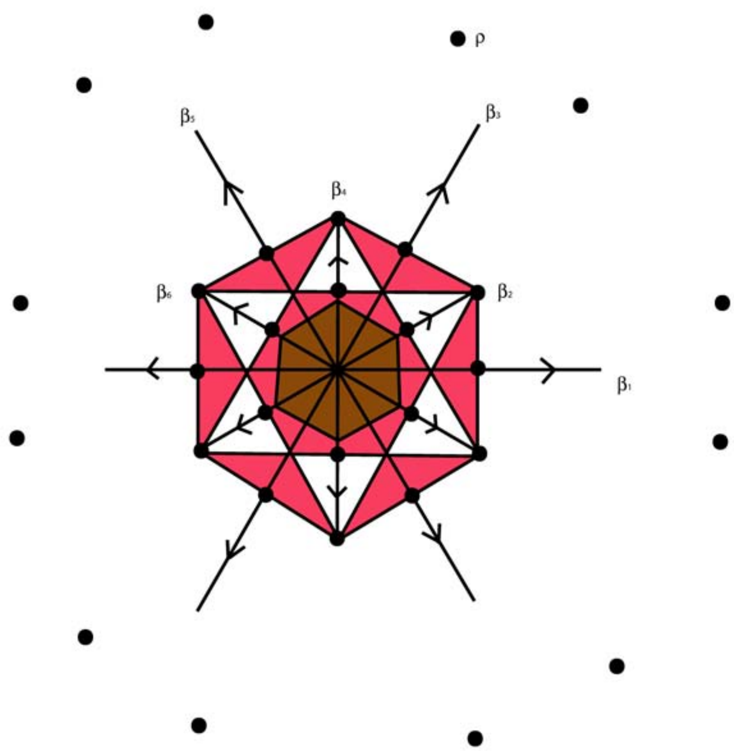



FIGURE 1. The shaded area together with the outside dots yield the spherical unitary dual of $G_2$. The dots (that is $\rho, \beta_6, \beta_1/2, \beta_6/2$ and 0 and their images under $W$) are in $\widehat{G_2(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}}$. The brown shaded region is an upper bound for the generic cuspidal part of $\widehat{G(\mathbb{R})}_{\mathrm{AUT}}$.

Note that points in $\mathfrak{a}_{\mathbb{R}}^*$ equivalent under $W$ correspond to the same point in $\widehat{G(\mathbb{R})}^{\mathrm{sph}}$. The point $\rho$ is half the sum of the positive roots and corresponds to the trivial representation of $G(\mathbb{R})$. Clearly it is isolated in $\widehat{G(\mathbb{R})}$ (as it should be since $G(\mathbb{R})$ has property $T$). Also note that for $0 \leq \sigma \leq 1/2$ and $j = 1, 2$, $I_{P_j}(\sigma, \mathbf{1})$ (which is real) corresponds to the point $\sigma\beta_j$ in $\mathfrak{a}_{\mathbb{R}}^*$.

We turn to $\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}}$. Let

$$C_0 = \{I_{P_0}(\chi) | \chi \text{ is unitary}\} \tag{49}$$

$$C_1 = \left\{I_{P_1}\left(\frac{1}{2}, \chi\right) | \chi \text{ is unitary}\right\} \tag{50}$$

and

$$C_2 = \left\{I_{P_2}\left(\frac{1}{2}, \chi\right) | \chi \text{ is unitary}\right\}. \tag{51}$$

We have the following lower bound

$$\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}} \supset C_0 \cup C_1 \cup C_2 \cup \{\beta_4\} \cup \{\rho\}. \tag{52}$$

Note that the set of points on the right-hand side of (52) meets $\mathfrak{a}_{\mathbb{R}}^*$ in the set of dotted points in Figure 1.

We explain the containment (52). Firstly, the point $\{\rho\}$ is self-evident. Since $C_0 = \widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{temp}}$ its inclusion in (52) follows from (28). One can show the containment of $C_1$ and $C_2$ by a variation of (26) where we allow $H$ to be a parabolic subgroup, specifically $P_1$ and $P_2$ in this case. However, the theory of Eisenstein series demonstrates this more explicitly. Form the Eisenstein series $E_{P_1}(g, s)$ on $G(\mathbb{Q})\backslash G(\mathbb{A}_{\mathbb{Q}})$ corresponding to $P_1$ and with the trivial representation on $M_1^{(1)}$ (where $L_1 = M_1^{(1)} A_1$). $E_{P_1}$ has a meromorphic continuation in $s$ and is analytic on $\Re(s) = 0$ where it furnishes continuous spectrum in $L^2(G(\mathbb{Q})\backslash G(\mathbb{A}_{\mathbb{Q}}))$. The corresponding spherical parameters fill out $C_1$ and place them in $\widehat{G}^{\mathrm{sph}}_{\mathrm{AUT}}$. Similarly the continuous spectrum corresponding to the Eisenstein series $E_{P_2}$ (with the trivial representation on $M_2^{(1)}$) yields $C_2$. The remaining point $\{\beta_4\}$ in (52) is more subtle. Again one can see that it is in $\widehat{G(\mathbb{R})}_{\mathrm{AUT}}$ using (26). The Lie subalgebra of $\mathfrak{g}$ generated by $\mathfrak{a}$ and the root vectors corresponding to the six long roots is of type $A_2$. The corresponding subgroup $H$ of $G$ is $SL_3$ and is defined over $\mathbb{Q}$. $H(\mathbb{R})$ and $G(\mathbb{R})$ are both of rank 2 and they share the split torus $A_0(\mathbb{R})$. Choosing $\beta_1$ and $\beta_5$ as simple positive roots of $\Delta(\mathfrak{h}, \mathfrak{a})$, we find that $\rho_H = \beta_3$. Now,

$$\beta_4 \in \mathrm{Ind}_{H(\mathbb{R})}^{G(\mathbb{R})} \left(\mathrm{Ind}_{H(\mathbb{Z})}^{H(\mathbb{R})} \mathbf{1}\right). \tag{48$'$}$$

