Simons Symposia

Werner Müller
Sug Woo Shin
Nicolas Templier *Editors*

Families of Automorphic Forms and the Trace Formula

Springer

Simons Symposia

Werner Müller • Sug Woo Shin • Nicolas Templier
Editors

Families of Automorphic Forms and the Trace Formula

Springer

Editors
Werner Müller
Mathematical Institute
University of Bonn
Bonn, Nordrhein-Westfalen
Germany

Sug Woo Shin
Department of Mathematics
University of California, Berkeley
Berkeley, CA, USA

Nicolas Templier
Department of Mathematics
Cornell University
Ithaca, NY, USA

ISSN 2365-9564 ISSN 2365-9572 (electronic)
Simons Symposia
ISBN 978-3-319-41422-5 ISBN 978-3-319-41424-9 (eBook)
DOI 10.1007/978-3-319-41424-9

Library of Congress Control Number: 2016953477

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

Preface

The Simons symposium on families of automorphic forms and the trace formula took place in Puerto Rico from January 26th through February 1st of 2014. It was an opportunity to study families of automorphic representations of higher-rank groups with the goal of paving the way for future developments. We explored the trace formula, spectra of locally symmetric spaces, p-adic families, and other recent techniques from harmonic analysis and representation theory. Experts of different specialties discussed these topics together.

There were 23 participants. Background material has circulated in advance of the symposium, with the idea of focusing during the symposium on recent developments and conjectures toward the frontier of current knowledge. In addition to regular talks, open discussion sessions were scheduled daily for 1 h to promote in-depth exchanges. A different moderator was assigned to each session. The respective themes were: counting cohomological forms, p-adic trace formulas, Hecke fields, slopes of modular forms, and orbital integrals. The goal of each session was to isolate key difficulties and assess the feasibility of diverse approaches.

We hope that the activities of the symposium and the resulting 13 articles of this proceedings volume will be inspiring to participants and researchers in the field. Each article has been thoroughly refereed. Some articles contain original results that have not appeared before, some articles are a synthesis of current knowledge and future directions, and others are survey articles.

The symposium was made possible by the endeavor of the Simons Foundation which we would like to thank again for its generous support. We thank Yuri Tschinkel and Meghan Fazzi for their constant assistance in the organization. We thank the authors for contributing articles to these proceedings and also wish to thank the anonymous referees. Finally we thank Springer-Verlag for their help in publishing these proceedings.

Introduction

The symposium explored analytic, p-adic, and geometric perspectives on families of automorphic forms and the trace formula. An emphasis was on promoting the study of families on higher-rank groups, which was timely in view of recent spectacular developments in the Langlands program.

The Arthur–Selberg trace formula is one of the most important and fundamental tools in the theory of automorphic forms. Besides its indispensable role in reciprocity and functoriality, the trace formula is used to count automorphic forms and to globalize local representations to global automorphic forms, which has numerous applications. It continues to motivate a wide range of techniques in representation theory, in differential and algebraic geometry, and in analysis.

It has been a fruitful idea to study families when solving difficult problems, even if the problem concerns a single object. In the context of number theory, one can study an object, whether it is a variety, a representation, or an L-function, by deforming it in families. In deforming automorphic forms, harmonic families arise such as Dirichlet characters, holomorphic modular forms, Maass forms, Siegel modular forms, and automorphic representations with prescribed local components. The trace formula is essential in conceptualizing harmonic families and establishing their structural properties, such as the Sato–Tate equidistribution which generalizes the Weyl law, and limit multiplicities.

The study of families has taken a new turn in the last two decades with the advent of the Katz–Sarnak heuristics. For this and in other numerous applications of families to sieving, arithmetic statistics, zero-density estimates, L-values, diophantine equations, equidistribution of arithmetic cycles, the trace formula is again a key tool. Already in its most primitive version for GL(1) as the Poisson summation formula, it enters the theory of the distribution of prime numbers. This proceedings volume contributes to sharpening our knowledge of families and the trace formula with the expectation that it will drive new applications.

The trace formula is essential in the local Langlands correspondence and functoriality, starting from the work of Jacquet–Langlands and culminating in the work of Arthur on classical groups. For other applications, such as the ones mentioned above and many others, it is essential to allow a large class of test functions in order to get the most spectral information out of the trace formula. To this end, a number of deep problems in analysis need to be solved. On the spectral side, one has to deal with logarithmic derivatives of intertwining operators, which are the main ingredients of the terms associated to the Eisenstein series. On the geometric side, the singularities of orbital integrals play an important role, in addition to the volume terms which carry much of the arithmetic.

Toward the long-term goals of the subject, it is important to develop systematic ways to work with the local and global trace formulas, orbital integrals, trace characters, Plancherel measures, and other techniques from harmonic analysis and geometry. These themes have been developed separately over the years, and are now coming together.

As a quick guide for the reader, we give below a brief overview of each article in this volume and group them into the following four broad categories: geometric side, local representation theory, harmonic families, and p-adic families.

The geometric side of the trace formula has a rich arithmetic, algebraic, and combinatorial structure which has been studied for several decades. Arthur's fine expansion in weighted orbital integrals has opened the way to stabilization, endoscopic classification, and the fundamental lemma, which all have been achieved recently. Many more questions are now under investigation such as uniform expansions for test functions of non-compact support, a description of the global constants which are weighted generalizations of Tamagawa numbers, relations with the local trace formula, and analogues for function fields.

The article by Werner Hoffmann presents an approach to partition the geometric side according to a new equivalence relation which is finer than geometric conjugacy. The terms are then expressed in terms of certain prehomogeneous zeta functions. Supported by evidence coming from low-rank groups, several conjectures are stated with a view toward future developments.

The article by Jasmin Matz constructs a zeta function associated to the adjoint action of $\mathrm{GL}(n)$ on its Lie algebra. This zeta function is related to the Arthur–Selberg trace formula applied to certain non-compactly supported test functions. For $n = 2$ it coincides with Shintani's zeta function, and for $n = 3$ it is used to obtain results toward an asymptotic formula for the sum of residues of Dedekind zeta functions of families of real cubic fields.

Global orbital integrals factor as a product of local orbital integrals, thus generating interesting problems over local fields, Archimedean and non-Archimedean, of zero and positive characteristics. The solution of these problems involves a variety of techniques at the crossroad of harmonic analysis, algebraic geometry, and geometric representation theory.

The article by Jim Arthur develops a theory of germ expansions for weighted orbital integrals for real groups, thereby extending the pioneering work of Harish-Chandra in the unweighted case. These results will be useful for future investigations of invariant distributions and weighted orbital integrals, objects that are crucial in understanding the trace formula.

Motivic integration has its roots in quantifier elimination, resolution of singularities, and analytic continuation of Igusa integrals. It can be used to prove the transfer principle for the fundamental lemma, asserting that the matching of orbital integrals over a local field of equal characteristic is equivalent to the one over a local field of mixed characteristic. The article by Raf Cluckers, Julia Gordon, and Immanuel Halupczok concerns a related problem of uniform bounds for orbital integrals on p-adic groups as one varies the prime p, the conjugacy class, and the test function. One motivation comes from establishing the Sato–Tate equidistribution for families.

The spectral side of the trace formula consists of characters and weighted characters, which may be studied by local methods, and also terms of genuinely global nature, most notably the multiplicity of automorphic representations. Thanks to the works by Arthur, Moeglin-Waldspurger and others, we have the stabilization of the (twisted) trace formula, opening the doors for the full endoscopic classification of automorphic representations. Such a classification has been accomplished for quasi-split classical groups and anticipated for more groups in the near future. As a consequence, we have a deeper understanding of characters of reductive groups over local fields by relating characters of two different groups via endoscopic identities. In a different direction, the trace formula has been an indispensable tool in the study of asymptotic behavior of spectral invariants as exemplified by the Weyl law, the limit multiplicity problem, and more generally the Sato–Tate equidistribution for families. This allows another useful perspective on characters of reductive groups over local fields, e.g., by studying quantitative aspects of discrete series and formal degrees.

Tasho Kaletha surveys the new theory of local and global rigid inner forms, which seems indispensable in stating and proving a precise version of the Langlands correspondence and functoriality for reductive groups which are not quasi-split. The data for rigid inner forms are natural in that they determine a canonical normalization of transfer factors as well as the coefficients in the endoscopic character identities. The main advantage of Kaletha's approach over the previous ones is that every inner form over a local or global field admits at least one rigidification as a rigid inner form.

The article by Julee Kim, Sug Woo Shin, and Nicolas Templier studies an asymptotic behavior of supercuspidal characters of p-adic groups. The idea is that one can get a somewhat explicit control of the characters of supercuspidal representations constructed by Yu (which exhaust all supercuspidal representations if p is large by Kim's theorem). The main conjecture and its partial confirmation in the paper are motivated by an asymptotic study of the trace formula and analogy with Harish-Chandra theory of characters for real groups.

The Weyl law, the limit multiplicity problem and Sato-Tate equidistribution are some of the basic questions one can ask about the asymptotic distribution of automorphic forms. Originally, the Weyl law is concerned with the counting of eigenvalues of the Laplace operator on a compact Riemannian manifold. In the context of automorphic forms, it means that for a given reductive group we consider a family of cusp forms with fixed level and count them with respect to the analytic conductor. The goal is the same as above, namely, to establish an asymptotic formula for the number of cusp forms with fixed level and analytic contuctor bounded by a given number. Since in general, the underlying locally symmetric spaces are non-compact, it is much more subtle to establish the Weyl law in this setting. For GL(2) this problem was first approached by Selberg using his trace formula. In the higher-rank case, the Selberg trace formula is replaced by the Arthur trace formula.

The limit multiplicity problem is concerned with the limiting behavior of the discrete spectrum associated to congruence subgroups of a reductive group. For a given congruence subgroup of a reductive group G, one counts automorphic representations in the discrete spectrum whose Archimedean component belongs to a fixed bounded subset of the unitary dual of $G(\mathbb{R})$. The normalized counting function is a measure on the unitary dual, and the problem is to show that it approximates the Plancherel measure if the level of the congruence subgroup converges to infinity. This is known to be true for congruence subgroups of $\mathrm{GL}(n)$.

The first aim of the article by Peter Sarnak, Sug Woo Shin, and Nicolas Templier is to give a working definition for a family of automorphic representations. The definition given includes all known families. It distinguishes between harmonic families which can be approached by the trace formula and geometric families which arise from diophantine equations. One of the main issues is to put forth the basic structural properties of families. The implication is that one can define various invariants, notably the Frobenius–Schur indicator, moments of the Sato–Tate measure, a Sato–Tate group of the family, and the symmetry type. Altogether this refines the Katz–Sarnak heuristics and provides a framework for studying families and their numerous applications to sieving, equidistribution, L-functions, and other problems in number theory.

The article by Steven J. Miller et al. is a survey on results and works in progress on low-lying zeros of families of L-functions attached to geometric families of elliptic curves. The emphasis is on extended supports in the Katz–Sarnak heuristics and on lower-order terms and biases. The article begins with a detailed treatment of Dirichlet characters, which serves as an introduction to the techniques and general issues for a reader wishing to enter the subject.

The article by Werner Müller discusses the Weyl law and recent joint work with Finis and Lapid on limit multiplicities. Currently, both the geometric and the spectral sides can only be dealt with for the groups $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$. Further research about the related problems is in progress. In the final section, the growth of analytic torsion is discussed. Analytic torsion is a sophisticated spectral invariant of an arithmetic group, whose growth with respect to the level aspect is related to the limit multiplicity problem and has consequences for the growth of torsion in the cohomology of arithmetic groups.

In her article on Hecke eigenvalues, Jasmin Matz discusses work concerning the asymptotic distribution of eigenvalues of Hecke operators on cusp forms for $\mathrm{GL}(n)$. Matz–Templier established the Sato–Tate equidistribution of Hecke eigenvalues for families of Hecke–Maass cusp forms on $\mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$. This has consequences for average estimates toward the Ramanujan conjecture and the distribution of low-lying zeros of each of the principal, symmetric square and exterior square L-functions. The Arthur–Selberg trace formula is used in the same way as in the case of the Weyl law.

A particular aspect of the limit multiplicity problem is the study of the growth of Betti numbers of congruence quotients of symmetric spaces if the level of the congruence subgroups tends to infinity. In his article, Simon Marshall establishes asymptotic upper bounds for the L^2-Betti numbers of the locally symmetric spaces associated to a quasi-split unitary group of degree 4, which improve the standard bounds. The main tool is the endoscopic classification of automorphic representations of quasi-split unitary groups by Mok.

Eigenvarieties and p-adic families of automorphic forms arose from the study of mod p and p-adic congruences of modular forms. They are the p-adic analogues of the harmonic families of automorphic forms in the context of the trace formula, but the p-adic version admits rigorous algebraic and geometric definitions and have been more thoroughly studied as such. Many analytic questions about families of automorphic forms can also be asked in the p-adic context. For instance the distribution of Hecke eigenvalues can be studied p-adically, and one could study families of p-adic L-functions instead of the usual L-functions. This could lead to novel and strong methods, especially if combined with the analytic approach.

Hida presents his results on the growth of Hecke fields in Hida families of Hilbert modular forms with motivation from Iwasawa theory. Hida's main theorem is that an irreducible component of the ordinary Hecke algebra is a CM-component, i.e., its associated Galois representation is dihedral, if and only if the Hecke field for that component has bounded degree over the p^∞-power cyclotomic extension over $\mathbb{Q}$ in some precise sense.

Buzzard and Gee introduce conjectures by Gouvêa, Gouvêa–Mazur, and Buzzard on the slopes of modular forms, namely, the p-adic valuations of the U_p-eigenvalues, for varying weights and fixed tame level. Despite computational evidence, the conjectures are largely open to date. The article points out a purely local phenomenon in the reduction of crystalline Galois representations motivated by the conjectures and proposes to make progress toward Buzzard's conjectures via modularity lifting theorems.

Bonn, Germany — Werner Müller
Berkeley, CA, USA — Sug Woo Shin
Ithaca, NY, USA — Nicolas Templier

Contents

Germ Expansions for Real Groups

James Arthur

Abstract We shall introduce an archimedean analogue of the theory of p-adic Shalika germs. These are the objects for p-adic groups that govern the singularities of invariant orbital integrals. More generally, we shall formulate an archimedean theory of germs for weighted orbital integrals. In the process we shall be led to some interesting questions on a general class of asymptotic expansions. Weighted orbital integrals are the parabolic terms on the geometric side of the trace formula. An understanding of their singularities is important for the comparison of trace formulas. It might also play a role in the deeper spectral analysis of a single trace formula.

Mathematics Subject Classification (2010). Primary 22E55, 11F66; Secondary 22E50

1 Introduction

Suppose that G is a connected reductive group over a local field F of characteristic 0. The study of harmonic analysis on $G(F)$ leads directly to interesting functions with complicated singularities. If the field F is p-adic, there is an important qualitative description of the behaviour of these functions near a singular point. It is given by the Shalika germ expansion, and more generally, its noninvariant analogue. The purpose of this paper is to establish similar expansions in the archimedean case $F = \mathbb{R}$.

The functions in question are the invariant orbital integrals and their weighted generalizations. They are defined by integrating test functions $f \in C_c^\infty\big(G(F)\big)$ over strongly regular conjugacy classes in $G(F)$. We recall that $\gamma \in G(F)$ is strongly regular if its centralizer G_γ in G is a torus, and that the set G_{reg} of strongly regular

J. Arthur (✉)
Department of Mathematics, University of Toronto, 40 St. George Street, Toronto, ON, Canada, M5S 2E4
e-mail: arthur@math.utoronto.ca

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_1

elements is open and dense in G. If $\gamma \in G_{\mathrm{reg}}(F)$ approaches a singular point c, the corresponding orbital integrals blow up. It is important to study the resulting behaviour in terms of both γ and f.

The invariant orbital integral

$$f_G(\gamma) = |D(\gamma)|^{1/2} \int_{G_\gamma(F)\backslash G(F)} f(x^{-1}\gamma x)dx, \qquad \gamma \in G_{\mathrm{reg}}(F),$$

is attached to the invariant measure dx on the conjugacy class of γ. Invariant orbital integrals were introduced by Harish-Chandra. They play a critical role in his study of harmonic analysis on $G(F)$. The weighted orbital integral

$$J_M(\gamma,f) = |D(\gamma)|^{1/2} \int_{G_\gamma(F)\backslash G(F)} f(x^{-1}\gamma x)v_M(x)dx, \qquad \gamma \in M(F) \cap G_{\mathrm{reg}}(F),$$

is defined by a noninvariant measure $v_M(x)dx$ on the class of γ. The factor $v_M(x)$ is the volume of a certain convex hull, which depends on both x and a Levi subgroup M of G. Weighted orbital integrals have an indirect bearing on harmonic analysis, but they are most significant in their role as terms in the general trace formula. In the special case that $M = G$, the definitions reduce to $v_G(x) = 1$ and $J_G(\gamma,f) = f_G(\gamma)$. Weighted orbital integrals therefore include invariant orbital integrals.

Suppose that c is an arbitrary semisimple element in $G(F)$. In Sect. 2, we shall introduce a vector space $\mathcal{D}_c(G)$ of distributions on $G(F)$. Let $\mathcal{U}_c(G)$ be the union of the set of conjugacy classes $\Gamma_c(G)$ in $G(F)$ whose semisimple part equals the conjugacy class of c. Then $\mathcal{D}_c(G)$ is defined to be the space of distributions that are invariant under conjugation by $G(F)$ and are supported on $\mathcal{U}_c(G)$. If F is p-adic, $\mathcal{D}_c(G)$ is finite dimensional. It has a basis composed of singular invariant orbital integrals

$$f \longrightarrow f_G(\rho), \qquad \rho \in \Gamma_c(G),$$

taken over the classes in $\Gamma_c(G)$. However if $F = \mathbb{R}$, the space $\mathcal{D}_c(G)$ is infinite dimensional. It contains normal derivatives of orbital integrals, as well as more general distributions associated with harmonic differential operators. In Sect. 2 (which like the rest of the paper pertains to the case $F = \mathbb{R}$), we shall describe a suitable basis $R_c(G)$ of $\mathcal{D}_c(G)$.

For p-adic F, the invariant orbital integral has a decomposition

$$f_G(\gamma) = \sum_{\rho\in\Gamma_c(G)} \rho^\vee(\gamma)f_G(\rho), \qquad f \in C_c^\infty\big(G(F)\big), \qquad (1)_p$$

into a finite linear combination of functions parametrized by conjugacy classes. This is the original expansion of Shalika. It holds for strongly regular points γ that are close to c, in a sense that depends on f. The terms

$$\rho^\vee(\gamma) = g_G^G(\gamma, \rho), \qquad \rho \in \Gamma_c(G),$$

are known as Shalika germs, since they are often treated as germs of functions of γ around c. One can in fact also treat them as functions, since they have a homogeneity property that allows them to be defined on a fixed neighbourhood of c. The role of the Shalika germ expansion is to free the singularities of $f_G(\gamma)$ from their dependence on f.

In Sect. 3, we introduce an analogue of the Shalika germ expansion for the archimedean case $F = \mathbb{R}$. The situation is now slightly more complicated. The sum in $(1)_p$ over the finite set $\Gamma_c(G)$ has instead to be taken over the infinite set $R_c(G)$. Moreover, in place of an actual identity, we obtain only an asymptotic formula

$$f_G(\gamma) \sim \sum_{\rho \in R_c(G)} \rho^\vee(\gamma) f_G(\rho), \qquad f \in C_c^\infty(G). \qquad (1)_\mathbb{R}$$

As in the p-adic case, however the terms

$$\rho^\vee(\gamma) = g_G^G(\gamma, \rho), \qquad \rho \in R_c(G),$$

can be treated as functions of γ, by virtue of a natural homogeneity property. The proof of $(1)_\mathbb{R}$ is not difficult, and is probably implicit in several sources. We shall derive it from standard results of Harish-Chandra and the characterization by Bouaziz [B2] of invariant orbital integrals.

Suppose now that M is a Levi subgroup of G, and that c is an arbitrary semisimple element in $M(F)$. It is important to understand something of the behaviour of the general weighted orbital integral $J_M(\gamma, f)$, for points γ near c. For example, in the comparison of trace formulas, one can sometimes establish identities among terms parametrized by strongly regular points γ. One would like to extend such identities to the more general terms parametrized by singular elements ρ.

In the p-adic case, there is again a finite expansion[1]

$$J_M(\gamma, f) = \sum_{L \in \mathcal{L}(M)} \sum_{\rho \in \Gamma_c(L)} g_M^L(\gamma, \rho) J_L(\rho, f), \quad f \in C_c^\infty\big(G(F)\big). \qquad (2)_p$$

The right-hand side is now a double sum, in which L ranges over the finite set $\mathcal{L}(M)$ of Levi subgroups containing M. The terms

$$g_M^L(\gamma, \rho), \qquad L \in \mathcal{L}(M),\ \rho \in \Gamma_c(L),$$

in the expansion are defined as germs of functions of γ in $M(F) \cap G_{\mathrm{reg}}(F)$ near c. The coefficients

$$J_L(\rho, f), \qquad L \in \mathcal{L}(M),\ \rho \in \Gamma_c(L),$$

[1] I thank Waldspurger for pointing this version of the expansion out to me. My original formulation [A3, Proposition 9.1] was less elegant.

are singular weighted orbital integrals. These objects were defined in [A3, (6.5)], for F real as well as p-adic, by constructing a suitable measure on the $G(F)$-conjugacy class of the singular element ρ. The role of $(2)_p$ is again to isolate the singularities of $J_M(\gamma,f)$ from their dependence on f.

The goal of this paper is to establish an analogue of $(2)_p$ in the archimedean case $F = \mathbb{R}$. We shall state the results in Sect. 6, in the form of two theorems. The main assertion is that there is an infinite asymptotic expansion

$$J_M(\gamma,f) \sim \sum_{L\in\mathcal{L}(M)} \sum_{\rho\in R_c(L)} g_M^L(\gamma,\rho) J_L(\rho,f), \qquad f \in C_c^\infty\big(G(\mathbb{R})\big). \tag*{$(2)_{\mathbb{R}}$}$$

The double sum here is essentially parallel to $(2)_p$, but its summands are considerably more complicated. The terms

$$g_M^L(\gamma,\rho), \qquad\qquad L \in \mathcal{L}(M),\ \rho \in R_c(L), \tag{3}$$

are "formal germs", in that they belong to completions of spaces of germs of functions. They determine asymptotic series, analogous (if more complicated) to the Taylor series of a smooth, nonanalytic function. The coefficients

$$J_L(\rho,f), \qquad\qquad L \in \mathcal{L}(M),\ \rho \in R_c(L), \tag{4}$$

have to be defined for *all* singular invariant distributions ρ, rather than just the singular orbital integrals spanned by $\Gamma_c(L)$. The definitions of [A3] are therefore not good enough. We shall instead construct the distributions $J_L(\rho,f)$ and the formal germs $g_M^L(\gamma,\rho)$ together, in the course of proving the two theorems. We refer the reader to the statement of Theorem 6.1 for a detailed list of properties of these objects.

As preparation for the theorems, we review the properties of general weighted orbital integrals in Sect. 4, with emphasis on the bounds they satisfy as γ approaches a singular point. These bounds provide motivation for the spaces of functions we introduce at the end of the section. Section 5 is one of the more complicated parts of the paper. However, the difficulties are largely formal (with apologies for the pun), for it is here that we introduce the spaces of formal germs that contain the coefficients on the right-hand side of $(2)_{\mathbb{R}}$. These are obtained from the spaces in Sect. 4 by a process of localization (which yields germs), followed by completion (which yields formal power series). The constructions are made more abstract, perhaps, by the need to account for the original singularities of the weighted orbital integrals on the left-hand side of $(2)_{\mathbb{R}}$. In any case, the various topological vector spaces are represented by a commutative diagram later in the section, which might be useful to the reader. At the end of Sect. 5 we introduce some simpler spaces, which act as a bridge between the invariant orbital integrals in $(1)_{\mathbb{R}}$ and the weighted orbital integrals in $(2)_{\mathbb{R}}$. The link is summarized in the sequence of inclusions (33).

The proof of Theorems 6.1 and 6.1* will occupy Sects. 7 through 10. The argument is by induction. We draw some preliminary inferences from our induction hypothesis in Sect. 7. However, our main inspiration is to be taken from the obvious source, the work of Harish-Chandra, specifically his ingenious use of differential equations to estimate invariant orbital integrals. One such technique is the foundation in Sect. 4 of some initial estimates for weighted orbital integrals around c. These estimates in turn serve as motivation for the general spaces of formal germs we introduce in Sect. 5. A second technique of Harish-Chandra will be the basis of our main estimate. We shall apply the technique in Sect. 8 to the differential equations satisfied by the asymptotic series on the right-hand side of $(2)_{\mathbb{R}}$, or rather, the difference between $J_M(\gamma,f)$ and that part of the asymptotic series that can be defined by our induction hypothesis. The resulting estimate will be used in Sect. 9 to establish two propositions. These propositions are really the heart of the matter. They will allow us to construct the remaining part of the asymptotic series in Sect. 10 and to show that it has the required properties.

In Sect. 11, we shall apply our theorems to invariant distributions. We are speaking here of the invariant analogues of weighted orbital integrals, the distributions

$$I_M(\gamma,f), \qquad \gamma \in M(\mathbb{R}) \cap G_{\mathrm{reg}}(\mathbb{R}),$$

that occur in the invariant trace formula. We shall derive an asymptotic expansion

$$I_M(\gamma,f) \sim \sum_{L\in\mathcal{L}(M)} \sum_{\rho\in R_c(L)} g_M^L(\gamma,\rho) I_L(\rho,f), \; f \in C_c^\infty\big(G(\mathbb{R})\big), \tag{5}$$

for these objects that is parallel to $(2)_{\mathbb{R}}$.

We shall conclude the paper in Sect. 12 with some supplementary comments on the new distributions. In particular, we shall show that the invariant distributions

$$I_L(\rho,f), \qquad L \in \mathcal{L}(M), \; \rho \in R_c(L), \tag{6}$$

in (5), as well as their noninvariant counterparts (4), satisfy a natural descent condition.

The distributions (6) are important objects in their own right. As we have noted, they should satisfy local transfer relations of the kind encountered in the theory of endoscopy. However, their definition is quite indirect. It relies on the construction of noninvariant distributions (4), which as we have noted is a consequence of our main theorems. Neither set of distributions is entirely determined by the given conditions. We shall frame this lack of uniqueness in terms of a choice of some element in a finite dimensional affine vector space. One can make the choice in either the noninvariant context (Proposition 9.3), or equivalently, the setting of the invariant distributions (as explained at the end of Sect. 11). When it comes to comparing invariant distributions (6) on different groups, it would of course be important to make the required choices in a compatible way. The question is related to the remarkable interpretation by Hoffmann [Ho] of the underlying differential equations, and the stabilization [A8] of these equations.

This paper is a revision of a preprint that was posted in 2004. My original interest was primarily in the endoscopic comparisons needed for the stabilization of the global trace formula. This problem has now been treated differently by Waldspurger [W], in his work on the stabilization of the more general twisted trace formula. He introduces a subset of the distributions we consider here, which are simpler to construct, but which still include all of the terms needed for the global stabilization. He is then able to establish the required endoscopic relations by a more direct approach. The endoscopic properties of the full set (6) would still be of interest, I think even in their own right. As for other global applications, the stable trace formula and another form of local transfer are at the heart of Beyond Endoscopy, the proposal of Langlands for studying the general principle of functoriality. However, it is too early to speculate whether this has any implications for singular (weighted) orbital integrals.

2 Singular Invariant Distributions

Let G be a connected reductive group over the real field $\mathbb{R}$. If c is a semisimple element in $G(\mathbb{R})$, we write $G_{c,+}$ for the centralizer of c in G, and $G_c = (G_{c,+})^0$ for the connected component of 1 in $G_{c,+}$. Both $G_{c,+}$ and G_c are reductive algebraic groups over $\mathbb{R}$. Recall that c is said to be strongly G-regular if $G_{c,+} = T$ is a maximal torus in G. We shall frequently denote such elements by the symbol γ, reserving c for more general semisimple elements. We write $\Gamma_{ss}(G) = \Gamma_{ss}\big(G(\mathbb{R})\big)$ and $\Gamma_{\mathrm{reg}}(G) = \Gamma_{\mathrm{reg}}\big(G(\mathbb{R})\big)$ for the set of conjugacy classes in $G(\mathbb{R})$ that are, respectively, semisimple and strongly G-regular.

We follow the usual practice of representing the Lie algebra of a group by a corresponding lowercase Gothic letter. For example, if c belongs to $\Gamma_{ss}(G)$,

$$\mathfrak{g}_c = \{X \in \mathfrak{g} : \ \mathrm{Ad}(c)X = X\}$$

denotes the Lie algebra of G_c. (We frequently do not distinguish between a conjugacy class and some fixed representative of the class.) Suppose that $\gamma \in \Gamma_{\mathrm{reg}}(G)$. Then $T = G_\gamma$ is a maximal torus of G over $\mathbb{R}$, with Lie algebra $\mathfrak{t} = \mathfrak{g}_\gamma$, and we write

$$D(\gamma) = D^G(\gamma) = \det\big(1 - \mathrm{Ad}(\gamma)\big)_{\mathfrak{g}/\mathfrak{t}}$$

for the Weyl discriminant of G. If γ is contained in G_c, we can of course also form the Weyl discriminant

$$D_c(\gamma) = D^{G_c}(\gamma) = \det\big(1 - \mathrm{Ad}(\gamma)\big)_{\mathfrak{g}_c/\mathfrak{t}}$$

of G_c. The function D_c will play an important role in formulating the general germ expansions of this paper.

Suppose that f is a function in the Schwartz space $\mathcal{C}(G) = \mathcal{C}\big(G(\mathbb{R})\big)$ on $G(\mathbb{R})$ [H3], and that γ belongs to $\Gamma_{\text{reg}}(G)$. The *invariant orbital integral* of f at γ is defined by the absolutely convergent integral

$$f_G(\gamma) = J_G(\gamma, f) = |D(\gamma)|^{1/2} \int_{G_\gamma(\mathbb{R})\backslash G(\mathbb{R})} f(x^{-1}\gamma x)dx.$$

One can regard $f_G(\gamma)$ as a function of f, in which case it is a tempered distribution. One can also regard $f_G(\gamma)$ is a function of γ, in which case it represents a transform from $\mathcal{C}(G)$ to a space of functions on either $\Gamma_{\text{reg}}(G)$ or

$$T_{\text{reg}}(\mathbb{R}) = G_\gamma(\mathbb{R}) \cap G_{\text{reg}}(\mathbb{R}).$$

(Recall that $G_{\text{reg}}(\mathbb{R})$ denotes the open dense subset of strongly G-regular elements in $G(\mathbb{R})$.) We shall generally take the second point of view. In the next section, we shall establish an asymptotic expansion for $f_G(\gamma)$, as γ approaches a fixed singular point.

Let $c \in \Gamma_{ss}(G)$ be a fixed semisimple conjugacy class. Keeping in mind that c also denotes a fixed element within the given class, we write $\mathcal{U}_c(G)$ for the union of those conjugacy classes in $G(\mathbb{R})$ whose semisimple Jordan component equals c. Then $\mathcal{U}_c(G)$ is a closed subset of $G(\mathbb{R})$ on which $G(\mathbb{R})$ acts by conjugation. We define $\mathcal{D}_c(G)$ to be the vector space of $G(\mathbb{R})$-invariant distributions that are supported on $\mathcal{U}_c(G)$. In this section, we shall introduce a suitable basis of $\mathcal{D}_c(G)$.

Elements in $\mathcal{D}_c(G)$ are easy to construct. Let $\mathcal{T}_c(G)$ be a fixed set of representatives of the $G_{c,+}(\mathbb{R})$-orbits of maximal tori on G_c over $\mathbb{R}$, or equivalently, a fixed set of representatives of the $G(\mathbb{R})$-orbits of maximal tori in G over $\mathbb{R}$ that contain c. We shall write $S_c(G)$ for the set of triplets

$$\sigma = (T, \Omega, X),$$

where T belongs to $\mathcal{T}_c(G)$, Ω belongs to the set $\pi_{0,c}\big(T_{\text{reg}}(\mathbb{R})\big)$ of connected components of $T_{\text{reg}}(\mathbb{R})$ whose closure contains c, and X is an invariant differential operator on $T(\mathbb{R})$. (By an invariant differential operator on $T(\mathbb{R})$, we of course mean a linear differential operator that is invariant under translation by $T(\mathbb{R})$.) Let σ be a triplet in $S_c(G)$. A deep theorem of Harish-Chandra [H2, H3] asserts that the orbital integral

$$f_G(\gamma), \qquad f \in \mathcal{C}(G),\ \gamma \in \Omega,$$

extends to a continuous linear map from $\mathcal{C}(G)$ to the space of smooth functions on the closure of Ω. It follows from this that the limit

$$f_G(\sigma) = \lim_{\gamma \to c} (Xf_G)(\gamma), \qquad \gamma \in \Omega, f \in \mathcal{C}(G),$$

exists, and is continuous in f. If f is compactly supported and vanishes on a neighbourhood of $\mathcal{U}_c(G)$, $f_G(\sigma)$ equals 0. The linear form $f \to f_G(\sigma)$ therefore belongs to $\mathcal{D}_c(G)$.

Bouaziz has shown that, conversely, the distributions $f \to f_G(\sigma)$ span $\mathcal{D}_c(G)$. To describe the result in more detail, we need to attach some familiar data to the tori T in $\mathcal{T}_c(G)$. Given T, we write $W_{\mathbb{R}}(G,T)$ for the subgroup of elements in the Weyl group $W(G,T)$ of (G,T) that are defined over $\mathbb{R}$, and $W\big(G(\mathbb{R}),T(\mathbb{R})\big)$ for the subgroup of elements in $W_{\mathbb{R}}(G,T)$ induced from $G(\mathbb{R})$. We also write $W_{\mathbb{R},c}(G,T)$ and $W_c\big(G(\mathbb{R}),T(\mathbb{R})\big)$ for the subgroups of elements in $W_{\mathbb{R}}(G,T)$ and $W\big(G(\mathbb{R}),T(\mathbb{R})\big)$, respectively, that map the element $c \in T(\mathbb{R})$ to itself. We then form the imaginary root sign character

$$\varepsilon_{c,I}(w) = (-1)^b, \qquad\qquad b = |w(\Sigma^+_{c,I}) \cap \Sigma^+_{c,I}|,\ w \in W_{\mathbb{R},c}(G,T),$$

on $W_{\mathbb{R},c}(G,T)$, where $\Sigma^+_{c,I}$ denotes the set of positive imaginary roots on (G_c,T) relative to any chamber. This allows us to define the subspace

$$S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I} = \big\{u \in S\big(\mathfrak{t}(\mathbb{C})\big) :\ wu = \varepsilon_{c,I}(w)u,\ w \in W_c\big(G(\mathbb{R}),T(\mathbb{R})\big\}$$

of elements in the symmetric algebra on $\mathfrak{t}(\mathbb{C})$ that transform under $W_c\big(G(\mathbb{R}),T(\mathbb{R})\big)$ according to the character $\varepsilon_{c,I}$. There is a canonical isomorphism $u \to \partial(u)$ from $S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}$ onto the space of $\varepsilon_{c,I}$-equivariant differential operators on $T(\mathbb{R})$.

For each $T \in \mathcal{T}_c(G)$, we choose a connected component $\Omega_T \in \pi_{0,c}\big(T_{\mathrm{reg}}(\mathbb{R})\big)$. For any $u \in S\big(T(\mathbb{C})\big)^{c,I}$ and $w \in W_{\mathbb{R},c}(G,T)$, the triplet

$$\sigma_{w,u} = \big(T, w\Omega_T, \partial(u)\big)$$

lies in $S_c(G)$. We obtain a linear transformation

$$\rho : \bigoplus_{T\in\mathcal{T}_c(G)} S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I} \longrightarrow \mathcal{D}_c(G) \tag{7}$$

by mapping u to the distribution

$$\rho_u : f \longrightarrow \sum_{w\in W_{\mathbb{R},c}(G,T)} \varepsilon_{c,I}(w) f_G(\sigma_{w,u}), \qquad\qquad f \in \mathcal{C}(G),$$

in $\mathcal{D}_c(G)$. For each T, we choose a basis $B\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}$ of $S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}$, whose elements we take to be homogeneous. We then form the subset

$$R_c(G) = \big\{\rho_u :\ T \in \mathcal{T}_c(G),\ u \in B\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}\big\}$$

of $\mathcal{D}_c(G)$.

Lemma 2.1. *The map (7) is an isomorphism, and $R_c(G)$ is a basis of $\mathcal{D}_c(G)$. In particular, $\mathcal{D}_c(G)$ consists of tempered distributions.*

Proof. Since $R_c(G)$ is the image under the linear transformation (7) of a basis, it would be enough to establish the assertion that (7) is an isomorphism. We could equally well deal with the mapping

$$\rho' : \bigoplus_{T \in \mathcal{T}_c(G)} S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I} \longrightarrow \mathcal{D}_c(G) \tag{7$'$}$$

that sends an element $u \in S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}$ to the distribution

$$\rho'_u : f \longrightarrow f_G(\sigma_{1,u}) = f_G\big(T, \Omega_T, \partial(u)\big), \qquad f \in \mathcal{C}(G).$$

For it is an easy consequence of Harish-Chandra's jump conditions for orbital integrals that there is an isomorphism of the domain of (7) to itself whose composition with (7) equals (7′). It would be enough to show that (7′) is an isomorphism.

That the mapping (7′) is an isomorphism is implicit in the papers [B1] and [B2] of Bouaziz. In the special case that $c = 1$, the corresponding result for the Lie algebra $\mathfrak{g}(\mathbb{R})$ was proved explicitly [B1, Proposition 6.1.1]. The assertion for $G(\mathbb{R})$, again in the special case that $c = 1$, follows immediately from properties of the exponential map. A standard argument of descent then reduces the general assertion for $G(\mathbb{R})$ to the special case, applied to the group $G_c(\mathbb{R})$. □

If $\rho = \big(T, \Omega_T, \partial(u)\big)$ belongs to $R_c(G)$, we set $\deg(\rho)$ equal to the degree of the homogeneous element $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$. Observe that for any nonnegative integer n, the subset

$$R_{c,n}(G) = \{\rho \in R_c(G) : \ \deg(\rho) \leq n\}$$

of $R_c(G)$ is finite. This set is in turn a disjoint union of subsets

$$R_{c,(k)}(G) = \{\rho \in R_c(G) : \ \deg(\rho) = k\}, \qquad 0 \leq k \leq n.$$

The sets $R_{c,(k)}(G)$ will be used in the next section to construct formal germ expansions of invariant orbital integrals.

Let $\mathcal{Z}(G)$ be the centre of the universal enveloping algebra of $\mathfrak{g}(\mathbb{C})$. For any torus $T \in \mathcal{T}_c(G)$, we write

$$h_T : \ \mathcal{Z}(G) \longrightarrow S\big(\mathfrak{t}(\mathbb{C})\big)^{W(G,T)}$$

for the Harish-Chandra isomorphism from $\mathcal{Z}(G)$ onto the space of $W(G,T)$-invariant elements in $S\big(\mathfrak{t}(\mathbb{C})\big)$. We then define an action $\sigma \to z\sigma$ of $\mathcal{Z}(G)$ on $\mathcal{D}_c(G)$ by setting

$$z\rho = \big(T, \Omega_T, \partial\big(h_T(z)u\big)\big), \qquad z \in \mathcal{Z}(G),$$

for any $\rho = \big(T, \Omega_T, \partial(u)\big)$ in the basis $R_c(G)$. It follows immediately from Harish-Chandra's differential equations

$$(zf)_G(\gamma) = \partial\big(h_T(z)\big)f_G(\gamma), \qquad f \in \mathcal{C}(G),\ \gamma \in T_{\mathrm{reg}}(\mathbb{R}), \tag{8}$$

for invariant orbital integrals that

$$f_G(z\rho) = (zf)_G(\rho). \tag{9}$$

There is no special reason to assume that $R_c(G)$ is stable under the action of $\mathcal{Z}(G)$. However, we do agree to identify any function ϕ on $R_c(G)$ with its linear extension to $\mathcal{D}_c(G)$, in order that the values

$$\phi(z\rho), \qquad\qquad z \in \mathcal{Z}(G),\ \rho \in R_c(G),$$

be defined. Moreover, for any $z \in \mathcal{Z}(G)$, we write $\hat{z}$ for the transpose of the linear operator $\sigma \to z\sigma$ on $\mathcal{D}_c(G)$, relative to the basis $R_c(G)$. In other words,

$$\sum_{\rho \in R_c(G)} \phi(\rho)\psi(\hat{z}\rho) = \sum_{\rho \in R_c(G)} \phi(z\rho)\psi(\rho), \tag{10}$$

for any functions ϕ and ψ of finite support on $R_c(G)$.

We note for future reference that as a $\mathcal{Z}(G)$-module, $\mathcal{D}_c(G)$ is free. To exhibit a free basis, we write $\mathcal{D}_{c,\mathrm{harm}}(G)$ for the finite dimensional subspace of $\mathcal{D}_c(G)$ spanned by triplets $\big(T, \Omega, \partial(u)\big)$ in $S_c(G)$ for which u belongs to the subspace $S_{\mathrm{harm}}\big(\mathfrak{t}(\mathbb{C})\big)$ of harmonic elements in $S\big(\mathfrak{t}(\mathbb{C})\big)$. (Recall that u is *harmonic* if as a polynomial on $\mathfrak{t}(\mathbb{C})^*$, $\partial(u^*)u = 0$ for every element $u^* \in S\big(\mathfrak{t}(\mathbb{C})^*\big)^{W(G,T)}$ with zero constant term.) It can be shown that

$$S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I} = S_{\mathrm{harm}}\big(\mathfrak{t}(\mathbb{C})\big)^{c,I} \otimes S\big(\mathfrak{t}(\mathbb{C})\big)^{W(G,T)},$$

where

$$S_{\mathrm{harm}}\big(\mathfrak{t}(\mathbb{C})\big)^{c,I} = S_{\mathrm{harm}}\big(\mathfrak{t}(\mathbb{C})\big) \cap S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}.$$

Any linear basis of $\mathcal{D}_{c,\mathrm{harm}}(G)$ is therefore a free basis of $\mathcal{D}_c(G)$ as a $\mathcal{Z}(G)$-module.

The remarks above are of course simple consequences of the isomorphism (7). Another implication of (7) is the existence of a canonical grading on the vector space $\mathcal{D}_c(G)$. The grading is compatible with the natural filtration on $\mathcal{D}_c(G)$ that is inherited from the underlying filtration on the space

$$\mathcal{I}(G) = \big\{f_G(\gamma) : f \in \mathcal{C}(G)\big\}.$$

We shall be a bit more precise about this, in order to review how subsets of $R_c(G)$ are related to Levi subgroups.

By a Levi subgroup M of G, we mean an $\mathbb{R}$-rational Levi component of a parabolic subgroup of G over $\mathbb{R}$. For any such M, we write A_M for the $\mathbb{R}$-split component of the centre of M. Then $A_M(\mathbb{R})^0$ is a connected abelian Lie group, whose Lie algebra can be identified with the real vector space

$$\mathfrak{a}_M = \mathrm{Hom}\big(X(M)_{\mathbb{R}}, \mathbb{R}\big).$$

We write

$$W(M) = W^G(M) = \mathrm{Norm}_G(M)/M$$

for the Weyl group of (G, A_M). We shall follow a standard convention of writing $\mathcal{L}(M) = \mathcal{L}^G(M)$ for the finite set of Levi subgroups of G that contain M, and $\mathcal{L}^0(M)$ for the complement of $\{G\}$ in $\mathcal{L}(M)$. Similarly, $\mathcal{F}(M) = \mathcal{F}^G(M)$ stands for the finite set of parabolic subgroups

$$P = M_P N_P, \qquad\qquad M_P \in \mathcal{L}(M),$$

of G over $\mathbb{R}$ that contain M, while

$$\mathcal{P}(M) = \mathcal{P}^G(M) = \big\{P \in \mathcal{F}(M) : \ M_P = M\big\}$$

stands for the subset of parabolic subgroups in $\mathcal{F}(M)$ with Levi component M. Again, $\mathcal{F}^0(M)$ denotes the complement of $\{G\}$ in $\mathcal{F}(M)$.

Suppose that M is a Levi subgroup of G. We write $\Gamma_{G\text{-reg}}(M)$ for the set of classes in $\Gamma_{\mathrm{reg}}(M)$ that are strongly G-regular. There is a canonical map from $\Gamma_{G\text{-reg}}(M)$ to $\Gamma_{\mathrm{reg}}(G)$ on whose fibres the group $W(M)$ acts. The dual restriction map of functions is a linear transformation $\phi_G \to \phi_M$ from $\mathcal{I}(G)$ to $\mathcal{I}(M)$. We define $F^M(\mathcal{I}(G))$ to be the space of functions ϕ_G in $\mathcal{I}(G)$ such that $\phi_L = 0$ for every Levi subgroup L of G that does not contain a conjugate of M. If $M = G$, $F^M(\mathcal{I}(G))$ is the space $\mathcal{I}_{\mathrm{cusp}}(G)$ of cuspidal functions in $\mathcal{I}(G)$. This space is nonzero if and only if G has maximal torus T over $\mathbb{R}$ that is elliptic, in the sense that $T(\mathbb{R})/A_G(\mathbb{R})$ is compact. Letting M vary, we obtain an order reversing filtration on $\mathcal{I}(G)$ over the partially ordered set of G-conjugacy classes of Levi subgroups. The graded vector space attached to the filtration has M-component equal to the quotient

$$G^M\big(\mathcal{I}(G)\big) = F^M\big(\mathcal{I}(G)\big) / \sum_{L \supsetneq M} F^L\big(\mathcal{I}(G)\big).$$

The map $\phi_G \to \phi_M$ is then an isomorphism from $G^M\big(\mathcal{I}(G)\big)$ onto the space $\mathcal{I}_{\mathrm{cusp}}(M)^{W(M)}$ of $W(M)$-invariant cuspidal functions in $\mathcal{I}(M)$. (See [A6]. The definition of $F^M\big(\mathcal{I}(G)\big)$ was unfortunately stated incorrectly on p. 508 of that paper, as was the definition of the corresponding stable space on p. 510.)

Since the distributions in $\mathcal{D}_c(G)$ factor through the projection $f \to f_G$ of $\mathcal{C}(G)$ onto $\mathcal{I}(G)$, they may be identified with linear forms on $\mathcal{I}(G)$. The decreasing filtration on $\mathcal{I}(G)$ therefore provides an increasing filtration on $\mathcal{D}_c(G)$. To be precise, $F^M\big(\mathcal{D}_c(G)\big)$ is defined to be the subspace of distributions in $\mathcal{D}_c(G)$ that annihilate any of the spaces $F^L\big(\mathcal{I}(G)\big)$ with $L \supsetneq M$. The M-component

$$G^M\big(\mathcal{D}_c(G)\big) = F^M\big(\mathcal{D}_c(G)\big) / \sum_{L \subsetneq M} F^L\big(\mathcal{D}_c(G)\big)$$

of the corresponding graded vector space can of course be zero. It is nonzero if and only if $M(\mathbb{R})$ contains some representative of c, and M_c contains a maximal torus T over $\mathbb{R}$ that is elliptic in M. The correspondence $M \to T$ in fact determines a bijection between the set of nonzero graded components of the filtration of $\mathcal{D}_c(G)$ and the set $\mathcal{T}_c(G)$. Moreover, the mapping (7) yields an isomorphism between the associated graded component $S\big(\mathfrak{t}(\mathbb{C})\big)^{c,I}$ and $G^M\big(\mathcal{D}_c(G)\big)$. We therefore obtain an isomorphism

$$\mathcal{D}_c(G) \xrightarrow{\sim} \bigoplus_{\{M\}} G^M\big(\mathcal{D}_c(G)\big), \tag{11}$$

where $\{M\} = \{M\}/G$ ranges over conjugacy classes of Levi subgroups of G. The construction does depend on the choice of chambers Ω_T that went into the original definition (7), but only up to a sign on each summand in (11).

The isomorphism (11) gives the grading of $\mathcal{D}_c(G)$. We should point out that there is also a natural grading on the original space $\mathcal{I}(G)$. For the elements f_G in $\mathcal{I}(G)$ can be regarded as functions on the set $\Pi_{\text{temp}}(G)$ of irreducible tempered representations of $G(\mathbb{R})$, rather than the set $\Gamma_{\text{reg}}(G)$. The space of functions on $\Pi_{\text{temp}}(G)$ so obtained has been characterized [A5], and has a natural grading that is compatible with the filtration above. (See [A6, §4] for the related p-adic case.) However, this grading on $\mathcal{I}(G)$ is not compatible with (11).

We shall say that an element in $\mathcal{D}_c(G)$ is *elliptic* if it corresponds under the isomorphism (11) to an element in the space $G^G\big(\mathcal{D}_c(G)\big)$. We write $\mathcal{D}_{c,\text{ell}}(G)$ for the subspace of elliptic elements in $\mathcal{D}_c(G)$, and we write

$$R_{c,\text{ell}}(G) = R_c(G) \cap \mathcal{D}_{c,\text{ell}}(G)$$

for the associated basis of $\mathcal{D}_{c,\text{ell}}(G)$. For any Levi subgroup M of G, we shall also write $\mathcal{D}_{c,\text{ell}}(M, G)$ for the subspace of distributions in $\mathcal{D}_{c,\text{ell}}(M)$ that are invariant under the action of the finite group $W(M)$. (We can assume that $M(\mathbb{R})$ contains a representative c of the given conjugacy class, since the space is otherwise zero.) The set

$$R_{c,\text{ell}}(M, G) = R_c(G) \cap G^M\big(\mathcal{D}_c(G)\big)$$

can then be identified with a basis of $\mathcal{D}_{c,\mathrm{ell}}(M, G)$. The grading (11) gives a decomposition

$$R_c(G) = \coprod_{\{M\}} R_{c,\mathrm{ell}}(M, G)$$

of the basis of $\mathcal{D}_c(G)$.

Suppose, finally, that θ is an $\mathbb{R}$-isomorphism from G to another reductive group $G_1 = \theta G$ over $\mathbb{R}$. Then $c_1 = \theta c$ is a class in $\Gamma_{ss}(G_1)$. For any $f \in \mathcal{C}(G)$, the function

$$(\theta f)(x_1) = f(\theta^{-1} x_1), \qquad x_1 \in G_1(\mathbb{R}),$$

belongs to $\mathcal{C}(G_1)$. The map that sends any $\rho \in \mathcal{D}_c(G)$ to the distribution $\theta\rho$ defined by

$$(\theta f)_G(\theta\rho) = f_G(\rho) \tag{12}$$

is an isomorphism from $\mathcal{D}_c(G)$ onto $\mathcal{D}_{c_1}(G_1)$. It of course maps the basis $R_c(G)$ of $\mathcal{D}_c(G)$ to the basis $R_{c_1}(G_1) = \theta R_c(G)$ of $\mathcal{D}_{c_1}(G_1)$.

3 Invariant Germ Expansions

Let c be a fixed element in $\Gamma_{ss}(G)$ as in Sect. 2. We are going to introduce an asymptotic approximation of the invariant orbital integral $f_G(\gamma)$, for elements γ near c. This will be a foundation for the more elaborate asymptotic expansions of weighted orbital integrals that are the main goal of the paper.

Suppose that V is an open, $G(\mathbb{R})$-invariant neighbourhood of c in $G(\mathbb{R})$. We write

$$\mathcal{I}(V) = \left\{ f_G : V_{\mathrm{reg}} \longrightarrow \mathbb{C},\ f \in \mathcal{C}(G) \right\}$$

for the space of functions on

$$V_{\mathrm{reg}} = V \cap G_{\mathrm{reg}}(\mathbb{R})$$

that are restrictions of functions in $\mathcal{I}(G)$. If $\sigma = (T, \Omega, X)$ belongs to the set $S_c(G)$ defined in Sect. 2, the intersection

$$V_\Omega = V_{\mathrm{reg}} \cap \Omega$$

is an open neighbourhood of c in the connected component Ω of $T_{\mathrm{reg}}(\mathbb{R})$. The functions ϕ in $\mathcal{I}(V)$ are smooth on V_Ω, and have the property that the seminorms

$$\|\phi\|_\sigma = \sup_{\gamma \in V_\Omega} |(X\phi)(\gamma)| \tag{13}$$

are finite. These seminorms make $\mathcal{I}(V)$ into a topological vector space. To deal with neighbourhoods that vary, it will be convenient to work with the algebraic direct limit

$$\mathcal{I}_c(G) = \varinjlim_{V} \mathcal{I}(V)$$

relative to the restriction maps

$$\mathcal{I}(V_1) \longrightarrow \mathcal{I}(V_2), \qquad V_1 \supset V_2.$$

The elements in $\mathcal{I}_c(G)$ are germs of $G(\mathbb{R})$-invariant, smooth functions on invariant neighbourhoods of c in $G_{\mathrm{reg}}(\mathbb{R})$. (We will ignore the topology on $\mathcal{I}_c(G)$ inherited from the spaces $\mathcal{I}(V)$, since it is not Hausdorff.)

As is customary in working with germs of functions, we shall generally not distinguish in the notation between an element in $\mathcal{I}_c(G)$ and a function in $\mathcal{I}(V)$ that represents it. The open neighbourhood V of c is of course not uniquely determined by the original germ. The convention is useful only in describing phenomena that do not depend on the choice of V. It does make sense, for example, for the linear forms ρ in $\mathcal{D}_c(G)$. By Lemma 2.1, ρ factors through the map $f \to f_G$. It can be evaluated at a function in any of the spaces $\mathcal{I}(V)$, and the value taken depends only on the image of the functions in $\mathcal{I}_c(G)$. In other words, the notation $\phi(\rho)$ is independent of whether we treat ϕ as a germ in $\mathcal{I}_c(G)$ or a function in $\mathcal{I}(V)$.

For a given V, Bouaziz characterizes the image of the space $C_c^\infty(V)$ under the mapping $f \to f_G$. He proves that the image is the space of $G(\mathbb{R})$-invariant, smooth functions on V_{reg} that satisfy the conditions $\mathrm{I}_1(G)$-$\mathrm{I}_4(G)$ on pp. 579–580 of [B2, §3]. Assume that the open invariant neighbourhood V of c is sufficiently small. The conditions can then be formulated in terms of triplets (T, Ω, X) in $S_c(G)$. Condition $\mathrm{I}_1(G)$ is simply the finiteness of the seminorm (13). Condition $\mathrm{I}_2(G)$ asserts that the singularities of ϕ in $T(\mathbb{R}) \cap V$ that do not come from noncompact imaginary roots are removable. Condition $\mathrm{I}_3(G)$ is Harish-Chandra's relation for the jump of $X\phi(\gamma)$ across any wall of V_Ω defined by a noncompact imaginary root. Condition $\mathrm{I}_4(G)$ asserts that the closure in $T(\mathbb{R}) \cap V$ of the support of ϕ is compact. The theorem of Bouaziz leads directly to a characterization of our space $\mathcal{I}_c(G)$.

Lemma 3.1. *$\mathcal{I}_c(G)$ is the space of germs of invariant, smooth functions $\phi \in C^\infty(V_{\mathrm{reg}})$ that for any $(T, \Omega, X) \in S_c(G)$ satisfy the conditions $\mathrm{I}_1(G) - \mathrm{I}_3(G)$ in [B2, §3].*

Proof. Suppose that ϕ belongs to $\mathcal{I}_c(G)$. Then ϕ has a representative in $\mathcal{I}(V)$, for some open invariant neighbourhood V of c. We can therefore identify ϕ with the restriction to V_{reg} of an orbital integral f_G of some function $f \in \mathcal{C}(G)$. It follows from the analytic results of Harish-Chandra that f_G satisfies the three conditions. (See [H3, Lemma 26] and [H4, Theorem 9.1].)

Conversely, suppose that for some small V, ϕ is an invariant function in $C^\infty(V_{\mathrm{reg}})$ that satisfies the three conditions. In order to accommodate the fourth condition, we modify the support of ϕ. Let $\psi_1 \in C^\infty\big(G(\mathbb{R})\big)$ be a smooth, $G(\mathbb{R})$-invariant function whose support is contained in V, and which equals 1 on some open, invariant neighbourhood $V_1 \subset V$ of c. For example, we can choose a positive, homogeneous, $G_{c,+}(\mathbb{R})$-invariant polynomial q_c on $\mathfrak{g}_c(\mathbb{R})$ whose zero set equals $c\,\mathcal{U}_1(G_c)$, as in the construction on p. 166 of [B1], together with a function $\alpha_1 \in C_c^\infty(\mathbb{R})$ that is supported on a small neighbourhood of 0, and equals 1 on an even smaller neighbourhood of 0. The function

$$\psi_1(x) = \alpha_1\big(q_c(\log \gamma)\big),$$

defined for any

$$x = y^{-1}c\gamma y, \qquad\qquad y \in G(\mathbb{R}),\ \gamma \in G_c(\mathbb{R}),$$

has the required property. Given ψ_1, we set

$$\phi_1(x) = \psi_1(x)\phi(x), \qquad\qquad x \in G(\mathbb{R}).$$

The function ϕ_1 then satisfies the support condition $\mathrm{I}_4(G)$ of [B2]. It is not hard to see that ϕ_1 inherits the other three conditions $\mathrm{I}_1(G) - \mathrm{I}_3(G)$ of [B2] from the corresponding conditions on ϕ. It follows from the characterization [B2, Théorème 3.2] that $\phi_1 = f_G$, for some function $f \in C_c^\infty(V)$. Since $C_c^\infty(V)$ is contained in $\mathcal{C}(G)$, and since ϕ takes the same values on $V_{1,\mathrm{reg}}$ as the function $\phi_1 = f_G$, the germ of ϕ coincides with the germ of f_G. In other words, the germ of ϕ lies in the image of $\mathcal{C}(G)$. It therefore belongs to $\mathcal{I}_c(G)$. □

In order to describe the asymptotic series of this paper, it will be convenient to fix a "norm" function that is defined on any small $G(\mathbb{R})$-invariant neighbourhood V of c in $G(\mathbb{R})$. We assume that V is small enough that

(i) any element in V is $G(\mathbb{R})$-conjugate to an element in $T(\mathbb{R})$, for some torus $T \in \mathcal{T}_c(G)$,

(ii) for any $T \in \mathcal{T}_c(G)$ and any w in the complement of $W_c\big(G(\mathbb{R}), T(\mathbb{R})\big)$ in $W\big(G(\mathbb{R}), T(\mathbb{R})\big)$, the intersection

$$w\big(V \cap T(\mathbb{R})\big) \cap \big(V \cap T(\mathbb{R})\big)$$

is empty, and

(iii) for any $T \in \mathcal{T}_c(G)$, the mapping

$$\gamma \longrightarrow \ell_c(\gamma) = \log(\gamma c^{-1}) \tag{14}$$

is a diffeomorphism from $\big(V \cap T(\mathbb{R})\big)$ to an open neighbourhood of zero in $\mathfrak{t}(\mathbb{R})$.

We can of course regard the mapping $\gamma \to \ell_c(\gamma)$ as a coordinate system around the point c in $T(\mathbb{R})$. Let us assume that the Cartan subalgebras $\{\mathfrak{t}(\mathbb{R}) : T \in \mathcal{T}_c(G)\}$ are all stable under a fixed Cartan involution θ_c of $\mathfrak{g}_c(\mathbb{R})$. We choose a $G_{c,+}(\mathbb{R})$-invariant bilinear form B on $\mathfrak{g}_c$ such that the quadratic form

$$\|X\|^2 = -B\big(X, \theta_c(X)\big), \qquad X \in \mathfrak{g}_c(\mathbb{R}),$$

is positive definite on $\mathfrak{g}_c(\mathbb{R})$. The function

$$\gamma \longrightarrow \|\ell_c(\gamma)\|,$$

defined a priori for γ in any of the sets $V \cap T(\mathbb{R})$, $T \in \mathcal{T}_c(G)$, then extends to a $G(\mathbb{R})$-invariant function on V. It will be used to describe the estimates implicit in our asymptotic series.

We have noted that the elements in $\mathcal{D}_c(G)$ can be identified with linear forms on the space $\mathcal{I}_c(G)$. Let us write $\mathcal{I}_{c,n}(G)$ for the annihilator in $\mathcal{I}_c(G)$ of the finite subset $R_{c,n}(G)$ of our basis $R_c(G)$ of $\mathcal{D}_c(G)$. It is obvious that

$$\mathcal{I}_{c,n}(G) = \varinjlim_{V} \mathcal{I}_{c,n}(V),$$

where $\mathcal{I}_{c,n}(V)$ is the subspace of $\mathcal{I}(V)$ annihilated by $R_{c,n}(G)$. We can think of $\mathcal{I}_{c,n}(G)$ as the subspace of functions in $\mathcal{I}_c(G)$ that vanish of order at least $(n+1)$ at c. For later use, we also set $\mathcal{C}_{c,n}(G)$ equal to the subspace of $\mathcal{C}(G)$ annihilated by $R_{c,n}(G)$. It is clear that the map $f \to f_G$ takes $\mathcal{C}_{c,n}(G)$ surjectively to $\mathcal{I}_{c,n}(G)$.

Suppose that ϕ is an element in $\mathcal{I}_c(G)$. We can take the Taylor series around c, relative to the coordinates $\ell_c(\gamma)$, of each of the functions

$$\phi(\gamma), \qquad \gamma 1 V_\Omega,\ T \in \mathcal{T}_c(G),\ \Omega \in \pi_{0,c}\big(T_{\mathrm{reg}}(\mathbb{R})\big),$$

that represent ϕ. For any nonnegative integer k, let $\phi^{(k)}$ be the term in the Taylor series of total degree k. Then $\phi^{(k)}$ can be regarded as an invariant, smooth function in $C^\infty(V_{\mathrm{reg}})$. We claim that it belongs to $\mathcal{I}_c(G)$.

Lemma 3.1 asserts that $\phi^{(k)}$ belongs to $\mathcal{I}_c(G)$ if and only if it satisfies the conditions $\mathrm{I}_1(G) - \mathrm{I}_3(G)$ of [B2, §3]. Condition $\mathrm{I}_1(G)$ is trivial. Conditions $\mathrm{I}_2(G)$ and $\mathrm{I}_3(G)$ are similar, since they both concern the jumps of ϕ about walls in V_Ω, for triplets $(T, \Omega, X) \in S_c(G)$. We shall check only $\mathrm{I}_3(G)$. Suppose that β is a noncompact imaginary root of (G_c, T) that defines a wall of $\Omega = \Omega_+$. Let Ω_- be the complementary component in $T_{\mathrm{reg}}(\mathbb{R})$ that shares this wall. By means of the Cayley transform associated with β, one obtains a second triplet $(T_\beta, \Omega_\beta, X_\beta) \in S_c(G)$ for which Ω_β also shares the given wall of Ω. Condition $I_3(G)$ for ϕ asserts that

$$(X\phi_{\Omega_+})(\gamma) - (X\phi_{\Omega_-})(\gamma) = d(\beta)(X_\beta \phi_{\Omega_\beta})(\gamma), \tag{15}$$

for γ on the given wall of Ω. Here, ϕ_{Ω_*} represents the restriction of ϕ to V_{Ω_*}, a smooth function that extends to the closure of V_{Ω_*}, while $d(\beta)$ is independent of ϕ. If X is a homogeneous invariant differential operator on $T_*(\mathbb{R})$ of degree d, and ϕ is homogeneous of degree k (in the coordinates $\ell_c(\gamma)$), then $(X\phi_{\Omega_*})(\gamma)$ is homogeneous of degree $k-d$ if $k \geq d$, and vanishes if $k < d$. The relation (15) for ϕ then implies the corresponding relation

$$(X\phi^{(k)}_{\Omega_+})(\gamma) - (X\phi^{(k)}_{\Omega_-})(\gamma) = d(\beta)(X_\beta \phi^{(k)}_{\Omega_\beta})(\gamma)$$

for the homogeneous components $\phi^{(k)}$ of ϕ. This is the condition $\mathrm{I}_3(G)$ for $\phi^{(k)}$. The claim follows.

We set

$$\mathcal{I}^{(k)}_c(G) = \{\phi \in \mathcal{I}_c(G) :\ \phi^{(k)} = \phi\},$$

for any nonnegative integer k. Suppose that n is another nonnegative integer. Then $\mathcal{I}^{(k)}_c(G)$ is contained in $\mathcal{I}_{c,n}(G)$ if $k > n$, and intersects $\mathcal{I}_{c,n}(G)$ only at 0 if $k \leq n$. It follows from what we have just proved that the quotient

$$\mathcal{I}^n_c(G) = \mathcal{I}_c(G)/\mathcal{I}_{c,n}(G)$$

has a natural grading

$$\mathcal{I}^n_c(G) \cong \bigoplus_{0 \leq k \leq n} \mathcal{I}^{(k)}_c(G).$$

But $\mathcal{I}_{c,n}(G)$ is the subspace of $\mathcal{I}_c(G)$ annihilated by the finite subset

$$R_{c,n}(G) = \coprod_{0 \leq k \leq n} R_{c,(k)}(G)$$

of $R_c(G)$. It follows that $R_{c,n}(G)$ is a basis of the dual space of $\mathcal{I}^n_c(G)$, and that $R_{c,(k)}(G)$ is a basis of the dual space of $\mathcal{I}^{(k)}_c(G)$.

Let

$$\{\rho^\vee :\ \rho \in R_{c,(k)}(G)\}$$

be the basis of $\mathcal{I}^{(k)}_c(G)$ that is dual to $R_{c,(k)}(G)$. If $T \in \mathcal{T}_c(G)$ and $\Omega \in \pi_{0,c}\big(T_{\mathrm{reg}}(\mathbb{R})\big)$, the restriction to V_Ω of any function $\rho^\vee$ in this set is a homogeneous polynomial

$$\gamma \longrightarrow \rho^\vee(\gamma), \qquad \gamma \in V_\Omega,$$

of degree k (in the coordinates $\ell_c(\gamma)$). In particular, $\rho^\vee$ has a canonical extension to the set of regular points in any invariant neighbourhood in V of c on which the coordinate functions (14) are defined. Thus, unlike a general element in $\mathcal{I}_c(G)$, $\rho^\vee$ really can be treated as a function, as well as a germ of functions.

The union over k of our bases of $\mathcal{I}_c^{(k)}(G)$ is a family of functions

$$\rho^\vee(\gamma), \qquad\qquad \gamma \in V_{\mathrm{reg}},\ \rho \in R_c(G),$$

with properties that are dual to those of $R_c(G)$. For example, the dual of the action (9) of $\mathcal{Z}(G)$ on $\mathcal{D}_c(G)$ is a differential equation

$$(\hat{z}\rho)^\vee = h(z)\rho^\vee, \tag{16}$$

for any $z \in \mathcal{Z}(G)$ and $\rho \in R_c(G)$. Here $\hat{z}$ represents the transpose action (10) of $\mathcal{Z}(G)$, and $h(z)$ is the $G(\mathbb{R})$-invariant differential operator on V_{reg} obtained from the various Harish-Chandra maps $z \to h_T(z)$. The dual of (12) is the symmetry condition

$$\theta\rho^\vee = (\theta\rho)^\vee, \tag{17}$$

for any isomorphism θ: $G \to \theta G$ over $\mathbb{R}$, and any $\rho \in R_c(G)$.

The main reason for defining the functions $\{\rho^\vee\}$ is that they represent germs of invariant orbital integrals. It is clear that

$$\phi^{(k)}(\gamma) = \sum_{\rho \in R_{c,(k)}(G)} \rho^\vee(\gamma)\phi(\rho), \qquad\qquad k \geq 0,$$

for any function $\phi \in \mathcal{I}_c(G)$. Suppose that f belongs to $\mathcal{C}(G)$. The Taylor polynomial of degree n attached to the function $f_G(\gamma)$ on V_{reg} (taken relative to the coordinates $\ell_c(\gamma)$) is then equal to the function

$$f_G^n(\gamma) = \sum_{0\leq k\leq n} f_G^{(k)}(\gamma) = \sum_{\rho\in R_{c,n}(G)} \rho^\vee(\gamma)f_G(\rho). \tag{18}$$

It follows from Taylor's theorem that there is a constant C_n for each n such that

$$|f_G(\gamma) - f_G^n(\gamma)| \leq C_n\|\ell_c(\gamma)\|^{n+1},$$

for any $\gamma \in V_{\mathrm{reg}}$. Otherwise said, $f_G(\gamma)$ has an asymptotic expansion

$$\sum_{\rho\in R_c(G)} \rho^\vee(\gamma)f_G(\rho),$$

in the sense that $f_G(\gamma)$ differs from the partial sum $f_G^n(\gamma)$ by a function in the class $O(\|\ell_c(\gamma)\|^{n+1})$.

The main points of Sects. 2 and 3 may be summarized as follows. There are invariant distributions

$$f \longrightarrow f_G(\rho), \qquad\qquad \rho \in R_c(G),$$

supported on $\mathcal{U}_c(G)$, and homogeneous germs

$$\gamma \longrightarrow \rho^\vee(\gamma), \qquad\qquad \rho \in R_c(G),$$

in $\mathcal{I}_c(G)$, which transform according to (9) and (16) under the action of $\mathcal{Z}(G)$, satisfy the symmetry conditions (12) and (17), and provide an asymptotic expansion

$$f_G(\gamma) \sim \sum_{\rho \in R_c(G)} \rho^\vee(\gamma) f_G(\rho), \qquad\qquad \gamma \in V_{\mathrm{reg}}, \tag{19}$$

around c for the invariant orbital integral $f_G(\gamma)$.

It is useful to have a formulation of (19) that is uniform in f.

Proposition 3.2. *For any $n \geq -1$, the mapping*

$$f \longrightarrow f_G(\gamma) - f_G^n(\gamma), \qquad\qquad f \in \mathcal{C}(G),$$

is a continuous linear transformation from $\mathcal{C}(G)$ to the space $\mathcal{I}_{c,n}(V)$.

Proof. We have interpreted $f_G^n(\gamma)$ as the Taylor polynomial of degree n for the function $f_G(\gamma)$. Since $\mathcal{I}_{c,n}(V)$ can be regarded as a closed subspace of functions in $\mathcal{I}(V)$ that vanish of order at least $(n+1)$ at c, the difference $f_G(\gamma) - f_G^n(\gamma)$ belongs to $\mathcal{I}_{c,n}(V)$. The continuity assertion of the lemma follows from the integral formula for the remainder in Taylor's theorem [D, (8.14.3)], and the continuity of the mapping $f \to f_G$. □

Remarks. 1. Proposition 3.2 could of course be formulated as a concrete estimate. Given $n \geq -1$, we simplify the notation by writing

$$(n, X) = \big(n + 1 - \deg(X)\big)_+ = \max\{(n + 1 - \deg X), 0\}, \tag{20}$$

for any differential operator X. The proposition asserts that for any $\sigma = (T, \Omega, X)$ in $S_c(G)$, there is a continuous seminorm μ_σ^n on $\mathcal{C}(G)$ such that

$$\big|X\big(f_G(\gamma) - f_G^n(\gamma)\big)\big| \leq \mu_\sigma^n(f)\|\ell_c(\gamma)\|^{(n,X)},$$

for any $\gamma \in V_\Omega$ and $f \in \mathcal{C}(G)$.

2. Invariant orbital integrals can be regarded as distributions that are dual to irreducible characters. In this sense, the asymptotic expansion (19) is dual to the character expansions introduced by Barbasch and Vogan near the beginning of [BV].

Our goal is to extend these results for invariant orbital integrals to weighted orbital integrals. As background for this, we observe that much of the discussion of Sects. 2 and 3 for G applies to the relative setting of a pair (M, G), for a fixed Levi subgroup M of G. In this context, we take c to be a fixed class in $\Gamma_{ss}(M)$. Then c represents a $W(M)$-orbit in $\Gamma_{ss}(M)$ (or equivalently, the intersection of M with a class in $\Gamma_{ss}(G)$), which we also denote by c. With this understanding, we take V to be a small open neighbourhood of c in $M(\mathbb{R})$ that is invariant under the normalizer

$$W(M)M(\mathbb{R}) = \mathrm{Norm}_{G(\mathbb{R})}\big(M(\mathbb{R})\big)$$

of $M(\mathbb{R})$ in $G(\mathbb{R})$.

Given V, we can of course form the invariant Schwartz space $\mathcal{I}(V)$ for M. If f belongs to $\mathcal{C}(G)$, the relative (invariant) orbital integral f_M around c is the restriction of f_G to the subset

$$V_{G\text{-reg}} = V \cap G_{\mathrm{reg}}(\mathbb{R})$$

of $G_{\mathrm{reg}}(\mathbb{R})$. It is easy to see that $f \to f_M$ is a continuous linear mapping from $\mathcal{C}(G)$ into the closed subspace

$$\mathcal{I}(V, G) = \mathcal{I}(V)^{W(M)}$$

of $W(M)$-invariant functions in $\mathcal{I}(V)$. (We identify functions in $\mathcal{I}(V, G)$ with their restrictions to $V_{G\text{-reg}}$.) Other objects defined earlier have obvious relative analogues. For example, $S_c(M, G)$ denotes the set of triplets (T, Ω, X), where T belongs to the set $\mathcal{T}_c(M)$ (defined for M as in Sect. 2), Ω is a connected component in $T_{G\text{-reg}}(\mathbb{R})$ (rather than $T_{M\text{-reg}}(\mathbb{R})$) whose closure contains c, and X is an invariant differential operator on $T(\mathbb{R})$ (as before). The elements in $S_c(M, G)$ yield continuous seminorms (13) that determine the topology on $\mathcal{I}(V, G)$. We can also define the direct limits

$$\mathcal{I}_c(M, G) = \varinjlim_{V} \mathcal{I}(V, G)$$

and

$$\mathcal{I}_{c,n}(V, G) = \varinjlim_{V} \mathcal{I}_{c,n}(V, G),$$

where $\mathcal{I}_{c,n}(V, G)$ denotes the subspace of $\mathcal{I}(V, G)$ annihilated by the finite subset $R_{c,n}(M)$ of the basis $R_c(M)$. We shall use these relative objects in Sect. 5, when we introduce spaces that are relevant to weighted orbital integrals.

We note that there is also a relative analogue of the space of harmonic distributions introduced in Sect. 2. We define the subspace $\mathcal{D}_{c,G\text{-harm}}(M)$ of *G-harmonic* distributions in $\mathcal{D}_c(M)$ be the space spanned by those triplets

$(T, \Omega, \partial(u))$ in $S_c(M, G)$ such that the element $u \in S(\mathfrak{t}(\mathbb{C}))$ is harmonic relative to G. Any linear basis of $\mathcal{D}_{c,G\text{-harm}}(M)$ is a free basis of $\mathcal{D}_c(M)$, relative to the natural $\mathcal{Z}(G)$-module structure on $\mathcal{D}_c(M)$. In our construction of certain distributions later in the paper, the elements in $\mathcal{D}_{c,G\text{-harm}}(M)$ will be the primitive objects to deal with.

4 Weighted Orbital Integrals

We now fix a maximal compact subgroup K of $G(\mathbb{R})$. We also fix a Levi subgroup M of G such that $\mathfrak{a}_M$ is orthogonal to the Lie algebra of K (with respect to the Killing form on $\mathfrak{g}(\mathbb{R})$). There is then a natural smooth function

$$v_M(x) = \lim_{\lambda \to 0} \left(\sum_{P \in \mathcal{P}(M)} \mathrm{e}^{-\lambda(H_P(x))} \theta_P(\lambda)^{-1} \right)$$

on $M(\mathbb{R})\backslash G(\mathbb{R})/K$, defined as the volume of a certain convex hull. This function provides a noninvariant measure on the $G(\mathbb{R})$-conjugacy class of any strongly G-regular point in $M(\mathbb{R})$, relative to which any Schwartz function $f \in \mathcal{C}(G)$ is integrable. The resulting integral

$$J_M(\gamma, f) = J_M^G(\gamma, f) = |D(\gamma)|^{1/2} \int_{G_\gamma(\mathbb{R})\backslash G(\mathbb{R})} f(x^{-1}\gamma x) v_M(x) dx$$

is a smooth, $M(\mathbb{R})$-invariant function of γ in the set

$$M_{G\text{-reg}}(\mathbb{R}) = M(\mathbb{R}) \cap G_{\text{reg}}(\mathbb{R}).$$

(See [A1, Lemma 8.1] and [A2, §6–7].) We recall a few of its basic properties.

For any γ, the linear form $f \to J_M(\gamma, f)$ is a tempered distribution. In contrast to the earlier special case

$$J_G(\gamma, f) = f_G(\gamma)$$

of $M = G$, however, it is not invariant. Let

$$f^y : x \longrightarrow f(yxy^{-1}), \qquad x \in G(\mathbb{R}),$$

be the conjugate of f by a fixed element $y \in G(\mathbb{R})$. The weighted orbital integral of f^y can then be expanded as

$$J_M(\gamma, f^y) = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\gamma, f_{Q,y}), \tag{21}$$

in the notation of [A2, Lemma 8.2]. The summand with $Q = G$ is equal to $J_M(\gamma, f)$. The expansion can therefore be written as an identity

$$J_M(\gamma, f^y - f) = \sum_{Q \in \mathcal{F}^0(M)} J_M^{M_Q}(\gamma, f_{Q,y})$$

that represents the obstruction to the distribution being invariant.

Weighted orbital integrals satisfy a generalization of the differential equations (8). If z belongs to $\mathcal{Z}(G)$, the weighted orbital integral of zf has an expansion

$$J_M(\gamma, zf) = \sum_{L \in \mathcal{L}(M)} \partial_M^L(\gamma, z_L) J_L(\gamma, f). \tag{22}$$

Here $z \to z_L$ denotes the canonical injective homomorphism from $\mathcal{Z}(G)$ to $\mathcal{Z}(L)$, while $\partial_M^L(\gamma, z_L)$ is an $M(\mathbb{R})$-invariant differential operator on $M(\mathbb{R}) \cap L_{\mathrm{reg}}(\mathbb{R})$ that depends only on L. If T is a maximal torus in $\mathcal{T}_c(M)$, $\partial_M^L(\gamma, z_L)$ restricts to an algebraic differential operator on the algebraic variety $T_{L\text{-reg}}$. Moreover, $\partial_M^L(\gamma, z_L)$ is invariant under the finite group $W^L(M)$ of outer automorphisms of M. We can therefore regard $\partial_M^L(\gamma, z_L)$ as a $W^L(M)M(\mathbb{R})$-invariant, algebraic differential operator on the algebraic variety $M_{G\text{-reg}}$. In the case that $L = M$, $\partial_M^M(\gamma, z_M)$ reduces to the invariant differential operator $\partial\big(h(z)\big)$ on $M(\mathbb{R})$ obtained from the Harish-Chandra isomorphism. The differential equation (22) can therefore be written as an identity

$$J_M(\gamma, zf) - \partial\big(h(z)\big) J_M(\gamma, f) = \sum_{L \neq M} \partial_M^L(\gamma, z_L) J_L(\gamma, f)$$

that is easier to compare with the simpler equations (8). (See [A1, Lemma 8.5] and [A3, §11–12].)

Suppose that $\theta \colon G \to \theta G$ is an isomorphism over $\mathbb{R}$, as in Sect. 2. We can then take weighted orbital integrals on $(\theta G)(\mathbb{R})$ with respect to θK and θM. They satisfy the relation

$$J_{\theta M}(\theta\gamma, \theta f) = J_M(\gamma, f) \tag{23}$$

[A7, Lemma 3.3]. In particular, suppose that $\theta = \mathrm{Int}(w)$, for a representative $w \in K$ of some element in the Weyl group $W(M)$. In this case, $J_M(\gamma, \theta f)$ equals $J_M(\gamma, f)$, and $\theta M = M$, from which it follows that

$$J_M(w\gamma w^{-1}, f) = J_M(\gamma, f).$$

Therefore $J_M(\gamma, f)$ is actually a $W(M)M(\mathbb{R})$-invariant function of γ.

At this point, we fix a class $c \in \Gamma_{ss}(M)$ and an open $W(M)M(\mathbb{R})$-invariant neighbourhood V of c in $M(\mathbb{R})$, as at the end of Sect. 3. We can assume that V is small. In particular, we assume that the intersection of V with any maximal torus in $M(\mathbb{R})$ is relatively compact.

We propose to study $J_M(\gamma, f)$ as a function of γ in $V_{G\text{-reg}}$. The behaviour of this function near the boundary is more complicated in general than it is in the invariant case $M = G$. In particular, if (T, Ω, X) lies in the set $S_c(M, G)$ introduced at the end of Sect. 3, the restriction of $J_M(\gamma, f)$ to the region

$$V_\Omega = V \cap \Omega$$

does not extend smoothly to the boundary of V_Ω. The function satisfies only the weaker estimate of the following lemma.

Lemma 4.1. *For every triplet $\sigma = (T, \Omega, X)$ in $S_c(M, G)$, there is a positive real number a such that the supremum*

$$\mu_\sigma(f) = \sup_{\gamma \in V_\Omega} \big(|X J_M(\gamma, f)| |D_c(\gamma)|^a \big), \qquad f \in \mathcal{C}(G),$$

is a continuous seminorm on $\mathcal{C}(G)$. In the case that $X = 1$, we can take a to be any positive number.

Proof. This lemma is essentially the same as Lemma 13.2 of [A3]. The proof is based on an important technique of Harish-Chandra for estimating invariant orbital integrals [H1, Lemma 48]. We shall recall a part of the argument, in order to persuade ourselves that it remains valid under the minor changes here (where, for example, $\mathcal{C}(G)$ replaces $C_c^\infty\big(G(\mathbb{R})\big)$, and $D_c(\gamma)$ takes the place of $D(\gamma)$), referring the reader to [A3] and [H1] for the remaining part.

We fix the first two components $T \in \mathcal{T}_c(M)$ and $\Omega \in \pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big)$ of a triplet σ. We require an estimate for every invariant differential operator X that can form a third component of σ. As in Harish-Chandra's treatment of invariant orbital integrals, one studies the general problem in three steps.

The first step is to deal with the identity operator $X = 1$. In this case, the required estimate is a consequence of Lemma 7.2 of [A1]. The lemma cited leads to a bound

$$|J_M(\gamma, f)| \leq \mu(f)\big(1 + L(\gamma)\big)^p, \qquad \gamma \in V_\Omega,$$

in which μ is a continuous seminorm on $\mathcal{C}(G)$. The function $L(\gamma)$ is defined at the bottom of p. 245 of [A1] as a supremum of functions

$$\Big| \log \big(|1 - \alpha(\gamma)|\big) \Big|, \qquad \gamma \in V_\Omega,$$

attached to roots α of (G, T). Since V is assumed to be small, the function attached to α is bounded on V_Ω unless α is a root of (G_c, T). It follows that for any $a_1 > 0$, we can choose a constant C_1 such that

$$\big(1 + L(\gamma)\big)^p \leq C_1 |D_c(\gamma)|^{-a_1}, \qquad \gamma \in V_\Omega.$$

Lemma 7.2 of [A1] therefore implies that

$$f \longrightarrow \sup_{\gamma \in V_\Omega} \big(|J_M(\gamma, f)||D_c(\gamma)|^{a_1} \big), \qquad f \in \mathcal{C}(G), \tag{24}$$

is a continuous seminorm on $\mathcal{C}(G)$. The required estimate is thus valid in the case $X = 1$, for any positive exponent $a = a_1$.

The next step concerns the case that X is the image under the Harish-Chandra map of a biinvariant differential operator. That is,

$$X = \partial\big(h_T(z)\big), \qquad z \in \mathcal{Z}(G).$$

In this case, the differential equation (22) yields an identity

$$\begin{aligned} XJ_M(\gamma, f) &= \partial\big(h_T(z)\big) J_M(\gamma, f) \\ &= J_M(\gamma, zf) - \sum_{L \supsetneq M} \partial_M^L(\gamma, z_L) J_L(\gamma, f) \end{aligned} \tag{25}$$

for the function we are trying to estimate. We have noted that for each L, $\partial_M^L(\gamma, z_L)$ is an algebraic differential operator on $T_{G\text{-reg}}$. In other words, the coefficients of $\partial_M^L(\gamma, z_L)$ are rational functions on T whose poles lie along singular hypersurfaces of T. Since V is small, any singular hypersurface of T that meets the closure of V_Ω is defined by a root of (G_c, T). It follows that for each L, there is a positive integer k_L such that the differential operator

$$D_c(\gamma)^{k_L} \partial_M^L(\gamma, z_L)$$

has coefficients that are bounded on V_Ω. We can assume inductively that Lemma 4.1 is valid if M is replaced by any $L \supsetneq M$. The estimate of the lemma clearly extends to differential operators with bounded coefficients. We can therefore choose $a_L > 0$ for each such L so that

$$f \longrightarrow \sup_{\gamma \in V_\Omega} \big(|D_c(\gamma)^{k_L} \partial_M^L(\gamma, z_L) J_L(\gamma, f)||D_c(\gamma)|^{a_L} \big)$$

is a continuous seminorm on $\mathcal{C}(G)$. We set a equal to the largest of the numbers $k_L + a_L$. The functional

$$f \longrightarrow \sum_{L \supsetneq M} \sup_{\gamma \in V_\Omega} \big(|\partial_M^L(\gamma, z_L) J_L(\gamma, f)||D_c(\gamma)|^{a} \big)$$

is then a continuous seminorm on $\mathcal{C}(G)$. According to the case (24) we have already established,

$$f \longrightarrow \sup_{\gamma \in V_\Omega} \big(|J_M(\gamma, zf)||D_c(\gamma)|^{a} \big)$$

is also a continuous seminorm on $\mathcal{C}(G)$. Applying these estimates to the differential equation for $XJ_M(\gamma,f)$ above, we conclude that

$$f \longrightarrow \sup_{\gamma\in V_\Omega}\big(|XJ_M(\gamma,f)||D_c(\gamma)|^a\big), \qquad f\in\mathcal{C}(G), \tag{26}$$

is a continuous seminorm on $\mathcal{C}(G)$. We have established the lemma for X of the form $\partial\big(h_T(z)\big)$.

The last step is to treat a general invariant differential operator X on $T(\mathbb{R})$. This is the main step, and the part of the argument that is based on [H1, Lemma 48]. In the proof of [A3, Lemma 13.2], we explained how to apply Harish-Chandra's technique to the weighted orbitals we are dealing with here. Used in this way, the technique reduces the required estimate for X to the case (26) obtained above. It thus establishes the assertion of the lemma for any X, and hence for any triplet σ in $S_c(M,G)$. We refer the reader to [A3] and [H1] for the detailed discussion of this step. □

With Lemma 4.1 as motivation, we now introduce some new spaces of functions. We first attach some spaces to any maximal torus T in M over $\mathbb{R}$ that contains c. Given T, let $\Omega\in\pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big)$ be a connected component whose closure contains c. Then $V_\Omega = V\cap\Omega$ is an open neighbourhood of c in Ω. If a is a nonnegative real number, we write $F_c^a(V_\Omega,G)$ for the Banach space of continuous functions ϕ_Ω on V_Ω such that the norm

$$\|\phi_\Omega\| = \sup_{\gamma\in V_\Omega}\big(|\phi_\Omega(\gamma)||D_c(\gamma)|^a\big)$$

is finite. More generally, if n is an integer with $n\geq -1$, and $\ell_c(\gamma)$ is the weight function (14), we define $F_{c,n}^a(V_\Omega,G)$ to be the Banach space of continuous functions ϕ_Ω on V_Ω such that the norm

$$\|\phi_\Omega\|_n = \sup_{\gamma\in V_\Omega}\big(|\phi_\Omega(\gamma)||D_c(\gamma)|^a\|\ell_c(\gamma)\|^{-(n+1)}\big)$$

is finite. The first space $F_c^a(V_\Omega,G)$ is of course the special case that $n=-1$. It consists of functions with specified growth near the boundary. In the second space $F_{c,n}^a(V_\Omega,G)$, n will vary, and will ultimately index terms in our asymptotic expansions around c.