This can be shown by considering the density of $H(\mathbb{Z})$ points in expanding regions in $G(\mathbb{R})$, first by examining $H(\mathbb{Z})$ as a lattice in $H(\mathbb{R})$ and second by using (48$'$) (see [Sa2]), the key points being that $\beta_2$ (or $\beta_4$) is an extreme point of the outer hexagon and that $\rho_G = 2\rho_H - \beta_2$. From (48$'$) and (26) it follows that $\beta_4 \in \widehat{G(\mathbb{R})}_{\mathrm{AUT}}$. As before, the Eisenstein series provides a more explicit automorphic realization of $\beta_4$. In fact it occurs as a residue (and hence in the discrete spectrum) of the minimal parabolic Eisenstein series $E_{P_0}(g,s)$ (here $s$ denotes two complex variables). See for example [K2].

The above account for the lower bound (52). It is interesting that there are other residual and even cuspidal spectra which contribute to various points on the right-hand side of (52). The Eisenstein series $E_{P_1,\pi}(g,s)$, where $\pi$ is an automorphic cuspidal representation on $M^{(1)} \cong PGL_2$, has a pole at $s = 1/2$ if the special value $L\left(\frac{1}{2}, \pi, \mathrm{sym}^3\right)$ of the symmetric cube $L$-function is not zero, see [K2]. If $\pi_\infty(\pi = \bigotimes_v \pi_v)$ is spherical and tempered then the corresponding residue on $G(\mathbb{A}_\mathbb{Q})$ produces a point in $C_2$ (for example if $\pi_\infty$ is spherical corresponding to a Maass cusp form with eigenvalue 1/4 then the corresponding point in $\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}}$ is $I_{P_2}\left(\frac{1}{2}, 1\right) = \beta_2/2$, that is,:w the point in the middle of the side of the inner hexagon). Similarly, the Eisenstein series $E_{P_2,\pi}(g,s)$, where $\pi$ is an automorphic cuspidal representation on $M_2^{(1)} \simeq PGL_2$ has a pole at $s = \frac{1}{2}$ if $L(\left(\frac{1}{2}, \pi\right) \neq 0$ (see [K2]). If $\pi_\infty$ is spherical and tempered the residue produces a point in $C_1$ (this time the eigenvalue 1/4 produces the point $\beta_1/2$, i.e. the midpoint of the outer hexagon).

It is a deeper fact that $\{\beta_4\}$ and a dense subset of points in $C_1$ can be produced cuspidally. In [G-G-J], Gan-Gurevich and Jiang show that $S_3 \times G$ can be realized as a dual pair in $H = \mathrm{Spin}(8) \ltimes S_3$. Restricting the automorphic minimal representation of $H(\mathbb{A})$ ([G-R-S]) to $S_3 \times G$ yields a correspondence between automorphic forms on $S_3$ and $G$. The spherical representation $\beta_4$ of $G(\mathbb{R})$ is a constituent of this restriction. Moreover, by comparing what they construct with the multiplicities of the residual spectrum, they show that $\beta_4$ occurs as an archimedian component of a cusp form in $L^2_{\mathrm{cusp}}(G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q}))$. A dense set of points in $C_1$ corresponding to cuspidal representations was constructed by Rallis and Schiffmann [R-S2] using the oscillator representation of $w$. They realize $G \times \widetilde{SL_2}$ as a subgroup of $Sp_{14}$. While this does not form a dual pair they show that nevertheless restricting $w$ to $G \times \widetilde{SL_2}$ yields a correspondence between forms on $\widetilde{SL_2}$ and $G$. In particular, suitable cuspidal representations $\sigma$ of $\widetilde{SL_2}$ are transferred to automorphic cusp forms $\pi(\sigma)$ on $G(\mathbb{A}_\mathbb{Q})$ and the corresponding $\pi(\sigma)_\infty$ lies in $C_1$ (assuming that $\sigma_\infty$ is tempered). For example, choosing $\sigma$ appropriately, one can produce the point $\beta_1/2$ in $\mathfrak{a}^*_\mathbb{R}$ cuspidally. In [G-G-J] and [G-G] the authors compute the Arthur parameters (see (53) below) explicitly corresponding to these cuspidal automorphic forms on $G(\mathbb{A}_\mathbb{Q})$. They find an excellent agreement with the Arthur Conjectures for $G$.

Our discussion above shows that the lower bound (52) is achieved by various parts of the spectrum. Unfortunately, I don't know of any nontrivial upper bounds for $\widehat{G(\mathbb{R})}^{\mathrm{sph}}_{\mathrm{AUT}}$ (either for this $G$ or any other exceptional group, though for generic representations upper bounds are given below). An interesting start would be to establish that $\beta_4$ is isolated in $\widehat{G(\mathbb{R})}_{\mathrm{AUT}}$. The natural conjecture here about this part of the automorphic dual of $G$ is that the inclusion (52) is an equality.

The above are typical examples of the use of dual pairs in constructing automorphic representations and in particular non-tempered ones. As a final example we consider the case of a group of type $F$. We fixate on the problem of cohomology in the minimal degree. Let $F_{4,4}(\mathbb{R})$ be the real split group of type $F_4$ and of rank 4 (see the description and notation in Helgason [He]). The corresponding symmetric space $F_{4,4}(\mathbb{R})\big/Sp(3) \times Sp(1)$ has dimension 28. For $\Gamma$ a co-compact lattice in $F_{4,4}(\mathbb{R})$ the cohomology groups $H^j(\Gamma, \mathbb{C})$ vanish for $0 < j < 8$, $j \neq 4$ (see [V-Z]).

For $j = 4$ the cohomology comes entirely from parallel forms (i.e. from the trivial representation in Matsushima's formula) and so $\dim H^4(\Gamma, \mathbb{C})$ is constant (i.e. independent of $\Gamma$). So the first interesting degree is 8. According to Vogan [Vo1] there is a non-tempered cohomological representation $\psi$, which is isolated in $\widehat{F_{4,4}(\mathbb{R})}$ and which contributes to $H^8(\Gamma, \mathbb{C})$. Now let $G$ be an algebraic group defined over $\mathbb{Q}$ (after restriction of scalars) such that $G(\mathbb{Q}_\infty) \simeq F_{4,4}(\mathbb{R}) \times$ compact and with $G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q})$ compact. Using the classification in [Ti] one can show (see [B-L-S2]) that $G$ contains a symmetric $\mathbb{Q}$ subgroup $H$ such that $H(\mathbb{Q}_\infty) \simeq \mathrm{Spin}_\mathbb{R}(5,4) \times$ compact. According to Oshima's computation of the spectra of the affine symmetric space $F_{4,4}(\mathbb{R})/\mathrm{Spin}_\mathbb{R}(5,4)$ one finds that $\psi$ occurs discretely in $\mathrm{Ind}_{\mathrm{Spin}_\mathbb{R}(5,4)}^{F_{4,4}(\mathbb{R})} 1$. Hence according to (26), $\psi \in \widehat{G(\mathbb{Q}_\infty)}_{\mathrm{AUT}}$. Since $\psi$ is isolated in $F_{4,4}(\mathbb{R})$ it follows that $\psi$ occurs in $L^2(\Gamma\backslash F_{4,4}(\mathbb{R}))$ for a suitable congruence subgroup $\Gamma$ of $G(\mathbb{Z})$. Using the classification of lattices $\Gamma$ (see [Ma]) in $F_{4,4}(\mathbb{R})$ one can show in this way that for any such lattice $\Gamma$ and any $N > 0$ there is a subgroup $\Gamma'$ of finite index in $\Gamma$ such that $\dim H^8(\Gamma, \mathbb{C}) > N$ (see [B-L-S2]). For a survey of results concerning nonvanishing of cohomology in the minimal degree see [Li-Sc].

This concludes our list of examples. We return to the general $G$. In Example 2, the upper bound (37) implies the useful fact that the trivial representation **1** is isolated in $\widehat{G_f(\mathbb{R})}_{\mathrm{AUT}}$. It was conjectured by Lubotzky and Zimmer that this feature is true in general. That is, if $G$ is a semi-simple group defined over $F$ then the trivial representation is isolated in $\widehat{G(F_v)}_{\mathrm{AUT}}$ for any place $v$ of $F$ (they called this property $\tau$). Of course, if $v$ is a place at which $G(F_v)$ has property $T$ then there is nothing to prove. Clozel [Cl1] has recently settled this property $\tau$ conjecture, this being the first general result of this kind. One proceeds by exhibiting (in all cases where $G(F_v)$ has rank 1 for some place $v$) an $F$ subgroup $H$ of $G$ for which the isolation property is known for $\widehat{H(F_v)}_{\mathrm{AUT}}$ and hence by the restriction principle (27) this allows one to deduce the isolation property for $\widehat{G(F_v)}_{AUT}$. For example if $G$ is isotropic then such an $H$ isomorphic to $SL_2$ (or $PGL_2$) can be found. Hence by (19) and (20) the result follows. If $G$ is anisotropic then he shows that $G$ contains an $F$ subgroup $H$ isomorphic to $SL(1, D)$ with $D$ a division algebra of prime degree over $F$ or $SU(D, \alpha)$, a unitary group corresponding to a division algebra $D$ of prime degree over a quadratic extension $E$ of $F$, and with $\alpha$ an involution of the second kind (cf Example 3 above). Thus one needs to show that the isolation property holds for these groups. For $SL(1, D)$ this follows the generalized Jacquet-Langlands correspondence [A-C] and the bounds (14) and (15) (for $GL_p(\mathbb{A}_F)$ with $p$ prime there are no non-trivial residual of Eisenstein series so the discrete spectrum is cuspidal). For the above unitary groups $G = SU(D, \alpha)$ of prime degree Clozel establishes the base change lift from $G$ over $F$ to $G$ over $E$ (this being based on earlier works by Kottwitz, Clozel and Labesse). Now $G$ over $E$ is essentially $SL(1, D)$ over $E$ so one can proceed as above. As Clozel points out, it is fortuitous that these basic cases that one lands up with are among the few for which one can stabilize the trace formula transfer at present.