Lemma 4.1 suggests that we introduce a space of smooth functions on the $W(M)M(\mathbb{R})$-invariant set $V_{G\text{-reg}}$ whose derivatives also have specified growth near the boundary. This entails choosing a function to measure the growth. By a *weight function*, we shall mean an assignment

$$\alpha:\ X\longrightarrow\alpha(X)$$

of a nonnegative real number $\alpha(X)$ to each invariant differential operator X on a maximal torus T of M. We assume that

$$\alpha(X) = \overline{\alpha}(\deg X),$$

for an increasing function $\overline{\alpha}$ on the set of nonnegative integers. The weight function is then defined independently of T.

Suppose that α is a weight function, and that V is as above, an open $W(M)M(\mathbb{R})$-invariant neighbourhood of c in $M(\mathbb{R})$. If ϕ is a function on $V_{G\text{-reg}}$, and $\sigma = (T, \Omega, X)$ is a triplet in the set $S_c(M, G)$ introduced in Sect. 3, we shall write ϕ_Ω for the restriction of ϕ to V_Ω. We define $\mathcal{F}_c^\alpha(V, G)$ to be the space of smooth, $W(M)M(\mathbb{R})$-invariant functions ϕ on $V_{G\text{-reg}}$ such that for every $\sigma = (T, \Omega, X)$ in $S_c(M, G)$, and every $\varepsilon > 0$, the derivative $X\phi_\Omega$ belongs to the space $F_c^{\alpha(X)+\varepsilon}(V_\Omega, G)$. More generally, suppose $n \geq -1$ is a given integer. We define $\mathcal{F}_{c,n}^\alpha(V, G)$ to be the subspace of functions ϕ in $\mathcal{F}_c^\alpha(V, G)$ such that for any $\sigma = (T, \Omega, X)$ and ε, $X\phi_\Omega$ belongs to the space

$$F_{c,n,X}^{\alpha(X)+\varepsilon}(V_\Omega, G) = F_{c,(n,X)}^{\alpha(X)+\varepsilon}(V_\Omega, G).$$

We recall here that

$$(n, X) = \max\left\{\big(n + 1 - \deg(X)\big), 0\right\}.$$

The seminorms

$$\|\phi\|_{\sigma,\varepsilon,n} = \|X\phi_\Omega\|_n$$

make $\mathcal{F}_{c,n}^\alpha(V, G)$ into a Fréchet space. The original space $\mathcal{F}_c^\alpha(V, G)$ is again the special case that $n = -1$. It is the Fréchet space of smooth, $W(M)M(\mathbb{R})$-invariant functions ϕ on $V_{G\text{-reg}}$ such that for every $\sigma = (T, \Omega, X)$ and ε, the seminorm

$$\|\phi\|_{\sigma,\varepsilon} = \sup_{x \in V_\Omega} \left(|(X\phi)(\gamma)||D_c(\gamma)|^{\alpha(X)+\varepsilon}\right)$$

is finite.

Lemma 4.1 is an assertion about the mapping that sends $f \in \mathcal{C}(G)$ to the function $J_M(\gamma, f)$ of $\gamma \in V_{G\text{-reg}}$. It can be reformulated as follows.

Corollary 4.2. *There is a weight function* α, *with* $\alpha(1) = 0$, *such that the mapping*

$$f \longrightarrow J_M(\gamma, f), \qquad f \in \mathcal{C}(G),$$

is a continuous linear transformation from $\mathcal{C}(G)$ *to* $\mathcal{F}_c^\alpha(V, G)$.

There are some obvious operations that can be performed on the spaces $\mathcal{F}_{c,n}^\alpha(V, G)$. Suppose that α_1 is second weight function, and that $n_1 \geq -1$ is a second integer. The multiplication of functions then provides a continuous bilinear map

$$\mathcal{F}^{\alpha}_{c,n}(V,G) \times \mathcal{F}^{\alpha_1}_{c,n_1}(V,G) \longrightarrow \mathcal{F}^{\alpha+\alpha_1}_{c,n+n_1+1}(V,G),$$

where $\alpha + \alpha_1$ is the weight function defined by

$$\overline{(\alpha+\alpha_1)}(d_+) = \max_{d+d_1=d_+} \big(\overline{\alpha}(d) + \overline{\alpha}_1(d_1)\big), \qquad d_+ \geq 0.$$

In particular, suppose that q is a $W(G,T)$-invariant rational function on a maximal torus T in M that is regular on $T_{G\text{-reg}}$. Then q extends to a $W(M)M(\mathbb{R})$-invariant function on $V_{G\text{-reg}}$ that lies in $\mathcal{F}^{\alpha_q}_{c}(V,G)$, for some weight function α_q. The multiplication map $\phi \to q\phi$ therefore sends $\mathcal{F}^{\alpha}_{c,n}(V,G)$ continuously to $\mathcal{F}^{\alpha+\alpha_q}_{c,n}(V,G)$. A similar observation applies to any (translation) invariant differential operator X on T that is also invariant under the action of $W(G,T)$. For X extends to a $W(M)M(\mathbb{R})$-invariant differential operator on $V_{G\text{-reg}}$, and if $X\alpha$ is the weight function $X' \to \alpha(XX')$, the map $\phi \to X\phi$ sends $\mathcal{F}^{\alpha}_{c,n}(V,G)$ continuously to $\mathcal{F}^{X\alpha}_{c,n,X}(V,G)$. More generally, suppose that $\partial(\gamma)$ is an algebraic differential operator on $T_{G\text{-reg}}$ that is invariant under $W(G,T)$. Then $\partial(\gamma)$ extends to a $W(M)M(\mathbb{R})$-invariant differential operator on $M_{G\text{-reg}}$. One sees easily that there is a weight function $\partial\alpha$ such that $\phi \to \partial\phi$ is a continuous mapping from $\mathcal{F}^{\alpha}_{c,n}(V,G)$ to $\mathcal{F}^{\partial\alpha}_{c,n,\partial}(V,G)$.

5 Spaces of Formal Germs

We fix a Levi subgroup M of G, and a class $c \in \Gamma_{ss}(M)$, as before. We again take V to be a small, open, $W(M)M(\mathbb{R})$-invariant neighbourhood of c in $M(\mathbb{R})$. In the last section, we introduced some spaces

$$\mathcal{F}^{\alpha}_{c,n}(V,G), \qquad n \geq 1,$$

of functions on $V_{G\text{-reg}}$. In this section, we shall examine the behaviour of these spaces under operations of localization and completion.

The most basic of these spaces $\mathcal{F}^{\alpha}_{c}(V,G) = \mathcal{F}^{\alpha}_{c,-1}(V,G)$ is a generalization of the relative invariant Schwartz space

$$\mathcal{I}(V,G) = \mathcal{I}(V)^{W(M)}$$

defined near the end of Sect. 3. It is an easy consequence of Lemma 3.1 that for each α, there is a continuous injection

$$\mathcal{I}(V,G) \hookrightarrow \mathcal{F}^{\alpha}_{c}(V,G)$$

defined also near the end of Sect. 3. As in the special case of $\mathcal{I}(V, G)$ from Sect. 3, we can localize the spaces $\mathcal{F}_c^\alpha(V, G)$ at c. We form the algebraic direct limit

$$\mathcal{G}_c^\alpha(M, G) = \varinjlim_V \mathcal{F}_c^\alpha(V, G), \tag{27}$$

relative to the restriction maps

$$\mathcal{F}_c^\alpha(V_1, G) \longrightarrow \mathcal{F}_c^\alpha(V_2, G), \qquad V_1 \supset V_2.$$

We shall call $\mathcal{G}_c^\alpha(M, G)$ the space of *α-germs* for (M, G) at c. The elements of this space are germs of smooth, $W(M)M(\mathbb{R})$-invariant functions on invariant neighbourhoods of c in $M_{G\text{-reg}}(\mathbb{R})$, with α-bounded growth near the boundary. The space has a decreasing filtration by the subspaces

$$\mathcal{G}_{c,n}^\alpha(M, G) = \varinjlim_V \mathcal{F}_{c,n}^\alpha(V, G), \qquad n \geq -1.$$

Asymptotic series are best formulated in terms of the completion of $\mathcal{G}_c^\alpha(M, G)$. For any α, and any $n \geq 0$, the quotient

$$\mathcal{G}_c^{\alpha,n}(M, G) = \mathcal{G}_c^\alpha(M, G)/\mathcal{G}_{c,n}^\alpha(M, G)$$

is a vector space that is generally infinite dimensional. We call it the space of (α, n)-*jets* for (M, G) at c. The completion of $\mathcal{G}_c^\alpha(M, G)$ is then defined as the projective limit

$$\widehat{\mathcal{G}}_c^\alpha(M, G) = \varprojlim_n \mathcal{G}_c^{\alpha,n}(M, G). \tag{28}$$

This space is obviously also isomorphic to a projective limit of quotients

$$\widehat{\mathcal{G}}_c^{\alpha,n}(M, G) = \widehat{\mathcal{G}}_c^\alpha(M, G)/\widehat{\mathcal{G}}_{c,n}^\alpha(M, G),$$

where $\widehat{\mathcal{G}}_{c,n}^\alpha(M, G)$ is the kernel of the projection of $\widehat{\mathcal{G}}_c^\alpha(M, G)$ onto $\mathcal{G}_c^{\alpha,n}(M, G)$. We call $\widehat{\mathcal{G}}_c^\alpha(M, G)$ the space of *formal α-germs* for (M, G) at c. The final step is to remove the dependence on α. We do so by forming the direct limit

$$\widehat{\mathcal{G}}_c(M, G) = \varinjlim_\alpha \widehat{\mathcal{G}}_c^\alpha(M, G), \tag{29}$$

relative to the natural partial order on the set of weight functions. The operations of multiplication and differentiation from the end of the last section clearly extend to this *universal space* of formal germs. In particular, any $W(M)M(\mathbb{R})$-invariant, algebraic differential operator $\partial(\gamma)$ on $M_{G\text{-reg}}$ has a linear action $g \to \partial g$ on $\widehat{\mathcal{G}}_c(M, G)$.

As an example, consider the case that $G = M = T$ is a torus. The function $D_c(\gamma)$ is then equal to 1, and the various spaces are independent of α. For each α, $\mathcal{G}_c(T) = \mathcal{G}_c^\alpha(M, G)$ is the space of germs of smooth functions on $T(\mathbb{R})$ at c, while $\mathcal{G}_{c,n}(T) = \mathcal{G}_{c,n}^\alpha(M, G)$ is the subspace of germs of functions that vanish at c of order at least $(n+1)$. The quotient $\mathcal{G}_c^n(T) = \mathcal{G}_c^{\alpha,n}(M, G)$ is the usual space of n-jets on $T(\mathbb{R})$ at c, while $\widehat{\mathcal{G}}_c(T) = \widehat{\mathcal{G}}_c^\alpha(M, G)$ is the space of formal Taylor series (in the coordinates $\ell_c(\gamma)$) at c.

If G is arbitrary, but c is G-regular, the group $T = G_c$ is a torus. In this case, the function $D_c(\gamma)$ is again trivial. The various spaces reduce to the ones above for T, or rather, the subspaces of the ones above consisting of elements invariant under the finite group $M_{c,+}(\mathbb{R})/M_c(\mathbb{R})$. We are of course mainly interested in the case that c is not G-regular. Then $D_c(\gamma)$ has zeros, and the spaces are more complicated. On the other hand, we can make use of the function D_c in this case to simplify the notation slightly. For example, given α and $\sigma = (T, \Omega, X)$, we can choose a positive number a such that for any $n \geq 0$, $\phi \to X\phi_\Omega$ is a continuous linear map from $\mathcal{F}_{c,n}^\alpha(V, G)$ to $F_{c,n}^a(V_\Omega, G)$ (rather than $F_{c,n,X}^a(V_\Omega, G)$). A similar result applies if X is replaced by an algebraic differential operator on $T_{G\text{-reg}}(\mathbb{R})$.

Lemma 5.1. *For any V, α and n, the map*

$$\mathcal{F}_c^\alpha(V, G) \longrightarrow \mathcal{G}_c^{\alpha,n}(M, G)$$

is surjective. In other words, any element g^n in $\mathcal{G}_c^{\alpha,n}(M, G)$ has a representative $g^n(\gamma)$ in $\mathcal{F}_c^\alpha(V, G)$.

Proof. Suppose that g^n belongs to $\mathcal{G}_c^{\alpha,n}(M, G)$. By definition, g^n has a representative $g_0^n(\gamma)$ in $\mathcal{F}_c^\alpha(V_0, G)$, for some $W(M)M(\mathbb{R})$-invariant neighbourhood V_0 of c in $M(\mathbb{R})$ with $V_0 \subset V$. Let ψ_0 be a smooth, compactly supported, $W(M)M(\mathbb{R})$-invariant function on V_0 that equals 1 on some neighbourhood of c. The product

$$g^n(\gamma) = \psi_0(\gamma) g_0^n(\gamma)$$

then extends by 0 to a function on V that lies in $\mathcal{F}_c^\alpha(V, G)$. On the other hand, both $g^n(\gamma)$ and $g_0^n(\gamma)$ represent the same germ in $\mathcal{G}_c(M, G)$. They both therefore have the same image g^n in $\mathcal{G}_c^{\alpha,n}(M, G)$. The function $g^n(\gamma)$ is the required representative. □

Lemma 5.2. *For any weight function α, the canonical map from $\widehat{\mathcal{G}}_c^\alpha(M, G)$ to $\widehat{\mathcal{G}}_c(M, G)$ is injective.*

Proof. It is enough to show that if α' is a weight function with $\alpha' \geq \alpha$, the map from $\widehat{\mathcal{G}}_c^\alpha(M, G)$ to $\widehat{\mathcal{G}}_c^{\alpha'}(M, G)$ is injective. Suppose that g is an element in $\widehat{\mathcal{G}}_c^\alpha(M, G)$ that maps to 0 in $\widehat{\mathcal{G}}_c^{\alpha'}(M, G)$. To show that $g = 0$, it would be enough to establish that for any $n \geq 0$, the image g^n of g in $\mathcal{G}_c^{\alpha,n}(M, G)$ equals 0.

Fix n, and let $g^n(\gamma)$ be a representative of g^n in the space of functions $\mathcal{F}_c^\alpha(V, G)$ attached to some V. We have to show that $g^n(\gamma)$ lies in $\mathcal{F}_{c,n}^\alpha(V, G)$. In other words, we must show that for any $\sigma = (T, \Omega, X)$ in $S_c(M, G)$, and any $\varepsilon > 0$, the derivative $(Xg^n_\Omega)(\gamma)$ lies in $F^{\alpha(X)+\varepsilon}_{c,n,X}(V_\Omega, G)$. This condition is of course independent of the choice of representative g^n. Given σ, we are free to assume that $g^n(\gamma)$ represents the image g^m of g in $\mathcal{G}^{\alpha,m}_c(V, G)$, for some large integer $m > n + \deg X$. Since g^m maps to zero in $\mathcal{G}^{\alpha',m}_c(M, G)$, $g^n(\gamma)$ lies in $\mathcal{F}^{\alpha'}_{c,m}(V, G)$. In other words, $(Xg^n_\Omega)(\gamma)$ lies in $F^{\alpha'(X)+\varepsilon'}_{c,n'}(V_\Omega, G)$, for the large integer $n' = m - \deg(X)$ and for any $\varepsilon' > 0$. But $(Xg^n_\Omega)(\gamma)$ also lies in $F^{\alpha(X)+\varepsilon'}_c(V_\Omega, G)$. We shall apply these two conditions successively to two subsets of V_Ω.

Given $\varepsilon > 0$, we choose $\varepsilon' > 0$ with $\varepsilon' < \varepsilon$. We then write $\delta = \varepsilon - \varepsilon'$, $a = \alpha(X) + \varepsilon$, and $a' = \alpha'(X) + \varepsilon'$. The two conditions amount to two inequalities

$$|(Xg^n_\Omega)(\gamma)| \leq C'|D_c(\gamma)|^{-a'}\|\ell_c(\gamma)\|^{n'}, \qquad \gamma \in V_\Omega,$$

and

$$|(Xg^n_\Omega)(\gamma)| \leq C_\delta|D_c(\gamma)|^{-(\alpha(X)+\varepsilon')} = C_\delta|D_c(\gamma)|^{-a}|D_c(\gamma)|^{\delta}, \qquad \gamma \in V_\Omega,$$

for fixed constants C' and C_δ. We can assume that $a' > a$, since there would otherwise be nothing to prove. (The functions $|D_c(\gamma)|$ and $\|\ell_c(\gamma)\|$ are of course bounded on V_Ω.) We apply the first inequality to the points γ in the subset

$$V\big(\delta, (n, X)\big) = \big\{\gamma \in V_\Omega : \ \|\ell_c(\gamma)\|^{(n,X)} \leq |D_c(\gamma)|^{\delta}\big\}$$

of V_Ω, and the second inequality to each γ in the complementary subset. We thereby deduce that if n' is sufficiently large, there is a constant C such that

$$|(Xg_\Omega)(\gamma)| \leq C|D_c(\gamma)|^{-a}\|\ell_c(\gamma)\|^{(n,X)},$$

for any point γ in V_Ω. In other words, $|(Xg_\Omega)(\gamma)|$ belongs to the space $F^{\alpha(X)+\varepsilon}_{c,n,X}(V_\Omega, G)$. It follows that the vector g^n in $\mathcal{G}^{\alpha,n}_c(M, G)$ vanishes. Since n was arbitrary, the original element g in $\widehat{\mathcal{G}}^{\alpha}_c(M, G)$ vanishes. The map from $\widehat{\mathcal{G}}^{\alpha}_c(M, G)$ to $\widehat{\mathcal{G}}^{\alpha'}_c(M, G)$ is therefore injective. □

The lemma asserts that $\widehat{\mathcal{G}}_c(M, G)$ is the union over all weight functions α of the spaces $\widehat{\mathcal{G}}^{\alpha}_c(M, G)$. Suppose that we are given a formal germ $g \in \mathcal{G}_c(M, G)$, and a positive integer n. We shall write $g^n = g^{\alpha,n}$ for the image of g in the quotient $\mathcal{G}^{\alpha,n}_c(M, G)$, for some fixed α such that $\mathcal{G}^{\alpha}_c(M, G)$ contains g. The choice of α will generally be immaterial to the operations we perform on g^n, so its omission from the notation is quite harmless. If $\phi(\gamma)$ is a function in one of the spaces $\mathcal{F}^{\alpha}_c(V, G)$, we shall sometimes denote the image of $\phi(\gamma)$ in $\widehat{\mathcal{G}}_c(M, G)$ simply by ϕ. This being the case, ϕ^n then stands for an element in $\mathcal{G}^{\alpha,n}_c(M, G)$. This element is of course equal to the projection of the original function $\phi(\gamma)$ onto $\mathcal{G}^{\alpha,n}_c(M, G)$.

We shall need to refer to two different topologies on $\widehat{\mathcal{G}}_c(M,G)$. The first comes from the discrete topology on each of the quotients $\mathcal{G}_c^{\alpha,n}(M,G)$. The corresponding projective limit topology over n, followed by the direct limit topology over α, yields what we call the *adic* topology on $\widehat{\mathcal{G}}_c(M,G)$. This is the usual topology assigned to a completion. A sequence (g_k) converges in the adic topology if there is an α such that each g_k is contained in $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$, and if for any n, the image g_k^n of g_k in $\mathcal{G}_c^{\alpha,n}(M,G)$ is independent of k, for all k sufficiently large.

To describe the second topology, we recall that the quotient spaces $\mathcal{G}_c^{\alpha,n}(M,G)$ are generally infinite dimensional. As an abstract vector space over $\mathbb{C}$, however, each $\mathcal{G}_c^{\alpha,n}(M,G)$ can be regarded as a direct limit of finite dimensional spaces. The standard topologies on these finite dimensional spaces therefore induce a direct limit topology on each $\mathcal{G}_c^{\alpha,n}(M,G)$. The corresponding projective limit topology over n, followed by the direct limit topology over α, yields what we call the *complex* topology on $\widehat{\mathcal{G}}_c(M,G)$. This is the appropriate topology for describing the continuity properties of maps from some space into $\widehat{\mathcal{G}}_c(M,G)$. A sequence (g_k) converges in the complex topology of $\widehat{\mathcal{G}}_c(M,G)$ if there is an α such that each g_k is contained in $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$, and if for each n, the sequence g_k^n is contained in a finite dimensional subspace of $\mathcal{G}_c^{\alpha,n}(M,G)$, and converges in the standard topology of that space. Unless otherwise stated, any limit in $\widehat{\mathcal{G}}_c(M,G)$ will be understood to be in the adic topology, while any assertion of continuity for a $\widehat{\mathcal{G}}_c(M,G)$-valued function will refer to the complex topology.

Suppose that g lies in $\widehat{\mathcal{G}}_c(M,G)$. We have agreed to write g^n for the image of g in the quotient $\mathcal{G}_c^{\alpha,n}(M,G)$ of $\mathcal{G}_c^{\alpha}(M,G)$. Here n is any nonnegative integer, and α is a fixed weight function such that g lies in $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$. We shall also write $g^n(\gamma)$ for an α-germ of functions in $\mathcal{G}_c^{\alpha}(M,G)$ that represents g^n, or as in Lemma 5.1, a function in $\mathcal{F}_c^{\alpha}(V,G)$ that represents the α-germ. The function $g^n(\gamma)$ is of course not uniquely determined by g. To see that this does not really matter, we recall that under the previous convention, g^n also denotes the image of $g^n(\gamma)$ in $\widehat{\mathcal{G}}_c(M,G)$. We are therefore allowing g^n to stand for two objects: an element in $\mathcal{G}_c^{\alpha,n}(M,G)$ that is uniquely determined by g (once α is chosen), and some representative of this element in $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$ that is not uniquely determined, and that in particular, need not map to g. With this second interpretation, however, the elements g^n can be chosen so that

$$g = \lim_{n\to\infty} (g^n),$$

in the adic topology.

These remarks can be phrased in terms of asymptotic series. Suppose that $g_k(\gamma)$ is a sequence of functions in $\mathcal{F}_c^{\alpha}(V,G)$ such that the corresponding elements $g_k \in \widehat{\mathcal{G}}_c(M,G)$ converge to zero (in the adic topology). In other words, for any n all but finitely many of the functions $g_k(\gamma)$ lie in the space $\mathcal{F}_{c,n}^{\alpha}(V,G)$. We shall denote the associated asymptotic series by

$$g(\gamma) \sim \sum_k g_k(\gamma), \tag{30}$$

where g is the element in $\widehat{\mathcal{G}}_c(M, G)$ such that

$$g = \sum_k g_k \tag{31}$$

(in the adic topology). Conversely, any formal germ can be represented in this way. For if g belongs to $\widehat{\mathcal{G}}_c(M, G)$, the difference

$$g^{(n)}(\gamma) = g^n(\gamma) - g^{n-1}(\gamma), \qquad\qquad n \geq 0,\ g^{-1}(\gamma) = 0,$$

stands for a function in a space $\mathcal{F}^{\alpha}_{c,n-1}(V, G)$. Therefore

$$g = \sum_{n=0}^{\infty} g^{(n)},$$

so we can represent g by the asymptotic series

$$g(x) \sim \sum_{n=0}^{\infty} g^{(n)}(x).$$

We can use the notation (30) also to denote a convergent sum of asymptotic series. In this more general usage, the terms in (30) stand for asymptotic series $g_k(\gamma)$ and $g(\gamma)$, which in turn represent elements g_k and g in $\widehat{\mathcal{G}}_c(M, G)$ that satisfy (31).

The objects we have introduced might be easier to keep track of if we view them within the following commutative diagram of topological vector spaces:

$$\begin{tikzcd}
0 \arrow[r] & F^{a}_{c,n}(V_\Omega, G) \arrow[r] & F^{a}_{c}(V_\Omega, G) & & \\
0 \arrow[r] & \mathcal{F}^{\alpha}_{c,n}(V, G) \arrow[r] \arrow[u] \arrow[d] & \mathcal{F}^{\alpha}_{c}(V, G) \arrow[u] \arrow[d] & & \\
0 \arrow[r] & \mathcal{G}^{\alpha}_{c,n}(M, G) \arrow[r] \arrow[d] & \mathcal{G}^{\alpha}_{c}(M, G) \arrow[r] \arrow[d] & \mathcal{G}^{\alpha,n}_{c}(M, G) \arrow[r] \arrow[d, "\wr"] & 0 \\
0 \arrow[r] & \widehat{\mathcal{G}}^{\alpha}_{c,n}(M, G) \arrow[r] & \widehat{\mathcal{G}}^{\alpha}_{c}(M, G) \arrow[r] \arrow[d] & \widehat{\mathcal{G}}^{\alpha,n}_{c}(M, G) \arrow[r] & 0 \\
 & & \widehat{\mathcal{G}}_{c}(M, G) & &
\end{tikzcd}$$

As always, M is a Levi subgroup of G and $c \in \Gamma_{ss}(M)$ is a semisimple conjugacy class in $M(\mathbb{R})$, while $V_\Omega = V \cap \Omega$ is as in Sect. 4, α is weight function, $n \geq -1$, and in the top row, a is a real number with $a > \alpha(n)$ (where $n = -1$ in the space on the right). The rows consist of exact sequences, and their constituents become more complex as we go down the columns. More precisely, the top row consists of Banach spaces, the second row consists of Fréchet spaces, the third row consists of LF-spaces (direct limits of Fréchet spaces), while the fourth row consists of ILF-spaces (inverse limits of LF-spaces). The final space $\widehat{\mathcal{G}}_c(M, G)$ is a supplementary direct limit.

The diagram may also give us a better sense of the notational conventions above. Once again, g is a formal germ in the space $\widehat{\mathcal{G}}_c(M, G)$ at the bottom, which we then identify with an element in (the injective image of) the space $\widehat{\mathcal{G}}_c^{\alpha}(M, G)$ immediately above it, for some chosen weight α. Its image in the quotient

$$\widehat{\mathcal{G}}_c^{\alpha,n}(M, G) \cong \mathcal{G}_c^{\alpha,n}(M, G)$$

immediately to the right is what we are denoting by g^n. Finally, $g^n(\gamma)$ stands either for a germ in $\mathcal{G}_c^{\alpha}(M, G)$ or a function in $\mathcal{F}_c^{\alpha}(V, G)$, taken from the two spaces in the middle column, that maps to g^n. This notation, as well as the spaces in the diagram themselves, might seem a bit overblown. However, we shall see that it provides an elegant way to formulate the central formula of this paper, the asymptotic expansion (43) of Theorem 6.1.

As a link between the (relative) invariant Schwartz space and the general spaces above, we consider a space $\mathcal{F}_c^{bd}(V, G)$ of bounded functions. For any integer $n \geq -1$, let $\mathcal{F}_{c,n}^{bd}(V, G)$ be the space of smooth, $W(M)M(\mathbb{R})$-invariant functions ϕ on $V_{G\text{-reg}}$ such that for each $\sigma = (T, \Omega, X)$ in $S_c(M, G)$, the derivative $X\phi_\Omega$ belongs to the space $F_{c,n}^0(V, G)$. If α is any weight function, $\mathcal{F}_{c,n}^{bd}(V, G)$ is contained in $\mathcal{F}_{c,n}^{\alpha}(V, G)$. In fact in the basic case of $n = -1$, the space

$$\mathcal{F}_c^{bd}(V, G) = \mathcal{F}_{c,-1}^{bd}(V, G)$$

is just the subspace of functions in $\mathcal{F}_c^{\alpha}(V, G)$ whose derivatives are all bounded. As above, we form the localizations

$$\mathcal{G}_{c,n}^{bd}(M, G) = \varinjlim_{V} \mathcal{F}_{c,n}^{bd}(V, G),$$

the quotients

$$\mathcal{G}_c^{bd,n}(M, G) = \mathcal{G}_c^{bd}(M, G)/\mathcal{G}_{c,n}^{bd}(M, G) = \mathcal{G}_{c,-1}^{bd}(M, G)/\mathcal{G}_{c,n}^{bd}(M, G)$$

and the completion

$$\widehat{\mathcal{G}}_c^{bd}(M, G) = \varprojlim_{n} \mathcal{G}_c^{bd,n}(M, G).$$

Lemma 5.3. *Suppose that α is a weight function with $\alpha(1) = 0$. Then for any nonnegative integer n, the canonical mapping*

$$\mathcal{G}_c^{bd,n}(M,G) \longrightarrow \mathcal{G}_c^{\alpha,n}(M,G)$$

is injective.

Proof. By Lemma 5.1, there is a canonical isomorphism

$$\mathcal{G}_c^{\alpha,n}(M,G) \cong \mathcal{F}_c^{\alpha}(V,G)/\mathcal{F}_{c,n}^{\alpha}(V,G).$$

On the other hand, any element in $\mathcal{G}_c^{bd,n}(M,G)$ can be identified with a family

$$\{\phi_\Omega^n : T \in \mathcal{T}_c(M),\ \Omega \in \pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big)\}$$

of Taylor polynomials of degree n (in the coordinates $\ell_c(\gamma)$). This is because the Ω-component of any function in $\mathcal{F}_c^{bd}(V,G)$ extends to a smooth function on the closure of V_Ω. In particular, each element in $\mathcal{G}_c^{bd,n}(M,G)$ has a canonical representative in $\mathcal{F}_c^{bd}(V,G)$, which of course also lies in $\mathcal{F}_c^{\alpha}(V,G)$. With this interpretation, we consider a function ϕ in the intersection

$$\mathcal{G}_c^{bd,n}(M,G) \cap \mathcal{F}_{c,n}^{\alpha}(V,G).$$

We have only to show that ϕ vanishes.

Suppose that $T \in \mathcal{T}_c(M)$ and $\Omega \in \pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big)$. As an element in $\mathcal{F}_{c,n}^{\alpha}(V,G)$, ϕ_Ω satisfies a bound

$$\sup_{\gamma \in V_\Omega} \big(|\phi_\Omega(\gamma)||D_c(\gamma)|^{\varepsilon}\|\ell_c(\gamma)\|^{-(n+1)}\big) < \infty,$$

for any $\varepsilon > 0$. As an element in $\mathcal{G}_c^{bd,n}(M,G)$, $\phi_\Omega = \phi_\Omega^n$ is a polynomial (in the coordinates $\ell_c(\gamma)$) of degree less than $(n+1)$. Taking ε to be close to zero, we see that no such polynomial can satisfy the bound unless it vanishes. It follows that $\phi_\Omega = 0$. We conclude that the function ϕ vanishes, and hence, that the original map is injective. □

Corollary 5.4. *For any weight function α, the canonical mapping*

$$\widehat{\mathcal{G}}_c^{bd}(M,G) \longrightarrow \widehat{\mathcal{G}}_c^{\alpha}(M,G)$$

is injective.

Proof. Given α, we choose a weight function $\alpha_0 \leq \alpha$ with $\alpha_0(1) = 0$. The lemma implies that $\widehat{\mathcal{G}}_c^{bd}(M, G)$ maps injectively into $\widehat{\mathcal{G}}_c^{\alpha_0}(M, G)$, while Lemma 5.2 tells us that $\widehat{\mathcal{G}}_c^{\alpha_0}(M, G)$ maps injectively into $\widehat{\mathcal{G}}_c^{\alpha}(M, G)$. The corollary follows. □

Remark. It is not hard to show that if the weight function α is bounded, the injection of Corollary 5.4 is actually an isomorphism. The completion $\widehat{\mathcal{G}}_c^{bd}(M, G)$ is therefore included among the general spaces defined earlier.

The (relative) invariant Schwartz space $\mathcal{I}(V, G)$ is the closed subspace of functions in $\mathcal{F}_c^{bd}(V, G)$ that satisfy the Harish-Chandra jump conditions. Its localization $\mathcal{I}_c(M, G)$ is therefore a subspace of $\mathcal{G}_c^{bd}(M, G)$. Recall that for any n, $\mathcal{I}_{c,n}(M, G)$ is the subspace of $\mathcal{I}_c(M, G)$ annihilated by the finite set of distributions $R_{c,n}(M)$. It follows easily from the discussion of Sect. 3 that

$$\mathcal{I}_{c,n}(M, G) = \mathcal{I}_c(M, G) \cap \mathcal{G}_{c,n}^{bd}(M, G).$$

The quotient

$$\mathcal{I}_c^n(M, G) = \mathcal{I}_c(M, G)/\mathcal{I}_{c,n}(M, G)$$

of $\mathcal{I}_c(M, G)$ therefore injects into the quotient $\mathcal{G}_c^{bd,n}(M, G)$ of $\mathcal{G}_c^{bd}(M, G)$. This in turn implies that the completion

$$\widehat{\mathcal{I}}_c(M, G) = \varprojlim_n \mathcal{I}_c^n(M, G) \tag{32}$$

injects into $\widehat{\mathcal{G}}_c^{bd}(M, G)$. We thus have embeddings

$$\widehat{\mathcal{I}}_c(M, G) \subset \widehat{\mathcal{G}}_c^{bd}(M, G) \subset \widehat{\mathcal{G}}_c^{\alpha}(M, G) \subset \widehat{\mathcal{G}}_c(M, G) \tag{33}$$

for any weight function α

As a subspace of $\widehat{\mathcal{G}}_c(M, G)$, the completion $\widehat{\mathcal{I}}_c(M, G)$ is particularly suited to the conventions above. If g belongs to $\widehat{\mathcal{I}}_c(M, G)$ and $n \geq 0$, we take $g^n(\gamma)$ to be the *canonical* representative of g^n in $\mathcal{I}_c(M, G)$ that is spanned by the finite set $\{\rho^\vee(\gamma) : \rho \in R_{c,n}(M)\}$. This means that $g^{(n)}(\gamma)$ is the canonical element in $\mathcal{I}_{c,n-1}(M, G)$ that is spanned by the set $\{\rho^\vee(\gamma) : \ \rho \in R_{c,(n)}(M)\}$. The formal germ g can therefore be represented by a canonical, adically convergent series

$$g = \sum_{\rho \in R_c(M)} g(\rho)\rho^\vee,$$

or if one prefers, a canonical asymptotic expansion

$$g(\gamma) \sim \sum_{\rho \in R_c(M)} g(\rho)\rho^\vee(\gamma),$$

for uniquely determined coefficients $g(\rho)$ in $\mathbb{C}$. In particular, suppose that g equals f_M, for a Schwartz function $f \in \mathcal{C}(G)$. The relative invariant orbital integral $f_M(\gamma)$ then has an asymptotic expansion

$$f_M(\gamma) \sim \sum_{\rho \in R_c(M)} f_M(\rho)\rho^\vee(\gamma).$$

We end this section by remarking that the $W(M)$-invariance we have built into the definitions is not essential. Its purpose is only to reflect the corresponding property for weighted orbital integrals. We shall sometimes encounter formal germs for which the property is absent (notably as individual terms in a finite sum that is $W(M)$-invariant). There is no general need for extra notation. However, one case of special interest arises when M_1 is a Levi subgroup of M, and c is the image of a class c_1 in $\Gamma_{ss}(M_1)$. Under these conditions, we let $\widehat{\mathcal{G}}_{c_1}(M_1 \mid M, G)$ denote the space of formal germs for (M_1, G) at c_1, defined as above, but with $W(M_1)$ replaced by the stabilizer $W(M_1 \mid M)$ of M in $W(M_1)$. There is then a canonical restriction mapping

$$g \longrightarrow g_{M_1}, \qquad g \in \widehat{\mathcal{G}}_c(M, G),$$

from $\widehat{\mathcal{G}}_c(M, G)$ to $\widehat{\mathcal{G}}_{c_1}(M_1 \mid M, G)$.

6 Statement of the General Germ Expansions

In Sect. 3, we introduced asymptotic expansions for the invariant orbital integrals $J_G(\gamma, f) = f_G(\gamma)$. Our goal is now to establish formal germ expansions for the more general weighted orbital integrals $J_M(\gamma, f)$. We shall state the general expansions in this section. The proof of the expansions will then take up much of the remaining part of the paper.

Recall that the weighted orbital integrals depend on a choice of maximal compact subgroup $K \subset G(\mathbb{R})$, as well as the Levi subgroup M. The formal germ expansions will of course also depend on a fixed element $c \in \Gamma_{ss}(M)$. The theorem we are about to state asserts the existence of two families of objects attached to the 4-tuple (G, K, M, c), which depend also on bases

$$R_c(L) \subset \mathcal{D}_c(L), \qquad L \in \mathcal{L}(M),$$

chosen as in Lemma 2.1.

The first family is a collection of tempered distributions

$$f \longrightarrow J_L(\rho, f), \qquad L \in \mathcal{L}(M),\ \rho \in R_c(L), \tag{34}$$

on $G(\mathbb{R})$, which reduce to the invariant distributions

$$J_G(\rho, f) = f_G(\rho), \qquad \rho \in R_c(G),$$

when $L = G$, and in general are supported on the closed, $G(\mathbb{R})$-invariant subset $\mathcal{U}_c(G)$ of $G(\mathbb{R})$. The second family is a collection of formal germs

$$g_M^L(\rho), \qquad L \in \mathcal{L}(M),\ \rho \in R_c(L), \tag{35}$$

in $\widehat{\mathcal{G}}_c(M, L)$, which reduce to the homogeneous germs

$$g_M^M(\rho) = \rho^\vee, \qquad \rho \in R_c(M),$$

when $L = M$, and in general have the convergence property

$$\lim_{deg(\rho)\to\infty} \big(g_M^L(\rho)\big) = 0. \tag{36}$$

This implies that the series

$$g_M^L\big(J_{L,c}(f)\big) = \sum_{\rho\in R_c(L)} g_M^L(\rho) J_L(\rho,f)$$

converges in (the adic topology of) $\widehat{\mathcal{G}}_c(M, L)$, for any $f \in \mathcal{C}(G)$. The continuity of the linear forms (34) also implies that the mapping

$$f \longrightarrow g_M^L\big(J_{L,c}(f)\big)$$

from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M, L)$ is continuous (in the complex topology of $g\mathcal{G}_c(M, L)$.) The objects will also have a functorial property, which can be formulated as an assertion that for any L and f,

$$g_M^L\big(J_{L,c}(f)\big) \text{ is independent of the choice of basis } R_c(L). \tag{37}$$

The two families of objects will have other properties, which are parallel to those of weighted orbital integrals. If y lies in $G(\mathbb{R})$, the distributions (34) are to satisfy

$$J_L(\rho,f^y) = \sum_{Q\in\mathcal{F}(L)} J_L^{M_Q}(\rho,f_{Q,y}), \qquad f \in \mathcal{C}(G). \tag{38}$$

If z belongs to $\mathcal{Z}(G)$, we require that

$$J_L(\rho, zf) = J_L(z_L\rho,f) \tag{39}$$

and

$$g_M^L(\hat{z}_L\rho) = \sum_{S\in\mathcal{L}^L(M)} \partial_M^S(z_S) g_S^L(\rho)_M. \tag{40}$$

Finally, suppose that θ: $G \to \theta G$ is an isomorphism over $\mathbb{R}$. The two families of objects are then required to satisfy the symmetry conditions

$$J_{\theta L}(\theta\rho, \theta f) = J_L(\rho, f) \tag{41}$$

and

$$g_{\theta M}^{\theta L}(\theta\rho) = \theta g_M^L(\rho), \tag{42}$$

relative to the basis $R_{\theta c}(\theta L) = \theta R_c(L)$ of $\mathcal{D}_{\theta c}(\theta L)$.

Given objects (34) and (35), consider the sum

$$g_{M,c}(f) = \sum_{L \in \mathcal{L}(M)} g_M^L\big(J_{L,c}(f)\big).$$

Then $g_{M,c}$ is, a priori, a continuous map from $\mathcal{C}(G)$ to a space of formal germs that lack the property of symmetry by $W(M)$. However, suppose that $\theta = \mathrm{Int}(w)$, for a representative $w \in K$ of some element in the Weyl group $W(M)$. Then

$$\begin{aligned} \theta g_{M,c}(f) &= \sum_L \sum_\rho \theta g_M^L(\rho) \cdot J_L(\rho, f) \\ &= \sum_L \sum_\rho g_{\theta M}^{\theta L}(\theta\rho) J_{\theta L}(\theta\rho, \theta f), \end{aligned}$$

by (41) and (42). Since θM equals M and $J_{\theta L}(\theta\rho, \theta f)$ equals $J_{\theta L}(\theta\rho, f)$, we obtain

$$\begin{aligned} \theta g_{M,c}(f) &= \sum_L \sum_\rho g_M^{\theta L}(\theta\rho) J_{\theta L}(\theta\rho, f) \\ &= \sum_L \sum_\rho g_M^L(\rho) J_L(\rho, f) = g_{M,c}(f), \end{aligned}$$

from the condition (37). It follows that $g_{M,c}(f)$ is symmetric under $W(M)$, and therefore that $g_{M,c}(f)$ lies in the space $\widehat{\mathcal{G}}_c(M, G)$. In other words, $g_{M,c}$ can be regarded as a continuous linear map from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M, G)$.

Theorem 6.1. *There are distributions (34) and formal germs (35) such that the conditions (36)–(42) hold, and such that for any $f \in \mathcal{C}(G)$, the weighted orbital integral $J_M(f)$ has a formal germ expansion given by the sum*

$$\sum_{L \in \mathcal{L}(M)} g_M^L\big(J_{L,c}(f)\big) = \sum_{L \in \mathcal{L}(M)} \sum_{\rho \in R_c(L)} g_M^L(\rho) J_L(\rho, f). \tag{43}$$

Theorem 6.1 asserts that the sum (43) represents the same element in $\widehat{\mathcal{G}}_c(M,G)$ as $J_M(f)$. In other words, the weighted orbital integral has an asymptotic expansion

$$J_M(\gamma,f) \sim \sum_L \sum_\rho g_M^L(\gamma,\rho) J_L(\rho,f).$$

This is the archimedean analogue of the germ expansion for weighted orbital integrals on a p-adic group [A3, A8]. We should note that the formal germs $g_M^L(\rho)$ are more complicated in general than in the special case of $L = M = G$ treated in Sect. 3. For example, if $M = G$, the formal germs can be identified with homogeneous functions $g_G^G(\gamma,\rho) = \rho^\vee(\gamma)$. In the general case, each $g_M^L(\gamma,\rho)$ does have to be treated as an asymptotic series.

The functorial condition (37) seems entirely natural in the light of the main assertion of Theorem 6.1. We observe that (37) amounts to a requirement that the individual objects (34) and (35) be functorial in ρ. More precisely, suppose that for each L, $R_c'(L) = \{\rho'\}$ is a second basis of $\mathcal{D}_c(L)$. The condition (37) is equivalent to the transformation formulas

$$J_L(\rho',f) = \sum_\rho a_L(\rho',\rho) J_L(\rho,f) \tag{44}$$

and

$$g_M^L(\rho') = \sum_\rho a_L^\vee(\rho',\rho) g_M^L(\rho), \tag{45}$$

where $A_L = \{a_L(\rho',\rho)\}$ is the transformation matrix for the bases $\{\rho'\}$ and $\{\rho\}$, and $A_L^\vee = \{a_L^\vee(\rho',\rho)\} = {}^tA_L^{-1}$ is the transformation matrix for the dual bases $\{(\rho')^\vee\}$ and $\{\rho^\vee\}$. In the special case that $M = G$, these formulas are consequences of the constructions in Sects. 2 and 3 (as is (37)). In general, they follow inductively from this special case and the condition (37) (with L taken to be either M or G). The two formulas tell us that for any L, the two families of objects are functorial in the following sense. The distributions (34) are given by a mapping $f \to J_{L,c}(f)$ from $\mathcal{C}(G)$ to the dual space $\mathcal{D}_c(L)'$ such that

$$J_L(\rho,f) = \langle \rho, J_{L,c}(f)\rangle, \qquad \rho \in R_c(L).$$

The formal germs (35) are given by an element g_M^L in the (adic) tensor product $\widehat{\mathcal{G}}_c(M,L) \otimes \mathcal{D}_c(L)$ such that

$$\langle g_M^L, \rho^\vee\rangle = g_M^L(\rho), \qquad \rho \in R_c(L).$$

The distribution in (37) can thus be expressed simply as a pairing

$$g_M^L\big(J_{L,c}(f)\big) = \langle g_M^L, J_{L,c}(f)\rangle.$$

However, we shall retain the basis dependent notation (34) and (35), in deference to the traditional formulation of p-adic germ expansions.

The formal germ expansion for $J_M(f)$ is the main result of the paper. We shall actually need a quantitative form of the expansion, which applies to partial sums in the asymptotic series, and is slightly stronger than the assertion of Theorem 6.1.

It follows from (36) and the definition of the adic topology on $\widehat{\mathcal{G}}_c(M,L)$ that there is a weight function α such that $g_M^L(\rho)$ belongs to $\widehat{\mathcal{G}}_c^{\alpha}(M,L)$, for all L and ρ. Given such an α, and any $n \geq 0$, our conventions dictate that we write $g_M^{L,n}(\rho)$ for the projection of $g_M^L(\rho)$ onto the quotient $\mathcal{G}_c^{\alpha,n}(M,L)$ of $\widehat{\mathcal{G}}_c^{\alpha}(M,L)$, and $g_M^{L,n}(\gamma,\rho)$ for a representative of $g_M^{L,n}(\rho)$ in $\mathcal{F}_c^{\alpha}(V,L)$. We assume that $g_M^{L,n}(\gamma,\rho) = 0$, if $g_M^{L,n}(\rho) = 0$. The sum

$$J_M^n(\gamma,f) = \sum_L \sum_\rho g^{L,n}(\gamma,\rho) J_L(\rho,f) \tag{46}$$

can then be taken over a finite set. Our second theorem will include a slightly sharper form of the symmetry condition (42), namely that the functions $g_M^{L,n}(\gamma,\rho)$ can be chosen so that

$$g_{\theta M}^{\theta L,n}(\theta\gamma,\theta\rho) = g_M^{L,n}(\gamma,\rho), \tag{42*}$$

for θ as in (42). This condition, combined with the remarks prior to the statement of Theorem 6.1, tells us that (46) is invariant under the action of $W(M)$ on γ. The function $J_M^n(\gamma,f)$ therefore belongs to $\mathcal{F}_c^{\alpha}(V,G)$. It is uniquely determined up to a finite sum

$$\sum_i \phi_i(\gamma) J_i(f), \tag{47}$$

for tempered distributions $J_i(f)$ and functions $\phi_i(\gamma)$ in $\mathcal{F}_{c,n}^{\alpha}(V,G)$.

According to Corollary 4.2, we can choose α so that the weighted orbital integral $J_M(\gamma,f)$ also belongs to $\mathcal{F}_c^{\alpha}(V,G)$.

Theorem 6.1*. *We can choose the weight function α above so that $\alpha(1)$ equals 0 and the symmetry condition 42* is valid, and so that for any n, the mapping*

$$f \longrightarrow J_M(\gamma,f) - J_M^n(\gamma,f), \qquad f \in \mathcal{C}(G),$$

is a continuous linear transformation from $\mathcal{C}(G)$ to the space $\mathcal{F}_{c,n}^{\alpha}(V,G)$.

Remarks. 1. The statement of Theorem 6.1* is well posed, even though the mapping is determined only up to a finite sum (47). For (47) represents a continuous linear mapping from $\mathcal{C}(G)$ to $\mathcal{F}_{c,n}^{\alpha}(V,G)$. In other words, the difference

$$K_M^n(\gamma,f) = J_M(\gamma,f) - J_M^n(\gamma,f)$$

is defined up to a function that satisfies the condition of the theorem.

2. In concrete terms, Theorem 6.1* asserts the existence of a continuous seminorm $\mu^{\alpha}_{\sigma,\varepsilon,n}$ on $\mathcal{C}(G)$, for each $\sigma = (T, \Omega, X)$ in $S_c(M, G)$, each $\varepsilon > 0$, and each $n \geq 0$, such that

$$|XK^n_M(\gamma,f)| \leq \mu^{\alpha}_{\sigma,\varepsilon,n}(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)},$$

for every $\gamma \in V_\Omega$ and $f \in \mathcal{C}(G)$. This can be regarded as the analogue of Taylor's formula with remainder. The germ expansion of Theorem 6.1 is of course analogous to the asymptotic series provided by Taylor's theorem. In particular, Theorem 6.1* implies the germ expansion of Theorem 6.1.

We are going to prove Theorems 6.1 and 6.1* together. The argument will be inductive. We fix the 4-tuple of objects (G, K, M, c), and assume inductively that the two theorems have been established for any other 4-tuple (G_1, K_1, M_1, c_1), with

$$\dim(A_{M_1}/A_{G_1}) < \dim(A_M/A_G).$$

In particular, we assume that the distributions $J_L(\cdot,f)$ have been defined for any $L \supsetneq M$, that the formal germs $g^L_M(\cdot)$ have been defined for any $L \subsetneq G$, and that both sets of objects satisfy conditions of the theorems. Our task will be to construct distributions $J_M(\cdot,f)$ and formal germs $g^G_M(\cdot)$ that also satisfy the required conditions.

We shall begin the proof in the next section. In what remains of this section, let us consider the question of how closely the conditions of Theorem 6.1 come to determining the distributions and formal germs uniquely. Assume that we have been able to complete the induction argument by constructing the remaining distributions $J_M(\cdot,f)$ and formal germs $g^G_M(\cdot)$. To what degree are these objects determined by the distributions and formal germs for lower rank whose existence we have postulated?

Suppose for a moment that $\rho \in R_c(M)$ is fixed. Let ${}^*J_M(\rho,f)$ be an arbitrary distribution on $G(\mathbb{R})$ that is supported on $\mathcal{U}_c(G)$, and satisfies (38) (with $L = M$). That is, we suppose that

$${}^*J_M(\rho,f^y) = {}^*J_M(\rho,f) + \sum_{Q\in\mathcal{F}^0(M)} J^{M_Q}_M(\rho,f_{Q,y}),$$

for any $y \in G(\mathbb{R})$. Applying (38) to $J_M(\rho,f)$, we deduce that the difference

$$f \longrightarrow {}^*J_M(\rho,f) - J_M(\rho,f)$$

is an invariant tempered distribution that is supported on $\mathcal{U}_c(G)$. It follows that

$${}^*J_M(\rho,f) = J_M(\rho,f) + \sum_{\rho_G\in R_c(G)} c(\rho,\rho_G)f_G(\rho_G), \tag{48}$$

for complex coefficients $\{c(\rho,\rho_G)\}$ that vanish for almost all ρ_G.

Suppose now that ${}^*J_M(\cdot,f)$ and ${}^*g^G_M(\cdot)$ are arbitrary families of objects that satisfy the relevant conditions of Theorem 6.1. For each $\rho \in R_c(M)$, the distributions

${}^{*}J_M(\rho,f)$ and $J_M(\rho,f)$ then satisfy an identity (48), for complex coefficients

$$c(\rho_M,\rho_G), \qquad \rho_M\in R_c(M),\ \rho_G\in R_c(G), \tag{49}$$

that for any ρ_M, have finite support in ρ_G. The terms with $L\neq M$, G in the formal germ expansion (43) are assumed to have been chosen. It follows that the difference

$$g_M^G\big(J_{G,c}(f)\big) - {}^{*}g_M^G\big(J_{G,c}(f)\big) = \sum_{\rho_G\in R_c(G)} \big(g_M^G(\rho_G) - {}^{*}g_M^G(\rho_G)\big)f_G(\rho_G)$$

equals

$$\begin{aligned} g_M^M\big({}^{*}J_{M,c}(f)\big) - g_M^M\big(J_{M,c}(f)\big) &= \sum_{\rho_M\in R_c(M)} \rho_M^\vee\big({}^{*}J_M(\rho_M,f) - J_M(\rho_M,f)\big) \\ &= \sum_{\rho_G\in R_c(G)}\left(\sum_{\rho_M\in R_c(M)} \rho_M^\vee c(\rho_M,\rho_G)\right)f_G(\rho_G). \end{aligned}$$

Comparing the coefficients of $f_G(\rho_G)$, we find that

$${}^{*}g_M^G(\rho) = g_M^G(\rho) - \sum_{\rho_M\in R_c(M)} \rho_M^\vee c(\rho_M,\rho), \tag{50}$$

for any $\rho\in R_c(G)$. The general objects ${}^{*}J_M(\cdot,f)$ and ${}^{*}g_M^G(\cdot)$ could thus differ from the original ones, but only in a way that is quite transparent. Moreover, the coefficients (49) are governed by the conditions of Theorem 6.1. If z belongs to $\mathcal{Z}(G)$, they satisfy the equation

$$c(\rho_M,\hat{z}\rho_G) = c(z_M\rho_M,\rho_G). \tag{51}$$

They also satisfy the symmetry condition

$$c(\theta\rho_M,\theta\rho_G) = c(\rho_M,\rho_G), \tag{52}$$

for any isomorphism $\theta\colon G\to\theta G$ over $\mathbb{R}$. Finally, they satisfy the transformation formula

$$c(\rho_M',\rho_G') = \sum_{\rho_M}\sum_{\rho_G} a_M(\rho_M',\rho_M)c(\rho_M,\rho_G)a_G^\vee(\rho_G',\rho_G) \tag{53}$$

for change of bases, with matrices $\{a_M(\rho_M',\rho_M)\}$ and $\{a_G^\vee(\rho_G',\rho_G)\}$ as in (44) and (45).

Conversely, suppose that ${}^{*}J_M(\cdot,f)$ and ${}^{*}g_M^G(\cdot)$ are defined in terms of $J_M(\cdot,f)$ and $g_M^G(\cdot)$ by (48) and (50), for coefficients (49) that satisfy (51)–(53). It is then easy to see that ${}^{*}J_M(\cdot,f)$ and ${}^{*}g_M^G(\cdot)$ satisfy the conditions of Theorems 6.1 and 6.1*. We obtain

Proposition 6.2. *Assume that Theorems 6.1 and 6.1* are valid for distributions $J_L(\cdot,f)$ and formal germs $g_M^L(\cdot)$. Let ${}^*J_L(\cdot,f)$ and ${}^*g_M^L(\cdot)$ be secondary families of such objects for which ${}^*J_L(\cdot,f) = J_L(\cdot,f)$ if $L \neq M$, and ${}^*g_M^L(\cdot) = g_M^L(\cdot)$ if $L \neq G$. Then Theorems 6.1 and 6.1* are valid for ${}^*J_L(\cdot,f)$ and ${}^*g_M^L(\cdot)$ if and only if the relations (48) and (50) hold, for coefficients (49) that satisfy the conditions (51)–(53).*

7 Some Consequences of the Induction Hypotheses

We shall establish Theorems 6.1 and 6.1* over the next four sections. In these sections, G, K, M and c will remain fixed. We are assuming inductively that the assertions of the theorems are valid for any (G_1, K_1, M_1, c_1), with

$$\dim(A_{M_1}/A_{G_1}) < \dim(A_M/A_G).$$

In this section, we shall see what can be deduced directly from this induction assumption.

Let L be a Levi subgroup of G in $\mathcal{L}(M)$ that is distinct from both M and G. The terms in the series

$$g_M^L\big(J_{L,c}(f)\big) = \sum_{\rho \in R_c(L)} g_M^L(\rho) J_L(\rho, f), \qquad f \in \mathcal{C}(G),$$

are then defined, according to our induction assumption. The series converges to a formal germ in $\widehat{\mathcal{G}}_c(M,L)$ that is independent of the basis $R_c(L)$, as we see by applying (36), (44) and (45) inductively to L. Moreover, the mapping

$$f \longrightarrow g_M^L\big(J_{L,c}(f)\big), \qquad f \in \mathcal{C}(G),$$

is a continuous linear transformation from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M,L)$. We begin by describing three simple properties of this mapping.

Suppose that $y \in G(\mathbb{R})$. We can then consider the value

$$g_M^L\big(J_{L,c}(f^y)\big) = \sum_{\rho \in R_c(L)} g_M^L(\rho) J_L(\rho, f^y)$$

of the mapping at the y-conjugate of f. Since

$$\dim(A_L/A_G) < \dim(A_M/A_G),$$

we can apply the formula (38) inductively to $J_L(\rho, f^y)$. We obtain

$$\begin{aligned}\sum_{\rho \in R_c(L)} g_M^L(\rho) J_L(\rho, f^y) &= \sum_{\rho \in R_c(L)} \sum_{Q \in \mathcal{F}(L)} g_M^L(\rho) J_L^{M_Q}(\rho, f_{Q,y}) \\ &= \sum_{Q} \bigg(\sum_{\rho} g_M^L(\rho) J_L^{M_Q}(\rho, f_{Q,y}) \bigg).\end{aligned}$$

It follows that

$$g_M^L\big(J_{L,c}(f^y)\big) = \sum_{Q\in\mathcal{F}(L)} g_M^L\big(J_{L,c}^{M_Q}(f_{Q,y})\big), \qquad f \in \mathcal{C}(G). \tag{54}$$

Suppose that $z \in \mathcal{Z}(G)$. Consider the value

$$g_M^L\big(J_{L,c}(zf)\big) = \sum_{\rho\in R_c(L)} g_M^L(\rho) J_L(\rho, zf)$$

of the mapping at the z-transform of f. Since

$$\dim(A_L/A_G) < \dim(A_M/A_G),$$

we can apply the formula (39) to $J_L(\rho, zf)$. We obtain

$$\begin{aligned}\sum_{\rho\in R_c(L)} g_M^L(\rho) J_L(\rho, zf) &= \sum_{\rho} g_M^L(\rho) J_L(z_L\rho, f)\\ &= \sum_{\rho\in R_c(L)} g_M^L(\hat{z}_L\rho) J_L(\rho, f),\end{aligned}$$

by the definition of the transpose $\hat{z}_L$. Since

$$\dim(A_M/A_L) < \dim(A_M/A_G),$$

we can apply the formula (40) inductively to $g_M^L(\hat{z}_L\rho)$. We obtain

$$\begin{aligned}\sum_{\rho\in R_c(L)} g_M^L(\hat{z}_L\rho) J_L(\rho, f) &= \sum_{\rho} \sum_{S\in\mathcal{L}^L(M)} \big(\partial_M^S(z_S) g_S^L(\rho)_M\big) J_L(\rho, f)\\ &= \sum_{S} \partial_M^S(z_S) \left(\sum_{\rho} g_S^L(\rho)_M J_L(\rho, f)\right).\end{aligned}$$

It follows that

$$g_M^L\big(J_{L,c}(zf)\big) = \sum_{S\in\mathcal{L}^L(M)} \partial_M^S(z_S) g_S^L\big(J_{L,c}(f)\big)_M, \qquad f \in \mathcal{C}(G). \tag{55}$$

Finally, suppose that $\theta\colon G \to \theta G$ is an isomorphism over $\mathbb{R}$. Consider the composition

$$\theta g_M^L\big(J_{L,c}(f)\big) = \sum_{\rho\in R_c(L)} \theta g_M^L(\rho) \cdot J_L(\rho, f)$$

of the mapping with θ. Since

$$\dim(A_L/A_G) < \dim(A_M/A_G),$$

we can apply (41) to $J_L(\rho, f)$. Since

$$\dim(A_M/A_L) < \dim(A_M/A_G),$$

we can apply (42) to $g_M^L(\rho)$. It follows that

$$g_{\theta M}^{\theta L}\big(J_{\theta L,\theta c}(\theta f)\big) = \theta g_M^L\big(J_{L,c}(f)\big). \tag{56}$$

The main assertion of Theorem 6.1 is that the difference

$$K_M(f) = J_M(f) - \sum_{L \in \mathcal{L}(M)} g_M^L\big(J_{L,c}(f)\big),$$

regarded as an element in $\widehat{\mathcal{G}}_c(M, G)$, vanishes. We are not yet in a position to investigate this question, since we have not defined the terms in the series with $L = M$ and $L = G$. We consider instead the partial difference

$$\widetilde{K}_M(f) = J_M(f) - \sum_{\{L \in \mathcal{L}(M):\, L \neq M, G\}} g_M^L\big(J_{L,c}(f)\big), \qquad f \in \mathcal{C}(G), \tag{57}$$

regarded again as an element in $\widehat{\mathcal{G}}_c(M, G)$.

Lemma 7.1. *Suppose that $f \in \mathcal{C}(G)$ and $y \in G(\mathbb{R})$. Then*

$$\widetilde{K}_M(f^y) - \widetilde{K}_M(f) = \sum_{Q \in \mathcal{F}^0(M)} g_M^M\big(J_{M,c}^{M_Q}(f_{Q,y})\big). \tag{58}$$

Proof. The left-hand side of (58) equals

$$\big(J_M(f^y) - J_M(f)\big) - \sum_{L \neq M, G} \Big(g_M^L\big(J_{L,c}(f^y)\big) - g_M^L\big(J_{L,c}(f)\big)\Big).$$

We apply (21) to the term on the left, and (54) to each of the summands on the right. The expression becomes

$$\sum_{Q \in \mathcal{F}^0(M)} J_M^{M_Q}(f_{Q,y}) - \sum_{L \neq M} \sum_{Q \in \mathcal{F}^0(L)} g_M^L\big(J_{L,c}^{M_Q}(f_{Q,y})\big).$$

We take the second sum over Q outside the sum over L. The new outer sum is then over $Q \in \mathcal{F}^0(M)$, while the new inner sum is over Levi subgroups $L \in \mathcal{L}^{M_Q}(M)$ with $L \neq M$. Since $Q \neq G$, the formal germ $g_M^M\big(J_{M,c}^{M_Q}(f_{Q,y})\big)$ is defined, according to the induction assumption. We can therefore take the new inner sum over all elements $L \in \mathcal{L}^{M_Q}(M)$, provided that we then subtract the term corresponding to $L = M$. The left-hand side of (58) thus equals the sum of

$$\sum_{Q \in \mathcal{F}^0(M)} \left(J_M^{M_Q}(f_{Q,y}) - \sum_{L \in \mathcal{L}^{M_Q}(M)} g_M^L\big(J_{L,c}^{M_Q}(f_{Q,y})\big) \right)$$

and

$$\sum_{Q \in \mathcal{F}^0(M)} g_M^M\big(J_{M,c}^{M_Q}(f_{Q,y})\big).$$

The first of these expressions reduces to a sum,

$$\sum_{Q \in \mathcal{F}^0(M)} K_M^{M_Q}(f_{Q,y}),$$

whose terms vanish by our induction assumption. The second expression is just the right-hand side of (58). The formula (58) follows. □

In stating the next lemma, we write $f_{G,c}$ for the function

$$J_{G,c}(f): \ \rho \longrightarrow J_G(\rho, f) = f_G(\rho), \qquad \rho \in R_c(G),$$

to remind ourselves that it is invariant in f.

Lemma 7.2. *Suppose that $f \in \mathcal{C}(G)$ and $z \in \mathcal{Z}(G)$. Then*

$$\widetilde{K}_M(zf) - \partial\big(h(z)\big)\widetilde{K}_M(f) = \sum_{\{L \in \mathcal{L}(M):\, L \neq M\}} \partial_M^L(z_L) g_L^G(f_{G,c})_M. \tag{59}$$

Proof. The left-hand side of (59) equals

$$\big(J_M(zf) - \partial\big(h(z)\big)J_M(f)\big) - \sum_{L \neq M,G} \Big(g_M^L\big(J_{L,c}(zf)\big) - \partial\big(h(z)\big)g_M^L\big(J_{L,c}(f)\big)\Big).$$

We apply (22) to the term on the left, and (55) to each of the summands on the right. The expression becomes

$$\sum_S \partial_M^S(z_S)J_S(f) - \sum_{L,S} \partial_M^S(z_S)g_S^L\big(J_{L,c}(f)\big)_M,$$

where the first sum is over Levi subgroups $S \in \mathcal{L}(M)$ with $S \neq M$, and the second sum is over groups L and S in $\mathcal{L}(M)$ with

$$M \subsetneq S \subset L \subsetneq G.$$

This second sum can obviously be represented as an iterated sum over elements $S \in \mathcal{L}(M)$ with $S \neq M$, and elements $L \in \mathcal{L}(S)$ with $L \neq G$. Since $S \neq M$, the formal germ $g_S^L\big(J_{L,c}(f)\big)_M$ is defined, according to the induction assumption. We can therefore sum L over all elements in $\mathcal{L}(S)$, provided that we then subtract the term corresponding to $L = G$. The left-hand side of (59) thus equals the sum of

$$\sum_{\{S\in\mathcal{L}(M):\, S\neq M\}} \partial_M^S(z_S)\left(J_S(f) - \sum_{L\in\mathcal{L}(S)} g_M^L\big(J_{L,c}(f)\big)_M\right)$$

and

$$\sum_{\{S\in\mathcal{L}(M):\, S\neq M\}} \partial_M^S(z_S) g_S^G\big(J_{G,c}(f)\big)_M.$$

The first of these expressions reduces to a sum,

$$\sum_{S\neq M} \partial_M^S(z_S) K_S(f)_M,$$

whose terms vanish by our induction assumption. The second expression equals

$$\sum_{\{L\in\mathcal{L}(M):\, L\neq M\}} \partial_M^L(z_L) g_L^G(f_{G,c})_M,$$

the right-hand side of (59). The formula (59) follows. □

Lemmas 7.1 and 7.2 can be interpreted as identities

$$\widetilde{K}_M(\gamma, f^y) - \widetilde{K}_M(\gamma, f) = \sum_{Q\in\mathcal{F}^0(M)} g_M^M\big(\gamma, J_{M,c}^{M_Q}(f_{Q,y})\big), \qquad f \in \mathcal{C}(G),$$

and

$$\widetilde{K}_M(\gamma, zf) - \partial\big(h(z)\big)\widetilde{K}_M(\gamma, f) = \sum_{L\neq M} \partial_M^L(\gamma, z_L) g_L^G(\gamma, f_{G,c}), \qquad f \in \mathcal{C}(G),$$

of asymptotic series. What do these identities imply about the partial sums in the series? The question is not difficult, but in the case of Lemma 7.2 at least, it will require a precise answer.

As in the preamble to Theorem 6.1*, we can choose a weight function α such that for each $L \neq G$ and $\rho \in R_c(L)$, $g_M^L(\rho)$ belongs to $\widehat{\mathcal{G}}_c^{\alpha}(M, L)$. By applying the first assertion of Theorem 6.1* inductively to $(L, K \cap L, M, c)$ (in place of (G, K, M, c)), we see that α may be chosen so that $\alpha(1)$ equals zero. By Corollary 4.2, we can also assume that α is such that $f \to J_M(\gamma, f)$ is a continuous linear transformation from

$\mathcal{C}(G)$ to $\mathcal{F}_c^\alpha(V,G)$. Having chosen α, we set

$$\widetilde{J}_M^n(\gamma,f) = \sum_{L\neq M,G}\ \sum_{\rho\in R_c(L)} g_M^{L,n}(\gamma,\rho)J_L(\rho,f), \tag{60}$$

for any $n \geq 0$. The sums in this expression can be taken over finite sets, while the functions $g_M^{L,n}(\gamma,\rho)$ can be assumed inductively to satisfy the symmetry condition (42*). The function

$$\widetilde{K}_M^n(\gamma,f) = J_M(\gamma,f) - \widetilde{J}_M^n(\gamma,f), \qquad f\in\mathcal{C}(G), \tag{61}$$

is then invariant under the action of $W(M)$ on γ, and is uniquely determined up to a continuous linear mapping from $\mathcal{C}(G)$ to $\mathcal{F}_{c,n}^\alpha(V,G)$.

The following analogue of Corollary 4.2 is an immediate consequence of these remarks.

Lemma 7.3. *There is a weight function α with $\alpha(1)=0$, such that for any n, the mapping*

$$f \longrightarrow \widetilde{K}_M^n(\gamma,f), \qquad f\in\mathcal{C}(G),$$

defines a continuous linear transformation from $\mathcal{C}(G)$ to $\mathcal{F}_c^\alpha(V,G)$.

We shall now state our sharper form of Lemma 7.2. We assume for simplicity that c is not G-regular, or in other words, that the function D_c is nontrivial. The identity in Lemma 7.2 concerns an element $z\in\mathcal{Z}(G)$. In order to estimate the terms in this identity, we fix a triplet $\sigma=(T,\Omega,X)$ in $S_c(M,G)$. For any $n\geq 0$, we then write $k_{z,\sigma}^n(\gamma,f)$ for the function

$$X\big(\widetilde{K}_M^n(\gamma,zf) - \partial\big(h_T(z)\big)\widetilde{K}_M^n(\gamma,f)\big) - \sum_{L\neq M}\partial_M^L(\gamma,z_L)g_L^{G,n}(\gamma,f_{G,c})$$

of $\gamma\in V_\Omega$.

Lemma 7.4. *Given z and σ, we can choose a positive number a with the property that for any $n\geq 0$, the functional*

$$\nu_{z,\sigma}^{a,n}(f) = \sup_{\gamma\in V_\Omega}\big(|k_{z,\sigma}^n(\gamma,f)||D_c(\gamma)|^a\|\ell_c(\gamma)\|^{-(n+1)}\big), \qquad f\in\mathcal{C}(G),$$

is a continuous seminorm on $\mathcal{C}(G)$.

Proof. The assertion is a quantitative reformulation of Lemma 7.2 that takes into account its dependence on f. The proof is in principle the same. However, we do require a few preliminary comments to allow us to interpret the earlier argument.

There is of course some ambiguity in the definition of $k_{z,\sigma}^n(\gamma,f)$. The definition is given in terms of the restrictions to V_Ω of the functions $\widetilde{K}_M^n(\gamma,zf)$, $\widetilde{K}_M^n(\gamma,f)$ and

$g_L^{G,n}(\gamma, f_{G,c})$ in $\mathcal{F}_c^\alpha(V, G)$. For a given n, these functions are each defined only up to a continuous linear map from $\mathcal{C}(G)$ to $\mathcal{F}_{c,n}^\alpha(V, G)$. It is actually the images of the three functions under three linear transformations

$$\phi \longrightarrow X\phi_\Omega,$$

$$\phi \longrightarrow X\partial\big(h_T(z)\big)\phi_\Omega,$$

and

$$\phi \longrightarrow X\partial_M^L(\gamma, z_L)\phi_\Omega, \qquad \phi \in \mathcal{F}_c^\alpha(V, G),\ L \neq M,$$

that occur in the definition of $k_{z,\sigma}^n(\gamma, f)$. Each transformation is given by a linear partial differential operator on $T_{G\text{-reg}}(\mathbb{R})$ whose coefficients are at worst algebraic. Since $D_c \neq 1$, the notation of Sect. 5 simplifies slightly. Recalling the remark preceding the statement of Lemma 5.1, we see that there is a positive number a_0 with the property that for any $a \geq a_0$, and any n, each of the three linear transformations maps $\mathcal{F}_{c,n}^\alpha(V, G)$ continuously to $F_{c,n}^a(V, G)$. It follows that $k_{z,\sigma}^n(\gamma, f)$ is determined up to a continuous linear mapping from $\mathcal{C}(G)$ to $F_{c,n}^a(V, G)$. In other words, $k_{z,\sigma}^n(\gamma, f)$ is well-defined up to a function that satisfies the condition of the lemma. This means that it would suffice to establish the lemma with any particular choice for each of the three functions.

The main ingredient in the proof of Lemma 7.2 was the formula (55), which we can regard as an identity

$$g_M^L\big(\gamma, J_{L,c}(zf)\big) = \sum_{S \in \mathcal{L}^L(M)} \partial_M^S(\gamma, z_S) g_S^L\big(\gamma, J_{L,c}(f)\big)$$

of asymptotic series. To prove the lemma at hand, we need a corresponding identity of partial sums. For each S, we choose a weight function β such that the formal germ $g_S^L\big(J_{L,c}(f)\big)$ lies in $\widehat{\mathcal{G}}_c^\beta(S, L)$. For any positive integer m, $g_S^{L,m}\big(\gamma, J_{L,c}(f)\big)$ then denotes a representative in $\mathcal{F}_c^\beta(V, L)$ of the corresponding m-jet. After a moment's thought, it is clear that we can assign an integer $m > n$ to every n such that

$$g_M^{L,n}\big(\gamma, J_{L,c}(zf)\big) = \sum_{S \in \mathcal{L}^L(M)} \partial_M^S(\gamma, z_S) g_S^{L,m}\big(\gamma, J_{L,c}(f)\big).$$

The left-hand side here stands for some particular representative of $g_M^{L,n}\big(J_{L,c}(zf)\big)$ in $\mathcal{F}_c^\alpha(V, L)$, rather than the general one. Its sum over Levi subgroups $L \neq M, G$ yields a particular choice for the function $\widetilde{K}_M^n(\gamma, f)$ that occurs in the definition of $k_{z,\sigma}^n(\gamma, f)$. As we have noted, this is good enough for the proof of the lemma.

Armed with the last formula, we have now only to copy the proof of Lemma 7.2. A review of the earlier argument leads us directly to a formula

$$k_{z,\sigma}^n(\gamma, f) = \sum_{S \neq M} X\partial_M^S(\gamma, z_S) K_S^m(\gamma, f), \qquad \gamma \in V_\Omega.$$

We are free to apply Theorem 6.1* inductively to the summand $K_S^m(\gamma,f)$, since $S \neq M$. We thereby observe that $f \to K_S^m(\gamma,f)$ is a continuous linear transformation from $\mathcal{C}(G)$ to $\mathcal{F}_{c,m}^{\beta}(V,G)$. Since $D_c \neq 1$, we know from the discussion in Sect. 5 that there is an $a \geq a_0$ such that for any n, and any $m \geq n$, the linear transformation

$$\phi \longrightarrow X\partial_M^S(\gamma, z_S)\phi_\Omega, \qquad \phi \in \mathcal{F}_{c,m}^{\beta}(V,G),$$

maps $\mathcal{F}_{c,m}^{\beta}(V,G)$ continuously to $F_{c,n}^a(V_\Omega, G)$. It follows that for any n, the map

$$f \longrightarrow k_{z,\sigma}^n(\gamma,f), \qquad f \in \mathcal{C}(G),$$

is a continuous linear transformation from $\mathcal{C}(G)$ to $F_{c,n}^a(V_\Omega, G)$. The assertion of the lemma then follows from the definition of $F_{c,n}^a(V_\Omega, G)$. □

Recall that for any nonnegative integer N, $\mathcal{C}_{c,N}(G)$ denotes the subspace of $\mathcal{C}(G)$ annihilated by $R_{c,N}(G)$. This subspace is of finite codimension in $\mathcal{C}(G)$, and is independent of the choice of basis $R_c(G)$.

Lemma 7.5. *For any $n \geq 0$, we can choose an integer N so that if f belongs to $\mathcal{C}_{c,N}(G)$, the function $k_{z,\sigma}^n(\gamma,f)$ simplifies to*

$$k_{z,\sigma}^n(\gamma,f) = X\big(\widetilde{K}_M^n(\gamma, zf) - \partial\big(h_T(z)\big)\widetilde{K}_M^n(\gamma,f)\big), \qquad \gamma \in V_\Omega.$$

Proof. We have to show that the summands

$$g_L^{G,n}(\gamma, f_{G,c}) = \sum_{\rho \in R_c(G)} g_L^{G,n}(\gamma,\rho) f_G(\rho), \qquad L \neq M,$$

in the original definition of $k_{z,\sigma}^n(\gamma,f)$ vanish for the given f. Applying (36) inductively (with (G,M) replaced by (G,L)), we see that the (α,n)-jet $g_L^G(\rho)$ vanishes for all but finitely many ρ. We can therefore choose N so that for each L, the function $g_L^{G,n}(\gamma,\rho)$ vanishes for any ρ in the complement of $R_{c,N}(G)$ in $R_c(G)$. The lemma follows. □

Finally, we note that $\widetilde{K}_M(f)$ transforms in the obvious way under any isomorphism θ: $G \to \theta G$ over $\mathbb{R}$. If we apply (23) and (56) to the definition (57), we see immediately that

$$\widetilde{K}_{\theta M}(\theta f) = \theta \widetilde{K}_M(f), \qquad f \in \mathcal{C}(G). \tag{62}$$

Moreover, for any $n \geq 0$, the function (61) satisfies the symmetry condition

$$\widetilde{K}_{\theta M}^n(\theta\gamma, \theta f) = \widetilde{K}_M^n(\gamma,f), \quad \gamma \in V_{G\text{-reg}},\ f \in \mathcal{C}(G). \tag{63}$$

This follows from our induction assumption that the relevant terms in (60) satisfy (42*).

8 An Estimate

We have been looking at some of the more obvious implications of our induction hypothesis. We are now ready to begin a construction that will eventually yield the remaining objects $g_M^G(\rho)$ and $J_M(\rho, f)$. We shall carry out the process in the next section. The purpose of this section is to establish a key estimate for the mapping $\widetilde{K}_M$, which will be an essential part of the construction. The estimate is based on an important technique [H1] that Harish-Chandra developed from the differential equations (8).

Recall that $\widetilde{K}_M$ is a continuous linear transformation from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M, G)$. We choose a weight function α as in Lemma 7.3. For any $n \geq 0, f \to \widetilde{K}_M^n(f)$ and $f \to \widetilde{K}_M^n(\gamma, f)$ then represent continuous linear mappings from $\mathcal{C}(G)$ onto the respective spaces $\mathcal{G}_c^{\alpha,n}(M, G)$ and $\mathcal{F}_c^{\alpha}(V, G)$. To focus the discussion, let us write ψ_M^n for the restriction of $\widetilde{K}_M^n$ to some given subspace $\mathcal{C}_{c,N}(G)$ of $\mathcal{C}(G)$. Then

$$f \longrightarrow \psi_M^n(\gamma, f) = \widetilde{K}_M^n(\gamma, f), \qquad f \in \mathcal{C}_{c,N}(G),$$

is a continuous linear mapping from $\mathcal{C}_{c,N}(G)$ to $\mathcal{F}_c^{\alpha}(V, G)$. For all intents and purposes, we shall take N to be any integer that is large relative to n, and that in particular has the property of Lemma 7.5. This will lead us to an estimate for $\psi_M^n(\gamma, f)$ that is stronger than the bound implied by the definition of $\mathcal{F}_c^{\alpha}(V, G)$.

To simplify the statement of the estimate, we may as well rule out the trivial case that c is G-regular, as we did in Lemma 7.4. In other words, we assume that $\dim(G_c/T) > 0$, for any maximal torus $T \in \mathcal{T}_c(M)$. We fix T, together with a connected component $\Omega \in \pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big)$. Consider the open subset

$$V_\Omega(a, n) = \big\{\gamma \in V_\Omega : \, |D_c(\gamma)|^{-a} \|\ell_c(\gamma)\|^n < 1\big\}$$

of $T_{G\text{-reg}}(\mathbb{R})$, defined for any $a > 0$ and any nonnegative integer n. Our interest will be confined to the case that the closure of $V_\Omega(a, n)$ contains c. This condition will obviously be met if n is large relative to a, or more precisely, if n is greater than the integer

$$a^+ = a \dim(G_c/T).$$

According to our definitions, any function in the space $\mathcal{F}_{c,n}^{\alpha}(V, G)$ will be bounded on $V_\Omega(a, n)$, for any $a > \alpha(1)$. (We assume of course that the invariant function $\ell_c(\gamma)$ is bounded on V.) The function $\psi_M^n(\gamma, f)$ above lies a priori only in the larger space $\mathcal{F}_c^{\alpha}(V, G)$. However, the next lemma asserts that for N large, the restriction of $\psi_M^n(\gamma, f)$ to $V_\Omega(a, n)$ is also bounded.

More generally, we shall consider the derivative $X\psi_M^n(\gamma, f)$, for any (translation) invariant differential operator X on $T(\mathbb{R})$. Given X, we assume that a is greater than the positive number

$$\alpha^+(X) = \alpha(X) + \deg(X) \dim(G_c/T)^{-1}.$$

Then if $n > a^+$, as above, and $\varepsilon > 0$ is small, n will be greater than $\deg(X)$, and

$$|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)} = |D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{n+1-\deg X}$$
$$\leq C|D_c(\gamma)|^{-a}\|\ell_c(\gamma)\|^{n+1}, \qquad \gamma \in V_{\mathrm{reg}},$$

for some constant C. It follows that the X-transform of any function in $\mathcal{F}^{\alpha}_{c,n}(V, G)$ is bounded by a constant multiple of $\|\ell_c(\gamma)\|$ on $V_\Omega(a, n)$, and is therefore absolutely bounded on $V_\Omega(a, n)$. We are going to show that for N large, the function $X\psi^n_M(\gamma, f)$ is also bounded on $V_\Omega(a, n)$.

Lemma 8.1. *Given the triplet $\sigma = (T, \Omega, X)$, we can choose a positive integer $a > \alpha^+(X)$ with the property that for any $n > a^+$, and for N large relative to n, the function*

$$f \longrightarrow \sup_{\gamma\in V_\Omega(a,n)} |X\psi^n_M(\gamma, f)|, \qquad f \in \mathcal{C}_{c,N}(G),$$

is a continuous seminorm on $\mathcal{C}_{c,N}(G)$.

Proof. We should first check that the statement of the lemma is well posed, even though the function $\psi^n_M(\gamma, f)$ is not uniquely determined. As in the remark preceding the statement of Theorem 6.1*, we observe that $\psi^n_M(\gamma, f)$ is defined only up to a finite sum

$$\sum_i \phi_i(\gamma)J_i(f), \qquad f \in \mathcal{C}_{c,N}(G),$$

for tempered distributions $J_i(f)$ and functions $\phi_i(\gamma)$ in $\mathcal{F}^{\alpha}_{c,n}(V, G)$. From the discussion above, we see that the function

$$\sum_i \big(X\phi_i(\gamma)\big)J_i(f)$$

is bounded on $V_\Omega(a, n)$, and in fact, can be bounded by a continuous seminorm in f. In other words, $X\psi^n_M(\gamma, f)$ is well defined up to a function that satisfies the condition of the lemma. The condition therefore makes sense for $X\psi^n_M(\gamma, f)$.

Let $u_1 = 1, u_2, \ldots, u_q$ be a basis of the G-harmonic elements in $S\big(\mathfrak{t}(\mathbb{C})\big)$. Any element in $S\big(\mathfrak{t}(\mathbb{C})\big)$ can then be written uniquely in the form

$$\sum_j u_j h_T(z_j), \qquad z_j \in \mathcal{Z}(G).$$

For any $n \geq 0, f \in \mathcal{C}_{c,N}(G)$, and $\gamma \in V_\Omega$, we write

$$\psi^n_i(\gamma, f) = \psi^n_{M,i}(\gamma, f) = \partial(u_i)\psi^n_M(\gamma, f), \qquad 1 \leq i \leq q.$$

Our aim is to estimate the functions

$$\partial(u)\psi^n_i(\gamma, f), \qquad u \in S\big(\mathfrak{t}(\mathbb{C})\big),\ 1 \leq i \leq q. \tag{64}$$

The assertion of the lemma will then follow from the case $i = 1$ and $X = \partial(u)$.

Consider a fixed element $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$. For any i, we can write

$$uu_i = \sum_{j=1}^{q} h_T(z_{ij})u_j,$$

for operators $z_{ij} = z_{u,ij}$ in $\mathcal{Z}(G)$. This allows us to write (64) as the sum of

$$\sum_j \psi_j^n(\gamma, z_{ij}f)$$

with

$$\sum_j \partial(u_j)\big(\partial\big(h_T(z_{ij})\big)\psi_M^n(\gamma,f) - \psi_M^n(\gamma, z_{ij}f)\big). \tag{65}$$

We shall estimate the two expressions separately.

The first step is to apply Lemma 7.4 to the summands in (65). For any given n, we choose N to be large enough that the summands have the property of Lemma 7.5. In other words, the expression (65) is equal to a sum of functions

$$-\sum_j k^n_{z_{ij},\sigma_j}(\gamma,f), \qquad \sigma_j = \big(T, \Omega, \partial(u_j)\big),\ f \in \mathcal{C}_{c,N}(G),$$

defined as in the preamble to Lemma 7.4. Applying Lemma 7.4 to each summand, we obtain a positive number a with the property that for any n, and for each i and j, the functional

$$\nu^{a,n}_{z_{ij},\sigma_j}(f) = \sup_{\gamma\in V_\Omega}\big(|k^n_{z_{ij},\sigma_j}(\gamma,f)||D_c(\gamma)|^a\|\ell_c(\gamma)\|^{-(n+1)}\big)$$

is a continuous seminorm on $\mathcal{C}_{c,N}(G)$. Given a, we write $\nu_u^{a,n}(f)$ for the supremum over $1 \leq i \leq q$ and γ in $V_\Omega(a,n)$ of the absolute value of (65). It then follows from the definition of $V_\Omega(a,n)$ that

$$\nu_u^{a,n}(f) \leq C_0 \sup_i\Big(\sum_j \nu^n_{z_{ij},\sigma_j}(f)\Big), \qquad f \in \mathcal{C}_{c,N}(G),$$

where

$$C_0 = \sup_{\gamma\in V_\Omega(a,n)} \|\ell_c(\gamma)\|.$$

We conclude that $\nu_u^{a,n}$ is a continuous seminorm on $\mathcal{C}_{c,N}(G)$. The exponent a depends on the elements $z_{ij} \in \mathcal{Z}(G)$, and these depend in turn on the original elements u. It will be best to express this dependence in terms of an arbitrary positive integer d.

For any such d, we can choose an exponent $a = a_d$ so that for any $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\deg(u) \leq d$, the functional $\nu_u^{a,n}(f)$ is a continuous seminorm on $\mathcal{C}_{c,N}(G)$.

The next step is to combine the estimate we have obtained for (65) with the estimate for the functions

$$\psi_j^n(\gamma, z_{ij}f) = \partial(u_j)\widetilde{K}_M^n(\gamma, z_{ij}f)$$

provided by Lemma 7.3. It is a consequence of this lemma that there is an integer b such that for any n, i, and j, the mapping

$$f \longrightarrow \psi_j^n(\gamma, z_{ij}f), \qquad f \in \mathcal{C}_{c,N}(G),$$

is a continuous linear transformation from $\mathcal{C}_{c,N}(G)$ to $F_c^b(V_\Omega, G)$. (Here, N can be any nonnegative integer.) In other words, each functional

$$\sup_{\gamma \in V_\Omega} \big(|D_c(\gamma)|^b |\psi_j^n(\gamma, z_{ij}f)|\big), \qquad f \in \mathcal{C}_{c,N}(G),$$

is a continuous seminorm on $\mathcal{C}_{c,N}(G)$. We can now handle both expressions in the original decomposition of (64). Our conclusion is that there is a continuous seminorm μ_u^n on $\mathcal{C}_{c,N}(G)$ such that

$$|\partial(u)\psi_i^n(\gamma, f)| \leq |D_c(\gamma)|^{-b}\mu_u^n(f) + \nu_u^{a,n}(f), \qquad f \in \mathcal{C}_{c,N}(G),$$

for every γ in $V_\Omega(a, n)$. In particular,

$$|\partial(u)\psi_i^n(\gamma, f)| \leq \mu_u^{a,n}(f)|D(\gamma)|^{-b}, \qquad \gamma \in V_\Omega(a, n),\ f \in \mathcal{C}_{c,N}(G),$$

where

$$\mu_u^{a,n}(f) = \mu_u^n(f) + \Big(\sup_{\gamma \in V_\Omega(a,n)} |D_c(\gamma)|^b\Big)\nu_u^{a,n}(f)$$

is a continuous seminorm on $\mathcal{C}_{c,N}(G)$. We need be concerned only with the index $i = 1$. We shall write

$$\psi_u^n(\gamma, f) = \partial(u)\psi_M^n(\gamma, f) = \partial(u)\psi_1^n(\gamma, f),$$

in this case. The last estimate is then

$$|\psi_u^n(\gamma, f)| \leq \mu_u^{a,n}(f)|D_c(\gamma)|^{-b}, \qquad \gamma \in V_\Omega(a, n),\ f \in \mathcal{C}_{c,N}(G),$$

for any element $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\deg(u) \leq d$. Our task is to establish a stronger estimate, in which $b = 0$. We emphasize that in the estimate we have already obtained, b is independent of u, and therefore also of d and $a = a_d$. It is this circumstance that allows an application of the technique of Harish-Chandra from [H1].

For any $\delta > 0$, set

$$V_{\Omega,\delta}(a,n) = \big\{\gamma \in V_\Omega(a,n) : \ \|\ell_c(\gamma)\| < \delta\big\}.$$

If γ belongs to the complement of $V_{\Omega,\delta}(a,n)$ in $V_\Omega(a,n)$, we have

$$|D_c(\gamma)|^a > \|\ell_c(\gamma)\|^n \geq \delta^n.$$

It follows that the function $|D_c(\gamma)|^b$ is bounded away from 0 on the complement of $V_{\Omega,\delta}(a,n)$ in $V_\Omega(a,n)$. We have therefore only to show that for some δ, the function

$$\sup_{\gamma \in V_{\Omega,\delta}(a,n)} \big(|\psi_u^n(\gamma, f)|\big), \qquad f \in \mathcal{C}_{c,N}(G), \tag{66}$$

is a continuous seminorm on $\mathcal{C}_{c,N}(G)$.