It is of interest (see comment 3 of Section 3) to know more generally which $\pi_v$'s are isolated in $\widehat{G(F_v)}_{\mathrm{AUT}}$? In this connection a natural conjecture is that if $G(F_v)$ is of rank 1 then every non-tempered point of $\widehat{G(F_v)}_{\mathrm{AUT}}$ is isolated.

At the conjectural level, Arthur's Conjectures [A2] give very strong restrictions (upper bounds) on $\widehat{G}_{\mathrm{AUT}}$. While these conjectures involve the problematic group $L_F$, they are functorial and localizing them involves the concrete group[§] $L_{F_v}$ and its representations. In this way these conjectures impose explicit restrictions on the automorphic spectrum. For example, if $G$ is a split group over $F$ then the local components $\pi_v$ of an automorphic representation $\pi$ occurring discretely in $L^2(G(F)\backslash G(\mathbb{A}_F))$ must correspond to certain Arthur parameters. In the unramified case these are morphisms of the local Weil group times $SL_2(\mathbb{C})$ into ${}^LG$ satisfying further properties. That is,

$$\psi : W_{F_v} \times SL(2,\mathbb{C}) \longrightarrow {}^LG \tag{53}$$

such that

(i) $\psi\big|_{W_{F_v}}$ is unramified and $\psi(\mathrm{Frob}_v)$ lies in a maximal compact subgroup of ${}^LG$ (c.f. (9)).

(ii) If $j$ is the unramified map of $W_{F_v} \longrightarrow SL(2,\mathbb{C})$ which sends $\mathrm{Frob}_v$ to

$$\begin{bmatrix} N(v)^{1/2} & 0 \\ 0 & N(v)^{1/2} \end{bmatrix},$$

then the corresponding Arthur parameter is the conjugacy class $\psi(\mathrm{Frob}_v,\ j\ (\mathrm{Frob}_v))$ in ${}^LG$.

Thus the $SL(2,\mathbb{C})$ factor in (49) allows for non-tempered parameters but they are highly restricted.

In many of these split examples these local restrictions are probably even sharp and hence yield precise conjectures for $\widehat{G(F_v)}_{\mathrm{AUT}}$. We note however that it is by no means clear that the upper bounds imposed on $\widehat{G}_{\mathrm{AUT}}$ by Arthur's parameters are consistent, for example, with the lower bound (26) which must hold for all subgroups $H$. Establishing this would be of interest. As Clozel [Cl3] has shown, the Arthur Conjectures together with (26) and (27) (in the form extended to $\widehat{G(S)}_{\mathrm{AUT}}$) lead to some apparently non-obvious statements and structures for unipotent representations of local groups. Assuming a general twisted form of the "Fundamental Lemma" (see Hales' lectures), Arthur [A4], using the trace formula, gives a precise transfer of automorphic forms from classical orthogonal and symplectic groups to the corresponding general linear group. Hence, if and when this fundamental lemma is established, one will be able to combine this transfer with the bounds of Section 1 to get new sharp upper bounds for $\widehat{G}_{\mathrm{AUT}}$ with $G$ classical.

In the meantime, when $G$ is split over $F$ and the representation of $\pi$ of $G(\mathbb{A}_F)$ is cuspidal and generic, there have been some impressive developments along the lines of such functorial lifts. Here $\pi$ being generic means that there is an $f$ in the space of $\pi$ such that $\int_{U(F)\backslash U(\mathbb{A})} f(ug)\,\psi(u)du \neq 0$ for some character $\psi$ of a maximal unipotent subgroup $U$ of $G$. Using these lifts one can deduce strong upper bounds for the part of $\widehat{G}_{\mathrm{AUT}}$ which corresponds to globally generic cuspidal $\pi$'s. Indeed, the formulation of the generalized Ramanujan Conjectures for such representations takes the simple form that it does in $GL_n$.

*GRC* (cuspidal generic): (see [H-PS])

[§] $L_{F_v}$ is simply $W_{F_v}$ if $v$ is archimedian and is $W_{F_v} \times SU(2,\mathbb{R})$ if $v$ is finite.

Let $G$ be a quasi-split group defined over $F$. If $\pi \simeq \underset{v}{\otimes} \pi_v$ is a globally generic automorphic cuspidal representation of $G(\mathbb{A})$ then $\pi_v$ is tempered.

The main progress for functoriality for generic representations is due to Cogdell-Kim-Piatetski-Shapiro and Shahidi ([C-K-PS-S], see also [K-K] and [A-S]). Their work is concerned with a split classical group $G$, that is one of $SO_{2n+1}$, $SO_{2n}$ or $Sp_{2n}$. The corresponding dual groups ${}^LG^0$ are $Sp_{2n}(\mathbb{C})$, $SO_{2n}(\mathbb{C})$ and $SO_{2n+1}(\mathbb{C})$ respectively. These dual groups have a standard representation in $GL_{2n}(\mathbb{C})$, $GL_{2n}(\mathbb{C})$ and $GL_{2n+1}(\mathbb{C})$ and hence there should be a corresponding functorial lift from $G(\mathbb{A})$ to $GL(\mathbb{A})$. In [C-K-PS-S] it is shown that if $\pi \simeq \underset{v}{\otimes} \pi_v$ is an automorphic cuspidal generic representation of $G(\mathbb{A})$ then this lift to an automorphic form on $GL(\mathbb{A})$ exists. The lift is explicit and one can analyze its local components. It follows that if one assumes the *GRC* for $GL_N$ then the $\pi_v$'s above are tempered (that is, the *GRC* (cuspidal generic) for $G$ follows from *GRC* for $GL_N$). Moreover, using the results described in Section 1, specifically (14) and (15) one obtains corresponding sharp bounds towards the *GRC* (generic cuspidal) for such $G$'s (see [C-K-PS-S], [A-S] and [K-K]).

Combining the results above with work of Ginzburg and Jiang ([G-J]) which establishes the functorial transfer of generic cusp forms on $G_2(\mathbb{A})$ (as in example 4 above) to $GSp_6$, one obtains similar upper bounds for the part of $\widehat{G_2(\mathbb{Q}_v)}_{\mathrm{AUT}}$ that comes from generic cuspidal automorphic representations $\pi$ of $G_2(\mathbb{A})$. For example, any such $\pi$ for which $\pi_\infty$ is spherical must have the real part of its parameters lie in the brown region shaded in Figure 1, which consists of points lying in the hexagon with vertices $\frac{49}{100}\beta$, where $\beta$ is a short root. This comes about from the functorial lift

$$G_2(\mathbb{C}) = ({}^LG_2^0) \hookrightarrow SO_7(\mathbb{C}) = ({}^LSp_6^0) \hookrightarrow GL_7(\mathbb{C}) = ({}^LGL_7^0)$$

and the bound (15) with $n = 7$. Note that the dotted points in Figure 1 which are all in $\widehat{G_2(\mathbb{R})}_{\mathrm{AUT}}$, are nongeneric (except for 0 which is tempered) and hence they do not contradict *GRC* (generic cuspidal).

These upper bounds on the generic cuspidal spectrum are quite a bit better than the local bounds that one gets by identifying the generic unitary duals of the classical groups [L-M-T] and of $G_2$ [Ko]. The proofs of the functorial lifts of generic cusp forms from the split classical groups to $GL$ are based on the Langlands Shahidi method and the converse theorem while the transfer of such forms from $G_2$ to $GSp_6$ relies on these forming a dual pair in $E_7$.

## 3. Applications

The Ramanujan Conjectures and their generalizations in the form that we have described them, and especially the upper bounds, have varied applications. We give a brief list of some recent ones.

(1) For $GL_2/F$ there are applications to the problem of estimation of automorphic $L$-functions on their critical lines and especially to the fundamental "sub-convexity" problem. See [I-S] and [Sa] for recent accounts as well as for a description of some of the applications of sub-convexity.

(2) The problem of counting asymptotically integral and rational points on homogeneous varieties for actions by semi-simple and reductive groups as well as the equi-distribution of "Hecke Orbits" on homogeneous spaces,

depends directly on the upper bounds towards $GRC$. For recent papers on these topics see [Oh], [C-O-U], [G-O], [S-T-T] and [G-M] and also [Sa2].

(3) There have been many works concerning geometric constructions of cohomology classes in arithmetic quotients of real and complex hyperbolic spaces. Bergeron and Clozel have shown that the injectivity of the inclusion and restriction of cohomology classes associated with $H < G$ (here $H$ and $G$ are $SO(n,1)$ or $SU(m,1)$) can be understood in terms of the isolation properties of these cohomological representations in $\widehat{G(F_\infty)}_{\mathrm{AUT}}$. This allows for an elegant and unified treatment of the constructions of cohomology classes as well as far reaching extensions thereof. They have also established the isolation property for some unitary groups. See [Be] and the references therein.

(4) Müller and Speh [M-S] have recently established the absolute convergence of the spectral side of the Arthur trace formula for $GL_n$. Their proof requires also the extension of (14) and (15) to ramified representations of $GL_n(F_v)$, which they provide. Their work has applications to the construction of cusp forms on $GL_n$ and in particular to establish that Weyl's Law holds for the cuspidal spectrum.

(5) An older application to topics outside of number theory is to the construction of highly connected but sparse graphs ("Ramanujan Graphs"). These applications as well as ones related to problems of invariant measures are described in the monograph of Lubotzky [Lu]. The property $\tau$ conjecture mentioned in Section 2 is related to such applications.

For a discussion of the automorphic spectral theory of $GL_2(\mathbb{A}_\mathbb{Q})$ in classical language see [Sa]. The recent article [Cl2] is close in flavor to these notes and should be consulted as it goes into more detail at various places.

Acknowledgements: I thank J. Arthur, J. Cogdell, N. Gurevich, E. Lapid, J.S. Li, H. Oh, S. Rallis and J.P. Serre for comments and discussions concerning these notes and G. Pecht for her technical assistance.

## 4. References

[A] J. Arthur, "The principle of functoriality," *BAMS* Vol. 40, No 1, (2003), 39-53.

[A2] J. Arthur, "Unipotent automorphic representations: conjectures. Orbites unipotentes et reprsentations, I," *Asterisque*, 171-172, (1989), 13-71.

[A3] J. Arthur, "A note on the automorphic Langlands group," *Canadian Math. Bull.*, **45**, (2002), 466-482.

[A4] J. Arthur, Lecture Series I.A.S., (2001).

[A-C] J. Arthur and L. Clozel, "Simple algebras, base change, and the advanced theory of the trace formula," *Ann. Math. Studies*, (1989).

[A-S] M. Asgari and F. Shahidi, "Generic transfer for general spin groups," preprint, (2004).

[Be] N. Bergeron, "Lefschetz properties for arithmetic real and complex hyperbolic manifolds," *IMRN*, **20**, (2003), 1089-1122.