Given a and n, we simply choose any $\delta > 0$ that is sufficiently small. We then assign a vector $H \in \mathfrak{t}(\mathbb{R})$ to each point γ in $V_{\Omega,\delta}(a,n)$ in such a way that the line segments

$$\psi_t = \gamma \exp tH, \qquad \gamma \in V_{\Omega,\delta}(a,n),\ 0 \leq t \leq 1, \tag{67}$$

are all contained in $V_\Omega(a,n)$, and the end points $\gamma_1 = \gamma \exp H$ all lie in the complement of $V_{\Omega,\delta}(a,n)$ in $V_\Omega(a,n)$. We can in fact arrange that the correspondence $\gamma \to H$ has finite image in $\mathfrak{t}(\mathbb{R})$. We can also assume that the points (67) satisfy an inequality

$$|D_c(\gamma_t)|^{-1} \leq C_1 t^{-\dim(G_c/T)}, \qquad 0 < t \leq 1,$$

where C_1 is a constant that is independent of the starting point γ in $V_{\Omega,\delta}(a,n)$. Setting p equal to the product of $\dim(G_c/T)$ with the integer b, and absorbing the constant C_1^b in the seminorm $\mu_u^{a,n}(f)$ above, we obtain an estimate

$$|\psi_u^n(\gamma_t, f)| \leq \mu_u^{a,n}(f) t^{-p}, \qquad f \in \mathcal{C}_{c,N}(G),$$

for each of the points γ_t in (67), and any $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\deg(u) \leq d$.

The last step is to apply the argument from [H1, Lemma 49]. Observe that

$$\frac{d}{dt}\psi_u^n(\gamma_t, f) = \partial(H)\psi_u^n(\gamma_t, f) = \psi_{Hu}^n(\gamma_t, f).$$

Therefore

$$\left|\frac{d}{dt}\psi_u^n(\gamma_t, f)\right| \leq \mu_{Hu}^{a,n}(f) t^{-p},$$

for any γ_t as in (67), and any $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\det(u) \leq d - 1$. Combining this estimate with the fundamental theorem of calculus, we obtain

$$\begin{aligned}|\psi_u^n(\gamma_t,f)| &\leq \left|\int_t^1 \Big(\frac{d}{ds}\psi_u^n(\gamma_s,f)\Big)ds\right| + |\psi_u^n(\gamma_1,f)| \\ &\leq \int_t^1 \mu_{Hu}^{a,n}(f)s^{-p}ds + \mu_u^{a,n}(f) \\ &\leq \Big(\frac{1}{p-1}\Big)\mu_{Hu}^{a,n}(f)(t^{-p+1}-1) + \mu_u^{a,n}(f).\end{aligned}$$

It follows that there is a continuous seminorm $\mu_{u,1}^{a,n}$ on $\mathcal{C}_{c,N}(G)$ such that

$$|\psi_u^n(\gamma_t,f)| \leq \mu_{u,1}^{a,n}(f)t^{-p+1},$$

for any γ_t as in (67), and any $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\deg(u) \leq d-1$. Following the proof of [H1, Lemma 49], we repeat this operation p times. We obtain a continuous seminorm $\mu_{u,p}^{a,n}$ on $\mathcal{C}_{c,N}(G)$ such that

$$|\psi_u^n(\gamma_t,f)| \leq \mu_{u,p}^{a,n}(f)|\log t|,$$

for any γ_t as in (67), and any $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\deg(u) \leq d-p$. Repeating the operation one last time, and using the fact that $\log t$ is integrable over $[0,1]$, we conclude that there is a continuous seminorm $\lambda_u^{a,n} = \mu_{u,p+1}^{a,n}$ on $\mathcal{C}_{c,N}(G)$ such that

$$|\psi_u^n(\gamma_t,f)| \leq \lambda_u^{a,n}(f), \qquad f \in \mathcal{C}_{c,N}(G),$$

for all γ_t as in (67), and any $u \in S\big(\mathfrak{t}(\mathbb{C})\big)$ with $\deg(u) \leq d-(p+1)$. Setting $t=0$, we see that the supremum (66) is bounded by $\lambda_u^{a,n}(f)$, and is therefore a continuous seminorm on $\mathcal{C}_{c,N}(G)$.

We have now finished. Indeed, for the given differential operator $X = \partial(u)$, we set

$$d = \deg(u) + b\,\dim(G_c/T) + 1 = \deg(u) + p + 1,$$

where b is the absolute exponent above. We then take a to be the associated number a_d. Given a, together with a positive integer n, we choose $\delta > 0$ as above. The functional (66) is then a continuous seminorm on $\mathcal{C}_{c,N}(G)$. As we have seen, this yields a proof of the lemma. □

Corollary 8.2. *Given the triplet $\sigma = (T,\Omega,X)$, we can choose a positive number $a > \alpha^+(X)$ with the property that for any $n > a^+$, and for N large relative to n, the limit*

$$\chi_M(\sigma,f) = \lim_{\gamma\to c} X\psi_M^n(\gamma,f), \qquad \gamma \in V_\Omega(a,n),\ f \in \mathcal{C}_{c,N}(G), \tag{68}$$

exists, and is continuous in f.

Proof. Once again, the statement is well posed, even though $\psi_M^n(\gamma,f)$ is defined only up to a function

$$\sum_i \phi_i(\gamma)J_i(f)$$

in $\mathcal{F}_{c,n}^{\alpha}(V,G)$. For it follows from the preamble to Lemma 8.1 that the X-transform of any function in $\mathcal{F}_{c,n}^{\alpha}(V,G)$ can be written as a product of $\ell_c(\gamma)$ with a function that is bounded on $V_\Omega(a,n)$. In particular, $X\psi_M^n(\gamma,f)$ is well defined up to a function on $V_\Omega(a,n)$ whose limit at c vanishes.

Given σ, and thus X, we choose a so that the assertion of the lemma holds for all the differential operators

$$\partial(H)X, \qquad H \in \mathfrak{t}(\mathbb{C}).$$

The first derivatives

$$\partial(H)X\psi_M^n(\gamma,f), \qquad \gamma \in V_\Omega(a,n),\ f \in \mathcal{C}_{c,N}(G),$$

of the function $X\psi_M^n(\gamma,f)$ are then all bounded on $V_\Omega(a,n)$ by a fixed, continuous seminorm in f. It follows that the function $\gamma \to X\psi_M^n(\gamma,f)$ extends continuously to the closure of $V_\Omega(a,n)$ in a way that is also continuous in f. The limit $\chi_M(\sigma,f)$ therefore exists, and is continuous in f. $\square$

Remarks. 1. As the notation suggests, the limit $\chi_M(\sigma,f)$ is independent of n. For if $m > n$, $\psi_M^m(\gamma,f)$ differs from $\psi_M^n(\gamma,f)$ by a function of γ that lies in $\mathcal{F}_{c,n}^{\alpha}(V,G)$. As we noted at the beginning of the proof of the corollary, the X-transform of any such function converges to 0 as γ approaches c in $V_\Omega(a,n)$. Of course n must be large relative to $\deg(X)$, and N has in turn to be large relative to n. The point is that for any $\sigma = (T,\Omega,X)$, and for N sufficiently large relative to $\deg(X)$, the limit

$$\chi_M(\sigma,f), \qquad f \in \mathcal{C}_{c,N}(G),$$

can be defined in terms of any appropriately chosen n.

2. Lemma 8.1 and Corollary 8.2 were stated under the assumption that $\dim(G_c/T) > 0$. The excluded case that $\dim(G_c/T) = 0$ is trivial. For in this case, the function $\psi_M^n(\gamma,f)$ on V_Ω extends to a smooth function in an open neighbourhood of c. The lemma and corollary then hold for any n.

9 The Mapping $\tilde{\chi}_M$

We fix a weight function α satisfying the conditions of Lemma 7.3, as we did in the last section. Then $\alpha(1)$ equals zero, and the continuous mapping

$$\widetilde{K}_M : \mathcal{C}(G) \longrightarrow \widehat{\mathcal{G}}_c(M,G)$$

takes values in the subspace $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$ of $\widehat{\mathcal{G}}_c(M,G)$. Recall that the space $\widehat{\mathcal{I}}_c(M,G)$ introduced in Sect. 5 is contained in $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$. The goal of this section is to construct a continuous linear mapping

$$\tilde{\chi}_M : \mathcal{C}(G) \longrightarrow \widehat{\mathcal{I}}_c(M,G)$$

that approximates $\widetilde{K}_M$.

The main step will be the next proposition, which applies to the restrictions

$$\psi_M^n = \widetilde{K}_M^n : \mathcal{C}_{c,N}(G) \longrightarrow \mathcal{G}_c^{\alpha,n}(M,G)$$

treated in the last section.

Proposition 9.1. *Suppose that $n \geq 0$, and that N is large relative to n. Then there is a uniquely determined continuous linear transformation*

$$\chi_M^n : \mathcal{C}_{c,N}(G) \longrightarrow \mathcal{I}_c^n(M,G),$$

such that for any $f \in \mathcal{C}_{c,N}(G)$, the image of $\chi_M^n(f)$ in $\mathcal{G}_c^{\alpha,n}(M,G)$ equals $\psi_M^n(f)$. More precisely, the mapping

$$f \longrightarrow \psi_M^n(\gamma,f) - \chi_M^n(\gamma,f), \qquad f \in \mathcal{C}_{c,N}(G), \qquad (69)$$

is a continuous linear transformation from $\mathcal{C}_{c,N}(G)$ to the space $\mathcal{F}_{c,n}^{\alpha}(V,G)$.

Proof. Recall that $\mathcal{I}_c(M,G)$ is contained in the space $\mathcal{G}_c^{bd}(M,G)$ of bounded germs. The first step is to construct χ_M^n as a mapping from $\mathcal{C}_{c,N}(G)$ to the quotient $\mathcal{G}_c^{bd,n}(M,G)$ of $\mathcal{G}_c^{bd}(M,G)$. As we observed in the proof of Lemma 5.3, any element in $\mathcal{G}_c^{bd,n}(M,G)$ can be identified with a family

$$\phi^n = \left\{\phi_\Omega^n : T \in \mathcal{T}_c(M),\ \Omega \in \pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big)\right\}$$

of Taylor polynomials of degree n (in the coordinates $\ell_c(\gamma)$) on the neighbourhoods V_Ω. In particular, $\mathcal{G}_c^{bd,n}(M,G)$ is finite dimensional. The subspace $\mathcal{I}_c^n(M,G)$ consists of those families that satisfy Harish-Chandra's jump conditions (15).

Suppose that f belongs to $\mathcal{C}_{c,N}(G)$, for some N that is large relative to n. We define $\chi_M^n(f)$ as a family of Taylor polynomials of degree n by means of the limits $\chi_M(\sigma,f)$ provided by Corollary 8.2. More precisely, we define

$$\chi_{M,\Omega}^n(\gamma,f), \qquad T \in \mathcal{T}_c(M),\ \Omega \in \pi_{0,c}\big(T_{G\text{-reg}}(\mathbb{R})\big),\ \gamma \in V_\Omega,$$

to be the polynomial of degree n such that

$$\lim_{\gamma \to c}\big(X\chi_{M,\Omega}^n(\gamma,f)\big) = \chi_M(\sigma,f), \qquad \gamma \in V_\Omega, \qquad (70)$$

where X ranges over the invariant differential operators on $T(\mathbb{R})$ of degree less than or equal to n, and where $\sigma = (T, \Omega, X)$. If (T, Ω) is replaced by a pair (T', Ω') that is $W(M)M(\mathbb{R})$-conjugate to (T, Ω), the corresponding polynomial $\chi^n_{M,\Omega'}(\gamma', f)$ is $W(M)M(\mathbb{R})$-conjugate to $\chi^n_{M,\Omega}(\gamma, f)$. This follows from Corollary 8.2 and the analogous property for the (α, n)-jet $\psi^n_M(\gamma, f)$. Therefore $\chi^n_M(f)$ is a well defined element in $\mathcal{G}^{bd,n}_c(M, G)$. Moreover, Corollary 8.2 tells us that each limit (70) is continuous in f. It follows that $f \to \chi^n_M(f)$ is a continuous linear map from $\mathcal{C}_{c,N}(G)$ to the finite dimensional space $\mathcal{G}^{bd,n}_c(M, G)$.

The main step will be to establish the continuity of the mapping (69), defined for some weight α that satisfies the conditions of Lemma 7.3. This amounts to showing that for any triplet $\sigma = (T, \Omega, X)$ in $S_c(M, G)$, and any $\varepsilon > 0$, there is a continuous seminorm $\mu(f)$ on $\mathcal{C}_{c,N}(G)$ such that

$$\left|X\big(\psi^n_M(\gamma, f) - \chi^n_M(\gamma, f)\big)\right| \leq \mu(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)}, \tag{71}$$

for $\gamma \in V_\Omega$. Observe that if deg $|X| > n$, (71) reduces to an inequality

$$|X\widetilde{K}^n_M(\gamma, f)| \leq \mu(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)},$$

since $X\chi^n_M(\gamma, f) = 0$ and $(n, X) = 0$. We know from Lemma 7.3 that such an inequality actually holds for any f in the full Schwartz space $\mathcal{C}(G)$. We may therefore assume that $\deg(X) \leq n$. We shall derive (71) in this case from four other inequalities, in which $\mu_1(f)$, $\mu_2(f)$, $\mu_3(f)$, and $\mu_4(f)$ denote four continuous seminorms on $\mathcal{C}_{c,N}(G)$.

We have first to combine Taylor's formula with Lemma 8.1. This lemma actually applies only to the case that $\dim(G_c/T) > 0$. However, if $\dim(G_c/T) = 0$, $D_c(\gamma)$ equals 1, and the weight function α plays no role. In this case, the estimate (71) is a direct application of Taylor's formula, which we can leave to the reader. We shall therefore assume that $\dim(G_c/T) > 0$.

We have fixed data n, $\sigma = (T, \Omega, X)$ and ε, with $\deg(X) \leq n$, for which we are trying to establish (71). For later use, we also fix a positive number ε', with $\varepsilon' < \varepsilon$. At this point, we have removed from circulation the symbols X and n in terms of which Lemma 8.1 was stated, so our application of the lemma will be to a pair of objects denoted instead by Y and m. We allow Y to range over the invariant differential operators on $T(\mathbb{R})$ with $\deg(Y) \leq n + 1$. If

$$a = a_n > \sup_Y \big(\alpha^+(Y)\big) = \overline{\alpha}(n+1) + (n+1)\dim(G_c/T)^{-1},$$

as in Lemma 8.1, we choose m with

$$m > a_n^+ = a_n \dim(G_c/T).$$

The lemma applies to functions $f \in \mathcal{C}_{c,N}(G)$, for N large relative to m (which is the same as being large relative to n, if m is fixed in terms of n). In combination with the fundamental theorem of calculus, it tells us that $\psi^m_M(\gamma, f)$ extends to a function on an open neighbourhood of the closure of $V_\Omega(a_n, m)$ that is continuously differentiable of order $(n+1)$. The derivatives of this function at $\gamma = c$ are the limits treated in Corollary 8.2. They are independent of m, and can be identified with the coefficients of the polynomial $\chi_M^n(\gamma, f)$ on V_Ω. We can therefore regard $\chi_M^n(\gamma, f)$ as the Taylor polynomial of degree n at $\gamma = c$ (relative to the coordinates $\ell_c(\gamma)$) of the function $\psi_M^m(\gamma, f)$ on $V_\Omega(a_n, m)$. Now if γ belongs to $V_\Omega(a_n, m)$, the set

$$\lambda_t(\gamma) = c \exp\big(t\ell_c(\gamma)\big), \qquad 0 < t < 1,$$

is contained in $V_\Omega(a_n, m)$, and may be regarded as the line segment joining c with γ. Applying the bound of Lemma 8.1 to the remainder term (of order $(n+1)$) in Taylor's theorem, we obtain an estimate

$$\big|X(\psi_M^m(\gamma, f) - \chi_M^n(\gamma, f))\big| \leq \mu_1(f)\|\ell_c(\gamma)\|^{(n,X)},$$

for any γ in $V_\Omega(a_n, m)$. We are assuming that $m > n$ and that α satisfies the conditions of Lemma 7.3. The projection of $\psi_M^m(f)$ onto $\mathcal{G}_c^{\alpha,n}(M, G)$ therefore exists, and is equal to $\psi_M^n(f)$. The definitions then yield a second estimate

$$\big|X(\psi_M^m(\gamma, f) - \psi_M^n(\gamma, f))\| \leq \mu_2(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)},$$

that is valid for any γ in V_Ω. Combining the two estimates, we see that

$$\big|X\big(\psi_M^n(\gamma, f) - \chi_M^n(\gamma, f)\big)\big| \leq \mu_3(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)}, \tag{72}$$

for any γ in $V_\Omega(a_n, m)$.

The functions $\psi_M^n(\gamma, f)$ and $\chi_M^n(\gamma, f)$ in (72) both belong to the space $\mathcal{F}_c^\alpha(V, G)$. Applying the estimate that defines this space to each of the functions, we obtain a bound

$$\big|X\big(\psi_M^n(\gamma, f) - \chi_M^n(\gamma, f)\big)\big| \leq \mu_4(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon')},$$

that holds for every γ in V_Ω. Suppose that γ lies in the complement of $V_\Omega(a_n, m)$ in V_Ω. Then

$$|D_c(\gamma)| \leq \|\ell_c(\gamma)\|^{m'},$$

for the exponent $m' = ma_n^{-1}$. Setting $\delta = \varepsilon - \varepsilon' > 0$, we write

$$|D_c(\gamma)|^{-(\alpha(X)+\varepsilon')} = |D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}|D_c(\gamma)|^{\delta} \leq |D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{\delta m'}.$$

We are free to choose m to be as large as we like. In particular, we can assume that

$$\delta m' \geq (n, X),$$

and therefore that

$$|D_c(\gamma)|^{-(\alpha(X)+\varepsilon')} \leq C'|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)},$$

for some constant C'. Absorbing C' in the seminorm $\mu_4(f)$, we conclude that

$$\left|X\big(\psi^n_M(\gamma,f) - \chi^n_M(\gamma,f)\big)\right| \leq \mu_4(f)|D_c(\gamma)|^{-(\alpha(X)+\varepsilon)}\|\ell_c(\gamma)\|^{(n,X)}, \tag{73}$$

for any γ in the complement of $V_\Omega(a_n, m)$ in V_Ω.

The estimates (72) and (73) account for all the points γ in V_Ω. Together, they yield an estimate of the required form (71), in which we can take

$$\mu(f) = \mu_3(f) + \mu_4(f).$$

We have established the required assertion that for N large relative to n,

$$f \longrightarrow \psi^n_M(\gamma,f) - \chi^n_M(\gamma,f), \qquad f \in \mathcal{C}_{c,N}(G),$$

is a continuous linear transformation from $\mathcal{C}_{c,N}(G)$ to $\mathcal{F}^\alpha_{c,n}(V,G)$. From this, it follows from the definitions that the image of $\chi^n_M(f)$ in $\mathcal{G}^{\alpha,n}_c(M,G)$ equals $\psi^n_M(f)$. In particular, $\psi^n_M(f)$ lies in the subspace $\mathcal{G}^{bd,n}_c(M,G)$ of $\mathcal{G}^n_c(M,G)$.

The space $\mathcal{I}^n_c(M,G)$ is in general a proper subspace of $\mathcal{G}^{bd,n}_c(M,G)$, by virtue of the extra constraints imposed by the jump conditions (15). The last step is to show that for suitable N, χ^n_M takes $\mathcal{C}_{c,N}(G)$ to the smaller space $\mathcal{I}^n_c(M,G)$. This will be an application of Lemma 7.1.

Let ξ be a linear form on the finite dimensional space $\mathcal{G}^{bd,n}_c(M,G)$ that vanishes on the subspace $\mathcal{I}^n_c(M,G)$. The mapping

$$J_\xi : f \longrightarrow \xi\big(\chi^n_M(f)\big) = \xi\big(\psi^n_M(f)\big), \qquad f \in \mathcal{C}_{c,N}(G),$$

is continuous, and is therefore the restriction to $\mathcal{C}_{c,N}(G)$ of a tempered distribution. Suppose that

$$f = h^y - h, \qquad h \in \mathcal{C}(G),\ y \in G(\mathbb{R}).$$

Then

$$\begin{aligned} J_\xi(f) &= \xi\big(\psi^n_M(h^y - h)\big) \\ &= \xi\big(\widetilde{K}^n_M(h^y) - \widetilde{K}^n_M(h)\big) \\ &= \sum_{Q\in\mathcal{F}^0(M)} \xi\big(g^{M,n}_M\big(J^{M_Q}_{M,c}(h_{Q,y})\big)\big), \end{aligned}$$

by Lemma 7.1. The sum of the n-jets

$$g^{M,n}_M\big(J^{M_Q}_{M,c}(h_{Q,y})\big), \qquad Q \in \mathcal{F}^0(M),$$

lies in the subspace $\mathcal{I}_c^n(M,G)$ on which ξ vanishes. The distribution J_ξ thus annihilates any function of the form $h^y - h$, and is therefore invariant. On the other hand, if $f_0 \in \mathcal{C}(G)$ is compactly supported, and vanishes on a neighbourhood of the closed invariant subset $\mathcal{U}_c(G)$ of $G(\mathbb{R})$, one sees easily from the definitions that $\widetilde{K}_M^n(f_0)$ equals 0. It follows that the distribution J_ξ is supported on $\mathcal{U}_c(G)$. We have established that J_ξ belongs to the space $\mathcal{D}_c(G)$, and is therefore a finite linear combination of distributions in the basis $R_c(G)$. Increasing N if necessary, we can consequently assume that for each ξ, J_ξ annihilates the space $\mathcal{C}_{c,N}(G)$. In other words, ψ_M^n takes any function $f \in \mathcal{C}_{c,N}(G)$ to the subspace $\mathcal{I}_c^n(M,G)$ of $\mathcal{G}_c^{bd,n}(M,G)$. Since $\psi_M^n(f)$ equals $\chi_M^n(f)$, the image of χ_M^n is also contained in $\mathcal{I}_c^n(M,G)$.

We have now proved that for N large relative to n, $f \to \chi_M^n(f)$ is a continuous linear mapping from $\mathcal{C}_{c,N}(G)$ to $\mathcal{I}_c^n(M,G)$. We have also shown that the image of $\chi_M^n(f)$ in $\mathcal{G}_c^{\alpha,n}(M,G)$ equals $\psi_M^n(f)$. But Lemma 5.3 implies that the mapping of $\mathcal{I}_c^n(M,G)$ into $\mathcal{G}_c^{\alpha,n}(M,G)$ is injective. We conclude that $\chi_M^n(f)$ is uniquely determined. With this last observation, the proof of the proposition is complete. □

The germs $\chi_M^n(f)$ share some properties with the (α,n)-jets $\widetilde{K}_M^n(f)$ from which they were constructed. For example, suppose that $m \geq n$, and that N is large relative to m. Then if f belongs to $\mathcal{C}_{c,N}(G)$, both $\chi_M^n(f)$ and $\chi_M^m(f)$ are defined. But $\psi_M^n(f) = \widetilde{K}_M^n(f)$ is the projection of $\psi_M^m(f) = \widetilde{K}_M^m(f)$ onto $\mathcal{G}_c^{\alpha,n}(M,G)$. It follows that $\chi_M^n(f)$ is the projection of $\chi_M^m(f)$ onto $\mathcal{I}_M^n(M,G)$.

We can reformulate this property in terms of the dual pairing between $\mathcal{D}_c(M)$ and $\widehat{\mathcal{I}}_c(M)$. Recall that $\widehat{\mathcal{I}}_c(M,G)$ is the subspace of $W(M)$-invariant elements in $\widehat{\mathcal{I}}_c(M)$. The pairing

$$\langle \sigma, \phi \rangle, \quad \sigma \in \mathcal{D}_c(M),\ \phi \in \widehat{\mathcal{I}}_c(M,G), \tag{74}$$

therefore identifies $\widehat{\mathcal{I}}_c(M,G)$ with the dual $\mathcal{D}_c(M)^*_{W(M)}$ of the space $\mathcal{D}_c(M)_{W(M)}$ of $W(M)$-covariants of $\mathcal{D}_c(M)$. If σ belongs to the finite dimensional subspace

$$\mathcal{D}_{c,n}(M) = \{\sigma \in \mathcal{D}_c(M) :\ \deg(\sigma) \leq n\}$$

of $\mathcal{D}_c(M)$ spanned by $R_{c,n}(M)$, the value

$$\langle \sigma, \phi^n \rangle = \langle \sigma, \phi \rangle$$

depends only on the image ϕ^n of ϕ in $\mathcal{I}_c^n(M,G)$. With this notation, we set

$$\langle \sigma, \chi_M(f) \rangle = \langle \sigma, \chi_M^n(f) \rangle, \qquad \sigma \in \mathcal{D}_c(M),\ f \in \mathcal{C}_{c,N}(G), \tag{75}$$

for any $n \geq \deg(\sigma)$ and for N large relative to n. In view of the projection property above, the pairing (75) is independent of the choice of n. It is defined for any N that is large relative to $\deg(\sigma)$.

Another property that $\chi_M^n(f)$ inherits is the differential equation (59) satisfied by $\widetilde{K}_M(f)$. We shall state it in terms of the pairing (75).

Lemma 9.2. *Suppose that $z \in \mathcal{Z}(G)$ and $\sigma \in \mathcal{D}_c(M)$, and that N is large relative to* $\deg(\sigma) + \deg(z)$. *Then*

$$\langle \sigma, \chi_M(zf)\rangle = \langle z_M\sigma, \chi_M(f)\rangle, \tag{76}$$

for any $f \in \mathcal{C}_{c,N}(G)$.

Proof. The assertion is a reformulation of Lemma 7.2 in terms of the objects $\chi_M^n(f)$. Its proof, like that of Lemmas 7.4 and 7.5, is quite straightforward. We can afford to be brief.

We choose positive integers $n_1 \geq \deg(\sigma)$ and $n \geq n_1 + \deg(z)$, and we assume that N is large relative to n. If f belongs to $\mathcal{C}_{c,N}(G)$, zf belongs to the space $\mathcal{C}_{c,N_1}(G)$, where $N_1 = N - \deg(z)$ is large relative to n_1. We can therefore take the pairing

$$\langle \sigma, \chi_M(zf)\rangle = \langle \sigma, \chi_M^{n_1}(zf)\rangle.$$

We can also form the pairing

$$\langle z_M\sigma, \chi_M(f)\rangle = \langle z_M\sigma, \chi_M^n(f)\rangle,$$

which can be written as

$$\big\langle \sigma, \partial\big(h(z)\big)\chi_M^n(f)\big\rangle = \big\langle \sigma, \big(\partial\big(h(z)\big)\chi_M^n(f)\big)^{n_1}\big\rangle,$$

since the action of z_M on $\mathcal{D}_c(M)$ is dual to the action of $\partial\big(h(z)\big)$ on $\widehat{\mathcal{I}}_c(M,G)$. We have to show that the difference

$$\langle \sigma, \chi_M(zf)\rangle - \langle z_M\sigma, \chi_M(f)\rangle = \big\langle \sigma, \chi_M^{n_1}(zf) - \big(\partial\big(h(z)\big)\chi_M^n(f)\big)^{n_1}\big\rangle$$

vanishes.

Combining Lemma 7.2 with the various definitions, we see that

$$\begin{aligned}
&\chi_M^{n_1}(zf) - \big(\partial\big(h(z)\big)\chi_M^n(f)\big)^{n_1}\\
=&\psi_M^{n_1}(zf) - \big(\partial\big(h(z)\big)\psi_M^n(f)\big)^{n_1}\\
=&\big(\widetilde{K}_M^{n_1}(zf) - \partial\big(h(z)\big)\widetilde{K}_M^n(f)\big)^{n_1}\\
=&\big(\widetilde{K}_M(zf) - \partial\big(h(z)\big)\widetilde{K}_M(f)\big)^{n_1}\\
=&\sum_{L\neq M}\big(\partial_M^L(z_L)g_L^G(f_{G,c})_M\big)^{n_1}\\
=&\sum_{L\neq M}\big(\partial_M^L(z_L)g_L^{G,n}(f_{G,c})_M\big)^{n_1}.
\end{aligned}$$

We apply (36) inductively to the formal germs

$$g_L^{G,n}(f_{G,c})_M = \sum_{\rho \in R_c(G)} g_L^{G,n}(\rho)_M f_G(\rho), \qquad L \neq M,$$

as in the proof of Lemma 7.5. Since N is large relative to n, we conclude that these objects all vanish. Equation (76) follows. □

Finally, it is clear that $\chi_M^n(f)$ inherits the symmetry property (62), relative to an isomorphism $\theta\colon G \to \theta G$ over $\mathbb{R}$. If N is large relative to a given $\sigma \in \mathcal{D}_c(M)$, we obtain

$$\langle \theta\sigma, \chi_{\theta M}(\theta f)\rangle = \langle \sigma, \chi_M(f)\rangle, \tag{77}$$

for any $f \in \mathcal{C}_{c,N}(G)$. This property will be of special interest in the case that θ belongs to the group $\mathrm{Aut}(G, K, M, c)$ of automorphisms of the 4-tuple (G, K, M, c).

We shall now construct $\tilde{\chi}_M$ as an extension of the family of mappings $\{\chi_M^n\}$.

Proposition 9.3. *There is a continuous linear mapping*

$$\tilde{\chi}_M : \mathcal{C}(G) \to \widehat{\mathcal{I}}_c(M, G) \tag{78}$$

that satisfies the restriction condition

$$\langle \sigma, \tilde{\chi}_M(f)\rangle = \langle \sigma, \chi_M(f)\rangle, \qquad f \in \mathcal{C}_{c,N}(G), \tag{79}$$

for any $\sigma \in \mathcal{D}_c(M)$ *and* N *large relative to* $\deg(\sigma)$*, the differential equation*

$$\langle \sigma, \tilde{\chi}_M(zf)\rangle = \langle z_M\sigma, \tilde{\chi}_M(f)\rangle, \qquad z \in \mathcal{Z}(G),\ f \in \mathcal{C}(G), \tag{80}$$

for any $\sigma \in \mathcal{D}_c(M)$*, and the symmetry condition*

$$\langle \theta\sigma, \tilde{\chi}_M(\theta f)\rangle = \langle \sigma, \tilde{\chi}_M(f)\rangle, \qquad f \in \mathcal{C}(G), \tag{81}$$

for any $\sigma \in \mathcal{D}_c(M)$ *and* $\theta \in \mathrm{Aut}(G, K, M, c)$.

Proof. Let

$$\mathcal{D}_{c,1}(M) = \mathcal{D}_{c,G\text{-harm}}(M)$$

be the space of G-harmonic elements in $\mathcal{D}_c(M)$. This is a finite dimensional subspace of $\mathcal{D}_c(M)$, which is invariant under the action of $W(M)$ and, more generally, the group $\mathrm{Aut}(G, K, M, c)$. (Here we are regarding c as a $W(M)$-orbit in $\Gamma_{ss}(M)$.) Choose a positive integer N_1 that is large enough that the pairing (75) is defined for every σ in $\mathcal{D}_{c,1}(M)$ and f in $\mathcal{C}_{c,N_1}(G)$. We thereby obtain a continuous linear map

$$\chi_{M,1} : \mathcal{C}_{c,N_1}(G) \longrightarrow \mathcal{D}_{c,1}(M)^*,$$

which by (77) is fixed under the action of the group $\mathrm{Aut}(G, K, M, c)$. Let

$$\tilde{\chi}_{M,1} : \mathcal{C}(G) \longrightarrow \mathcal{D}_{c,1}(M)^* \tag{82}$$

be any linear extension of this mapping to $\mathcal{C}(G)$ that remains fixed under the action of $\mathrm{Aut}(G, K, M, c)$. Since $\mathcal{C}_{c,N_1}(G)$ is of finite codimension in $\mathcal{C}(G)$, $\tilde{\chi}_{M,1}$ is automatically continuous. With this mapping, we obtain a pairing $\langle \sigma, \tilde{\chi}_{M,1}(f) \rangle$, for elements $\sigma \in \mathcal{D}_{c,1}(M)$ and functions $f \in \mathcal{C}(G)$, that satisfies (79) and (81).

The extension of the pairing to all elements $\sigma \in \mathcal{D}_c(M)$ is completely determined by the differential equations (80). According to standard properties of harmonic polynomials, the map

$$z \otimes \sigma \longrightarrow z_M \sigma, \qquad z \in \mathcal{Z}(G),\ \sigma \in \mathcal{D}_{c,1}(M),$$

is a linear isomorphism from $\mathcal{Z}(G) \otimes \mathcal{D}_{c,1}(M)$ onto $\mathcal{D}_c(M)$. Any element in $\mathcal{D}_c(M)$ is therefore a finite linear combination of elements

$$\sigma = z_{1,M}\sigma_1, \qquad z_1 \in \mathcal{Z}(G),\ \sigma_1 \in \mathcal{D}_{c,1}(M).$$

For any σ of this form, we define

$$\langle \sigma, \tilde{\chi}_M(f) \rangle = \langle \sigma_1, \tilde{\chi}_{M,1}(z_1 f) \rangle, \qquad f \in \mathcal{C}(G).$$

Since $z_{1,M}$ is $W(M)$-invariant, the values taken by this pairing at a given f determine a linear form on the quotient $\mathcal{D}_c(M)_{W(M)}$ of $\mathcal{D}_c(M)$. The pairing therefore defines a continuous mapping (78) that satisfies the differential equation (80). The restriction condition (79) follows from (80), the associated differential equation (76) for $\langle \sigma, \chi_M(f) \rangle$, and the special case of $\sigma \in \mathcal{D}_{c,1}(M)$ that was built into the definition. The symmetry condition (81) follows from the compatibility of $\mathrm{Aut}(G, K, M, c)$ with the action of $\mathcal{Z}(G)$, and again the special case of $\sigma \in \mathcal{D}_{c,1}(M)$. Our construction is complete. □

Remarks. 1. The three properties of $\tilde{\chi}_M$ can of course be stated without recourse to the pairing (74). The restriction condition (79) can be formulated as the commutativity of the diagram

$$\begin{array}{ccc}
\mathcal{C}_{c,N}(G) & \xrightarrow{\chi_M^n} & \mathcal{I}_c^n(M,G) \\
\big\downarrow & & \big\uparrow \\
\mathcal{C}(G) & \xrightarrow{\tilde{\chi}_M} & \widehat{\mathcal{I}}_c(M,G),
\end{array}$$

for any $n \geq 0$ and N large relative to n. The differential equation (80) has a dual version

$$\tilde{\chi}_M(zf) = \partial\big(h(z)\big)\,\tilde{\chi}_M(f), \qquad z \in \mathcal{Z}(G),\ f \in \mathcal{C}(G),$$

that is similar to (8). The symmetry condition (81) is essentially just the equation

$$\theta\big(\tilde{\chi}_M(f)\big) = \tilde{\chi}_M(\theta f), \qquad \theta \in \mathrm{Aut}(G, K, M, c),\ f \in \mathcal{C}(G).$$

2. The mapping $\tilde{\chi}_M$ is completely determined up to translation by an $\mathrm{Aut}(G, K, M, c)$-fixed linear transformation

$$C:\ \mathcal{C}(G)/\mathcal{C}_{c,N_1}(G) \longrightarrow \mathcal{D}_{c,G\text{-harm}}(M)^*. \tag{83}$$

The space of such linear transformations is of course finite dimensional.

10 Completion of the Proof

We shall now complete the proof of Theorems 6.1 and 6.1*. We have to construct distributions (34) with $L = M$, and formal germs (35) with $L = G$, that satisfy the conditions (36)–(42). The key to the construction is the mapping

$$\tilde{\chi}_M:\ \mathcal{C}(G) \longrightarrow \widehat{\mathcal{I}}_c(M, G)$$

of Proposition 9.3.

The distributions (34) are in fact built into $\tilde{\chi}_M$. If ρ belongs to $R_c(M)$, we simply set

$$J_M(\rho, f) = \langle \rho, \tilde{\chi}_M(f)\rangle, \qquad f \in \mathcal{C}(G). \tag{84}$$

Since $\tilde{\chi}_M$ is continuous, the linear forms

$$f \longrightarrow J_M(\rho, f), \qquad \rho \in R_c(M),$$

are tempered distributions. We must check that they are supported on $\mathcal{U}_c(G)$.

Suppose that f_0 is a function in $\mathcal{C}(G)$ that is compactly supported, and vanishes on a neighbourhood of $\mathcal{U}_c(G)$. As we noted near the end of the proof of Proposition 9.1, the definitions imply that the (α, n)-jet

$$\widetilde{K}^n_M(f_0) = J^n_M(f_0) - \sum_{L \neq M, G}\ \sum_{\rho_L \in R_c(L)} g^{L,n}_M(\rho_L) J_L(\rho_L, f_0)$$

vanishes for $n \geq 0$. Indeed, the weighted orbital integral $J_M(\gamma, f_0)$ vanishes for γ near $\mathcal{U}_c(G)$, while our induction hypothesis includes the assumption that the distributions $J_L(\rho_L, f)$ are supported on $\mathcal{U}_c(G)$. Given $\rho \in R_c(M)$, we choose any $n \geq \deg(\rho)$. Then

$$\begin{aligned} J_M(\rho, f_0) &= \langle \rho, \tilde{\chi}_M(f_0) \rangle = \langle \rho, \chi_M(f_0) \rangle \\ &= \langle \rho, \chi_M^n(f_0) \rangle = \langle \rho, \psi_M^n(f_0) \rangle \\ &\qquad\qquad = \langle \rho, \widetilde{K}_M^n(f_0) \rangle = 0, \end{aligned}$$

by (79), (75) and Proposition 9.1. The distribution $J_M(\rho, f)$ is therefore supported on $\mathcal{U}_c(G)$.

We have now constructed the distributions (34), in the remaining case that $L = M$. The required conditions (37)–(39) (with $L = M$) amount to properties of $\tilde{\chi}_M$ we have already established. The functorial condition (37) concerns the (adically) convergent series

$$g_M^M\big(J_{M,c}(f)\big) = \sum_{\rho \in R_c(M)} \rho^\vee J_M(\rho, f).$$

Observe that

$$\rho\big(g_M^M\big(J_{M,c}(f)\big)\big) = J_M(\rho, f) = \langle \rho, \tilde{\chi}_M(f) \rangle,$$

for any $\rho \in R_c(M)$. It follows that

$$g_M^M\big(J_{M,c}(f)\big) = \tilde{\chi}_M(f), \qquad f \in \mathcal{C}(G). \tag{85}$$

Since $\tilde{\chi}_M(f)$ was constructed without recourse to the basis $R_c(M)$, the same is true of $g_M^M\big(J_{M,c}(f)\big)$.

To check the variance condition (38), we note that for $f \in \mathcal{C}(G)$ and $y \in G(\mathbb{R})$, the function $f^y - f$ belongs to each of the spaces $\mathcal{C}_{c,N}(G)$. Given $\rho \in R_c(M)$, we may therefore write

$$\begin{aligned} J_M(\rho, f^y - f) &= \langle \rho, \tilde{\chi}_M(f^y - f) \rangle = \langle \rho, \chi_M(f^y - f) \rangle \\ &= \langle \rho, \psi_M(f^y - f) \rangle = \langle \rho, \widetilde{K}_M(f^y - f) \rangle \\ &= \sum_{Q \in \mathcal{F}^0(M)} \rho\big(g_M^M\big(J_{M,c}^{M_Q}(f_{Q,y})\big)\big) \\ &= \sum_{Q \in \mathcal{F}^0(M)} J_M^{M_Q}(\rho, f_{Q,y}), \end{aligned}$$

by Lemma 7.1. The formula (38) follows.

The differential equation (39) is even simpler. If $z \in \mathcal{Z}(G)$ and $\rho \in R_c(M)$, we use (80) to write

$$\begin{aligned} J_M(\rho, zf) &= \langle \rho, \tilde{\chi}_M(zf) \rangle \\ &= \langle z_M \rho, \tilde{\chi}_M(f) \rangle = J_M(z_M \rho, f). \end{aligned}$$

This is the required equation.

To deal with the other assertions of the two theorems, we set

$$K'_M(f) = \widetilde{K}_M(f) - \tilde{\chi}_M(f) = \widetilde{K}_M(f) - g_M^M\big(J_{M,c}(f)\big).$$

Then $f \to K'_M(f)$ is a continuous linear transformation from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M, G)$, which can be expanded in the form

$$K'_M(f) = J_M(f) - \sum_{L \in \mathcal{L}^0(M)} \sum_{\rho \in R_c(L)} g_M^L(\rho) J_L(\rho, f).$$

Suppose that α is as before, a weight function that satisfies the conditions of Lemma 7.3. Then $K'_M(f)$ lies in the subspace $\widehat{\mathcal{G}}_c^{\alpha}(M, G)$ of $\widehat{\mathcal{G}}_c(M, G)$. For any n, $K'_M(f)$ projects to the (α, n)-jet

$$K'^{,n}_M(f) = \widetilde{K}^n_M(f) - \tilde{\chi}^n_M(f)$$

in $\mathcal{G}_c^{\alpha,n}(M, G)$, which in turn comes with a representative

$$\begin{aligned} K'^{,n}_M(\gamma, f) &= \widetilde{K}^n_M(\gamma, f) - \tilde{\chi}^n_M(\gamma, f) \\ &= J_M(\gamma, f) - \sum_{L \in \mathcal{L}^0(M)} \sum_{\rho \in R_c(L)} g_M^{L,n}(\gamma, \rho) J_L(\rho, f) \end{aligned}$$

in $\mathcal{F}_c^{\alpha}(V, G)$. We shall use the mappings $K'^{,n}_M$ to construct the remaining germs (35). The argument at this point is quite similar to that of the p-adic case [A3, §9].

Since we are considering the case $L = G$ of (35), we take ρ to be an element in the basis $R_c(G)$. If N is a large positive integer, let f_ρ^N denote a function in $\mathcal{C}(G)$ with the property that for any ρ_1 in the subset $R_{c,N}(G)$ of $R_c(G)$, the condition

$$f^N_{\rho,G}(\rho_1) = \begin{cases} 1, \text{ if } \rho_1 = \rho, \\ 0, \text{ if } \rho_1 \neq \rho, \end{cases} \tag{86}$$

holds. Suppose that $n \geq 0$. Taking N to be large relative to n, we define

$$g_M^{G,n}(\rho) = K'^{,n}_M(f^N_\rho). \tag{87}$$

Then $g_M^{G,n}(\rho)$ is an element in $\mathcal{G}_c^{\alpha,n}(M,G)$. Suppose that N' is another integer, with $N' \geq N$, and that $f_\rho^{N'}$ is a corresponding function (86). The difference

$$f_\rho^{N,N'} = f_\rho^N - f_\rho^{N'}$$

then lies in $\mathcal{C}_{c,N}(G)$. From Propositions 9.1 and 9.3, we see that

$$\begin{aligned} K_M^{\prime,n}(f_\rho^N) - K_M^{\prime,n}(f_\rho^{N'}) &= K_M^{\prime,n}(f_\rho^{N,N'}) \\ &= \widetilde{K}_M^n(f_\rho^{N,N'}) - \tilde{\chi}_M^n(f_\rho^{N,N'}) \\ &= \psi_M^n(f_\rho^{N,N'}) - \chi_M^n(f_\rho^{N,N'}) = 0. \end{aligned}$$

It follows that the (α,n)-jet $g_M^{G,n}(\rho)$ depends only on ρ and n. It is independent of both N and the choice of function f_ρ^N.

Suppose that $m \geq n$, and that N is large relative to m. Then

$$g_M^{G,m}(\rho) = K_M^{\prime,m}(f_\rho^N)$$

is an element in $\mathcal{G}_c^{\alpha,m}(M,G)$. Since the image of $K_M^{\prime,m}(f_\rho^N)$ in $\mathcal{G}_c^{\alpha,n}(M,G)$ equals $K_M^{\prime,n}(f_\rho^N)$, by definition, the image of $g_M^{G,m}(\rho)$ in $\mathcal{G}_c^{\alpha,n}(M,G)$ equals $g_M^{G,n}(\rho)$. We conclude that the inverse limit

$$g_M^G(\rho) = \varprojlim_n g_M^{G,n}(\rho)$$

exists, and defines an element in $\widehat{\mathcal{G}}_c^\alpha(M,G)$. This completes our construction of the formal germs (35), in the remaining case that $L = G$. As we agreed in Sect. 5, we can represent them by asymptotic series

$$g_M^G(\gamma,\rho) = \sum_{n=0}^{\infty} g_M^{G,(n)}(\gamma,\rho),$$

where

$$g_M^{G,(n)}(\gamma,\rho) = g_M^{G,n}(\gamma,\rho) - g_M^{G,n-1}(\gamma,\rho),$$

and

$$g_M^{G,n}(\gamma,\rho) = K_M^{\prime,n}(\gamma, f_\rho^N) = \widetilde{K}_M^n(\gamma, f_\rho^N) - \tilde{\chi}_M^n(\gamma, f_\rho^N). \tag{88}$$

Suppose that N is large relative to n, and that ρ lies in the complement of $R_{c,N}(G)$ in $R_c(G)$. Taking $f_\rho^N = 0$ in this case, we deduce that

$$g_M^{G,n}(\rho) = K_M^{\prime,n}(f_\rho^N) = 0.$$

In other words, $g_M^{G,n}(\rho)$ vanishes whenever $\deg(\rho)$ is large relative to n. This is the property (36). It implies that for any $f \in \mathcal{C}(G)$, the series

$$g_M^G(f_{G,c}) = g_M^G\big(J_{G,c}(f)\big) = \sum_{\rho \in R_c(G)} g_M^G(\rho) f_{G,c}(\rho)$$

converges in (the adic topology of) $\widehat{\mathcal{G}}_c(M, G)$.

Before we establish other properties of the formal germs $g_M^G(\rho)$, let us first prove the main assertion of Theorem 6.1*. Having defined the series $g_M^G(f_{G,c})$, we set

$$\begin{aligned} K_M(f) &= K'_M(f) - g_M^G(f_{G,c}) \\ &= J_M(f) - \sum_{L \in \mathcal{L}(M)} \sum_{\rho \in R_c(L)} g_M^L(\rho) J_L(\rho, f). \end{aligned}$$

Then $f \to K_M(f)$ is a continuous linear mapping from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M, G)$. For any n, $K_M(f)$ projects to the element

$$K_M^n(f) = K_M'^{,n}(f) - g_M^{G,n}(f_{G,c})$$

in $\mathcal{G}_c^{\alpha,n}(M, G)$, which in turn comes with a representative

$$\begin{aligned} K_M^n(\gamma, f) &= K_M'^{,n}(\gamma, f) - g_M^{G,n}(\gamma, f_{G,c}) \\ &= K_M'^{,n}(\gamma, f) - \sum_{\rho \in R_c(G)} g_M^{G,n}(\gamma, \rho) f_G(\rho) \end{aligned}$$

in $\mathcal{F}_c^\alpha(V, G)$. We can also write

$$\begin{aligned} K_M^n(\gamma, f) &= J_M(\gamma, f) - \sum_{L \in \mathcal{L}(M)} \sum_{\rho \in R_c(L)} g_M^{L,n}(\gamma, \rho) J_L(\rho, f) \\ &= J_M(\gamma, f) - J_M^n(\gamma, f), \end{aligned}$$

in the notation (46). By construction, $f \to K_M^n(\gamma, f)$ is a continuous linear mapping from $\mathcal{C}(G)$ to $\mathcal{F}_c^\alpha(V, G)$. Theorem 6.1* asserts that the mapping actually sends $\mathcal{C}(G)$ continuously to the space $\mathcal{F}_{c,n}^\alpha(V, G)$.

Given n, we once again choose N to be large. It is a consequence of (36) that the sum in the first formula for $K_M^n(\gamma, f)$ may be taken over the finite subset $R_{c,N}(G)$ of $R_c(G)$. It follows that

$$\begin{aligned} K_M^n(\gamma, f) &= K_M'^{,n}(\gamma, f) - \sum_{\rho \in R_{c,N}(G)} g_M^{G,n}(\gamma, \rho) f_G(\rho) \\ &= K_M'^{,n}(\gamma, f) - \sum_{\rho \in R_{c,N}(G)} K_M'^{,n}(\gamma, f_\rho^N) f_G(\rho). \end{aligned}$$

The mapping

$$f \longrightarrow K_M^{\prime,n}(\gamma,f) = \widetilde{K}_M^n(\gamma,f) - \tilde{\chi}_M^n(\gamma,f)$$

from $\mathcal{C}(G)$ to $\mathcal{F}_c^\alpha(V, G)$ is of course linear. Consequently,

$$K_M^n(\gamma,f) = K_M^{\prime,n}(\gamma,f^{c,N}),$$

where

$$f^{c,N} = f - \sum_{\rho\in R_{c,N}(G)} f_G(\rho) f_\rho^N.$$

Observe that the mapping

$$f \longrightarrow f^{c,N}, \qquad f \in \mathcal{C}(G), \tag{89}$$

is a continuous linear operator on $\mathcal{C}(G)$ that takes values in the subspace $\mathcal{C}_{c,N}(G)$ of $\mathcal{C}(G)$. In particular, the function

$$K_M^n(\gamma,f) = \widetilde{K}_M^n(\gamma,f^{c,N}) - \tilde{\chi}_M^n(\gamma,f^{c,N})$$

equals

$$\psi_M^n(\gamma,f^{c,N}) - \chi_M^n(\gamma,f^{c,N}),$$

by the restriction condition (79). Composing (89) with the mapping (69) of Proposition 9.1, we conclude that

$$f \longrightarrow K_M^n(\gamma,f), \qquad f \in \mathcal{C}(G),$$

is a continuous linear mapping from $\mathcal{C}(G)$ to $\mathcal{F}_{c,n}^\alpha(V, G)$. This is the main assertion of Theorem 6.1*.

We have finished our inductive construction of the objects (34) and (35). We have also established the continuity assertion of Theorem 6.1*. As we noted in Sect. 6, this implies the assertion of Theorem 6.1 that the weighted orbital integral $J_M(f)$ represents the same element in $\widehat{\mathcal{G}}_c(M, G)$ as the formal germ (43). We have therefore an identity

$$g_M^G\big(J_{G,c}(f)\big) = J_M(f) - \sum_{L\in\mathcal{L}^0(M)} g_M^L\big(J_{L,c}(f)\big) \tag{90}$$

of formal germs, which holds for any function $f \in \mathcal{C}(G)$. According to our induction assumption, the summands in (90) with $L \neq M$ are independent of the choice of bases $R_c(L)$. The same is true of the summand with $L = M$, as we observed earlier in this section. Since the other term on the right-hand side of (90) is just the weighted orbital integral $J_M(f)$, the left-hand side of (90) is also independent of any choice of bases. We have thus established the functorial condition (37) in the remaining case that $L = G$.

We can also use (90) to prove the differential equation (40). Suppose that $z \in \mathcal{Z}(G)$. In Sect. 7, we established the identity (55) as a consequence of the two sets of Eqs. (39) and (40). Since we now have these equations for any $L \neq G$, we can assume that (55) also holds for any $L \neq G$. It follows that

$$\begin{aligned} &\sum_{L\in\mathcal{L}^0(M)} g_M^L\big(J_{L,c}(zf)\big) \\ &\quad = \sum_{L\in\mathcal{L}^0(M)} \sum_{S\in\mathcal{L}^L(M)} \partial_M^S(z_S) g_S^L\big(J_{L,c}(f)\big)_M \\ &\quad = \sum_{S\in\mathcal{L}(M)} \sum_{L\in\mathcal{L}^0(S)} \partial_M^S(z_S) g_S^L\big(J_{L,c}(f)\big)_M. \end{aligned}$$

We combine this with (90) (f being replaced by zf), and the differential equation (22) for $J_M(zf)$. We obtain

$$\begin{aligned} &g_M^G\big(J_{G,c}(zf)\big) \\ &\quad = \sum_{S\in\mathcal{L}(M)} \partial_M^S(z_S)\Big(J_S(f) - \sum_{L\in\mathcal{L}^0(S)} g_S^L\big(J_{L,c}(f)\big)\Big)_M \\ &\quad = \sum_{S\in\mathcal{L}(M)} \partial_M^S(z_S) g_S^G\big(J_{G,c}(f)\big)_M \\ &\quad = \sum_{\rho\in R_c(G)} \Big(\sum_{S\in\mathcal{L}(M)} \partial_M^S(z_S) g_S^G(\rho)_M\Big) f_G(\rho). \end{aligned}$$

But

$$g_M^G\big(J_{G,c}(zf)\big) = \sum_{\rho\in R_c(G)} g_M^G(\rho)(zf)_G(\rho) = \sum_{\rho\in R_c(G)} g_M^G(\hat{z}\rho) f_G(\rho).$$

Comparing the coefficients of $f_G(\rho)$ in the two expressions, we see that

$$g_M^G(\hat{z}\rho) = \sum_{S\in\mathcal{L}(M)} \partial_M^S(z_S) g_S^G(\rho)_M, \qquad \rho \in R_c(G).$$

This is Eq. (40) in the remaining case that $L = G$.

It remains only to check the symmetry conditions (41), (42) and (42*). Given an isomorphism $\theta\colon G \to \theta G$ over $\mathbb{R}$, we need to prescribe the mapping

$$\tilde{\chi}_{\theta M} : \mathcal{C}(\theta G) \longrightarrow \widehat{\mathcal{I}}_{\theta G}(\theta M, \theta G)$$

of Proposition 9.3 for the 4-tuple $(G_1, K_1, M_1, c_1) = (\theta G, \theta K, \theta M, \theta c)$ in terms of the chosen mapping $\tilde{\chi}_M$ for (G, K, M, c). We do so in the obvious way, by setting

$$\tilde{\chi}_{\theta M}(\theta f) = \theta\tilde{\chi}_M(f), \qquad f \in \mathcal{C}(G).$$

This mapping depends of course on (G_1, K_1, M_1, c_1), but by the symmetry condition (81) for G, it is independent of the choice of θ. The conditions (79)–(81) for θG follow from (77) and the corresponding conditions for G. Having defined the mapping $\tilde{\chi}_{\theta M}$, we then need only appeal to the earlier discussion of this section. If ρ belongs to $R_c(M)$, we obtain

$$\begin{aligned} J_{\theta M}(\theta\rho, \theta f) &= \langle \theta\rho, \tilde{\chi}_{\theta M}(\theta f)\rangle \\ &= \langle \theta\rho, \theta\tilde{\chi}_M(f)\rangle = J_M(\rho, f), \end{aligned}$$

from (84). This is the condition (41) in the remaining case that $L = M$. For $n \geq 0$, we also obtain

$$\begin{aligned} g_{\theta M}^{\theta G, n}(\theta\gamma, \theta\rho) &= \widetilde{K}_{\theta M}^n(\theta\gamma, f_{\theta\rho}^N) - \tilde{\chi}_M^n(\theta\gamma, f_{\theta\rho}^N) \\ &= \widetilde{K}_{\theta M}^n(\theta\gamma, \theta f_\rho^N) - \tilde{\chi}_M^n(\theta\gamma, \theta f_\rho^N) \\ &= \widetilde{K}_M^n(\gamma, f_\rho^N) - \tilde{\chi}_M^n(\gamma, f_\rho^N) \\ &= g_M^{G,n}(\gamma, \rho), \end{aligned}$$

from (88), (63) and the definition of $\tilde{\chi}_{\theta M}(\theta f)$ above. This is condition (42*) in the remaining case that $L = G$. Finally, we observe that

$$\begin{aligned} g_{\theta M}^{\theta G}(\theta\rho) &= \varprojlim_n g_{\theta M}^{\theta G, n}(\theta\rho) \\ &= \varprojlim_n \theta g_M^{G,n}(\rho) = \theta g_M^G(\rho). \end{aligned}$$

This is the third symmetry condition (42), in the remaining case $L = G$.

We have now established the last of the conditions of Theorems 6.1 and 6.1*. This brings us to the end of the induction argument begun in Sect. 7, and completes the proof of the two theorems.

We observed in Sect. 6 that the objects we have now constructed are not unique. The definitions of this section do depend canonically on the mapping $\tilde{\chi}_M$ of Proposition 9.3, which is in turn determined by the mapping $\tilde{\chi}_{M,1}$ in (82). But $\tilde{\chi}_{M,1}$ is uniquely determined only up to translation by the $\mathrm{Aut}(G, K, M, c)$-fixed linear transformation C in (83). The coefficients $c(\rho_M, \rho_G)$ in (49), which were used in Proposition 6.2 to describe the lack of uniqueness, are of course related to C. Suppose that the basis $R_c(M)$ is chosen so that the subset

$$R_{c,G\text{-harm}}(M) = R_c(M) \cap \mathcal{D}_{c,G\text{-harm}}(M)$$

is a basis of $\mathcal{D}_{c,G\text{-harm}}(M)$. It then follows that

$$c(\rho_M, \rho_G) = \langle {}^t C\rho_M, \rho_G\rangle, \qquad \rho_M \in R_{c,G\text{-harm}}(M),\ \rho_G \in R_{c,N_1}(G).$$

For general ρ_M and ρ_G, the coefficient $c(\rho_M, \rho_G)$ is determined from this special case by the relation (51).

11 Invariant Distributions $I_M(\rho, f)$

Weighted orbital integrals have the obvious drawback of not being invariant. Their dependence on the maximal compact subgroup K is also not entirely agreeable. However, there is a natural way to construct a parallel family of distributions with better properties. We shall show that these distributions satisfy the same formal germ expansions as the weighted orbital integrals.

As we recalled in Sect. 2, elements in $\mathcal{I}(G)$ can be regarded as functions

$$f_G :\ \pi \longrightarrow f_G(\pi) = \mathrm{tr}\big(\pi(f)\big), \qquad f \in \mathcal{C}(G),\ \pi \in \Pi_{\mathrm{temp}}(G),$$

on $\Pi_{\mathrm{temp}}(G)$ (the set of irreducible tempered representations of $G(\mathbb{R})$), rather than $\Gamma_{\mathrm{reg}}(G)$ (the set of strongly regular conjugacy classes in $G(\mathbb{R})$). The two interpretations are related by the formula

$$f_G(\pi) = \int_{\Gamma_{\mathrm{reg}}(G)} f_G(\gamma)|D(\gamma)|^{\frac{1}{2}}\Theta_\pi(\gamma)d\gamma, \qquad f \in \mathcal{C}(G),\ \gamma \in \Gamma_{\mathrm{reg}}(G),$$

where Θ_π is the character of π, and $d\gamma$ is a measure on $\Gamma_{\mathrm{reg}}(G)$ provided by the Weyl integration formula. We have also noted that any invariant, tempered distribution I on $G(\mathbb{R})$ factors through the space $\mathcal{I}(G)$. In other words, there is a continuous linear form $\widehat{I}$ on $\mathcal{I}(G)$ such that

$$I(f) = \widehat{I}(f_G), \qquad f \in \mathcal{C}(G).$$

This can be proved either by analysing elements in $\mathcal{I}(G)$ directly as functions on $\Gamma_{\mathrm{reg}}(G)$ [B2] or by using the characterization [A5] of elements in $\mathcal{I}(G)$ as functions on $\Pi_{\mathrm{temp}}(G)$.

We fix a Levi subgroup $M \subset G$ and a maximal compact subgroup $K \subset G(\mathbb{R})$, as in Sect. 4. For each Levi subgroup $L \in \mathcal{L}(M)$, one can define a continuous linear transformation

$$\phi_L = \phi_L^G :\ \mathcal{C}(G) \longrightarrow \mathcal{I}(L)$$

in terms of objects that are dual to weighted orbital integrals. If f belongs to $\mathcal{C}(G)$, the value of $\phi_L(f)$ at $\pi \in \Pi_{\mathrm{temp}}(L)$ is the weighted character

$$\phi_L(f, \pi) = \mathrm{tr}\big(\mathcal{M}_L(\pi, P)\mathcal{I}_P(\pi, f)\big), \qquad P \in \mathcal{P}(L),$$

defined on p. 38 of [A7]. In particular, $\mathcal{I}_P(\pi)$ is the usual induced representation of $G(\mathbb{R})$, while

$$\mathcal{M}_L(\pi, P) = \lim_{\lambda \to 0} \left(\sum_{Q \in \mathcal{P}(L)} \mathcal{M}_Q(\lambda, \pi, P) \theta_Q(\lambda)^{-1} \right)$$

is the operator built out of Plancherel densities and unnormalized intertwining operators between induced representations, as on p. 37 of [A7]. Weighted characters behave in many ways like weighted orbital integrals. In particular, $\phi_L(f)$ depends on K, and transforms under conjugation of f by $y \in G(\mathbb{R})$ by a formula

$$\phi_L(f^y) = \sum_{Q \in \mathcal{F}(L)} \phi_L^{M_Q}(f_{Q,y}) \tag{91}$$

that is similar to (21).

The role of the mappings ϕ_L is to make weighted orbital integrals invariant. One defines invariant linear forms

$$I_M(\gamma, f) = I_M^G(\gamma, f), \qquad \gamma \in M_{G\text{-reg}}(\mathbb{R}),\ f \in \mathcal{C}(G),$$

on $\mathcal{C}(G)$ inductively by setting

$$J_M(\gamma, f) = \sum_{L \in \mathcal{L}(M)} \widehat{I}_M^L\big(\gamma, \phi_L(f)\big).$$

In other words,

$$I_M(\gamma, f) = J_M(\gamma, f) - \sum_{L \in \mathcal{L}^0(M)} \widehat{I}_M^L\big(\gamma, \phi_L(f)\big).$$

This yields a family of tempered distributions, which are parallel to weighted orbital integrals, but which are invariant and independent of K. (See, for example, [A7, §3].) We would like to show that they satisfy the formal germ expansions of Theorems 6.1 and 6.1*.

We fix the conjugacy class $c \in \Gamma_{ss}(M)$, as before. We must first attach invariant linear forms to the noninvariant distributions $J_M(\rho, f)$ in (34). Following the prescription above, we define invariant distributions

$$I_M(\rho, f) = I_M^G(\rho, f), \qquad \rho \in R_c(M),\ f \in \mathcal{C}(G),$$

inductively by setting

$$J_M(\rho, f) = \sum_{L \in \mathcal{L}(M)} \widehat{I}_M^L\big(\rho, \phi_L(f)\big).$$

In other words,

$$I_M(\rho,f) = J_M(\rho,f) - \sum_{L\in\mathcal{L}^0(M)} \widehat{I}_M^L\big(\rho,\phi_L(f)\big).$$

The invariance of $I_M(\gamma,f)$ follows inductively in the usual way from (21) and (91). As a general rule, the application of harmonic analysis improves one property only at the expense of another. In the case at hand, the price to pay for making $J_M(\rho,f)$ invariant is that the new distribution $I_M(\rho,f)$ is no longer supported on $\mathcal{U}_c(G)$.

We have in any case replaced the family (34) with a family

$$f \longrightarrow I_L(\rho,f), \qquad\qquad L \in \mathcal{L}(M),\ \ \rho \in R_c(L), \tag{92}$$

of invariant tempered distributions. These new objects do have many properties in common with the original ones. They satisfy the differential equation

$$I_L(\rho, zf) = I_L(z_L\rho,f), \tag{93}$$

for each $z \in \mathcal{Z}(G)$. They also satisfy the symmetry condition

$$I_{\theta L}(\theta\rho, \theta f) = I_L(\rho,f), \tag{94}$$

for any isomorphism θ: $G \to \theta G$ over $\mathbb{R}$. In addition, the distributions satisfy the transformation formula

$$I_L(\rho',f) = \sum_{\rho} a_L(\rho', \rho) I_L(\rho,f), \tag{95}$$

for $\{\rho'\}$ and $A_L = \{a_L(\rho', \rho)\}$ as in (5.4.1). We leave the reader to check that these properties are direct consequences of the corresponding properties in Sect. 6.

It follows from (36) that the series

$$g_M^L\big(I_{L,c}(f)\big) = \sum_{\rho\in R_c(L)} g_M^L(\rho) I_L(\rho,f)$$

converges in (the adic topology of) $\widehat{\mathcal{G}}_c(M,L)$. The continuity of the linear forms (92) implies, moreover, that the mapping

$$f \longrightarrow g_M^L\big(I_{L,c}(f)\big)$$

from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c(M,L)$ is continuous (in the complex topology of $\widehat{\mathcal{G}}_c(M,L)$). Finally, (45) and (95) yield the functorial property that for any L and f,

$$g_M^L\big(J_{L,c}(f)\big) \text{ is independent of the choice of basis } R_c(L). \tag{96}$$

The distributions (92) play the role of coefficients in a formal germ expansion of the function $I_M(\gamma, f)$. Following Sect. 6, we set

$$I_M^n(\gamma, f) = \sum_L \sum_\rho g_M^{L,n}(\gamma, \rho) I_L(\rho, f), \tag{97}$$

for any $n \geq 0$, and for fixed representatives $g_M^{L,n}(\gamma, \rho)$ of $g_M^{L,n}(\rho)$ in $\mathcal{F}_c^\alpha(M, L)$ as in (46). We then obtain the following corollary of Theorems 6.1 and 6.1*.

Corollary 11.1. *We can choose the weight function* α *so that* $\alpha(1)$ *equals zero, and so that for any n, the mapping*

$$f \longrightarrow I_M(\gamma, f) - I_M^n(\gamma, f), \qquad f \in \mathcal{C}(G),$$

is a continuous linear mapping from $\mathcal{C}(G)$ *to the space* $\mathcal{F}_{c,n}^\alpha(V, G)$*. In particular,* $I_M(f)$ *has a formal germ expansion given by the sum*

$$\sum_{L \in \mathcal{L}(M)} g_M^L\big(I_{L,c}(f)\big) = \sum_{L \in \mathcal{L}(M)} \sum_{\rho \in R_c(L)} g_M^L(\rho) I_L(\rho, f). \tag{98}$$

Proof. The second assertion follows immediately from the first, in the same way that the corresponding assertion of Theorem 6.1 follows from Theorem 6.1*. To establish the first assertion, we write

$$I_M(\gamma, f) - I_M^n(\gamma, f)$$

as the difference between

$$J_M(\gamma, f) - J_M^n(\gamma, f)$$

and

$$\sum_{L \in \mathcal{L}^0(M)} \big(\widehat{I}_M^L\big(\gamma, \phi_L(f)\big) - \widehat{I}_M^{L,n}\big(\gamma, \phi_L(f)\big)\big).$$

The assertion then follows inductively from Theorem 6.1*. □

Corollary 11.1 tells us that the sum (98) represents the same element in $\widehat{\mathcal{G}}_c(M, G)$ as $I_M(f)$. In other words, the invariant distributions attached to weighted orbital integrals satisfy asymptotic expansions

$$I_M(\gamma, f) \sim \sum_L \sum_\rho g_M^L(\gamma, \rho) I_L(\rho, f).$$

The invariant distributions $I_M(\rho, f)$ ultimately depend on our choice of the mapping $\tilde{\chi}_M$. It is interesting to note that this mapping has an invariant formulation, which leads to posteriori to a more direct construction of the distributions. To see this, we first set

$$I\widetilde{K}_M(f) = I_M(f) - \sum_{\{L \in \mathcal{L}(M):\, L \neq M,G\}} g_M^L\big(I_{L,c}(f)\big), \quad f \in \mathcal{C}(G). \tag{99}$$

Let α be a fixed weight function that satisfies the conditions of Lemma 7.3. Then $f \to I\widetilde{K}_M(f)$ is a continuous linear transformation from $\mathcal{C}(G)$ to $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$.

Lemma 11.2. *Suppose that f belongs to $\mathcal{C}(G)$. Then*

$$\widetilde{K}_M(f) - I\widetilde{K}_M(f) = g_M^M\big(J_{M,c}(f)\big) - g_M^M\big(I_{M,c}(f)\big). \tag{100}$$

Proof. The proof is similar to that of Lemmas 7.1 and 7.2, so we shall be brief. The left-hand side of (100) equals

$$\sum_{L_1} \widehat{I}_M^{L_1}\big(\phi_{L_1}(f)\big) - \sum_{L,L_1} g_M^L\big(\widehat{I}_{L,c}^{L_1}(\phi_{L_1}(f))\big),$$

where the first sum is over Levi subgroups $L_1 \in \mathcal{L}^0(M)$, and the second sum is over pairs $L, L_1 \in \mathcal{L}(M)$ with

$$M \subsetneq L \subset L_1 \subsetneq G.$$

Taking the second sum over L_1 outside the sum over L, we obtain an expression

$$\sum_{L_1 \in \mathcal{L}^0(M)} \left(\Big(\widehat{I}_M^{L_1}\big(\phi_{L_1}(f)\big) - \sum_{L \in \mathcal{L}^{L_1}(M)} g_M^L\big(\widehat{I}_{L,c}^{L_1}(\phi_{L_1}(f))\big)\Big) + g_M^M\big(\widehat{I}_{M,c}^{L_1}(\phi_{L_1}(f))\big) \right).$$

By Corollary 11.1, the formal germ

$$\widehat{I}K_M^{L_1}\big(\phi_{L_1}(f)\big) = \widehat{I}_M^{L_1}\big(\phi_{L_1}(f)\big) - \sum_{L \in \mathcal{L}^{L_1}(M)} g_M^L\big(\widehat{I}_{L,c}^{L_1}(\phi_{L_1}(f))\big)$$

vanishes for any L_1. The left-hand side of (100) therefore equals

$$\sum_{L_1 \in \mathcal{L}^0(M)} g_M^M\big(\widehat{I}_{M,c}^{L_1}(\phi_{L_1}(f))\big).$$

By definition, this in turn equals the right-hand side of the required formula (100).

□

The lemma implies that the mapping

$$f \longrightarrow I\widetilde{K}_M(f) - \widetilde{K}_M(f), \qquad f \in \mathcal{C}(G),$$

takes values in the subspace $\widehat{\mathcal{I}}_c(M,G)$ of $\widehat{\mathcal{G}}_c^{\alpha}(M,G)$. We shall use this property to give invariant versions of the constructions of Sect. 9. For any $n \geq 0$, and N large relative to n, we can write $I\psi_M^n$ for the restriction of $I\widetilde{K}_M^n$ to the subspace $\mathcal{C}_{c,N}(G)$ of $\mathcal{C}(G)$. If f is a function in $\mathcal{C}_{c,N}(G)$, the (α,n)-jet

$$I\psi_M^n(f) = \psi_M^n(f) - \big(\widetilde{K}_M^n(f) - I\widetilde{K}_M^n(f)\big)$$

then belongs to the image of $\mathcal{I}_c^n(M,G)$ in $\mathcal{G}_c^{\alpha,n}(M,G)$. This yields the invariant analogue of Proposition 9.1. In particular, there is a uniquely determined, continuous linear mapping

$$I\chi_M^n : \mathcal{C}_{c,N}(G) \longrightarrow \mathcal{I}_c^n(M,G)$$

such that for any $f \in \mathcal{C}_{c,N}(G)$, the image of $I\chi_M^n(f)$ in $\mathcal{G}_c^{\alpha,n}(M,G)$ equals $I\psi_M^n(f)$. Following (75), we set

$$\langle \sigma, I\chi_M(f)\rangle = \langle \sigma, I\chi_M^n(f)\rangle, \qquad \sigma \in \mathcal{D}_c(M),\ f \in \mathcal{C}_{c,N}(G),$$

for any $n \geq \deg(\sigma)$ and N large relative to n. Then

$$\langle \sigma, I\chi_M(f)\rangle = \langle \sigma, \chi_M(f)\rangle - \langle \sigma, \widetilde{K}_M(f) - I\widetilde{K}_M(f)\rangle.$$

Given the mapping $\tilde{\chi}_M$ of Proposition 9.3, we set

$$I\tilde{\chi}_M(f) = \tilde{\chi}_M(f) - \big(\widetilde{K}_M(f) - I\widetilde{K}_M(f)\big), \qquad f \in \mathcal{C}(G). \tag{101}$$

Then $I\tilde{\chi}_M$ is a continuous linear mapping from $\mathcal{C}(G)$ to $\widehat{\mathcal{I}}_c(M,G)$ that satisfies the invariant analogue of the restriction property (79). Moreover, it follows easily from the lemma that $I\tilde{\chi}_M$ also satisfies the analogues of (80) and (81). Conversely, suppose that $I\tilde{\chi}_M$ is any continuous mapping from $\mathcal{C}(G)$ to $\widehat{\mathcal{I}}_c(M,G)$ that satisfies the invariant analogues of (79)–(81). Then the mapping $\tilde{\chi}_M(f)$ defined by (101) satisfies the hypotheses of Proposition 9.3. Thus, instead of choosing the extension $\tilde{\chi}_M$ of mappings $\{\chi_M^n\}$, as in Proposition 9.3, we could equally well choose an extension $I\tilde{\chi}_M$ of invariant mappings $\{I\chi_M^n\}$. To see the relationship of the latter with our invariant distributions, we take any element $\rho \in R_c(M)$, and write

$$\begin{aligned}
&I_M(\rho,f) - J_M(\rho,f)\\
&\quad = \big\langle \rho, g_M^M\big(I_{M,c}(f)\big)\big\rangle - \big\langle \rho, g_M^M\big(J_{M,c}(f)\big)\big\rangle\\
&\quad = \big\langle \rho, I\widetilde{K}_M(f) - \widetilde{K}_M(f)\big\rangle\\
&\quad = \big\langle \rho, I\tilde{\chi}_M(f)\big\rangle - \big\langle \rho, \tilde{\chi}_M(f)\big\rangle,
\end{aligned}$$

by Lemma 11.2 and the definition (101). It follows from the definition (84) that

$$I_M(\rho,f) = \langle \rho, I\tilde{\chi}_M(f)\rangle, \qquad f \in \mathcal{C}(G). \tag{102}$$

The invariant distributions can therefore be defined directly in terms of the mapping $I\tilde{\chi}_M$.

12 Supplementary Properties

There are further constraints that one could impose on the mapping $\tilde{\chi}_M$ of Proposition 9.3 (or equivalently, the invariant mapping (101)). Any new constraint makes the construction more rigid. It puts extra conditions on the families of coefficients (49) and linear transformations (83), either of which describes the lack of uniqueness of the construction. A suitable choice of $\tilde{\chi}_M$ will also endow our distributions and formal germs with new properties.

The most important property is that of parabolic descent. Suppose that M_1 is a Levi subgroup of M, chosen so that $\mathfrak{a}_{M_1}$ is orthogonal to the Lie algebra of K. Any element γ_1 in $M_{1,G\text{-reg}}(\mathbb{R})$ obviously maps to an element $\gamma = \gamma_1^M$ in $M_{G\text{-reg}}(\mathbb{R})$. The associated weighted orbital integral satisfies the descent formula

$$J_M(\gamma,f) = \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1) J_{M_1}^{G_1}(\gamma_1, f_{Q_1}), \tag{103}$$

in the notation of [A4, Corollary 8.2]. The coefficient $d_{M_1}^G(M, G_1)$ is defined on p. 356 of [A4], while the section

$$G_1 \longrightarrow Q_1 = Q_{G_1}, \qquad G_1 \in \mathcal{L}(M_1),\ Q_{G_1} \in \mathcal{P}(G_1),$$

is defined on p. 357 of [A4]. We would like to establish similar formulas for our singular distributions and our formal germs.

Suppose that c is the image in $\Gamma_{ss}(M)$ of a class $c_1 \in \Gamma_{ss}(M_1)$. If L belongs to $\mathcal{L}(M)$, and L_1 lies in the associated set $\mathcal{L}^L(M_1)$, we shall denote the image of c_1 in $\Gamma_{ss}(L_1)$ by c_1 as well. For any such L and L_1, there is a canonical induction mapping $\sigma_1 \to \sigma_1^L$ from $\mathcal{D}_{c_1}(L_1)$ to $\mathcal{D}_c(L)$ such that

$$h_L(\sigma_1^L) = h_{L_1}(\sigma_1), \qquad \sigma_1 \in \mathcal{D}_{c_1}(L_1),\ h \in \mathcal{C}(L).$$

Since we can view $J_L(\cdot,f)$ as a linear form on $\mathcal{D}_c(L)$, the tempered distribution

$$J_L(\sigma_1^L, f), \qquad f \in \mathcal{C}(G),$$

is defined for any σ_1. We also write $\sigma \to \sigma_{L_1}$ for the adjoint restriction mapping from $\mathcal{D}_c(L)$ to $\mathcal{D}_{c_1}(L_1)$, relative to the bases $R_c(L)$ and $R_{c_1}(L_1)$. In other words,

$$\sum_{\rho \in R_c(L)} \phi_1(\rho_{L_1})\phi(\rho) = \sum_{\rho_1 \in R_{c_1}(L_1)} \phi_1(\rho_1)\phi(\rho_1^L),$$

for any linear functions ϕ_1 and ϕ on $\mathcal{D}_{c_1}(L_1)$ and $\mathcal{D}_c(L)$, respectively, for which the sums converge. (The restriction mapping comes from a canonical linear transformation $\widehat{\mathcal{I}}_c(L) \to \widehat{\mathcal{I}}_{c_1}(L_1)$ between the dual spaces of $\mathcal{D}_c(L)$ and $\mathcal{D}_{c_1}(L_1)$. Its basis dependent formulation as a mapping from $\mathcal{D}_c(L)$ to $\mathcal{D}_{c_1}(L_1)$ is necessitated by our notation for the formal germs $g_M^L(\rho)$.) We recall that as an element in $\widehat{\mathcal{G}}_c(M, L)$, $g_M^L(\rho)$ can be mapped to the formal germ $g_M^L(\rho)_{M_1}$ in $\widehat{\mathcal{G}}_{c_1}(M_1 \mid M, L)$.

Proposition 12.1. *We can choose the mapping $\tilde{\chi}_M$ of Proposition 9.3 so that for any M_1 and c_1, the distributions (34) satisfy the descent formula*

$$J_L(\rho_1^L, f) = \sum_{G_1 \in \mathcal{L}(L_1)} d_{L_1}^G(L, G_1) J_{L_1}^{G_1}(\rho_1, f_{Q_1}), \quad L_1 \in \mathcal{L}^L(M_1), \quad \rho_1 \in R_{c_1}(L_1), \tag{104}$$

while the formal germs (35) satisfy the descent formula

$$g_M^L(\rho)_{M_1} = \sum_{L_1 \in \mathcal{L}^L(M_1)} d_{M_1}^L(M, L_1) g_{M_1}^{L_1}(\rho_{L_1}), \qquad \rho \in R_c(L). \tag{105}$$

Proof. We have to establish the two formulas for any $L \in \mathcal{L}(M)$. We can assume inductively that for each M_1 and c_1, (104) holds for $L \neq M$, and (105) holds for $L \neq G$. In particular, both formulas hold for any L in the complement $\widetilde{\mathcal{L}}(M)$ of $\{G, M\}$ in $\mathcal{L}(M)$. We shall use this property to establish a descent formula for the formal germ

$$\widetilde{K}_M(f)_{M_1} = J_M(f)_{M_1} - \sum_{L \in \widetilde{\mathcal{L}}(M)} g_M^L\big(J_{L,c}(f)\big)_{M_1}.$$

The original identity (103) leads immediately to a descent formula

$$J_M(f)_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1) J_{M_1}^{G_1}(f_{Q_1})$$

for the first term in the last expression for $\widetilde{K}_M(f)_{M_1}$. We apply (104) and (105) inductively to the summands in the second term

$$\sum_{L \in \widetilde{\mathcal{L}}(M)} g_M^L\big(J_{L,c}(f)\big)_{M_1}. \tag{106}$$

We obtain

$$\begin{aligned}
& g_M^L\big(J_{L,c}(f)\big)_{M_1} \\
&= \sum_{\rho\in R_c(L)} g_M^L(\rho)_{M_1} J_L(\rho,f) \\
&= \sum_{L_1\in\mathcal{L}^L(M_1)} d_{M_1}^L(M,L_1) \sum_{\rho\in R_c(L)} g_{M_1}^{L_1}(\rho_{L_1}) J_L(\rho,f) \\
&= \sum_{L_1\in\mathcal{L}^L(M_1)} d_{M_1}^L(M,L_1) \sum_{\rho_1\in R_{c_1}(L_1)} g_{M_1}^{L_1}(\rho_1) J_L(\rho_1^L,f) \\
&= \sum_{L_1\in\mathcal{L}^L(M_1)} \sum_{G_1\in\mathcal{L}(L_1)} d_{M_1}^L(M,L_1) d_{L_1}^G(L,G_1)\Big(\sum_{\rho_1\in R_{c_1}(L_1)} g_{M_1}^{L_1}(\rho_1) J_{L_1}^{G_1}(\rho_1,f_{Q_1})\Big).
\end{aligned}$$

Therefore (106) equals the sum over $L\in\widetilde{\mathcal{L}}(M)$ of the expression

$$\sum_{L_1\in\mathcal{L}^L(M_1)} \sum_{G_1\in\mathcal{L}(L_1)} \big(d_{M_1}^L(M,L_1) d_{L_1}^G(L,G_1)\big) g_{M_1}^{L_1}\big(J_{L_1,c_1}^{G_1}(f_{Q_1})\big). \tag{107}$$

We can of course sum L over the larger set $\mathcal{L}(M)$, provided that we subtract the values of (107) taken when $L = M$ and $L = G$. If $L = M$, $d_{M_1}^L(M,L_1)$ vanishes unless $L_1 = M_1$, in which case $d_{M_1}^L(M,L_1) = 1$. The value of (107) in this case is

$$\sum_{G_1\in\mathcal{L}(M_1)} d_{M_1}^G(M,G_1) g_{M_1}^{M_1}\big(J_{M_1,c_1}^{G_1}(f_{Q_1})\big). \tag{108}$$

If $L = G$, $d_{L_1}^G(L,G_1)$ vanishes unless $G_1 = L_1$, in which case $d_{L_1}^G(L,G_1) = 1$. The value of (107) in this case is

$$\sum_{G_1\in\mathcal{L}(M_1)} d_{M_1}^G(M,G_1) g_{M_1}^{G_1}\big(J_{G_1,c_1}^{G_1}(f_{Q_1})\big). \tag{109}$$

Thus (106) equals the sum over $L\in\mathcal{L}(M)$ of (107) minus the sum of (108) and (109). The only part of (107) that depends on L is the product of coefficients in the brackets. We shall therefore take the sum over L inside the two sums over L_1 and G_1, which at the same time, we interchange. Then G_1, L_1 and L will be summed over $\mathcal{L}(M_1)$, $\mathcal{L}^{G_1}(M_1)$, and $\mathcal{L}(L_1)$, respectively. The resulting interior sum

$$\sum_{L\in\mathcal{L}(L_1)} d_{M_1}^L(M,L_1) d_{L_1}^G(L,G_1)$$

simplifies. According to [A4, (7.11)], this sum is just equal to $d_{M_1}^G(M,G_1)$, and in particular, is independent of L_1. We can therefore write (106) as the difference between the expression

$$\sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1) \sum_{L_1 \in \mathcal{L}^{G_1}(M_1)} g_{M_1}^{L_1}\big(J_{L_1,c_1}^{G_1}(f_{Q_1})\big)$$

and the sum of (108) and (109). But (108) is equal to contribution to the last expression of $L_1 = M_1$, while (109) equals the contribution of $L_1 = G_1$. We conclude that (106) equals

$$\sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1) \sum_{L_1 \in \widetilde{\mathcal{L}}^{G_1}(M_1)} g_{M_1}^{L_1}\big(J_{L_1,c_1}^{G_1}(f_{Q_1})\big).$$

We have established that $\widetilde{K}_M(f)_{M_1}$ equals

$$\sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1)\Big(J_{M_1}^{G_1}(f_{Q_1}) - \sum_{L_1 \in \widetilde{\mathcal{L}}^{G_1}(M_1)} g_{M_1}^{L_1}\big(J_{L_1,c_1}^{G_1}(f_{Q_1})\big)\Big).$$

Since the expression in the brackets equals $\widetilde{K}_{M_1}^{G_1}(f_{Q_1})$, we obtain a descent formula

$$\widetilde{K}_M(f)_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1)\widetilde{K}_{M_1}^{G_1}(f_{Q_1}). \tag{110}$$

Suppose that $n \geq 0$, and that N is large relative to n. The mapping χ_M^n of Proposition 9.1 then satisfies

$$\begin{aligned}
\chi_M^n(f)_{M_1} &= \psi_M^n(f)_{M_1} = \widetilde{K}_M^n(f)_{M_1} \\
&= \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1)\widetilde{K}_{M_1}^{G_1,n}(f_{Q_1}) \\
&= \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1)\psi_{M_1}^{G_1,n}(f_{Q_1}) \\
&= \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1)\chi_{M_1}^{G_1,n}(f_{Q_1}),
\end{aligned}$$

for any $f \in \mathcal{C}_{c,N}(G)$. This implies that

$$\langle \sigma, \chi_M(f)\rangle = \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1)\langle \sigma_1, \chi_{M_1}^{G_1}(f_{Q_1})\rangle, \qquad f \in \mathcal{C}_{c,N}(G),$$

for any induced element $\sigma = \sigma_1^M$ with $\sigma_1 \in \mathcal{D}_{c_1}(M_1)$, and for N large relative to σ_1.

We are now in a position to choose the mapping

$$\tilde{\chi}_M : \mathcal{C}(G) \longrightarrow \widehat{\mathcal{I}}_c(M, G)$$

of Proposition 9.3. More precisely, we shall specify that part of the mapping that is determined by its proper restrictions $\tilde{\chi}_M(f)_{M_1}$. We do so by making the inductive definition

$$\langle \sigma, \tilde{\chi}_M(f)\rangle = \sum_{G_1 \in \mathcal{L}(M_1)} d^G_{M_1}(M, G_1)\langle \sigma_1, \tilde{\chi}^{G_1}_{M_1}(f_{Q_1})\rangle, \ f \in \mathcal{C}(G), \tag{111}$$

for any properly induced element

$$\sigma = \sigma_1^M, \qquad \sigma_1 \in \mathcal{D}_{c_1}(M_1),\ M_1 \subsetneq M,$$

in $\mathcal{D}_c(M)$. The right-hand side of this expression is easily seen to depend only on σ, as opposed to the inducing data (M_1, σ_1). In fact, using the grading (11) of $\mathcal{D}_c(M)$, we can choose (M_1, σ_1) so that σ_1 belongs to $\mathcal{D}_{c_1,\mathrm{ell}}(M_1)$. The condition (79) of Proposition 9.3 follows from the formula above for $\langle \sigma, \chi_M(f)\rangle$. The conditions (80) and (81) follow inductively from the corresponding conditions for the terms $\langle \sigma_1, \tilde{\chi}^{G_1}_{M_1}(f_{Q_1}\rangle$. The formula (111) thus gives a valid definition of the linear form $\tilde{\chi}_M(f)$ on the subspace $\mathcal{D}_{c,\mathrm{par}}(M)$ spanned by elements in $\mathcal{D}_c(M)$ that are properly induced. For elements σ in the complementary subspace $\mathcal{D}_{c,\mathrm{ell}}(M)$, we remain free to define $\langle \sigma, \tilde{\chi}_M(f)\rangle$ in any way that satisfies the conditions (80)–(81) of Proposition 9.3.

Having chosen $\tilde{\chi}_M(f)$, we have only to apply the appropriate definitions. The first descent formula (104), in the remaining case that $L = M$, follows as directly from (111) and (84). Notice that (104) implies a similar formula

$$g^M_M\big(J^G_{M,c}(f)\big)_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d^G_{M_1}(M, G_1) g^{M_1}_{M_1}\big(J^{G_1}_{M_1,c_1}(f_{Q_1})\big)$$

for the formal germ $g^M_M\big(J^G_{M,c}(f)\big)$. Notice also that

$$K_M(f)_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d^G_{M_1}(M, G_1) K^{G_1}_{M_1}(f_{Q_1}),$$

since both sides vanish by Theorem 6.1. Combining these two observations with (110), we see that

$$g^G_M(f_{G,c})_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d^G_{M_1}(M, G_1) g^{G_1}_{M_1}(f_{Q_1,c_1}).$$

Now as a linear form in f, each side of this last formula is a linear combination of distributions $f_G(\rho)$ in the basis $R_c(M)$. We can therefore compare the coefficients of $f_G(\rho)$. The resulting identity is the second descent formula (105), in the remaining case that $L = G$. This completes the proof of the proposition. □

Corollary 12.2. *Suppose that the mapping $\tilde{\chi}_M$ is chosen as in the proposition. Then for any M_1 and c_1, the invariant distributions* (92) *satisfy the descent formula*

$$I_L(\rho_1^L, f) = \sum_{G_1 \in \mathcal{L}(L_1)} d_{L_1}^G(L, G_1) \widehat{I}_{L_1}^{G_1}(\rho_1, f_{G_1}), \ \ L_1 \in \mathcal{L}^L(M_1),$$

$$\rho_1 \in R_{c_1}(L_1). \tag{112}$$

Proof. We can assume inductively that (112) holds for any $L \in \mathcal{L}(M)$ with $L \neq M$, so it will be enough to treat the case that $L = M$. This frees the symbol L for use in the definition

$$I_M(\rho_1^L, f) = J_M(\rho_1^L, f) - \sum_{L \in \mathcal{L}^0(M)} \widehat{I}_M^L\big(\rho_1^L, \phi_L(f)\big)$$

from Sect. 11. We apply (104) to the first term $J_M(\rho_1^L, f)$. To treat the remaining summands $\widehat{I}_M^L\big(\rho_1^L, \phi_L(f)\big)$, we combine an inductive application of (112) to $I_M^L(\rho_1^L)$ with the descent formula

$$\phi_L(f)_{L_1} = \sum_{G_1 \in \mathcal{L}(L_1)} d_{L_1}^G(L, G_1) \phi_{L_1}^{G_1}(f_{Q_1}), \qquad L_1 \in \mathcal{L}^L(M_1),$$

established as, for example, in [A2, (7.8)]. We can then establish (112) (in the case $L = M$) by following the same argument that yielded the descent formula (110) in the proof of the proposition. (See also the proof of [A4, Theorem 8.1].) □

For the conditions of Proposition 12.1 and its corollary to hold, it is necessary and sufficient that the mapping $\tilde{\chi}_M = \tilde{\chi}_M^G$ satisfy its own descent formula. That is,

$$\tilde{\chi}_M(f)_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1) \tilde{\chi}_{M_1}^{G_1}(f_{Q_1})$$

for each M_1 and c_1. This in turn is equivalent to asking that the corresponding invariant mapping $I\tilde{\chi}_M = I\tilde{\chi}_M^G$ satisfy the descent formula

$$I\tilde{\chi}_M(f)_{M_1} = \sum_{G_1 \in \mathcal{L}(M_1)} d_{M_1}^G(M, G_1) \widehat{I}\tilde{\chi}_{M_1}^{G_1}(f_{G_1}),$$

again for each M_1 and c_1. Recall that $\tilde{\chi}_M(f)$ can be identified with a $W(M)$-fixed linear form on $\mathcal{D}_c(M)$. Its value at any element in $\mathcal{D}_c(M)$ is determined by the descent condition and the differential equation (80), once we have defined $\tilde{\chi}_M(f)$ as a linear form on the subspace

$$\mathcal{D}_{c,\text{ell},G\text{-harm}}(M) = \mathcal{D}_{c,\text{ell}}(M) \cap \mathcal{D}_{c,G\text{-harm}}(M).$$

The mapping $\tilde{\chi}_M$ is then uniquely determined up to an $\mathrm{Aut}(G, K, M, c)$-fixed linear transformation

$$C : \mathcal{C}(G)/\mathcal{C}_{c,N}(G) \longrightarrow \mathcal{D}_{c,\mathrm{ell},G\text{-harm}}(M)^*.$$

There is another kind of descent property we could impose on our distributions and formal germs. This is geometric descent with respect to c, the aim of which would be to reduce the general study to the case of $c = 1$. One would try to find formulas that relate the objects attached to (G, M, c) with corresponding objects for $(G_c, M_c, 1)$. This process has been carried out for p-adic groups. Geometric descent formulas for distributions were given in Theorem 8.5 of [A3] and its corollaries, while the descent formula for p-adic germs was in [A3, Proposition 10.2]. These formulas have important applications to the stable formula. In the archimedean case, however, geometric descent does not seem to play a role in the trace formula. Since it would entail a modification of our construction in the case of $c \neq 1$, we shall not pursue the matter here.

Finally, it is possible to build the singular weighted orbital integrals of [A3] into the constructions of this paper. Suppose that γ_c is a conjugacy class in $M(\mathbb{R})$ that is contained in $\mathcal{U}_c(M)$, and has been equipped with $M(\mathbb{R})$-invariant measure. The associated invariant integral gives a distribution

$$h \longrightarrow h_M(\gamma_c), \qquad h \in \mathcal{C}(M),$$

in $\mathcal{D}_c(M)$. We write $\mathcal{D}_{c,\mathrm{orb}}(M)$ for the subspace of $\mathcal{D}_c(M)$ spanned by such distributions. Any element in $\mathcal{D}_{c,\mathrm{orb}}(M)$ is known to be a finite linear combination of distributions $h \to h_M(\sigma)$, for triplets $\sigma = \big(T, \Omega, \partial(u)\big)$ in $S_c(M)$ such that u is M_c-harmonic. Since $W(M_c, T)$ is contained in $W(G, T)$, any M_c-harmonic element is automatically G-harmonic. The space $\mathcal{D}_{c,\mathrm{orb}}(M)$ is therefore contained in $\mathcal{D}_{c,G\text{-harm}}(M)$. The point is that one can define a *canonical* distribution $J_M(\sigma, f)$, for any σ in $\mathcal{D}_{c,\mathrm{orb}}(M)$ [A3, (6.5)]. This distribution is supported on $\mathcal{U}_c(G)$, and satisfies the analogue

$$J_M(\sigma, f^y) = \sum_{Q \in \mathcal{F}(M)} J_M^{M_Q}(\sigma, f_{Q,y}), \qquad f \in C_c^\infty(G),\ y \in G(\mathbb{R}),$$

of (38). It follows from (48) that $J_M(\sigma, f)$ can be chosen to represent an element in the family (34) (and in particular, is a tempered distribution). Otherwise said, the constructions of [A3] provide a canonical definition for a part of the operator $\tilde{\chi}_M$ of Proposition 9.3. They determine the restriction of each linear form $\tilde{\chi}_M(f)$ to the subspace $\mathcal{D}_{c,\mathrm{orb}}(M)$ of $\mathcal{D}_c(M)$. The conditions of Proposition 9.3 and [A3, (6.5)] therefore reduce the choice of $\tilde{\chi}_M$ to that of an $\mathrm{Aut}(G, K, M, c)$-fixed linear transformation that fits into a diagram

$$\mathcal{C}_{c,N_1}(G) \hookrightarrow \mathcal{C}(G) \xrightarrow{\tilde{\chi}_M} \mathcal{D}_{c,G\text{-harm}}(M)^* \longrightarrow\!\!\!\!\rightarrow \mathcal{D}_{c,\mathrm{orb}}(M)^*,$$

in which the composition of any two arrows is predetermined. The mapping $\tilde{\chi}_M$ is thus uniquely determined up to an $\mathrm{Aut}(G, K, M, c)$-fixed linear transformation

$$C:\ \mathcal{C}(G)/\mathcal{C}_{c,N_1}(G) \longrightarrow \big(\mathcal{D}_{c,G\text{-harm}}(M)/\mathcal{D}_{c,\mathrm{orb}}(M)\big)^*.$$

However, the last refinement of our construction is not compatible with that of Proposition 12.1. This is because an induced distribution $\rho = \rho_1^M$ in $\mathcal{D}_c(M)$ may be orbital without the inducing distribution ρ_1 being so. The conditions of Proposition 12.1 and of [A5] are thus to be regarded as two separate constraints. We are free to impose either one of them on the general construction of Proposition 9.3, but not both together. The decision of which one to choose in any given setting would depend of course upon the context.

Acknowledgements Supported in part by a Guggenheim Fellowship, the Institute for Advanced Study, and an NSERC Operating Grant.

List of Symbols

This is a partial list of symbols that occur in the paper, arranged according to the order in which they first appear. I hope that the logic of the order (such as it is) might be of some help to a reader in keeping track of their meaning. There are some overlapping symbols, due primarily to the notational conventions described prior to the commutative diagram in Sect. 5.

Section 2

Section 3

Section 4

Section 5

Section 6

Section 7

Section 8

Section 9

Section 10

Section 11

References

[A1] J. Arthur, The characters of discrete series as orbital integrals. Invent. Math. **32**, 205–261 (1976)
[A2] J. Arthur, The trace formula in invariant form. Ann. Math. **114**, 1–74 (1981)
[A3] J. Arthur, The local behaviour of weighted orbital integrals. Duke Math. J. **56**, 223–293 (1988)
[A4] J. Arthur, The invariant trace formula I. Local theory. J. Am. Math. Soc. **1**, 323–383 (1988)
[A5] J. Arthur, The trace Paley-Wiener theorem for Schwartz functions. Contemp. Math. **177**, 171–180 (1994)
[A6] J. Arthur, On local character relations. Selecta Math. **2**, 501–579 (1996)
[A7] J. Arthur, Canonical normalization of weighted characters and a transfer conjecture. C.R. Math. Rep. Acad. Sci. Can. **20**, 33–52 (1998)
[A8] J. Arthur, Stabilization of a family of differential equations. Proc. Symp. Pure Math. **158**, 77–98 (2000)
[B1] A. Bouaziz, Intégrales orbitales sur les algèbres de Lie réductives. Invent. Math. **115**, 163–207 (1994)
[B2] A. Bouaziz, Intégrales orbitales sur les groupes de Lie réductives. An. Scient. Éc. Norm. Sup. **27**, 573–609 (1994)
[BV] D. Barbasch, D. Vogan, The local structure of characters. J. Func. Anal. **37**, 27–55 (1980)
[D] J. Dieudonné, *Foundations of Modern Analysis* (Academic Press, New York, 1969)
[H1] Harish-Chandra, Invariant eigen distributions on a semisimple Lie group. Trans. Am. Math. Soc. **119**, 457–508 (1965)
[H2] Harish-Chandra, Two theorems on semisimple Lie groups. Ann. Math. **83**, 74–128 (1966)
[H3] Harish-Chandra, Discrete series for semisimple Lie groups, II. Explicit determination of the characters. Acta. Math. **116**, 1–111 (1966)
[H4] Harish-Chandra, Harmonic analysis on real reductive groups I. The theory of the constant term. J. Funct. Anal. **19**, 104–204 (1975)
[Ho] W. Hoffmann, On the differential equations satisfied by weighted orbital integrals. Ann. Math. **155**, 759–802 (2002)
[W] J.-L. Waldspurger, Stabilisation de la formule des traces tordues V: intégrales orbitales et endoscopie sur le corps réel. Preprint

Slopes of Modular Forms

Kevin Buzzard and Toby Gee

Abstract We survey the progress (or lack thereof!) that has been made on some questions about the p-adic slopes of modular forms that were raised by the first author in Buzzard (Astérisque 298:1–15, 2005), discuss strategies for making further progress, and examine other related questions.

1 Introduction

1.1 Overview

The question of the distribution of the local components of automorphic representations at finite places has received a great deal of attention.

In the case of fixing an automorphic representation and varying the finite place, we now have the recently proved Sato–Tate conjecture for elliptic curves over totally real fields [HSBT10, CHT08, Tay08]. More recently, there has been much progress on questions where the automorphic representation varies, but the finite place is fixed; see [Shi12], and the references discussed in its introduction, for a detailed history of the question. Still more recently, there has been the fascinating work of Shin and Templier [ST12] on hybrid problems, where both the finite place and the automorphic representation are allowed to vary, but we will have nothing to say about this here.

In this survey we will consider some other variants of this basic question, including p-adic ones. Just as in the classical setting, there are really several questions here, which will have different answers depending on what is varying: for example, if one fixes a weight 2 modular form corresponding to a non-CM elliptic curve, then it is ordinary for a density one set of primes; however, if one fixes a prime and a level and considers eigenforms of all weights, then almost none of them are ordinary (the dimension of the ordinary part remains bounded by Hida theory as the weight gets bigger).

K. Buzzard (✉) • T. Gee
Department of Mathematics, Imperial College London, London, UK
e-mail: k.buzzard@imperial.ac.uk; toby.gee@imperial.ac.uk

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*,
Simons Symposia, DOI 10.1007/978-3-319-41424-9_2

We will for the most part limit ourselves to the case of classical modular forms for several reasons. The questions we consider are already interesting (and largely completely open) in this case, and in addition, there appear to be interesting phenomena that we do not expect to generalise in any obvious way (see Remark 4.1.9 below.) However, it seems worth recording a natural question (from the point of view of the p-adic Langlands program) about the distribution of local parameters as the tame level varies; for concreteness, we phrase the question for GL_n over a CM field, but the same question could be asked in greater generality in an obvious fashion.

Fix a CM field F, and consider regular algebraic essentially conjugate self-dual cuspidal automorphic representations π of GL_n/F. Fix an isomorphism between $\overline{\mathbb{Q}}_p$ and $\mathbb{C}$, and a place $v|p$ of F. Assume that π_v is unramified (one could instead consider π_v lying on a particular Bernstein component). To such a π is associated a Galois representation $\rho_\pi : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_p)$, and $\rho_\pi|_{\mathrm{Gal}(\overline{F_v}/F_v)}$ is crystalline, with Hodge–Tate weights determined by π_∞. (See the introduction to [CH13] for this result, and a discussion of the history of its proof. Thanks to the work of Harris et al. [HLTT13] and Varma [Var14], the result is now known without the assumption of essentially conjugate self-duality; but the cuspidal automorphic representations of a fixed regular algebraic infinite type which are *not* essentially conjugate self-dual are expected to be rather sparse, and in particular precise asymptotics for the number of such representations as the level varies are unknown, and it therefore seems unwise to speculate about equidistribution questions for them. Note that in the essentially conjugate self-dual case, these automorphic representations arise via endoscopy from automorphic representations on unitary groups which are discrete series at infinity, and can thus be counted by the trace formula.) If we now run over π' of the same infinity type, which have π'_v unramified, and which furthermore have $\overline{\rho}_{\pi'} \cong \overline{\rho}_\pi$ (the bar denoting reduction to $\mathrm{GL}_n(\overline{\mathbb{F}}_p)$), then the local representations $\rho_{\pi'}|_{\mathrm{Gal}(\overline{F}_v/F_v)}$ naturally give rise to points of the corresponding (framed) deformation ring for crystalline lifts of $\overline{\rho}_\pi|_{\mathrm{Gal}(\overline{F}_v/F_v)}$ of the given Hodge–Tate weights. The existence of level-raising congruences mean that one can often prove that this (multi)set is infinite (and it is expected to always be infinite), and one could ask whether some form of equidistribution of the $\rho_{\pi'}|_{\mathrm{Gal}(\overline{F}_v/F_v)}$ holds in the rigid-analytic generic fibre of the crystalline deformation ring.

Unfortunately, this appears to be a very hard problem. Indeed, we do not in general even know that every irreducible component of the generic fibre of the local deformation space contains even a single $\rho_{\pi'}|_{\mathrm{Gal}(\overline{F}_v/F_v)}$; it is certainly expected that this holds, and a positive solution would yield a huge improvement on the existing automorphy lifting theorems (*cf.* the introduction to [CEG$^+$13]). Automorphy lifting theorems can sometimes be used to show that if an irreducible component contains an automorphic point, then it contains a Zariski-dense set of automorphic points, but they do not appear to be able to say anything about p-adic density, or about possible equidistribution.

More generally, one could allow the weight (and, if one wishes, the level at p) to vary (as well as, or instead of, allowing the level to vary) and ask about equidistribution in the generic fibre of the full deformation ring, with no

p-adic Hodge theoretic conditions imposed. The points arising will necessarily lie on the sublocus of crystalline (or more generally, if the level at p varies, potentially semistable) representations, but as these are expected to be Zariski dense (indeed, this is known in most cases by the results of Chenevier [Che13] and Nakamura [Nak14]), it seems reasonable to conjecture that the points will also be Zariski dense.

One could also consider the case of a place $v \nmid p$, where very similar questions could be asked (except that there are no longer any p-adic Hodge-theoretic conditions), and we are similarly ignorant (although the automorphy lifting machinery can often be used to show that each irreducible component contains an automorphic point, using the Khare–Wintenberger method [KW09, Theorem 3.3] and Taylor's Ihara-avoidance result [Tay08]; see [Gee11, §5]).

In the case of modular forms (over $\mathbb{Q}$) one can make all of this rather more concrete, due to a pleasing low-dimensional coincidence: an irreducible two-dimensional crystalline representation of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ is almost always completely determined by its Hodge–Tate weights and the trace of the crystalline Frobenius (because there is almost always a unique weakly admissible filtration on the associated filtered ϕ-module—see Sect. 4.1 below). This means that if we work with modular forms of weight k and level prime to p, the local p-adic Galois representation is almost always determined by the Hecke eigenvalue a_p (modulo the issue of semisimplicity in the ordinary case), and the question above reduces to the question of studying the p-adic behaviour of a_p. Such questions were studied computationally (and independently) by Gouvêa and one of us (KB), for the most part in level 1, when the era of computation of modular forms was in its infancy. Gouvêa noticed (see the questions in §2 of [Gou01]) that in weight k, the p-adic valuation $v_p(a_p)$ of a_p (normalised so that $v_p(p) = 1$) was almost always at most $(k-1)/(p+1)$, an observation which at the time did not appear to be predicted by any conjectures. Gouvêa and Buzzard also noticed that $v_p(a_p)$ was almost always an integer, an observation which even now is not particularly well understood. Furthermore, in level 1, the primes p for which there existed forms with $v_p(a_p) > (k-1)/(p+1)$ seemed to *coincide* with the primes for which there existed forms with $v_p(a_p) \notin \mathbb{Z}$. These led Buzzard in §1 of [Buz05] to formulate the notion of an $\mathrm{SL}_2(\mathbb{Z})$-irregular prime, a prime for which there exists a level 1 non-ordinary eigenform of weight at most $p+1$. Indeed one might even wonder whether the following are equivalent:

- p is $\mathrm{SL}_2(\mathbb{Z})$-irregular;
- There exists a level 1 eigenform with $v_p(a_p) \notin \mathbb{Z}$;
- There exists a level 1 eigenform of weight k with $v_p(a_p) > (k-1)/(p+1)$.

One can check whether a given prime p is $\mathrm{SL}_2(\mathbb{Z})$-regular or not in finite time (one just needs to compute the determinant of the action of T_p on level 1 modular forms of weight k for each $k \le p+1$ and check if it is always a p-adic unit; in fact, one only has to check cusp forms of weights $4 \le k \le (p+3)/2$ because of known results about θ-cyles); one can also verify with machine computations that the second or third conditions hold by exhibiting an explicit eigenform with the property in question. The authors do not know how to verify with machine computations that the second or third conditions fail; equivalently, how to prove for a given p either that all T_p-eigenvalues a_p of all level 1 forms of all weights have integral p-adic valuations, or that they all satisfy $v_p(a_p) \leq (k-1)/(p+1)$. In particular it is still logically possible that for *every* prime number there will be some level 1 eigenforms satisfying $v_p(a_p) \notin \mathbb{Z}$ or $v_p(a_p) > (k-1)/(p+1)$. However this seems very unlikely—for example $p = 2$ is an $\mathrm{SL}_2(\mathbb{Z})$-regular prime, and the first author has computed $v_p(a_p)$ for $p = 2$ and for all $k \leq 2048$ and has found no examples where $v_2(a_2) \notin \mathbb{Z}$ or $v_2(a_2) > (k-1)/3$. Gouvea also made substantial calculations for all other $p < 100$ which add further weight to the idea that the conditions are equivalent.

There are precisely two $\mathrm{SL}_2(\mathbb{Z})$-irregular primes less than 100, namely 59 and 79, and it does not appear to be known whether there are infinitely many $\mathrm{SL}_2(\mathbb{Z})$-regular primes or whether there are infinitely many $\mathrm{SL}_2(\mathbb{Z})$-irregular primes. (However, Frank Calegari has given https://galoisrepresentations.wordpress.com/2015/03/03/review-of-buzzard-gee/ an argument which shows that under standard conjectures about the existence of prime values of polynomials with rational coefficients, then there are infinitely many $\mathrm{SL}_2(\mathbb{Z})$-irregular primes.) Note that for $p = 59$ and $p = 79$ eigenforms with $v_p(a_p) \notin \mathbb{Z}$ and $v_p(a_p) > (k-1)/(p+1)$ do exist, but any given eigenform will typically satisfy at most one of these conditions, and we do not even know how to show that the second and third conditions are equivalent.

Buzzard conjectured that for an $\mathrm{SL}_2(\mathbb{Z})$-regular prime, $v_p(a_p)$ was integral for all level 1 eigenforms, and even conjectured an algorithm to compute these valuations in all weights. Similar conjectures were made at more general levels $N > 1$ prime to p, and indeed Buzzard formulated the notion of a $\Gamma_0(N)$-regular prime—for $p > 2$ this is a prime $p \nmid N$ such that all eigenforms of level $\Gamma_0(N)$ and weight at most $p+1$ are ordinary, although here one has to be a little more careful when $p = 2$ (and even for $p > 2$ some care needs to be taken when generalising this notion to $\Gamma_1(N)$ because allowing odd weights complicates the picture somewhat; see Remark 4.1.5.)

These observations of Buzzard and Gouvêa can be thought of as saying something about the behaviour of the Coleman–Mazur eigencurve near the centre of weight space. Results of Buzzard and Kilford [BK05], Roe [Roe14], Wan et al. [WXZ14], and Liu et al. [LWX14] indicate that there is even more structure near the boundary of weight space; this structure translates into concrete assertions about $v_p(a_p)$ when a_p is the U_p-eigenvalue of a newform of level $\Gamma_1(Np^r)$ and character of conductor Mp^r for some $M \mid N$ coprime to p. We make precise conjectures in Sect. 4.2. On the other hand, perhaps these results are intimately related to the p-adic Hodge-theoretic coincidence alluded to above—that in this low-dimensional situation there is usually only one (up to isomorphism) weakly admissible filtration on the Weil–Deligne representation in question. In particular such structure might not be so easily found in a general unitary group eigenvariety.

Having formulated these conjectures, in Sect. 5 we discuss a potential approach to them via modularity lifting theorems.

1.2 Acknowledgements

We would like to thank Christophe Breuil, Frank Calegari, and Matthew Emerton for many helpful conversations over the past 14 years about the subjects of this survey. We would like to thank Vytautas Paškūnas for a helpful conversation about the Breuil–Mézard conjecture, and Liang Xiao for explaining some aspects of the papers [WXZ14, LWX14] to us. We would like to thank Frank Calegari, Matthew Emerton, and James Newton for their helpful comments on an earlier draft of this paper; in addition, we would like to thank Frank Calegari for refereeing the paper (https://galoisrepresentations.wordpress.com/2015/03/03/review-of-buzzard-gee/) and for making several helpful comments in doing so. The second author would like to thank the organisers of the Simons Symposium on Automorphic Forms and the Trace Formula for the invitation to speak at the symposium, and for the suggestion of turning the rough notes for the talk into this survey.

2 Limiting Distributions of Eigenvalues

In this section we briefly review some conjectures and questions about the limiting distributions of eigenvalues of Hecke operators in the p-adic context. These questions will not be the main focus of our discussions, but as they are perhaps the most natural analogues of the questions considered in [ST12], it seems worth recording them.

2.1 $\ell = p$: Conjectures of Gouvêa

The reference for this section is the paper [Gou01]. Fix a prime p, an integer $N \geq 1$ coprime to p, and consider the operator U_p on the spaces of classical modular forms $S_k(\Gamma_0(Np))$ for varying weights $k \geq 2$. The characteristic polynomial of U_p has integer coefficients so it makes sense to consider the slopes of the eigenvalues—by definition, these are the p-adic valuations of the eigenvalues considered as elements of $\overline{\mathbb{Q}}_p$. The eigenvalues themselves fall into two categories. The ones corresponding to eigenforms which are new at p (corresponding to Steinberg representations) have U_p-eigenvalues $\pm p^{(k-2)/2}$, and thus slope $(k-2)/2$. The other eigenvalues come in pairs, each pair being associated with an eigenvalue of T_p on $S_k(\Gamma_0(N))$, and if the T_p-eigenvalue is a_p (considered as an element of $\overline{\mathbb{Q}}_p$), then the corresponding two p-oldforms have eigenvalues given by the roots of $x^2 - a_px + p^{k-1}$; so the slopes $\alpha, \beta \in [0, k-1]$ satisfy $\alpha + \beta = k-1$. Note that $\min\{\alpha, \beta\} = \min\{v_p(a_p), \frac{k-1}{2}\}$ by the theory of the Newton polygon, and in particular if $v_p(a_p) < \frac{k-1}{2}$ then $v_p(a_p)$ can be read off from α and β.

Now consider the (multi-)set of slopes of p-oldforms, normalised by dividing by $k-1$ to lie in the range $[0, 1]$. More precisely we could consider the measure (a finite sum of point measures, normalised to have total mass 1) attached to this multiset in weight k. Let k tend to ∞ and consider how these measures vary. Is there a limiting measure?

Conjecture 2.1.1. (Gouvêa) The slopes converge to the measure which is uniform on $[0, \frac{1}{p+1}] \cup [\frac{p}{p+1}, 1]$ and 0 elsewhere.

This is supported by the computational evidence, which is particularly convincing in the $\Gamma_0(N)$-regular case. This conjecture implies that if a_p runs through the eigenvalues of T_p on $S_k(\Gamma_0(N))$ then we "usually" have $v(a_p) \leq (k-1)/(p+1)$. This appears to be the case, although the reasons why are not well understood. If $p > 2$ is $\Gamma_0(N)$-regular however, the (purely local—see below) main result of [BLZ04] shows that $v(a_p) \leq \lfloor (k-2)/(p-1) \rfloor$. One might hope that the main result of [BLZ04] could be strengthened to show that in fact $v(a_p) \leq \frac{k-1}{p+1}$; it seems likely that the required local statement is true, but Berger tells us that the proof in [BLZ04] does not seem to extend to this more general range. (This problem is carefully examined in Mathieu Vienney's unpublished PhD thesis.)

2.2 $\ell \neq p$

In the previous subsection we talked about the distribution of a_p, the eigenvalues of T_p on $S_k(\Gamma_0(N))$, considered as elements of $\overline{\mathbb{Q}}_p$. The Ramanujan bounds and the Sato–Tate conjecture give us information about the eigenvalues of T_p as elements of the complex numbers. What about the behaviour of the a_p as elements of $\overline{\mathbb{Q}}_l$ for $\ell \neq p$ prime? We have very little idea what to expect. In this short section we merely present a sample of some computational results concerning the even weaker question of the distribution of the reductions of the a_p as elements of $\overline{\mathbb{F}}_l$. In contrast to the previous section we here vary N and keep $k = 2$ fixed. More precisely, we fix distinct ℓ and p, and then loop over $N \geq 1$ coprime to ℓp and compute the eigenvalues $\overline{a}_p$ of T_p acting on $S_2(\Gamma_0(N); \overline{\mathbb{F}}_l)$. On the next page is a sample of the results with $p = 5$ and $\ell = 3$, looping over the first 5,533,155 newforms. The first numbers in the second column of this table are *not* decreasing, which is perhaps not what one might initially guess; Frank Calegari observed that this could perhaps be explained by observing that if you choose a random element of a finite field $\mathbb{F}_q$ then the field it generates over $\mathbb{F}_p$ might be strictly smaller than $\mathbb{F}_q$, and the heuristics are perhaps complicated by this.

Size of $\mathbb{F}_3[\overline{a}_5]$	Number
3^1	80,656
3^2	38,738
3^3	35,880
3^4	32,968
3^5	35,330
3^6	33,372
3^7	34,601
3^8	33,896
3^9	35,262
3^{10}	33,600

3 The Gouvêa–Mazur Conjecture/Buzzard's Conjectures

Coleman theory (see Theorem D of [Col97]) tells us that for a fixed prime p and tame level N, there is a function $M(n)$ such that if $k_1, k_2 > n + 1$ and $k_1 \equiv k_2 \pmod{p^{M(n)}(p-1)}$, then the sequences of slopes (with multiplicities) of classical modular forms of level Np and weights k_1, k_2 agree up to slope n. A more geometric way to think about this theorem is that given a point on the eigencurve of slope $\alpha \leq n$, there is a small neighbourhood of that point in the eigencurve, which maps in a finite manner down to a disc in weight space of some explicit radius $p^{-M(n)}$ and such that all the points in the neighbourhood have slope α. Gouvêa and Mazur [GM92] conjectured that we could take $M(n) = n$; for $n = 0$, this is a theorem of Hida (his ordinary families are finite over entire components of weight space). Wan [Wan98] deduced from Coleman's results that $M(n)$ could be taken to be quadratic (with the implicit constants depending on both p and N; as far as we know, it is still an open problem to obtain a quadratic bound independent of either p or N). However, Buzzard and Calegari [BC04] found an explicit counterexample to the conjecture that $M(n) = n$ always works.

On the other hand, Buzzard [Buz05] accumulated a lot of numerical evidence that whenever p is $\Gamma_0(N)$-regular, many (but not all) families of eigenforms seemed to have slopes which were locally equal to n on discs of size $p^{-L(n)}$ with $L(n)$ seemingly linear in $\log(n)$—a much stronger bound than the Gouvêa–Mazur conjectures. For example, if $p = 2$, $N = 1$, then the classical slopes at weight $k = 2^d$ (the largest of which is approximately $k/3$) seem to be an initial segment of the classical slopes at weight 2^{d+1}. For example, the 2-adic slopes in level 1 and weight $128 = 2^7$ are

$$3, 7, 13, 15, 17, 25, 29, 31, 33, 37$$

and the Gouvêa–Mazur conjectures would predict that the slopes which were at most 7 should show up in weight $256 = 2^8$. However in weight 256 the slopes are

$$3, 7, 13, 15, 17, 25, 29, 31, 33, 37, 47, 49, 51, \ldots$$

and more generally the slopes at weight equal to a power of 2 all seem to be initial segments of the infinite slope sequence on overconvergent 2-adic forms of weight 0, a sequence explicitly computed in Corollary 1 of [BC05]. In particular, if one were to restrict to $p = 2$, $N = 1$ and k a power of 2, then $M(n)$ can be conjecturally taken to be the base 2 logarithm of $3n$. Note also that the counterexamples at level $\Gamma_0(N)$ to the Gouvêa–Mazur conjecture in [BC04] were all $\Gamma_0(N)$-irregular. It may well be the case that the Gouvêa–Mazur conjectures are true at level $\Gamma_0(N)$ if one restricts to $\Gamma_0(N)$-regular primes—indeed the numerical examples above initially seem to lend credence to the hope that something an order of magnitude stronger than the Gouvêa–Mazur conjectures might be true in the $\Gamma_0(N)$-regular case. However life is not quite so easy—numerical evidence seems to indicate that near to a newform for $\Gamma_0(Np)$ on the eigencurve, the behaviour of slopes seems to be broadly speaking behaving in the same sort of way as predicted by the Gouvêa–Mazur conjectures. For example, again with $p = 2$ and $N = 1$, computer calculations give that the slopes in weight $38 + 2^8$ are

$$5, 8, 16, 18, 18, 20, 29, 32, 37, 40, 45, 50, 50, 56, 61, 64, 70, \ldots$$

whereas in weight $38 + 2^9$ they are

$$5, 8, 17, 18, 18, 19, 29, 32, 37, 40, 45, 50, 50, 56, 61, 64, 70, \ldots.$$

Again one sees evidence of something far stronger than the Gouvêa–Mazur conjecture going on (the Gouvêa–Mazur conjecture only predicts equality of slopes which are at most 8); however, there seems to be a family which has slope 16 in weight $38 + 2^8$ and slope 17 in weight $38 + 2^9$. This family could well be passing through a classical newform of level $\Gamma_0(2)$ in weight 38, and newforms in weight 38 have slope $(38 - 2)/2 = 18$, so one sees that for just this one family $M(n)$ is behaving much more like something linear in n.

Staying in the $\Gamma_0(N)$-regular case, Buzzard found a lot of evidence for a far more precise conjecture than the Gouvêa–Mazur conjecture—one that gives a complete description of the slopes in the $\Gamma_0(N)$-regular case, in terms of a recursive algorithm, which is purely combinatorial in nature and uses nothing about modular forms at all.[1] Then (see [Buz05, §3] for a more detailed discussion) the algorithm can for the most part be deduced from various heuristic assumptions about families of p-adic

[1] A preprint "Slopes of modular forms and the ghost conjecture" by John Bergdall and Robert Pollack gives a much more natural conjectural algorithm for the slopes, the output of which presumably coincides with Buzzard's algorithm.

modular forms, for example the very strong “logarithmic” form of the Gouvêa–Mazur conjecture mentioned above, plus some heuristics about behaviour of slopes near newforms that seem hard to justify. Unfortunately, essentially nothing is known about these conjectures, even in the simplest case $N = 1$ and $p = 2$, where the slopes are all conjectured to be integers but even this is not known.

In fact it does not even seem to be known that the original form of the Gouvêa–Mazur conjecture (in the $\Gamma_0(N)$-regular case) is a consequence of Buzzard's conjectures; see [Buz05, Question 4.11]. It would also be of interest to examine Buzzard's original data to try to formulate a precise conjecture about the best possible value of $M(n)$ in the $\Gamma_0(N)$-regular case. The following are combinatorial questions, and are presumably accessible.

Question 3.1. Say p is $\Gamma_0(N)$-regular.

(1) Does the Gouvêa–Mazur conjecture for (p, N), or perhaps something even stronger, follow from Buzzard's conjectures?
(2) Does Conjecture 2.1.1 follow from Buzzard's conjectures?

One immediate consequence of Buzzard's conjectures is that in the $\Gamma_0(N)$-regular case, all of the slopes should be integers. This can definitely fail in the $\Gamma_0(N)$-irregular case (and is a source of counterexamples to the Gouvêa–Mazur conjecture), and we suspect that understanding this phenomenon could be helpful in proving the full conjectures (see the discussion in Sect. 5 below). In Sect. 4.1 we will explain a purely local conjecture that would imply this integrality.

Note that Lisa Clay's PhD thesis also studies this problem and makes the observation that the combinatorial recipes seem to remain valid when restricting to the subset of eigenforms with a fixed mod p Galois representation which is reducible locally at p.

4 Local Questions

4.1 The Centre of Weight Space

In this section we discuss some purely local conjectures and questions about p-adic Galois representations that are motivated by the conjectures of Sect. 3. We briefly recall the relevant local Galois representations and their relationship to the global picture, referring the reader to the introduction to [BG09] for further details. If $k \geq 2$ and $a_p \in \overline{\mathbb{Q}}_p$ with $v(a_p) > 0$, then there is a two-dimensional crystalline representation V_{k,a_p} with Hodge–Tate weights $0, k-1$, with the property that the crystalline Frobenius of the corresponding weakly admissible module has characteristic polynomial $X^2 - a_pX + p^{k-1}$. Furthermore, if $a_p^2 \neq 4p^{k-1}$, then V_{k,a_p} is uniquely determined up to isomorphism. This is easily checked by directly computing the possible Hodge filtrations on the weakly admissible module; see for example, [BB10, Proposition 2.4.5]. This is a low-dimensional coincidence however—a certain parameter space of flags is connected of dimension zero in this situation.

The relevance of this representation to the questions of Sect. 3 is that if $f \in S_k(\Gamma_0(N), \overline{\mathbb{Q}}_p)$ is an eigenform with $a_p^2 \neq 4p^{k-1}$ (which is expected to always hold; in the case $N = 1$ it holds by Theorem 1 of [Gou01], and the paper [CE98] proves that it holds for general N if $k = 2$, and for general k, N if one assumes the Tate conjecture) then $\rho_f|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)} \cong V_{k,a_p}$.

As explained in [Buz05, §1], $p > 2$ is $\Gamma_0(N)$-regular if and only if $\overline{\rho}_f|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is reducible for every $f \in S_k(\Gamma_0(N))$ (and every $k \geq 2$). This suggests that the problem of determining when $\overline{V}_{k,a_p}$ (the reduction of V_{k,a_p} modulo p) is reducible could be relevant to the conjectures of Sect. 3. To this end, we have the following conjecture.

Conjecture 4.1.1. If p is odd, k is even and $v(a_p) \notin \mathbb{Z}$ then $\overline{V}_{k,a_p}$ is irreducible.

Remark 4.1.2. Any modular form of level $\Gamma_0(N)$ necessarily has even weight, and this conjecture would therefore imply for $p > 2$ that in the $\Gamma_0(N)$-regular case, all slopes are integral, as Buzzard's conjectures predict (see Sect. 3 above).

Remark 4.1.3. This conjecture is arguably "folklore" but seems to originate in emails between Breuil, Buzzard, and Emerton in 2005.

Remark 4.1.4. The conjecture is of course false without the assumption that $v(a_p) \notin \mathbb{Z}$; indeed, if $v(a_p) = 0$, then we are in the ordinary case, and V_{k,a_p} is reducible (and so $\overline{V}_{k,a_p}$ is certainly reducible).

Remark 4.1.5. If k is allowed to be odd, then the conjecture would be false—for global reasons! There are p-newforms of level $\Gamma_1(N) \cap \Gamma_0(p)$ and odd weight k_0, which automatically have slope $(k_0-2)/2 \notin \mathbb{Z}$, and in computational examples these forms give rise to both reducible and irreducible local mod p representations. The corresponding local p-adic Galois representations are now semistable rather than crystalline, and depend on an additional parameter, the *$\mathcal{L}$-invariant*; the reduction of the Galois representation depends on this $\mathcal{L}$-invariant in a complicated fashion, see, for example, the calculations of Breuil and Mézard [BM02]. Considering oldforms which are sufficiently p-adically close to such newforms (and these will exist by the theory of the eigencurve) produces examples of V_{k,a_p} with $v(a_p) = (k_0 - 2)/2$ and $\overline{V}_{k,a_p}$ reducible. If k and k_0 are close in weight space, then k will also be odd.

The main result of Buzzard and Gee [BG13] determines, for odd p, exactly for which a_p with $0 < v(a_p) < 1$ the representation $\overline{V}_{k,a_p}$ is irreducible; it is necessary that $k \equiv 3 \pmod{p-1}$, that $k \geq 2p + 1$, and that $v(a_p) = 1/2$, and there are examples for all k satisfying these conditions.

Remark 4.1.6. If $p = 2$, then the conjecture is also false for the trivial reason that if $k \equiv 4 \bmod 6$ then $\overline{V}_{k,0}$ is reducible and hence $\overline{V}_{k,a}$ is reducible for $v(a)$ sufficiently large (whether or not it is integral) by the main result of Berger et al. [BLZ04]. In particular, the conjecture does not offer a local explanation for the global phenomenon that thousands of slopes of cusp forms have been computed for $N = 1$ and $p = 2$, and not a single non-integral one has been found (and the conjectures of [Buz05] predict that the slopes will all be integral).

Remark 4.1.7. Conjecture 4.1.1 is known if $v(a_p) \in (0, 1)$, which is the main result of Buzzard and Gee [BG09]. It is also known if $v(a_p) > \lfloor (k-2)/(p-1) \rfloor$, by the main result of Berger et al. [BLZ04]. In the case that $k \leq (p^2+1)/2$, it is expected to follow from work in progress of Yamashita and Yasuda.

The result of Berger et al. [BLZ04] is proved by constructing an explicit family of (ϕ, Γ)-modules which are p-adically close to the representation $V_{k,0}$. Since $V_{k,0}$ is induced from a Lubin–Tate character, it has irreducible reduction if k is not congruent to 1 modulo $p+1$, and in particular has irreducible reduction when $p > 2$ and k is even, which implies the result.

In contrast, the papers [BG09, BG13] use the p-adic local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$ to compute $\overline{V}_{k,a_p}$ more or less explicitly. Despite the simplicity of the calculations of Buzzard and Gee [BG09], which had originally made us optimistic about the prospects of proving Conjecture 4.1.1 in general, it seems that when $v(a_p) > 1$ the calculations involved in computing $\overline{V}_{k,a_p}$ are very complicated, and without having some additional structural insight we are pessimistic that Conjecture 4.1.1 can be directly proved by this method.

In the light of the previous remark, we feel that it is unlikely that Conjecture 4.1.1 will be proved without some gaining some further understanding of why it should be true. We therefore regard the following question as important.

Question 4.1.8. Are there any local or global reasons that we should expect Conjecture 4.1.1 to hold, other than the computational evidence of the second author discussed in [Buz05]?

Remark 4.1.9. It seems unlikely that any analogue of Conjecture 4.1.1 will hold in a more general setting (i.e., for higher-dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$, or for representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/F)$ of dimension > 1, where $F/\mathbb{Q}_p$ is a non-trivial extension). The reason for this is that there is no analogue of the fact that V_{k,a_p} is completely determined by k and a_p; in these more general settings, additional parameters are needed to describe the p-adic Hodge filtration, and it is highly likely that the reduction mod p of the crystalline Galois representations will depend on these parameters. (Indeed, as remarked above, this already happens for semistable 2-dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$.)

For this reason we are sceptical that there is any simple generalisation of the conjectures of Sect. 3, except to the case of Hilbert modular forms over a totally real field in which p splits completely. For example, Table 5 in [Loe08] and the comments below it show that non-integral slopes appear essentially immediately when one computes with $U(3)$.

4.2 *The Boundary of Weight Space*

Perhaps surprisingly, near the boundary of weight space, the combinatorics of the eigencurve seem to become simpler. For example, if $N = 1$ and $p = 2$, one can compare Corollary 1 of [BC05] (saying that in weight 0 all overconvergent slopes are determined by a complicated combinatorial formula) with Theorem B of [BK05] (saying that at the boundary of weight space the slopes form an arithmetic progression).

Now let f be a newform of weight $k \geq 2$ and level $\Gamma_1(Np^r)$, with $r \geq 2$, and with character whose p-part χ has conductor p^r. For simplicity, fix an isomorphism $\mathbb{C} = \overline{\mathbb{Q}}_p$. Say f has U_p-eigenvalue α. One checks that the associated smooth admissible representation of $\mathrm{GL}_2(\mathbb{Q}_p)$ attached to f must be principal series associated with two characters of $\mathbb{Q}_p^\times$, one unramified (and sending p to α) and the other of conductor p^r. Now say ρ_f is the p-adic Galois representation attached to f.

By local-global compatibility (the main theorem of [Sai97]), and the local Langlands correspondence, the F-semisimplified Weil–Deligne representation associated with ρ_f at p will be the direct sum of two characters, one unramified and the other of conductor p^r. Moreover, the p-adic Hodge-theoretic coincidence still holds: there is at most one possible weakly admissible filtration on this Weil–Deligne representation with jumps at 0 and $k-1$, by Proposition 2.4.5 of [BB10] (or by a direct calculation).

The resulting weakly admissible module depends only on k, α, and χ, and so we may call its associated Galois representation $V_{k,\alpha,\chi}$; the local–global assertion is then that this representation is the restriction of ρ_f to the absolute Galois group of $\mathbb{Q}_p$. Let $\overline{V}_{k,\alpha,\chi}$ denote the semisimplification of the mod p reduction of $V_{k,\alpha,\chi}$. We propose a conjecture which would go some way towards explaining the results of Buzzard and Kilford [BK05], Roe [Roe14], Kilford [Kil08], and Kilford and McMurdy [KM12]. We write v_χ for the p-adic valuation v on $\overline{\mathbb{Q}}_p$ normalised so that the image of v_χ on $\mathbb{Q}_p(\chi)^\times$ is $\mathbb{Z}$ (so for $p > 2$ we have $v_\chi(p) = 1/(p-1)p^{r-2}$.)

Conjecture 4.2.1. If $v_\chi(\alpha) \notin \mathbb{Z}$, then $\overline{V}_{k,\alpha,\chi}$ is irreducible.

This is a local assertion so does not follow directly from the results in the global papers cited above. The four papers above prove that $v_\chi(\alpha) \in \mathbb{Z}$ if α is an eigenvalue of U_p on a space of modular forms of level 2^r, 3^r, 5^2, and 7^2, respectively; note that in all these cases, all the local mod p Galois representations which show up are reducible locally at p, for global reasons. In fact, slightly more is true in the special case $p = 2$ and $r = 2$: in this case $\mathbb{Q}_p(\chi) = \mathbb{Q}_2$ so the conjecture predicts that if $v(\alpha) \notin \mathbb{Z}$ then $\overline{V}_{k,\alpha,\chi}$ is irreducible; yet in [BK05] it is proved that eigenforms of odd weight, level 4, and character of conductor 4, all have slopes in $2\mathbb{Z}$.

It is furthermore expected that in the global setting the sequence of slopes is a finite union of arithmetic progressions; see [WXZ14, Conjecture 1.1]. Indeed, a version of this statement (sufficiently close to the boundary of weight space, in the setting of the eigenvariety for a definite quaternion algebra with $p > 2$) is proved by Liu et al. in [LWX14].

5 A Strategy to Prove Buzzard's Conjectures

The following strategy for attacking the conjectures of Sect. 3 was explained by the second author to the first author in 2005, and was the motivation for the research reported on in the papers [BG09, BG13] (which we had originally hoped would result in a proof of Conjecture 4.1.1).

Assume that $p > 2$, and fix a continuous odd, irreducible (and thus modular, by Serre's conjecture), representation $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$. Assume further that $\overline{\rho}$ satisfies the usual Taylor–Wiles condition that $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_p))}$ is irreducible.

Let $R_k^{\mathrm{loc}}(\overline{\rho})$ be the (reduced and p-torsion free) universal framed deformation ring for lifts of $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ which are crystalline with Hodge–Tate weights $0, k-1$. This connects to the global setting via the following consequence of the results of Kisin [Kis09].

Proposition 5.1. *Maintain the assumptions and notation of the previous two paragraphs, so that $p > 2$, and $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ is a continuous, odd, irreducible representation with $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_p))}$ irreducible.*

Let N be an integer not divisible by p such that $\overline{\rho}$ is modular of level $\Gamma_1(N)$. If $p = 3$, assume further that $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is not a twist of the direct sum of the mod p cyclotomic character and the trivial character. Fix an irreducible component of $\mathrm{Spec}\, R_k^{\mathrm{loc}}(\overline{\rho})[1/p]$. Then there is a newform $f \in S_k(\Gamma_1(N), \overline{\mathbb{Q}}_p)$ such that $\overline{\rho}_f \cong \overline{\rho}$, and $\rho_f|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ corresponds to a point of $\mathrm{Spec}\, R_k^{\mathrm{loc}}(\overline{\rho})[1/p]$ lying on our chosen component.

Proof. This follows almost immediately from the results of Kisin [Kis09], exactly as in the proof of [Cal12, Proposition 3.7]. (Note that the condition that f is a newform of level $\Gamma_1(N)$ can be expressed in terms of the conductor of ρ_f, and thus in terms of the components of the local deformation rings at primes dividing N.)

More precisely, this argument immediately gives the result in the case that $\overline{\rho}_f|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is not a twist of an extension of the trivial representation by the mod p cyclotomic character. However, this assumption on $\overline{\rho}_f|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is needed only in the proof of [Kis09, Corollary 2.2.17], where this assumption guarantees that the Breuil–Mézard conjecture holds for $\overline{\rho}_f|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ (indeed, the Breuil–Mézard conjecture is proved under this assumption in [Kis09]). The Breuil–Mézard conjecture is now known for $p > 2$, except in the case that $p = 3$ and $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is a twist of the direct sum of the mod p cyclotomic character and the trivial character, so the result follows. (The case that $p \geq 3$ and $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is a twist of a non-split extension of the trivial character by the mod p cyclotomic character is treated in [Paš15], and the case that $p > 3$ and $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is a twist of the direct sum of the mod p cyclotomic character and the trivial character is proved in [HT13].) □

Suppose that $\overline{\rho}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is reducible, and that Conjecture 4.1.1 holds. Consider a_p as a rigid-analytic function on $\mathrm{Spec}\, R_k^{\mathrm{loc}}(\overline{\rho})[1/p]$; since $v(a_p) \in \mathbb{Z}$ by assumption, we see that $v(a_p)$ is in fact constant on connected (equivalently, irreducible) components of $\operatorname{Spec} R_k^{\mathrm{loc}}(\overline{\rho})[1/p]$.

Corollary 5.2. *Maintain the assumptions of Proposition 5.1, and assume further that $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is reducible. Assume Conjecture 4.1.1. Then the set of slopes (without multiplicities) of T_p on newforms $f \in S_k(\Gamma_1(N), \overline{\mathbb{Q}}_p)$ with $\overline{\rho}_f \cong \overline{\rho}$ is determined purely by k and $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$; more precisely, it is the set of slopes of (the crystalline Frobenius of the Galois representations corresponding to) components of* $\operatorname{Spec} R_k^{\mathrm{loc}}(\overline{\rho})[1/p]$.

Proof. This is immediate from Proposition 5.1 (and the discussion in the preceding paragraph). □

Remark 5.3. The conclusion of Corollary 5.2 seems unlikely to hold if $\overline{\rho}$ is allowed to be (globally) reducible; for example, if $p = 2$, it is known that the slopes of all cusp forms for $\mathrm{SL}_2(\mathbb{Z})$ are at least 3, but there are local crystalline representations of slope 1 (for example, the local 2-adic representation attached to the unique weight 6 level 3 cuspidal eigenform). We do not know if there is any reasonable "local to global principle" when $\overline{\rho}$ is reducible.

It would be very interesting to be able to have some control on the multiplicities with which slopes occur in $S_k(\Gamma_1(N), \overline{\mathbb{Q}}_p)$ (for example, to show that these multiplicities agree for two weights which are sufficiently p-adically close, as predicted by the Gouvêa–Mazur conjecture), but it is not clear to us how such results could be extracted from the modularity lifting machinery. If all the irreducible components of $R_k^{\mathrm{loc}}(\overline{\rho})$ were regular, it would presumably be possible to use the argument of Diamond [Dia97] to relate the multiplicities of the same slope in different weights, but we do not expect this to hold in any generality.

Not withstanding this difficulty, one could still hope to prove the conjectures of [Buz05] up to multiplicity. If Conjecture 4.1.1 were known, the main obstruction to doing this would be obtaining a strong local constancy result for slopes as k varies p-adically. More precisely, we would like to prove the following purely local conjecture for some function $M(n)$ as in Sect. 3 above.

Conjecture 5.4. Let $\bar{r} : \mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be reducible. If $n \geq 0$ is an integer, and $k, k' \geq n + 1$ have $k \equiv k' \pmod{(p-1)p^{M(n)}}$, then there is a crystalline lift of $\bar{r}$ with Hodge–Tate weights $0, k - 1$ and slope n if and only if there is a crystalline lift of $\bar{r}$ with Hodge–Tate weights $0, k' - 1$ and slope n.

It might well be possible to prove a weak result in the direction of Conjecture 5.4 by the methods of Berger [Ber12] (more precisely, to prove the conjecture with a much worse bound on $M(n)$ than would be needed for interesting applications to the conjectures of [Buz05], but without any assumption on the reducibility of $\bar{r}$).

Corollary 5.2 (which shows, granting as always Conjecture 4.1.1, that the set of slopes which occur globally is the same as the set of slopes that occur locally) shows that it would be enough to prove the global version of this statement, and it is possible that the methods of Wan [Wan98] could allow one to deduce a local constancy result where the dependence on n in "sufficiently close" is quadratic in n. (Note that while it is not immediately clear how to adapt the methods of Wan [Wan98] to allow $\overline{\rho}$ to be fixed, it seems plausible that the methods used to prove [WXZ14, Theorem D] will be able to do this.) Note again that the computations of Buzzard and Calegari [BC04] (which in particular disprove the original Gouvêa–Mazur conjecture) mean that we cannot expect to deduce Conjecture 5.4 (for an optimal function $M(n)$ of the kind suggested by Buzzard's conjectures) from any global result that does not use the hypothesis that $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is reducible.

However, it seems plausible to us that a weak local constancy result of this kind, also valid in the case that $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is irreducible, could be bootstrapped to give the required strong constancy, provided that Conjecture 4.1.1 is proved. The idea is as follows: under the assumption of Conjecture 4.1.1, $v(a_p)$ is constrained to be an integer when $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ is reducible. If one could prove a result (with no hypothesis on the reducibility of $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$) saying that if k, k' are sufficiently close in weight space, then the small slopes of crystalline lifts of $\overline{\rho}|_{\mathrm{Gal}\,(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ of Hodge–Tate weights $0, k-1$ and $0, k'-1$ are also close, then the fact that the slopes are constrained to be integers could then be used to deduce that the slopes are equal (because two integers which differ by less than 1 must be equal.)

Acknowledgements Kevin Buzzard and Toby Gee were supported in part by EPSRC grant EP/L025485/1. The second author was additionally supported by a Leverhulme Prize, Marie Curie Career Integration Grant 303605, and ERC Starting Grant 306326.

References

[BB10] L. Berger, C. Breuil, Sur quelques représentations potentiellement cristallines de $\mathrm{GL}_2(\mathbf{Q}_p)$. Astérisque **330**, 155–211 (2010)

[BC04] K. Buzzard, F. Calegari, A counterexample to the Gouvêa-Mazur conjecture. C. R. Math. Acad. Sci. Paris **338**(10), 751–753 (2004)

[BC05] K. Buzzard, F. Calegari, Slopes of overconvergent 2-adic modular forms. Compos. Math. **141**(3), 591–604 (2005)

[Ber12] L. Berger, Local constancy for the reduction mod$p$ of 2-dimensional crystalline representations. Bull. Lond. Math. Soc. **44**(3), 451–459 (2012)

[BG09] K. Buzzard, T. Gee, Explicit reduction modulo p of certain two-dimensional crystalline representations. Int. Math. Res. Not. IMRN **12**, 2303–2317 (2009)

[BG13] K. Buzzard, T. Gee, Explicit reduction modulo p of certain 2-dimensional crystalline representations, II. Bull. Lond. Math. Soc. **45**(4), 779–788 (2013)

[BK05] K. Buzzard, L.J.P. Kilford, The 2-adic eigencurve at the boundary of weight space. Compos. Math. **141**(3), 605–619 (2005)

[BLZ04] L. Berger, H. Li, H.J. Zhu, Construction of some families of 2-dimensional crystalline representations. Math. Ann. **329**(2), 365–377 (2004)

[BM02] C. Breuil, A. Mézard, Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbf{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ en $l = p$. Duke Math. J. **115**(2), 205–310 (2002). With an appendix by Guy Henniart

[Buz05] K. Buzzard, Questions about slopes of modular forms. Astérisque **298**, 1–15 (2005). Automorphic forms. I

[Cal12] F. Calegari, Even Galois representations and the Fontaine–Mazur conjecture. II. J. Am. Math. Soc. **25**(2), 533–554 (2012)

[CE98] R.F. Coleman, B. Edixhoven, On the semi-simplicity of the U_p-operator on modular forms. Math. Ann. **310**(1), 119–127 (1998)

[CEG$^+$13] A. Caraiani, M. Emerton, T. Gee, D. Geraghty, V. Paškūnas, S.W. Shin, Patching and the p-adic local Langlands correspondence. Camb. J. Math. **4**(2), 197–287 (2016). DOI: http://dx.doi.org/10.4310/CJM.2016.v4.n2.a2

[CH13] G. Chenevier, M. Harris, Construction of automorphic Galois representations, II. Camb. J. Math. **1**(1), 53–73 (2013)

[Che13] G. Chenevier, Sur la densité des représentations cristallines de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$. Math. Ann. **355**(4), 1469–1525 (2013)

[CHT08] L. Clozel, M. Harris, R. Taylor, Automorphy for some l-adic lifts of automorphic mod l Galois representations. Publ. Math. Inst. Hautes Études Sci. **108**, 1–181 (2008). With Appendix A, summarizing unpublished work of Russ Mann, and Appendix B by Marie-France Vignéras

[Col97] R.F. Coleman, p-adic Banach spaces and families of modular forms. Invent. Math. **127**(3), 417–479 (1997)

[Dia97] F. Diamond, The Taylor-Wiles construction and multiplicity one. Invent. Math. **128**(2), 379–391 (1997)

[Gee11] T. Gee, Automorphic lifts of prescribed types. Math. Ann. **350**(1), 107–144 (2011)

[GM92] F. Gouvêa, B. Mazur, Families of modular eigenforms. Math. Comput. **58**(198), 793–805 (1992)

[Gou01] F.Q. Gouvêa, Where the slopes are? J. Ramanujan Math. Soc. **16**(1), 75–99 (2001)

[HLTT13] M. Harris, K.-W. Lan, R. Taylor, J. Thorne, On the rigid cohomology of certain Shimura varieties. Preprint (2013)

[HSBT10] M. Harris, N. Shepherd-Barron, R. Taylor, A family of Calabi-Yau varieties and potential automorphy. Ann. Math. (2) **171**(2), 779–813 (2010)

[HT13] Y. Hu, F. Tan, The Breuil–Mézard conjecture for non-scalar split residual representations. arXiv preprint. arXiv:1309.1658 (2013)

[Kil08] L.J.P. Kilford, On the slopes of the U_5 operator acting on overconvergent modular forms. J. Théor. Nombres Bordeaux **20** (1), 165–182 (2008)

[Kis09] M. Kisin, The Fontaine-Mazur conjecture for GL_2. J. Am. Math. Soc. **22**(3), 641–690 (2009)

[KM12] L.J.P. Kilford, K. McMurdy, Slopes of the U_7 operator acting on a space of overconvergent modular forms. LMS J. Comput. Math. **15**, 113–139 (2012)

[KW09] C. Khare, J.-P. Wintenberger, On Serre's conjecture for 2-dimensional mod p representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Ann. Math. (2) **169**(1), 229–253 (2009)

[Loe08] D. Loeffler, Explicit calculations of automorphic forms for definite unitary groups. LMS J. Comput. Math. **11**, 326–342 (2008)

[LWX14] R. Liu, D. Wan, L. Xiao, Eigencurve over the boundary of the weight space (2014). https://arxiv.org/abs/1412.2584

[Nak14] K. Nakamura, Zariski density of crystalline representations for any p-adic field. J. Math. Sci. Univ. Tokyo **21**(1), 79–127 (2014)

[Paš15] V. Paškūnas, On the Breuil–Mézard conjecture. Duke Math. J. **164**(2), 297–359 (2015)

[Roe14] D. Roe, The 3-adic eigencurve at the boundary of weight space. Int. J. Number Theory **10**(7), 1791–1806 (2014)

[Sai97] T. Saito, Modular forms and p-adic Hodge theory. Invent. Math. **129**(3), 607–620 (1997)

[Shi12] S.W. Shin, Automorphic Plancherel density theorem. Isr. J. Math. **192**(1), 83–120 (2012)

[ST12] S.W. Shin, N. Templier, Sato-Tate Theorem for Families and low-lying zeros of automorphic L-functions. With Appendix 1 by R. Kottwitz and Appendix 2 by R. Cluckers, J. Gordon and I. Halupczok. Invent. Math. **203**(1), 1–177 (2016)

[Tay08] R. Taylor, Automorphy for some l-adic lifts of automorphic mod l Galois representations. II. Publ. Math. Inst. Hautes Études Sci. **108**, 183–239 (2008)

[Var14] I. Varma, Local-global compatibility for regular algebraic cuspidal automorphic representation when $\ell \neq p$ (2014). arXiv:1411.2520

[Wan98] D. Wan, Dimension variation of classical and p-adic modular forms. Invent. Math. **133**(2), 449–463 (1998)

[WXZ14] D. Wan, L. Xiao, J. Zhang, Slopes of eigencurves over boundary disks (2014). https://arxiv.org/abs/1407.0279

Transfer Principles for Bounds of Motivic Exponential Functions

Raf Cluckers, Julia Gordon, and Immanuel Halupczok

Abstract We study transfer principles for upper bounds of motivic exponential functions and for linear combinations of such functions, directly generalizing the transfer principles from Cluckers and Loeser (Ann Math 171:1011–1065, 2010) and Shin and Templier (Invent Math, 2015, Appendix B). These functions come from rather general oscillatory integrals on local fields, and can be used to describe, e.g., Fourier transforms of orbital integrals. One of our techniques consists in reducing to simpler functions where the oscillation only comes from the residue field.

2000 *Mathematics Subject Classification.* Primary 14E18; Secondary 22E50, 40J99

1 Introduction

After recalling concrete motivic exponential functions and their stability under taking integral transformations, we study transfer principles for bounds of motivic exponential functions and their linear combinations. In this context, *transfer* means switching between local fields with isomorphic residue field (in particular between positive and mixed characteristic). By the word *concrete* (in the first sentence),

R. Cluckers (✉)
Université de Lille, Laboratoire Painlevé, CNRS - UMR 8524, Cité Scientifique, 59655 Villeneuve d'Ascq Cedex, France

Department of Mathematics, KU Leuven, Celestijnenlaan 200B, B-3001 Leuven, Belgium
e-mail: Raf.Cluckers@math.univ-lille1.fr
http://math.univ-lille1.fr/~cluckers

J. Gordon
Department of Mathematics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2
e-mail: gor@math.ubc.ca
http://www.math.ubc.ca/~gor

I. Halupczok
School of Mathematics, University of Leeds, Leeds LS2 9JT, UK
e-mail: math@karimmi.de
http://www.immi.karimmi.de/en.math.html

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_3

we mean that we work uniformly in all local fields of large enough residue field characteristic, as opposed to genuinely motivic as done in [CLe]; this setting is perfectly suited for transfer principles, which are, indeed, about local fields.

Our results relate to previously known transfer principles (from [CLe, CGH], and [ShTe, Appendix B]) as follows. The principle given by Theorem 3.1 below, which allows to transfer bounds on motivic exponential functions, generalizes both the transfer principle of [CLe, Proposition 9.2.1], where, one can say, the upper bound was identically zero, and the transfer principle of [ShTe, Theorem B.7], where the case without oscillation is treated. A generalization to $\mathscr{C}^{\mathrm{exp}}$ (instead of $\mathscr{C}$) of Theorem B.6 of [ShTe] (which contains a statement about uniformity across all completions of a given number field rather than a transfer principle) is left to future work in [CGH5], since it requires different, and deeper, proof techniques.

The results in this paper are independent of the transfer principles of [CGH] about e.g., loci of integrability, and in fact, our proofs are closer to the ones of [CLe], and can avoid the heavier machinery from [CGH].

After Theorem 3.1, we give some further generalizations which treat $\mathbb{C}$-linear combinations of motivic exponential functions, uniformly in the complex scalars. Specifically, we obtain transfer principles for linear (in-)dependence and for upper bounds of linear combinations of motivic exponential functions (or rather, their specializations for any local field F with large residue field characteristic), see Theorem 3.2, Proposition 3.3 and Corollary 3.4.

A key proof technique that we share with [CLe] consists in reducing from general motivic exponential functions to simpler functions where the oscillation only comes from additive characters on the residue field. We recall these classes of functions with their respective oscillatory behavior in Sect. 2.

Let us finally mention that the transfer principles of [CLe] have been applied in [CHL] and [YGo, Appendix] to obtain the Fundamental Lemma of the Langlands program in characteristic zero (see also [Nad]), and the ones of [CGH] have been used in [CGH2] to show local integrability of Harish-Chandra characters in large positive characteristic. The results of this paper may apply to a wide class of p-adic integrals, e.g. orbital integrals and their Fourier transforms. We will leave the study of such applications to future work.

2 Motivic Exponential Functions

In a nutshell, motivic functions are a natural class of functions from (subsets of) valued fields to $\mathbb{C}$, built from functions on the valued fields that are definable in the Denef-Pas language; the class is closed under integration. Motivic *exponential* functions are a bigger such class, incorporating additive characters of the valued field. These functions were introduced in [CLe], and the strongest form of stability under integration for these functions was proved in [CGH]. (Constructible functions without oscillation and on a fixed p-adic field were introduced earlier by Denef in [Den1].) We start with recalling three classes of functions, $\mathscr{C}$, $\mathscr{C}^{\mathrm{e}}$, and $\mathscr{C}^{\mathrm{exp}}$, which have, so to speak, increasing oscillatory richness, and each one is stable under integration, see Theorem 2.8.

2.1 Motivic Functions

We recall some terminology of [CLo] and [CLe], with the same focus as in [CGH3] (namely uniform in the local field, as opposed to an approach with Grothendieck rings).

Fix a ring of integers Ω of a number field, as base ring.

Definition 2.1. Let Loc_Ω be the collection of all triples (F, ι, ϖ), where F is a non-Archimedean local field which allows at least one ring homomorphism from Ω to F, the map $\iota : \Omega \to F$ is such a ring homomorphism, and ϖ is a uniformizer for the valuation ring of F. Here, by a non-Archimedean local field we mean a finite extension of $\mathbb{Q}_p$ or $\mathbb{F}_p((t))$ for any prime p.

Given an integer M, let $\mathrm{Loc}_{\Omega,M}$ be the collection of (F, ι, ϖ) in Loc_Ω such that the residue field of F has characteristic at least M.

For a non-Archimedean local field F, write $\mathcal{O}_F$ for its valuation ring with maximal ideal $\mathcal{M}_F$ and residue field k_F with q_F elements.

We will use the Denef-Pas language with coefficients from $\Omega[[t]]$ for our fixed ring of integers Ω. We denote this language by $\mathcal{L}_\Omega$.

Definition 2.2. The language $\mathcal{L}_\Omega$ has three sorts, VF for the valued field, RF for the residue field, and a sort for the value group which we simply call $\mathbb{Z}$, since we will only consider structures where it is actually equal to $\mathbb{Z}$. On VF, one has the ring language and coefficients from the ring $\Omega[[t]]$. On RF, one has the ring language. On $\mathbb{Z}$, one has the Presburger language, namely the language of ordered abelian groups together with constant symbols 0, 1, and symbols $\equiv_n$ for each $n > 0$ for the congruence relation modulo n. Finally, one has the symbols ord for the valuation map from the valued field minus 0 to $\mathbb{Z}$, and $\overline{\mathrm{ac}}$ for an angular component map from the valued field to the residue field.

It was an important insight of Denef that one has elimination of valued field quantifiers for first order formulas in this language $\mathcal{L}_\Omega$, and this was worked out by his student Pas in [Pas]. Indeed, quantifier elimination is a first step to understanding the geometry of the definable sets and functions. Another geometrical key result and insight by Denef [Den2, Pas, CLo] is the so-called cell decomposition, which is behind Proposition 4.4.

The language $\mathcal{L}_\Omega$ is interpreted in any (F, ι, ϖ) in Loc_Ω in the obvious way, where t is interpreted as ϖ and where $\overline{\mathrm{ac}}$ is defined by

$$\overline{\mathrm{ac}}(u\varpi^\ell) = \bar{u} \text{ and } \overline{\mathrm{ac}}(0) = 0$$

for any $u \in \mathcal{O}_F^\times$ and $\ell \in \mathbb{Z}$, $\bar{u}$ being reduction modulo $\mathcal{M}_F$. We will abuse notation by notationally identifying F and $(F, \iota, \varpi) \in \mathrm{Loc}_\Omega$.

Any $\mathcal{L}_\Omega$-formula φ gives a subset $\varphi(F)$ of $F^n \times k_F^m \times \mathbb{Z}^r$ for $F \in \mathrm{Loc}_\Omega$ for some n, m, r only depending on φ, by taking the F-rational points on φ in the sense of model theory (see Sect. 2.1 of [CGH3] for more explanation). This leads us to the following handy definition.

Definition 2.3. A collection $X = (X_F)_{F \in \mathrm{Loc}_{\Omega,M}}$ of subsets $X_F \subset F^n \times k_F^m \times \mathbb{Z}^r$ for some M, n, m, r is called a *definable set* if there is an $\mathcal{L}_\Omega$-formula φ such that $X_F = \varphi(F)$ for each F in $\mathrm{Loc}_{\Omega,M}$ (see Remark 2.5).

By Definition 2.3, a "definable set" is actually a collection of sets indexed by $F \in \mathrm{Loc}_{\Omega,M}$; such practice is often used in model theory and also in algebraic geometry. A particularly simple definable set is $(F^n \times k_F^m \times \mathbb{Z}^r)_F$, for which we use the simplified notation $\mathrm{VF}^n \times \mathrm{RF}^m \times \mathbb{Z}^r$. We apply the typical set-theoretical notation to definable sets X, Y, e.g., $X \subset Y$ (if $X_F \subset Y_F$ for each $F \in \mathrm{Loc}_{\Omega,M}$ for some M), $X \times Y$, and so on, which may increase M if necessary.

Definition 2.4. For definable sets X and Y, a collection $f = (f_F)_F$ of functions $f_F : X_F \to Y_F$ for $F \in \mathrm{Loc}_{\Omega,M}$ for some M is called a definable function and denoted by $f : X \to Y$ if the collection of graphs of the f_F is a definable set.

Remark 2.5. For a definable set X as in Definition 2.3, we are usually only interested in $(X_F)_{F \in \mathrm{Loc}_{\Omega,M}}$ for M sufficiently big, and thus, we often allow ourselves to replace M by a larger number if necessary, without saying so explicitly; also the uniform objects defined below in Definitions 2.6, 2.7, and so on, are only interesting for M sufficiently large. In model theoretic terms, we are using the theory of all non-archimedean local fields, together with, for each $M > 0$, an axiom stating that the residue characteristic is at least M. Note, however, that a more general theory of uniform integration which works uniformly in all local fields of mixed characteristic (but not in local fields of small positive characteristic) is under development in [CHa] and will generalize [CLb].

For motivic functions, definable functions are the building blocks, as follows.

Definition 2.6. Let $X = (X_F)_{F \in \mathrm{Loc}_{\Omega,M}}$ be a definable set. A collection $H = (H_F)_F$ of functions $H_F : X_F \to \mathbb{R}$ is called *a motivic function* on X if there exist integers N, N', and N'', nonzero integers $a_{i\ell}$, definable functions $\alpha_i : X \to \mathbb{Z}$ and $\beta_{ij} : X \to \mathbb{Z}$, and definable sets $Y_i \subset X \times \mathrm{RF}^{r_i}$ such that for all $F \in \mathrm{Loc}_{\Omega,M}$ and all $x \in X_F$

$$H_F(x) = \sum_{i=1}^{N} \#Y_{i,F,x} \cdot q_F^{\alpha_{iF}(x)} \cdot \left(\prod_{j=1}^{N'} \beta_{ijF}(x) \right) \cdot \left(\prod_{\ell=1}^{N''} \frac{1}{1 - q_F^{a_{i\ell}}} \right),$$

where $Y_{i,F,x}$ is the finite set $\{y \in k_F^{r_i} \mid (x, y) \in Y_{i,F}\}$.

We write $\mathscr{C}(X)$ to denote the ring of motivic functions on X.

The precise form of this definition is motivated by the property that motivic functions behave well under integration (see Theorem 2.8).

2.2 Motivic Exponential Functions

For any local field F, let $\mathcal{D}_F$ be the set of the additive characters ψ on F that are trivial on the maximal ideal $\mathcal{M}_F$ of $\mathcal{O}_F$, nontrivial on $\mathcal{O}_F$, and such that, for $x \in \mathcal{O}_F$, one has

$$\psi(x) = \mathbf{e}(\mathrm{Tr}_{k_F/\mathbb{F}_p}(\bar{x})) \tag{1}$$

with $\bar{x}$ the reduction of x modulo $\mathcal{M}_F$ and where q_F is an integer power of the prime number p, and where $\mathbf{e} : \mathbb{F}_p \to \mathbb{C}$ sends $a \in \{0, \dots, p-1\}$ to $\exp(\frac{2\pi i a}{p})$ for some fixed complex square root i of -1. Expressions involving additive characters of p-adic fields often give rise to exponential sums, and this explains the term "exponential" in the definition below.

Definition 2.7. Let $X = (X_F)_{F \in \mathrm{Loc}_{\Omega,M}}$ be a definable set. A collection $H = (H_{F,\psi})_{F,\psi}$ of functions $H_{F,\psi} : X_F \to \mathbb{C}$ for $F \in \mathrm{Loc}_{\Omega,M}$ and $\psi \in \mathcal{D}_F$ is called *a motivic exponential function* on X if there exist integers $N > 0$ and $r_i \geq 0$, motivic functions $H_i = (H_{iF})_F$ on X, definable sets $Y_i \subset X \times \mathrm{RF}^{r_i}$ and definable functions $g_i : Y_i \to \mathrm{VF}$ and $e_i : Y_i \to \mathrm{RF}$ for $i = 1, \dots, N$, such that for all $F \in \mathrm{Loc}_{\Omega,M}$, all $\psi \in \mathcal{D}_F$ and all $x \in X_F$

$$H_{F,\psi}(x) = \sum_{i=1}^{N} H_{iF}(x) \left(\sum_{y \in Y_{i,F,x}} \psi\big(g_{iF}(x,y) + e_{iF}(x,y)\big) \right), \tag{2}$$

where $\psi(a + v)$ for $a \in F$ and $v \in k_F$, by abuse of notation, is defined as $\psi(a+u)$, with u any unit in $\mathcal{O}_F$ such that $\bar{u} = v$, which is well defined by (1). We write $\mathscr{C}^{\exp}(X)$ to denote the ring of motivic exponential functions on X. Define the subring $\mathscr{C}^{\mathrm{e}}(X)$ of $\mathscr{C}^{\exp}(X)$ consisting of those functions H as in (2) such that all g_{iF} are identically vanishing. Note that for $H \in \mathscr{C}^{\mathrm{e}}(X)$, $H_{F,\psi}$ does not depend on $\psi \in \mathcal{D}_F$ because of (1), so we will just write H_F instead.

Compared to Definition 2.6, the counting operation # has been replaced by taking exponential sums, which makes the motivic exponential functions a richer class than the motivic functions. Indeed, note that the sum as above gives just $\#(Y_{iF})_x$ in the case that $g_{iF} = 0$ and $e_{iF} = 0$.

2.3 Integration

To integrate a motivic function f on a definable set X, we need a uniformly given family of measures on each X_F. For $X = \mathrm{VF}$, we put the Haar measure on $X_F = F$ so that $\mathcal{O}_F$ has measure 1; on k_F and on $\mathbb{Z}$, we use the counting measure and for $X \subset \mathrm{VF}^n \times \mathrm{RF}^m \times \mathbb{Z}^r$ we use the measure on X_F induced by the product measure on

$F^n \times k_F^m \times \mathbb{Z}^r$. To obtain other motivic measures on definable sets X, one can also use measures associated with "definable volume forms," see Sect. 2.5 of [CGH3], [CLo, Sect. 8], and Sect. 12 of [CLb].

Maybe the most important aspect of these motivic functions is that they have nice and natural properties related to integration, see, e.g., the following theorem about stability, which generalizes Theorem 9.1.4 of [CLe] (see also Theorem 4.1.1 of [CLe]).

Theorem 2.8 ([CGH, Theorem 4.3.1]). *Let f be in $\mathscr{C}(X \times Y)$, resp. in $\mathscr{C}^{\mathrm{e}}(X \times Y)$ or in $\mathscr{C}^{\exp}(X \times Y)$, for some definable sets X and Y, with Y equipped with a motivic measure μ_Y. Then there exist a function I in $\mathscr{C}(X)$, resp. in $\mathscr{C}^{\mathrm{e}}(X)$ or $\mathscr{C}^{\exp}(X)$ and an integer $M > 0$ such that for each $F \in \mathrm{Loc}_{\Omega,M}$, each $\psi \in \mathcal{D}_F$ and for each $x \in X_F$ one has*

$$I_F(x) = \int_{y \in Y_F} f_F(x, y)\, d\mu_{Y_F}, \text{ resp. } I_{F,\psi}(x) = \int_{y \in Y_F} f_{F,\psi}(x, y)\, d\mu_{Y_F},$$

whenever the function $Y_F \to \mathbb{C} : y \mapsto f_F(x, y)$, resp. $y \mapsto f_{F,\psi}(x, y)$, is in L^1.

Proof. The cases $\mathscr{C}$ and $\mathscr{C}^{\exp}$ are treated in [CGH, Theorems 4.3.1 and 4.4.3]. The proof for $\mathscr{C}^{\exp}$ in [CGH] goes through also for $\mathscr{C}^{\mathrm{e}}$. (A more direct and simpler proof for $\mathscr{C}^{\mathrm{e}}$ can also be given, by reducing to the case for $\mathscr{C}$ using residual parameterizations as in Definition 4.5.1 of [CGH].) □

3 Transfer Principles for Bounds and Linear Combinations

In this section, we state the main results of this article.

The following statement allows one to transfer bounds which are known for local fields of characteristic zero to local fields of positive characteristic, and vice versa.

Theorem 3.1 (Transfer Principle for Bounds). *Let X be a definable set, let H be in $\mathscr{C}^{\exp}(X)$, and let G be in $\mathscr{C}^{\mathrm{e}}(X)$. Then there exist M and N such that, for any $F \in \mathrm{Loc}_{\Omega,M}$, the following holds. If*

$$|H_{F,\psi}(x)|_{\mathbb{C}} \leq |G_F(x)|_{\mathbb{C}} \text{ for all } (\psi, x) \in \mathcal{D}_F \times X_F \tag{3}$$

then, for any local field F' with the same residue field as F, one has

$$|H_{F',\psi}(x)|_{\mathbb{C}} \leq N \cdot |G_{F'}(x)|_{\mathbb{C}} \text{ for all } (\psi, x) \in \mathcal{D}_{F'} \times X_{F'}. \tag{4}$$

Moreover, one can take $N = 1$ if H lies in $\mathscr{C}^{\mathrm{e}}(X)$.

As mentioned in the introduction, the case where $G = 0$ is [CLe, Proposition 9.2.1], and the case that both H and G lie in $\mathscr{C}(X)$ is [ShTe, Theorem B.7].

We also show the following strengthening of Theorem 3.1, for linear combinations.

Theorem 3.2 (Transfer Principle for Bounds of Linear Combinations). *Let X be a definable set, let H_i be in $\mathscr{C}^{\exp}(X)$ for $i = 1 \ldots, \ell$, and let G be in $\mathscr{C}^{\mathrm{e}}(X)$. Then there exist M and N such that, for any $F \in \mathrm{Loc}_{\Omega,M}$, the following holds for any $c = (c_i)_i$ in $\mathbb{C}^\ell$. If*

$$|\sum_{i=1}^{\ell} c_i H_{i,F,\psi}(x)|_{\mathbb{C}} \leq |G_F(x)|_{\mathbb{C}} \text{ for all } (\psi, x) \in \mathcal{D}_F \times X_F \tag{5}$$

then, for any local field F' with the same residue field as F, one has

$$|\sum_{i=1}^{\ell} c_i H_{i,F',\psi}(x)|_{\mathbb{C}} \leq N \cdot |G_{F'}(x)|_{\mathbb{C}} \text{ for all } (\psi, x) \in \mathcal{D}_{F'} \times X_{F'}. \tag{6}$$

Moreover, one can take $N = 1$ if the H_i lie in $\mathscr{C}^{\mathrm{e}}(X)$.

The key improvement of Theorem 3.2 (compared to Theorem 3.1) is that the choice of M and N works uniformly in c.

Although in our proofs, the integer N of Theorems 3.1 and 3.2 appears naturally, it is not unconceivable that one can take N close to 1 even when H does not lie in $\mathscr{C}^{\mathrm{e}}(X)$.

The following proposition was motivated by the application to the transfer of linear (in-)dependence of Shalika germs in [GoHa]. The first part of the proposition gives a transfer principle for linear (in-)dependence of motivic exponential functions, which is deduced quite directly from transfer of identical vanishing of motivic functions (the $G = 0$ case of Theorem 3.1). The second part describes how the coefficients in a linear relation can depend on the local field, the additive character, and parameters; see below for more explanation.

Proposition 3.3 (Transfer Principle for Linear Dependence). *Let X and Y be definable sets and let H_i be in $\mathscr{C}^{\exp}(X \times Y)$ for $i = 1 \ldots, \ell$.*

(1) *There exists M such that, for any $F, F' \in \mathrm{Loc}_{\Omega,M}$ with $k_F \cong k_{F'}$, the following holds:*

If for each $\psi \in \mathcal{D}_F$ and each $y \in Y_F$ the functions $H_{i,F,\psi}(\cdot, y) : X_F \to \mathbb{C}$ for $i = 1, \ldots, \ell$ are linearly dependent, then, also for each $\psi \in \mathcal{D}_{F'}$ and each $y \in Y_{F'}$, the functions $H_{i,F',\psi}(\cdot, y)$ are linearly dependent.

(2) *Let moreover G be in $\mathscr{C}^{\exp}(X \times Y)$. Then there exists a definable set W and functions C_i and D in $\mathscr{C}^{\exp}(W \times Y)$ such that the following holds for M sufficiently big. For every $F \in \mathrm{Loc}_{\Omega,M}$, for every $\psi \in \mathcal{D}_F$, and for every $y \in Y_F$, if the functions $H_{i,F,\psi}(\cdot, y)$ (on X_F) are linearly independent and*

$$G_{F,\psi}(\cdot, y) = \sum_{i=1}^{\ell} c_i H_{i,F,\psi}(\cdot, y)$$

for some $c_i \in \mathbb{C}$*, then* $D_{F,\psi}(\cdot, y)$ *is not identically zero on* W_F*, and for all* $w \in W_F$

$$D_{F,\psi}(w, y)c_i = C_{i,F,\psi}(w, y).$$

The second part of the proposition, essentially, states that the coefficients c_i are ratios of motivic exponential functions. However, our proof needs an additional parameter w to write the c_i as ratios: both, C_i and D depend on w and only their quotient is independent of w, for w with $D_{F,\psi}(w, y)$ nonzero. Note that despite this complication, the proposition permits to apply transfer principles to the constants c_i. (One of course does not need w if there is a definable function $h: Y \to W$ such that $D_{F,\psi} \circ h_F$ is nowhere zero.)

Proposition 3.3 naturally applies also in case that the H_i and G are in $\mathscr{C}^{\exp}(Z)$ for some definable subset Z of $X \times Y$, instead of in $\mathscr{C}^{\exp}(X \times Y)$. Indeed, one can extend the H_i by zero outside Z and apply the proposition to these extensions.

Finally, we note the following corollary of Theorem 3.2, showing that the complex coefficients of a linear relation between motivic exponential functions stay *the same* (regardless of their motivic interpretation as in Proposition 3.3 above) in situations where these coefficients are independent of the additive character. This independence is a strong assumption, but note that it in particular applies to arbitrary linear relations of motivic (non-exponential) functions.

Corollary 3.4 (Transfer Principle for Coefficients of Linear Relations). *Let* X *be a definable set and let* H_i *be in* $\mathscr{C}^{\exp}(X)$ *for* $i = 1 \ldots, \ell$*. Then there exists* M *such that, for any* $F \in \mathrm{Loc}_{\Omega,M}$*, the following holds for any* $c = (c_i)_i$ *in* $\mathbb{C}^\ell$*.*

If

$$\sum_{i=1}^{\ell} c_i H_{i,F,\psi} = 0 \text{ on } X_F \text{ for all } \psi \in \mathcal{D}_F,$$

then, for any $F' \in \mathrm{Loc}_{\Omega,M}$ *with* $k_F \cong k_{F'}$*, one also has*

$$\sum_{i=1}^{\ell} c_i H_{i,F',\psi'} = 0 \text{ on } X_{F'} \text{ for all } \psi' \in \mathcal{D}_{F'}.$$

Proof. Just apply Theorem 3.2 with $G = 0$. □

4 Proofs of the Transfer Principles

Before proving Theorem 3.1, we give a proposition relating the square of the complex modulus of a motivic exponential function to the complex modulus of a function where the oscillation only comes from the residue field.

Proposition 4.1. *Let H be in $\mathscr{C}^{\exp}(X)$ for some definable set X. Then there exist $\tilde{H}$ in $\mathscr{C}^{\mathrm{e}}(X)$ and integers M and N such that for all $F \in \mathrm{Loc}_{\Omega,M}$ the following hold for all $x \in X_F$.*

(1) *There is ψ_1 in $\mathcal{D}_F$ (depending on x) such that*

$$\frac{1}{N}|\tilde{H}_F(x)|_{\mathbb{C}} \leq |H_{F,\psi_1}(x)|^2_{\mathbb{C}}.$$

(2) *For all ψ in $\mathcal{D}_F$, one has*

$$|H_{F,\psi}(x)|^2_{\mathbb{C}} \leq |\tilde{H}_F(x)|_{\mathbb{C}}.$$

The proof of Proposition 4.1 uses an elementary result from Fourier analysis, which we now recall.

Lemma 4.2. *Consider a finite abelian group G with dual group $\hat{G}$ and with $|G|$ elements. For any function $f : G \to \mathbb{C}$ one has*

$$\frac{1}{|G|}\|\hat{f}\|_{\sup} \leq \|f\|_{\sup} \leq \|\hat{f}\|_{\sup}$$

where $\|\cdot\|_{\sup}$ is the supremum norm and $\hat{f}$ the Fourier transform of f, namely

$$\hat{f}(\varphi) = \sum_{x\in G} f(x)\varphi(x) \text{ for } \varphi \in \hat{G}.$$

Proof. Clearly one has

$$\|f\|_{\sup} \leq \|f\|_2 \leq \sqrt{|G|}\|f\|_{\sup},$$

and similarly for $\hat{f}$, where $\|\cdot\|_2$ is the L_2-norm, namely $\|f\|_2 = \sqrt{\sum_{g\in G}|f(g)|^2_{\mathbb{C}}}$. By Plancherel identity one has

$$\sqrt{|G|}\|f\|_2 = \|\hat{f}\|_2.$$

The lemma follows. □

Corollary 4.3. *Consider a finite abelian group G with dual group $\hat{G}$. Consider a function*

$$f : \hat{G} \to \mathbb{C} : \varphi \mapsto \sum_{j=1}^{s} c_j\varphi(y_j)$$

for some complex numbers c_j and some distinct $y_j \in G$. Then there exists $\varphi_0 \in \hat{G}$ with

$$\sup_{1\le j\le s} |c_j|_{\mathbb{C}} \le |f(\varphi_0)|_{\mathbb{C}}.$$

Note that Corollary 4.3 generalizes Lemma 9.2.3 of [CLe], a basic ingredient for proving the transfer principle [CLe, Proposition 9.2.1].

We will use the simple fact that for n complex numbers a_i, one has

$$\sum_{i=1}^{n} |a_i|_{\mathbb{C}}^2 \le (\sum_{i=1}^{n} |a_i|_{\mathbb{C}})^2 \le n \cdot \sum_{i=1}^{n} |a_i|_{\mathbb{C}}^2. \tag{7}$$

Proof of Proposition 4.1. Recall that we allow ourselves to increase M whenever necessary without further mentioning. By "for every F" we shall always mean for $F \in \mathrm{Loc}_{\Omega,M}$.

Consider a general H in $\mathscr{C}^{\exp}(X)$ and write it as in (2):

$$H_{F,\psi}(x) = \sum_i H_{iF}(x)\Big(\sum_{y\in Y_{i,F,x}} \psi\big(g_{iF}(x,y) + e_{iF}(x,y)\big)\Big). \tag{8}$$

We will start by grouping the summands of the sum over y according to the value of $g_{iF}(x,y)$ modulo $\mathcal{O}_F$. This is done as follows. For each $x \in X_F$, the union of the images $A_{F,x} := \bigcup_i g_{iF}(Y_{i,F,x})$ is finite. Therefore, the cardinality $\#A_{F,x}$ is bounded by some $N' > 0$ (independently of x and F), and by cell decomposition (in the form of Theorem 7.2.1 of [CLo]), there exists a definable set $X' \subset X \times \mathrm{RF}^t$ (for some $t \ge 0$) and a definable function $g' : X' \to \mathrm{VF}$ inducing a bijection $X'_{F,x} \to A_{F,x}$ for every F and x (where $X'_{F,x}$ is the fiber of X'_F over $x \in X_F$). This allows us to write H as

$$H_{F,\psi}(x) = \sum_{x'\in X'_{F,x}} \psi(g'_F(x'))H'_F(x'). \tag{9}$$

for a suitable $H' \in \mathscr{C}^{\mathrm{e}}(X')$; indeed, we can take H' such that

$$H'_F(x') = \sum_i H_{iF}(\pi(x')) \sum_{\substack{y\in Y_{i,F,x}\\ g_{iF}(x,y)=g'_F(x')\\ \pi(x')=x}} \psi\big(e_{iF}(x,y)\big),$$

where $\pi : X' \to X$ is the projection and with notation as in (2) concerning $\psi(\xi)$ for $\xi \in k_F$, which does not depend on ψ since it is fixed by (1).