[B-S] M. Burger and P. Sarnak, "Ramanujan duals. II," *Invent. Math.*, **106**, (1991), 1-11.

[B-L-S1] M. Burger, J.S. Li and P.Sarnak, "Ramanujan duals and automorphic spectrum," *BAMS*, (1992), 253-257.

[B-L-S2] M. Burger, J.S. Li and P. Sarnak, "Ramanujan duals and automorphic spectrum III," (in preparation).

[B-W] A. Borel and N. Wallach, Continuous cohomology, discrete subgroups, and representations of reductive groups, *Ann. Math. Studies*, **94**, (1980).

[B-C] N. Bergeron and L. Clozel, "Spectre et homologie des varits hyperboliques complexes de congruence," *C.R. Acad. Sci.* Paris Ser. I, **334** (2002), 995-998.

[Cl1] L. Clozel, "Dmonstration de la conjecture $\tau$," *Invent. Math.*, **151**, (2003), 297-328.

[C12] L. Clozel, "Spectral theory of automorphic forms," (2003), (preprint).

[C13] L. Clozel, "Combinatorial consequences of Arthur's conjectures and the Burger-Sarnak method." *Int. Math. Res. Not.* (2004), **11**, 511-523.

[C-U] L. Clozel and E. Ullmo, "Equidistribution des points de Hecke,", *Contributions to automorphic forms, geometry and number theory*, Johns Hopkins University Press, (2004), 193-254.

[C-O-U] L. Clozel, H. Oh and E. Ullmo, "Hecke operators and equidistribution of Hecke points," *Invent. Math.*, **144**, (2001), 327-351.

[Co] J. Cogdell, Fields Institute Lectures on $GL_n$ and converse theorems, (2003).

[Con] B. Conrad, "Modular forms, cohomology and the Ramanujan Conjectures," *CUP*, (to appear).

[C-K-PS-S] J. Cogdell, H. Kim, I. Piatetski-Shapiro and F. Shahidi, "Functoriality for the classical groups," Publ. IHES, **99**, (2004), 163-233.

[de-W] D. de George and N. Wallach, "Limit formulas for multiplicities in $L^2(\Gamma\backslash G)$," *Annals of Math.*, **107**, (1978), 133-150.

[D-I] W. Duke and H. Iwaniec, "Automorphic forms and analytic number theory," (1989), *CRM*, Montreal, (1990), 43-47.

[Di] J. Dixmier, "Les $C^*$-algebras et leurs representations," Paris, (1969).

[G-M] D. Goldstein and A. Mayer, "On the equidistribution of Hecke points," *Forum Math.*, **15**, (2003), 165-189.

[G-O] W. Gan and H. Oh, "Equidistribution of integer points on a family of homogeneous varieties: a problem of Linnik," *Compositio Math.*, Vol. 323, (2003), 323-352.

[Ge] S. Gelbart, "Automorphic forms on adle groups," *Annal. of Math. Studies*, Vol. 83, (1975).

[G-G] W. Gan and N. Gurevich, "Non-tempered $A$ packets of $G_2$: Liftings from $\overline{SL_2}$," (2003), (preprint).

[G-G-J] W. Gan, N. Gurevich and D. Jiang, "Cubic unipotent Arthur parameters and multiplicities of square integrable automorphic forms," *Invent. Math.*, **149**, (2002), 225-265.

[G-J] D. Ginzburg and D. Jiang, "Periods and liftings from $G_2$ to $C_3$," *Israel Jnl. Math.*, **123**, (2001), 29-59.

[G-R-S] D. Ginzburg, D. Soudry and S. Rallis, "On the automorphic theta representation for simply laced groups," *Israel Jnl. Math.*, **100**, (1997), 61-116.

[H-T] M. Harris and R. Taylor, The Geometry and Cohomology of Some Simple Shimura Varieties, *Annals. of Math. Studies*, **151**, (2001).

[H-M] R. Howe and C. Moore, "Asymptotic properties of unitary representations," *Jnl. Funct. Anal.*, **32**, (1979), 72-96.

[Ho] R. Howe, Notes on the oscillator representation, (unpublished).

[H-PS] R. Howe and I. Piatetski-Shapiro, *Proc. Symp. Pure Math.*, **33**, (1975), 315-322.

[He] S. Helgason, "Differential Geometry, Lie Groups and Symmetric Spaces," A.P., (1978).

[I-S] H. Iwaniec and P. Sarnak, "Perspectives on the analytic theory of $L$-functions," *GAFA*, (2000), 705-741.

[J-PS-S] H. Jacquet, and J. Shalika, "On Euler products and the classification of automorphic forms. II," *Am. Jnl. of Math.*, **103**, (1981), 499-558.

[J-W] K.D. Johnson and N. Wallach, "Composition series and intertwining operators for the spherical principal series. I," *T.A.M.S.*, **229**, (1977), 137-173.

[K-K] H. Kim and M. Krishnamurthy, "Stable base change lift from unitary groups to $GL_N$," preprint, (2004).

[K-S] H. Kim and F. Shahidi, "Functorial products for $\mathrm{GL}_2 \times \mathrm{GL}_3$ and the symmetric cube for $\mathrm{GL}_2$," *Ann. of Math.*, **155**, (2002), 837-893.

[K] H. Kim, " Functoriality for the exterior square of $GL_4$ and the symmetric fourth of $GL_2$," *JAMS*, Vol. 16, **1**, (2002), 139-183.

[K2] H. Kim, "The residual spectrum of $G_2$." *Can. J. Math.*, Vol. 48, (1996), 1245-1272.

[Ko] B. Kostant, "On Whittaker vectors and representation theory." *Invent. Math.*, (1978), 101-184.

[Ku] N. Kurokawa, "Examples of eigenvalues of Hecke operators on Siegel cusp forms of degree two." *Invent. Math.*, **49**, (1978),149-165.

[Ka-Sav] D. Kazhdan and G. Savin, "The smallest representation of simply laced groups." *Israel Math. Conf. Proc.*, **2**, (1990), 209-223.

[Ki-Sa] H. Kim and P. Sarnak, Appendix to [K].

[K-Z] A. Knapp and G. Zuckerman, "Classification of irreducible tempered representations of semisimple groups." *Ann. of Math.*, **116**, (1982), 389-500.

[Li1] J.S. Li, "The minimal decay of matrix coefficients for classical groups." *Math. Appl.*, **327**, Kluwer, (1995), 146-169.

[Li2] J.S. Li, "Minimal Representations and Reductive Dual Pairs," *IAS*, Park City Math. Series, Vol. 8, (2000).

[Lang1] R. Langlands "Automorphic representations, Shimura varieties, and motives" in A.M.S., *Proc. Sym. Pure Math.*, Vol. XXXIII, part 2, (1979), 205-246.

[Lang2] R. Langlands, "Problems in the theory of automorphic forms," Springer Lecture Notes, Vol. 170, (1970), 18-61.

[Lu] A. Lubotzky, "Discrete groups, expanding graphs and invariant measures," Birkhauser, (1994).

[L-R-S1] W. Luo, Z. Rudnick and P. Sarnak, "On the generalized Ramanujan conjecture for GL($n$)." *Proc. Sym. Pure Math.*, **66**, Vol. II, (1999), 301-311.

[L-R-S2] W. Luo, Z. Rudnick and P. Sarnak, "On Selberg's eigenvalue conjecture." *GAFA*, **5**, (1995), 387-401.

[Lan] E. Landau, *Nachr. Ge Wiss Göttingen*, (1915), 209-243.

[Li-Sc] J.S. Li and J. Schwermer, "Automorphic representations and cohomology of arithmetic groups," In Challenges for the 21st Century, *World Scientific*, (2001), 102-138.

[L-M-T] E. Lapid, G. Muić and M. Tadić, "On the generic unitary dual of quasi-split classical groups," (2003), preprint.

[Ma] G. Margulis, *Ergeb. Math.*, Grenzeb(3), Vol. 17, Springer, (1989).

[M-S] W. Müller and B. Speh, "Absolute convergence of the spectral side of the Arthur trace formula for $GL_n$." *Geom. Funct. Anal.*, **14**, (2004), 58–93.

[M-W] C. Moeglin and L. Waldspurger, "Le spectre rsiduel de GL($n$)." *Ann. Ecole Norm. Sup*, **22**, (1989), 605-674.

[Oh] H. Oh, "Harmonic Analysis and Hecke Operators," in Rigidity in Dynamics and Geometry, (2002).

[O-M] T. Oshima and T. Matsuki, "A description of discrete series for semisimple symmetric spaces." *Adv. Studies in Pure Math*, **4**, (1984), 331-390.

[Ro] J. Rogawski, Automorphic representations of unitary groups in three variables. *Ann. Math. Studies*, **123**, (1990).

[R-S1] S. Rallis and Schiffmann, "Discrete spectrum of the Weil representation,"*BAMS*, **83**, (1977),Theta correspondence associated to $G_2$ 267-270.

[R-S2] S. Rallis and G. Schiffmann, "Theta correspondence associated to $G_2$." *Amer. Jnl. of Math.*, **111**, (1989), 801-849.

[Roh] D. Rohrlich, "Nonvanishing of $L$-functions for GL(2)." *Invent. Math.*, **97**, (1989), 381-403.

[Sa] P. Sarnak, www.math.princeton.edu/˜sarnak file baltimore.

[Sa2] P. Sarnak, "Diophantine problems and linear groups," In Proceedings of the *I.C.M.*, Kyoto, (1990), Vol. 1, 459-471.

[Sel] A. Selberg, "On the estimation of Fourier coefficients of modular forms." *Proc. Symp. Pure Math.*, VIII, (1965), 1-15.

[Ser] J. P. Serre, Letter to Jacquet, (1981).

[Ser2] J.P. Serre, " Cohomologie des groupes discrets." Annals of Math Studies, no. 70, (1971), 77-169, see also Collected Works Vol II, 593-685.

[Sh] F. Shahidi, "Automorphic L-functions: A Survey," in Automorphic Forms, Shimura Varieties and L-functions, Editors, L. Clozel and J. S. Milne, Perspectives in Math., Academic Press, (1990), 415-437

[Sat] I. Satake, *Proc. Symp. AMS*, **9**, (1971), 258-264.

[Shi] G. Shimura, "On the holomorphy of certain Dirichlet series." *Proc. London Math. Soc.*, **31**, (1975), 79-98.

[S-T-T] J. Shalika, R. Takloo-Bighash and Y. Tschinkel, "Rational points and automorphic forms," Contributions to automorphic forms, geometry, and number theory, 733–742, Johns Hopkins Univ. Press, Baltimore, MD, 2004.

[Ta] J. Tate, "Number theoretic background." Automorphic forms, representations and $L$-functions *Proc. Symp. Pure Math.*, Vol. XXXIII, part 2, (1979), 3-26.