This construction ensures that for $x', x'' \in X'_{F,x}$ with $x' \ne x''$, we have $g'_F(x') \ne g'_F(x'')$. We can even achieve that for such x', x'' we have

$$\mathrm{ord}(g'_F(x') - g'_F(x'')) < 0, \tag{10}$$

by modifying g_{iF} and e_{iF} in (8) in such a way that $g'_F(x') \neq g'_F(x'')$ already implies (10). To this end, replace $g_{iF}(x, y)$ by the arithmetic mean of the (finite) set $A_{F,x} \cap (g_{iF}(x, y) + \mathcal{O}_F)$ and change e_{iF} , using the additivity of ψ, to make up for this modification.

Let G' be a function in $\mathscr{C}^{\mathrm{e}}(X')$ such that for all F,

$$G'_F = |H'_F|^2_{\mathbb{C}}. \tag{11}$$

Such G' exists by multiplying (uniformly in F) H'_F with its complex conjugate which is constructed by replacing the arguments (appearing in H') of the additive character on the residue field by their additive inverses, similarly to the proof of Lemma 4.5.9 of [CGH]. Now define $\tilde{H}$ such that

$$\tilde{H}_F(x) = N' \cdot \sum_{x',\ \pi_F(x')=x} G'_F(x') \tag{12}$$

for each F and each $x \in X_F$, and let N be N'^2. We claim that $\tilde{H}$ and N are as desired. Firstly, $\tilde{H}$ lies in $\mathscr{C}^{\mathrm{e}}(X)$ by Theorem 2.8. From (7), (9), and (11) it follows that

$$|H_{F,\psi}(x)|^2_{\mathbb{C}} \leq |\tilde{H}_F(x)|_{\mathbb{C}} \text{ for all } (\psi, x) \text{ in } \mathcal{D}_F \times X_F.$$

We now show that for each $x \in X_F$ there is ψ_1 in $\mathcal{D}_F$ such that

$$\frac{1}{N}|\tilde{H}_F(x)|_{\mathbb{C}} \leq |H_{F,\psi_1}(x)|^2_{\mathbb{C}}. \tag{13}$$

Fix F and $x \in X_F$. From Corollary 4.3, applied to a large enough finite subgroup G of $F/\mathcal{O}_F$ so that G contains $g'_F(x')$ mod $\mathcal{O}_F$ for all x' with $\pi(x') = x$, one finds ψ_1 in $\mathcal{D}_F$ such that

$$\sup_{x',\ \pi_F(x')=x} |H'_F(x')|_{\mathbb{C}} \leq |H_{F,\psi_1}(x)|_{\mathbb{C}}.$$

Hence, from (7) again,

$$\sum_{x',\ \pi_F(x')=x} |H'_F(x')|^2_{\mathbb{C}} \leq N'|H_{F,\psi_1}(x)|^2_{\mathbb{C}},$$

and thus

$$\frac{1}{N}\tilde{H}_F(x) = \frac{N'}{N} \sum_{x',\ \pi_F(x')=x} |H'_F(x')|^2_{\mathbb{C}} \leq \frac{N'^2}{N}|H_{F,\psi_1}(x)|^2_{\mathbb{C}} = |H_{F,\psi_1}(x)|^2_{\mathbb{C}}.$$

This shows (13). □

We will also use the following generalization of Proposition B.8 of the appendix B of [ShTe]. Intuitively, it says that functions in $\mathscr{C}^{\mathrm{e}}(S)$ (for arbitrary definable S) only depend on value group and residue field information.

Proposition 4.4. *Let H be in $\mathscr{C}^{\mathrm{e}}(S \times B)$ for some definable sets S and B. Then there exist a definable function $f : S \times B \to \mathrm{RF}^m \times \mathbb{Z}^r \times B$ for some $m \geq 0$ and $r \geq 0$, which makes a commutative diagram with both projections to B, and a function G in $\mathscr{C}^{\mathrm{e}}(\mathrm{RF}^m \times \mathbb{Z}^r \times B)$ such that, for some M and all F in $\mathrm{Loc}_{\Omega,M}$, the function H_F equals the function $G_F \circ f_F$, and such that G_F vanishes outside the range of f_F.*

Proof. The proof is similar to the one for Proposition B.8 in Appendix B of [ShTe]. Let us write $S \subset \mathrm{VF}^n \times \mathrm{RF}^a \times \mathbb{Z}^b$ for some integers n, a and b. It is enough to prove the lemma when $n = 1$ by a finite recursion argument. The case $n = 1$ follows from the Cell Decomposition Theorem 7.2.1 from [CLo]. Indeed, this result can be used to push the domains of all appearing definable functions in the build-up of H into a set of the form $\mathrm{RF}^m \times \mathbb{Z}^r$, forcing them to have only residue field variables and value group variables. □

Proof of Theorem 3.1. By Proposition 4.1 it is enough to consider the case that H lies in $\mathscr{C}^{\mathrm{e}}(X)$ and to show that one can take $N = 1$ in this case. Suppose that X is a definable subset of $\mathrm{VF}^n \times \mathrm{RF}^m \times \mathbb{Z}^r$. In the case that $n = 0$, the proof goes as follows. By quantifier elimination, any finite set of formulas needed to describe H and G can be taken to be without valued field quantifiers. It follows that

$$H_{F_1} = H_{F_2} \text{ and } G_{F_1} = G_{F_2} \tag{14}$$

for F_1 and F_2 in $\mathrm{Loc}_{\Omega,M}$ with $k_{F_1} \cong k_{F_2}$ and M large enough, and up to identifying k_{F_1} with k_{F_2}. This implies the case $n = 0$ with $N = 1$.

Now assume $n > 0$. By Proposition 4.4, there is a definable function

$$f : X \to \mathrm{RF}^{m'} \times \mathbb{Z}^{r'} \tag{15}$$

for some m', r', and $\tilde{H} \in \mathscr{C}^{\mathrm{e}}(\mathrm{RF}^{m'} \times \mathbb{Z}^{r'})$ and $\tilde{G} \in \mathscr{C}^{\mathrm{e}}(\mathrm{RF}^{m'} \times \mathbb{Z}^{r'})$, such that $H = \tilde{H} \circ f$ and $G = \tilde{G} \circ f$ and such that $\tilde{H}$ and $\tilde{G}$ vanish outside the range of f. We finish the case of H in $\mathscr{C}^{\mathrm{e}}(X)$ by applying the case $n = 0$ to $\tilde{H}$ and $\tilde{G}$. □

In order to prove Theorem 3.2, we will need the corresponding strengthening of Proposition 4.1, which goes as follows.

Proposition 4.5. *Let H_i be in $\mathscr{C}^{\exp}(X)$ for some definable set X and for $i = 1, \ldots, \ell$ for some $\ell > 0$. Then there exist integers M and N, and functions $\tilde{H}_{i,s}$ in $\mathscr{C}^{\mathrm{e}}(X)$ for $i, s = 1, \ldots, \ell$, such that for all $F \in \mathrm{Loc}_{\Omega,M}$ the following conditions hold for all $x \in X_F$ and all $c = (c_i)_i$ in $\mathbb{C}^\ell$.*

(1) *There is ψ_1 in $\mathcal{D}_F$ (depending on x and c) such that*

$$\frac{1}{N}\Big|\sum_{i,s=1}^{\ell} c_i \bar{c}_s \tilde{H}_{i,s,F}(x)\Big|_{\mathbb{C}} \leq \Big|\sum_i c_i H_{i,F,\psi_1}(x)\Big|_{\mathbb{C}}^2.$$

(2) *For all ψ in $\mathcal{D}_F$, one has*

$$\left|\sum_i c_i H_{i,F,\psi}(x)\right|_{\mathbb{C}}^2 \leq \left|\sum_{i,s=1}^{\ell} c_i \bar{c}_s \tilde{H}_{i,s,F}(x)\right|_{\mathbb{C}}.$$

Proof. We start by applying the construction from the beginning of the proof of Proposition 4.1 to each of our functions $H_{i,F,\psi}$, i.e., we write each of them in the form

$$H_{i,F,\psi}(x) = \sum_{x' \in X'_{F,x}} \psi(g'_F(x')) H'_{i,F}(x'), \tag{16}$$

where $X' \subset X \times \mathrm{RF}^t$ has finite fibers $X'_{F,x}$ which are bounded uniformly in $x \in X_F$ and in F, H'_i lies in $\mathscr{C}^{\mathrm{e}}(X')$, $g' : X' \to \mathrm{VF}$ is definable, and such that

$$\mathrm{ord}(g'_F(x') - g'_F(x'')) < 0 \tag{17}$$

for any $x', x'' \in X'_{F,x}$ with $x' \neq x''$.

We can do this in such a way that neither X' nor g' depends on i. Indeed, first do the construction for each $H_{i,F,\psi}$ separately, yielding sets X'_i and functions g'_i. Then let $X' := \dot{\bigcup}_i X'_i$ be the disjoint union, set $g'_F(x') := g'_{i,F}(x')$ if $x' \in X'_i$ and extend $H'_{i,F}(x')$ from X'_i to X' by 0. Finally, note that the same construction as in the proof of Proposition 4.1 allows us to assume that (17) holds on the whole of X'.

Let $G'_{i,s}$ be functions in $\mathscr{C}^{\mathrm{e}}(X')$ such that

$$\sum_{i,s=1}^{\ell} c_i \bar{c}_s G'_{i,s,F}(x') = |\sum_{i=1}^{\ell} c_i H'_{i,F}(x')|_{\mathbb{C}}^2. \tag{18}$$

(for all F and all $x' \in X'_F$). Such $G'_{i,s}$ exist by a similar argument to the one explained for G' in the proof of Proposition 4.1. Now use Theorem 2.8 for each i to define $\tilde{H}_{i,s}$ in $\in \mathscr{C}^{\mathrm{e}}(X)$ satisfying

$$\tilde{H}_{i,s,F}(x) = N' \cdot \sum_{x' \in X'_{F,x}} G'_{i,s,F}(x'), \tag{19}$$

where $N' \in \mathbb{N}$ is some constant which we will fix later.

We claim that for a suitable choice of N', the functions $\tilde{H}_{i,s}$ are as desired. Indeed, we have the following, where the relations "$\approx_1$" and "$\approx_2$" are explained below.

$$\left|\sum_i c_i H_{i,F,\psi}(x)\right|_{\mathbb{C}}^2 \overset{(16)}{=} \left|\sum_{x' \in X'_{F,x}} \psi(g'_F(x')) \sum_i c_i H'_{i,F}(x')\right|_{\mathbb{C}}^2$$

$$\approx_1 \left| \sum_{x' \in X'_{F,x}} \left| \sum_i c_i H'_{i,F}(x') \right| \right|_{\mathbb{C}}^2$$

$$\approx_2 \sum_{x' \in X'_{F,x}} \left| \sum_i c_i H'_{i,F}(x') \right|_{\mathbb{C}}^2$$

$$\overset{(18)}{\underset{(19)}{=}} \frac{1}{N'} \left| \sum_{i,s=1}^{\ell} c_i \bar{c}_s \tilde{H}_{i,s,I,F}(x) \right|_{\mathbb{C}}$$

The meaning of the symbol "$\approx_2$" is the following. For the left-hand side L and the right-hand side R of "$\approx_2$", there is a constant c such that $L \le cR$ and $R \le cL$ by the simple fact (7) and since the sets $X'_{F,x}$ are finite sets which are bounded uniformly in $x \in X_F$ and F. At "$\approx_1$", we have "$\le$", which already implies (2) of the proposition for a suitable choice of N', and we obtain an estimate in the other direction in the same way as in the proof of Proposition 4.1: By Corollary 4.3, and using (17), for each F and each x, there exists a $\psi_1 \in \mathcal{D}_F$ such that

$$\left| \sum_{x' \in X'_{F,x}} \psi_1(g'_F(x')) \sum_i c_i H'_{i,F}(x') \right|_{\mathbb{C}} \ge \sup_{x' \in X'_{F,x}} \left| \sum_i c_i H'_{i,F}(x') \right|;$$

now use once more that the cardinality of $X'_{F,x}$ is uniformly bounded to replace the supremum over x' by the sum, and to obtain (1) of the Proposition. □

Proof of Theorem 3.2. By Proposition 4.5 it is enough to consider the case that the H_i lie in $\mathscr{C}^{\mathrm{e}}(X)$ and to show that one can take $N = 1$ in this case. But this case is proved as the proof for the corresponding case of Theorem 3.1. □

It remains to prove Proposition 3.3. We do this by reducing to the transfer principle of [CLe, Proposition 9.2.1]. The main ingredient for this reduction is the following classical result, which shows that a finite collection of functions being linearly dependent is equivalent to some other function that can be constructed from this collection being constantly zero.

Lemma 4.6. *Let f_i be complex-valued functions on some set A for $i = 1, \dots, n$. Then there exists nonzero $c = (c_i)_{i=1}^n$ in $\mathbb{C}^n$ such that the function $\sum_{i=1}^n c_i f_i$ is identically vanishing on A if and only if the determinant of the matrix*

$$(f_i(z_j))_{i,j}$$

is identically vanishing on A^n, where the z_j are distinct variables, running over A for $j = 1, \dots, n$.

Proof. The implication "⇒" is easy, so let us assume that the given determinant is identically vanishing on A^n. Choose as many points $z_1, \dots, z_r$ in A as possible such that the rows

$$(f_1(z_1), \dots, f_n(z_1))$$
$$\vdots$$
$$(f_1(z_r), \dots, f_n(z_r))$$

are linearly independent. By the assumption on the determinant D, we have $r < n$, hence there exists a linear dependence between the columns, i.e., there are complex numbers $a_1, \dots, a_n$, not all zero, such that

$$a_1 f_1(z_j) + \cdots + a_n f_n(z_j) = 0 \tag{20}$$

for every $j \leq r$.

Now we claim that this implies

$$\sum a_i f_i = 0 \text{ on } A, \tag{21}$$

with a_i as in (20). To verify this, choose any other point z in A. By the choice of $z_1, \dots, z_r$, the row

$$(f_1(z), \dots, f_n(z))$$

can be written as a linear combination of the rows

$$(f_1(z_j), \dots, f_n(z_j)).$$

This implies that (20) also holds for

$$(f_1(z), \dots, f_n(z)),$$

but this implies (21). □

Proof of Proposition 3.3. (1) Consider the function D in $\mathscr{C}^{\exp}(X^\ell \times Y)$ given by

$$D_{F,\psi}(x_1, \dots, x_\ell, y) = \det((H_{i,F,\psi}(x_j, y))_{ij}).$$

For each F, ψ and y, by Lemma 4.6, $D_{F,\psi}(\cdot, y)$ is identically zero on X_F^ℓ iff the $H_{i,F,\psi}(\cdot, y)$ for $i = 1, \dots, \ell$ are linearly dependent. Thus the statement we want to transfer is that $D_{F,\psi}$ is identically zero on $X_F^\ell \times Y_F$ for all ψ. This follows from [CLe, Proposition 9.2.1] (which is the case of Theorem 3.1 with $G = 0$).

(2) Set $W := X^\ell$ and define D in $\mathscr{C}^{\exp}(W \times Y)$ as in (1).

Consider F, ψ, $w = (x_1, \ldots, x_\ell)$, y such that $d := D_{F,\psi}(w, y) \neq 0$. Then there exist unique $c_1, \ldots, c_\ell \in \mathbb{C}$ such that

$$G_{F,\psi}(x_j, y) = \sum_i c_i H_{i,F,\psi}(x_j, y) \qquad \text{for } 1 \leq j \leq \ell. \tag{22}$$

By Cramer's rule, the products $c_i \cdot d$ are polynomials in $G_{F,\psi}(x_j, y)$ and $H_{i,F,\psi}(x_j, y)$, so there exist functions C_i in $\mathscr{C}^{\exp}(W \times Y)$ such that $c_i = C_{i,F,\psi}(w, y)/D_{F,\psi}(w, y)$. These C_i (and this D) are as required: As noted in the proof of (1), if F, ψ and y are such that the $H_{i,F,\psi}(\cdot, y)$ are linearly independent, then there exists a $w \in W_F$ such that $D_{F,\psi}(w, y) \neq 0$, and if $G_{F,\psi}(\cdot, y)$ is a linear combination of the $H_{i,F,\psi}(\cdot, y)$, then for such a w, the coefficients c_i from (22) are the desired ones. □

Acknowledgements The authors are grateful to T. Hales, F. Loeser, and J. Denef whose influence on the subject of motivic integration is and has always been very important. Special thanks to T. Hales for suggesting to look at a statement like Proposition 3.3.

The author Raf Cluckers was supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013) with ERC Grant Agreement nr. 615722 MOTMELSUM, by the Labex CEMPI (ANR-11-LABX-0007-01), and would like to thank both the Forschungsinstitut für Mathematik (FIM) at ETH Zürich and the IHÉS for the hospitality during part of the writing of this paper. Julia Gordon was supported by NSERC, and Immanuel Halupczok was supported by the SFB 878 of the Deutsche Forschungsgemeinschaft.

References

[CGH] R. Cluckers, J. Gordon, I. Halupczok, Integrability of oscillatory functions on local fields: transfer principles. Duke Math. J. **163**(8), 1549–1600 (2014)

[CGH2] R. Cluckers, J. Gordon, I. Halupczok, Local integrability results in harmonic analysis on reductive groups in large positive characteristic. Ann. Sci. École Norm. Sup. (4) **47**(6), 1163–1195 (2014)

[CGH3] R. Cluckers, J. Gordon, I. Halupczok, Motivic functions, integrability, and uniform in p bounds for orbital integrals. Electron. Res. Announc. Math. **21**, 137–152 (2014)

[CGH5] R. Cluckers, J. Gordon, I. Halupczok, *Uniform Analysis on Local Fields and Applications to Orbital Integrals* (2016)

[CHa] R. Cluckers, I. Halupczok, Integration, uniform in all fields of p-adic numbers (2015), preprint available at arXiv:1510.08250

[CHL] R. Cluckers, T. Hales, F. Loeser, Transfer principle for the fundamental lemma, in *Stabilisation de la formule des traces, variétés de Shimura, et applications arithmétiques*, vol. I, ed. by L. Clozel, M. Harris, J.-P. Labesse, B.-C. Ngô (International Press, Boston, 2011)

[CLb] R. Cluckers, F. Loeser, Motivic integration in all residue field characteristics for Henselian discretely valued fields of characteristic zero. J. Reine Angew. Math. **701**, 1–31 (2015)

[CLe] R. Cluckers, F. Loeser, Constructible exponential functions, motivic Fourier transform and transfer principle. Ann. Math. **171**, 1011–1065 (2010)

[CLo] R. Cluckers, F. Loeser, Constructible motivic functions and motivic integration. Invent. Math. **173**(1), 23–121 (2008)

[Den1] J. Denef, *Arithmetic and Geometric Applications of Quantifier Elimination for Valued Fields*. MSRI Publications, vol. 39 (Cambridge University Press, Cambridge, 2000), pp. 173–198

[Den2] J. Denef, p-adic semialgebraic sets and cell decomposition. J. Reine Angew. Math. **369**, 154–166 (1986)

[GoHa] J. Gordon, T. Hales, Endoscopic transfer of orbital integrals in large residual characteristic. Am. J. Math. **138**(1) (2016)

[Nad] D. Nadler, The geometric nature of the fundamental lemma. Bull. Am. Math. Soc. **49**(1), 1–50 (2012)

[Pas] J. Pas, Uniform p-adic cell decomposition and local zeta functions. J. Reine Angew. Math. **399**, 137–172 (1989)

[ShTe] S.-W. Shin, N. Templier, Sato-Tate theorem for families and low-lying zeros of automorphic l-functions, with Appendix A by R. Kottwitz and Appendix B by R. Cluckers, J. Gordon and I. Halupczok. Invent. Math. (2015). doi:10.1007/s00222–015-0583-y

[YGo] Z. Yun, The fundamental lemma of Jacquet and Rallis. Duke Math. J. **156**(2), 167–227 (2011). Appendix by J. Gordon

Growth of Hecke Fields Along a p-Adic Analytic Family of Modular Forms

Haruzo Hida

Abstract Fix a nearly ordinary non-CM p-adic analytic family of Hilbert modular Hecke eigenforms (over a totally real field F). We prove existence of a density one set Ξ of primes of the field F such that the degree of the field over $\mathbb{Q}(\mu_{p^\infty})$ generated by the Hecke eigenvalue of the Hecke operator $T(\mathfrak{l})$ grows indefinitely over the family for each prime $\mathfrak{l}$ in the set Ξ.

2010 *Mathematics Subject Classification.* Primary 11E25, 11F41, 11F80; Secondary 11F27, 11F33, 11E16

We generalize in this paper all the principal results obtained in [H14] for the one variable cyclotomic p-ordinary Hecke algebra to the full nearly p-ordinary Hecke algebra of fixed central character. This algebra is finite flat over the m variable Iwasawa algebra for the degree m totally real base field F. The restriction coming from fixing a central character is essentially harmless as we can bring one central character to another by character twists (up to finite order character of bounded order).

Take the field $\overline{\mathbb{Q}}$ of all numbers in $\mathbb{C}$ algebraic over $\mathbb{Q}$. Fix a prime p and a field embedding $\overline{\mathbb{Q}} \overset{i_p}{\hookrightarrow} \overline{\mathbb{Q}}_p \subset \mathbb{C}_p$. Fix a totally real number field F (of degree m over $\mathbb{Q}$) inside $\overline{\mathbb{Q}}$ with integer ring O (as the base field for Hilbert modular forms). We use the symbol O exclusively for the integer ring of F, and for a general number field L, we write O_L for the integer ring of L. We choose and fix an O-ideal $\mathfrak{n}$ prime to p (as the level of modular form). We define an algebraic group G (resp. T_L) by $\mathrm{Res}_{O/\mathbb{Z}}\mathrm{GL}(2)$ (resp. $\mathrm{Res}_{O_L/\mathbb{Z}}\mathbb{G}_m$); so, $G(R) = \mathrm{GL}_2(R \otimes_{\mathbb{Z}} O)$ and $T_L(R) = (R \otimes_{\mathbb{Z}} O_L)^\times$. We write $T_F^\Delta \cong T_F^2$ for the diagonal torus of G; so, writing T^Δ for the diagonal torus of $\mathrm{GL}(2)_{/O}$, $T_F^\Delta = \mathrm{Res}_{O/\mathbb{Z}} T^\Delta$.

Let $S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$ denote the space of weight κ adelic Hilbert cusp forms $\mathbf{f} : G(\mathbb{Q})\backslash G(\mathbb{A}) \to \mathbb{C}$ of level $\mathfrak{n}$ with character ϵ modulo $\mathfrak{n}$, where $\mathfrak{n}$ is a non-zero ideal of O. Here the weight $\kappa = (\kappa_1, \kappa_2)$ is the Hodge weight of the rank 2 pure

H. Hida (✉)
Department of Mathematics, UCLA, Los Angeles, CA 90095-1555, USA
e-mail: hida@math.ucla.edu

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_4

motive $M(\mathbf{f})$ with coefficient in the Hecke field $\mathbb{Q}(\mathbf{f})$ associated with any Hecke eigenform $\mathbf{f} \in S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$ (see [BR93]). Though $M(\mathbf{f})$ is possibly defined over a quadratic extension F' of F (depending on $\mathbf{f}$), the Hodge weight is well defined over F independent of the infinity places over a given place of F. For each field embedding $\sigma : F \hookrightarrow \overline{\mathbb{Q}}$, taking an extension $\tilde{\sigma}$ of σ to F', $M(\mathbf{f}) \otimes_{F', \iota_\infty \circ \tilde{\sigma}} \mathbb{C}$ has Hodge weights $(\kappa_{1,\sigma}, \kappa_{2,\sigma})$ and $(\kappa_{2,\sigma}, \kappa_{1,\sigma})$, and the motivic weight $[\kappa] := \kappa_{1,\sigma} + \kappa_{2,\sigma}$ is independent of σ. We normalize the weight imposing an inequality $\kappa_{1,\sigma} \leq \kappa_{2,\sigma}$. This normalization is the one in [HMI, (SA1–3)]. Writing I (resp. I_p) for the set of all field embeddings into $\overline{\mathbb{Q}}$ (resp. p-adic places) of F, we identify κ_j with $\sum_{\sigma \in I} \kappa_{j,\sigma}\sigma \in \mathbb{Z}[I]$. Sometimes we identify I_p and I regarding I_p as a set of p-adic places induced by $i_p \circ \sigma$ for $\sigma \in I$. Often we use I to denote $\sum_\sigma \sigma \in \mathbb{Z}[I]$. If the Hodge weight is given by $\kappa = (0, kI)$ for an integer $k \geq 1$, traditionally, the integer $k+1$ is called the weight [of the cusp forms in $S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$] at all σ, but we use here the Hodge weight κ.

The "Neben character" ϵ we use is again not a traditional one (but the one introduced in [HMI]). It is a set of three characters $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_+)$, where $\epsilon_+ : F_{\mathbb{A}}^\times / F^\times \to \mathbb{C}^\times$ is the central character of the automorphic representation $\pi_{\mathbf{f}}$ of $G(\mathbb{A})$ generated by any Hecke eigenform $0 \neq \mathbf{f} \in S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$. The character ϵ_+ has infinity type $I - \kappa_1 - \kappa_2$, and therefore its finite part has values in $\overline{\mathbb{Q}}^\times$. The finite order characters ϵ_j are $\overline{\mathbb{Q}}$-valued continuous characters of $\hat{O}^\times = \varprojlim_{0<N\in\mathbb{Z}} (O/NO)^\times$ with $\epsilon_1\epsilon_2 = \epsilon_+|_{\hat{O}^\times}$. These characters ϵ_j ($j = 1, 2$) factor through $(O/\mathfrak{N})^\times$ for an integral ideal $\mathfrak{N}$. The two given data $\{\epsilon_1, \epsilon_2\}$ are purely local and may not extend to Hecke characters of the idele class group $F_{\mathbb{A}}^\times / F^\times$. Put $\epsilon^- := \epsilon_1\epsilon_2^{-1}$, and we assume that ϵ^- factors through $(O/\mathfrak{n})^\times$; so, the conductor of ϵ^- is a factor of $\mathfrak{n}$ and $\mathfrak{N}$ (which could be a proper factor of $\mathfrak{n}$). Then for the level group

$$U = U_0(\mathfrak{n}) = \{u = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\widehat{\mathbb{Z}}) \text{ with } c \in \hat{\mathfrak{n}} = \mathfrak{n}\widehat{O}\},$$

we have $\mathbf{f}(gu) = \epsilon(u)\mathbf{f}(g)$ for all $g \in G(\mathbb{A})$ and $u \in U$, where

$$\epsilon(u) = \epsilon_2(\det(u))\epsilon^-(a_{\mathfrak{n}}) = \epsilon_1(\det(u))(\epsilon^-)^{-1}(d_{\mathfrak{n}})$$

for the projection $d_{\mathfrak{n}}$ of d to $\prod_{\mathfrak{l}|\mathfrak{n}} F_{\mathfrak{l}}$. The characters ϵ_j for $j = 1, 2$ factor through $(O/\mathfrak{n}_j)^\times$ for some multiple $\mathfrak{n}_j$ of $\mathfrak{n}$ but we do not insist on $\mathfrak{n} = \mathfrak{n}_j$. As long as the local component $\pi_{\mathfrak{l}}$ of $\pi_{\mathbf{f}}$ at a prime $\mathfrak{l}|\mathfrak{n}_j|\mathfrak{N}$ is principal of the form $\pi(\alpha, \beta)$ or Steinberg of the form $\sigma(\alpha, \beta)$, we choose the data so that $\{\epsilon_1, \epsilon_2\} = \{\alpha|_{O_{\mathfrak{l}}^\times}, \beta|_{O_{\mathfrak{l}}^\times}\}$ (see [H89, Sect. 2]). In other words, for a suitable choice of (ϵ_1, ϵ_2), we have a unique minimal form $\mathbf{f}^\circ \in S_\kappa(\mathfrak{n}^\circ, \epsilon; \mathbb{C})$ in $\pi_{\mathbf{f}}$ with minimal level $\mathfrak{n}^\circ|\mathfrak{n}$. This minimal level $\mathfrak{n}^\circ$ of $\pi_{\mathbf{f}}$ is a factor of the conductor of $\pi_{\mathbf{f}}$ but could be a **proper** factor of it. These minimal forms are $\mathfrak{p}$-adically interpolated (the interpolation property is not always true for new forms). A detailed description of cusp forms in $S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$ will be recalled in Sect. 1.9 from [HMI].

Throughout the paper, $\mathfrak{n}$ denotes an O-ideal prime to p, and we work with cusp forms of (minimal) level $\mathfrak{n}p^{r+I_p}$ ($r = \sum_{\mathfrak{p}\in I_p} r_{\mathfrak{p}}\mathfrak{p} \in \mathbb{Z}[I_p]$ with $r_{\mathfrak{p}} \geq 0$ and $p^{r+I_p} = \prod_{\mathfrak{p}|p} \mathfrak{p}^{r_{\mathfrak{p}}+1}$, symbolically). Extend ϵ_j to a character of the finite adele group $(F_{\mathbb{A}}^{(\infty)})^{\times}$ (trivial outside the level $\mathfrak{n}_j$ and trivial at a choice of uniformizer $\varpi_{\mathfrak{l}}$ at each prime $\mathfrak{l}$), and extend the character ϵ of U to the semi-group

$$\Delta_0(\mathfrak{n}) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\mathbb{A}^{(\infty)}) \cap M_2(\widehat{O}) \middle| d\widehat{O} + \hat{\mathfrak{n}} = \widehat{O},\ c \in \hat{\mathfrak{n}}\right\}$$

by $\epsilon\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \epsilon_1(ad - bc)(\epsilon^-)^{-1}(d_{\mathfrak{n}})$. The Hecke operator $T(y)$ of the double coset $U\left(\begin{smallmatrix} y & 0 \\ 0 & 1 \end{smallmatrix}\right)U = \bigsqcup_\delta \delta U$ is defined by $\mathbf{f}|T(y)(g) = \sum_\delta \epsilon(\delta)^{-1}\mathbf{f}(g\delta)$ [see (14)]. For a Hecke eigenform $\mathbf{f}$, the eigenvalue $a(y, \mathbf{f})$ of $T(y)$ depends only on the ideal $\mathfrak{y} = y\widehat{O} \cap F$ [see (19)]; so, for each prime $\mathfrak{l}$ of F, we write $a(\mathfrak{l}, \mathbf{f})$ for $a(\varpi_{\mathfrak{l}}, \mathbf{f})$ and put $T(\mathfrak{l}) := T(\varpi_{\mathfrak{l}})$, choosing a prime element $\varpi_{\mathfrak{l}}$ of the $\mathfrak{l}$-adic completion $O_{\mathfrak{l}}$. Therefore the yth Fourier coefficient $c(y, \mathbf{f})$ of $\mathbf{f}$ is $\epsilon_1(y)a(y, \mathbf{f})$ for each Hecke eigenform $\mathbf{f}$ normalized so that $c(1, \mathbf{f}) = 1$, and the Fourier coefficient depends on y (if $\epsilon_1 \neq 1$) not just on the ideal $\mathfrak{y}$. For $\mathfrak{l}|\mathfrak{n}p$, we often write $U(\mathfrak{l})$ for $T(\mathfrak{l})$. For a Hecke eigenform $\mathbf{f} \in S_\kappa(\mathfrak{n}p^{r+I_p}, \epsilon; \mathbb{C})$ ($p \nmid \mathfrak{n}$) and a subfield H of $\overline{\mathbb{Q}}$, the Hecke field $H(\mathbf{f})$ inside $\mathbb{C}$ is generated over H by the eigenvalues $a(\mathfrak{l}, \mathbf{f})$ of $\mathbf{f}$ for the Hecke operators $T(\mathfrak{l})$ for all prime ideals $\mathfrak{l}$ and the values of ϵ over finite ideles. If $H \subset \overline{\mathbb{Q}}$, then $H(\mathbf{f})$ is a finite extension of H sitting inside $\overline{\mathbb{Q}}$.

Let W be a sufficiently large discrete valuation ring flat over $\mathbb{Z}_p$. Let $\mathbf{\Gamma} \cong \mathbb{Z}_p^m$ ($m = [F : \mathbb{Q}]$) be the maximal torsion-free quotient of $O_p^\times$ for $O_p = O \otimes_{\mathbb{Z}} \mathbb{Z}_p$. We use this symbol $\mathbf{\Gamma}$ exclusively for the base totally real field F. Later in Sect. 1.12, for a CM quadratic extension M/F, we write $\mathbf{\Gamma}_M$ for the maximal p-profinite torsion free quotient of the anti-cyclotomic quotient of the ray class group $Cl_M(p^\infty) = \varprojlim_n Cl_M(p^n)$ of M modulo p^∞ (i.e., the projective limit of the ray class group $Cl_M(p^n)$ modulo p^n). Here the anti-cyclotomic quotient is the maximal quotient on which the generator c of $\mathrm{Gal}(M/F)$ acts by -1. Note that we have a natural inclusion: $\mathbf{\Gamma} \to \mathbf{\Gamma}_M$ but it could have finite cokernel. We fix once and for all a splitting of the projection: $O_p^\times \twoheadrightarrow \mathbf{\Gamma}$ and decompose $O_p^\times = \mathbf{\Gamma} \times \Delta$ for a finite group Δ.

We fix a $\mathbb{Z}_p$-basis $\{\gamma_j\}_{j=1,\dots,m} \subset \mathbf{\Gamma}$ so that $\mathbf{\Gamma} = \prod_j \gamma_j^{\mathbb{Z}_p}$ and identify the Iwasawa algebra $\mathbf{\Lambda} = \mathbf{\Lambda}_W := W[[\mathbf{\Gamma}]]$ with the power series ring $W[[T]]$ ($T = \{T_j\}_{j=1,\dots,m}$) by $\mathbf{\Gamma} \ni \gamma_j \mapsto t_j := (1 + T_j) \in \mathbf{\Lambda}$. We have $W[[T]] = \varprojlim_n W[t, t^{-1}]/(t^{p^n} - 1)$, where $t = (t_j)_j$, $t^{-1} = (t_j^{-1})_j$ and $(t^{p^n} - 1)$ is the ideal $(t_1^{p^n} - 1, \dots, t_m^{p^n} - 1)$ in $W[[T]]$. In this way, we identify the formal spectrum $\mathrm{Spf}(\mathbf{\Lambda})$ with $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ for $\mathbf{\Gamma}^* = \mathrm{Hom}_{\mathbb{Z}_p}(\mathbf{\Gamma}, \mathbb{Z}_p)$, as t_j giving the character of $\mathbf{\Gamma}^*$ corresponding $t_j(\gamma_i^*) = \delta_{ij}$ for the dual basis $\{\gamma_j^*\}_j$ of $\{\gamma_j\}_j$. Here $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ sends a local p-profinite ring R to the p-profinite group $(1 + \mathfrak{m}_R) \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ as a group functor (for the maximal ideal $\mathfrak{m}_R$ of R).

A p-adic nearly ordinary analytic family of eigenforms $\mathcal{F} = \{\mathbf{f}_P | P \in \mathrm{Spec}(\mathbb{I})(\mathbb{C}_p)\}$ is indexed by points of $\mathrm{Spec}(\mathbb{I})(\mathbb{C}_p)$, where $\mathrm{Spec}(\mathbb{I})$ is an irreducible component of the spectrum of the big nearly ordinary Hecke algebra $\mathbf{h}$ and is a torsion-free domain of finite rank over $\mathbf{\Lambda}$ (in this sense, we call $\mathrm{Spec}(\mathbb{I})$ a finite torsion-free covering of $\mathrm{Spec}(\mathbf{\Lambda})$). For each $P \in \mathrm{Spec}(\mathbb{I})(\mathbb{C}_p)$, $\mathbf{f}_P$ is a p-adic Hecke eigenform of slope 0 and level $\mathfrak{n}p^\infty$ for a fixed prime to p-level $\mathfrak{n}$. The family is called analytic because $P \mapsto a(y, \mathbf{f}_P)$ is a p-adic analytic function on the rigid analytic space associated with the formal spectrum $\mathrm{Spf}(\mathbb{I})$ in the sense of Berthelot (cf. [dJ95, Sect. 7], see also [dJ98]). We call $P \in \mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$ *arithmetic* of weight $\kappa = \kappa(P) \in \mathbb{Z}[I]^2$ with character $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_+)$ if $\kappa_{2,\sigma} - \kappa_{1,\sigma} \geq 1$ for all $\sigma \in I$ (we write this condition as $\kappa_2 - \kappa_1 \geq I$), $\epsilon_2|_{\mathbf{\Gamma}}$ has values in $\mu_{p^\infty}(\overline{\mathbb{Q}}_p)$ and $P(t_j - \epsilon_2^{-1}(\gamma_j)\gamma_j^{\kappa_2}) = 0$ for all j (regarding P as a W-algebra homomorphism $P : \mathbb{I} \to \overline{\mathbb{Q}}_p$). Here $\gamma^k = \prod_{\sigma \in I} \sigma(\gamma)^{k_\sigma}$ for $\gamma \in O_p$ and $k = \sum_{\sigma \in I} k_\sigma \sigma$, and $k \geq I$ means $k_\sigma \geq 1$ for all $\sigma \in I$. If P is arithmetic, $\mathbf{f}_P$ is a classical p-stabilized Hecke eigenform (not just a p-adic modular form). In order to make the introduction succinct, we put off, to Sect. 1.9, recalling the theory of analytic families of eigenforms including the definition and necessary properties of CM families. We only remark that each universal nearly ordinary family comes from an irreducible component $\mathrm{Spec}(\mathbb{I})$ of the spectrum $\mathrm{Spec}(\mathbf{h})$ of the big nearly ordinary Hecke algebra $\mathbf{h}$, and we assume now that $\mathrm{Spec}(\mathbb{I})$ is one of such irreducible components.

In this paper, we prove the following theorem.

Theorem. *Let* $\mathrm{Spec}(\mathbb{I})$ *be an irreducible (reduced) component of* $\mathrm{Spec}(\mathbf{h})$ *and* $K = \mathbb{Q}(\mu_{p^\infty})$. *Then* $\mathbb{I}$ *is a non-CM component if there exists a prime* $\mathfrak{l}$ *of* F *and an infinite set* $\mathcal{A}$ *of arithmetic points in* $\mathrm{Spec}(\mathbb{I})$ *of a fixed weight* κ *with* $\kappa_2 - \kappa_1 \geq I$ *such that*

$$\limsup_{P \in \mathcal{A}} [K(a(\mathfrak{l}, \mathbf{f}_P)) : K] = \infty.$$

Indeed, if $\mathbb{I}$ *is a CM component, the degree* $[K(\mathbf{f}_P) : K]$ *is bounded independently of arithmetic* P. *Conversely, if* $\mathbb{I}$ *is a non-CM component, there exists a set of primes* Ξ *of* F *with Dirichlet density one such that for any infinite set* $\mathcal{A}$ *of arithmetic points in* $\mathrm{Spec}(\mathbb{I})$ *of a fixed weight* κ *with* $\kappa_2 - \kappa_1 \geq I$, *we have*

$$\limsup_{P \in \mathcal{A}} [K(a(\mathfrak{l}, \mathbf{f}_P)) : K] = \infty \text{ for each } \mathfrak{l} \in \Xi.$$

In particular, for any bound $B > 0$, *the set of arithmetic primes* P *of given weight* κ *in* $\mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$ *with* $[K(\mathbf{f}_P) : K] < B$ *is finite for a non-CM component* $\mathbb{I}$.

The first assertion and the boundedness of $[K(\mathbf{f}_P) : K]$ (for a CM component $\mathbb{I}$) independently of arithmetic P follow from the construction of CM families in Sects. 1.12 and 1.13 (see [H11, Corollary 4.2] for the argument for $F = \mathbb{Q}$ which holds without modification for general F). We prove in this paper a slightly stronger statement than the converse in the theorem. The formulation of Theorem 3.1 is a bit different from the above theorem asserting that boundedness of $[K(a(\mathfrak{l}, \mathbf{f}_P)) : K]$ $(P \in \mathcal{A})$ over $\mathfrak{l} \in \Sigma$ implies that $\mathbb{I}$ is a CM component as long as Σ has positive upper density.

We could have written the assertion of the theorem as $\lim_{P\in\mathcal{A}}[K(a(\mathfrak{l},\mathbf{f}_P)):K]=\infty$ for the "limit" with respect to the filter of $\mathcal{A}$ given by the complement of all finite subsets of $\mathcal{A}$ instead of $\limsup_{P\in\mathcal{A}}[K(a(\mathfrak{l},\mathbf{f}_P)) : K]=\infty$. Hereafter we use this filter and use $\lim_{P\in\mathcal{A}}$ instead of $\limsup_{P\in\mathcal{A}}$. In [H11] (and [H13]), we proved a similar result for $K[a(\mathfrak{p},\mathbf{f}_P)]$ for $\mathfrak{p}|p$. Here the point is to study the same phenomena for $a(\mathfrak{l},\mathbf{f}_P)$ for $\mathfrak{l}$ outside $\mathfrak{n}p$. Indeed, we proved in [H14] the above fact replacing the nearly ordinary Hecke algebra by the smaller cyclotomic ordinary Hecke algebra of one variable. The many variable rigidity lemma (see Lemma 4.1) enabled us to extend our result in [H14] to the many variable setting here. We expect that, assuming $\kappa_2-\kappa_1\geq I$,

$$\lim_{P\in\mathcal{A}}[K(a(\mathfrak{l},\mathbf{f}_P)):K]=\infty \textit{ for any single } \mathfrak{l}\nmid\mathfrak{n}p \textit{ if } \mathbb{I} \textit{ is a non CM component}$$

as in the case of $\mathfrak{p}|p$ (see Conjecture 3.5). As in [H11], the proof of the above theorem is based on the elementary finiteness of Weil l-numbers of given weight in any extension of $\mathbb{Q}(\mu_{p^\infty})$ of bounded degree up to multiplication by roots of unity and rigidity lemmas (in Sect. 4) asserting that a geometrically connected formal subscheme in a formal split torus stable under the (central) action $t\mapsto t^z$ of z in an open subgroup of $\mathbb{Z}_p^\times$ is a formal subtorus. Another key tool is the determination by Rajan [Rj98] of compatible systems by trace of Frobeniai for primes of positive density (up to character twists).

Infinite growth of the absolute degree of Hecke fields (under different assumptions) was proven in [Se97] for growing level N, and Serre's analytic method is now generalized to (almost) an optimal form to automorphic representations of classical groups by Shin and Templier [ST13]). Our proof is purely algebraic, and the degree we study is over the infinite cyclotomic field $\mathbb{Q}[\mu_{p^\infty}]$ (while the above papers use non-elementary analytic trace formulas and Plancherel measures in representation theory). Our result applies to any thin infinite set $\mathcal{A}$ of slope 0 non-CM cusp forms, while in [Se97] and [ST13], they studied the set of all automorphic representations of given infinity type (and given central character), growing the level. Note here the Zariski closure of $\mathcal{A}$ could be a transcendental formal subscheme of $\widehat{\mathbb{G}}_m\otimes\mathbf{\Gamma}^*$ relative to the rational structure coming from T_F and could have the smallest positive dimension 1, while $\dim\widehat{\mathbb{G}}_m\otimes\mathbf{\Gamma}^*=m=[F:\mathbb{Q}]$. Another distinction from earlier works is that we are now able to prove that the entire $\mathbb{I}$ has CM if the degrees $[K(\mathbf{f}_P) : K]$ are bounded only over arithmetic points P of a possibly very small closed subscheme in $\mathrm{Spf}(\mathbb{I})$.

To state transcendence results of Hecke operators, let L/F be a finite field extension inside $\mathbb{C}_p$ with integer ring O_L, and look into the torus $T_L=\mathrm{Res}_{O_L/\mathbb{Z}}\mathbb{G}_m$. Write $\mathbb{Z}_{(p)}=\mathbb{Q}\cap\mathbb{Z}_p$ and $O_{L,(p)}=O_L\otimes_{\mathbb{Z}}\mathbb{Z}_{(p)}\subset L^\times$. Consider an algebraic homomorphism $\nu\in\mathrm{Hom}_{\text{gp scheme}}(T_L,T_F)$. We regard $\nu : T_L(\mathbb{Z}_p)=O_{L,p}^\times=(O_L\otimes_{\mathbb{Z}}\mathbb{Z}_p)^\times\to T_F(\overline{\mathbb{Q}}_p)\supset T_F(\mathbb{Z}_p)=O_p^\times$. Project $\nu(T_L(\mathbb{Z}_p))\cap T_F(\mathbb{Z}_p)\subset O_p^\times$ to the maximal torsion free quotient $\mathbf{\Gamma}$ of $O_p^\times$. As an example of $\mathbb{Q}_p$-rational ν (so, $\nu(O_{L,(p)})\subset T_F(\mathbb{Z}_p)=O_p^\times$), we have the norm character $N_{L/\mathbb{Q}}$ or, if L is a CM field with a p-adic CM type Φ (in the sense of [HT93]), $\nu:(L\otimes_{\mathbb{Q}}\mathbb{Q}_p)^\times\to\mathbb{Q}_p^\times$ given by

$\nu(\xi) = \prod_{\varphi\in\Phi} \xi^{\varphi}$. Define an integral domain $R = R_\nu$ to be the subalgebra of $\mathbf{\Lambda}_{\mathbb{Z}_p}$ generated over $\mathbb{Z}_{(p)}$ by the projected image G of $\nu(T_L(\mathbb{Z}_{(p)})) \cap O_p^\times$ in $\mathbf{\Gamma}$. Note that for any $\xi \in R_\nu$ and any arithmetic point P, $P(\xi) = \xi_P$ is in $L^{\mathrm{gal}}(\mu_N, \mu_{p^\infty})$ for the Galois closure L^{gal} of $L/\mathbb{Q}$ and for a sufficiently large $0 < N \in \mathbb{Z}$ for which μ_N receives all the values of characters of Δ (e.g., $N = |\Delta|$). The field $L^{\mathrm{gal}}(\mu_N, \mu_{p^\infty})$ is a finite extension of $\mathbb{Q}(\mu_{p^\infty})$. For an integral domain A, write $Q(A)$ for the quotient field of A. By definition, R_ν is isomorphic to the group algebra $\mathbb{Z}_{(p)}[G]$ of the torsion-free group G. Unless $G = \{1\}$, the quotient field $Q(R_\nu)$ has infinite transcendental degree over $\mathbb{Q}$.

If the family (associated with $\mathbb{I}$) contains a theta series of weight κ with $\kappa_2 - \kappa_1 \geq I$ of the norm form of a quadratic extension $M_{/F}$, M is a CM field, and all forms indexed by $\mathrm{Spec}(\mathbb{I})$ have CM by the same CM field M (see Sects. 1.12 and 1.13). In particular, the ring generated over $\mathbb{Z}_{(p)}$ by $a(\mathfrak{l})$ for primes $\mathfrak{l}$ of F in any CM component is a finite extension of R_ν taking $L = M$ for ν given by a CM type of M; so, the Hecke field has bounded degree over $\mathbb{Q}(\mu_{p^\infty})$ for any CM component. Fix an algebraic closure $\overline{Q}$ of the quotient field $Q = Q(\mathbf{\Lambda}_{\mathbb{Z}_p})$ of $\mathbf{\Lambda}_{\mathbb{Z}_p}$. We regard $\mathbb{I}$ as a subring of $\overline{Q}$. As a corollary of Theorem 3.1, we prove

Corollary I. *Let the notation be as above; in particular,* $\mathrm{Spec}(\mathbb{I})$ *is an irreducible (reduced) component of* $\mathrm{Spec}(\mathbf{h})$. *We regard* $\mathbb{I} \subset \overline{Q}$ *as* $\mathbf{\Lambda}$*-algebras and* $R_\nu \subset \mathbf{\Lambda} \subset \overline{Q}$. *Take a set* Σ *of prime ideals of* F *prime to* $p\mathfrak{n}$. *Suppose that* Σ *has positive upper density. If* $Q(R_\nu)[a(\mathfrak{l})] \subset \overline{Q}$ *for all* $\mathfrak{l} \in \Sigma$ *is a finite extension of* $Q(R_\nu)$ *for the quotient field* $Q(R_\nu)$ *of* R_ν*, then* $\mathbb{I}$ *is a component having complex multiplication by a CM quadratic extension* $M_{/F}$.

An obvious consequence of the above corollary is

Corollary II. *Let the notation be as in the above theorem. If* $\mathbb{I}$ *is a non-CM component, for a density one subset* Ξ *of primes of* F*, the subring* $Q(R_\nu)[a(\mathfrak{l})]$ *of* $\overline{Q}$ *for all* $\mathfrak{l} \in \Xi$ *has transcendental degree* 1 *over* $Q(R_\nu)$.

We could have made a conjecture on a mod p version of the above corollary as was done in [H14, Sect. 0], but we do not have an explicit application (as discussed in [H14]) to the Iwasawa μ-invariant of the generalized version; so, we do not formulate formally the obvious conjecture. We denote by a Gothic letter an ideal of a number field (in particular, any lowercase Gothic letter denotes an ideal of F). The corresponding Roman letter denotes the residual characteristic if a Gothic letter is used for a prime ideal. Adding superscript "(∞)", we indicate finite adeles; so, for example, $(F_{\mathbb{A}}^{(\infty)})^\times = \{x \in F_{\mathbb{A}}^\times | x_\infty = 1\}$. Similarly, $\mathbb{A}^{(p\infty)}$ is made of adeles without p and ∞-components.

The author would like to thank the referees for their careful reading.

1 Hilbert Modular Forms

We recall the arithmetic theory of Hilbert modular forms limiting ourselves to what we need later. The purpose of giving fair detail of the moduli theoretic interpretation of Hilbert modular forms here is twofold: (1) to make this article essentially self-contained and (2) because most account of this theory was written before the publication of the paper of Deligne–Pappas [DP94] and because there seems no detailed account available explaining that the correction to the original treatment in [Rp78] does not affect much the theory of p-adic modular forms.

Though most results in this section are used implicitly in the rest of the paper, the author also thought that it would be good to give a summary of the theory as this conference participants are very diverse and some of the people are quite far from the author's area of research. The reader who is familiar with the theory can go directly to Sect. 1.13 where a characterization of CM components is given (which is essential to the sequel). We keep the notation used in the introduction.

1.1 Abelian Varieties with Real Multiplication

Put $O^* = \{x \in F | \mathrm{Tr}(xO) \subset \mathbb{Z}\}$ (which is the different inverse $\mathfrak{d}^{-1}$). Recall the level ideal $\mathfrak{n}$, and fix a fractional ideal $\mathfrak{c}$ of F prime to $p\mathfrak{n}$. We write A for a fixed base commutative algebra with identity, in which the absolute norm $N(\mathfrak{c})$ and the prime-to-p part of $N(\mathfrak{n})$ are invertible. To include the case where p ramifies in the base field F, we use the moduli problem of Deligne–Pappas in [DP94] to define Hilbert modular varieties. As explained in [Z14, Sects. 2 and 3], if p is unramified in F, the resulting p-integral model of the Hilbert modular Shimura variety is canonically isomorphic to the one defined by Rapoport [Rp78] and Kottwitz [Ko92] (see also [PAF, Chap. 4]). Writing $\mathfrak{c}_+$ for the monoid of totally positive elements in $\mathfrak{c}$, giving data $(\mathfrak{c}, \mathfrak{c}_+)$ is equivalent to fix a strict ideal class of $\mathfrak{c}$. The Hilbert modular variety $\mathfrak{M} = \mathfrak{M}(\mathfrak{c}; \mathfrak{n})$ of level $\mathfrak{n}$ classifies triples $(X, \Lambda, i)_{/S}$ formed by

- An abelian scheme $\pi : X \to S$ of relative dimension $m = [F : \mathbb{Q}]$ over an A-scheme S (for the fixed algebra A) with an embedding: $O \hookrightarrow \mathrm{End}(X_{/S})$;
- An O-linear polarization $X^t := \mathrm{Pic}^0_{X/S} \underset{\sim}{\overset{\Lambda}{\to}} X \otimes \mathfrak{c}$ inducing an isomorphism $(\mathfrak{c}, \mathfrak{c}_+) \cong (\mathrm{Hom}^{Sym}_S(X_{/S}, X^t_{/S}), \mathcal{P}(X, X^t_{/S}))$, where $\mathrm{Hom}^{Sym}_S(X_{/S}, X^t_{/S})$ is the O-module of symmetric O-linear homomorphisms and $\mathcal{P}(X, X^t_{/S}) \subset \mathrm{Hom}^{Sym}_S(X_{/S}, X^t_{/S})$ is the positive cone made up of O-linear polarizations;
- A closed O-linear immersion $i = i_{\mathfrak{n}} : (\mathbb{G}_m \otimes_{\mathbb{Z}} O^*)[\mathfrak{n}] \hookrightarrow X$ for the group $(\mathbb{G}_m \otimes_{\mathbb{Z}} O^*)[\mathfrak{n}]$ of $\mathfrak{n}$-torsion points of the multiplicative O-module scheme $\mathbb{G}_m \otimes_{\mathbb{Z}} O^*$.

By Λ, we identify the O-module $\mathrm{Hom}^{Sym}_S(X_{/S}, X^t_{/S})$ of symmetric O-linear homomorphisms inside $\mathrm{Hom}_S(X_{/S}, X^t_{/S})$ with $\mathfrak{c}$. Then we require that the (multiplicative) monoid of symmetric O-linear isogenies induced locally by ample invertible sheaves be identified with the set of totally positive elements $\mathfrak{c}_+ \subset \mathfrak{c}$. The quasi-projective scheme $\mathfrak{M} = \mathfrak{M}(\mathfrak{c};\mathfrak{n})_{/A}$ is the coarse moduli scheme of the following functor $\wp$ from the category of A-schemes into the category *SETS*:

$$\wp(S) = \left[(X, \Lambda, i)_{/S}\right],$$

where $[\ ] = \{\ \}/\cong$ is the set of isomorphism classes of the objects inside the brackets, and we say $(X, \Lambda, i) \cong (X', \Lambda', i')$ if we have an O-linear isomorphism $\phi : X_{/S} \to X'_{/S}$ such that $\Lambda' = \phi \circ \Lambda \circ \phi^t$ and $i'^* \circ \phi = i^* (\Leftrightarrow \phi \circ i = i')$. The scheme $\mathfrak{M}$ is a fine moduli scheme if $\mathfrak{n}$ is sufficiently deep (see [DP94]).

1.2 Geometric Hilbert Modular Forms

In the definition of the functor $\wp$ in Sect. 1.1, we could impose local $\mathcal{O}_S \otimes_{\mathbb{Z}} O$-freeness of the $\mathcal{O}_S \otimes_{\mathbb{Z}} O$-module $\pi_*(\Omega_{X/S})$ as was done by Rapoport in [Rp78]. We consider an open subfunctor $\wp^R$ of $\wp$ which is defined by imposing local freeness of $\pi_*(\Omega_{X/S})$ over $\mathcal{O}_S \otimes_{\mathbb{Z}} O$. Over $\mathbb{Z}[\frac{1}{D_F}]$ for the discriminant D_F of F, the two functors $\wp^R$ and $\wp$ coincide (see [DP94]). We write $\mathfrak{M}^R(\mathfrak{c};\mathfrak{n})$ for the open subscheme of $\mathfrak{M}(\mathfrak{c};\mathfrak{n})$ representing $\wp^R$. For ω with $\pi_*(\Omega_{X/S}) = (\mathcal{O}_S \otimes_{\mathbb{Z}} O)\omega$, we consider the functor classifying quadruples (X, Λ, i, ω):

$$\mathcal{Q}(S) = \left[(X, \Lambda, i, \omega)_{/S}\right].$$

We let $a \in T_F(S) = H^0(S, (\mathcal{O}_S \otimes_{\mathbb{Z}} O)^\times)$ act on $\mathcal{Q}(S)$ by $(X, \Lambda, i, \omega) \mapsto (X, \Lambda, i, a\omega)$. By this action, $\mathcal{Q}$ is a T_F-torsor over the open subfunctor $\wp^R$ of $\wp$; so, $\mathcal{Q}$ is representable by an A-scheme $\mathcal{M} = \mathcal{M}(\mathfrak{c};\mathfrak{n})$ affine over $\mathfrak{M}^R = \mathfrak{M}^R(\mathfrak{c};\mathfrak{n})_{/A}$. For each weight $k \in X^*(T_F) = \mathrm{Hom}_{\text{gp-sch}}(T_F, \mathbb{G}_m)$, if $F \neq \mathbb{Q}$, the k^{-1}-eigenspace of $H^0(\mathcal{M}_{/A}, \mathcal{O}_{\mathcal{M}/A})$ is the space of modular forms of weight k integral over a ring A. We write $G_k(\mathfrak{c}, \mathfrak{n}; A)$ for this space of A-integral modular forms, which is an A-module of finite type. Thus $f \in G_k(\mathfrak{c}, \mathfrak{n}; A)$ is a functorial rule (i.e., a natural transformation $f : \mathcal{Q} \to \mathbb{G}_a$) assigning a value in B to each isomorphism class of $(X, \Lambda, i, \omega)_{/B}$ (defined over an A-algebra B) satisfying the following four conditions:

(G0) the value f at every cusp is finite (see below for its precise meaning);
(G1) $f(X, \Lambda, i, \omega) \in B$ if (X, Λ, i, ω) is defined over B;
(G2) $f((X, \Lambda, i, \omega) \otimes_B B') = \rho(f(X, \Lambda, i, \omega))$ for each morphism $\rho : B_{/A} \to B'_{/A}$;
(G3) $f(X, \Lambda, i, a\omega) = k(a)^{-1} f(X, \Lambda, i, \omega)$ for $a \in T_F(B)$.

Strictly speaking, the condition (G0) is only necessary when $F = \mathbb{Q}$ by the Koecher principle (see below at the end of this subsection for more details). By abusing the language, we consider f to be a function of isomorphism classes of test objects $(X, \Lambda, i, \omega)_{/B}$ hereafter. The sheaf of k^{-1}-eigenspace $\mathcal{O}_{\mathcal{M}}[k^{-1}]$ under the action of T_F is an invertible sheaf on $\mathfrak{M}^R_{/A}$. We write this sheaf as $\underline{\omega}^k$ (imposing (G0) when $F = \mathbb{Q}$). Then we have

$$G_k(\mathfrak{c}, \mathfrak{n}; A) \cong H^0(\mathfrak{M}^R(\mathfrak{c};\mathfrak{n})_{/A}, \underline{\omega}^k_{/A}) \ \text{ canonically},$$

as long as $\mathfrak{M}^R(\mathfrak{c};\mathfrak{n})$ is a fine moduli space. Writing $\underline{\mathbb{X}} := (\mathbb{X}, \boldsymbol{\lambda}, \mathbf{i}, \boldsymbol{\omega})$ for the universal abelian scheme over $\mathfrak{M}^R$, $s = f(\underline{\mathbb{X}})\boldsymbol{\omega}^k$ gives rise to the section of $\underline{\omega}^k$. Conversely, for any section $s \in H^0(\mathfrak{M}^R(\mathfrak{c};\mathfrak{n}), \underline{\omega}^k)$, taking the unique morphism $\phi : \mathrm{Spec}(B) \to \mathfrak{M}^R$ such that $\phi^*\underline{\mathbb{X}} = \underline{X}$ for $\underline{X} := (X, \Lambda, i, \omega)_{/B}$, we can define $f \in G_k$ by $\phi^* s = f(\underline{X})\omega^k$.

We suppose that the fractional ideal $\mathfrak{c}$ is prime to $\mathfrak{n}p$, and take two ideals $\mathfrak{a}$ and $\mathfrak{b}$ prime to $\mathfrak{n}p$ such that $\mathfrak{a}\mathfrak{b}^{-1} = \mathfrak{c}$. To $(\mathfrak{a}, \mathfrak{b})$, we attach the Tate AVRM $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$ defined over the completed group ring $\mathbb{Z}((\mathfrak{a}\mathfrak{b}))$ made of formal series $f(q) = \sum_{\xi \gg -\infty} a(\xi) q^{\xi}$ $(a(\xi) \in \mathbb{Z})$. Here ξ runs over all elements in $\mathfrak{a}\mathfrak{b}$, and there exists a positive integer n (dependent on f) such that $a(\xi) = 0$ if $\sigma(\xi) < -n$ for some $\sigma \in I$. We write $A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$ for the subring of $A[[\mathfrak{a}\mathfrak{b}]]$ made of formal series f with $a(\xi) = 0$ for all ξ with $\sigma(\xi) < 0$ for at least one embedding $\sigma : F \hookrightarrow \mathbb{R}$. Actually, we skipped a step of introducing the toroidal compactification of $\mathfrak{M}^R$ and $\mathfrak{M}$ (done in [Rp78] and [DP94]), and the universal abelian scheme over the moduli scheme degenerates to $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$ over the spectrum of (completed) stalk at the cusp corresponding to $(\mathfrak{a}, \mathfrak{b})$. The toroidal compactification of the scheme $\mathfrak{M}^R_{/A}$ is proper normal by Deligne and Pappas [DP94] and hence by Zariski's connected theorem, it is geometrically connected. Since $\mathfrak{M}^R$ is open dense in each fiber of $\mathfrak{M}$ (as shown by Deligne and Pappas [DP94]), it is geometrically connected. Therefore the q-expansion principle holds for $H^0(\mathfrak{M}^R(\mathfrak{c};\mathfrak{n}), \underline{\omega}^k)$. We refer details of these facts to [K78, Chap. I], [C90, DT04, Di03, DP94] [HT93, Sect. 1] and [PAF, Sect. 4.1.4]. The scheme $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$ can be extended to a semi-abelian scheme over $\mathbb{Z}[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$ adding the fiber $\mathbb{G}_m \otimes \mathfrak{a}^*$ over the augmentation ideal $\mathfrak{A}$. Since $\mathfrak{a}$ is prime to p, $\mathfrak{a}_p = O_p$. Thus if A is a $\mathbb{Z}_p$-algebra, we have the identity: $A \otimes_{\mathbb{Z}} \mathfrak{a}^* = A \otimes_{\mathbb{Z}_p} \mathfrak{a}_p^* = A \otimes_{\mathbb{Z}_p} O_p^* = A \otimes_{\mathbb{Z}} O^*$, and we have a canonical isomorphism:

$$Lie(\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q) \mod \mathfrak{A}) = Lie(\mathbb{G}_m \otimes \mathfrak{a}^*) \cong A \otimes_{\mathbb{Z}} \mathfrak{a}^* = A \otimes_{\mathbb{Z}} O^*.$$

By duality, we have $\Omega_{\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)/A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]} \cong A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$. Indeed we have a canonical generator ω_{can} of $\Omega_{\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)}$ induced by $\frac{dt}{t} \otimes 1$ on $\mathbb{G}_m \otimes \mathfrak{a}^*$. Since $\mathfrak{a}$ is prime to $\mathfrak{n}$, we have a canonical inclusion $(\mathbb{G}_m \otimes O^*)[\mathfrak{n}] = (\mathbb{G}_m \otimes \mathfrak{a}^*)[\mathfrak{n}]$ into $\mathbb{G}_m \otimes \mathfrak{a}^*$, which induces a canonical closed immersion $i_{can} : (\mathbb{G}_m \otimes O^*)[\mathfrak{n}] \hookrightarrow \mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$. As described in [K78, (1.1.14)] and [HT93, p. 204], $\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q)$ has a canonical $\mathfrak{c}$-polarization Λ_{can}. Thus we can evaluate $f \in G_k(\mathfrak{c}, \mathfrak{n}; A)$ at $(\mathrm{Tate}_{\mathfrak{a},\mathfrak{b}}(q), \Lambda_{can}, i_{can}, \omega_{can})$. The value $f(q) = f_{\mathfrak{a},\mathfrak{b}}(q)$ actually falls in $A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$ (if $F \neq \mathbb{Q}$: Koecher principle) and is called the q-expansion at the cusp $(\mathfrak{a}, \mathfrak{b})$. Finiteness at cusps in the condition (G0) can be stated as

(G0′) $f_{\mathfrak{a},\mathfrak{b}}(q) \in A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$ for all $(\mathfrak{a}, \mathfrak{b})$.

1.3 p-Adic Hilbert Modular Forms of Level $\mathfrak{n}p^\infty$

Suppose that $A = \varprojlim_n A/p^nA$ (such a ring is called a p-adic ring) and that $\mathfrak{n}$ is prime to p. We consider a functor into sets

$$\widehat{\wp}(A) = \left[(X, \Lambda, i_p, i_{\mathfrak{n}})_{/S}\right]$$

defined over the category of p-adic A-algebras $B = \varprojlim_n B/p^nB$. An important point is that we consider an embedding of ind-group schemes $i_p : \mu_{p^\infty} \otimes_{\mathbb{Z}_p} O_p^* \hookrightarrow X[p^\infty]$ (in place of a differential ω), which induces $\widehat{\mathbb{G}}_m \otimes O_p^* \cong \widehat{X}$ for the formal completion $\widehat{X}$ along the identity section of the characteristic p-fiber of the abelian scheme X over A.

We call an AVRM X over a characteristic p ring A p-ordinary if the Barsotti–Tate group $X[p^\infty]$ is ordinary; in other words, its (Frobenius) Newton polygon has only two slopes 0 and 1. In the moduli space $\mathfrak{M}(\mathfrak{c};\mathfrak{n})_{/\overline{\mathbb{F}}_p}$, locally under Zariski topology, the p-ordinary locus is an open subscheme of $\mathfrak{M}(\mathfrak{c};\mathfrak{n})$. Indeed, the locus is obtained by inverting the Hasse invariant (over $\mathfrak{M}(\mathfrak{c};\mathfrak{n})_{/\overline{\mathbb{F}}_p}$). So, the p-ordinary locus inside $\mathfrak{M}^R(\mathfrak{c};\mathfrak{n})$ is open in $\mathfrak{M}^R(\mathfrak{c};\mathfrak{n})$. In the same way as was done by Deligne–Ribet and Katz for the level p^∞-structure, we can prove that this functor is representable by the formal completion $\widehat{\mathfrak{M}}^R(\mathfrak{c};\mathfrak{n})$ of $\mathfrak{M}^R(\mathfrak{c};\mathfrak{n})$ along the p-ordinary locus of the modulo p fiber (e.g., [PAF, Sect. 4.1.9]).

Take a character $k \in \mathbb{Z}[I]$. A p-adic modular form $f_{/A}$ over a p-adic ring A is a function (strictly speaking, a functorial rule) of isomorphism classes of $(X, \Lambda, i_p, i_{\mathfrak{n}})_{/B}$ ($i_{\mathfrak{n}} : \mathbb{G}_m \otimes_{\mathbb{Z}} O^*[\mathfrak{n}] \hookrightarrow X$) satisfying the following three conditions:

(P1) $f(X, \Lambda, i_p, i_{\mathfrak{n}}) \in B$ if $(X, \Lambda, i_p, i_{\mathfrak{n}})$ is defined over B;
(P2) $f((X, \Lambda, i_p, i_{\mathfrak{n}}) \otimes_B B') = \rho(f(X, \Lambda, i_p, i_{\mathfrak{n}}))$ for each continuous A-algebra homomorphism $\rho : B \to B'$;
(P3) $f_{\mathfrak{a},\mathfrak{b}}(q) \in A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$ for all $(\mathfrak{a}, \mathfrak{b})$ prime to $\mathfrak{n}p$.

We write $V(\mathfrak{c}, \mathfrak{n}p^\infty; A)$ for the space of p-adic modular forms satisfying (P1-3). By definition, this space $V(\mathfrak{c}, \mathfrak{n}p^\infty; A)$ is a p-adically complete A-algebra.

The q-expansion principle is valid both for classical modular forms and p-adic modular forms f:

(q-exp) *The q-expansion: $f \mapsto f_{\mathfrak{a},\mathfrak{b}}(q) \in A[[(\mathfrak{a}\mathfrak{b})_{\geq 0}]]$ determines f uniquely.*

This follows from the irreducibility of the level p^∞ Igusa tower, which was proven in [DR80] (see also [PAF, Sect. 4.2.4] for another argument).

Fix a generator d of O_p^*. Since $\widehat{\mathbb{G}}_m \otimes O^*$ has a canonical invariant differential $\frac{dt}{t} \otimes d$, we have $\omega_p = i_{p,*}(\frac{dt}{t} \otimes d)$ on $X_{/B}$ [under the notation of (P1–3)]. This allows us to regard $f \in G_k(\mathfrak{c}, \mathfrak{n}; A)$ as a p-adic modular form by

$$f(X, \Lambda, i_p, i_{\mathfrak{n}}) := f(X, \Lambda, i_{\mathfrak{n}}, \omega_p).$$

By (q-exp), this gives an injection of $G_k(\mathfrak{c}, \mathfrak{n}; A)$ into $V(\mathfrak{c}, \mathfrak{n}p^\infty; A)$ preserving q-expansions.

1.4 Complex Analytic Hilbert Modular Forms

Over $\mathbb{C}$, the category of test objects (X, Λ, i, ω) is equivalent to the category of triples $(\mathcal{L}, \Lambda, i)$ made of the following data (by the theory of theta functions): $\mathcal{L}$ is an O-lattice in $O \otimes_{\mathbb{Z}} \mathbb{C} = \mathbb{C}^I$, an alternating pairing $\Lambda : \mathcal{L} \wedge_O \mathcal{L} \cong \mathfrak{c}^*$ and $i : \mathfrak{n}^*/O^* \hookrightarrow F\mathcal{L}/\mathcal{L}$. The alternating form Λ is supposed to be positive in the sense that $\Lambda(u, v)/\operatorname{Im}(uv^c)$ is totally positive definite. The differential ω can be recovered by $\iota : X(\mathbb{C}) = \mathbb{C}^I/\mathcal{L}$ so that $\omega = \iota^* du$ where $u = (u_\sigma)_{\sigma \in I}$ is the variable on $\mathbb{C}^I$. Conversely

$$\mathcal{L}_X = \left\{ \int_\gamma \omega \in O \otimes_{\mathbb{Z}} \mathbb{C} \middle| \gamma \in H_1(X(\mathbb{C}), \mathbb{Z}) \right\}$$

is a lattice in $\mathbb{C}^I$, and the polarization $\Lambda : X^t \cong X \otimes \mathfrak{c}$ induces $\mathcal{L}_X \wedge \mathcal{L}_X \cong \mathfrak{c}^*$.

Using this equivalence, we can relate our geometric definition of Hilbert modular forms with the classical analytic definition. Define $\mathfrak{Z}$ by the product of I copies of the upper half complex plane $\mathfrak{H}$. We regard $\mathfrak{Z} \subset F \otimes_{\mathbb{Q}} \mathbb{C} = \mathbb{C}^I$. For each $z \in \mathfrak{Z}$, we define

$$\mathcal{L}_z = 2\pi\sqrt{-1}(\mathfrak{b}z + \mathfrak{a}^*),\ \Lambda_z(2\pi\sqrt{-1}(az+b), 2\pi\sqrt{-1}(cz+d)) = -(ad-bc) \in \mathfrak{c}^*$$

with $i_z : \mathfrak{n}^*/O^* \to \mathbb{C}^I/\mathcal{L}_z$ given by $i_z(a \mod O^*) = (2\pi\sqrt{-1}a \mod \mathcal{L}_z)$.

Consider the following congruence subgroup $\Gamma_1^1(\mathfrak{n}; \mathfrak{a}, \mathfrak{b})$ given by

$$\left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(F) \middle| a, d \in O,\ b \in (\mathfrak{a}\mathfrak{b})^*,\ c \in \mathfrak{n}\mathfrak{a}\mathfrak{b}\mathfrak{d} \text{ and } d - 1 \in \mathfrak{n} \right\}.$$

Write $\Gamma_1^1(\mathfrak{c}; \mathfrak{n}) = \Gamma_1^1(\mathfrak{n}; O, \mathfrak{c}^{-1})$. We let $g = (g_\sigma) \in SL_2(F \otimes_{\mathbb{Q}} \mathbb{R}) = SL_2(\mathbb{R})^I$ act on $\mathfrak{Z}$ by linear fractional transformation of g_σ on each component z_σ. It is easy to verify

$$(\mathcal{L}_z, \Lambda_z, i_z) \cong (\mathcal{L}_w, \Lambda_w, i_w) \iff w = \gamma(z) \text{ for } \gamma \in \Gamma_1^1(\mathfrak{n}; \mathfrak{a}, \mathfrak{b}).$$

The set of pairs $(\mathfrak{a}, \mathfrak{b})$ with $\mathfrak{a}\mathfrak{b}^{-1} = \mathfrak{c}$ is in bijection with the set of cusps (unramified over ∞) of $\Gamma_1^1(\mathfrak{n}; \mathfrak{a}, \mathfrak{b})$. Two cusps are equivalent if they transform to each other by an element in $\Gamma_1^1(\mathfrak{n}; \mathfrak{a}, \mathfrak{b})$. The standard choice of the cusp is $(O, \mathfrak{c}^{-1})$, which we call the infinity cusp of $\mathfrak{M}(\mathfrak{c}; \mathfrak{n})$. For each ideal $\mathfrak{t}$, $(\mathfrak{t}, \mathfrak{t}\mathfrak{c}^{-1})$ gives another cusp. The two cusps $(\mathfrak{t}, \mathfrak{t}\mathfrak{c}^{-1})$ and $(\mathfrak{s}, \mathfrak{s}\mathfrak{c}^{-1})$ are equivalent under $\Gamma_1^1(\mathfrak{c}; \mathfrak{n})$ if $\mathfrak{t} = \alpha\mathfrak{s}$ for an element $\alpha \in F^\times$ with $\alpha \equiv 1 \mod \mathfrak{n}$ in $F_{\mathfrak{n}}^\times$. We have

$$\mathfrak{M}(\mathfrak{c}; \mathfrak{n})(\mathbb{C}) \cong \Gamma_1^1(\mathfrak{c}; \mathfrak{n})\backslash\mathfrak{Z}, \text{ canonically.}$$

Recall $G := \operatorname{Res}_{O/\mathbb{Z}}\mathrm{GL}(2)$. Take the following open compact subgroup of $G(\mathbb{A}^{(\infty)})$:

$$U_1^1(\mathfrak{n}) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\widehat{\mathbb{Z}}) \middle| c \in \mathfrak{n}\widehat{O} \text{ and } a \equiv d \equiv 1 \mod \mathfrak{n}\widehat{O} \right\},$$

and put $K = K_1^1(\mathfrak{n}) = \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1} U_1^1(\mathfrak{n}) \left(\begin{smallmatrix} d & 0 \\ 0 & 1 \end{smallmatrix}\right)$ for an idele d with $d\widehat{O} = \widehat{\mathfrak{d}}$ and $d_{\mathfrak{d}} = 1$. Here for an idele and an O-ideal $\mathfrak{a} \neq 0$, we write $x_{\mathfrak{a}}$ for the projection of x to $\prod_{\mathfrak{l}|\mathfrak{a}} F_{\mathfrak{l}}^{\times}$ and $x^{(\mathfrak{a})} = xx_{\mathfrak{a}}^{-1}$. Then taking an idele c with $c\widehat{O} = \widehat{\mathfrak{c}}$ and $c_{\mathfrak{c}} = 1$, we see that

$$\Gamma_1^1(\mathfrak{c};\mathfrak{n}) \subset \left(\left(\begin{smallmatrix} c & 0 \\ 0 & 1 \end{smallmatrix}\right) K \left(\begin{smallmatrix} c & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1} \cap G(\mathbb{Q})_+\right)$$

for $G(\mathbb{Q})_+$ made up of all elements in $G(\mathbb{Q})$ with totally positive determinant. Choosing a complete representative set $\{c\} \subset F_{\mathbb{A}}^{\times}$ for the strict ray class group $Cl_F^+(\mathfrak{n})$ modulo $\mathfrak{n}$, we find by the approximation theorem that

$$G(\mathbb{A}) = \bigsqcup_{c \in Cl_F^+(\mathfrak{n})} G(\mathbb{Q}) \left(\begin{smallmatrix} c & 0 \\ 0 & 1 \end{smallmatrix}\right) K \cdot G(\mathbb{R})^+$$

for the identity connected component $G(\mathbb{R})^+$ of the Lie group $G(\mathbb{R})$. This shows

$$G(\mathbb{Q})\backslash G(\mathbb{A})/KC_{\mathbf{i}} \cong G(\mathbb{Q})_+\backslash G(\mathbb{A})_+/KC_{\mathbf{i}} \cong \bigsqcup_{\mathfrak{c} \in Cl_F^+(\mathfrak{n})} \mathfrak{M}(\mathfrak{c};\mathfrak{n})(\mathbb{C}), \tag{1}$$

where $G(\mathbb{A})_+ = G(\mathbb{A}^{(\infty)})G(\mathbb{R})^+$ and $C_{\mathbf{i}}$ is the stabilizer of $\mathbf{i} = (\sqrt{-1}, \ldots, \sqrt{-1}) \in \mathfrak{Z}$ in $G(\mathbb{R})^+$. By (1), a $Cl_F^+(\mathfrak{n})$-tuple $(f_{\mathfrak{c}})_{\mathfrak{c}}$ with $f_{\mathfrak{c}} \in G_k(\mathfrak{c},\mathfrak{n};\mathbb{C})$ can be viewed as a single automorphic form defined on $G(\mathbb{A})$.

Recall the identification $X^*(T_F)$ with $\mathbb{Z}[I]$ so that $k(x) = \prod_\sigma \sigma(x)^{k_\sigma}$ for $k = \sum_\sigma k_\sigma \sigma \in \mathbb{Z}[I]$. Regarding $f \in G_k(\mathfrak{c},\mathfrak{n};\mathbb{C})$ as a function of $z \in \mathfrak{Z}$ by $f(z) = f(\mathcal{L}_z, \Lambda_z, i_z)$, it satisfies the following automorphic property:

$$f(\gamma(z)) = f(z) \prod_\sigma (c^\sigma z_\sigma + d^\sigma)^{k_\sigma} \quad \text{for all } \gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1^1(\mathfrak{c};\mathfrak{n}). \tag{2}$$

The holomorphy of f follows from the functoriality (G2). The function f has the Fourier expansion

$$f(z) = \sum_{\xi \in (\mathfrak{ab})_{\geq 0}} a(\xi)\mathbf{e}_F(\xi z)$$

at the cusp corresponding to $(\mathfrak{a},\mathfrak{b})$. Here $\mathbf{e}_F(\xi z) = \exp(2\pi\sqrt{-1}\sum_\sigma \xi^\sigma z_\sigma)$. This Fourier expansion gives the q-expansion $f_{\mathfrak{a},\mathfrak{b}}(q)$ substituting q^ξ for $\mathbf{e}_F(\xi z)$.

1.5 Γ_0-Level Structure and Hecke Operators

We now assume that the base algebra A is a W-algebra. Choose a prime $\mathfrak{q}$ of F. We are going to define Hecke operators $U(\mathfrak{q}^n)$ and $T(1,\mathfrak{q}^n)$ assuming for simplicity that $\mathfrak{q} \nmid p\mathfrak{n}$, though we may extend the definition to arbitrary $\mathfrak{q}$ (see [PAF, Sect. 4.1.10]).

Then $X[\mathfrak{q}^r]$ is an étale group scheme over B if X is an abelian scheme over an A-algebra B. We call a subgroup $C \subset X$ cyclic of order $\mathfrak{q}^r$ if $C \cong O/\mathfrak{q}^r$ over an étale faithfully flat extension of B.

We can think of quintuples $(X, \Lambda, i, C, \omega)_{/S}$ adding an additional information C of a cyclic subgroup scheme $C \subset X$ cyclic of order $\mathfrak{q}^r$. We define the space of classical modular forms $G_k(\mathfrak{c}, \mathfrak{n}, \Gamma_0(\mathfrak{q}^r); A)$ (resp. the space $V(\mathfrak{c}, \mathfrak{n}p^\infty, \Gamma_0(\mathfrak{q}^r); A)$ of p-adic modular forms) of prime-to-p level $(\mathfrak{n}, \Gamma_0(\mathfrak{q}^r))$ by (G0-3) [resp. (P1-3)] replacing test objects (X, Λ, i, ω) [resp. $(X, \Lambda, i_\mathfrak{n}, i_p)$] by $(X, \Lambda, i, C, \omega)$ [resp. $(X, \Lambda, i_\mathfrak{n}, C, i_p)$].

Our Hecke operators are defined on the space of prime-to-p level $(\mathfrak{n}, \Gamma_0(\mathfrak{q}^r))$. The operator $U(\mathfrak{q}^n)$ is defined only when $r > 0$ and $T(1, \mathfrak{q}^n)$ is defined only when $r = 0$. For a cyclic subgroup C' of $X_{/B}$ of order $\mathfrak{q}^n$, we can define the quotient abelian scheme X/C' with projection $\pi : X \to X/C'$. The polarization Λ and the differential ω induce a polarization $\pi_*\Lambda$ and a differential $(\pi^*)^{-1}\omega$ on X/C'. If $C' \cap C = \{0\}$ (in this case, we call that C' and C are *disjoint*), $\pi(C)$ gives rise to the level $\Gamma_0(\mathfrak{q}^r)$-structure on X/C'. Then we define $U(\mathfrak{q})$-operators acting on $f \in V(\mathfrak{c}\mathfrak{q}^n; \mathfrak{n}p^\infty, \Gamma_0(\mathfrak{q}^r); A)$ by

$$f|U(\mathfrak{q}^n)(X, \Lambda, C, i_\mathfrak{n}, C, i_p) = \frac{1}{N(\mathfrak{q}^n)} \sum_{C'} f(X/C', \pi_*\Lambda, \pi \circ i_\mathfrak{n}, \pi(C), \pi \circ i_p) \qquad (3)$$

where C' runs over all cyclic subgroups of order $\mathfrak{q}^n$ *disjoint* from C. Since $\pi_*\Lambda = \pi \circ \Lambda \circ \pi^t$ is a $\mathfrak{c}\mathfrak{q}^n$-polarization, the modular form f has to be defined for abelian varieties with $\mathfrak{c}\mathfrak{q}^n$-polarization.

As for $T(1, \mathfrak{q}^n)$, since $\mathfrak{q} \nmid \mathfrak{n}$, forgetting the $\Gamma_0(\mathfrak{q}^n)$-structure, we define $T(1, \mathfrak{q}^n)$ acting on $f \in V(\mathfrak{c}\mathfrak{q}^n; \mathfrak{n}p^\infty; A)$ by

$$f|T(1, \mathfrak{q}^n)(X, \Lambda, i_\mathfrak{n}, i_p) = \frac{1}{N(\mathfrak{q}^n)} \sum_{C'} f(X/C', \pi_*\Lambda, \pi \circ i_\mathfrak{n}, \pi \circ i_p) \text{ if } f \in V(A), \qquad (4)$$

where C' runs over all cyclic subgroups of order $\mathfrak{q}^n$. We check that $f|U(\mathfrak{q}^n)$ [resp. $T(1, \mathfrak{q}^n)$] belongs to $V(\mathfrak{c}\mathfrak{q}^n; \mathfrak{n}p^\infty, \Gamma_0(\mathfrak{q}^r); A)$ [resp. $V(\mathfrak{c}\mathfrak{q}^n; \mathfrak{n}p^\infty; A)$], and compatible with the natural inclusion $G_k(\mathfrak{c}, \mathfrak{n}, \Gamma_0(\mathfrak{q}^r); A) \hookrightarrow V(\mathfrak{c}\mathfrak{q}^n; \mathfrak{n}p^\infty, \Gamma_0(\mathfrak{q}^r); A)$ [resp. $G_k(\mathfrak{c}, \mathfrak{n}; A) \hookrightarrow V(\mathfrak{c}\mathfrak{q}^n; \mathfrak{n}p^\infty; A)$] defined at the end of Sect. 1.3; so, these Hecke operators preserve classicality. We have

$$U(\mathfrak{q}^n) = U(\mathfrak{q})^n.$$

1.6 Hilbert Modular Shimura Varieties

We extend the level structure i limited to $\mathfrak{n}$-torsion points to far bigger structure $\eta^{(p)}$ including all prime-to-p torsion points. Let $\mathbb{Z}_{(p)} = \mathbb{Q} \cap \mathbb{Z}_p$ (the localization of $\mathbb{Z}$ at (p)). Triples $(X, \overline{\Lambda}, \eta^{(p)})_{/S}$ for $\mathbb{Z}_{(p)}$-schemes S are classified by an integral model $Sh^{(p)}_{/\mathbb{Z}_{(p)}}$ (cf. [Ko92]) of the Shimura variety $Sh_{/\mathbb{Q}}$ associated with the algebraic $\mathbb{Z}_{(p)}$-group G (in the sense of Deligne [D71, 4.22] interpreting Shimura's original definition in [Sh70] as a moduli of abelian schemes up to isogenies). Here the classification is up to prime-to-p isogenies, and $\overline{\Lambda}$ is an equivalence class of polarizations up to multiplication by totally positive elements in F prime to p.

To give a description of the functor represented by $Sh^{(p)}$, we introduce some more notations. We consider the fiber category $\mathcal{A}^{(p)}_F$ over schemes defined by

(Object) abelian schemes X with real multiplication by O;
(Morphism) $\mathrm{Hom}_{\mathcal{A}^{(p)}_F}(X, Y) = \mathrm{Hom}(X, Y) \otimes_{\mathbb{Z}} \mathbb{Z}_{(p)}$.

Isomorphisms in this category are isogenies with degree prime to p (called "prime-to-p isogenies"), and hence the degree of polarization Λ is supposed to be also prime to p. Two polarizations are equivalent if $\Lambda = c\Lambda' = \Lambda' \circ i(c)$ for a totally positive c prime to p. We fix an O-lattice $L \subset V = F^2$ with O-hermitian alternating pairing $\langle \cdot, \cdot \rangle$ inducing a self-duality on $L_p = L \otimes_{\mathbb{Z}} \mathbb{Z}_p$.

For an open-compact subgroup K of $G(\mathbb{A}^{(\infty)})$ *maximal at* p (i.e., $K = G(\mathbb{Z}_p) \times K^{(p)}$), we consider the following functor from $\mathbb{Z}_{(p)}$-schemes into *SETS*:

$$\wp^{(p)}_K(S) = \left[(X, \overline{\Lambda}, \overline{\eta}^{(p)})_{/S} \text{ with (det)}\right]. \tag{5}$$

Here $\overline{\eta}^{(p)} : L \otimes_{\mathbb{Z}} \mathbb{A}^{(p\infty)} \cong V^{(p)}(X) = T(X) \otimes_{\mathbb{Z}} \mathbb{A}^{(p\infty)}$ is an equivalence class of $\eta^{(p)}$ modulo multiplication $\eta^{(p)} \mapsto \eta^{(p)} \circ k$ by $k \in K^{(p)}$ for the Tate module $T(X) = \varprojlim_{\mathfrak{n}} X[\mathfrak{n}]$ (in the sheafified sense that $\eta^{(p)} \equiv (\eta')^{(p)} \mod K$ étale-locally), and a $\Lambda \in \overline{\Lambda}$ induces the self-duality on L_p. As long as $K^{(p)}$ is sufficiently small, $\wp^{(p)}_K$ is representable over any $\mathbb{Z}_{(p)}$-algebra A (cf. [Ko92, DP94] and [Z14, Sect. 3]) by a scheme $Sh_{K/A} = Sh/K$, which is smooth over $\mathrm{Spec}(\mathbb{Z}_{(p)})$ if p is unramified in $F_{/\mathbb{Q}}$ and singular if $p|D_F$ but is smooth outside a closed subscheme of codimension 2 in the p-fiber $Sh^{(p)} \times_{\mathbb{Z}_{(p)}} \mathbb{F}_p$ by the result of [DP94]. We let $g \in G(\mathbb{A}^{(p\infty)})$ act on $Sh^{(p)}_{/\mathbb{Z}_{(p)}}$ by

$$x = (X, \overline{\Lambda}, \eta) \mapsto g(x) = (X, \overline{\Lambda}, \eta \circ g),$$

which gives a right action of $G(\mathbb{A})$ on $Sh^{(p)}$ through the projection $G(\mathbb{A}) \twoheadrightarrow G(\mathbb{A}^{(p\infty)})$.

By the universality, we have a morphism $\mathfrak{M}(\mathfrak{c}; \mathfrak{n}) \to Sh^{(p)}/\widehat{\Gamma}^1_1(\mathfrak{c}; \mathfrak{n})$ for the open compact subgroup: $\widehat{\Gamma}^1_1(\mathfrak{c}; \mathfrak{n}) = \left(\begin{smallmatrix} c & 0 \\ 0 & 1 \end{smallmatrix}\right) K^1_1(\mathfrak{n}) \left(\begin{smallmatrix} c & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1} = \left(\begin{smallmatrix} cd^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right) U^1_1(\mathfrak{n}) \left(\begin{smallmatrix} cd^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1}$ maximal at p. The image of $\mathfrak{M}(\mathfrak{c}; \mathfrak{n})$ gives a geometrically irreducible component of $Sh^{(p)}/\widehat{\Gamma}^1_1(\mathfrak{c}; \mathfrak{n})$. If $\mathfrak{n}$ is sufficiently deep, we can identify $\mathfrak{M}(\mathfrak{c}; \mathfrak{n})$ with its image in $Sh^{(p)}/\widehat{\Gamma}^1_1(\mathfrak{c}; \mathfrak{n})$. By the action on the polarization $\Lambda \mapsto \alpha\Lambda$ for a suitable totally positive $\alpha \in F$, we can bring $\mathfrak{M}(\mathfrak{c}; \mathfrak{n})$ into $\mathfrak{M}(\alpha\mathfrak{c}; \mathfrak{n})$; so, the image of $\varprojlim_{\mathfrak{n}} \mathfrak{M}(\mathfrak{c}; \mathfrak{n})$ in $Sh^{(p)}$ only depends on the strict ideal class of $\mathfrak{c}$ in $\varprojlim_{\mathfrak{n}:\mathfrak{n}+(p)=O} Cl^+_F(\mathfrak{n})$.

1.7 Level Structure with "Neben" Character

In order to make a good link between classical modular forms and adelic automorphic forms (which we will describe in the following subsection), we would like to introduce "Neben" characters. We fix an integral ideal $\mathfrak{n}' \subset O$. We think of the following level structure on an AVRM X:

$$i : (\mathbb{G}_m \otimes O^*)[\mathfrak{n}'] \hookrightarrow X[\mathfrak{n}'] \text{ and } i' : X[\mathfrak{n}'] \twoheadrightarrow O/\mathfrak{n}', \tag{6}$$

where the sequence

$$1 \to (\mathbb{G}_m \otimes O^*)[\mathfrak{n}'] \xrightarrow{i} X[\mathfrak{n}'] \xrightarrow{i'} O/\mathfrak{n}' \to 0 \tag{7}$$

is exact and is required to induce a canonical duality between $(\mathbb{G}_m \otimes O^*)[\mathfrak{n}']$ and $O/\mathfrak{n}'$ under the polarization Λ. Here, if $\mathfrak{n}' = (N)$ for an integer $N > 0$, a canonical duality pairing

$$\langle \cdot, \cdot \rangle : (\mathbb{G}_m \otimes O^*)[N] \times O/N \to \mu_N$$

is given by $\langle \zeta \otimes \alpha, m \otimes \beta \rangle = \zeta^{m\mathrm{Tr}(\alpha\beta)}$ for $(\alpha, \beta) \in O^* \times O$ and $(\zeta, m) \in \mu_N \times \mathbb{Z}/N$ identifying $(\mathbb{G}_m \otimes O^*)[N] = \mu_N \otimes O^*$ and $O/N = (\mathbb{Z}/N\mathbb{Z}) \otimes_{\mathbb{Z}} O$. In general, taking an integer $0 < N \in \mathfrak{n}'$, the canonical pairing between $(\mathbb{G}_m \otimes O^*)[\mathfrak{n}']$ and $O/\mathfrak{n}'$ is induced by the one for (N) via the canonical inclusion $(\mathbb{G}_m \otimes O^*)[\mathfrak{n}'] \hookrightarrow (\mathbb{G}_m \otimes O^*)[N]$ and the quotient map $O/(N) \twoheadrightarrow O/\mathfrak{n}'$.

We fix two characters $\epsilon_1 : (O/\mathfrak{n}')^\times \to A^\times$ and $\epsilon_2 : (O/\mathfrak{n}')^\times \to A^\times$, and we insist for $f \in G_k(\mathfrak{c}, \mathfrak{n}; A)$ on the version of (G0-3) for quintuples $(X, \Lambda, i \cdot a, d \cdot i', \omega)$ and the equivariancy:

$$f(X, \overline{\Lambda}, i \cdot d, a \cdot i', \omega) = \epsilon_1(d)\epsilon_2(a) f(X, \overline{\Lambda}, i, i', \omega) \text{ for } a, d \in (O/\mathfrak{n})^\times. \tag{Neben}$$

Here the order $\epsilon_1(d)\epsilon_2(a)$ is correct as the diagonal matrix $\left(\begin{smallmatrix} d & 0 \\ 0 & a \end{smallmatrix}\right)$ in $T^\Delta(O/\mathfrak{n}') \subset \mathrm{GL}_2(O/\mathfrak{n}')$ acts on the quotient $O/\mathfrak{n}'$ by a and the submodule $(\mathbb{G}_m \otimes O^*)[\mathfrak{n}']$ by d. The ordering of ϵ_1, ϵ_2 is normalized with respect to the Galois representation local at p of f (when f is a p-ordinary Hecke eigenform so that ϵ_1 as a Galois character corresponds to the quotient character of the local Galois representation; see (Ram) in Sect. 1.11). Here $\overline{\Lambda}$ is the polarization class modulo equivalence relation given by multiplication by totally positive numbers in F prime to p. We write $G_k(\mathfrak{c}, \Gamma_0(\mathfrak{n}), \epsilon; A)$ ($\epsilon = (\epsilon_1, \epsilon_2)$) for the A-module of geometric modular forms satisfying these conditions.

1.8 Adelic Hilbert Modular Forms

Let us interpret what we have said so far in automorphic language and give a definition of the adelic Hilbert modular forms and their Hecke algebra of level $\mathfrak{n}$ (cf. [H96, Sects. 2.2–4] and [PAF, Sects. 4.2.8–4.2.12]).

We consider the following open compact subgroup of $G(\mathbb{A}^{(\infty)})$:

$$U_0(\mathfrak{n}) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\widehat{\mathbb{Z}}) \middle| c \equiv 0 \mod \mathfrak{n}\widehat{O}\right\},$$
$$U_1^1(\mathfrak{n}) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in U_0(\mathfrak{n}) \middle| a \equiv d \equiv 1 \mod \mathfrak{n}\widehat{O}\right\}, \tag{8}$$

where $\widehat{O} = O \otimes_{\mathbb{Z}} \widehat{\mathbb{Z}}$ and $\widehat{\mathbb{Z}} = \prod_\ell \mathbb{Z}_\ell$. Then we introduce the following semi-group

$$\Delta_0(\mathfrak{n}) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\mathbb{A}^{(\infty)}) \cap M_2(\widehat{O}) \middle| c \equiv 0 \mod \mathfrak{n}\widehat{O}, d_{\mathfrak{n}} \in O_{\mathfrak{n}}^\times\right\}, \tag{9}$$

where $d_{\mathfrak{n}}$ is the projection of $d \in \widehat{O}$ to $O_{\mathfrak{n}} := \prod_{\mathfrak{q}|\mathfrak{n}} O_{\mathfrak{q}}$ for prime ideals $\mathfrak{q}$. Recall the maximal diagonal torus T^Δ of $\mathrm{GL}(2)_{/O}$. Putting

$$D_0 = \left\{\mathrm{diag}[a, d] = \left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right) \in T^\Delta(F_{\mathbb{A}^{(\infty)}}) \cap M_2(\widehat{O}) \middle| d_{\mathfrak{n}} = 1\right\}, \tag{10}$$

we have (e.g., [MFG, 3.1.6] and [PAF, Sect. 5.1])

$$\Delta_0(\mathfrak{n}) = U_0(\mathfrak{n}) D_0 U_0(\mathfrak{n}). \tag{11}$$

In this section, the group U is assumed to be a subgroup of $U_0(\mathfrak{n}p^\alpha)$ with $U \supset U_1^1(\mathfrak{n}p^\alpha)$ for some $0 < \alpha \leq \infty$. Formal finite linear combinations $\sum_\delta c_\delta U\delta U$ of double cosets of U in $\Delta_0(\mathfrak{n}p^\alpha)$ form a ring $R(U, \Delta_0(\mathfrak{n}p^\alpha))$ under convolution product (see [IAT, Chap. 3] or [MFG, Sect. 3.1.6]). Recall the prime element $\varpi_{\mathfrak{q}}$ of $O_{\mathfrak{q}}$ for each prime $\mathfrak{q}$ fixed in the introduction. The algebra is commutative and is isomorphic to the polynomial ring over the group algebra $\mathbb{Z}[U_0(\mathfrak{n}p^\alpha)/U]$ with variables $\{T(\mathfrak{q}), T(\mathfrak{q},\mathfrak{q})\}_{\mathfrak{q}}$. Here $T(\mathfrak{q})$ (resp. $T(\mathfrak{q},\mathfrak{q})$ for primes $\mathfrak{q} \nmid \mathfrak{n}p^\alpha$) corresponds to the double coset $U\left(\begin{smallmatrix} \varpi_{\mathfrak{q}} & 0 \\ 0 & 1 \end{smallmatrix}\right)U$ (resp. $U_0\varpi_{\mathfrak{q}}U$). The group element $u \in U_0(\mathfrak{n}p^\alpha)/U$ in the group algebra $\mathbb{Z}[U_0(\mathfrak{n}p^\alpha)/U]$ corresponds to the double coset UuU (cf. [H95, Sect. 2]).

As in the introduction, we extend ϵ_j to a character of $(F_{\mathbb{A}}^{(\infty)})^\times \subset \widehat{O}^\times \times \prod_{\mathfrak{q}} \varpi_{\mathfrak{q}}^{\mathbb{Z}}$ trivial on the factor $\prod_{\mathfrak{q}} \varpi_{\mathfrak{q}}^{\mathbb{Z}}$, and denote the extended character by the same symbol ϵ_j. In [HMI, (ex0–3)], ϵ_2 is extended as above, but the extension of ϵ_1 taken there is to keep the identity $\epsilon_+ = \epsilon_1\epsilon_2$ over $(F_{\mathbb{A}}^{(\infty)})^\times$. The present extension is more convenient in this paper.

The double coset ring $R(U, \Delta_0(\mathfrak{n}p^\alpha))$ naturally acts on the space of modular forms on U. We now recall the action (which is a slight simplification of the action of $[UxU]$ given in [HMI, (2.3.14)]). Recall the diagonal torus T^Δ of $\mathrm{GL}(2)_{/O}$; so,

$T^\Delta = \mathbb{G}^2_{m/O}$. Since $T^\Delta(O/\mathfrak{n}')$ is canonically a quotient of $U_0(\mathfrak{n}')$ for an ideal $\mathfrak{n}'$, a character $\epsilon : T^\Delta(O/\mathfrak{n}') \to \mathbb{C}^\times$ can be considered as a character of $U_0(\mathfrak{n}')$. If ϵ_j is defined modulo $\mathfrak{n}_j$, we can take $\mathfrak{n}'$ to be any multiple of $\mathfrak{n}_1 \cap \mathfrak{n}_2$. Writing $\epsilon\left(\left(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right)\right) = \epsilon_1(a)\epsilon_2(d)$, if $\epsilon^- = \epsilon_1\epsilon_2^{-1}$ factors through $(O/\mathfrak{n})^\times$ for for an ideal $\mathfrak{n}|\mathfrak{n}'$, then we can extend the character ϵ of $U_0(\mathfrak{n}')$ to $\Delta_0(\mathfrak{n})$ by putting $\epsilon(\delta) = \epsilon_1(\det(\delta))(\epsilon^-)^{-1}(d_\mathfrak{n})$ for $\delta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Delta_0(\mathfrak{n})$ (as before). In this sense, we hereafter assume that ϵ is defined modulo $\mathfrak{n}$ and regard ϵ as a character of the group $U_0(\mathfrak{n})$ and the semi-group $\Delta_0(\mathfrak{n})$. Recall that $\epsilon_+ : F^\times_\mathbb{A} \to \mathbb{C}^\times$ is a Hecke character trivial on $F^\times$ with infinity type $(1-[\kappa])I$ (for an integer $[\kappa]$) such that $\epsilon_+(z) = \epsilon_1(z)\epsilon_2(z)$ for $z \in \widehat{O}^\times$.

Recall the set I of all embeddings of F into $\overline{\mathbb{Q}}$ and T^Δ_F for $\mathrm{Res}_{O/\mathbb{Z}}T^\Delta$ (the diagonal torus of G). Then the group of geometric characters $X^*(T^\Delta_F)$ is isomorphic to $\mathbb{Z}[I]^2$ so that $(m,n) \in \mathbb{Z}[I]^2$ send $\mathrm{diag}[x,y] \in T^\Delta_F$ to $x^m y^n = \prod_{\sigma\in I}(\sigma(x)^{m_\sigma}\sigma(y)^{n_\sigma})$. Taking $\kappa = (\kappa_1,\kappa_2) \in \mathbb{Z}[I]^2$, we assume $[\kappa]I = \kappa_1 + \kappa_2$, and we associate with κ a factor of automorphy:

$$J_\kappa(g,\tau) = \det(g_\infty)^{\kappa_1 - I} j(g_\infty,\tau)^{\kappa_2-\kappa_1+I} \text{ for } g \in G(\mathbb{A}) \text{ and } \tau \in \mathfrak{Z}. \tag{12}$$

We define $S_\kappa(U,\epsilon;\mathbb{C})$ for an open subgroup $U \subset U_0(\mathfrak{n})$ by the space of functions $\mathbf{f} : G(\mathbb{A}) \to \mathbb{C}$ satisfying the following three conditions (e.g., [HMI, (SA1–3)] and [PAF, Sect. 4.3.1]):

(S1) $\mathbf{f}(\alpha x u z) = \epsilon(u)\epsilon_+(z)\mathbf{f}(x)J_\kappa(u,\mathbf{i})^{-1}$ for $\alpha \in G(\mathbb{Q})$, $u \in U \cdot C_\mathbf{i}$ and $z \in Z(\mathbb{A})$.

(S2) Choose $u \in G(\mathbb{R})$ with $u(\mathbf{i}) = \tau$ for $\tau \in \mathfrak{Z}$, and put $\mathbf{f}_x(\tau) = \mathbf{f}(xu)J_\kappa(u,\mathbf{i})$ for each $x \in G(\mathbb{A}^{(\infty)})$ (which only depends on τ). Then $\mathbf{f}_x$ is a holomorphic function on $\mathfrak{Z}$ for all x.

(S3) $\mathbf{f}_x(\tau)$ for each x is rapidly decreasing as $\eta_\sigma \to \infty$ ($\tau = \xi + \mathbf{i}\eta$) for all $\sigma \in I$ uniformly.

If we replace the expression "rapidly decreasing" in (S3) by "slowly increasing," we get the definition of the space $G_\kappa(U,\epsilon;\mathbb{C})$. It is easy to check (e.g., [HMI, (2.3.5)] that the function $\mathbf{f}_x$ in (S2) satisfies

$$f(\gamma(\tau)) = \epsilon^{-1}(x^{-1}\gamma x)f(\tau)J_\kappa(\gamma,\tau) \text{ for all } \gamma \in \Gamma_x(U), \tag{13}$$

where $\Gamma_x(U) = xUx^{-1}G(\mathbb{R})^+ \cap G(\mathbb{Q})$. Also by (S3), $\mathbf{f}_x$ is rapidly decreasing towards all cusps of Γ_x; so, it is a cusp form. If we restrict $\mathbf{f}$ as above to $SL_2(F_\mathbb{A})$, the determinant factor $\det(g)^{\kappa_1-I}$ in the factor $J_\kappa(g,\tau)$ disappears, and the automorphy factor becomes only dependent on $k = \kappa_2 - \kappa_1 + I \in \mathbb{Z}[I]$; so, the classical modular form in G_k has single digit weight $k \in \mathbb{Z}[I]$. Via (1), we have an embedding of $S_\kappa(U_0(\mathfrak{n}'),\epsilon;\mathbb{C})$ into $G_k(\Gamma_0(\mathfrak{n}'),\epsilon;\mathbb{C}) = \bigoplus_{[\mathfrak{c}]\in Cl^+_F} G_k(\mathfrak{c},\Gamma_0(\mathfrak{n}'),\epsilon;\mathbb{C})$ ($\mathfrak{c}$ running over a complete representative set prime to $\mathfrak{n}'$ for the strict ideal class group Cl^+_F) bringing $\mathbf{f}$ into $(\mathbf{f}_\mathfrak{c})_{[\mathfrak{c}]}$ for $\mathbf{f}_\mathfrak{c} = \mathbf{f}_x$ [as in (S3)] with $x = \left(\begin{smallmatrix} cd^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right)$ (for $d \in F^\times_\mathbb{A}$ with $d\widehat{O} = \widehat{\mathfrak{d}}$). The cusp form $\mathbf{f}_\mathfrak{c}$ is determined by the restriction of $\mathbf{f}$ to $x \cdot SL_2(F_\mathbb{A})$. Though in (13), ϵ^{-1} shows up, the Neben character of the direct factor $G_k(\mathfrak{c},\Gamma_0(\mathfrak{n}'),\epsilon;\mathbb{C})$ is given by ϵ, since in (Neben), the order of (a,d) is reversed to have $\epsilon_1(d)\epsilon_2(a)$. If we vary the weight κ keeping $k = \kappa_2 - \kappa_1 + I$, the image of S_κ in $G_k(\Gamma_0(\mathfrak{n}'), \epsilon; \mathbb{C})$ transforms accordingly. By this identification, the Hecke operator $T(\mathfrak{q})$ for non-principal $\mathfrak{q}$ makes sense as an operator acting on a single space $G_\kappa(U, \epsilon; \mathbb{C})$, and its action depends on the choice of κ.

It is known that $G_\kappa = 0$ unless $\kappa_1+\kappa_2 = [\kappa_1+\kappa_2]I$ for $[\kappa_1+\kappa_2] \in \mathbb{Z}$, because $I-(\kappa_1+\kappa_2)$ is the infinity type of the central character of automorphic representations generated by G_κ. We write simply $[\kappa]$ for $[\kappa_1+\kappa_2] \in \mathbb{Z}$ assuming $G_\kappa \neq 0$. The $SL(2)$-weight of the central character of an irreducible automorphic representation π generated by $\mathbf{f} \in G_\kappa(U, \epsilon; \mathbb{C})$ is given by k (which specifies the infinity type of π_∞ as a discrete series representation of $SL_2(F_\mathbb{R})$).

In the introduction, we have extended ϵ_j to $(F_\mathbb{A}^{(\infty)})^\times$ and ϵ to $\Delta_0(\mathfrak{n})$ (as long as ϵ^- is defined modulo $\mathfrak{n}$), and we have $\epsilon(\delta) = \epsilon_1(\det(\delta))(\epsilon^-)^{-1}(d_\mathfrak{n})$ for $\delta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Delta_0(\mathfrak{n})$. Let $\mathcal{U}$ be the unipotent algebraic subgroup of $\mathrm{GL}(2)_{/O}$ defined by $\mathcal{U}(A) = \left\{\left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right) \middle| a \in A\right\}$. Note here that $\mathcal{U}(\widehat{O}) \subset \mathrm{Ker}(\epsilon)$; so, $\epsilon(tu) = \epsilon(t)$ if $t \in D_0$ and $u \in \mathcal{U}(\widehat{O})$. For each $UyU \in R(U, \Delta_0(\mathfrak{n}p^\alpha))$, we decompose $UyU = \bigsqcup_{t \in D_0, u \in \mathcal{U}(\widehat{O})} utU$ for finitely many u and t (see [IAT, Chap. 3] or [MFG, Sect. 3.1.6]) and define

$$\mathbf{f}|[UyU](x) = \sum_{t,u} \epsilon(t)^{-1}\mathbf{f}(xut). \tag{14}$$

We check that this operator preserves the spaces of automorphic forms: $G_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$ and $S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C})$, and depends only on UyU not the choice of y as long as $y \in D_0$. However it depends on the choice of $\varpi_\mathfrak{q}$ as the character ϵ (extended to $\Delta_0(\mathfrak{n})$) depends on $\varpi_\mathfrak{q}$. This action for y with $y_\mathfrak{n} = 1$ is independent of the choice of the extension of ϵ to $T^\Delta(F_\mathbb{A})$. When $y_\mathfrak{n} \neq 1$, we may assume that $y_\mathfrak{n} \in D_0 \subset T^\Delta(F_\mathbb{A})$, and in this case, t can be chosen so that $t_\mathfrak{n} = y_\mathfrak{n}$ (so $t_\mathfrak{n}$ is independent of single right cosets in the double coset). If we extend ϵ to $T^\Delta(F_\mathbb{A}^{(\infty)})$ by choosing another prime element $\varpi'_\mathfrak{q}$ and write the extension as ϵ', then we have

$$\epsilon(t_\mathfrak{n})[UyU] = \epsilon'(t_\mathfrak{n})[UyU]',$$

where the operator on the right-hand side is defined with respect to ϵ'. Thus the sole difference is the root of unity $\epsilon(t_\mathfrak{n})/\epsilon'(t_\mathfrak{n}) \in \mathrm{Im}(\epsilon|_{T^\Delta(O/\mathfrak{n}')})$. Since it depends on the choice of $\varpi_\mathfrak{q}$, we make the choice once and for all, and write $T(\mathfrak{q})$ for $\left[U\left(\begin{smallmatrix} \varpi_\mathfrak{q} & 0 \\ 0 & 1 \end{smallmatrix}\right)U\right]$ (if $\mathfrak{q} \nmid \mathfrak{n}$), which coincides with $T(1, \mathfrak{q})$ in (4) if $\mathfrak{q} \nmid \mathfrak{n}'$. By linearity, these actions of double cosets extend to the ring action of the double coset ring $R(U, \Delta_0(\mathfrak{n}p^\alpha))$.

To introduce rationality of modular forms, we recall Fourier expansion of adelic modular forms (cf. [HMI, Proposition 2.26]). Recall the embedding $\iota_\infty : \overline{\mathbb{Q}} \hookrightarrow \mathbb{C}$, and identify $\overline{\mathbb{Q}}$ with the image of ι_∞. Recall also the differential idele $d \in F_\mathbb{A}^\times$ with $d^{(\mathfrak{d})} = 1$ and $d\widehat{O} = \mathfrak{d}\widehat{O}$. Each member $\mathbf{f}$ of $S_\kappa(U, \epsilon; \mathbb{C})$ has its Fourier expansion:

$$\mathbf{f}\left(\begin{smallmatrix} y & x \\ 0 & 1 \end{smallmatrix}\right) = |y|_\mathbb{A} \sum_{0 \ll \xi \in F} c(\xi y d, \mathbf{f})(\xi y_\infty)^{-\kappa_1}\mathbf{e}_F(i\xi y_\infty)\mathbf{e}_F(\xi x), \tag{15}$$

where $\mathbf{e}_F : F_{\mathbb{A}}/F \to \mathbb{C}^\times$ is the additive character with $\mathbf{e}_F(x_\infty) = \exp(2\pi i \sum_{\sigma\in I} x_\sigma)$ for $x_\infty = (x_\sigma)_\sigma \in \mathbb{R}^I = F \otimes_{\mathbb{Q}} \mathbb{R}$. Here $y \mapsto c(y, \mathbf{f})$ is a function defined on $y \in F_{\mathbb{A}}^\times$ only depending on its finite part $y^{(\infty)}$. The function $c(y, \mathbf{f})$ is supported by the set $(\widehat{O} \times F_\infty) \cap F_{\mathbb{A}}^\times$ of *integral* ideles.

Let $F[\kappa]$ be the field fixed by $\{\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/F) | \kappa\sigma = \kappa\}$, over which the character $\kappa \in X^*(T_F^\Delta)$ is rational. Write $O[\kappa]$ for the integer ring of $F[\kappa]$. We also define $O[\kappa, \epsilon]$ for the integer ring of the field $F[\kappa, \epsilon]$ generated by the values of ϵ over $F[\kappa]$. For any $F[\kappa, \epsilon]$-algebra A inside $\mathbb{C}$, we define

$$S_\kappa(U, \epsilon; A) = \left\{\mathbf{f} \in S_\kappa(U, \epsilon; \mathbb{C}) \middle| c(y, \mathbf{f}) \in A \text{ as long as } y \text{ is integral}\right\}. \tag{16}$$

As we have seen, we can interpret $S_\kappa(U, \epsilon; A)$ as the space of A-rational global sections of a line bundle of a variety defined over A; so, by the flat base-change theorem (e.g., [GME, Lemma 1.10.2]),

$$S_\kappa(\mathfrak{n}, \epsilon; A) \otimes_A \mathbb{C} = S_\kappa(\mathfrak{n}, \epsilon; \mathbb{C}). \tag{17}$$

The Hecke operators preserve A-rational modular forms (cf. (23) below). We define the Hecke algebra $h_\kappa(U, \epsilon; A) \subset \mathrm{End}_A(S_\kappa(U, \epsilon; A))$ by the A-subalgebra generated by the Hecke operators of $R(U, \Delta_0(\mathfrak{n}p^\alpha))$. Thus for any $\overline{\mathbb{Q}}_p$-algebras A, we may consistently define

$$S_\kappa(U, \epsilon; A) = S_\kappa(U, \epsilon; \overline{\mathbb{Q}}) \otimes_{\overline{\mathbb{Q}}, \iota_p} A. \tag{18}$$

By linearity, $y \mapsto c(y, \mathbf{f})$ extends to a function on $F_{\mathbb{A}}^\times \times S_\kappa(U, \epsilon; A)$ with values in A. For $u \in \widehat{O}^\times$, we know from [HMI, Proposition 2.26]

$$c(yu, \mathbf{f}) = \epsilon_1(u) c(y, \mathbf{f}). \tag{19}$$

If $\mathbf{f}$ is a normalized Hecke eigenform, its eigenvalue $a(y, \mathbf{f})$ of $T(y)$ is given by $\epsilon_1(y)^{-1} c(y, \mathbf{f})$ which depends only on the ideal $\mathfrak{y} := y\widehat{O} \cap F$ by the above formula as claimed in the introduction. We define the q-expansion coefficients (at p) of $\mathbf{f} \in S_\kappa(U, \epsilon; A)$ by

$$\mathbf{c}_p(y, \mathbf{f}) = y_p^{-\kappa_1} c(y, \mathbf{f}). \tag{20}$$

The formal q-expansion of an A-rational $\mathbf{f}$ has values in the space of functions on $F_{\mathbb{A}^{(\infty)}}^\times$ with values in the formal monoid algebra $A[[q^\xi]]_{\xi \in F_+}$ of the multiplicative semi-group F_+ made up of totally positive elements, which is given by

$$\mathbf{f}(y) = \mathcal{N}(y)^{-1} \sum_{\xi \gg 0} \mathbf{c}_p(\xi y d, \mathbf{f}) q^\xi, \tag{21}$$

where $\mathcal{N} : F_{\mathbb{A}}^\times / F^\times \to \overline{\mathbb{Q}}_p^\times$ is the character given by $\mathcal{N}(y) = y_p^{-I} |y^{(\infty)}|_{\mathbb{A}}^{-1}$.

We now define for any p-adically complete $O[\kappa,\epsilon]$-algebra A in $\mathbb{C}_p$

$$S_\kappa(U,\epsilon;A)=\left\{\mathbf{f}\in S_\kappa(U,\epsilon;\mathbb{C}_p)\middle|\mathbf{c}_p(y,\mathbf{f})\in A \text{ for integral } y\right\}. \tag{22}$$

As we have already seen, these spaces have geometric meaning as the space of A-integral global sections of a line bundle defined over A of the Hilbert modular variety of level U, and the q-expansion above for a fixed $y=y^{(\infty)}$ gives rise to the geometric q-expansion at the infinity cusp of the classical modular form $\mathbf{f}_x$ for $x=\left(\begin{smallmatrix} y & 0\\ 0 & 1\end{smallmatrix}\right)$ (see [H91, (1.5)] and [PAF, (4.63)]).

We have chosen a complete representative set $\{c_i\}_{i=1,\dots,h}$ in finite ideles for the strict idele class group $F^\times\backslash F^\times_{\mathbb{A}}/\widehat{O}^\times F^\times_{\infty+}$, where h is the strict class number of F. Let $\mathfrak{c}_i=c_iO$. Write $t_i=\begin{pmatrix} c_id^{-1} & 0\\ 0 & 1\end{pmatrix}$ and consider $\mathbf{f}_i=\mathbf{f}_{t_i}$ as defined in (S2). The collection $(\mathbf{f}_i)_{i=1,\dots,h}$ determines f, because of the approximation theorem. Then $\mathbf{f}(c_id^{-1})$ gives the q-expansion of $\mathbf{f}_i$ at the Tate abelian variety with $\mathfrak{c}_i$-polarization $\mathrm{Tate}_{\mathfrak{c}_i^{-1},O}(q)$ ($\mathfrak{c}_i=c_iO$). By (q-exp), the q-expansion $\mathbf{f}(y)$ determines $\mathbf{f}$ uniquely.

We write $T(y)$ for the Hecke operator acting on $S_\kappa(U,\epsilon;A)$ corresponding to the double coset $U\left(\begin{smallmatrix} y & 0\\ 0 & 1\end{smallmatrix}\right)U$ for an integral idele y. We renormalize $T(y)$ to have a p-integral operator $\mathbb{T}(y)$: $\mathbb{T}(y)=y_p^{-\kappa_1}T(y)$. Since this only affects $T(y)$ with $y_p\neq 1$, $\mathbb{T}(\mathfrak{q})=T(\varpi_{\mathfrak{q}})=T(\mathfrak{q})$ if $\mathfrak{q}\nmid p$. However depending on weight, we can have $\mathbb{T}(\mathfrak{p})\neq T(\mathfrak{p})$ for primes $\mathfrak{p}|p$. The renormalization is optimal to have the stability of the A-integral spaces under Hecke operators. We define $\langle\mathfrak{q}\rangle=N(\mathfrak{q})T(\mathfrak{q},\mathfrak{q})$ with $T(\mathfrak{q},\mathfrak{q})=[U\varpi_{\mathfrak{q}}U]$ for $\mathfrak{q}\nmid\mathfrak{n}'p^\alpha$ ($\mathfrak{n}'=\mathfrak{n}_1\cap\mathfrak{n}_2$), which is equal to the central action of a prime element $\varpi_{\mathfrak{q}}$ of $O_{\mathfrak{q}}$ times $N(\mathfrak{q})=|\varpi_{\mathfrak{q}}|_{\mathbb{A}}^{-1}$. We have the following formula of the action of $\mathbb{T}(\mathfrak{q})$ (e.g., [HMI, (2.3.21)] or [PAF, Sect. 4.2.10]):

$$\mathbf{c}_p(y,\mathbf{f}|\mathbb{T}(\mathfrak{q}))=\begin{cases}\mathbf{c}_p(y\varpi_{\mathfrak{q}},\mathbf{f})+\mathbf{c}_p(y\varpi_{\mathfrak{q}}^{-1},\mathbf{f}|\langle\mathfrak{q}\rangle) & \text{if } \mathfrak{q}\nmid\mathfrak{n}p\\ \mathbf{c}_p(y\varpi_{\mathfrak{q}},\mathbf{f}) & \text{otherwise,}\end{cases} \tag{23}$$

where the level $\mathfrak{n}$ of U is the ideal maximal under the condition: $U_1^1(\mathfrak{n})\subset U\subset U_0(\mathfrak{n})$. Thus $\mathbb{T}(\varpi_{\mathfrak{q}})=(\varpi_{\mathfrak{q}})_p^{-\kappa_1}U(\mathfrak{q})$ when $\mathfrak{q}$ is a factor of the level of U (even when $\mathfrak{q}|p$; see [PAF, (4.65–66)]). Writing the level of U as $\mathfrak{n}p^\alpha$, we assume

$$\text{either } p|\mathfrak{n}p^\alpha \text{ or } [\kappa]\geq 0, \tag{24}$$

since $\mathbb{T}(\mathfrak{q})$ and $\langle\mathfrak{q}\rangle$ preserve the space $S_\kappa(U,\epsilon;A)$ under this condition (see [PAF, Theorem 4.28]). We define the Hecke algebra $h_\kappa(U,\epsilon;A)$ [resp. $h_\kappa(\mathfrak{n},\epsilon_+;A)$] with coefficients in A by the A-subalgebra of the A-linear endomorphism algebra $\mathrm{End}_A(S_\kappa(U,\epsilon;A))$ [resp. $\mathrm{End}_A(S_\kappa(\mathfrak{n},\epsilon_+;A))$] generated by the action of the finite group $U_0(\mathfrak{n}p^\alpha)/U$, $\mathbb{T}(\mathfrak{q})$ and $\langle\mathfrak{q}\rangle$ for all $\mathfrak{q}$.

1.9 Hecke Algebras

We have canonical projections:

$$R(U_1^1(\mathfrak{n}p^\alpha, \Delta_0(\mathfrak{n}p^\alpha)) \twoheadrightarrow R(U, \Delta_0(\mathfrak{n}p^\alpha)) \twoheadrightarrow R(U_0(\mathfrak{n}p^\beta), \Delta_0(\mathfrak{n}p^\beta))$$

for all $\alpha \geq \beta$ taking canonical generators to the corresponding ones, which are compatible with inclusions

$$S_\kappa(U_0(\mathfrak{n}p^\beta), \epsilon; A) \hookrightarrow S_\kappa(U, \epsilon; A) \hookrightarrow S_\kappa(U_1^1(\mathfrak{n}p^\alpha), \epsilon; A).$$

We decompose $O_p^\times = \mathbf{\Gamma} \times \Delta$ as in the introduction and hence $\mathbf{G} = \mathbf{\Gamma} \times \Delta \times (O/\mathfrak{n}')^\times$. We fix κ and ϵ_+ and the initial $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_+)$. We suppose that ϵ_j $(j = 1, 2)$ factors through $\mathbf{G}/\mathbf{\Gamma} = \Delta \times (O/\mathfrak{n}')^\times$ for $\mathfrak{n}'$ prime to p. We write $\mathfrak{n}$ for a factor of $\mathfrak{n}'$ such that ϵ^- is defined modulo $\mathfrak{n}p^{r_0+I_p}$ for $p^{r_0+I_p} = \prod_{\mathfrak{p}|p} \mathfrak{p}^{r_{0,\mathfrak{p}}+1}$ for a multi-index $r_0 = (r_{0,\mathfrak{p}})_\mathfrak{p}$ with $\mathfrak{p}$ running over prime factors of p. Then we get a projective system of Hecke algebras $\{h_\kappa(U, \epsilon; A)\}_U$ (U running through open subgroups of $U_0(\mathfrak{n}p^{r_0+1})$ containing $U_1^1(\mathfrak{n}p^\infty)$), whose projective limit (when $\kappa_2 - \kappa_1 \geq I$) gives rise to the universal Hecke algebra $\mathbf{h}(\mathfrak{n}, \epsilon; A)$ for a complete p-adic algebra A. We have a continuous character $T : \widehat{O}^\times \to \mathbf{h}(\mathfrak{n}, \epsilon; A)$ given by $u \mapsto T(u)$ where $\mathbf{f}|T(u)(x) = \epsilon_1(u)^{-1}\mathbf{f}\left(x\left(\begin{smallmatrix} u & 0 \\ 0 & 1 \end{smallmatrix}\right)\right)$ for $u \in \widehat{O}^\times$ (here $T(u)$ is the Hecke operator $T(y)$ taking $y = u$ as the double coset $U\left(\begin{smallmatrix} u & 0 \\ 0 & 1 \end{smallmatrix}\right)U$ is equal to the single coset $U\left(\begin{smallmatrix} u & 0 \\ 0 & 1 \end{smallmatrix}\right)$). This character T factors through $\mathbf{\Gamma} = \mathbf{G}/(\Delta \times (O/\mathfrak{n}')^\times)$ and induces a canonical algebra structure of $\mathbf{h}(\mathfrak{n}, \epsilon; A)$ over $A[[\mathbf{\Gamma}]]$.

Let W be a sufficiently large complete discrete valuation ring inside $\overline{\mathbb{Q}}_p$ (as before). Define $W[\epsilon] \subset \overline{\mathbb{Q}}_p$ by the W-subalgebra generated by the values of ϵ (over the finite adeles). It has canonical generators $\mathbb{T}(y)$ over $\mathbf{\Lambda} = W[[\mathbf{\Gamma}]]$. Here note that the operator $\langle \mathfrak{q} \rangle$ acts via multiplication by $N(\mathfrak{q})\epsilon_+(\mathfrak{q})$ for the fixed central character ϵ_+, where $N(\mathfrak{q}) = |O/\mathfrak{q}|$.

The (nearly) p-ordinary projector $e = \lim_n \mathbb{T}(p)^{n!}$ gives an idempotent of the Hecke algebras $h_\kappa(U, \epsilon; W)$, $h_\kappa(\mathfrak{n}p^\alpha, \epsilon_+; W)$ and $\mathbf{h}(\mathfrak{n}, \epsilon_+; W)$. By adding superscript "n.ord," we indicate the algebra direct summand of the corresponding Hecke algebra cut out by e; e.g., $h_\kappa^{\text{n.ord}}(\mathfrak{n}p^\alpha, \epsilon_+; W) = e(h_\kappa(\mathfrak{n}p^\alpha, \epsilon_+; W))$. We simply write $\mathbf{h}$ for $\mathbf{h}^{\text{n.ord}} = \mathbf{h}^{\text{n.ord}}(\mathfrak{n}, \epsilon_+; W)$. The algebra $\mathbf{h}^{\text{n.ord}}$ is by definition the universal nearly p-ordinary Hecke algebra over $\mathbf{\Lambda}$ of level $\mathfrak{n}p^\infty$ with "Neben character" ϵ. This algebra $\mathbf{h}^{\text{n.ord}}(\mathfrak{n}, \epsilon; W)$ is exactly the one $\mathbf{h}(\psi^+, \psi')$ employed in [HT93, p. 240] (note that in [HT93] we assumed $\kappa_1 \geq \kappa_2$ reversing our normalization here).

The algebra $\mathbf{h}^{\text{n.ord}}(\mathfrak{n}, \epsilon; W)$ is a torsion-free $\mathbf{\Lambda}$-algebra of finite rank. Take a point $P \in \mathrm{Spf}(\mathbf{\Lambda})(\overline{\mathbb{Q}}_p)$. If P is arithmetic, $\epsilon_P = P\kappa(P)^{-1}$ is a character of $\mathbf{\Gamma}$. By abusing a symbol, we write ϵ_P for the character $(\epsilon_{P,1}, \epsilon_{P,2}, \epsilon_+)$ given by $\epsilon_{P,j}$ on $\mathbf{\Gamma}$ and ϵ_j on $\Delta \times (O/\mathfrak{n}')^\times$. Writing the conductor of $\epsilon_P^-|_{O_p^\times}$ as $p^{f(P)}$, we define $r(P) \geq 0$ by $p^{r(P)+I_p} = p^{f(P)} \cap \mathfrak{p}$. Here $r(P)$ is an element of $\mathbb{Z}[I_p]$; so, $r(P) = \sum_{\mathfrak{p}|p} r(P)_\mathfrak{p}\mathfrak{p}$ indexed by prime factors $\mathfrak{p}|p$, and we write I_p for $\{1\}_{\mathfrak{p}|p}$. Therefore $r(P) + I_p = \sum_\mathfrak{p}(r(P)_\mathfrak{p} + 1)\mathfrak{p}$. As long as P is arithmetic, we have a canonical specialization morphism:

$$\mathbf{h}^{\text{n.ord}}(\mathfrak{n}, \epsilon_+; W) \otimes_{\mathbf{\Lambda},P} W[\epsilon_P] \twoheadrightarrow h^{\text{n.ord}}_{\kappa(P)}(\mathfrak{n}p^{r(P)+I_p}, \epsilon_+; W[\epsilon_P]),$$

which is an isogeny and is an isomorphism if $\mathbf{h}^{\text{n.ord}}(\mathfrak{n}, \epsilon_+; W)$ is $\mathbf{\Lambda}$-free [PAF, Sect. 4.2.11] (note in [PAF] the order of κ_j is reversed so that $\kappa_1 > \kappa_2$). The specialization morphism takes the generators $\mathbb{T}(y)$ to $\mathbb{T}(y)$.

1.10 Analytic Families of Hecke Eigenforms

In summary, for a fixed κ and ϵ_+, we have the algebra $\mathbf{h} = \mathbf{h}^{\text{n.ord}}(\mathfrak{n}, \epsilon_+; W)$ characterized by the following two properties:

(C1) $\mathbf{h}$ is torsion-free of finite rank over $\mathbf{\Lambda}$ equipped with $\mathbb{T}(\mathfrak{l}) = \mathbb{T}(\varpi_{\mathfrak{l}}), \mathbb{T}(y) \in \mathbf{h}$ for all primes $\mathfrak{l}$ prime to p and $y \in O_p \cap F_p^\times$,

(C2) if $\kappa_2 - \kappa_1 \geq I$ and P is an arithmetic point of $\text{Spec}(\mathbf{\Lambda})(\overline{\mathbb{Q}}_p)$, we have a surjective W-algebra homomorphism: $\mathbf{h} \otimes_{\mathbf{\Lambda},P} W[\epsilon_P]) \to h^{\text{n.ord}}_{\kappa(P)}(\mathfrak{n}p^{r(P)+I_p}, \epsilon_+; W[\epsilon_P])$ with finite kernel, sending $\mathbb{T}(\mathfrak{l}) \otimes 1$ to $\mathbb{T}(\mathfrak{l})$ (and $\mathbb{T}(y) \otimes 1$ to $\mathbb{T}(y)$).

Actually, if $p \geq 5$ and $p \nmid |\Delta|$, in (C1), quite plausibly, $\mathbf{h}$ would be free over $\mathbf{\Lambda}$ (not just torsion-free), and we would have an isomorphism in (C2) (this fact holds true under unramifiedness of $p \geq 5$ in $F/\mathbb{Q}$; see [PAF, Corollary 4.31]), but we do not need this stronger fact.

By fixing an isomorphism $\mathbf{\Gamma} \cong \mathbb{Z}_p^m$ with $m = [F_p : \mathbb{Q}_p]$, we have identified $\mathbf{\Lambda} = \mathbf{\Lambda}_W$ with $W[[T_1, \ldots, T_m]]$ for $\{t_i = 1 + T_i\}_{i=1,\ldots,m}$ corresponding to a $\mathbb{Z}_p$-basis $\{\gamma_i\}_{i=1,\ldots,m}$ of $\mathbf{\Gamma}$. Regard κ_2 as a character of $O_p^\times$ whose value at $\gamma \in O_p^\times$ is

$$\gamma^{\kappa_2} = \prod_{\sigma \in I} \sigma(\gamma)^{\kappa_{2,\sigma}}.$$

We may write an arithmetic prime P as a prime $\mathbf{\Lambda}$-ideal

$$P = (t_i - \epsilon_2(\gamma_i)^{-1}\gamma_i^{\kappa_2})\mathbf{\Lambda}_{W[\epsilon]} \cap \mathbf{\Lambda}_W.$$

When $\kappa_2 = kI$ for an integer k, $\gamma \mapsto \gamma^{\kappa_2}$ is given by $\gamma \mapsto N(\gamma)^k$ for the norm map $N = N_{F_p/\mathbb{Q}_p}$ on $O_p^\times$. For a point $P \in \text{Spec}(\mathbf{\Lambda})(\overline{\mathbb{Q}}_p)$ killing $(t_i - \zeta_i^{-1}\gamma_i^{\kappa_2})$ for $\zeta_i \in \mu_{p^\infty}(W)$, we make explicit the character ϵ_P. First we define a character $\epsilon_{P,2,\mathbf{\Gamma}} : O_p^\times \to \mu_{p^\infty}(W)$ factoring through $\mathbf{\Gamma} = O_p^\times/\Delta$ by $\epsilon_{P,2,\mathbf{\Gamma}}(\gamma_i) = \zeta_i$ for all i. Then for the fixed ϵ_+, we put $\epsilon_{P,1,\mathbf{\Gamma}} = (\epsilon_+|_{\mathbf{\Gamma}})\epsilon_{P,2,\mathbf{\Gamma}}^{-1}$. With the fixed data $\epsilon_1^{(\mathbf{\Gamma})} := \epsilon_1|_{(O/\mathfrak{n}')^\times \times \Delta}$ and $\epsilon_2^{(\mathbf{\Gamma})} := \epsilon_2|_{(O/\mathfrak{n}')^\times \times \Delta}$, we put $\epsilon_{P,j} = \epsilon_{j,P,\mathbf{\Gamma}}\epsilon_j^{(\mathbf{\Gamma})}$. In this way, we form $\epsilon_P = (\epsilon_{P,1}, \epsilon_{P,2}, \epsilon^+)$.

Let $\text{Spec}(\mathbb{I})$ be a reduced irreducible component $\text{Spec}(\mathbb{I}) \subset \text{Spec}(\mathbf{h})$. Since $\mathbf{h}$ is torsion-free of finite rank over $\mathbf{\Lambda}$, $\text{Spec}(\mathbb{I})$ is a finite torsion-free covering of $\text{Spec}(\mathbf{\Lambda})$. Write $a(y)$ and $a(\mathfrak{l})$ for the image of $T(y)$ and $T(\mathfrak{l})$ in $\mathbb{I}$ (so, $a(\varpi_{\mathfrak{p}})$ is the image of $T(\varpi_{\mathfrak{p}})$). We also write $\mathbf{a}(y)$ for the image of $\mathbb{T}(y)$; so, $\mathbf{a}(y) = y_p^{-\kappa_1} a(y)$. If $P \in \mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$ induces an arithmetic point P_0 of $\mathrm{Spec}(\boldsymbol{\Lambda})$, we call it again an *arithmetic* point of $\mathrm{Spec}(\mathbb{I})$, and put $\kappa_j(P) = \kappa_j(P_0)$. If P is arithmetic, by (C2), we have a Hecke eigenform $\mathbf{f}_P \in S_{\kappa(P)}(U_0(\mathfrak{n}p^{r(P)+I_p}), \epsilon_P; \overline{\mathbb{Q}}_p)$ such that its eigenvalue for $\mathbb{T}(\mathfrak{l})$ and $\mathbb{T}(y)$ is given by $a_P(\mathfrak{l}) := P(a(\mathfrak{l})), a_P(y) := P(a(y)) \in \overline{\mathbb{Q}}_p$ for all $\mathfrak{l}$ and $y \in F_p^\times$. Thus $\mathbb{I}$ gives rise to a family $\mathcal{F} = \mathcal{F}_{\mathbb{I}} = \{\mathbf{f}_P | \text{arithmetic } P \in \mathrm{Spec}(\mathbb{I})\}$ of classical Hecke eigenforms. We call this family a *p-adic analytic family of p-slope* 0 (with coefficients in $\mathbb{I}$) associated with an irreducible component $\mathrm{Spec}(\mathbb{I}) \subset \mathrm{Spec}(\mathbf{h})$. There is a sub-family corresponding to any closed integral subscheme $\mathrm{Spec}(\mathbb{J}) \subset \mathrm{Spec}(\mathbb{I})$ as long as $\mathrm{Spec}(\mathbb{J})$ has densely populated arithmetic points. Abusing our language slightly, for any covering $\pi : \mathrm{Spec}(\widetilde{\mathbb{I}}) \twoheadrightarrow \mathrm{Spec}(\mathbb{I})$, we will consider the pulled back family $\mathcal{F}_{\widetilde{\mathbb{I}}} = \{\mathbf{f}_P = \mathbf{f}_{\pi(P)} | \text{arithmetic } P \in \mathrm{Spec}(\widetilde{\mathbb{I}})\}$. The choice of $\widetilde{\mathbb{I}}$ is often the normalization of $\mathbb{I}$ or the integral closure of $\mathbb{I}$ in a finite extension of the quotient field of $\mathbb{I}$.

Identify $\mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$ with $\mathrm{Hom}_{W\text{-alg}}(\mathbb{I}, \overline{\mathbb{Q}}_p)$ so that each element $a \in \mathbb{I}$ gives rise to a "function" $a : \mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p) \to \overline{\mathbb{Q}}_p$ whose value at $(P : \mathbb{I} \to \overline{\mathbb{Q}}_p) \in \mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$ is $a_P := P(a) \in \overline{\mathbb{Q}}_p$. Then a is an analytic function of the rigid analytic space associated with $\mathrm{Spf}(\mathbb{I})$. We call such a family p-slope 0 because $|\mathbf{a}_P(\varpi_{\mathfrak{p}})|_p = 1$ for the p-adic absolute value $|\cdot|_p$ of $\overline{\mathbb{Q}}_p$ for all $\mathfrak{p}|p$ (it is also called a p-ordinary family).

1.11 Modular Galois Representations

Each (reduced) irreducible component $\mathrm{Spec}(\mathbb{I})$ of the Hecke spectrum $\mathrm{Spec}(\mathbf{h})$ has a 2-dimensional semi-simple (actually absolutely irreducible) continuous representation $\rho_{\mathbb{I}}$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/F)$ with coefficients in the quotient field of $\mathbb{I}$ (see [H86a] and [H89]). The representation $\rho_{\mathbb{I}}$ restricted to the $\mathfrak{p}$-decomposition group $D_{\mathfrak{p}}$ (for each prime factor $\mathfrak{p}|p$) is reducible (see [HMI, Sect. 2.3.8]). Define the p-adic avatar $\widehat{\epsilon}_+ : (F_{\mathbb{A}}^{(\infty)})^\times / F^\times \to \overline{\mathbb{Q}}_p^\times$ by $\widehat{\epsilon}_+(y) = \epsilon_+(y) y_p^{I-\kappa_1-\kappa_2}$ (note here $y_\infty = 1$ as $F_{\mathbb{A}}^{(\infty)}$ is made of finite adales in $F_{\mathbb{A}}$). We write $\rho_{\mathbb{I}}^{ss}$ for its semi-simplification over D_p. As is well known now (e.g., [HMI, Sect. 2.3.8]), $\rho_{\mathbb{I}}$ is unramified outside $\mathfrak{n}p$ and satisfies

$$\mathrm{Tr}(\rho_{\mathbb{I}}(Frob_{\mathfrak{l}})) = a(\mathfrak{l}) \ \text{ for all prime } \mathfrak{l} \nmid p\mathfrak{n}. \tag{Gal}$$

By (Gal) and Chebotarev density, $\mathrm{Tr}(\rho_{\mathbb{I}})$ has values in $\mathbb{I}$; so, for any integral closed subscheme $\mathrm{Spec}(\mathbb{J}) \subset \mathrm{Spec}(\mathbb{I})$ with projection $\pi : \mathbb{I} \to \mathbb{J}$, $\pi \circ \mathrm{Tr}(\rho_{\mathbb{I}}) : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathbb{J}$ gives rise to a pseudo-representation of Wiles (e.g., [MFG, Sect. 2.2]). Then by a theorem of Wiles, we can make a unique 2-dimensional semi-simple continuous representation $\rho_{\mathbb{J}} : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathrm{GL}_2(Q(\mathbb{J}))$ unramified outside $\mathfrak{n}p$ with $\mathrm{Tr}(\rho_{\mathbb{J}}(Frob_{\mathfrak{l}})) = \pi(a(\mathfrak{l}))$ for all primes $\mathfrak{l} \nmid \mathfrak{n}p$, where $Q(\mathbb{J})$ is the quotient field of $\mathbb{J}$. If $\mathrm{Spec}(\mathbb{J})$ is one point $P \in \mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$, we write ρ_P for $\rho_{\mathbb{J}}$. This is the Galois representation associated with the Hecke eigenform $\mathbf{f}_P$ (given in [H89]). As for p-ramification, the restriction of $\rho_{\mathbb{I}}$ to the decomposition group at a prime

$\mathfrak{p}|p$ is reducible. Taking $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}_p/F_\mathfrak{p})$ whose restriction to the maximal abelian extension of $F_\mathfrak{p}$ is the Artin symbol $[u, F_\mathfrak{p}]$, we have by Hida [H89]

$$\rho_P(\sigma) \sim \begin{pmatrix} \epsilon_{2,P}(u)u^{-\kappa_2} & * \\ 0 & \epsilon_{1,P}(u)u^{-\kappa_1} \end{pmatrix} \text{ for } u \in O_\mathfrak{p}^\times \text{ and } \rho_P(\sigma) \sim \begin{pmatrix} * & * \\ 0 & \mathbf{a}_P(u) \end{pmatrix} \text{ for } u \in O_\mathfrak{p} - \{0\}. \tag{Ram}$$

Thus $[u, F_\mathfrak{p}] \mapsto \epsilon_{1,P}(u)u^{-\kappa_1}$ is the quotient character at $\mathfrak{p}$ (and in this way, ϵ_j ($j = 1, 2$) are ordered).

1.12 CM Theta Series

Following the description in [H06, Sect. 6.2], we construct CM theta series with p-slope 0 and describe the CM component which gives rise to such theta series (the construction was first made in [HT93]). We first recall a cusp form $\mathbf{f}$ on $G(\mathbb{A})$ with complex multiplication by a CM field M top down without much proof. By computing its classical Fourier expansion, we can confirm that $\mathbf{f}$ is a cusp form. Let M/F be a CM field with integer ring O_M and choose a CM type Σ:

$$I_M = \mathrm{Hom}_{\mathrm{field}}(M, \overline{\mathbb{Q}}) = \Sigma \sqcup \Sigma c$$

for complex conjugation c. To assure the p-slope 0 condition, we need to assume that the CM type Σ is p-ordinary, that is, the set Σ_p of p-adic places induced by $\iota_p \circ \sigma$ for $\sigma \in \Sigma$ is disjoint from Σ_{pc} (its conjugate by the generator c of $\mathrm{Gal}(M/F)$). The existence of such a p-ordinary CM type implies that each prime factor $\mathfrak{p}|p$ of F split in M/F. Thus the set $I_{M,p}$ of p-adic places of M is given by $\Sigma_p \sqcup \Sigma_p^c$. Write $\mathfrak{p} = \mathfrak{P}\mathfrak{P}^c$ in O_M for two primes $\mathfrak{P} \neq \mathfrak{P}^c$ such that $\mathfrak{P} \in \Sigma_p$ is induced by $\iota_p \circ \sigma$ on M for $\sigma \in \Sigma$. For each $k \in \mathbb{Z}[I]$ and $X = \Sigma, I_M$, we write $kX = \sum_{\sigma \in X} k_{\sigma|F}\sigma$.

We choose $\kappa_2 - \kappa_1 \geq I$ with $\kappa_1 + \kappa_2 = [\kappa]I$ for an integer $[\kappa]$. We then choose a Hecke ideal character λ of conductor $\mathfrak{C}\mathfrak{P}^e$ ($\mathfrak{C}$ prime to $\mathfrak{p}$) such that

$$\lambda((\alpha)) = \alpha^{c\kappa_1\Sigma + \kappa_2\Sigma} \text{ for } \alpha \in M^\times \text{ with } \alpha \equiv 1 \mod \mathfrak{C}\mathfrak{P}^e O_{M,\mathfrak{C}\mathfrak{P}^e} \text{ in } \prod_{\mathfrak{l}|\mathfrak{C}\mathfrak{P}^e} M_\mathfrak{l},$$

where $\mathfrak{P}^e = \prod_{\mathfrak{P} \in \Sigma_p} \mathfrak{P}^{e(\mathfrak{P})}\mathfrak{P}^{ce(\mathfrak{P}^c)}$ for $e = \sum_{\mathfrak{P} \in \Sigma_p}(e(\mathfrak{P})\mathfrak{P} + e(\mathfrak{P}^c)\mathfrak{P}^c)$ and $O_{M,\mathfrak{a}} = \prod_{\mathfrak{l}|\mathfrak{a}} O_{M_\mathfrak{l}}$ for an integral ideal $\mathfrak{a}$ of O_M.

We now recall a very old idea of Weil (and history) to lift the ideal character λ to an "idele" Hecke character: $\tilde{\lambda} : M_\mathbb{A}^\times/M^\times \to \mathbb{C}^\times$ following to Weil (who invented this identification of two types of Hecke characters in [W55] as a part of the theory of complex multiplication of abelian varieties, established by himself together with Shimura and Taniyama in the Tokyo–Nikko symposium in 1955). For the moment, we write $\tilde{\lambda}$ for the lifted idele character following [W55], but once it is defined, we just write simply λ for the idele and the ideal characters removing the tilde "~", following the more recent tradition. We write $(M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times := \{x \in M_{\mathbb{A}}^\times | x_\infty = x_{\mathfrak{l}} = 1\}$ for all primes $\mathfrak{l}|\mathfrak{C}\mathfrak{P}^e$. For an idele $x \in (M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times$ whose $\mathfrak{C}\mathfrak{P}^e\infty$-component is trivial, we require $\tilde{\lambda}(x) := \lambda(xO_M)$, where $xO_M = M \cap x\widehat{O}_M$ inside $(M_{\mathbb{A}}^{(\infty)})^\times = \{x \in M_{\mathbb{A}}^\times | x_\infty = 1\}$ (which is a fractional ideal prime to $\mathfrak{C}\mathfrak{P}^e$). At the infinity component $M_\infty^\times = (M \otimes_{\mathbb{A}} \mathbb{R})^\times = \prod_{\sigma\in\Sigma} \mathbb{C}^\times$, for $x_\infty = (x_\sigma)_{\sigma\in\Sigma}$ requiring

$$\tilde{\lambda}(x_\infty) = x_\infty^{-\kappa_2\Sigma - c\kappa_1\Sigma} := \prod_{\sigma\in\Sigma} x_\sigma^{-\kappa_{2,\sigma} - c\kappa_{2,\sigma c}},$$

we get a continuous character $\tilde{\lambda} : (M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times \times M_\infty^\times \to \mathbb{C}^\times$. We consider $M^\times (M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times M_\infty^\times \subset M_{\mathbb{A}}^\times$ which is a dense subgroup of $M_{\mathbb{A}}^\times$, and in particular, we have $M_{\mathbb{A}}^\times = U(\mathfrak{C}\mathfrak{P}^e)(M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times M_\infty^\times$, where $U(\mathfrak{a}) = \widehat{O}_M^\times \cap 1 + \mathfrak{a}\widehat{O}_M$ for an O_M-ideal $\mathfrak{a}$. We can extend $\tilde{\lambda}$ to the entire idele group $M_{\mathbb{A}}^\times$ so that $\tilde{\lambda}(M^\times) = 1$. To verify this point, we only need to show $\tilde{\lambda}(\alpha) = 1$ for $\alpha \in M^\times \cap U(\mathfrak{C}\mathfrak{P}^e)M_\infty^\times$ inside $M_{\mathbb{A}}^\times$. Since the $\mathfrak{C}\mathfrak{P}^e$ component of $\alpha \in M_{\mathbb{A}}^\times$ is in $U(\mathfrak{C}\mathfrak{P}^e)$, we check $\alpha_{\mathfrak{C}\mathfrak{P}^e} \equiv 1 \mod \mathfrak{C}\mathfrak{P}^e$, and hence, writing $(\alpha) = xO_M$ for $x = \alpha^{(\mathfrak{C}\mathfrak{P}^e\infty)} \in (M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times$ (the projection of $\alpha \in M_{\mathbb{A}}^\times$ to $(M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times$), we have $\tilde{\lambda}(x\alpha_\infty) = \lambda((\alpha))\alpha^{-\kappa_2\Sigma - c\kappa_1\Sigma} = 1$. By continuity, this extension $\tilde{\lambda}$ of λ to the dense subgroup $M^\times (M_{\mathbb{A}}^{(\mathfrak{C}\mathfrak{P}^e\infty)})^\times M_\infty^\times$ extends uniquely to the entire idele group $M_{\mathbb{A}}^\times$ which is trivial on $M^\times U(\mathfrak{C}\mathfrak{P}^e)$. Hereafter, we just use the symbol λ for $\tilde{\lambda}$ (as identifying the ideal character λ with the corresponding idele character $\tilde{\lambda}$).

If we need to indicate that $\mathfrak{C}$ is the prime-to-$\mathfrak{p}$ conductor of λ, we write $\mathfrak{C}(\lambda)$ for $\mathfrak{C}$. We also decompose $\mathfrak{C} = \prod_{\mathfrak{L}} \mathfrak{L}^{e(\mathfrak{L})}$ for prime ideals $\mathfrak{L}$ of M. We extend λ to a p-adic idele character $\widehat{\lambda} : M_{\mathbb{A}}^\times / M^\times M_\infty^\times \to \overline{\mathbb{Q}}_p^\times$ so that $\widehat{\lambda}(a) = \lambda(aO)a_p^{-\kappa_2\Sigma - c\kappa_1\Sigma}$. By class field theory, for the topological closure $\overline{M^\times M_\infty^\times}$ in $M_{\mathbb{A}}^\times$, $M_{\mathbb{A}}^\times / \overline{M^\times M_\infty^\times}$ is canonically isomorphic to the Galois group of the maximal abelian extension of M; so, this is the first occurrence in the history (again due to Weil [W55]) of the correspondence between an automorphic representation $\lambda = \tilde{\lambda}$ of $GL_1(M_{\mathbb{A}})$ and the Galois representation $\widehat{\lambda}$. Pulling back to $\mathrm{Gal}(\overline{F}/M)$, we may regard $\widehat{\lambda}$ as a character of $\mathrm{Gal}(\overline{F}/M)$. Any character φ of $\mathrm{Gal}(\overline{F}/M)$ of the form $\widehat{\lambda}$ as above is called "of weight κ". For a prime ideal $\mathfrak{L}$ of M outside p, we write $\lambda_{\mathfrak{L}}$ for the restriction of $\widehat{\lambda}$ to $M_{\mathfrak{L}}^\times$; so, $\lambda_{\mathfrak{L}}(x) = \widehat{\lambda}(x) = \lambda(x)$ for $x \in M_{\mathfrak{L}}^\times$. For a prime ideal $\mathfrak{P}|p$ of M, we put $\lambda_{\mathfrak{P}}(x) = \widehat{\lambda}(x)x^{\kappa_2\Sigma + c\kappa_1\Sigma} = \lambda(x)$ for $x \in M_{\mathfrak{P}}^\times$. In particular, for the prime $\mathfrak{P}|\mathfrak{p}$ with $\mathfrak{P} \in \Sigma_p$, we have $\lambda_{\mathfrak{P}}(x) = \widehat{\lambda}(x)x^{\kappa_2\Sigma_{\mathfrak{p}}}$ for $x \in M_{\mathfrak{P}}^\times$, and $\lambda_{\mathfrak{P}^c}(x) = \widehat{\lambda}(x)x^{c\kappa_1\Sigma_{\mathfrak{p}}}$ for $x \in M_{\mathfrak{P}^c}^\times$. Then $\lambda_{\mathfrak{L}}$ for all prime ideals $\mathfrak{L}$ (including those above p) is a continuous character of $M_{\mathfrak{L}}^\times$ with values in $\overline{\mathbb{Q}}$ whose restriction to the $\mathfrak{L}$-adic completion $O_{M,\mathfrak{L}}^\times$ of O_M is of finite order. By the condition $\kappa_1 \neq \kappa_2$, $\widehat{\lambda}$ cannot be of the form $\widehat{\lambda} = \phi \circ N_{M/F}$ for an idele character $\phi : F_{\mathbb{A}}^\times / F^\times F_{\infty+}^\times \to \overline{\mathbb{Q}}_p^\times$.

We define a function $(F_{\mathbb{A}}^{(\infty)})^\times \ni y \mapsto c(y, \theta(\lambda))$ supported by integral ideles by

$$c(y, \theta(\lambda)) = \sum_{x\in(M_{\mathbb{A}}^{(\infty)})^\times, xx^c = y} \lambda(x) \quad \text{if } y \text{ is integral,} \tag{25}$$

where x runs over elements in $M_{\mathbb{A}}^{\times}/(\widehat{O}_M^{(\mathfrak{C}\mathfrak{P}^e)})^{\times}$ satisfying the following four conditions: (0) $x_\infty = 1$, (1) xO_M is an integral ideal of M, (2) $N_{M/F}(x) = y$ and (3) $x_{\mathfrak{Q}} = 1$ for prime factors $\mathfrak{Q}$ of the conductor $\mathfrak{C}\mathfrak{P}^e$. The q-expansion determined by the coefficients $c(y, \theta(\lambda))$ gives a unique element $\theta(\lambda) \in S_\kappa(\mathfrak{n}_\theta, \epsilon'_\lambda; \overline{\mathbb{Q}})$ ([HT93, Theorem 6.1] and [HMI, Theorem 2.72]), where $\mathfrak{n}_\theta = N_{M/F}(\mathfrak{C}\mathfrak{P}^e)d(M/F)$ for the discriminant $d(M/F)$ of M/F and ϵ'_λ is a suitable "Neben" character. We have

(C) The central character $\epsilon_{\lambda+}$ of the automorphic representation $\pi(\lambda)$ generated by $\theta(\lambda)$ is given by the product: $x \mapsto \lambda(x)|x|_{\mathbb{A}}\left(\frac{M/F}{x}\right)$ for $x \in F_{\mathbb{A}}^{\times}$ and the quadratic character $\left(\frac{M/F}{}\right)$ of the CM quadratic extension M/F.

Recall here that $\lambda : M_{\mathbb{A}}^{\times} \to \mathbb{C}^{\times}$ is trivial on $M^{\times}$ as $\lambda_\infty(x_\infty) = x_\infty^{-\kappa_2\Sigma - c\kappa_1\Sigma}$, and hence $\epsilon_{\lambda+}$ is a continuous character of the idele class group $F_{\mathbb{A}}^{\times}/F^{\times}$.

We describe the Neben character $\epsilon_\lambda = (\epsilon_{\lambda,1}, \epsilon_{\lambda,2}, \epsilon_{\lambda+})$ of the minimal form $\mathbf{f}(\lambda)$ in the automorphic representation $\pi(\lambda)$. For that, we choose a decomposition $\mathfrak{C} = \mathfrak{F}\mathfrak{F}_c\mathfrak{I}$ so that $\mathfrak{F}\mathfrak{F}_c$ is a product of split primes and $\mathfrak{I}$ for the product of inert or ramified primes, $\mathfrak{F} + \mathfrak{F}_c = O_M$ and $\mathfrak{F} \subset \mathfrak{F}_c^c$, where $\mathfrak{F}$ could be strictly smaller than $\mathfrak{F}_c^c$. If we need to make the dependence on λ of these symbols explicit, we write $\mathfrak{F}(\lambda) = \mathfrak{F}$, $\mathfrak{F}_c(\lambda) = \mathfrak{F}_c$ and $\mathfrak{I}(\lambda) = \mathfrak{I}$. We put $\mathfrak{f} = \mathfrak{F} \cap F$ and $\mathfrak{i} = \mathfrak{I} \cap F$. Define $\lambda^-(\mathfrak{a}) = \lambda(\mathfrak{a}^{c-1})$ (with $\mathfrak{a}^{c-1} = \mathfrak{a}^c\mathfrak{a}^{-1}$), and write its conductor as $\mathfrak{C}(\lambda^-)$. Decompose as above $\mathfrak{C}(\lambda^-) = \mathfrak{F}(\lambda^-)\mathfrak{F}^c(\lambda^-)\mathfrak{I}(\lambda^-)$ so that we have the following divisibility of radicals $\sqrt{\mathfrak{F}(\lambda^-)}|\sqrt{\mathfrak{F}(\lambda)}$ and $\sqrt{\mathfrak{F}_c(\lambda^-)}|\sqrt{\mathfrak{F}_c(\lambda)}$. Let $\mathcal{T}_M = \mathrm{Res}_{O_M/O}\mathbb{G}_m$. The $\mathfrak{l}$-component $\epsilon_{\lambda,j,\mathfrak{l}}$ ($j = 1, 2$) of the character $\epsilon_{\lambda,j}$ is given as follows:

(hk1) For $\mathfrak{l}|\mathfrak{f}$, we identify $\mathcal{T}_M(O_\mathfrak{l}) = O_{M,\mathfrak{L}}^{\times} \times O_{M,\mathfrak{L}^c}^{\times}$ with this order for the prime ideal $\mathfrak{L}|(\mathfrak{l}O_M \cap \mathfrak{F})$ and define $\epsilon_{\lambda,1,\mathfrak{l}} \times \epsilon_{\lambda,2,\mathfrak{l}}$ by the restriction of $\lambda_{\mathfrak{L}} \times \lambda_{\mathfrak{L}^c}$ to $\mathcal{T}_M(O_\mathfrak{l})$.

(hk2) For $\mathfrak{P} \in \Sigma_p$, we identify $\mathcal{T}_M(O_\mathfrak{p}) = \mathfrak{O}_{M_\mathfrak{P}}^{\times} \times \mathfrak{O}_{M_{\mathfrak{P}^c}}^{\times}$ and define $\epsilon_{\lambda,1,\mathfrak{p}} \times \epsilon_{\lambda,2,\mathfrak{p}}$ by the restriction of $\lambda_\mathfrak{P} \times \lambda_{\mathfrak{P}^c}$ to $\mathcal{T}_M(O_\mathfrak{p})$.

(hk3) For $\mathfrak{l}|(\mathfrak{I}(\lambda) \cap O)d(M/F)$ but $\mathfrak{l} \nmid (\mathfrak{I}(\lambda^-) \cap O)$, we can choose a character $\phi_\mathfrak{l} : F_\mathfrak{l}^{\times} \to \mathbb{C}^{\times}$ such that $\lambda_\mathfrak{L} = \phi_\mathfrak{l} \circ N_{M_\mathfrak{L}/F_\mathfrak{l}}$. Then we define $\epsilon_{\lambda,1,\mathfrak{l}}(a) = \left(\frac{M_\mathfrak{L}/F_\mathfrak{l}}{a}\right)\phi_\mathfrak{l}(a)$ and $\epsilon_{\lambda,2,\mathfrak{l}}(d) = \phi_\mathfrak{l}(d)$, where $\mathfrak{L}$ is the prime factor of $\mathfrak{l}$ in M and $\left(\frac{M_\mathfrak{L}/F_\mathfrak{l}}{d}\right)$ is the character of $M_\mathfrak{L}/F_\mathfrak{l}$.

(hk4) For $\mathfrak{l}|(\mathfrak{I}(\lambda^-) \cap O)$, $\epsilon_{\lambda,1,\mathfrak{l}} = \epsilon_{\lambda+,\mathfrak{l}}|_{O_\mathfrak{l}^{\times}}$ and $\epsilon_{\lambda,2,\mathfrak{l}} = 1$ for the central character $\epsilon_{\lambda+}$ given in (C).

We now give an explicit description of the automorphic representation $\pi(\lambda)$. In Cases (hk1–3), taking a prime $\mathfrak{L}|\mathfrak{l}$ in M, we have

$$\pi_\mathfrak{p}(\lambda) \cong \begin{cases} \pi(\lambda_\mathfrak{L}, \lambda_{\mathfrak{L}^c}) & \text{in Case (hk1)}, \\ \pi(\lambda_\mathfrak{P}, \lambda_{\mathfrak{P}^c}) & \text{in Case (hk2)}, \\ \pi(\left(\frac{M_\mathfrak{L}/F_\mathfrak{l}}{}\right)\phi_\mathfrak{l}, \phi_\mathfrak{l}) & \text{in Case (hk3)}. \end{cases} \tag{26}$$

In Case (hk4), $\pi_{\mathfrak{l}}(\lambda)$ is the super-cuspidal representation giving rise to $\mathrm{Ind}_{M_{\mathfrak{l}}}^{F_{\mathfrak{l}}} \widehat{\lambda}|_{\mathrm{Gal}(\overline{F}_{\mathfrak{l}}/M_{\mathfrak{l}})}$.

To describe of $\mathbf{f}(\lambda)$, we split $\mathfrak{n}_\theta$ into a product of co-prime ideals $\mathfrak{n}_{nc}$ and $\mathfrak{n}_{cusp}$ so that $\mathfrak{n}_{nc}$ is made up of primes in Cases (hk1–3). For $\mathfrak{l}|\mathfrak{n}_{nc}$, writing $\pi_{\mathfrak{l}}(\lambda) = \pi(\eta_{\mathfrak{l}}, \eta'_{\mathfrak{l}})$ for characters $\eta_{\mathfrak{l}}, \eta'_{\mathfrak{l}} : F_{\mathfrak{l}}^\times \to \mathbb{C}^\times$, we write $C_{\mathfrak{l}}$ for the conductor of $\eta_{\mathfrak{l}}^{-1}\eta'_{\mathfrak{l}}$. Define the minimal level of $\pi(\lambda)$ by

$$\mathfrak{n}(\lambda) = \mathfrak{n}_{cusp} \prod_{\mathfrak{l}|\mathfrak{n}_{nc}} C_{\mathfrak{l}},$$

where $\mathfrak{l}$ runs over primes satisfying one of the three conditions (hk1–3). Put

$$\Xi = \{\mathfrak{L} | \mathfrak{L} \supset \mathfrak{F} \prod_{\mathfrak{P} \in \Sigma_p} \mathfrak{P}, \mathfrak{L} \supset \mathfrak{n}(\lambda)\}$$

for primes $\mathfrak{L}$ of M. Then the minimal form $\mathbf{f}(\lambda)$ has the following q-expansion coefficient:

$$\mathbf{c}_p(y, \mathbf{f}(\lambda)) = \begin{cases} \sum_{xx^c = y, x_\Xi = 1} \widehat{\lambda}(x) & \text{if } y \text{ is integral,} \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

where x runs over $(\widehat{O}_M \cap M^\times_{\mathbb{A}^{(\infty)}}/(O_M^{(\Xi)})^\times$ with $x_{\mathfrak{L}} = 1$ for $\mathfrak{L} \in \Xi$. See [H06, Sect. 6.2] for more details of this construction (though in [H06], the order of (κ_1, κ_2) is interchanged so that $\kappa_1 > \kappa_2$).

1.13 CM Components

We fix a Hecke character λ of type κ as in the previous subsection, and we continue to use the symbols defined above. We may regard the Galois character $\widehat{\lambda}$ as a character of $Cl_M(\mathfrak{C}p^\infty)$.

We consider the ray class group $Cl_M(\mathfrak{C}(\lambda^-)p^\infty)$ modulo $\mathfrak{C}(\lambda^-)p^\infty$. Since $\lambda^-(\mathfrak{a}^c) = (\lambda^-)^{-1}(\mathfrak{a})$, we have $\mathfrak{C}(\lambda^-) = \mathfrak{C}(\lambda^-)^c$. Thus $\mathrm{Gal}(M/F) = \langle c \rangle$ acts naturally on $Cl_M(\mathfrak{C}(\lambda^-)p^\infty)$. We define the anticyclotomic quotient of $Cl_M(\mathfrak{C}(\lambda^-)p^\infty)$ by

$$Cl_M^-(\mathfrak{C}(\lambda^-)p^\infty) := Cl_M(\mathfrak{C}(\lambda^-)p^\infty)/Cl_M(\mathfrak{C}(\lambda^-)p^\infty)^{1+c}.$$

We have canonical identities:

$$O_{M,\mathfrak{p}}^\times = O_{M,\mathfrak{P}}^\times \times O_{M,\mathfrak{P}^c}^\times = O_{\mathfrak{p}}^\times \times O_{\mathfrak{p}}^\times \text{ and } O_{M,p}^\times := (O_M \otimes_{\mathbb{Z}} \mathbb{Z}_p)^\times = O_{M,\Sigma_p}^\times \times O_{M,\Sigma_p^c}^\times = O_p^\times \times O_p^\times$$

on which c acts by interchanging the components. Here $O_{M,X} = \prod_{\mathfrak{P}\in X} O_{M,\mathfrak{P}}$ for $X = \Sigma$ and Σ^c. The natural inclusion $O_{M,p}^{\times}/\overline{O_M^{\times}} \hookrightarrow Cl(\mathfrak{C}(\lambda^-)p^{\infty})$ induces an inclusion $\mathbf{\Gamma} \hookrightarrow Cl_M^-(\mathfrak{C}(\lambda^-)p^{\infty})$. Decompose $Cl_M^-(\mathfrak{C}(\lambda^-)p^{\infty}) = \mathbf{\Gamma}_M \times \Delta_M$ with the maximal finite subgroup Δ_M so that $\mathbf{\Gamma}_M \supset \mathbf{\Gamma}$. Then $\mathbf{\Gamma}$ is an open subgroup in $\mathbf{\Gamma}_M$. In particular, $W[[\mathbf{\Gamma}_M]]$ is a regular domain finite flat over $\mathbf{\Lambda}_W$. Thus we call $P \in \mathrm{Spec}(W[[\mathbf{\Gamma}_M]])(\overline{\mathbb{Q}}_p)$ arithmetic if P is above an arithmetic point of $\mathrm{Spec}(\mathbf{\Lambda}_W)(\overline{\mathbb{Q}}_p)$. Regard the tautological character

$$\upsilon : Cl_M(\mathfrak{C}p^{\infty}) \xrightarrow{\text{projection}} \mathbf{\Gamma}_M \hookrightarrow W[\mathbf{\Gamma}_M]]^{\times}$$

as a Galois character $\upsilon : \mathrm{Gal}(\overline{M}/M) \to W[\mathbf{\Gamma}_M]]^{\times}$.

The composite $\upsilon_P = P \circ \upsilon$ for an arithmetic point $P \in \mathrm{Spec}(W[[\mathbf{\Gamma}_M]])$ is of the form $\widehat{\varphi}_P$ for a Hecke character φ_P with p-type $\kappa'_{P,2}\Sigma_p + \kappa'_{P,1}\Sigma_p^c$ for $\kappa'_P = (\kappa'_{P,1}, \kappa'_{P,2}) \in \mathbb{Z}[I_p]^2$ satisfying $\kappa_2 + \kappa'_{P,2} - (\kappa_1 + \kappa'_{P,1}) \geq I_p$. Assume that $\widehat{\lambda}$ has values in $W^{\times}$ (enlarging W if necessary). We then consider the product $\widehat{\lambda}\upsilon$: $\mathrm{Gal}(\overline{M}/M) \to W[[\mathbf{\Gamma}_M]]^{\times}$ and $\rho_{W[[\mathbf{\Gamma}_M]]} := \mathrm{Ind}_M^F \widehat{\lambda}\upsilon : \mathrm{Gal}(\overline{M}/M) \to \mathrm{GL}_2(W[[\mathbf{\Gamma}_M]])$. Define $\mathbb{I}_M \subset W[[\mathbf{\Gamma}_M]]$ by the $\mathbf{\Lambda}_W$-subalgebra generated by $\mathrm{Tr}(\rho_{W[[\Gamma_M]]})$. Then we have the localization identity $\mathbb{I}_{M,P} = W[[\mathbf{\Gamma}_M]]_P$ for any arithmetic point P (this follows from the irreducibility of $\rho_P = P \circ \rho_{W[[\mathbf{\Gamma}_M]]} = \mathrm{Ind}_M^F \widehat{\lambda}\upsilon_P$; e.g., [H86b, Theorem 4.3]).

Let $\mathbf{h} = \mathbf{h}^{\mathrm{n.ord}}(\mathfrak{n}(\lambda), \epsilon_{\lambda+}; W)$, which is a torsion-free finite $\mathbf{\Lambda}_W$-algebra. We have a surjective projection $\pi_\lambda : \mathbf{h} \to \mathbb{I}_M$ sending $T(\mathfrak{l})$ to $\mathrm{Tr}(\rho_{W[[\mathbf{\Gamma}]]}(Frob_{\mathfrak{l}}))$ for primes $\mathfrak{l}$ outside $\mathfrak{n}(\lambda)$. Thus $\mathrm{Spec}(\mathbb{I}_M)$ is an irreducible component of $\mathrm{Spec}(\mathbf{h})$. In particular, $\rho_{\mathbb{I}_M} = \rho_{W[[\mathbf{\Gamma}_M]]}$. In the same manner as in [HMI, Proposition 3.78], we prove the following fact:

Proposition 1.1. *Let the notation be as above. Then for the reduced part* $\mathbf{h}^{red}$ *of* $\mathbf{h}$ *and each arithmetic point* $P \in \mathrm{Spec}(\mathbf{\Lambda}_W)(\overline{\mathbb{Q}}_p)$, $\mathrm{Spec}(\mathbf{h}_P^{red})$ *is finite étale over* $\mathrm{Spec}(\mathbf{\Lambda}_P)$. *In particular, no irreducible components cross each other at a point above an arithmetic point of* $\mathrm{Spec}(\mathbf{\Lambda}_W)$.

A component $\mathbb{I}$ is called a *CM component* if there exists a nontrivial character $\chi : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathbb{I}^{\times}$ such that $\rho_{\mathbb{I}} \cong \rho_{\mathbb{I}} \otimes \chi$. We also say that $\mathbb{I}$ has *complex multiplication* if $\mathbb{I}$ is a CM component. In this case, we call the corresponding family $\mathcal{F}$ a CM family (or we say $\mathcal{F}$ has complex multiplication). It is known essentially by deformation theory of Galois characters (cf. [H11, Sect. 4]) that any CM component is given by $\mathrm{Spec}(\mathbb{I}_M)$ as above for a specific choice of λ.

If $\mathcal{F}$ is a CM family associated with $\mathbb{I}$ with $\rho_{\mathbb{I}} \cong \rho_{\mathbb{I}} \otimes \chi$, then χ is a quadratic character of $\mathrm{Gal}(\overline{\mathbb{Q}}/F)$ which cuts out a CM quadratic extension M/F, i.e., $\chi = \left(\frac{M/F}{}\right)$. Write $\widetilde{\mathbb{I}}$ for the integral closure of $\mathbf{\Lambda}_W$ inside the quotient field of $\mathbb{I}$. The following three conditions are known to be equivalent:

(CM1) *$\mathcal{F}$ has CM and* $\rho_{\mathbb{I}} \cong \rho_{\mathbb{I}} \otimes \left(\frac{M/F}{}\right)$ *(*$\Leftrightarrow$ $\rho_{\mathbb{I}} \cong \mathrm{Ind}_M^F \Psi$ *for a character* $\Psi := \widehat{\lambda}\upsilon : \mathrm{Gal}(\overline{\mathbb{Q}}/M) \to Q(\mathbb{I})^{\times}$ *for the quotient field* $Q(\mathbb{I})$ *of* $\mathbb{I}$*);*

(CM2) *For all arithmetic P of* $\mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$*,* $\mathbf{f}_P$ *is a binary theta series of the norm form of M/F;*

(CM3) *For some arithmetic P of* $\mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$*,* $\mathbf{f}_P$ *is a binary theta series of the norm form of M/F.*

Since the characteristic polynomial of $\rho_{\mathbb{I}}(\sigma)$ has coefficients in $\mathbb{I}$, its eigenvalues fall in $\widetilde{\mathbb{I}}$; so, the character Ψ has values in $\widetilde{\mathbb{I}}^\times$ (see [H86b, Corollary 4.2]). Then, (CM1) is equivalent to $\rho_{\mathbb{I}} \cong \mathrm{Ind}_M^F \Psi$ for a character $\Psi : \mathrm{Gal}(\overline{\mathbb{Q}}/M) \to \widetilde{\mathbb{I}}^\times$ unramified outside Np (e.g., [MFG, Lemma 2.15]). Then by (Gal) and (Ram), $\Psi_P = P \circ \Psi : \mathrm{Gal}(\overline{\mathbb{Q}}/M) \to \overline{\mathbb{Q}}_p^\times$ for an arithmetic $P \in \mathrm{Spec}(\widetilde{\mathbb{I}})(\overline{\mathbb{Q}}_p)$ is a locally algebraic p-adic character, which is the p-adic avatar of a Hecke character $\lambda_P : M_{\mathbb{A}}^\times/M^\times \to \mathbb{C}^\times$ of type A_0 of the quadratic extension $M_{/F}$. Then by the characterization (Gal) of $\rho_{\mathbb{I}}$, $\mathbf{f}_P$ is the theta series $\mathbf{f}(\lambda)$, where $\mathfrak{a}$ runs over all integral ideals of M. By $\kappa_2(P)-\kappa_1(P) \geq I$ (and (Gal)), M has to be a CM field in which $\mathfrak{p}$ is split (as the existence of Hecke characters of infinity type corresponding to such $\kappa(P)$ forces that M/F is a CM quadratic extension). This shows (CM1)⇒(CM2)⇒(CM3). If (CM2) is satisfied, we have an identity $\mathrm{Tr}(\rho_{\mathbb{I}}(Frob_{\mathfrak{l}})) = a(\mathfrak{l}) = \chi(\mathfrak{l})a(\mathfrak{l}) = \mathrm{Tr}(\rho_{\mathbb{I}} \otimes \chi(Frob_{\mathfrak{l}}))$ with $\chi = \left(\frac{M/F}{}\right)$ for all primes $\mathfrak{l}$ outside a finite set of primes (including prime factors of $\mathfrak{n}(\lambda)p$). By Chebotarev density, we have $\mathrm{Tr}(\rho_{\mathbb{I}}) = \mathrm{Tr}(\rho_{\mathbb{I}} \otimes \chi)$, and we get (CM1) from (CM2) as $\rho_{\mathbb{I}}$ is semi-simple. If a component $\mathrm{Spec}(\mathbb{I})$ contains an arithmetic point P with theta series $\mathbf{f}_P$ of M/F as above, either $\mathbb{I}$ is a CM component or otherwise P is in the intersection in $\mathrm{Spec}(\mathbf{h})$ of a component $\mathrm{Spec}(\mathbb{I})$ not having CM by M and another component having CM by M (as all families with CM by M are made up of theta series of M by the construction of CM components as above). The latter case cannot happen as two distinct components never cross at an arithmetic point in $\mathrm{Spec}(\mathbf{h})$ (i.e., the reduced part of the localization $\mathbf{h}_P$ is étale over $\mathbf{\Lambda}_P$ for any arithmetic point $P \in \mathrm{Spec}(\mathbf{\Lambda})(\overline{\mathbb{Q}}_p)$; see Proposition 1.1). Thus (CM3) implies (CM2). We call a binary theta series of the norm form of a CM quadratic extension of F a *CM theta series*.

Remark 1.2. If $\mathrm{Spec}(\mathbb{J})$ is an integral closed subscheme of $\mathrm{Spec}(\mathbb{I})$, we write the associated Galois representation as $\rho_{\mathbb{J}}$. By abuse of language, we say $\mathbb{J}$ has CM by M if $\rho_{\mathbb{J}} \cong \rho_{\mathbb{J}} \otimes \left(\frac{M/F}{}\right)$. Thus (CM3) is equivalent to having ρ_P with CM for some arithmetic point P. More generally, if we find some arithmetic point P in $\mathrm{Spec}(\mathbb{J})$ and ρ_P has CM, $\mathbb{J}$ and $\mathbb{I}$ have CM.

2 Weil Numbers

Since $\overline{\mathbb{Q}}$ sits inside $\mathbb{C}$, it has "the" complex conjugation c. For a prime l, a Weil l-number $\alpha \in \overline{\mathbb{Q}}$ of integer weight $k \geq 0$ is defined by the following two properties:

(1) α is an algebraic integer;

(2) $|\alpha^\sigma| = l^{k/2}$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/F)$ for the complex archimedean absolute value $|\cdot|$.