[Ti] J. Tits, *Proc. Symp. Pure Math.*, **IX**, (1967), 33-62.

[Ven] T.N. Venkataramana, "A remark on extended congruence subgroups." *IMRN*, No. 15, (1999), 835-838.

[Vo1] D. Vogan, "Isolated unitary representations," (1994), (preprint).

[Vo2] D. Vogan, "Irreducibility of discrete series representations for semisimple symmetric spaces." *Advanced Studies in Pure Math*, **14**, (1988), 191-221.

[Vo3] D. Vogan, "The unitary dual of $G_2$." *Invent. Math.*, **116**, (1994), 677-791.

[V-Z] D. Vogan and G. Zuckerman, "Unitary representations with nonzero cohomology." *Compositio Math.*, **53**, (1984), 51-90.

[Wa] N. Wallach, "On the Selberg trace formula in the case of compact quotient." *BAMS*, Vol. 82, No. 2, (1976), 171-195.

[Wal] L. Waldspurger, "Sur les coefficients de Fourier des formes modulaires de poids demi-entier." *J. Math. Pures et Appl.*, **60**, (1981), 375-484.

[We] A. Weil, "Sur certains groupes d'oprateurs unitaires." *Acta Math.*, **111**, (1964), 143-211.



Department of Mathematics, Princeton University and the Courant Institute of Mathematical Sciences.

*E-mail address:* `sarnak@math.princeton.edu`

## List of Participants

Feryal Alayont
University of Minnesota

Andrew Archibald
McGill University

Mahdi Asgari
Institute for Advanced Study

Baskar Balasubramanyam
Brandeis University

Dubravka Ban
Southern Illinois University

Eliot Brenner
Yale University

Farrell Brumley
Princeton University

Ping Shun Chan
The Ohio State University

Benjamin Charbord
Université de Strasbourg

Jy-Ying Chen
Harvard University

CheeWhye Chin
Massachusetts Institute of Technology

Alina Cojocaru
Princeton University

Paul-Olivier Dehaye
Stanford University

Mladen Dimitrov
Université de Paris 13

Jyotsna Diwadkar
University of Pittsburgh

Brent Doran
Princeton University

Jens Funke
New Mexico State University

Hans Gmasz
University of Vienna

Julia Gordon
University of Toronto

Ulrich Görtz
Universität Bonn

Gerald Gotsbacher
University of Vienna

Shamgar Gurevich
Tel-Aviv University

Ronny Hadani
Tel-Aviv University

Tobias Hahn
Humboldt-Universität zu Berlin

Harald Helfgott
Yale University

Fritz Hoermann
University of Heidelberg

Benjamin Howard
Harvard University

Joseph Hundley
Pennsylania State University

Joshua Kantor
Arizona State University

Payman Kassaei
McGill University

Jukka Keränen
University of Pittsburgh

Wook Kim
Purdue University

Bruno Klingler
Yale University

Andrew Knightly
University of Maine

Jonathan Korman
University of Michigan

Muthukrishnan Krishnamurthy
University of Michigan

Abhinav Kumar
Harvard University

Wentang Kuo
University of Waterloo

Joshua Lansky
Bucknell University

Erez Lapid
Hebrew University of Jerusalem

Kyu-Hwan Lee
University of Toronto

Catherine Leigh Lucarelli
University of Chicago

Chun Che Li
University of California, Los Angeles

Hristina Lokvenec-Guleska
University Utrecht

Luis Lomeli
Purdue University

Vincent Lucarelli
University of Chicago

Christopher Malon
Massachusetts Institute of Technology

Kevin McGerty
Institute for Advanced Study

Mark McKee
Purdue University

Paul Mezo
Carleton University

Alberto Minguez
Université de Paris 11

David Nadler
University of Chicago

Gia-Vuong Nguyen-Chu
Université de Paris 7

Marc-Hubert Nicole
McGill University

Chu-Feng Nien
University of Minnesota

Omer Offen
Max-Planck Institute for Mathematics, Bonn

David Ogilvie
University of Chicago

Aftab Pande
Brandeis University

Jeehoon Park
Boston University

Manish Patnaik
Yale University

David Pollack
Wesleyan University

Amritanshu Prasad
Université de Montreal

Yannan Qiu
Columbia University

Nicholas Ramsey
Harvard University

Yiannis Sakellaridis
Stanford University

Sug Woo Shin
Harvard University

Lior Silberman
Princeton University

Brian Smithling
University of Chicago

Claus Sorensen
California Institute of Technology

Steven Spallone
University of Chicago

Jonathan Sparling
University of Toronto

Loren Spice
University of Michigan

Alexander Stasinski
University of Nottingham

Lee Stemkoski
Dartmouth College

Binyong Sun
Hong Kong University of Science and Technology

Taliesin Sutton
University of Wisconsin, Madison

Haryono Tandra
The University of Adelaide

Joaquin-Teruji Thomas
University of Chicago

Ngo Dac Tuan
Université de Paris 11

Ilya Tyomkin
Tel Aviv University

Anandavardhanan U.K.
University of Hyderabad

Akshay Venkatesh
Massachusetts Institute of Technology

Filippo Viviani
University of Rome "Tor Vergata"

Inken Vollaard
University of Bonn

Eric Wambach
Columbia University

Martin Weissman
University of California, Berkeley

Chia-Fu Yu
Academia Sinica

Qin Yujun
Hong Kong University of Science and Technology