Note that $\mathbb{Q}(\alpha)$ is in a CM field finite over $\mathbb{Q}$ (e.g., [Ho68, Proposition 4]), and the Weil l-number is realized as the Frobenius eigenvalue of a CM abelian variety over a finite field of characteristic l. We call two nonzero numbers $a, b \in \overline{\mathbb{Q}}$ equivalent (written as $a \sim b$) if a/b is a root of unity. We say that Weil numbers α and β are p-equivalent if $\alpha/\beta \in \mu_{p^\infty}(\overline{\mathbb{Q}})$. Here is an improvement of [H11, Corollary 2.5] proved as [H14, Corollary 2.2]:

Proposition 2.1. *Let d be a positive integer. Let $\mathcal{K}_d$ be the set of all finite extensions of $K = \mathbb{Q}[\mu_{p^\infty}]$ of degree d inside $\overline{\mathbb{Q}}$. If $l \neq p$, there are only finitely many Weil l-numbers of a given weight in the set-theoretic union $\bigcup_{L \in \mathcal{K}_d} L^\times$ (in $\overline{\mathbb{Q}}^\times$) up to p-equivalence.*

Let $L_{/F}$ be a finite field extension inside $\mathbb{C}_p$ with integer ring O_L as in the introduction. Recall $T_L = \mathrm{Res}_{O_L/\mathbb{Z}}\mathbb{G}_m$ (in the sense of [NMD, Sect. 7.6, Theorem 4]) and a morphism $\nu \in \mathrm{Hom}_{\text{gp scheme}}(T_L, T_F)$ in the introduction. Define an integral domain $R = R_\nu$ by the subalgebra of $\mathbf{\Lambda}$ generated over $\mathbb{Z}_{(p)}$ by the image G of $\nu(O_{L,(p)}^\times) \cap T_F(\mathbb{Z}_p)$ projected down to $\mathbf{\Gamma}$. If $\nu \neq 1$, $\nu(O_{L,(p)}^\times) \cap T_F(\mathbb{Z}_p)$ contains $G_0 := \{\xi^N | \xi \in \mathbb{Z}_{(p)}^\times\}$ for some $0 < N \in \mathbb{Z}$. Replacing N by its suitable multiple, G_0 is a free $\mathbb{Z}$-module of infinite rank. Since $R_\nu \cong \mathbb{Z}_{(p)}[G]$ (the group algebra of G), R_ν contains a polynomial ring over $\mathbb{Z}_{(p)}$ (isomorphic to $\mathbb{Z}_{(p)}[G_0]$) with infinitely many variables, and $Q(R_\nu)$ has infinite transcendental degree over $\mathbb{Q}$ (if $\nu \neq 1$). For any arithmetic point P and $\xi \in R_\nu$, the value $\xi_P \in \mathbb{C}_p$ falls in $L^{\mathrm{gal}}[\mu_N, \mu_{p^\infty}]$ for the Galois closure L^{gal} of $L/\mathbb{Q}$ and $N = |\Delta|$. For example, if $F = \mathbb{Q}$ and $L = \mathbb{Q}$ with the identity $\nu : \mathbb{G}_m \cong \mathbb{G}_m$, taking $\gamma_1 = 1 + \mathbf{p}$ for $\mathbf{p} = 4$ if $p = 2$ and $\mathbf{p} = p$ if $p > 2$, we have $G = \{t^{\log_p(\xi)/\log_p(\gamma_1)} | \xi \in \mathbb{Z}_{(p)}\}$; so, $P(t^{\log_p(\xi)/\log_p(\gamma_1)}) = \xi^{\kappa_2}\omega(\xi^{\kappa_2})^{-1}\zeta$ for $P = (t - \zeta\gamma_1^{\kappa_2})$, where ω is the Teichmüller character ($N = p - 1$ for $F = \mathbb{Q}$ and odd p). Note that ξ^{κ_2} has values in L^{gal} instead of L. Recall the algebraic closure $\overline{Q}$ (we fixed) of the quotient field Q of $\mathbf{\Lambda}$.

Proposition 2.2. *Let $\mathbb{I}$ be a finite normal extension of $\mathbf{\Lambda}$ inside $\overline{Q}$ and regard $R = R_\nu \subset \mathbf{\Lambda}$ as a subalgebra of $\mathbb{I}$. Let $A \subset \mathbb{I}$ be an R-subalgebra of finite type whose quotient field $Q(A)$ is a finite extension of the quotient field $Q(R)$ of R. Regarding an arithmetic point $P \in \mathrm{Spec}(\mathbb{I})$ as an algebra homomorphism $P : \mathbb{I} \to \overline{\mathbb{Q}}_p$, write A_P (resp. R_P) for the composite of the image $P(A)$ [resp. $P(R)$] with $\mathbb{Q}(\mu_{p^\infty})$ inside $\overline{\mathbb{Q}}_p$. Then there exists a closed subscheme E of codimension at least 1 of $\mathrm{Spec}(\mathbb{I})$ such that there are finitely many Weil l-numbers of a given weight in $\bigcup_{P \notin E} A_P \subset \overline{\mathbb{Q}}$ up to p-power roots of unity, where P runs over all arithmetic points of $\mathrm{Spec}(\mathbb{I})$ outside E.*

Proof. We may assume that $A = R[a]$ (i.e., A is generated over R by a single element a). The generator $a \in A$ satisfies an equation $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n \in R[x]$ with $a_0 \neq 0$. Then the zero locus E of a_0 is a closed formal subscheme of codimension at least 1. Since arithmetic points are Zariski dense in $\mathrm{Spec}(\mathbb{I})$, we have a plenty of arithmetic points outside E (i.e., the set arithmetic points outside

E is infinite). Thus as long as $P(a_0) \neq 0$, we have $[A_P : R_P] \leq n$. Since $R_P \subset L^{\mathrm{gal}}[\mu_N, \mu_{p^\infty}]$, we have $[R_P : \mathbb{Q}(\mu_{p^\infty})] \leq B$ for a constant B independent of arithmetic P outside E. Thus $[A_P : \mathbb{Q}(\mu_{p^\infty})]$ is bounded independently by $d := nB$ for all arithmetic $P \notin E$. Then we can apply Proposition 2.1 and get the desired result. □

3 Theorems and Conjectures

Hereafter,

(W) we fix $\kappa \in \mathbb{Z}[I]^2$ with $\kappa_2 - \kappa_1 \geq I$.

Though the weight κ is fixed, the character ϵ_P is a variable (so, we have densely populated arithmetic points $P \in \mathrm{Spec}(\mathbb{I})$ with $\kappa(P) = \kappa$). Let $\mathbf{f} \in S_\kappa(\mathfrak{n}p^{r+I_p}, \epsilon; W)$ be a Hecke eigenform with $\mathbf{f}|T(y) = a(y, \mathbf{f})\mathbf{f}$ for all y. We normalize $\mathbf{f}$ so that $c(1, \mathbf{f}) = 1$. For a prime $\mathfrak{l} \nmid p$, we write $\mathbf{f}|T(\mathfrak{l}) = (\alpha_\mathfrak{l} + \beta_\mathfrak{l})\mathbf{f}$ and $\alpha_\mathfrak{l}\beta_\mathfrak{l} = \epsilon(\mathfrak{l})l^{f_\mathfrak{l}}$ if $\mathfrak{l} \nmid \mathfrak{n}p^{r+1}$ ($\alpha_\mathfrak{l}, \beta_\mathfrak{l} \in \overline{\mathbb{Q}}$), where $f_\mathfrak{l}$ is the degree of the field $O/\mathfrak{l}$ over the prime field $\mathbb{F}_l$. If $l|\mathfrak{n}$, we put $\beta_\mathfrak{l} = 0$ and define $\alpha_\mathfrak{l} \in \overline{\mathbb{Q}}$ by $\mathbf{f}|U(\mathfrak{l}) = \alpha_\mathfrak{l}\mathbf{f}$. Then the Hecke polynomial $H_\mathfrak{l}(X) = (1 - \alpha_\mathfrak{l}X)(1 - \beta_\mathfrak{l}X)$ gives the Euler $\mathfrak{l}$-factor of $L(s, \mathbf{f}) = \sum_\mathfrak{n} a(\mathfrak{n}, \mathbf{f})N(\mathfrak{n})^{-s}$ after replacing X by $|O/\mathfrak{l}|^{-s} = N(\mathfrak{l})^{-s}$ and inverting the resulted factor. Here $\mathfrak{n}$ runs over all integral ideals of F.

Let $\mathcal{F} = \{\mathbf{f}_P\}_{P \in \mathrm{Spec}(\mathbb{I})(\mathbb{C}_p)}$ be a p-adic analytic family of p-ordinary Hecke eigen cusp forms of p-slope 0. The function $P \mapsto a(\mathfrak{y}, \mathbf{f}_P)$ is a function on $\mathrm{Spec}(\mathbb{I})$ in the structure sheaf $\mathbb{I}$; so, it is a formal (and analytic) function of P. We write $\alpha_{\mathfrak{l},P}, \beta_{\mathfrak{l},P}$ for $\alpha_\mathfrak{l}, \beta_\mathfrak{l}$ for $\mathbf{f}_P$. We write $\alpha_{\mathfrak{p},P}$ for $a(\mathfrak{p}, \mathbf{f}_P) = a(\varpi_\mathfrak{p}, \mathbf{f}_P)$. In particular, the field $F[\kappa][\mu_{Np^\infty}][\alpha_{\mathfrak{p},P}]$ (for the field $F[\kappa]$ of rationality of κ defined in Sect. 1.8) is independent of the choice of $\varpi_\mathfrak{p}$ (as long as $\varpi_\mathfrak{p}$ is chosen in F). By a result of Blasius [B02] (and by an earlier work of Brylinski–Labesse), writing $|\kappa_1| := \max_\sigma(|\kappa_{1,\sigma}|)$, $N(\mathfrak{l})^{|\kappa_1|}\alpha_{\mathfrak{l},P}$ is a Weil l-number of weight $([\kappa] + 2|\kappa_1|)f_\mathfrak{l}$ for $f_\mathfrak{l}$ given by $|O/\mathfrak{l}| = l^{f_\mathfrak{l}}$. Thus $\alpha_{\mathfrak{l},P}$ is a generalized Weil number in the sense of [H13, Sect. 2].

We state the horizontal theorem in a form different from the theorem in the introduction:

Theorem 3.1. *Let $K = \mathbb{Q}(\mu_{p^\infty})$. Suppose that there exist a subset Σ of primes of F with positive upper density outside $\mathfrak{n}p$ and an infinite set $\mathcal{A}_\mathfrak{l} \subset \mathrm{Spec}(\mathbb{I})(\overline{\mathbb{Q}}_p)$ of arithmetic points P of the fixed weight κ as in (W) such that $[K(\alpha_{\mathfrak{l},P}) : K] \leq B_\mathfrak{l}$ for all $P \in \mathcal{A}_\mathfrak{l}$ with a bound $B_\mathfrak{l}$ for each $\mathfrak{l} \in \Sigma$ (possibly dependent on $\mathfrak{l}$). If the Zariski closure $\overline{\mathcal{A}}_\mathfrak{l}$ in* $\mathrm{Spec}(\mathbb{I})$ *contains an irreducible subscheme* $\mathrm{Spec}(\mathbb{J})$ *of dimension $r \geq 1$ independent of $\mathfrak{l} \in \Sigma$ with Zariski-dense $\mathcal{A}_\mathfrak{l} \cap$* $\mathrm{Spec}(\mathbb{J})$ *in* $\mathrm{Spec}(\mathbb{J})$*, then $\mathbb{I}$ has complex multiplication.*

In the above theorem, κ is independent of $\mathfrak{l}$ but $B_\mathfrak{l}$ and $\mathcal{A}_\mathfrak{l}$ can be dependent on $\mathfrak{l}$. By replacing $\mathcal{A}_\mathfrak{l}$ by a suitable infinite subset of $\mathcal{A}_\mathfrak{l} \cap \mathrm{Spec}(\mathbb{J})$, we may assume that $\overline{\mathcal{A}}_\mathfrak{l}$ is irreducible with dimension r independent of $\mathfrak{l}$. By extending W if necessary, we may assume that $\mathrm{Spec}(\mathbb{J})$ is geometrically irreducible. From the proof of this theorem given in Sect. 6, it will be clear that we can ease the assumption of the theorem so that κ is also dependent on $\mathfrak{l}$.

Let R_ν be as in Proposition 2.2 for a number field L. Then we have the following result which implies Corollary I in the introduction:

Corollary 3.2. *Let the notation be as in* Proposition 2.2 *and in the above theorem. Let Σ be a set of primes of F with positive upper density. Let* $\mathrm{Spec}(\mathbb{I})$ *be a reduced irreducible component of* $\mathrm{Spec}(\mathbf{h})$*, and assume that $\mathbb{I}$ is a finite extension of $\mathbf{\Lambda}$ inside $\overline{\mathbb{Q}}$. If there exists a pair (L, ν) of a finite extension $L_{/F}$ and a homomorphism $\nu \in$* $\mathrm{Hom}_{gpscheme}(T_L, T_F)$ *such that the ring $R_\nu[a(\mathfrak{l})]$ generated over R_ν by $a(\mathfrak{l})$ inside $\overline{\mathbb{Q}}$ has quotient field $Q(R_\nu[a(\mathfrak{l})])$ finite over the quotient field $Q(R_\nu)$ for all $\mathfrak{l} \in \Sigma$, then $\mathbb{I}$ has complex multiplication.*

Proof. Applying Proposition 2.2 to $A_\mathfrak{l} = R_\nu[a(\mathfrak{l})]$, we take $\mathcal{A}_\mathfrak{l}$ to be the set of the arithmetic points outside the closed subscheme $E_\mathfrak{l}$ for $R_\nu[a(\mathfrak{l})]$ in Proposition 2.2. Then the Zariski closure of $\mathcal{A}_\mathfrak{l}$ is the entire $\mathrm{Spec}(\mathbb{I})$ as $E_\mathfrak{l}$ has codimension at least 1. Thus the assumption of the theorem is satisfied for $\mathcal{A}_\mathfrak{l}$ for all $\mathfrak{l} \in \Sigma$. Therefore, the above theorem tells us that $\mathbb{I}$ has CM. □

This corollary implies

Corollary 3.3. *Suppose that $\mathbb{I}$ is a non-CM component. Let (L, ν) be a pair of finite extension of F and $\nu \in$* $\mathrm{Hom}_{gpscheme}(T_L, T_F)$*. Then, for a density one set of primes Ξ of F outside $p\mathfrak{n}$, the ring $R_\nu[a(\mathfrak{l})] \subset \overline{\mathbb{Q}}$ for each $\mathfrak{l} \in \Xi$ generated over $R_\nu \subset \overline{\mathbb{Q}}$ by $a(\mathfrak{l})$ inside $\overline{\mathbb{Q}}$ has quotient field of transcendental degree one over $Q(R_\nu)$ in $\overline{\mathbb{Q}}$.*

Proof. Let Ξ be the set of primes $\mathfrak{l}$ of F made up of $\mathfrak{l}$ with $a(\mathfrak{l})$ transcendental over $Q(R_\nu)$ (as $a(\mathfrak{l}) \notin W$: non-constancy). Let Σ be the complement of Ξ outside $p\mathfrak{n}$. If Σ has positive upper density, by Corollary 3.2, $\mathbb{I}$ has complex multiplication by a subfield of L, a contradiction. Thus Σ has upper density 0, and hence Ξ has density 1. □

By Theorem 3.1, we get the following corollary:

Corollary 3.4. *Let $\mathcal{A}$ be an infinite set of arithmetic points of* $\mathrm{Spec}(\mathbb{I})$ *of fixed weight κ. Then there exists a subset Σ of primes of F with upper positive density such that $[K(a(\mathfrak{l}, \mathbf{f}_P)) : K]$ for $\mathfrak{l} \in \Sigma$ is bounded over $\mathcal{A}$ if and only if $\mathbf{f}_P$ is a CM theta series for an arithmetic P with $k(P) \geq I$.*

By the argument given after [H11, Conjecture 3.4], one can show $[K(a(\mathfrak{l}, \mathbf{f}_P)) : K]$ is bounded independently of arithmetic points $P \in \mathrm{Spec}(\mathbb{I})$ if $\mathbf{f}_{P_0}$ is square-integrable at a prime $\mathfrak{l} \nmid p$ (so, $\mathfrak{l}|\mathfrak{n}$) for one arithmetic P_0. Further, if a prime $\mathfrak{l}$ is a factor of $\mathfrak{n}$ (so $\mathfrak{l} \nmid p$) and $\mathbf{f}_P$ (or more precisely the automorphic representation generated by $\mathbf{f}_P$) is Steinberg (resp. super-cuspidal) at $\mathfrak{l}$ for an arithmetic point P, then all members of $\mathcal{F}$ are Steinberg (resp. super-cuspidal) at $\mathfrak{l}$ (see the remark after Conjecture 3.4 in [H11]). Take a prime $\mathfrak{l} \nmid \mathfrak{n}$ of O with $\alpha_{\mathfrak{l},P} \neq 0$ for some P (so, $\mathfrak{l}$ can be equal to $\mathfrak{p}$). If $\mathfrak{l} \nmid \mathfrak{n}p$, replacing $\mathbb{I}$ by a finite extension, we assume that $\det(T - \rho_\mathbb{I}(Frob_l)) = 0$ has roots in $\mathbb{I}$. Since $\alpha_{\mathfrak{l},P} \neq 0$ for some P (and hence $\alpha_{\mathfrak{l},P}$ is a p-adic unit), $\mathbf{f}_P$ is not super-cuspidal at $\mathfrak{l}$ for any arithmetic P.

Conjecture 3.5. *Let the notation be as in* Corollary 3.4. *Let $\mathcal{A}$ be an infinite subset of arithmetic points in* $\mathrm{Spec}(\mathbb{I})$ *of fixed weight κ. Then* $\lim_{P\in\mathcal{A}}[K(a(\mathfrak{l},\mathbf{f}_P)):K]<\infty$ *for a single prime $\mathfrak{l}$ of F if and only if either $\mathbb{I}$ has complex multiplication or the automorphic representation generated by $\mathbf{f}_P$ is square integrable at $\mathfrak{l}\nmid p$ for a single $P\in\mathcal{A}$.*

4 Rigidity Lemmas

We study formal subschemes of $\widehat{G}:=\widehat{\mathbb{G}}_m^n$ stable under the action of $t\mapsto t^z$ for all z in an open subgroup U of $\mathbb{Z}_p^\times$. The following lemma and its corollary were proven in [H13]. For the reader's convenience (and to make the paper self-contained), we recall the statements and their proof.

Lemma 4.1. *Let $X=\mathrm{Spf}(\mathcal{X})$ be a closed formal subscheme of $\widehat{G}=\widehat{\mathbb{G}}^n_{m/W}$ flat geometrically irreducible over W (i.e., $\mathcal{X}\cap\overline{\mathbb{Q}}_p=W$). Suppose there exists an open subgroup U of $\mathbb{Z}_p^\times$ such that X is stable under the action $\widehat{G}\ni t\mapsto t^u\in\widehat{G}$ for all $u\in U$. If there exists a subset $\Omega\subset X(\mathbb{C}_p)\cap\mu^n_{p^\infty}(\mathbb{C}_p)$ Zariski dense in X, then $\zeta^{-1}X$ is a formal subtorus for some $\zeta\in\Omega$.*

Proof. Let X^{sh} be the scheme associated with X given by $\mathrm{Spec}(\mathcal{X})$. Define X_s to be the singular locus of $X^{sh}=\mathrm{Spec}(\mathcal{X})$ over W, and put $X^\circ=X^{sh}\setminus X_s$. The scheme X_s is actually a closed formal subscheme of X. To see this, we note, by the structure theorem of complete noetherian rings, that $\mathcal{X}$ is finite over a power series ring $W[[X_1,\dots,X_d]]\subset\mathcal{X}$ for $d:=\dim_W X$ (cf. [CRT, Sect. 29]). The sheaf of continuous differentials $\Omega_{\mathcal{X}/\mathrm{Spf}(W[[X_1,\dots,X_d]])}$ with respect to the formal Zariski topology of $\mathcal{X}$ is a torsion $\mathcal{X}$-module, and X_s is the support of the sheaf of $\Omega_{\mathcal{X}/\mathrm{Spf}(W[[X_1,\dots,X_d]])}$ (which is a closed formal subscheme of X). The regular locus of X° is open dense in the generic fiber $X^{sh}_{/K}:=X^{sh}\times_W K$ of X^{sh} (for the field K of fractions of W). Then $\Omega^\circ:=X^\circ\cap\Omega$ is Zariski dense in $X^{sh}_{/K}$.

In this proof, by making scalar extension, we always assume that W is sufficiently large so that for $\zeta\in\Omega$ we focus on, we have $\zeta\in\widehat{G}(W)$ and that we have a plenty of elements of infinite order in $X(W)$ and in $X^\circ(K)\cap X(W)$, which we simply write as $X^\circ(W):=X^\circ(K)\cap X(W)$.

Note that the stabilizer U_ζ of $\zeta\in\Omega$ in U is an open subgroup of U. Indeed, if the order of ζ is equal to p^a, then $U_\zeta=U\cap(1+p^a\mathbb{Z}_p)$. Thus making a variable change $t\mapsto t\zeta^{-1}$ (which commutes with the action of U_ζ), we may assume that the identity $\mathbf{1}$ of $\widehat{G}$ is in Ω°.

Let $\widehat{G}^{an}$, X_{an}, and X^s_{an} be the rigid analytic spaces associated with $\widehat{G}$, X, and X_s (in Berthelot's sense in [dJ95, Sect. 7]). We put $X^\circ_{an}=X_{an}\setminus X^s_{an}$, which is an open rigid analytic subspace of X_{an}. Then we apply the logarithm $\log:\widehat{G}^{an}(\mathbb{C}_p)\to\mathbb{C}_p^n=Lie(\widehat{G}^{an}_{/\mathbb{C}_p})$ sending $(t_j)_j\in\widehat{G}^{an}(\mathbb{C}_p)$ (the p-adic open unit ball centered at $\mathbf{1}=(1,1,\dots,1)$) to $(\log_p(t_j))_j)\in\mathbb{C}_p^n$ for the p-adic Iwasawa logarithm map $\log_p$:

$\mathbb{C}_p^\times \to \mathbb{C}_p$. Then for each smooth point $x \in X^\circ(W)$, taking a small analytic open neighborhood V_x of x (isomorphic to an open ball in W^d for $d = \dim_W X$) in $X^\circ(W)$, we may assume that $V_x = G_x \cap X^\circ(W)$ for an n-dimensional open ball G_x in $\widehat{G}(W)$ centered at $x \in \widehat{G}(W)$. Since $\Omega^\circ \neq \emptyset$, $\log(X^\circ(W))$ contains the origin $0 \in \mathbb{C}_p^n$. Take $\zeta \in \Omega^\circ$. Write T_ζ for the Tangent space at ζ of X. Then $T_\zeta \cong W^d$ for $d = \dim_W X$. The space $T_\zeta \otimes_W \mathbb{C}_p$ is canonically isomorphic to the tangent space T_0 of $\log(V_\zeta)$ at 0.

If $\dim_W X = 1$, there exists an infinite order element $t_1 \in X(W)$. We may (and will) assume that $U = (1 + p^b\mathbb{Z}_p)$ for $0 < b \in \mathbb{Z}$. Then X is the (formal) Zariski closure $\overline{t_1^U}$ of

$$t_1^U = \{t_1^{1+p^b z} | z \in \mathbb{Z}_p\} = t_1\{t_1^{p^b z} | z \in \mathbb{Z}_p\},$$

which is a coset of a formal subgroup Z. The group Z is the Zariski closure of $\{t_1^{p^b z} | z \in \mathbb{Z}_p\}$; in other words, regarding t_1^u as a W-algebra homomorphism $t_1^u : \mathcal{X} \to \mathbb{C}_p$, we have $t_1 Z = \mathrm{Spf}(\mathcal{Z})$ for $\mathcal{Z} = \mathcal{X}/\bigcap_{u \in U} \mathrm{Ker}(t_1^u)$. Since t_1^U is an infinite set, we have $\dim_W Z > 0$. From geometric irreducibility and $\dim_W X = 1$, we conclude $X = t_1 Z$ and $Z \cong \widehat{\mathbb{G}}_m$. Since X contains roots of unity $\zeta \in \Omega \subset \mu_{p^\infty}^n(W)$, we confirm that $X = \zeta Z$ for $\zeta \in \Omega \cap \mu_{p^{b'}}^n$ for $b' \gg 0$. This finishes the proof in the case where $\dim_W X = 1$.

We prepare some result (still assuming $d = 1$) for an induction argument on d in the general case. Replacing t_1 by $t_1^{p^b}$ for b as above if necessary, we have the translation $\mathbb{Z}_p \ni s \mapsto \zeta t_1^s \in Z$ of the one parameter subgroup $\mathbb{Z}_p \ni s \mapsto t_1^s$. Thus we have $\log(t_1) = \frac{dt_1^s}{ds}|_{s=0} \in T_\zeta$, which is sent by "$\log : \widehat{G} \to \mathbb{C}_p^n$" to $\log(t_1) \in T_0$. This implies that $\log(t_1) \in T_0$ and hence $\log(t_1) \in T_\zeta$ for any $\zeta \in \Omega^\circ$ (under the identification of the tangent space at any $x \in \widehat{G}$ with $Lie(\widehat{G})$). Therefore T_ζ's over $\zeta \in \Omega^\circ$ can be identified canonically. This is natural as Z is a formal torus, and the tangent bundle on Z is constant, giving $Lie(Z)$.

Suppose now that $d = \dim_W X > 1$. Consider the Zariski closure Y of t^U for an infinite order element $t \in V_\zeta$ (for $\zeta \in \Omega^\circ$). Since U permutes finitely many geometrically irreducible components, each component of Y is stable under an open subgroup of U. Therefore $Y = \bigcup \zeta' \mathcal{T}_{\zeta'}$ is a union of formal subtori $\mathcal{T}_{\zeta'}$ of dimension ≤ 1, where ζ' runs over a finite set inside $\mu_{p^\infty}^n(\mathbb{C}_p) \cap X(\mathbb{C}_p)$. Since $\dim_W Y = 1$, we can pick $\mathcal{T}_{\zeta'}$ of dimension 1 which we denote simply by $\mathcal{T}$. Then $\mathcal{T}$ contains t^u for some $u \in U$. Applying the argument in the case of $\dim_W X = 1$ to $\mathcal{T}$, we find $u \log(t) = \log(t^u) \in T_\zeta$; so, $\log(t) \in T_\zeta$ for any $\zeta \in \Omega^\circ$ and $t \in V_\zeta$. Summarizing our argument, we have found

(T) The Zariski closure of t^U in X for an element $t \in V_\zeta$ of infinite order contains a coset $\xi\mathcal{T}$ of one dimensional subtorus $\mathcal{T}$, $\xi^{p^b} = 1$ and $t^{p^b} \in \mathcal{T}$ for some $b > 0$;

(D) Under the notation as above, we have $\log(t) \in T_\zeta$.

Moreover, the image $\overline{V}_\zeta$ of V_ζ in $\widehat{G}/\mathcal{T}$ is isomorphic to $(d-1)$-dimensional open ball. If $d > 1$, therefore, we can find $\overline{t}' \in \overline{V}_\zeta$ of infinite order. Pulling back $\overline{t}'$ to

$t' \in V_\zeta$, we find $\log(t), \log(t') \in T_\zeta$, and $\log(t)$ and $\log(t')$ are linearly independent in T_ζ. Inductively arguing this way, we find infinite order elements $t_1, \dots, t_d$ in V_ζ such that $\log(t_i)$ span over the quotient field $\mathbb{K} = Q(W)$ of W the tangent space $T_{\zeta/\mathbb{K}} = T_\zeta \otimes_W \mathbb{K} \hookrightarrow T_0$ (for any $\zeta \in \Omega^\circ$). We identify $T_{\mathbf{1}/\mathbb{K}} \subset T_0$ with $T_{\zeta/\mathbb{K}} \subset T_0$. Thus the tangent bundle over $X^\circ_{/\mathbb{K}}$ is constant as it is constant over the Zariski dense subset Ω°. Therefore X° is something close to an open dense subscheme of a coset of a formal subgroup. We pin-down this fact that X° is a coset of a formal scheme.

Take $t_j \in V_\zeta$ as above ($j = 1, 2, \dots, d$) which give rise to a basis $\{\partial_j = \log(t_j)\}_j$ of the tangent space of $T_{\zeta/\mathbb{K}} = T_{\mathbf{1}/\mathbb{K}}$. Note that $t_j^u \in X$ and $u\partial_j = \log(t_j^u) = u\log(t_j) \in T_{\mathbf{1}/\mathbb{K}}$ for $u \in U$. The embedding $\log : V_\zeta \hookrightarrow T_{\mathbf{1}} \subset Lie(\widehat{G}_{/W})$ is surjective onto a open neighborhood of $0 \in T_{\mathbf{1}}$ (by extending scalars if necessary). For $t \in V_\zeta$, as $t \to \zeta$, $\log(t) \to 0$. Thus by replacing $t_1, \dots, t_d$ inside V_ζ with elements in V_ζ closer to ζ, we may assume that $\log(t_i) \pm \log(t_j)$ for all $i \neq j$ belong to $\log(V_\zeta)$.

So, for each pair $i \neq j$, we can find $t_{i\pm j} \in V_\zeta$ such that $\log(t_i t_j^{\pm 1}) = \log(t_i) \pm \log(t_j) = \log(t_{i\pm j})$. The element $\log(t_{i\pm j})$ is uniquely determined in $\log(\widehat{G}_{an}(\mathbb{C}_p)) \cong \widehat{G}_{an}(\mathbb{C}_p)/\mu^n_{p^\infty}(\mathbb{C}_p)$. Thus we conclude $\zeta'_{i\pm j} t_i t_j^{\pm 1} = t_{i\pm j}$ for some $\zeta'_{i\pm j} \in \mu^n_{p^N}$ for sufficiently large N. Replacing X by its image under the p-power isogeny $\widehat{G} \ni t \mapsto t^{p^N} \in \widehat{G}$ and t_i by $t_i^{p^N}$, we may assume that $t_i t_j^{\pm 1} = t_{i\pm j}$ all in X. Since $t_i^U \subset X$, by (T), for a sufficiently large $b \in \mathbb{Z}$, we find a one dimensional subtorus $\widehat{H}_i$ containing $t_i^{p^b}$ such that $\zeta_i \widehat{H}_i \subset X$ with some $\zeta_i \in \mu^n_{p^b}$ for all i. Thus again replacing X by the image of the p-power isogeny $\widehat{G} \ni t \mapsto t^{p^b} \in \widehat{G}$, we may assume that the subgroup $\widehat{H}$ (Zariski) topologically generated by $t_1, \dots, t_d$ is contained in X. Since $\{\log(t_i)\}_i$ is linearly independent, we conclude $\dim_W \widehat{H} \geq d = \dim_W X$, and hence X must be the formal subgroup $\widehat{H}$ of $\widehat{G}$. Since X is geometrically irreducible, $\widehat{H} = X$ is a formal subtorus. Pulling it back by the p-power isogenies we have used, we conclude $X = \zeta\widehat{H}$ for the original X and $\zeta \in \mu^n_{p^{bN}}(W)$. Since Ω is Zariski dense in X, we may assume that $\zeta \in \Omega$. This finishes the proof. $\square$

Corollary 4.2. *Let W be a complete discrete valuation ring in $\mathbb{C}_p$. Write $W[[T]] = W[[T_1, \dots, T_n]]$ for the tuple of variables $T = (T_1, \dots, T_n)$. Let*

$$\widehat{G} := \widehat{\mathbb{G}}^n_m = \mathrm{Spf}(\widehat{W[t_1, t_1^{-1} \dots, t_n, t_n^{-1}]}),$$

and identify $\widehat{W[t_1, t_1^{-1} \dots, t_n, t_n^{-1}]}$ with $W[[T]]$ for $t_j = 1 + T_j$. Let $\Phi(T_1, \dots, T_n) \in W[[T]]$. Suppose that there is Zariski dense subset $\Omega \subset \mu^n_{p^\infty}(\mathbb{C}_p)$ in $\widehat{G}(\mathbb{C}_p)$ such that $\Phi(\zeta - 1) \in \mu_{p^\infty}(\mathbb{C}_p)$ for all $\zeta \in \Omega$. Then there exists $\zeta_0 \in \mu_{p^\infty}(W)$ and $z = (z_j)_j \in \mathbb{Z}_p^n$ with $z_j \in \mathbb{Z}_p$ such that $\zeta_0^{-1}\Phi(t) = \prod_j (t_j)^{z_j}$, where $(1 + T)^x = \sum_{j=0}^{\infty} \binom{x}{j} T^j$ with $x \in \mathbb{Z}_p$.

Proof. Pick $\eta = (\eta_j) \in \Omega$. Making variable change $T \mapsto \eta^{-1}(T + 1) - 1$ (i.e., $T_j \mapsto \eta_j^{-1}(T_j + 1) - 1$ for each j) replacing W by its finite extension if necessary, we may replace Ω by $\eta^{-1}\Omega \ni 1$; so, rewriting $\eta^{-1}\Omega$ as Ω, we may assume that $\mathbf{1} \in \Omega$.

Then $\Phi(0) = \zeta_0 \in \mu_{p^\infty}$. Thus again replacing Φ by $\zeta_0^{-1}\Phi$, we may assume that $\Phi(0) = 1$.

For $\sigma \in \mathrm{Gal}(\mathbb{K}(\mu_{p^\infty})/\mathbb{K})$ with the quotient field $\mathbb{K}$ of W, $\Phi(\zeta^\sigma - 1) = \Phi(\zeta - 1)^\sigma$. Writing $\phi(\zeta) = \Phi(\zeta - 1)$, the above identity means $\phi(\zeta^\sigma) = \phi(\zeta)^\sigma$. Identify $\mathrm{Gal}(\mathbb{K}(\mu_{p^\infty})/\mathbb{K})$ with an open subgroup U of $\mathbb{Z}_p^\times$. This is possible as W is a discrete valuation ring, while $W[\mu_{p^\infty}]$ is not. Writing $\sigma_u \in \mathrm{Gal}(\mathbb{K}(\mu_{p^\infty})/\mathbb{K})$ for the element corresponding to $u \in U$, we find that

$$\Phi \circ u(\zeta - 1) = \Phi(\zeta^u - 1) = \Phi(\zeta^{\sigma_u} - 1) = \Phi(\zeta - 1)^{\sigma_u} = u \circ \Phi(\zeta - 1).$$

We find that $u \circ \phi = \phi \circ u$ is valid on the Zariski dense subset Ω of $\mathrm{Spec}(W[[T]])$; so, ϕ as a scheme morphism of $\widehat{G} = \widehat{\mathbb{G}}_m^n$ into $\widehat{\mathbb{G}}_m$ commutes with the action of $u \in U$.

Note that $u \in \mathbb{Z}_p^\times$ acts on $\widehat{\mathbb{G}}_m$ as a group automorphism induced by a W-bialgebra automorphism of $W[[T]]$ sending $t = (1+T) \mapsto t^u = (1+T)^u = \prod_j (1+T_j)^u$. Take the morphism of formal schemes $\phi \in \mathrm{Hom}_{SCH/W}(\widehat{\mathbb{G}}_m^n, \widehat{\mathbb{G}}_m)$, which sends 1 to 1. Put $\widehat{\mathbf{G}} := \widehat{\mathbb{G}}_m^n \times \widehat{\mathbb{G}}_{m/W}$. We consider the graph Γ_ϕ of ϕ which is an irreducible formal subscheme $\Gamma_\phi \subset \widehat{\mathbb{G}}_m^n \times \widehat{\mathbb{G}}_m$ smooth over W. Writing the variable on $\widehat{\mathbf{G}}$ as (T, T'), Γ_ϕ is the geometrically irreducible closed formal subscheme containing the identity $\mathbf{1} \in \widehat{\mathbf{G}}$ defined by the principal ideal $(t' - \phi(t))$. Since $\phi \circ u = u \circ \phi$ for all u in an open subgroup U of $\mathbb{Z}_p^\times$ (where U acts on the source $\widehat{\mathbb{G}}_m^n$ and on the target $\widehat{\mathbb{G}}_m$ by $t \mapsto t^u$), Γ_ϕ is stable under the diagonal action of U on $\widehat{\mathbf{G}}$ and is finite flat over $\widehat{\mathbb{G}}_m^n$ (the left factor of $\widehat{\mathbf{G}}$). Then, applying Lemma 4.1 to Γ_ϕ, we find that Γ_ϕ is a subtorus of rank n surjecting down to the last factor $\widehat{\mathbb{G}}_m$. Since any subtorus of rank n in $\widehat{\mathbf{G}}$ whose projection to the last factor is defined by the equation $t' = (1+T)^z$, $t' = \Phi(T)$, we have the power series identity $\Phi(T) = t' = (1+T)^z$ in $W[[T]]$ identifying $\Gamma_\phi = \mathrm{Spf}(W[[T]])$. $\square$

5 Frobenius Eigenvalue Formula

Recall the fixed weight κ with $\kappa_2 - \kappa_1 \geq I$. We assume the following conditions and notations:

(J1) Let $\mathrm{Spec}(\mathbb{J})$ be a closed reduced geometrically irreducible subscheme of $\mathrm{Spec}(\mathbb{I})$ flat over $\mathrm{Spec}(W)$ of relative dimension r with Zariski dense set $\mathcal{A}$ of arithmetic points of the fixed weight κ.

(J2) We identify $\mathrm{Spf}(\mathbf{\Lambda})$ for $\mathbf{\Lambda} = W[[\mathbf{\Gamma}]]$ with $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ for $\mathbf{\Gamma}^* := \mathrm{Hom}_{\mathbb{Z}_p}(\mathbf{\Gamma}, \mathbb{Z}_p)$ naturally.

Then for any direct $\mathbb{Z}_p$-summand $\Gamma \subset \mathbf{\Gamma}$, $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*$ is a closed formal torus of $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$. We insert here a lemma (essentially) proven in [H13, Lemma 5.1].

Lemma 5.1. *Let the notation and the assumption be as in (J1–2). Then, after making extension of scalars to a sufficiently large complete discrete valuation ring*

$W \subset \mathbb{C}_p$, we can find a $\mathbb{Z}_p$-direct summand Γ of $\mathbf{\Gamma}$ with rank $\dim_W \mathrm{Spf}(\mathbb{J})$ and an arithmetic point $P_0 \in \mathcal{A} \cap \mathrm{Spec}(\mathbb{J})(W)$ such that we have the following commutative diagram:

$$\begin{array}{ccc} \mathrm{Spf}(\mathbb{J}) & \longrightarrow & P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*) \\ \cap\downarrow & & \downarrow\cap \\ \mathrm{Spf}(\mathbb{I}) & \longrightarrow & \widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^* = \mathrm{Spf}(\mathbf{\Lambda}), \end{array}$$

which becomes Cartesian after localizing at each arithmetic point of $\mathrm{Spf}(\mathbb{J})$*, and* $\mathrm{Spf}(\mathbb{J})$ *gives a geometrically irreducible component of* $\mathrm{Spf}(\mathbb{I}) \times_{\mathrm{Spf}(\mathbf{\Lambda})} P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*)$*. Here $P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*)$ is the image of the multiplication by the point $P_0 \in \widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ inside $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$.*

In [H13, Lemma 5.1], it was claimed the diagram is Cartesian, which is wrong (as the fiber product could have several components). The correct statement is as above. This correction does not affect the results obtained in [H13].

Proof. Let $\pi : \mathrm{Spec}(\mathbb{J}) \to \mathrm{Spec}(\mathbf{\Lambda})$ be the projection. Then the smallest reduced closed subscheme $Z \subset \mathrm{Spec}(\mathbf{\Lambda})$ containing the topological image of π contains an infinitely many arithmetic points of weight κ. Since $\mathbb{J}$ is a domain with geometrically irreducible $\mathrm{Spec}(\mathbb{J})$, Z is geometrically irreducible. Take a basis $\{\gamma_1, \ldots, \gamma_m\}$ of $\mathbf{\Gamma}$, and write $\widehat{G} := \widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ as $\mathrm{Spf}(\widehat{W[t_j, t_j^{-1}]}_{j=1,\ldots,m})$ for the variable t_j corresponding to the dual basis $\{\gamma_j^*\}_j$ of $\mathbf{\Gamma}^*$. Let $P_1 \in Z$ be an arithmetic point of weight κ under $P \in \mathrm{Spec}(\mathbb{J})(W)$ (after replacing W by its finite extension, we can find a W-point P). Then by the variable change $t \mapsto P_1^{-1} \cdot t$ (which can be written as $t_j \mapsto \zeta_j \gamma_j^{-\kappa_2} t_j$ for suitable $\zeta_j \in \mu_{p^\infty}(W)$), the image of arithmetic points of $\mathrm{Spec}(\mathbb{J})$ of weight κ in Z are contained in $\mu_{p^\infty}^m(\overline{\mathbb{Q}}_p)$. Since Z is defined over W, $\Omega := Z(\mathbb{C}_p) \cap \mu_{p^\infty}^m(\mathbb{C}_p)$ is stable under $\mathrm{Gal}(\mathbb{K}[\mu_{p^\infty}]/\mathbb{K})$ for the quotient field $\mathbb{K}$ of W. Identify $\mathrm{Gal}(\mathbb{K}[\mu_{p^\infty}]/\mathbb{K})$ with a closed subgroup U of $\mathbb{Z}_p^\times$ by the p-adic cyclotomic character. Since W is a discrete valuation ring, U has to be also open in $\mathbb{Z}_p^\times$. Since $u \in U$ acts on Ω by $\zeta \mapsto \zeta^u$, Z is stable under the central action $\widehat{G} \ni t \mapsto t^u \in \widehat{G}$. Then by Lemma 4.1, we may assume, after making further variable change $t \mapsto \eta^{-1} t$ for $\eta \in \mu_{p^\infty}^m(W)$ (again replacing W by a finite extension if necessary), that Z is a formal subtorus; i.e., $Z = \widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*$ for a direct summand Γ of $\mathbf{\Gamma}$. Since $\mathbb{J}$ is an integral extension of the normal domain $\Lambda := W[[\Gamma]]$, by Matsumura [CRT, Theorems 9.4 and 15.2–3], we conclude $\dim_W \mathbb{J} = \dim_W Z = \mathrm{rank}_{\mathbb{Z}_p} \Gamma$. Then putting $P_0 = P_1 \cdot \eta$, we get the commutative diagram. Thus we have a natural closed immersion $\mathrm{Spf}(\mathbb{J}) \hookrightarrow \mathrm{Spf}(\mathbb{I}) \times_{\mathrm{Spf}(\mathbf{\Lambda}_W)} P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*) \subset \mathrm{Spf}(\mathbb{I})$ by the universality of the fiber product. Since $\mathbb{I}$ is an integral extension of the normal domain $\mathbf{\Lambda}$, by Matsumura [CRT, Theorem 15.1], we have $\dim_W \mathrm{Spf}(\mathbb{I}) \times_{\mathrm{Spf}(\mathbf{\Lambda}_W)} P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*) = \mathrm{rank}_{\mathbb{Z}_p} \Gamma = \dim_W \mathbb{J}$. Thus $\mathrm{Spec}(\mathbb{J})$ is an irreducible component of $\mathrm{Spf}(\mathbb{I}) \times_{\mathrm{Spf}(\mathbf{\Lambda}_W)} P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*)$.

We can see that $\mathrm{Spec}(\mathbb{J})$ is an irreducible component of the fiber product in a more concrete way. At each arithmetic point $P \in \mathrm{Spf}(\mathbb{I})$, the localized ring extension $\mathbb{I}_P/\mathbf{\Lambda}_P$ is an étale extension (cf. [HMI, Proposition 3.78]). The morphism $\mathrm{Spec}(\mathbb{J}) \to Z$ is dominant of equal dimension; so, it is generically étale. Thus $\Omega_{\mathrm{Spf}(\mathbb{J})/Z}$ is a torsion $\mathbb{J}$-module. Hence the étale locus of $\mathrm{Spec}(\mathbb{J})^{\text{ét}}$ over Z is equal to the complement of the support of $\Omega_{\mathrm{Spf}(\mathbb{J})/Z}$. In particular, $\mathrm{Spec}(\mathbb{J})^{\text{ét}}$ is an open dense subscheme of $\mathrm{Spec}(\mathbb{J})$. Since arithmetic points are dense in $\mathrm{Spec}(\mathbb{J})$, we can find an arithmetic point $P \in \mathrm{Spec}(\mathbb{J})^{\text{ét}}$. Then we have the commutative diagram localized at P:

$$\begin{tikzcd} \mathbb{J}_P & \Lambda_P \arrow[l, "P_0^*"'] \\ \mathbb{I}_P \arrow[u] & \mathbf{\Lambda}_P \arrow[l] \arrow[u] \end{tikzcd}$$

By our choice of P, all horizontal morphisms in the above diagram are smooth (and all members of the diagram are integral domains). Thus the above diagram is Cartesian. In particular, $\mathrm{Spf}(\mathbb{J})$ is a geometrically irreducible component of the fiber of $\mathrm{Spf}(\mathbb{I})$ over $P_0 \cdot (\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*)$. $\square$

Take Γ as in Lemma 5.1 given for $\mathbb{J}$, and write $\Lambda = W[[\Gamma]]$. Fix a basis $\gamma_1, \dots, \gamma_r \in \Gamma$ and identify Λ with $W[[T]](T = (T_i)_{i=1,\dots,r})$ by $\gamma_i \leftrightarrow t_i = 1 + T_i$. Let Q be the quotient field of Λ and fix its algebraic closure $\overline{Q}$. We embed $\mathbb{J}$ into $\overline{Q}$. We introduce one more notation:

(J3) If $\mathfrak{l}|p$, let $A_{\mathfrak{l}}$ be the image $a(\varpi_{\mathfrak{l}})$ in $\mathbb{J}$, and if $\mathfrak{l} \nmid \mathfrak{n}p$, fix a root $A_{\mathfrak{l}}$ in $\overline{Q}$ of $\det(T - \rho_{\mathbb{J}}(Frob_{\mathfrak{l}})) = 0$. Replacing $\mathbb{J}$ by a finite extension, we assume that $A_{\mathfrak{l}} \in \mathbb{J}$.

If the prime $\mathfrak{l}$ is clearly understood in the context, we simply write A for $A_{\mathfrak{l}}$. Recall the notation $A_P = P(A)$. Take and fix p^nth root t_i^{1/p^n} of t_i in $\overline{Q}$ ($i = 1, 2, \dots, r$) and consider

$$W[\mu_{p^n}][[T]][t^{1/p^n}] := W[\mu_{p^n}][[T_1, \dots, T_r]][t_1^{1/p^n}, \dots, t_r^{1/p^n}] \subset \overline{Q}$$

which is independent of the choice of t^{1/p^n}. Take a basis $\{\gamma = \gamma_1, \dots, \gamma_m\}$ of $\mathbf{\Gamma}$ over $\mathbb{Z}_p$ (containing $\{\gamma_1, \dots, \gamma_r\}$). We write t_j for the variable of $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \mathbf{\Gamma}^*$ corresponding to the dual basis of $\{\gamma_j\}_j$ of $\mathbf{\Gamma}^*$. We recall another result from [H13, Proposition 5.2] and its proof (to make the paper self-contained and also by the request of one of the referees):

Proposition 5.2 (Frobenius Eigenvalue Formula). *Let the notation and the assumption be as in (J1–3), and fix a prime ideal $\mathfrak{l}$ prime to $\mathfrak{n}$ as in (J3). Write $K := \mathbb{Q}[\mu_{p^\infty}]$ and $L_P = K(A_P)$ for each arithmetic point P with $\kappa(P) = \kappa$. Suppose*

($\mathrm{BT}_{\mathfrak{l}}$) *$L_P/K$ is a finite extension of degree bounded (independently of $P \in \mathcal{A}$) by a bound $B_{\mathfrak{l}} > 0$ dependent on $\mathfrak{l}$.*

Then, after making extension of scalars to a sufficiently large W, we have

$$A = A_{\mathfrak{l}} \in W[\mu_{p^n}][[T_1, \dots, T_r]][t_1^{1/p^n}, \dots, t_r^{1/p^n}] \cap \mathbb{J}$$

in $\overline{Q}$ for $0 \leq n \in \mathbb{Z}$, and there exists $s = (s_i) \in \mathbb{Q}_p^r$ and a constant $c \in W^\times$ such that $A(T) = ct^s = c\prod_i t_i^{s_i}$ $(t_i = 1 + T_i)$.

To simplify the notation, for $k = r$ or m, we often write $(\zeta\gamma^{-\kappa_2}t - 1)$ for the ideal in $W[[T_1, \dots, T_k]]$ generated by a tuple $(\zeta_j\gamma_j^{-\kappa_2}t_j - 1)$ for $j = 1, 2, \dots, k$ (where $\zeta = (\zeta_j)$ is also a tuple in $\mu_{p^\infty}^k(\overline{\mathbb{Q}}_p)$). The value of k should be clear in the context.

Proof. Since $\mathcal{A}$ is Zariski dense in $\mathrm{Spec}(\mathbb{J})$, for any $\mathrm{Gal}(\mathbb{K}[\mu_{p^\infty}]/\mathbb{K})$ for the field $\mathbb{K}$ of fractions of W, $\mathcal{A}_{st} := \bigcup_{\sigma \in \mathrm{Gal}(\mathbb{K}[\mu_{p^\infty}]/\mathbb{K})} \mathcal{A}^\sigma$ is Zariski dense in $\mathrm{Spec}(\mathbb{J})$ and stable under $\mathrm{Gal}(\mathbb{K}[\mu_{p^\infty}]/\mathbb{K})$. We replace $\mathcal{A}$ by $\mathcal{A}_{st}$. Let $Z = \mathrm{Spec}(\mathbf{\Lambda}/\mathfrak{a})$ for $\mathfrak{a} := \mathrm{Ker}(\mathbf{\Lambda} \to \mathbb{J})$ be the image of $\mathrm{Spec}(\mathbb{J})$ in $\mathrm{Spec}(\mathbf{\Lambda})$, and identify $\mathcal{A}$ with its image in Z. By Proposition 2.1 (and by a remark just above Theorem 3.1), we have only a finite number of generalized Weil l-numbers α of weight $[\kappa]f_{\mathfrak{l}}$ with bounded l-power denominator (i.e., $l^B\alpha$ is a Weil number of weight $([\kappa] + 2B)f_{\mathfrak{l}}$ for some $B > 0$) in $\bigcup_{P \in \mathcal{A}} L_P$ up to multiplication by p-power roots of unity. Here we can take $B = |\kappa_1|$. Hence, replacing $\mathcal{A}$ by a subset, we may assume that A_P for all $P \in \mathcal{A}$ hits one α of such generalized Weil l-numbers of weight $[\kappa]f_{\mathfrak{l}}$, up to p-power roots of unity, since the automorphic representation generated by $\mathbf{f}_P$ is not Steinberg because $\mathfrak{l} \nmid \mathfrak{n}$.

Let P_0 be as in Lemma 5.1 for this $\mathcal{A}$. By making a variable change $t \mapsto P_0 \cdot t$, we may assume that $P_0 = (t_j - 1)_{j=1,\dots,m}$, and $\mathcal{A}$ sits above $\mu_{p^\infty}^r(K)$, where we regard $\mu_{p^\infty}^r = \mu_{p^\infty} \otimes_{\mathbb{Z}_p} \Gamma^*$ as a subgroup of $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*$ (for $\Gamma \cong \mathbb{Z}_p^r$ as in Lemma 5.1) isomorphic to $\mathrm{Spf}(W[[\Gamma]]) = \mathrm{Spf}(\widehat{W[t_1, t_1^{-1}, \dots, t_r, t_r^{-1}]}) = \mathrm{Spf}(W[[T_1 \dots, T_r]])$ with $t_j = 1 + T_j$.

After the variable change $t \mapsto P_0 \cdot t$ ($\Leftrightarrow T_j \mapsto Y_j$) described above, suppose for the moment $\mathbb{J} \cong \widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*$ (i.e., P_0 goes to the identity of $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*$ with $\mathbb{J} = W[[Y_1 \dots, Y_r]] = \Lambda$ (writing y_j for the variable corresponding to t_j and $y_j = 1 + Y_j$ and hence $A \in \Lambda$). Choosing $\gamma_1, \dots, \gamma_r$ to be a generator of Γ for $r = \mathrm{rank}_{\mathbb{Z}_p} \Gamma$, we may assume that the projection $\mathbf{\Lambda} \to \mathbb{J}$ has kernel $(t_{r+1} - 1, \dots, t_m - 1)$. In down to earth terms, for $A_{\mathfrak{l}} = A(T)$ in (J3), the variable change $t \mapsto P_0 \cdot t$ is the variable change $T_j \mapsto Y_j = \zeta_j\gamma_j^{-\kappa_2}(1 + T_j) - 1$ with $Y = (Y_1, \dots, Y_m)$, and we have $A(Y)|_{Y=0} = A(T)|_{T_j = \zeta_j\gamma_j^{\kappa_2} - 1}$. Let

$$\Phi_1(Y) := \alpha^{-1}A(Y) = \alpha^{-1}A(\gamma^{-\kappa_2}(1 + T) - 1) \in W[[Y]]$$

and $\mathbf{L}$ be the composite of L_P for P running through $\mathcal{A}$. By this variable change, $\mathcal{A}$ is brought into a Zariski dense subset Ω_1 of $\mu_{p^\infty}^r(\overline{\mathbb{Q}}_p) \subset \widehat{\mathbb{G}}_m^r = \widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^*$ made up of ζ such that $\Phi_1(\zeta - 1)$ is a root of unity in $\mathbf{L}$. It is easy to see (e.g., [H11, Lemma 2.6]) that the group of roots of unity of $\mathbf{L}$ contains $\mu_{p^\infty}(K)$ as a subgroup of finite index, and we find a subset $\Omega \subset \Omega_1$ Zariski dense in $\widehat{\mathbb{G}}_m \otimes_{\mathbb{Z}_p} \Gamma^* = \mathrm{Spec}(\mathbb{J})$ and a root of unity ζ_1 such that $\{\Phi_1(\zeta - 1) | \zeta \in \Omega\} \subset \zeta_1\mu_{p^\infty}(K)$. Then $\Phi = \zeta_1^{-1}\Phi_1$ satisfies the assumption of Corollary 4.2, and for a root of unity ζ, we have $A(Y) = \zeta\alpha(1+Y)^s$ for $s \in \mathbb{Z}_p^r$, and $A(T) = \zeta\alpha(\gamma^{-\kappa_2}(1+T))^s$. Thus $A(T) = c(1+T)^s$ for a non-zero p-adic unit $c = \zeta\alpha\gamma^{-\kappa_2 s} \in W^\times$ as desired.

More generally, we now assume that $A \in W[[T]][t^{1/p^n}]$ (so, $\mathbb{J}$ is an extension of $W[[T_1 \dots, T_r]]$ and $A \in \mathbb{J} \cap W[[T_1 \dots, T_r]][t_1^{1/p^n}, \dots, t_r^{1/p^n}]$). Since

$$\mathrm{Spf}(W[[T]][t^{1/p^n}]]) \cong \widehat{\mathbb{G}}_m^r \xrightarrow{t \mapsto t^{p^n}} \widehat{\mathbb{G}}_m^r = \mathrm{Spf}(W[[T]]),$$

by applying the same argument as above to $W[[T]][t^{1/p^n}]]$, we get $A(T) = c(1+T)^{s/p^n}$ for $s \in \mathbb{Z}_p^r$ and a constant $c \neq 0$.

We thus need to show $A \in W[\mu_{p^n}][[T]][t^{1/p^n}]$ for sufficient large n, and then the result follows from the above argument. Again we make the variable change $T \mapsto Y$ we have already done. Replacing A by $\alpha^{-1}A$ for a suitable Weil l-number α of weight k (up to $\mu_{p^\infty}(\overline{\mathbb{Q}}_p)$), we may assume that there exists a Zariski dense set $\mathcal{A}_0 \subset \mathrm{Spec}(\mathbb{J})(\overline{\mathbb{Q}}_p)$ such that $P \cap \Lambda = (1 + Y - \zeta_P)$ for $\zeta_P \in \mu_{p^\infty}^r(\overline{\mathbb{Q}}_p)$ and $A_P \in \mu_{p^\infty}(\overline{\mathbb{Q}}_p)$ for all $P \in \mathcal{A}_0$. By another variable change $(1+Y) \mapsto \zeta(1+Y)$ for a suitable $\zeta \in \mu_{p^\infty}^r(\overline{\mathbb{Q}}_p)$, we may further assume that we have $P_0 \in \mathcal{A}_0$ with $\zeta_{P_0} = 1$ and $A_{P_0} = 1$ (i.e., choosing α well in $\alpha \cdot \mu_{p^\infty}(\overline{\mathbb{Q}}_p)$). We now write $\mathbb{J}'$ for the subalgebra of $\mathbb{J}$ topologically generated by A over $\Lambda = W[[Y]]$. Then we have $\mathbb{J}' := \Lambda[A] \subset \mathbb{J}$. Since $\mathbb{J}$ is geometrically irreducible, the base ring W is integrally closed in $\mathbb{J}'$. Since A is a unit in $\mathbb{J}$, we may embed the irreducible formal scheme $\mathrm{Spf}(\mathbb{J}')$ into $\widehat{\mathbb{G}}_m^r \times \widehat{\mathbb{G}}_m = \mathrm{Spf}(\widehat{W[y, y^{-1}, t', t'^{-1}]})$ by the surjective W-algebra homomorphism $\pi : \widehat{W[y, y^{-1}, t', t'^{-1}]} \twoheadrightarrow \mathbb{J}'$ sending (y, t') to $(1+Y, A)$. Write $Z \subset \widehat{\mathbb{G}}_m^r \times \widehat{\mathbb{G}}_m$ for the reduced image of $\mathrm{Spf}(\mathbb{J}')$. Thus we are identifying Λ with $\widehat{W[y, y^{-1}]}$ by $y \leftrightarrow 1+Y$. Then $P_0 \in Z$ is the identity element of $(\widehat{\mathbb{G}}_m^r \times \widehat{\mathbb{G}}_m)(\overline{\mathbb{Q}}_p)$. Since A is integral over Λ, it is a root of a monic polynomial $\Phi(t') = \Phi(y, t') = t'^d + a_1(y)t'^{d-1} + \cdots + a_d(y) \in \Lambda[t']$ irreducible over the quotient field Q of Λ, and we have $\mathbb{J}' \cong \Lambda[t']/(\Phi(y, t'))$. Thus $\mathbb{J}$ is free of rank, say d, over Λ; so, $\pi : Z \to \widehat{\mathbb{G}}_m^r = \mathrm{Spf}(\Lambda)$ is a finite flat morphism of degree d. We let $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ act on Λ by $\sum_{n=0}^\infty a_n Y^n \mapsto \sum_{n=0}^\infty a_n^\sigma Y^n$ and on $\Lambda[t']$ by $\sum_j A_j(Y)t'^j \mapsto \sum_j A_j^\sigma(Y)t'^j$ for $A_j(Y) \in \Lambda$. Note that $\Phi(\zeta_P, A_P) = 0$ for $P \in \mathcal{A}_0$. Since $A_P \in \mu_{p^\infty}(\overline{\mathbb{Q}}_p)$, $A_P^\sigma = A_P^{\nu(\sigma)}$ for the p-adic cyclotomic character $\nu : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathbb{Z}_p^\times$. Since W is a discrete valuation ring, for its quotient field F, the image of ν on $\mathrm{Gal}(\overline{\mathbb{Q}}_p/F)$ is an open subgroup U of $\mathbb{Z}_p^\times$. Thus we have $\Phi^\sigma(\zeta_P^{\nu(\sigma)}, A_P^{\nu(\sigma)}) = \Phi(\zeta_P, A_P)^\sigma = 0$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ and if $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}_p/F)$, $\Phi^\sigma = \Phi$. Thus we get

$$\Phi(\zeta_P^{\nu(\sigma)}, A_P^{\nu(\sigma)}) = \Phi(\zeta_P, A_P)^\sigma = 0 \ \text{ for all } P \in \mathcal{A}_0.$$

For $s \in \mathbb{Z}_p^\times$, consider the integral closed formal subscheme $Z_s \subset \widehat{\mathbb{G}}_m^r \times \widehat{\mathbb{G}}_m$ defined by $\Phi(y^s, t'^s) = 0$. If $s \in U$, we have $\mathcal{A}_0 \subset Z \cap Z_s$. Since Z and Z_s are finite flat over Λ and $\mathcal{A}_0$ is Zariski dense, we conclude $Z = Z_s$. Thus $Z \subset \widehat{\mathbb{G}}_m^r \times \widehat{\mathbb{G}}_m$ is stable under the diagonal action $(y, t') \mapsto (y^s, t'^s)$ for $s \in U$. By Lemma 4.1, Z is a formal multiplicative group and is a formal subtorus of $\widehat{\mathbb{G}}_m^r \times \widehat{\mathbb{G}}_m$, because $\mathbf{1} = P_0 \in Z$. The projection $\pi : Z \to \mathrm{Spf}(\Lambda) = \widehat{\mathbb{G}}_m^r$ is finite flat of degree d. So $\pi : Z \to \widehat{\mathbb{G}}_m^r$ is an isogeny. Thus we conclude $\mathrm{Ker}(\pi) \cong \prod_{j=1}^r \mu_{p^{m_j}}$ and hence $d = p^m$ for $m = \sum_j m_j \geq 0$. This implies $\mathbb{J}' = \Lambda[A] \subset W[\mu_{p^n}][[Y]][(1+Y)^{p^{-n}}] = W[\mu_{p^n}][[T]][t^{p^{-n}}]$ for $n = \max(m_j | j)$, as desired. $\square$

6 Proof of Theorem 3.1

Let the notation be as in the previous section; so, $K := \mathbb{Q}[\mu_{p^\infty}]$. Put $L_{\mathfrak{l},P} = K(\alpha_{\mathfrak{l},P})$. Suppose that there exist a set Σ of primes of positive upper density as in Theorem 3.1. By the assumption of the theorem, we have an infinite set $\mathcal{A}_{\mathfrak{l}}$ of arithmetic points of a fixed weight κ with $\kappa_2 - \kappa_1 \geq I$ of $\mathrm{Spec}(\mathbb{I})$ (independent of $\mathfrak{l} \in \Sigma$) such that

(B) if $\mathfrak{l} \in \Sigma$, $L_{\mathfrak{l},P}/K$ is a finite extension of bounded degree independent of $P \in \mathcal{A}_{\mathfrak{l}}$.

Let $\overline{\mathcal{A}}_{\mathfrak{l}}$ be the Zariski closure of $\mathcal{A}_{\mathfrak{l}}$ in $\mathrm{Spec}(\mathbb{I})$. As remarked after stating Theorem 3.1, we may assume that $\overline{\mathcal{A}}_{\mathfrak{l}}$ is geometrically irreducible of dimension $r \geq 1$ independent of $\mathfrak{l}$. Thus (J1) is satisfied for $(\mathcal{A}_{\mathfrak{l}}, \mathrm{Spec}(\mathbb{J}) := \overline{\mathcal{A}}_{\mathfrak{l}})$ for all $\mathfrak{l} \in \Sigma$.

Since we want to find a CM quadratic extension M/F in which p splits such that the component $\mathbb{I}$ has complex multiplication by M, by absurdity, we assume that $\mathbb{I}$ is a non-CM component and try to get a contradiction.

By (B) and Proposition 5.2 applied to $\mathfrak{l} \in \Sigma$, for $A_{\mathfrak{l}}$ in (J3), we have

$$A_{\mathfrak{l}}(t) = c_{\mathfrak{l}} \prod_{i=1}^r t_i^{s_{i,\mathfrak{l}}} \text{ for } s_{\mathfrak{l}} = (s_{i,\mathfrak{l}}) \in \mathbb{Q}_p^r \text{ and } c_{\mathfrak{l}} \in W^\times. \tag{28}$$

As proved in Proposition 5.2, we have $A_{\mathfrak{l}} \in W[\mu_{p^n}][[T_1, \dots, T_r]][t_1^{p^{-n}} - 1, \dots, t_r^{p^{-n}} - 1]]$. Since $\mathrm{rank}_\Lambda \mathbb{J} \geq \mathrm{rank}_\Lambda \Lambda[A_{\mathfrak{l}}]$ with $A_{\mathfrak{l}} \in \mathbb{J} \cap W[\mu_{p^n}][[T_1, \dots, T_r]][t_1^{p^{-n}} - 1, \dots, t_r^{p^{-n}} - 1]]$, the integer n is also bounded independent of $\mathfrak{l}$. Thus by the variable change $t_i \mapsto t_i^{p^n}$, we may assume that $A_{\mathfrak{l}} \in W[[T_1, \dots, T_r]]$ for all $\mathfrak{l} \in \Sigma$ (and hence $s_i \in \mathbb{Z}_p$). Up until this point, we only used the existence of $\mathcal{A}_{\mathfrak{l}}$ whose weight $\kappa_{\mathfrak{l}}$ depends on $\mathfrak{l}$ to conclude the above explicit form (28) of $A_{\mathfrak{l}}$. Since $A_{\mathfrak{l}}$ in (28) is independent of weight $\kappa_{\mathfrak{l}}$, we may now take any weight κ (with $\kappa_2 - \kappa_1 \geq I$) discarding the original choice $\kappa_{\mathfrak{l}}$ dependent on $\mathfrak{l}$ (as remarked after stating Theorem 3.1 that κ is allowed to be dependent on $\mathfrak{l}$). Once κ is chosen, we can take $\mathcal{A}$ to be all the arithmetic points of weight κ of $\mathrm{Spec}(\mathbb{J})$ (so, we may assume that $\mathcal{A} = \mathcal{A}_{\mathfrak{l}}$ is also independent of $\mathfrak{l}$). We use the symbols introduced in the proof of Proposition 5.2. We now vary $\mathfrak{l} \in \Sigma$.

Pick a p-power root of unity $\zeta \neq 1$ of order $1 < a = p^e$ and consider $\underline{\zeta} := (\zeta, \zeta, \dots, \zeta) \in \mu^r_{p^\infty}$, and write $\alpha_{f,\mathfrak{l}} = \alpha_{\mathfrak{l}} = A_{\mathfrak{l}}(\underline{\gamma}^{\kappa_2-1})$ for $\underline{\gamma}^{\kappa_2-1} := (\gamma_1^{\kappa_2-1}, \dots, \gamma_r^{\kappa_2-1})$ and $\alpha_{g,\mathfrak{l}} = \beta_{\mathfrak{l}} = A_{\mathfrak{l}}(\underline{\zeta\gamma^{\kappa_2-1}})$ for $\underline{\zeta\gamma^{\kappa_2-1}} := (\zeta\gamma_1^{\kappa_2-1}, \dots, \zeta\gamma_r^{\kappa_2-1})$. They are generalized Weil l-numbers of weight $[\kappa]\underline{f_{\mathfrak{l}}}$. Write $f = \mathbf{f}_P$ for

$$P = (\underline{t} - \underline{\gamma}^{\kappa_2-1}) := (t_1 - \gamma_1^{\kappa_2-1}, \dots, t_r - \gamma_r^{\kappa_2-1})$$

and g for the cusp form $\mathbf{f}_{P'}$ for $P' = (\underline{t} - \underline{\zeta\gamma^{\kappa_2-1}})$. Consider the compatible system of Galois representation associated with f and g. Pick a prime $\mathfrak{Q}$ of $\mathbb{Q}(f, g) = \mathbb{Q}(f)(g)$ (with residual characteristic q sufficiently large) split over $\mathbb{Q}$. Write $\rho_{f,\mathfrak{Q}}$ (resp. $\rho_{g,\mathfrak{Q}}$) for the $\mathfrak{Q}$-adic member of the system associated with f (resp. g). Thus $\rho_{?,\mathfrak{Q}}$ has values in $\mathrm{GL}_2(\mathbb{Z}_q)$. Since proper compact subgroups of $SL_2(\mathbb{Z}_q)$ are either finite, open in a normalizer of a torus, open in a Borel subgroup or open in a unipotent subgroup, the non-CM property of f and g tells us that $\mathrm{Im}(\rho_{?,\mathfrak{Q}})$ contains an open subgroup of $SL_2(\mathbb{Z}_q)$ (e.g., [Di05, Sect. 0.1] or [CG14, Corollary 4.4]).

For a continuous representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \mathrm{GL}_2(R)$ (for $R = \overline{\mathbb{Q}}_q$ or any other topological ring), let $\rho^{sym\otimes j}$ denote the jth *symmetric* tensor representation into $\mathrm{GL}_{j+1}(R)$. Suppose that f [and hence g by the equivalence of (CM2–3)] does not have complex multiplication. Then by openness of $\mathrm{Im}(\rho_{?,\mathfrak{Q}})$ in $\mathrm{GL}_2(\mathbb{Z}_q)$, $\rho^{sym\otimes j}_{?,\mathfrak{Q}}$ is absolutely irreducible for all $j \geq 0$, and also the Zariski closure of $\mathrm{Im}(\rho^{sym\otimes j}_{?,\mathfrak{Q}})$ is connected isomorphic to a quotient of GL(2) by a finite subgroup in the center. Since $\beta_{\mathfrak{l}} = \zeta_{\mathfrak{l}}\alpha_{\mathfrak{l}}$ for a root of unity $\zeta_{\mathfrak{l}} = \prod_{i=1}^r \zeta^{s_{i,\mathfrak{l}}}$ (for $s_i \in \mathbb{Q}_p$ as in Proposition 5.2), we have $\beta_{\mathfrak{l}}^a = \alpha_{\mathfrak{l}}^a$ [for a p-power a with $\zeta^{s_{i,\mathfrak{l}}a} = 1$ $(j = 1, 2, \dots, r)$]. Thus $\mathrm{Tr}(\rho^a_{f,\mathfrak{Q}}(Frob_{\mathfrak{l}})) = \mathrm{Tr}(\rho^a_{g,\mathfrak{Q}}(Frob_{\mathfrak{l}}))$ for all prime $\mathfrak{l} \in \Sigma$ prime to $p\mathfrak{n}$, where $\mathrm{Tr}(\rho^a_{?,\mathfrak{Q}})(g)$ is just the trace of ath matrix power $\rho^a_{?,\mathfrak{Q}}(g)$. Since the continuous functions $\mathrm{Tr}(\rho^a_{f,\mathfrak{Q}})$ and $\mathrm{Tr}(\rho^a_{g,\mathfrak{Q}})$ match on $\widetilde{\Sigma} := \{Frob_{\mathfrak{l}} | \mathfrak{l} \in \Sigma\}$, we find that $\mathrm{Tr}(\rho^a_{f,\mathfrak{Q}}) = \mathrm{Tr}(\rho^a_{g,\mathfrak{Q}})$ on the closure of $\widetilde{\Sigma}$. Since we have

$$\mathrm{Tr}(\rho^a) = \mathrm{Tr}(\rho^{sym\otimes a}) - \mathrm{Tr}(\rho^{sym\otimes(a-2)} \otimes \det(\rho)),$$

we get over $\widetilde{\Sigma}$,

$$\mathrm{Tr}(\rho^{sym\otimes a}_{f,\mathfrak{Q}}) - \mathrm{Tr}(\rho^{sym\otimes(a-2)}_{f,\mathfrak{Q}} \otimes \det(\rho_{f,\mathfrak{Q}})) = \mathrm{Tr}(\rho^{sym\otimes a}_{g,\mathfrak{Q}}) - \mathrm{Tr}(\rho^{sym\otimes(a-2)}_{g,\mathfrak{Q}} \otimes \det(\rho_{g,\mathfrak{Q}})).$$

which implies

$$\mathrm{Tr}(\rho^{sym\otimes a}_{f,\mathfrak{Q}} \oplus (\rho^{sym\otimes(a-2)}_{g,\mathfrak{Q}} \otimes \det(\rho_{g,\mathfrak{Q}}))) = \mathrm{Tr}(\rho^{sym\otimes a}_{g,\mathfrak{Q}} \oplus (\rho^{sym\otimes(a-2)}_{f,\mathfrak{Q}} \otimes \det(\rho_{f,\mathfrak{Q}})))$$

over $\widetilde{\Sigma}$. Since Σ has positive upper Dirichlet density, by Rajan [Rj98, Theorem 2], there exists an open subgroup $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/F)$ such that as representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$

$$\rho^{sym\otimes a}_{f,\mathfrak{Q}} \oplus (\rho^{sym\otimes(a-2)}_{g,\mathfrak{Q}} \otimes \det(\rho_{g,\mathfrak{Q}})) \cong \rho^{sym\otimes a}_{g,\mathfrak{Q}} \oplus (\rho^{sym\otimes(a-2)}_{f,\mathfrak{Q}} \otimes \det(\rho_{f,\mathfrak{Q}})).$$

Since $\mathrm{Im}(\rho_{?,\mathfrak{Q}})$ contains open subgroup of $SL_2(\mathbb{Z}_q)$, $\rho_{?,\mathfrak{Q}}^{sym\otimes j}$ restricted to $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$ is absolutely irreducible for all $j \geq 0$. Therefore, as representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$, we conclude $\rho_{f,\mathfrak{Q}}^{sym\otimes a} \cong \rho_{g,\mathfrak{Q}}^{sym\otimes a}$ from the difference of the dimensions of absolutely irreducible factors in the left and right-hand side. By Calegari and Gee [CG14, Corollary 4.4 and Theorem 7.1], each member of $\rho_f^{sym\otimes a}$ and $\rho_g^{sym\otimes a}$ is absolutely irreducible over $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$. Thus the ath symmetric tensor product of the two compatible systems ρ_f and ρ_g are isomorphic to each other over $\mathrm{Gal}(\overline{\mathbb{Q}}/K)$. Again by Rajan [Rj98, Theorem 2], as compatible systems of Galois representations of the entire group $\mathrm{Gal}(\overline{\mathbb{Q}}/F)$, we find $\rho_f^{sym\otimes a} \cong \rho_g^{sym\otimes a} \otimes \chi$ for a finite order character $\chi : \mathrm{Gal}(\overline{\mathbb{Q}}/F) \to \overline{\mathbb{Q}}^\times$. In particular, we get the identity of their $\mathfrak{P}$-adic members

$$\rho_{f,\mathfrak{P}}^{sym\otimes a} \cong \rho_{g,\mathfrak{P}}^{sym\otimes a} \otimes \chi.$$

Note that $F_p := F \otimes_{\mathbb{Q}} \mathbb{Q}_p \cong \prod_{\mathfrak{p}|p} F_{\mathfrak{p}}$ for the $\mathfrak{p}$-adic completion $F_{\mathfrak{p}}$ of F at prime factors $\mathfrak{p}$ of p. Pick a prime $\mathfrak{p}|p$ of F. Then $\mathfrak{p} = \{x \in O : |i_p(\sigma(x))|_p < 1\}$ for an embedding $\sigma : F \hookrightarrow \overline{\mathbb{Q}}$. Then $i_p \circ \sigma$ embeds $F_{\mathfrak{p}}$ into $\overline{\mathbb{Q}}_p$ continuously. Write $I_{\mathfrak{p}}$ for the set of all continuous embeddings of $F_{\mathfrak{p}}$ into $\overline{\mathbb{Q}}_p$ (including $i_p \circ \sigma$). By (Ram), we can write the restriction $\rho_{?,\mathfrak{P}}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/F_{\mathfrak{p}})}$ in an upper triangular form $\begin{pmatrix} \epsilon_{?,\mathfrak{p}} & * \\ 0 & \delta_{?,\mathfrak{p}} \end{pmatrix}$ (up to isomorphisms) with

$$\delta_{?,\mathfrak{p}}([u, F_{\mathfrak{p}}]) = u^{-\kappa_1} \text{ and } \epsilon_{?,\mathfrak{p}}([u, F_{\mathfrak{p}}]) = u^{-\kappa_2} \text{ for } u \in O_{\mathfrak{p}}^\times \text{ sufficiently close to 1.} \tag{29}$$

Here $u^k = \prod_{i_p\circ\tau\in I_{\mathfrak{p}}} \tau(u)^{k_\tau}$ for $k = \sum_{\tau\in I} k_\tau$ (as the component of u in $F_{\mathfrak{p}'}^\times$ at $\mathfrak{p}' \neq \mathfrak{p}$ for other primes $\mathfrak{p}'|p$ is trivial in $F_p^\times$). This property distinguishes $\delta_{?,\mathfrak{p}}$ from $\epsilon_{?,\mathfrak{p}}$. Regard $\delta_{?,\mathfrak{p}}$ and $\epsilon_{?,\mathfrak{p}}$ as characters of $F_{\mathfrak{p}}^\times$ by local class field theory, and put $\delta_?((u_{\mathfrak{p}})_{\mathfrak{p}}) = \prod_{\mathfrak{p}} \delta_{?,\mathfrak{p}}(u_{\mathfrak{p}})$ and $\epsilon_?((u_{\mathfrak{p}})_{\mathfrak{p}}) = \prod_{\mathfrak{p}} \epsilon_{?,\mathfrak{p}}(u_{\mathfrak{p}})$ for $(u_{\mathfrak{p}})_{\mathfrak{p}} \in \prod_{\mathfrak{p}} F_{\mathfrak{p}}^\times$ as characters of $F_p^\times = \prod_{\mathfrak{p}} F_{\mathfrak{p}}^\times$ (in order to regard these characters as those of $F_p^\times$ not of the single $F_{\mathfrak{p}}^\times$). Then more precisely than (29), we have from our choice of f and g

$$\epsilon_f(\gamma_i) = \gamma_i^{-\kappa_2}, \epsilon_g(\gamma_i) = \zeta\gamma_i^{-\kappa_2}, \delta_f(\gamma_i) = \gamma_i^{-\kappa_1} \text{ and } \delta_g(\gamma_i) = \zeta^{-1}\gamma_i^{-\kappa_1} \tag{30}$$

as $\epsilon_{P'}(\gamma_i) = \zeta$ and $\epsilon_P(\gamma_i) = 1$ for all i. Since $\mathbf{\Gamma} \subset O_p^\times \subset F_p^\times$, and hence we may consider $\delta_?(\gamma_i)$ and $\epsilon_?(\gamma_i)$. Then we have from $\rho_{f,\mathfrak{P}}^{sym\otimes a} \cong \rho_{g,\mathfrak{P}}^{sym\otimes a} \otimes \chi$

$$\{\epsilon_f^j\delta_f^{a-j} | j = 0, \ldots, a\} = \{\epsilon_g^j\delta_g^{a-j}\chi | j = 0, \ldots, a\}.$$

Therefore we conclude from $\kappa_2 - \kappa_1 \geq I$ and (29) that $\epsilon_f^j\delta_f^{a-j} = \epsilon_g^j\delta_g^{a-j}\chi$. This means

$$\gamma_i^{-\kappa_2 j-\kappa_1(a-j)} = \epsilon_f^j\delta_f^{a-j}(\gamma_i) = \epsilon_g^j\delta_g^{a-j}\chi(\gamma_i) = \gamma_i^{-\kappa_2 j-\kappa_1(a-j)}\zeta^{2j-a}\chi(\gamma_i).$$

Therefore we get $\chi(\gamma_i) = \zeta^{a-2j}$ which has to be independent of j, a contradiction, as we can choose the p-power order of ζ as large as we want. Thus f and hence g must have complex multiplication by the same CM quadratic extension $M_{/F}$ by (CM1–3), and hence $\mathbb{I}$ is a CM component. $\square$

Acknowledgements The author is partially supported by the NSF grants: DMS 0753991 and DMS 1464106.

References

[B02] D. Blasius, Hilbert modular forms and the Ramanujan conjecture, in *Noncommutative Geometry and Number Theory*. Aspects of Mathematics, vol. E37 (Vieweg, Wiesbaden, 2006), pp. 35–56
[BR93] D. Blasius, J.D. Rogawski, Motives for Hilbert modular forms. Invent. Math. **114**, 55–87 (1993)
[C90] C.-L. Chai, Arithmetic minimal compactification of the Hilbert-Blumenthal moduli spaces. Ann. Math. **131**, 541–554 (1990)
[CG14] F. Calegari, T. Gee, Irreducibility of automorphic Galois representations of GL(n), n at most 5. Ann. Inst. Fourier **63**, 1881–1912 (2013)
[CRT] H. Matsumura, *Commutative Ring Theory* (Cambridge University Press, Cambridge, 1986)
[D71] P. Deligne, Travaux de Shimura, in *Sem. Bourbaki, Exp. 389*. Lecture Notes in Mathematics, vol. 244 (Springer, Berlin, 1971), pp. 123–165
[Di03] M. Dimitrov, Compactifications arithmétiques des variétés de Hilbert et formes modulaires de Hilbert pour $\Gamma_1(\mathfrak{c}, \mathfrak{n})$, in *Geometric Aspects of Dwork Theory*, vol. I, II (Walter de Gruyter, Berlin, 2004), pp. 527–554
[Di05] M. Dimitrov, Galois representations modulo p and cohomology of Hilbert modular varieties. Ann. Sci. École Norm. Sup. (4) **38**, 505–551 (2005)
[dJ95] A.J. De Jong, Crystalline Dieudonné module theory via formal and rigid geometry. Inst. Hautes Études Sci. Publ. Math. **82**, 5–96 (1995)
[dJ98] A.J. De Jong, Erratum to: "Crystalline Dieudonné module theory via formal and rigid geometry". Inst. Hautes Études Sci. Publ. Math. **87**, 175 (1998)
[DP94] P. Deligne, G. Pappas, Singularités des espaces de modules de Hilbert, en les caractéristiques divisant le discriminant. Compos. Math. **90**, 59–79 (1994)
[DR80] P. Deligne, K.A. Ribet, Values of abelian L-functions at negative integers over totally real fields. Invent. Math. **59**, 227–286 (1980)
[DT04] M. Dimitrov, J. Tilouine, Variété de Hilbert et arithmétique des formes modulaires de Hilbert pour $\Gamma_1(\mathfrak{c}, N)$, in *Geometric Aspects of Dwork's Theory*, vol. I, II (Walter de Gruyter, Berlin, 2004), pp. 555–614
[GME] H. Hida, *Geometric Modular Forms and Elliptic Curves*, 2nd edn. (World Scientific, Singapore, 2012)
[H06] H. Hida, Anticyclotomic main conjectures. Ducumenta Math. **2006**, 465–532 (2006). Extra Volume Coates
[H11] H. Hida, Hecke fields of analytic families of modular forms. J. Am. Math. Soc. **24**, 51–80 (2011)
[H13] H. Hida, Hecke fields of Hilbert modular analytic families. Contemp. Math. **614**, 97–138 (2014)
[H14] H. Hida, Transcendence of Hecke operators in the big Hecke algebra. Duke Math. J. **163**, 1655–1681 (2014)

[H86a] H. Hida, Galois representations into $GL_2(\mathbb{Z}_p[[X]])$ attached to ordinary cusp forms. Invent. Math. **85**, 545–613 (1986)

[H86b] H. Hida, Hecke algebras for GL_1 and GL_2, in *Sém. de Théorie des Nombres, Paris 1984–85*. Progress in Mathematics, vol. 63 (Birkhäuser, Boston, 1986), pp. 131–163

[H89] H. Hida, Nearly ordinary Hecke algebras and Galois representations of several variables, in *Proc. JAMI Inaugural Conference*, Supplement to Am. J. Math. 115–134 (1989)

[H91] H. Hida, On p-adic L-functions of $GL(2) \times GL(2)$ over totally real fields. Ann. Inst. Fourier **41**, 311–391 (1991)

[H95] H. Hida, Control theorems of p-nearly ordinary cohomology groups for $SL(n)$. Bull. Soc. Math. Fr. **123**, 425–475 (1995)

[H96] H. Hida, On the search of genuine p-adic modular L-functions for $GL(n)$. Mém. Soc. Math. Fr. **67** (1996)

[HMI] H. Hida, *Hilbert Modular Forms and Iwasawa Theory*. Oxford Mathematical Monographs (Oxford University Press, Oxford, 2006). (a list of errata posted at www.math.ucla.edu/~hida)

[Ho68] T. Honda, Isogeny classes of abelian varieties over finite fields. J. Math. Soc. Jpn. **20**, 83–95 (1968)

[HT93] H. Hida, J. Tilouine, Anticyclotomic Katz p-adic L-functions and congruence modules. Ann. Sci. École Norm. Sup. 4th series **26**, 189–259 (1993)

[IAT] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions* (Princeton University Press, Princeton; Iwanami Shoten, Tokyo, 1971)

[K78] N.M. Katz, p-adic L-functions for CM fields. Invent. Math. **49**, 199–297 (1978)

[Ko92] R. Kottwitz, Points on Shimura varieties over finite fields. J. Am. Math. Soc. **5**, 373–444 (1992)

[MFG] H. Hida, *Modular Forms and Galois Cohomology* (Cambridge University Press, Cambridge, 2000). (a list of errata posted at www.math.ucla.edu/~hida)

[NMD] S. Bosch, W. Lütkebohmert, M. Raynaud, *Néron Models* (Springer, New York, 1990)

[PAF] H. Hida, *p-Adic Automorphic Forms on Shimura Varieties* (Springer, New York, 2004). (a list of errata posted at www.math.ucla.edu/~hida)

[Rj98] C.S. Rajan, On strong multiplicity one for l-adic representations. Int. Math. Res. Not. **1998**(3), 161–172 (1998)

[Rp78] M. Rapoport, Compactifications de l'espace de modules de Hilbert-Blumenthal. Compos. Math. **36**, 255–335 (1978)

[Se97] J.-P. Serre, Répartition asymptotique des valeurs propres de l'opérateur de Hecke T_p. J. Am. Math. Soc. **10**, 75–102 (1997)

[Sh70] G. Shimura, On canonical models of arithmetic quotients of bounded symmetric domains. Ann. Math. **91**, 144–222 (1970); II **92**, 528–549 (1970)

[ST13] S.-W. Shin, N. Templier, On fields of rationality for automorphic representations. Compos. Math. **150**, 2003–2053 (2014)

[W55] A. Weil, On a certain type of characters of the idéle-class group of an algebraic number-field, in *Proceedings of the International Symposium on Algebraic Number Theory, Tokyo & Nikko, 1955* (Science Council of Japan, Tokyo, 1956), pp. 1–7 (the proceedings is posted at www.jmilne.org/math/Documents/TokyoNikko1955.pdf)

[Z14] B. Zhao, Local indecomposability of Hilbert modular Galois representations. Ann. Inst. Fourier **64**, 1521–1560 (2014)

The Trace Formula and Prehomogeneous Vector Spaces

Werner Hoffmann

Abstract We describe an approach to express the geometric side of the Arthur–Selberg trace formula in terms of zeta integrals attached to prehomogeneous vector spaces. This will provide explicit formulas for weighted orbital integrals and for the coefficients by which they are multiplied in the trace formula. We implement this programme for the principal unipotent conjugacy class. The method relies on certain convergence results and uses the notions of induced conjugacy classes and canonical parabolic subgroups. So far, it works for certain types of conjugacy classes, which covers all classes appearing in classical groups of absolute rank up to two.

MSC: Primary 11F72; Secondary 11S90, 11M41

1 Introduction

The trace formula is an equality between two expansions of a certain distribution on an adelic group. The spectral side of the formula encodes valuable information about automorphic representations of the group. Although the geometric side is regarded to be the source of information, it is far from explicit. It is a sum of so-called weighted orbital integrals, each multiplied with a coefficient that carries global arithmetic information. So far, those coefficients have only been evaluated in some special cases. Arthur remarked on p. 112 of [A-int] that "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". In the meantime, as reflected in the present proceedings, further applications have emerged which revive the interest in more detailed information on those coefficients and the weight factors of weighted orbital integrals.

The problem stems from the fact that the trace distribution is defined by an integral that does not converge without regularisation. The most successful

W. Hoffmann (✉)
Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany
e-mail: hoffmann@math.uni-bielefeld.de

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*,
Simons Symposia, DOI 10.1007/978-3-319-41424-9_5

method to accomplish this is Arthur's truncation [A-trI]. However, it does not yield useful formulas for the contributions from non-semisimple conjugacy classes to the geometric side. In the original rank-one trace formula (e.g., [A-rk1]), they were regularised by damping factors, which led to an expression containing zeta integrals. Shintani [Sh] observed that such integrals would also appear in the dimension formula for Siegel modular forms, which can be regarded as a special case of the trace formula, if one were able to prove convergence. The same method was applied by Flicker [Fl] to the group GL_3, but for groups of higher rank, the difficulties piled up. Arthur bypassed them by a clever invariance argument, which worked for unipotent conjugacy classes, and by reducing the general case to the unipotent one [A-mix]. The price to pay was that most coefficients and weight factors remained undetermined.

We take up the original approach and remove some of the obstacles on the way to express the regularised terms on the geometric side by zeta integrals. In many cases, these integrals are supported on prehomogeneous vector spaces which appear as subquotients of canonical parabolic subgroups of unipotent elements. Moreover, just as induced representations play an important role on the spectral side, we systematically apply the notion of induced conjugacy classes on the geometric side. So far, this approach has been successful for certain types of conjugacy classes, which suffice for a complete treatment of classical groups of absolute rank up to 2. The details, including the necessary estimates, can be found in a joint paper [HoWa] with Wakatsuki.

Over several years of work on this project, something like a general formula was gradually emerging, changing shape as more and more conjugacy classes with new features were covered. Incomplete as the results may be, they should perhaps be made available to a wider audience now together with an indication of the remaining difficulties.

Let us describe the setting in more detail. We consider a connected reductive linear algebraic group G defined over a number field F. The group $G(\mathbb{A})$ of points with coordinates in the ring $\mathbb{A}$ of adeles of F acts by right translations on the homogeneous space $G(F)\backslash G(\mathbb{A})$, which carries an invariant measure coming from a Haar measure on $G(\mathbb{A})$ and the counting measure on $G(F)$. The resulting unitary representation R_G of $G(\mathbb{A})$ on the Hilbert space $L^2(G(F)\backslash G(\mathbb{A}))$ can be integrated to a representation of the Banach algebra $L^1(G(\mathbb{A}))$, and for an element f of the latter, $R_G(f)$ is an integral operator with kernel

$$K_G(x,y) = \sum_{\gamma \in G(F)} f(x^{-1}\gamma y).$$

If G is F-anisotropic, then $G(F)\backslash G(\mathbb{A})$ is compact, so the integral

$$J(f) = \int_{G(F)\backslash G(\mathbb{A})} K_G(x,x)\, dx$$

converges for smooth compactly supported functions f and defines a distribution J on $G(\mathbb{A})$. Now we have the geometric expansion

$$J(f) = \sum_{[\gamma]} \int_{G^\gamma(F)\backslash G(\mathbb{A})} f(x^{-1}\gamma x)\, dx,$$

where G^γ is the centraliser of γ, and the spectral expansion

$$\operatorname{tr} R_G(f) = \sum_{\pi} a^G(\pi) \operatorname{tr} \pi(f),$$

where $a^G(\pi)$ is the multiplicity of the irreducible representation π of $G(\mathbb{A})$ in $L^2(G(F)\backslash G(\mathbb{A}))$. The Selberg trace formula in this case is the identity

$$\operatorname{tr} R_G(f) = J(f).$$

If the centre of $G(\mathbb{A})$ is non-compact, then $R_G(f)$ has no discrete spectrum, hence its trace is not defined. Either one has to fix a central character or one has to replace the group by its largest closed normal subgroup $G(\mathbb{A})^1$ with compact centre. If G has proper parabolic subgroups P defined over F, both sides of the formula will still diverge. One has to take into account the analogous unitary representations R_P of $G(\mathbb{A})^1$ on the spaces $L^2(N(\mathbb{A})P(F)\backslash G(\mathbb{A})^1)$, where the letter N will always denote the unipotent radical of the group P in the current context. By choosing a Levi component M of P, one can view R_P as the representation induced from the representation R_M, after the latter has been inflated to a representation of $P(\mathbb{A})$ by composing it with the projection $P(\mathbb{A}) \to M(\mathbb{A})$. The kernel function for $R_P(f)$ with $f \in C_c^\infty(G(\mathbb{A})^1)$ is

$$K_P(x, y) = \sum_{\gamma \in P(F)/N(F)} \int_{N(\mathbb{A})} f(x^{-1}\gamma n y)\, dn,$$

where we normalise the Haar measure on the group $N(\mathbb{A})$ in such a way that $N(F)\backslash N(\mathbb{A})$ has measure 1. This can be written as a single integral over $P(F)N(\mathbb{A})$, whose integrand is compactly supported locally uniformly in x and y. The trace distribution is defined as

$$J^T(f) = \int_{G(F)\backslash G(\mathbb{A})^1} \sum_P K_P(x, x)\hat{\tau}_P^T(x)\, dx,$$

where P runs over all parabolic F-subgroups including G itself. The functions $\hat{\tau}_P^T$ are, up to sign, certain characteristic functions on $G(\mathbb{A})$ depending on a truncation parameter T and on the choice of a maximal compact subgroup $\mathbf{K}$ of $G(\mathbb{A})$. We will recall their definition in Sect. 3.1 below, noting for the moment that $\hat{\tau}_G^T(x) = 1$. Their alternating signs are responsible for cancellations that make the integrand rapidly decreasing and allowed Arthur to prove absolute convergence [A-trI].

Actually, his argument was more subtle and led to a geometric expansion of $J^T(f)$, later called the coarse geometric expansion. It represents an intermediate stage on the way to the fine geometric expansion [A-mix]. The latter depends on the choice of a finite set S of valuations of F including the Archimedean ones and has the shape

$$J^T(f) = \sum_{[M]} \sum_{[\gamma]_{M,S}} a^M(S,\gamma) J_M^T(\gamma,f).$$

Here f is a smooth compactly supported function on $G(F_S)^1$ suitably extended to $G(\mathbb{A})^1$, where F_S is the product of completions F_v of F with respect to $v \in S$. The summation runs over the conjugacy classes of Levi F-subgroups M of G and, for each such class, over the classes of elements γ with respect to the finest equivalence relation with the following properties. Elements with $M(F)$-conjugate semisimple components are equivalent, and elements with the same semisimple component σ and $M_\sigma(F_S)$-conjugate unipotent components are also equivalent. The weighted orbital integral $J_M^T(\gamma,f)$ is an integral with respect to a certain non-invariant measure that is undetermined in general. It is supported on the F_S-valued points of the conjugacy class of G induced from that of γ in M. The coefficients $a^M(S,\gamma)$ do not depend on the ambient group G. They have been determined for semisimple elements [A-mix], for M of F-rank one [Ho-rk1], for $M = \mathrm{GL}_3$ [Fl, Ma] and, with the methods presented here, for the symplectic group of rank two [HoWa].

2 Prerequisites

In this section we collect some results in order to avoid interruptions of the arguments to follow. Unless stated otherwise, all affine varieties and linear algebraic groups that appear are assumed to be connected and defined over a given field F. When we speak of orbits in a G-variety V defined over F, we mean geometric orbits defined over F, i.e., minimal invariant F-subvarieties O such that $O(F)$ is non-empty. This applies, in particular, to conjugacy classes. By Proposition 12.1.2 of [Sp], every element of $V(F)$ belongs to an orbit, and an orbit remains a single orbit under base change to an extension field.

2.1 Induction of Conjugacy Classes

The following well-known result has been proved by Lusztig and Spaltenstein [LS] for unipotent conjugacy classes, and its extension to general conjugacy classes can be found in [Ho-ind].

Theorem 1. *Let P be a parabolic subgroup of a reductive group G with unipotent radical N and C a conjugacy class in a Levi component M of P. Then there is a unique dense P-conjugacy class C' in CN and a unique conjugacy class $\tilde{C}$ in G such that $\tilde{C} \cap P = C'$.*

We will write $\tilde{C} = \mathrm{Ind}_P^G C$ and $C' = \mathrm{Infl}_M^P C$. The map Ind_P^G is called induction of conjugacy classes from M to G via P, and the map Infl_M^P will be called inflation of conjugacy classes from M to P. The Levi components of P are naturally isomorphic to P/N and will be called Levi subgroups of G. We denote by P^{infl} the set of all elements $\gamma \in P$ for which the range of the endomorphism $\mathrm{Ad}\,\gamma - \mathrm{id}$ of the Lie algebra of P contains the Lie algebra of N.

Theorem 2. *(i) If M is a Levi component of two parabolic subgroups P and Q of G, then $\mathrm{Ind}_P^G C = \mathrm{Ind}_Q^G C$, whence this set can be denoted by $\mathrm{Ind}_M^G C$.*

(ii) If $M \subset M'$ are Levi subgroups of G, then $\mathrm{Ind}_M^G C = \mathrm{Ind}_{M'}^G \mathrm{Ind}_M^{M'} C$.

(iii) The union of all the sets $C'(F)$ with $C' = \mathrm{Infl}_M^P C$ for conjugacy classes C in M over F equals $P^{\mathrm{infl}}(F)$.

(iv) Given $\gamma \in G(F)$, the set $\mathcal{P}_\gamma^{\mathrm{infl}}$ of parabolic subgroups P such that $\gamma \in P^{\mathrm{infl}}$ is a finite algebraic subset of the flag variety defined over F.

The first two assertions have been proved in [LS], the other ones in [Ho-ind].

2.2 Prehomogeneous Varieties

Let G be a linear algebraic group. A prehomogeneous G-variety is an irreducible G-variety V possessing a dense G-orbit O. The "generic" stabilisers G^ξ (which may be non-connected) of elements $\xi \in O$ are then conjugate in G. A nonzero rational function p on V is relatively G-invariant if there exists a character χ of G such that, for all $g \in G$ and $x \in V$,

$$p(gx) = \chi(g)p(x).$$

A prehomogeneous G-variety V is called special if every relative invariant (defined over any extension field of F) is constant. This is the case if and only if the restriction homomorphism from the group $X(G)$ of algebraic characters of G to $X(G^\xi)$ is an isomorphism.

Theorem 3. *Let P be a parabolic subgroup of the reductive group G with unipotent radical N and let $N' \subset N''$ be normal unipotent subgroups of P.*

(i) For any $\gamma \in P^{\mathrm{infl}}$, the affine space $\gamma N''/N'$ is prehomogeneous under the action of the trivial connected component $P_{\gamma N''}$ of the stabiliser of $\gamma N''$ in P by conjugation.

(ii) If C' is the P-conjugacy class of γ, then the generic orbit is the projection of $C' \cap \gamma N''$, viz. $(C' \cap \gamma N'')N'/N'$.

(iii) The prehomogeneous variety $\gamma N/N'$ is special if and only if $\gamma N/N''$ and $\gamma N''/N'$ are special.

This follows from Proposition 5 of [Ho-ind]. Note that the action on the affine spaces in question is not always given by affine transformations.

A prehomogeneous G-variety is called a prehomogeneous vector space if it is a vector space and the action of G is linear. A prehomogeneous vector space is called regular if the dual space V^* is prehomogeneous for the contragredient action and the map $dp/p : O \to V^*$ is a dominant morphism for some relative invariant p. The notion of F-regularity is defined in the obvious way.

Prehomogeneous vector spaces that are regular over a number field F have been intensively studied because they give rise to zeta integrals

$$Z(\varphi, s_1, \ldots, s_n) = \int_{G(\mathbb{A})/G(F)} |\chi_1(g)|^{s_1} \cdots |\chi_n(g)|^{s_n} \sum_{\xi \in O(F)} \varphi(g\xi)\, dg,$$

where φ is a Schwartz–Bruhat function on $V(\mathbb{A})$ and the characters χ_i correspond to relative invariants p_i which extend to regular functions on V and form a basis of the group of all relative invariants defined over F. Here we preclude that the connected generic stabilisers G_ξ have nontrivial F-rational characters, as the integral is otherwise divergent. (We will encounter prehomogeneous vector spaces, of incomplete type in the terminology of [Yu], where this happens and one has to truncate the integrand.) A typical result of the classical theory is the following.

Theorem 4. *Suppose in addition that G is F-anisotropic modulo centre. Let $V = \bigoplus_{i=1}^n V_i$ be the splitting obtained by diagonalisation of the largest F-split torus in the centre of G and choose p_i depending only on the ith component.*

(i) The zeta integral converges absolutely when $\operatorname{Re} s_i > r_i$ for all i, where $r_i = \dim V_i / \deg p_i$, and extends to a meromorphic function on $\mathbb{C}^n$. Its only singularities are at most simple poles along the hyperplanes $s_i = r_i$ and $s_i = 0$.

(ii) For each splitting of the index set $\{1, \ldots, n\}$ into a disjoint union $I' \cup I''$ and the corresponding splitting $V = V' \oplus V''$, we have

$$\lim_{s' \to r'} Z(\varphi, s', s'') \prod_{i \in I'} (s_i - r_i) = Z''(\varphi'', s''),$$

where Z'' is the zeta integral over

$$\{g \in G(\mathbb{A}) \,\big|\, |\chi_i(g)| = 1 \;\forall\, i \in I'\}/G(F)$$

of the function

$$\varphi''(x'') = \int_{V'} \varphi(x', x'')\, dx'.$$

(iii) For each splitting as above, we have the functional equation

$$Z(\varphi, s', s'') = Z(\mathcal{F}'\varphi, r' - s', s''),$$

where $\mathcal{F}'$ denotes the partial Fourier transform with respect to V'.

The convergence for large $\operatorname{Re} s_i$ has been proved in a rather general situation by Saito [Sai]. The present situation is much easier, since $G(\mathbb{A})^1/G(F)$ is compact and the centre acts by componentwise multiplication. The proof of the remaining assertions goes hand in hand and proceeds as in [Sa].

If we fix a finite set S of places of F containing the archimedean ones, and a lattice in $V(\mathbb{A}^S)$ with respect to the maximal compact subring of $\mathbb{A}^S$, then every Schwartz–Bruhat function φ on $V(F_S)$ can be canonically extended to $V(\mathbb{A})$. For such functions, one obtains a decomposition

$$Z(\varphi, s) = \sum_{[\xi]_S} \zeta(\xi, s) \int_{G(F_S)/G^{\xi}(F_S)} |\chi_1(g)|^{s_1} \cdots |\chi_n(g)|^{s_n} \varphi(g\xi)\, dg$$

over the finitely many $G(F_S)$-orbits $[\xi]_S$ in $O(F_S)$, where the zeta functions $\zeta(\xi, s)$ encode valuable arithmetic information (see [Ki] for the case $F = \mathbb{Q}$, $S = \{\infty\}$). We will not go into details here but rather describe a similar procedure for conjugacy classes in Sect. 5.1.

2.3 Canonical Parabolic Subgroups

From now on, we assume that G is reductive and F has characteristic zero. Then we have mutually inverse F-morphisms log and exp between the unipotent subvariety of the group G and the nilpotent subvariety of its Lie algebra $\mathfrak{g}$. By the Jacobson–Morozov theorem, for every nilpotent element $X \in \mathfrak{g}$, there is a homomorphism $\mathfrak{sl}_2 \to \mathfrak{g}$ such that X is the image of $\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$. Let H be the image of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ and set $\mathfrak{g}_n = \{Z \in \mathfrak{g} \mid [H, Z] = nZ\}$, so that $X \in \mathfrak{g}_2$. We consider the subalgebras

$$\mathfrak{q} = \bigoplus_{n \geq 0} \mathfrak{g}_n, \qquad \mathfrak{u} = \bigoplus_{n > 0} \mathfrak{g}_n, \qquad \mathfrak{u}' = \bigoplus_{n > 1} \mathfrak{g}_n, \qquad \mathfrak{u}'' = \bigoplus_{n > 2} \mathfrak{g}_n$$

and the subgroups

$$Q = \operatorname{Norm}_G \mathfrak{q}, \qquad U = \exp \mathfrak{u}, \qquad U' = \exp \mathfrak{u}', \qquad U'' = \exp \mathfrak{u}''.$$

It is well known that $\mathfrak{q}$ is a parabolic subalgebra with ideals $\mathfrak{u}$, $\mathfrak{u}'$ and $\mathfrak{u}''$, where $[X, \mathfrak{q}] = \mathfrak{u}'$ and $[X, \mathfrak{u}] = \mathfrak{u}''$, and that Q is a parabolic subgroup of G with unipotent radical U and normal subgroups U' and U''. By results of Kostant (see Theorem 3.4.10 of [CG] or Sect. 11.1 of [Bou]), those subalgebras and hence the corresponding subgroups are independent of the choice of the homomorphism $\mathfrak{sl}_2 \to \mathfrak{g}$ used in the definition. Moreover, $L = \mathrm{Cent}_G H$ is a Levi component of Q. One calls $\mathfrak{q}$ the canonical parabolic subalgebra of X. If $\exp X$ is the unipotent component in the Jordan decomposition of an element $\gamma \in G$, we call Q the canonical parabolic subgroup of γ. Moreover, we denote by Q^{can} the set of elements of G whose canonical parabolic is Q.

Theorem 5. *(i) If $\gamma \in G(F)$, then Q, U, U' and U'' are defined over F. We can choose $H \in \mathfrak{g}(F)$, and then L is defined over F.*
(ii) The vector space $\mathfrak{u}'/\mathfrak{u}''$ with the adjoint action of $L \cong Q/U$ is a regular prehomogeneous vector space. In the situation of (i) it is F-regular.
(iii) If Q is the canonical parabolic and C the conjugacy class of an element γ, then $C \cap Q^{\mathrm{can}}$ is the conjugacy class of γ in Q. If γ is unipotent, then this set is open and dense in U' and invariant under translations by elements of U''.

If γ is F-rational, so is its unipotent component $\exp X$, and hence $X \in \mathfrak{g}(F)$. The Jacobson–Morozov theorem (see Sect. 11.2 of [Bou]) provides a homomorphism $\mathfrak{sl}_2(F) \to \mathfrak{g}(F)$ and a subalgebra $\mathfrak{q}(F)$. Thus Q and L are defined over F. It is well known that so are the unipotent radical U and its upper central series. Assertion (ii) is proved in the reference, too, even though it can be considered folklore. Theorem 2 of that paper also contains a version of the last statement for mixed elements, but that seems to be less useful for our purposes.

Conjecture 0. *If γ is a unipotent element of P^{infl} for a parabolic subgroup P of G, then $U \subset P$.*

This will be needed in Lemma 7. I thank the referee for pointing out that the proof presented in a previous version of this paper was incorrect. Conjecture 1 below has already been cited and cannot be renumbered.

2.4 Mean Values

A mean value formula has been proved by Siegel for the action of $\mathrm{SL}_n(\mathbb{R})$ on $\mathbb{R}^n$ ($n > 1$), generalised by Weil [We] to the adelic setting and by Ono [Ono] to the following general case.

Theorem 6. *If O is a special G-homogeneous variety over a number field F with trivial groups $\pi_1(O(\mathbb{C}))$, $\pi_2(O(\mathbb{C}))$ and $X(G)$, then*

$$\int_{G(\mathbb{A})/G(F)} \sum_{\xi \in O(F)} h(g\xi)\, dg = \int_{G(\mathbb{A})O(F)} h(x)\, dx$$

for $h \in C_c^\infty(G(\mathbb{A})O(F))$ and a suitable normalisation of invariant measures.

Actually, Ono imposed the additional assumption that $[G(\mathbb{A})\xi \cap O(F) : G(F)]$ be independent of $\xi \in O(F)$, but this is automatically satisfied by Proposition 2.3 of [MoWa]. Moreover, he used the term "special" only under the assumption that the group $X(G)$ is trivial. With our wider definition, the theorem is still valid if we replace G by its derived subgroup G', because the map $G'/G'^{\xi} \to G/G^{\xi}$ is an isomorphism.

If O is the generic orbit in a special prehomogeneous affine space V, then the first two homotopy groups are automatically trivial. In fact, the complement of O is a subvariety W of codimension greater than one by Lemma 7 of [Ho-ind]. For any Lipschitz map $\phi : S^i \to O(\mathbb{C})$, the map $\psi : W(\mathbb{C}) \times S^i \times \mathbb{R} \to V(\mathbb{C})$ given by $\psi(w, s, t) = tw + (1-t)\phi(s)$ has range of Hausdorff dimension at most $\dim_{\mathbb{R}} W + i + 1$. For $i \leq 2$, this is less than $\dim_{\mathbb{R}} V(\mathbb{C})$, so we can choose $x \in V(\mathbb{C})$ not in those ranges and get a null-homotopy $\phi_t(s) = tx + (1-t)\phi(s)$ in $O(\mathbb{C})$.

We need a slightly different version of the above theorem.

Theorem 7. *If V is a torsor under a unipotent group N and the group G with trivial $X(G)$ acts on the pair (N, V) by automorphisms, so that V is a special prehomogeneous G-space with generic orbit O and the orbit map $G \to O$ has local sections, then*

$$\int_{G(\mathbb{A})/G(F)} \sum_{\xi \in O(F)} h(g\xi)\, dg = \int_{G(\mathbb{A})/G(F)} \int_{V(\mathbb{A})} h(gx)\, dx\, dg$$

for $h \in C_c^\infty(V(\mathbb{A}))$, provided we normalise the measure on $V(\mathbb{A})$ so that $V(\mathbb{A})/N(F)$ has measure 1.

Proof. Since the orbit map $g \mapsto gv$ for $v \in O(F)$ has local sections, which are defined over F according to our standing assumption, it maps $G(F)$ onto $O(F)$ and $G(\mathbb{A})$ onto $O(\mathbb{A})$. In particular, Ono's additional condition is trivially satisfied. Indeed, the complement W of O is an algebraic subset, hence a null set for the $N(\mathbb{A})$-invariant measure on $V(\mathbb{A})$. That measure is also $G(\mathbb{A})$-invariant, hence its restriction to $O(\mathbb{A})$ coincides with the measure in Ono's theorem, in which we may replace the domain of integration on the right-hand side by $V(\mathbb{A})$. We may also replace the integrand $h(x)$ by $h(g_1 x)$, where $g_1 \in G(\mathbb{A})$ is arbitrary, and then integrate the right-hand side over g_1, as the measure of $G(F)\backslash G(\mathbb{A})$ is finite due to $X(G) = \{1\}$. This proves the claim up to the normalisation of measures and the extension to $C_c^\infty(V(\mathbb{A}))$.

There is an alternative, though less elegant, proof, which provides these facts. One reduces the assertion to the case of abelian N using a central series of a general unipotent group N and Proposition 5 of [Ho-ind]. In the abelian case one proceeds as in [We]. □

In the situation of Theorem 3, $\gamma N/N'$ is an N/N'-torsor, on which $P_{\gamma N}$ acts by automorphisms. In order to apply Theorem 7, we need the following hypothesis about a parabolic subgroup P of a reductive group G with unipotent radical N and a conjugacy class C in P/N:

Hypothesis 1. *There is a normal unipotent subgroup N^C of P such that, for $\gamma \in C' = \operatorname{Infl}^P C$,*

(i) the prehomogeneous affine space $\gamma N/N^C$ is special under $P_{\gamma N}$ and the generic orbit map has local sections,
(ii) all elements of $\gamma N^C \cap C'$ have the same canonical parabolic.

We call this a hypothesis rather than a conjecture because if it is not generally true, we may at least treat those conjugacy classes to which it applies. In fact, it has been checked for all classical groups up to rank 3.

As the notation suggests, there should be a canonical choice for N^C. By Lemma 8 of [Ho-ind], there is a largest normal unipotent subgroup of P with property (ii), and under Hypothesis 1 it will then also have property (i) in view of Theorem 3. In general, however, it seems not to be the correct choice for our purposes. We certainly assume, as we may, that $(\gamma N\gamma^{-1})^{\gamma C\gamma^{-1}} = \gamma N^C\gamma^{-1}$ for all $\gamma \in G(F)$.

2.5 (G, Q)-Families

In Sect. 6, we will need an analogue of the notion of (G, M)-families (see Sect. 17 of [A-int]) in which the Levi subgroup M is replaced by a parabolic subgroup Q. First we recall the pertinent notation.

For every connected linear algebraic group P defined over F, we denote by $\mathfrak{a}_P$ the real vector space of all homomorphisms from the group $X(P)_F$ of F-rational characters of P to the group $\mathbb{R}$. If $P = MN$ is a Levi decomposition and A the largest split torus in the centre of M, then the natural homomorphisms $\mathfrak{a}_A \to \mathfrak{a}_M \to \mathfrak{a}_P$ are isomorphisms, and the set Δ_P of fundamental roots of A in $\mathfrak{n}$ can be regarded as a subset of the dual space $\mathfrak{a}_P^*$ independent of the choice of A. Moreover, if $Q \subset P$ are parabolic subgroups of a reductive group G, we obtain natural maps $\mathfrak{a}_P \rightleftarrows \mathfrak{a}_Q$, which induce a splitting $\mathfrak{a}_Q = \mathfrak{a}_Q^P \oplus \mathfrak{a}_P$. The coroots $\check{\alpha}$ are originally only defined for roots α of a maximal split torus, hence for the elements of Δ_Q, when Q is a minimal parabolic, but if $\beta = \alpha|_{\mathfrak{a}_P}$ is nonzero, we may define $\check{\beta}$ as the projection of $\check{\alpha}$ to $\mathfrak{a}_P$. These coroots form a basis of $\mathfrak{a}_P^G$, and we denote the dual basis of $(\mathfrak{a}_P/\mathfrak{a}_G)^*$ by $\hat{\Delta}_P$, whose elements ϖ are called fundamental weights. The basis dual to Δ_P, whose elements are called fundamental coroots, is in bijection with Δ_P and hence with $\hat{\Delta}_P$. Following [A-int], the fundamental coroot corresponding to $\varpi \in \hat{\Delta}_P$ will be denoted by $\check{\varpi}$.

The characteristic functions of the chamber $\mathfrak{a}_P^+ = \{H \in \mathfrak{a}_P \mid \alpha(H) > 0 \ \forall\, \alpha \in \Delta_P\}$ and the dual cone ${}^+\mathfrak{a}_P = \{H \in \mathfrak{a}_P \mid \varpi(H) > 0 \ \forall\, \varpi \in \hat{\Delta}_P\}$ are denoted by τ_P and $\hat{\tau}_P$, respectively. Their Fourier transforms

$$\hat{\theta}_P(-\lambda)^{-1} = \int_{(\mathfrak{a}_P^G)^+} e^{\langle\lambda, H\rangle}\, dH, \qquad \theta_P(-\lambda)^{-1} = \int_{{}^+(\mathfrak{a}_P^G)} e^{\langle\lambda, H\rangle}\, dH$$

(mind the swap of the accent) are defined for complex-valued linear functions λ on $\mathfrak{a}_P$ with positive real part on the support. An easy computation yields

$$\hat{\theta}_P(\lambda) = \hat{\eta}_P \prod_{\varpi \in \hat{\Delta}_P} \langle \lambda, \check{\varpi} \rangle, \qquad \theta_P(\lambda) = \eta_P \prod_{\alpha \in \Delta_P} \langle \lambda, \check{\alpha} \rangle,$$

where the constants $\hat{\eta}_P$ and η_P depend on the Haar measure on $\mathfrak{a}_P^G$.

The parabolic subgroups R of a Levi component M of P are in bijection with the parabolic subgroups Q of P via $R \mapsto RN$, $Q \mapsto M \cap Q$. Prompted by the equality $\mathfrak{a}_R^M = \mathfrak{a}_Q^P$, one indexes the objects associated with the pair (M, R) in place of (G, Q) by the pair (P, Q) of parabolics of G, like Δ_Q^P, etc. The superscript G may sometimes be omitted, e. g. in τ_P^G.

For parabolics $P \subset P'$ containing Q, we denote the restriction of a linear function λ on $\mathfrak{a}_Q$ to the subspace $\mathfrak{a}_P^{P'}$ by $\lambda_P^{P'}$, where the upper index G and the lower index Q may be omitted. The relative versions of the above Fourier transforms are extended to all λ by setting

$$\hat{\theta}_P^{P'}(\lambda) = \hat{\theta}_P^{P'}(\lambda_P), \qquad \theta_P^{P'}(\lambda) = \theta_P^{P'}(\lambda_P),$$

and similar remarks apply to the functions $\tau_P^{P'}$ and $\hat{\tau}_P^{P'}$. We assume that the measures on all the spaces $\mathfrak{a}_P^{P'}$ are normalised in a compatible way.

With notational matters out of the way, we now define a (G, Q)-family to be a family of holomorphic functions $c_P(\lambda)$ indexed by the parabolic subgroups P containing Q and defined for $\operatorname{Re}\lambda$ in a neighbourhood of zero in $(\mathfrak{a}_Q)_{\mathbb{C}}^*$ such that, for any two parabolics $P \subset P'$ containing Q,

$$\lambda_P^{P'} = 0 \quad \Rightarrow \quad c_P(\lambda) = c_{P'}(\lambda).$$

This condition does not get weaker if we require it only for P, P' with $\dim \mathfrak{a}_P^{P'} = 1$. We say that the (G, Q)-family is frugal (resp. cofrugal) if $c_P(\lambda) = c_Q(\lambda_P)$ (resp. $c_P(\lambda) = c_G(\lambda^P)$) for all P, where λ_P (resp. λ^P) is extended to $\mathfrak{a}_Q$ so that it vanishes on $\mathfrak{a}_Q^P$ (resp. on $\mathfrak{a}_P$). Every holomorphic function c_Q on $\mathfrak{a}_Q$ (resp. c_G on $\mathfrak{a}_Q^G$) determines a frugal (resp. cofrugal) (G, Q)-family.

Lemma 1. *For each (G, Q)-family of functions c_P, the meromorphic function*

$$c'_Q(\lambda) = \sum_{P \supset Q} \epsilon_Q^P c_P(\lambda) \hat{\theta}_Q^P(\lambda)^{-1} \theta_P(\lambda)^{-1},$$

where $\epsilon_Q^P = (-1)^{\dim \mathfrak{a}_Q^P}$, is holomorphic for $\operatorname{Re}\lambda$ *in a neighbourhood of zero.*

The special case for frugal families is Lemma 6.1 of [A-inv] (with the roles of P and Q interchanged). As in that source, we could also prove a version for smooth functions defined for purely imaginary λ only, although it does not seem to have applications. One may compute the value $c'_Q(0)$ by setting $\lambda = z\lambda_0$ for any fixed λ_0 not on any singular hyperplane and applying l'Hospital's rule to the resulting function of $z \in \mathbb{C}$, reduced to a common denominator.

Proof. For each fundamental root $\alpha \in \Delta_Q$, we denote the corresponding fundamental weight by ϖ_α. Let P be a parabolic subgroup containing Q. For each $\alpha \in \Delta_Q \setminus \Delta_Q^P$, the projection of $\check{\alpha}$ along $\mathfrak{a}_Q^P$ onto $\mathfrak{a}_P^G$ is a fundamental coroot $\check{\alpha}_P$, and for each $\alpha \in \Delta_Q^P$, the projection of $\check{\varpi}_\alpha$ along $\mathfrak{a}_P^G$ onto $\mathfrak{a}_Q^P$ is a fundamental coweight $\check{\varpi}_\alpha^P$. All the fundamental coroots of $\mathfrak{a}_P^G$ in P and fundamental coweights of $\mathfrak{a}_Q^P$ in Q/N arise in this way.

We fix a fundamental root $\beta \in \Delta_Q$ and denote by Q' the parabolic with $\Delta_Q^{Q'} = \{\beta\}$. Then there is a unique bijection $P \mapsto P'$ from $\{P \supset Q \mid P \not\supset Q'\}$ onto $\{P \supset Q'\}$ such that $\Delta_Q^{P'} = \Delta_Q^P \cup \{\beta\}$. The elements $\check{\beta}_P$ and $\check{\varpi}_\beta^{P'}$ as well as the differences $\check{\alpha}_P - \check{\alpha}_{P'}$ for every $\alpha \in \Delta_Q \setminus \Delta_Q^{P'}$ and $\check{\varpi}_\alpha^P - \check{\varpi}_\alpha^{P'}$ for every $\alpha \in \Delta_Q^P$ lie in the one-dimensional subspace $\mathfrak{a}_P^{P'}$. Together with the defining property of the (G, Q)-family this implies that the difference of

$$c_P(\lambda) \prod_{\alpha \in \Delta_Q^P} \lambda(\check{\varpi}_\alpha^{P'}) \prod_{\alpha \in \Delta_Q \setminus \Delta_Q^{P'}} \lambda(\check{\alpha}_{P'})$$

and

$$c_{P'}(\lambda) \prod_{\alpha \in \Delta_Q^P} \lambda(\check{\varpi}_\alpha^{P}) \prod_{\alpha \in \Delta_Q \setminus \Delta_Q^{P'}} \lambda(\check{\alpha}_{P})$$

vanishes on the hyperplane defined by $\lambda|_{\mathfrak{a}_P^{P'}} = 0$ and is therefore a multiple of the proportional linear forms $\lambda(\check{\beta}_P)$ and $\lambda(\check{\varpi}_\beta^{P'})$. Dividing by

$$\hat{\theta}_Q^P(\lambda)\theta_P(\lambda) \cdot \hat{\theta}_Q^{P'}(\lambda)\theta_{P'}(\lambda),$$

and remembering its compatible normalisation, we see that

$$c_P(\lambda)\hat{\theta}_Q^P(\lambda)\theta_P(\lambda) - c_{P'}(\lambda)\hat{\theta}_Q^{P'}(\lambda)\theta_{P'}(\lambda)$$

is singular at most for those λ which vanish on some one-dimensional subspace $\mathfrak{a}_R^{R'}$ with $\Delta_Q^{R'} \setminus \Delta_Q^R = \{\alpha\}$ for some $\alpha \neq \beta$. Multiplying by ϵ_Q^P and summing over P, we see that the same is true of $c'_Q(\lambda)$. Since β was arbitrary, we are done. □

Lemma 2. *Let $X \in \mathfrak{a}_Q^G$.*

(i) If $c_P = e^{\langle \lambda, X_P \rangle}$, then $c'_Q(\lambda)$ is the Fourier transform of the function

$$\Gamma'_Q(H, X) = \sum_{P \supset Q} \epsilon_P \tau_Q^P(H)\hat{\tau}_P(H - X).$$

(ii) If $c_P = e^{\langle\lambda, X^P\rangle}$, then $c'_Q(\lambda)$ is the Fourier transform of the function

$$\Gamma''_Q(H,X) = \sum_{P\supset Q} \epsilon_P \tau_Q^P(H-X)\hat{\tau}_P(H).$$

(iii) For H outside a finite union of hyperplanes, we have

$$\Gamma''_Q(H,X) = \epsilon_Q^G \Gamma'_Q(X-H,X).$$

Proof. Assertion (i) is Lemma 2.2 of [A-inv], and the proof of assertion (ii) is analogous. The substitution of $X-H$ for H has on the Fourier transform the effect of substituting $-\lambda$ for λ and multiplying by $e^{\langle\lambda,X\rangle}$. Since $\hat{\theta}_Q^P(\lambda)\theta_P(\lambda)$ is homogeneous of degree $\dim \mathfrak{a}_Q^G$ and $X = X_P + X^P$, the two sides of the asserted equality are characteristic functions of polyhedra with equal Fourier transforms. □

Given a (G,Q)-family and a parabolic $P \supset Q$, we obtain a (G,P)-family by restricting the functions $c_{P'}$ with $P' \supset P$ to the subspace $\mathfrak{a}_P$, and we obtain an $(M, M\cap Q)$-family, where M is a Levi component of P, by setting

$$c^P_{M\cap P'}(\lambda) = c_{P'}(\lambda).$$

Checking the condition for such families is straightforward.

Lemma 3. *(i) For frugal (G,Q)-families, the definition of c'_Q is equivalent to the identity*

$$c_Q(\lambda)\theta_Q(\lambda)^{-1} = \sum_{P\supset Q} c'_P(\lambda_P)\theta_Q^P(\lambda)^{-1}.$$

(ii) For cofrugal (G,Q)-families, the definition of c'_Q is equivalent to the identity

$$c_G(\lambda)\hat{\theta}_Q(\lambda)^{-1} = \sum_{P\supset Q} \epsilon_Q^P (c^P)'_{M\cap Q}(\lambda)\hat{\theta}_P(\lambda)^{-1}.$$

Note that both identities in (ii) can be read as recursive definitions by isolating the term with $P = Q$ or $P = G$, resp. See Eq. (17.9) in [A-int] and Eq. (6.2) in [A-inv] for the frugal case.

Proof. For each of the four required implications, one starts with the right-hand side of the equation to be proved and plugs in the hypothesis. Then one interchanges summations and uses the fact that the expressions

$$\sum_{P\supset Q} \epsilon_P \hat{\theta}_Q^P(\lambda)^{-1}\theta_P(\lambda)^{-1}, \qquad \sum_{P\supset Q} \epsilon_P \theta_Q^P(\lambda)^{-1}\hat{\theta}_P(\lambda)^{-1}$$

are 1 for $Q = G$ and 0 otherwise, which follows from Eqs. (8.10) and (8.11) of [A-int]. □

The elementwise product of two (G, Q)-families is again a (G, Q)-family.

Lemma 4. *Given two (G, Q)-families of functions c_P and d_P, of which the former family is cofrugal or the latter family is frugal, we have the splitting formula*

$$(cd)'_Q(\lambda) = \sum_{P \supset Q} (c^P)'_{M \cap Q}(\lambda) d'_P(\lambda_P).$$

Proof. The proof for frugal d is analogous to the case of (G, L)-families. Using the relative version of the first identity of Lemma 3(ii) with P' in place of Q, we get

$$\begin{aligned}(cd)'_Q(\lambda) &= \sum_{P' \supset Q} \epsilon_Q^{P'} c_{P'}(\lambda) \hat{\theta}_Q^{P'}(\lambda)^{-1} \sum_{P \supset P'} d'_P(\lambda_P) \theta_{P'}^P(\lambda)^{-1} \\ &= \sum_{P \supset Q} d'_P(\lambda_P) \sum_{P' : Q \subset P' \subset P} \epsilon_Q^{P'} c_{P'}(\lambda) \hat{\theta}_Q^{P'}(\lambda)^{-1} \theta_{P'}^P(\lambda)^{-1}.\end{aligned}$$

Similarly, using the second identity of Lemma 3(ii) with M' in place of G, we get

$$\begin{aligned}(cd)'_Q(\lambda) &= \sum_{P' \supset Q} \epsilon_Q^{P'} d_{P'}(\lambda) \theta_{P'}(\lambda)^{-1} \sum_{P : Q \subset P \subset P'} \epsilon_Q^P (c^P)'_{M \cap Q}(\lambda) \hat{\theta}_P^{P'}(\lambda)^{-1} \\ &= \sum_{P \supset Q} (c^P)'_{M \cap Q}(\lambda) \sum_{P' \supset P} \epsilon_P^{P'} d_{P'}(\lambda) \hat{\theta}_P^{P'}(\lambda)^{-1} \theta_{P'}(\lambda)^{-1}.\end{aligned}$$

Now it remains to apply the definitions of $(c^P)'_Q$ and d'_P, resp. □

3 The Geometric Side of the Trace Formula

3.1 Truncation

Before introducing a new sort of geometric expansion, let us supply the details omitted in the introduction. We fix a number field F and denote the product of its archimedean completions by F_∞. For any linear algebraic group G, tacitly assumed to be connected and defined over F, the group of continuous homomorphisms from $G(\mathbb{A})$ to the additive group $\mathbb{R}$ has the structure of a real vector space. We denote its dual space by $\mathfrak{a}_G$ and define a continuous homomorphism $H_G : G(\mathbb{A}) \to \mathfrak{a}_G$ by $\chi(g) = \langle \chi, H_G(g) \rangle$ for all $\chi \in \mathfrak{a}_G^*$. The kernel of H_G is then the group $G(\mathbb{A})^1$. From now on, the letter G will be reserved for a reductive group.

In order to save space, the set $V(F)$ of F-rational points of any affine F-variety V will henceforth simply be denoted by V and the set of its adelic points by the corresponding boldface letter $\mathbf{V}$. As an exception, the letter $\mathbf{K}$ will denote a maximal compact subgroup of our adelic group $\mathbf{G}$ such that $G(F_\infty)\mathbf{K}$ is open and $\mathbf{G} = \mathbf{PK}$ for every parabolic subgroup P. One extends the map $H_P : \mathbf{P} \to \mathfrak{a}_P$ to $\mathbf{G}$ by setting $H_P(pk) = H_P(p)$ for $p \in \mathbf{P}$ and $k \in \mathbf{K}$, not indicating the dependence on the choice of $\mathbf{K}$ in the notation.

A truncation parameter T for the pair $(\mathbf{G}, \mathbf{K})$ is a family of elements $T_P \in \mathfrak{a}_P$ indexed by the parabolic subgroups such that the modified maps $H_P^T(x) = H_P(x) - T_P$ satisfy $H^T_{\gamma P \gamma^{-1}}(\gamma x) = H_P^T(x)$ for all $\gamma \in G$ and such that $H_{P'}^T(x)$ is the projection of $H_P^T(x)$ for arbitrary parabolics $P \subset P'$. Thereby we have eliminated the need for standard parabolic subgroups. The set of truncation parameters has the structure of an affine space such that the evaluation at any minimal parabolic P_0 is an isomorphism onto $\mathfrak{a}_{P_0}$.

For any maximal parabolic P', let $-\tau_{P'}^T$ be the characteristic function of

$$\{x \in \mathbf{G} \mid H_{P'}^T(x) \in \mathfrak{a}_{P'}^+\},$$

where $\mathfrak{a}_{P'}^+$ denotes the positive chamber in $\mathfrak{a}_{P'}$. For general P, set

$$\tau_P^T(x) = \prod_{\substack{P' \supset P \\ \text{max.}}} \tau_{P'}^T(x).$$

Thereby the usual sign factors $\epsilon_P = (-1)^{\dim \mathfrak{a}_P^G}$ in the integrand of $J^T(f)$ have been incorporated into these cut-off functions. We mention that the integral converges for all values of T (see [Ho-est]) and depends polynomially on T (see [A-inv]).

3.2 Expansion in Terms of Geometric Conjugacy Classes

In the distribution $J^T(f)$, one cannot isolate the contribution of a group-theoretic conjugacy class in $G(F)$ because the representatives of a coset γN appearing in K_P belong to various conjugacy classes. In the coarse geometric expansion (see [A-trI]), conjugacy has therefore been replaced by a coarser equivalence relation for which all elements in such cosets are equivalent. The finest such relation turns out to be just conjugacy of semisimple components. The fine geometric expansion is based on an intermediate refinement that depends on a choice of a finite set of places of F, but it is still not fully explicit. We propose to use geometric conjugacy for a start, deferring the finer expansions to the later step of stabilisation. It is the induction of conjugacy classes that makes this work.

Let P be a parabolic subgroup of G with unipotent radical N. Recall that induction as defined in Theorem 1 is actually a map from the set of conjugacy classes in P/N to conjugacy classes in G that does not depend on the choice of a Levi component. We define the contribution of a geometric conjugacy class C in G to the kernel function K_P as

$$K_{P,C}(x,y) = \sum_{\substack{D \subset P/N \\ \mathrm{Ind}_P^G D = C}} \sum_{\gamma \in D} \int_{\mathbf{N}} f(x^{-1}\gamma n y)\, dn,$$

where we denote conjugacy classes in P/N by D, the letter C being reserved for conjugacy classes in G. We stick to tradition and avoid the awkward expression $\gamma N \in D$ under the summation sign. Alternatively, we can write

$$K_{P,C}(x,y) = \sum_{\gamma \in (C \cap P)N/N} \int_{\mathbf{N}} f(x^{-1}\gamma n y)\, dn.$$

The convergence of K_P implies that of its subsum $K_{P,C}$, and it is obvious that

$$K_P = \sum_C K_{P,C}.$$

The contribution of the class C to the trace distribution is defined formally as

$$J_C^T(f) = \int_{G \backslash \mathbf{G}^1} \sum_P K_{P,C}(x,x)\tau_P^T(x)\, dx,$$

but its convergence and the validity of the expansion

$$J^T(f) = \sum_C J_C^T(f)$$

depends on the following condition.

Conjecture 1. *For* $f \in C_c^\infty(G(\mathbb{A})^1)$*, we have*

$$\sum_C \int_{G \backslash \mathbf{G}^1} \left| \sum_P K_{P,C}(x,x)\tau_P^T(x) \right| dx < \infty.$$

This statement would also make sense for function fields F. It is a version of the convergence theorem of [A-trI]. Its analogue for Lie algebras over number fields has recently been proved by Chaudouard [Cha], and his methods should carry over to the group case. The merit of this result will depend on our ability to find a useful alternative description of the distributions $J_C^T(f)$. We will see that on the way to this goal even more subtle convergence results are needed. Conjecture 1 has meanwhile been proved in [FL].

4 Rearranging the Geometric Side

4.1 Replacing Integrals by Sums

In the term of $K_{P,C}$ corresponding to a conjugacy class D in P/N, the integral over $\mathbf{N}$ can be split into an integral over $\mathbf{N}/\mathbf{N}^D$ and an integral over $\mathbf{N}^D$ in the notation of Hypothesis 1. We want to replace the first of these integrals by a sum. This is analogous to Theorem 8.1 in [A-trI].

Note that D defines a P-conjugacy class $D' = \mathrm{Infl}^P D$ and, conversely, each P-conjugacy class D' in $C \cap P$ determines a conjugacy class $D = D'N/N$ in P/N. We define the modified kernel function

$$\tilde{K}_{P,C}(x,y) = \sum_{\substack{D \subset P/N \\ \mathrm{Ind}_P^G D = C}} \sum_{\gamma \in D'N^D/N^D} \int_{\mathbf{N}^D} f(x^{-1}\gamma n' x)\, dn'$$

and, formally, the modified distribution

$$\tilde{J}_C^T(f) = \int_{G\backslash \mathbf{G}^1} \tilde{K}_{P,C}(x,y) \tau_P^T(x)\, dx.$$

Hypothesis 2. *(i) The analogue of Conjecture 1 is true for $\tilde{J}_C^T(f)$.*
(ii) For all parabolic subgroups P and conjugacy classes $D \subset P/N$, we have

$$\int_{P\backslash \mathbf{G}^1} \sum_{\delta \in D} \left| \sum_{\gamma \in (C \cap \delta N)N^D/N^D} \int_{\mathbf{N}^D} f(x^{-1}\gamma n' x)\, dn' - \int_{\mathbf{N}} f(x^{-1}\delta n x)\, dn \right| \hat{\tau}_P^T(x)\, dx < \infty.$$

This hypothesis is automatic for groups of F-rank one, as $N^D = N$ in that case, and has been checked for classical groups of absolute rank 2 in [HoWa].

Lemma 5. *Under Conjecture 1 and Hypotheses 1 and 2, we have*

$$J_C^T(f) = \tilde{J}_C^T(f).$$

Proof. Granting Hypothesis 1, Theorem 7 yields the vanishing of

$$\int_{P'_{\delta N}\backslash \mathbf{P}'_{\delta N}} \left(\sum_{\gamma \in (C \cap \delta N)N^D/N^D} h(p_1^{-1}\gamma p_1) - \int_{\mathbf{N}/\mathbf{N}^D} h(p_1^{-1}\delta n'' p_1)\, dn'' \right) dp_1$$

for all $\delta \in D$ and $h \in C_c^\infty(\delta\mathbf{N}/\mathbf{N}^D)$, where $P'_{\delta N}$ denotes the derived group of $P_{\delta N}$. We plug in

$$h(v) = \tau_P^T(y) \int_{\mathbf{N}^D} f(y^{-1} p_2^{-1} v n p_2 y)\, dn$$

with $p_2 \in \mathbf{P}^1$ and $y \in \mathbf{G}^1$, and substitute $p_1 n = n' p_1$ in the integral over $\mathbf{N}^D$. Observing that the domain of integration over p_1 can be written as $P\backslash P\mathbf{P}'_{\delta N}$, we integrate over $p_2 \in P\mathbf{P}'_{\delta N}\backslash\mathbf{P}^1$. Then we sum over $\delta \in D$ and take the combined integral over $p = p_1 p_2 \in P\backslash\mathbf{P}^1$ outside the sum. Combining the summation over γ with that over δ into one summation and combining the integral over n'' with that over n' into one integral, we obtain

$$\tau_P^T(y) \int_{P\backslash\mathbf{P}^1} \left(\sum_{\gamma \in (C\cap DN)N^D/N^D} \int_{\mathbf{N}^D} f(y^{-1} p^{-1} \gamma n' p y)\, dn' \right.$$
$$\left. - \sum_{\delta \in D} \int_{\mathbf{N}} f(y^{-1} p^{-1} \delta n p y)\, dn \right) dp = 0.$$

Then we integrate over $y \in \mathbf{P}^1\backslash\mathbf{G}^1$ and combine this integral with that over p, observing that $\tau_P^T(y) = \tau_P^T(py)$, to get

$$\int_{P\backslash\mathbf{G}^1} \left(\sum_{\gamma \in D' N^D/N^D} \int_{\mathbf{N}^D} f(x^{-1} \gamma n' x)\, dn' \right.$$
$$\left. - \sum_{\delta \in D} \int_{\mathbf{N}} f(x^{-1} \delta n x)\, dn \right) \tau_P^T(x)\, dx = 0.$$

All of these operations are justified under Hypothesis 2(i).

We sum this expression over the finitely many standard parabolics P and respective classes D. Then we split the integral into an integral over $G\backslash\mathbf{G}^1$ and a sum over $P\backslash G$ and interchange the latter integral with the former sum. The latter sum can be replaced by a sum over all parabolic subgroups conjugate to P, because the relevant objects attached to different parabolics (the unipotent radical N, the set of conjugacy classes D in P/N with $\operatorname{Ind}^P D = C$ and the subgroups N^D) correspond to each other under conjugation. Finally, we split the integral into the difference of $J_C^T(f)$ and $\tilde{J}_C^T(f)$, which is justified by Conjecture 1 and Hypothesis 2(ii). □

4.2 Ordering Terms According to Canonical Parabolics

We continue rewriting our formula for $J_C^T(f)$. The basic idea for the next step is that a sum over all elements of G can be written as a sum over all parabolic subgroups Q of partial sums over those elements whose canonical parabolic is Q. This applies to $K_G(x,x)$, but $J_C^T(f)$ contains in addition terms with $P \neq G$, which are indexed by cosets γN rather than elements. This is why the previous transformation was necessary.

Lemma 6. *Let Q be the canonical parabolic of some element of C. Under Hypotheses 1, 2(i) and 3 (the latter to be stated in the course of the proof), we have*

$$\tilde{J}_C^T(f) = \int_{Q\backslash \mathbf{G}^1} \sum_{N' \subset Q} \sum_{P} \sum_{\substack{\gamma \in (C\cap Q^{\mathrm{can}})N'/N' \\ N^{[\gamma]}=N' \\ \gamma \in P^{\mathrm{infl}}}} \int_{\mathbf{N}'} f(x^{-1}\gamma n' x)\, dn'\, \tau_P^T(x)\, dx.$$

Here the representative γ is chosen in $C \cap Q^{\mathrm{can}}$, and $N^{[\gamma]}$ is a notation for N^D, where D is the conjugacy class of γN in P/N. If C is unipotent, the condition $N' \subset Q$ can be sharpened to $N' \subset U'$.

Recall that the sets Q^{can} and U' were introduced in connection with Theorem 5(iii).

The summation over subgroups N' may look weird. Of course, we need only consider subgroups which appear as N^D in Hypothesis 1. For unipotent conjugacy classes D it is often the case that N^D is the unipotent radical of a parabolic subgroup P^D, namely the smallest parabolic which contains P and whose unipotent radical is contained in U'. In this case, the sum over N' can be written as a sum over parabolics P' containing Q.

Proof. Let us fix a conjugacy class C. The definition of $\tilde{J}_C^T(f)$ involves, for each P, a sum over conjugacy classes D in P/N. We get the same result if we take the partial sum over those D for which N^D equals a given group N' and add up those partial sums for all possible subgroups N' of G.

For each P, N' and D, we are now facing a sum over cosets $\gamma N' \in (C \cap P)N'/N'$. By the property (ii) of N^D according to Hypothesis 1, the elements of $\gamma N' \cap C$ have the same canonical parabolic. Thus, we may similarly take the partial sum over those cosets for which that canonical parabolic equals a given group Q and add up the partial sums for all possible parabolics Q in G. As a result, we see that $\tilde{J}_C^T(f)$ equals

$$\int_{G\backslash \mathbf{G}^1} \sum_{P} \sum_{N'} \sum_{\substack{D\subset P/N \\ N^D=N'}} \sum_{Q} \sum_{\substack{\gamma \in D'N'/N' \\ \gamma N' \cap C \subset Q^{\mathrm{can}}}} \int_{\mathbf{N}'} f(x^{-1}\gamma n' x)\, dn'\, \tau_P^T(x)\, dx.$$

We want to move the summations over Q and N' leftmost. This is permitted under the following

Hypothesis 3. *The integral*

$$\int_{G\backslash \mathbf{G}^1} \sum_{Q} \sum_{N'} \left| \sum_{P} \sum_{\substack{D\subset P/N \\ N^D=N'}} \sum_{\substack{\gamma\in D'N'/N' \\ \gamma N'\cap C\subset Q^{\mathrm{can}}}} \int_{\mathbf{N}'} f(x^{-1}\gamma n'x)\,dn'\,\tau_P^T(x) \right| dx$$

is convergent.

When P is fixed, D runs over a finite set, hence the order of the two inner summations is irrelevant. They can be written as a single sum over all pairs $(D, \gamma N')$ satisfying the conditions

(i) $N^D = N'$,
(ii) $\gamma \in D'N'/N'$, where $D' = C \cap DN$,
(iii) $\gamma N' \cap C \subset Q^{\mathrm{can}}$.

Since the set $\gamma N' \cap C$ is dense in $\gamma N'$, quotients of its elements form a dense subset of N'. Therefore, condition (iii) can only be satisfied if $N' \subset Q$. If C is unipotent, then $Q \cap C \subset U'$ by Theorem 5(iii), and we must even have $N' \subset U'$.

Condition (iii) also implies that

(iii′) $\gamma \in (C \cap Q^{\mathrm{can}})N'/N'$.

Condition (ii) shows that the representative γ of $\gamma N'$ can be chosen in D', hence

(ii′) $\gamma \in P^{\mathrm{infl}}$,

and that D is uniquely determined by P and $\gamma N'$. If we denote N^D by $N^{[\gamma]}$, condition (i) can be rewritten as

(i′) $N^{[\gamma]} = N'$.

Conversely, suppose that we are given a coset $\gamma N'$ satisfying conditions (i′), (iii′) and (ii′), the latter for a choice of γ in $C \cap Q^{\mathrm{can}}$. Let D be the conjugacy class of the image of γ in P/N. Then condition (i) is satisfied, and hence $N' \subset P$. Therefore D is independent of the choice of the representative γ. Condition (ii′) shows in view of Proposition 2(iii) that γ lies in $D' = \mathrm{Infl}^P D$, and in view of $\gamma \in C$ we have $D' = C \cap DN$. Thus condition (ii) is satisfied. In hindsight we see that any other representative with the same properties lies in the P-orbit D', hence it also satisfies condition (ii′). Since $N' \subset Q$, we have $\gamma N' \subset Q$, and condition (iii) follows by Theorem 5(iii).

The equivalence of the two sets of conditions shows that $\tilde{J}_C^T(f)$ equals

$$\int_{G\backslash \mathbf{G}^1} \sum_{Q} \sum_{N'\subset Q} \sum_{P} \sum_{\substack{\gamma\in(C\cap Q^{\mathrm{can}})N'/N' \\ N^{[\gamma]}=N' \\ \gamma\in P^{\mathrm{infl}}}} \int_{\mathbf{N}'} f(x^{-1}\gamma n'x)\,dn'\,\tau_P^T(x)\,dx.$$

If Q is the canonical parabolic of some element of C, we can obtain those of the other ones by conjugating with elements of $Q\backslash G$. Thus, rather than summing over all parabolics Q, we may fix one of them, replace x by δx, and insert a summation over $\delta \in Q\backslash G$. Combining that summation with the exterior integral, we obtain our result. $\square$

If we are allowed to interchange the summations over P and $\gamma N'$, then $\tilde{J}_C^T(f)$ becomes

$$\int_{Q\backslash \mathbf{G}^1} \sum_{N'\subset Q} \sum_{\gamma\in(C\cap Q^{\mathrm{can}})N'/N'} \sum_{\substack{P\in\mathcal{P}_\gamma^{\mathrm{infl}} \\ N^{[\gamma]}=N'}} \int_{\mathbf{N}'} f(x^{-1}\gamma n' x)\, dn'\, \tau_P^T(x)\, dx.$$

The set $\mathcal{P}_\gamma^{\mathrm{infl}}$ was introduced in Theorem 2(iv). According to our notational conventions in this section, $\mathcal{P}_\gamma^{\mathrm{infl}}$ actually stands for $\mathcal{P}_\gamma^{\mathrm{infl}}(F)$. The integral over $\mathbf{N}'$ is independent of P and can be extracted from the sum over P.

Corollary 1. *If only finitely many parabolic subgroups P occur in the formula of Lemma 6, then*

$$\tilde{J}_C^T(f) = \int_{Q\backslash \mathbf{G}^1} \sum_{N'\subset Q} \sum_{\gamma\in(C\cap Q^{\mathrm{can}})N'/N'} \int_{\mathbf{N}'} f(x^{-1}\gamma n' x)\, dn'\, \chi_{\gamma N'}^T(x)\, dx,$$

where

$$\chi_{\gamma N'}^T(x) = \sum_{\substack{P\in\mathcal{P}_\gamma^{\mathrm{infl}} \\ N^{[\gamma]}=N'}} \tau_P^T(x).$$

There is apparently no uniform argument justifying such an interchange of summations in general. Below, we will describe approaches to this problem and solutions in partial cases.

4.3 Truncation Classes

We have sticked to geometric conjugacy classes so far because they afford a clean notion of induction. However, we are forced to split them up as evidence shows that various elements of the same class may behave differently in our formulas. For a moment, let us distinguish notationally between varieties and their sets of F-rational points again. While the sets $\mathcal{P}_\gamma^{\mathrm{infl}}$ for various elements γ in the same geometric conjugacy class C are in bijection with each other under conjugation, the sets $\mathcal{P}_\gamma^{\mathrm{infl}}(F)$ for $\gamma \in C(F)$ may be different if the elements are not $G(F)$-conjugate. This may happen even for unipotent classes.

For the lack of a better idea, we call elements γ_1, γ_2 of $G(F)$ truncation equivalent if they belong to the same geometric conjugacy class and if every inner automorphism mapping γ_1 to γ_2 will map $\mathcal{P}^{\mathrm{infl}}_{\gamma_1}(F)$ onto $\mathcal{P}^{\mathrm{infl}}_{\gamma_2}(F)$. This is an equivalence relation, because conjugate F-rational parabolics are $G(F)$-conjugate. The equivalence classes for this relation will be called truncation classes.

It follows from the definition that an element of $P^{\mathrm{infl}}(F)$ will also belong to $P'^{\,\mathrm{infl}}(F)$ for any parabolic P' containing P. Thus, if $P \in \mathcal{P}^{\mathrm{infl}}_{\gamma}(F)$, then $P' \in \mathcal{P}^{\mathrm{infl}}_{\gamma}(F)$. The inclusion relation among the sets $\mathcal{P}^{\mathrm{infl}}_{\gamma}(F)$ therefore defines a partial order on the finite set of truncation classes O in a given geometric conjugacy class $C(F)$.

In order to split up $J^T_C(f)$ into contributions of truncation classes, we have to do so for the kernel functions $K_{P,C}$. Thus, to every F-rational coset γN meeting C, we have to assign a truncation class. Evidence suggests that we should pick the minimal truncation class O meeting $\gamma N(F)$. Its uniqueness will have to be proved.

We expect that this definition produces the correct grouping of terms so that all of the previous discussion applies to the resulting distributions $J^T_O(f)$. In Lemma 6, we have to replace $P^{\mathrm{infl}}(F)$ by the subset $P^{\min \mathrm{infl}}(F)$ of elements whose truncation class is minimal among those meeting P. In the corollary, we have to replace $\mathcal{P}^{\mathrm{infl}}_{\gamma}(F)$ by the subset $\mathcal{P}^{\min \mathrm{infl}}_{\gamma}(F)$ of those parabolics P for which $\gamma \in P^{\min \mathrm{infl}}(F)$.

5 Relation to Zeta Integrals

5.1 Damping Factors

A classical artifice to make the full integral-sum in Lemma 6 (or its analogue for a truncation class O) absolutely convergent, primarily in the case of unipotent orbits, is to insert a damping factor $e^{-\langle\lambda, H_Q(x)\rangle}$ into the integrand of $\tilde{J}^T_O(f)$, where λ is a complex-valued linear function on $\mathfrak{a}_Q$, to obtain a distribution $J^T_O(f,\lambda)$. This idea goes back to Selberg and has been applied in [A-rk1, Ho-rk1, HoWa] and other papers. It works in the following situation.

Hypothesis 4. *Let O be a truncation class. Then the integral-sum*

$$J^T_O(f,\lambda) = \int_{Q\backslash\mathbf{G}^1} e^{-\langle\lambda, H_Q(x)\rangle} \sum_{N'\subset Q} \sum_{\gamma\in(O\cap Q^{\mathrm{can}})N'/N'} \int_{\mathbf{N}'} f(x^{-1}\gamma n' x)\, dn'\, \chi^T_{\gamma N'}(x)\, dx$$

is absolutely convergent for $\operatorname{Re}\lambda$ *in a neighbourhood of zero. If a group N' occurring here is the unipotent radical of parabolic subgroup P', then its contribution $J^T_{O,P'}(f,\lambda)$ is absolutely convergent for* $\operatorname{Re}\lambda$ *in a certain positive chamber and has a meromorphic continuation to a domain including the point $\lambda = 0$.*

Assume that all subgroups N' occurring in $J_O^T(f,\lambda)$ are unipotent radicals of parabolics P', so that the sum over N' is finite. If we choose λ_0 such that $\mathbb{C}\lambda_0$ is not contained in the singular set of any of these Dirichlet series, then the value of the regular function $J_O^T(f,\lambda)$ at $\lambda = 0$ is

$$J_O^T(f) = \sum_{P' \supset Q} \underset{z=0}{\text{f.p.}}\, J_{O,P'}^T(f, z\lambda_0),$$

where f.p. denotes the finite part in the Laurent expansion.

Although the above distribution depends on the choice of Q, this information has not been indicated in the notation since it is encoded in λ being a linear function on $\mathfrak{a}_Q$. If we want to avoid a preference, we have to consider a family λ of linear functions $\lambda(Q)$ on all the spaces $\mathfrak{a}_Q$ which is coherent in the sense that $\lambda(\delta^{-1}Q\delta) = \lambda(Q) \circ \operatorname{Ad}\delta$ for all $\delta \in G(F)$. Denoting the canonical parabolic subgroup of an element $\gamma \in O$ by $Q(\gamma)$, we get in the special case $P' = G$

$$J_{O,G}^{T,T'}(f,\lambda) = \int_{G\backslash \mathbf{G}^1} \sum_{\gamma \in O} e^{-\langle \lambda(Q(\gamma)), H_{Q(\gamma)}^{T'}(x)\rangle} f(x^{-1}\gamma x)\, \chi_\gamma^T(x)\, dx.$$

Here T' is a partial truncation parameter in the sense that it has only components indexed by parabolics in one conjugacy class. We recover the previous distribution by setting $T'_Q = 0$ for the chosen parabolic Q.

In order to explain how the Hypothesis gives rise to weighted orbital integrals, we need some preparation. One fixes a finite set S of places of F including all archimedean ones and decomposes the ring of adeles as a direct product $\mathbb{A} = F_S\mathbb{A}^S$ and the group as $G(\mathbb{A}) = G(F_S)G(\mathbb{A}^S)$. Then every function $f \in C_c^\infty(G(F_S))$ can be extended to $G(\mathbb{A})$ by multiplying it with the characteristic function of $\mathbf{K}^S = \mathbf{K} \cap G(\mathbb{A}^S)$. The group $G(F_S)^1$ acts from the right on $G(F)\backslash G(\mathbb{A})^1/\mathbf{K}^S$ with finitely many orbits, and the stabiliser of the orbit with representative $g \in G(\mathbb{A}^S)^1$ is the S-arithmetic subgroup

$$\Gamma_g = \{\delta_S \mid \delta \in G(F),\ \delta^S \in g\mathbf{K}^S g^{-1}\},$$

where δ_S and δ^S are the images of δ in $G(F_S)$ and $G(\mathbb{A}^S)$, resp. Its subset

$$O_g = \{\gamma_S \mid \gamma \in O,\ \gamma^S \in g\mathbf{K}^S g^{-1}\}$$

is invariant under conjugacy. Let us restrict to the case $P' = G$ for simplicity. If we substitute $x = gy$ with $y \in G(F_S)$, then the integrand vanishes unless $\gamma^S \in g\mathbf{K}^S g^{-1}$, and the distribution $J_{O,G}^{T,T'}(f,\lambda)$ becomes

$$\sum_g \int_{\Gamma_g\backslash G(F_S)^1} \sum_{\gamma \in O_g} e^{-\langle \lambda(Q(\gamma)), H_{Q(\gamma)}^{T'(g)}(y)\rangle} f(y^{-1}\gamma y)\, \chi_\gamma^{T(g)}(y)\, dy,$$

where $T(g)_P = T_P - H_P(g)$, because $H_P(gy) = H_P(g) + H_P(y)$. Ordering the elements according to canonical parabolics, we obtain the S-arithmetic version of the original formula

$$J_{O,G}^{T,T'}(f,\lambda) = \sum_g \sum_{[Q]_{\Gamma_g}} \int_{\Gamma_g \cap Q(F)\backslash G(F_S)^1} e^{-\langle \lambda(Q), H_Q^{T'(g)}(y)\rangle} \sum_{\gamma \in O_g \cap Q^{\mathrm{can}}} f(y^{-1}\gamma y)\, \chi_\gamma^{T(g)}(y)\, dy.$$

Here we split the inner sum into subsums over $\Gamma_g \cap Q(F)$-conjugacy classes in $O_g \cap Q^{\mathrm{can}}(F)$, written as sums over $\Gamma_g \cap Q^\gamma(F)\backslash \Gamma_g \cap Q(F)$, which can be combined with the integral once the summation over the classes has been moved outside:

$$J_{O,G}^{T,T'}(f,\lambda) = \sum_g \sum_{[Q]_{\Gamma_g}} \sum_{[O_g \cap Q^{\mathrm{can}}]_{\Gamma_g \cap Q}} \mathrm{vol}(\Gamma_g \cap Q^\gamma(F)\backslash Q^\gamma(F_S)^1) \int_{Q^\gamma(F_S)^1\backslash G(F_S)^1} e^{-\langle \lambda(Q), H_Q^{T'(g)}(y)\rangle} f(y^{-1}\gamma y)\, \chi_\gamma^{T(g)}(y)\, dy.$$

If $\gamma = q^{-1}\gamma_0 q$ with $q \in Q(F_S)$, then the substitution $qy = z$ transforms the integral into its analogue for γ_0 times $\exp\langle \lambda(Q), H_Q(q)\rangle$. The set $C(F_S)$ consists of finitely many $G(F_S)^1$-conjugacy classes, and each of them intersects $C(F_S) \cap Q^{\mathrm{can}}(F_S)$ in a $Q(F_S)$-conjugacy class. Choosing representatives $\gamma_0 \in Q^{\mathrm{can}}(F_S)$, we can write

$$J_{O,G}^{T,T'}(f,\lambda) = \sum_g \sum_{[Q]_{\Gamma_g}} \sum_{[\gamma_0]_S} \zeta_G(g, Q, \gamma_0, \lambda) \int_{Q^{\gamma_0}(F_S)^1\backslash G(F_S)^1} e^{-\langle \lambda(Q), H_Q^{T'(g)}(y)\rangle} f(y^{-1}\gamma_0 y)\, \chi_{\gamma_0}^{T(g)}(y)\, dy,$$

where $\zeta_G(g, Q, \gamma_0, \lambda)$ is a certain Dirichlet series in the variable λ. Since the parabolics Q are conjugate, suitable substitutions reduce the integrals to multiples of

$$J_G^T(\gamma, f, \lambda) = \int_{Q^\gamma(F_S)^1\backslash G(F_S)^1} e^{-\langle \lambda, H_Q(y)\rangle} f(y^{-1}\gamma y)\, \chi_\gamma^T(y)\, dy$$

for a fixed Q, where we have set $T'_Q = 0$, and we get

$$J_{O,G}^T(f,\lambda) = \sum_{[\gamma]_S} \zeta_G(S, \gamma, \lambda) J_G^T(\gamma, f, \lambda).$$

We do not indicate the dependence of the weighted orbital integral on S as this information is encoded in the argument f. Similarly one can show that, if N' is the unipotent radical of $P' \supset Q$,

$$J^T_{O,P'}(f,\lambda) = \sum_{[\gamma N']_S} \zeta_{P'}(S,\gamma,\lambda) J^T_{P'}(\gamma N', f, \lambda),$$

where $[\gamma N']_S$ are the $Q(F_S)$-conjugacy classes in $(C(F_S) \cap Q^{\mathrm{can}}(F_S))N'(F_S)/N'(F_S)$,

$$J^T_{P'}(\gamma N', f, \lambda) = \int_{Q^{\gamma N'}(F_S)^1 \backslash G(F_S)^1} e^{-\langle \lambda, H_Q(x) \rangle} \int_{N'(F_S)} f(x^{-1}\gamma n' x)\, dn'\, \chi^T_{\gamma N'}(x)\, dx$$

with the notation $Q^{\gamma N'}$ for the stabiliser of $\gamma N'$ in Q, and where $\zeta_{P'}(S,\gamma,\lambda)$ are certain Dirichlet series. They depend only on the coset $\gamma N'(F_S)$, but we prefer to write them as functions of γ for reasons that become clear at the end of Sect. 6.

If these zeta functions can be meromorphically continued, one obtains a formula with explicit weight factors, because the Laurent expansion of a product can be expressed by those of its factors. In Sect. 6 we will carry this out for the principal unipotent conjugacy class.

One has to regroup the result in terms of conjugacy classes of Levi subgroups M and to relate the resulting explicit weighted orbital integrals to Arthur's distributions $J_M(\gamma, f)$ if one wishes to compute the coefficients $a^M(S,\gamma)$ in the fine geometric expansion [A-mix]. So far, this has only been done in special cases (like in [HoWa]) by an ad-hoc computation.

5.2 Reduction to Vector Spaces

Let O be a truncation class in a unipotent conjugacy class C and Q the canonical parabolic of one of its elements. We perform a further transformation of $J^T_O(f,\lambda)$ which is in a way contrary to that in Lemma 5, because this time we are replacing sums by integrals.

The group Q/U, which can be identified with any Levi subgroup L of Q defined over F, acts on the group $V = U'/U''$ by conjugation. This action is linear if we endow U' and U'' with the structure of vector spaces defined over F using the exponential maps. By Theorem 5(ii), V is an F-regular prehomogeneous vector space, and the generic orbit is $C \cap U'/U''$. For each parabolic subgroup $P' \supset Q$, we have the vector subspace $V_{P'} = N'U''/U''$ and quotient space $V^{P'} = V/V_{P'}$, both prehomogeneous by Theorem 3(i).

Now we switch back to the simplified notation for adelic and rational points of varieties introduced in Sect. 3.1. For each L-invariant subquotient W of U, we denote by δ_W the modular character for the action of $\mathbf{L}$ on $\mathbf{W}$ by inner automorphisms. It can be interpreted as an element of $\mathfrak{a}^*_L$, so that $\delta_W(l) = e^{\langle \delta_W, H_L(l) \rangle}$. We have to be cautious since the tradition of the adelic trace formula imposes upon us the right action by inverses of inner automorphisms.

Hypothesis 5. *In the situation of Hypothesis 4, for a unipotent truncation class O and every Schwartz–Bruhat function φ on $\mathbf{V}$, the integral-sum*

$$Z_O^T(\varphi,\lambda)=\int_{L\backslash \mathbf{L}\cap\mathbf{G}^1} e^{-\langle\lambda+\delta_{U/U''},H_L(l)\rangle}\sum_{N'\subset Q}\sum_{\nu\in(O\cap U')N'/U''N'}\int_{\mathbf{N}'\mathbf{U}''/\mathbf{U}''}\varphi(l^{-1}\nu n' l)\,dn'\,\chi_{\nu N'}^T(l)\,dl,$$

is absolutely convergent for $\operatorname{Re}\lambda$ *in a neighbourhood of the closure of* $(\mathfrak{a}_Q^*)^+$.

If a group N' occurring here is the unipotent radical of a parabolic subgroup P', then for every Schwartz–Bruhat function ψ on $\mathbf{V}^{P'}$, the truncated zeta integral

$$Z_{O,P'}^T(\psi,\lambda)=\int_{L\backslash \mathbf{L}\cap\mathbf{G}^1} e^{-\langle\lambda+\delta_{U/N'U''},H_L(l)\rangle}\sum_{\nu\in(O\cap U')N'/U''N'}\psi(l^{-1}\nu l)\,\chi_{\nu N'}^T(l)\,dl,$$

is absolutely convergent for $\operatorname{Re}\lambda\in(\mathfrak{a}_Q^*)^+$ *and extends meromorphically to a neighbourhood of the closure of that domain.*

Let us look at the special case $P'=G$. As in Theorem 4, such integrals usually converge when the parameter in the exponent is $\lambda+\delta_V$ with $\operatorname{Re}\lambda$ positive on the chamber $\mathfrak{a}_Q^+$. Our parameter shift differs by $\delta_{U/U'}$, and if this point is contained in the domain of convergence, no terms with $P'\neq G$ are needed for regularisation.

Due to the restriction to $\mathbf{G}^1$, the usual convergence condition takes the form $X(L_\nu)_F=X(G)_F$ or, equivalently, $A_{L_\nu}=A_G$, where A_G denotes the largest F-split torus in the centre of G. It may be violated, but the truncation function χ_ν^T should save the convergence.

Lemma 7. *Under Hypotheses 4 and 5 and Conjecture 0, we have*

$$J_O^T(f,\lambda)=Z_O^T(f_V,\lambda),\qquad J_{O,P'}^T(f,\lambda)=Z_{O,P'}^T(f_V^{P'},\lambda),$$

where

$$f_V(v)=\int_{\mathbf{K}}\int_{\mathbf{U}''}f(k^{-1}vu''k)\,du''\,dk$$

is a smooth compactly supported function on $\mathbf{V}$ *and, for each Schwartz–Bruhat function* φ *on* $\mathbf{V}$,

$$\varphi^{P'}(v) = \int_{\mathbf{V}_{P'}} \varphi(vv')\,dv'$$

is a Schwartz–Bruhat functions on $\mathbf{V}^{P'}$.

Proof. For each $\gamma \in O \cap U'$, the map $U_\gamma \backslash U \to \gamma U''$ given by $\delta \mapsto \delta^{-1}\gamma\delta$ is an isomorphism due to the representation theory of $\mathfrak{sl}_2$. An isomorphism between affine F-varieties induces a bijection between their sets of F-rational points, and the set of F-rational points of U'/U'' is $U'(F)/U''(F)$. Since $O \cap U(F)$ is normalised by $U(F)$, all elements of a $U''(F)$-coset in $U'(F)$ belong to the same truncation class. Writing again U for $U(F)$ etc., we get for finitely supported functions g on U' and h on $\nu U''$

$$\sum_{\gamma\in O\cap U'} g(\gamma) = \sum_{\nu\in O\cap U'/U''}\sum_{\eta\in U''} g(\nu\eta), \quad \sum_{\eta\in U''} h(\nu\eta) = \sum_{\delta\in U_\nu\backslash U} h(\delta^{-1}\nu\delta).$$

This argument also works if we replace U' by U'/N' and U'' by its image $U''N'/N'$ in that quotient. It shows that, for g on V,

$$\sum_{\gamma\in(O\cap U')N'/N'} g(\gamma) = \sum_{\nu\in(O\cap U')N'/U''N'}\ \sum_{\delta\in U_{\nu N'}\backslash U} g(\delta^{-1}\nu\delta).$$

As a by-product, we see that the sum in the Lemma is well defined. These identities have adelic versions, too, of which we only need

$$\int_{\mathbf{U}''\mathbf{N}'/\mathbf{N}'} h(\nu u')\,du' = \int_{\mathbf{U}_{\nu N'}\backslash\mathbf{U}} h(u^{-1}\nu u)\,du$$

for continuous compactly supported functions h on $\nu\mathbf{U}''\mathbf{N}'/\mathbf{N}'$.

By definition, $J_O^T(f,\lambda)$ is given by the expression

$$\int_{Q\backslash\mathbf{G}^1} e^{-\langle\lambda, H_Q(x)\rangle} \sum_{N'\subset Q}\ \sum_{\gamma\in(O\cap U')N'/N'} \int_{\mathbf{N}'} f(x^{-1}\gamma n'x)\,dn'\,\chi^T_{\gamma N'}(x)\,dx.$$

Upon applying the identity we have just proved, the inner sum becomes

$$\sum_{\nu\in(O\cap U')N'/U''N'}\ \sum_{\delta\in U_{\nu N'}\backslash U} \int_{\mathbf{N}'} f(x^{-1}\delta^{-1}\nu\delta n'x)\,dn'\,\chi^T_{\delta^{-1}\nu\delta N'}(x).$$

Note that

$$\chi^T_{\delta^{-1}\nu\delta N'}(x) = \chi^T_{\nu N'}(\delta x).$$

Substituting $\delta n' = n\delta$, decomposing the exterior integral according to $\mathbf{G} = \mathbf{ULK}$ and combining the integral over $U\backslash\mathbf{U}$ with the sum over $U_{\nu N'}\backslash U$, we get

$$J_O^T(f,\lambda) = \int_{\mathbf{K}}\int_{L\backslash\mathbf{L}\cap\mathbf{G}^1} e^{-\langle\lambda,H_L(l)\rangle}\sum_{N'\subset Q}\sum_{\nu\in(O\cap U')N'/U''N'}$$
$$\int_{U_{\nu N'}\backslash\mathbf{U}}\int_{\mathbf{N}'} f(k^{-1}l^{-1}u^{-1}\nu nulk)\,dn\,\chi_{\nu N'}^T(ul)\,du\,\delta_U(l^{-1})\,dl\,dk.$$

We split the integral over $U_{\nu N'}\backslash\mathbf{U}$ into integrals over $\mathbf{U}_{\nu N'}\backslash\mathbf{U}$ and $U_{\nu N'}\backslash\mathbf{U}_{\nu N'}$. The latter one drops out, since the integral over $\mathbf{N}'$ as a function of u and the function $\chi_{\nu N'}^T$ are left-invariant under $\mathbf{U}_{\nu N'}$, and with the usual normalisation, the measure of $U_{\nu N'}\backslash\mathbf{U}_{\nu N'}$ equals 1.

Granting Conjecture 0, H_P is left $\mathbf{U}$-invariant for all $P\in\mathcal{P}_\gamma^{\mathrm{infl}}$, hence so is $\chi_{\nu N'}^T$ and can be extracted from the integral over u. Substituting $nu = un'$, applying the adelic version of the above identity to

$$h(\nu u') = \int_{\mathbf{N}'} f(k^{-1}l^{-1}\nu u'n'lk)\,dn'$$

and combining the integrals over u' and n', we obtain

$$J_O^T(f,\lambda) = \int_{\mathbf{K}}\int_{L\backslash\mathbf{L}\cap\mathbf{G}^1} e^{-\langle\lambda,H_L(l)\rangle}\delta_U(l^{-1})\sum_{N'\subset Q}$$
$$\sum_{\nu\in(O\cap U')N'/U''N'}\int_{\mathbf{U}''\mathbf{N}'} f(k^{-1}l^{-1}\nu nlk)\,dn\,\chi_{\nu N'}^T(l)\,dl\,dk.$$

It remains to move the integral over $\mathbf{K}$ under the sum, to split the integral over $\mathbf{U}''\mathbf{N}'$ into integrals over $n'\in\mathbf{N}'\mathbf{U}''/\mathbf{U}''$ and $u''\in\mathbf{U}''$ and to substitute $u''l = lu'$, which produces a factor $\delta_{U''}(l)$. If we treat only the contribution from a fixed group P', we may substitute $n'l = lv'$ in the integral over $\mathbf{V}_{P'} = \mathbf{N}'\mathbf{U}''/\mathbf{U}''$, which produces a factor $\delta_{N'U''/U''}(l)$ and allows us to express everything in terms of $\psi = \varphi^{P'}$. □

6 The Principal Unipotent Contribution

6.1 Reduction to the Trivial Parabolic

A final formula can be obtained for the contribution of the principal unipotent conjugacy class in G, which we denote by G^{prin}. Let γ be an element. Its canonical parabolic Q is then a minimal parabolic, for which we choose a Levi component L. By Theorem 5, $G^{\mathrm{prin}}\cap Q^{\mathrm{can}}$ is a dense Q-conjugacy class in $U = U'$. The set

$\mathcal{P}_\gamma^{\mathrm{infl}}$ consists of all parabolics P containing Q. For each of them, $G^{\mathrm{prin}} \cap P$ is dense in $M^{\mathrm{prin}}N$, where N is the unipotent radical of P and M the Levi component containing L. By Theorem 3, γN is prehomogeneous under the action of $Q_{\gamma N} \subset P_{\gamma N}$ with generic orbit contained in the Q-conjugacy class of γ, hence in Q^{can}. Therefore we can take $N^{M^{\mathrm{prin}}} = N$, and one could even show that this is the only choice satisfying Hypothesis 1.

Thus, the definition given in Hypothesis 4 simplifies to

$$J^T_{G^{\mathrm{prin}},P}(f,\lambda) = \int_{Q\backslash \mathbf{G}^1} e^{-\langle \lambda, H_Q(x)\rangle} \sum_{\gamma \in (G^{\mathrm{prin}}\cap Q)N/N} \int_{\mathbf{N}} f(x^{-1}\gamma n' x)\, dn'\, \hat{\tau}_P^T(x)\, dx,$$

which depends only on the restriction of λ to $\mathfrak{a}_Q^G$. These distributions for various P can be expressed in terms of the one with $P = G$ (which does not depend on the truncation parameter), in which the ambient group G is replaced by M (indicated by a superscript M).

Lemma 8. *If Hypothesis 4 applies to principal unipotent orbits, then*

$$J^T_{G^{\mathrm{prin}},P}(f,\lambda) = \epsilon_P J^M_{M^{\mathrm{prin}},M}(f^P,\lambda^P)\theta_P^T(\lambda)^{-1},$$

where

$$f^P(m) = \int_{\mathbf{K}}\int_{\mathbf{N}} f(k^{-1}mnk)\, dn\, dk$$

is a compactly supported smooth function on $\mathbf{M}^1$ *and, in the notation of Sect. 2.5,* $\theta_P^T(\lambda) = e^{\langle \lambda, T_P\rangle}\theta_P(\lambda)$.

Proof. The natural map $M^{\mathrm{prin}} \cap Q \to (G^{\mathrm{prin}} \cap Q)N/N$ is a bijection, and with the usual integration formula for the decomposition $\mathbf{G} = \mathbf{NMK}$, the above expression can be written as

$$\int_{\mathbf{K}}\int_{M\cap Q\backslash \mathbf{M}\cap\mathbf{G}^1}\int_{N\backslash\mathbf{N}} e^{-\langle\lambda, H_{M\cap Q}(m)\rangle} \sum_{\gamma\in M^{\mathrm{prin}}\cap Q}\int_{\mathbf{N}} f(k^{-1}m^{-1}n^{-1}\gamma n' nmk)\, dn'\, \hat{\tau}_P^T(m)\, dn\, \delta_N(m)^{-1} dm\, dk,$$

where the integral over $N\backslash\mathbf{N}$ drops out. Now we substitute $n'm = mn$, thereby cancelling the factor $\delta_N(m)$, and split the integral over $M\cap Q\backslash \mathbf{M}\cap\mathbf{G}^1$ into integrals over $\mathbf{M}^1\backslash\mathbf{M}\cap\mathbf{G}^1 \cong \mathfrak{a}_M^G = \mathfrak{a}_P^G$ and $M\cap Q\backslash\mathbf{M}^1$. Since the elements $\gamma \in M$ act trivially on $\mathfrak{a}_M^G$, we get

$$\int_{\mathbf{K}} \int_{M\cap Q\backslash \mathbf{M}^1} e^{-\langle \lambda, H_{M\cap Q}(m)\rangle} \sum_{\gamma \in M^{\mathrm{prin}} \cap Q} \int_{\mathbf{N}} f(k^{-1}m^{-1}\gamma mnk)\, dn\, dm\, dk$$
$$\epsilon_P \int_{\mathfrak{a}_P^G} e^{-\langle \lambda, H\rangle} \hat{\tau}_P(H - T_P)\, dH,$$

where we have used that $\hat{\tau}_P^T(x) = \epsilon_P \hat{\tau}_P(H_P^T(x))$. □

6.2 Singularities of Zeta Integrals

Under Hypotheses 4 and 5, Lemma 7 shows that the distributions on both sides of the equality

$$J^T_{G^{\mathrm{prin}}}(f, \lambda) = \sum_{P \supset Q} J^T_{G^{\mathrm{prin}},P}(f^P, \lambda)$$

as well as in Lemma 8 can be expressed in terms of zeta integrals. Since the two possible interpretations of f_V^P coincide, there will be parallel formulas for zeta integrals. We prove them unconditionally.

Lemma 9. *For every Schwartz–Bruhat function φ on $\mathbf{V}$ and $\lambda \in (\mathfrak{a}_Q^*)^+$, we have*

$$Z^T_{G^{\mathrm{prin}}}(\varphi, \lambda) = \sum_{P \supset Q} Z^T_{G^{\mathrm{prin}},P}(\varphi^P, \lambda),$$

$$Z^T_{G^{\mathrm{prin}},P}(\varphi, \lambda) = \epsilon_P Z^M_{M^{\mathrm{prin}},M}(\varphi^P, \lambda^P)\theta_P^T(\lambda)^{-1}.$$

These functions are defined by convergent integral-sums for $\mathrm{Re}\,\lambda \in (\mathfrak{a}_Q^*)^+$ *and extend meromorphically to all λ with* $\mathrm{Re}\,\lambda$ *in a neighbourhood of zero. The function $Z^T_{G^{\mathrm{prin}}}(\varphi, \lambda)$ is holomorphic there.*

Proof. The space V is the direct sum of the prehomogeneous vector spaces V^P corresponding to the minimal parabolics P properly containing Q. For each such P, the space V^P is the isomorphic image, under the exponential map, of a root space for a fundamental root α of the maximal split torus in L, and we get a bijection between the set of F-irreducible summands of V and the set Δ_Q. The group L is F-anisotropic modulo centre, and its basic characters corresponding to the relative invariants of V, when restricted to the maximal split torus in the centre of L, are nothing but the elements of Δ_Q. This prompts us to write λ as a linear combination of those fundamental roots with certain coefficients s_α. These coefficients are then the values of λ on the elements of the dual basis, viz. the fundamental coweights $\check{\varpi}$.

The distribution $Z_{G^{\mathrm{prin}},G}(\varphi,\lambda)$ is a zeta integral without truncation on the prehomogeneous vector space V. By Theorem 4, it converges absolutely if $\langle \mathrm{Re}\,\lambda, \check{\varpi}\rangle > 0$ for all $\varpi \in \hat{\Delta}_Q$ and extends meromorphically to the whole space. Its singularities for $\mathrm{Re}\,\lambda$ in some neighbourhood of zero are at most simple poles along the hyperplanes where $\langle \lambda, \check{\varpi}\rangle = 0$ for some $\varpi \in \hat{\Delta}_Q$, and the multiple residue at any point λ_0 in that neighbourhood is described as follows. If P is the smallest parabolic containing Q such that λ_0 vanishes on $\mathfrak{a}_P^G$ (i.e., the singular hyperplanes containing λ_0 are indexed by $\hat{\Delta}_P$), then

$$\lim_{\lambda\to\lambda_0} Z_{G^{\mathrm{prin}},G}(\varphi,\lambda)\hat{\theta}_P(\lambda) = Z^M_{M^{\mathrm{prin}},M}(\varphi^P,\lambda_0).$$

An argument as in the proof of Lemma 8 shows the second asserted identity, which provides the convergence and meromorphic continuation of its left-hand side. The manipulations at the end of the proof of Lemma 7 are now valid unconditionally, thus proving the first identity for λ in the domain of convergence and hence for the meromorphically continued functions.

The theory of (G,Q)-families cannot be applied to meromorphic functions. One may remove the singularities of the zeta integral at $\lambda = 0$ either by multiplying with linear functions or by subtracting the principal part. The first method leads to the modified distribution

$$\tilde{Z}_{G^{\mathrm{prin}},G}(\varphi,\lambda) = Z_{G^{\mathrm{prin}},G}(\varphi,\lambda)\hat{\theta}_Q(\lambda)$$

and its analogues for Levi subgroups. Since the elements of the dual basis of Δ_Q^P are the projections of the $\check{\varpi}$ with $\varpi \in \hat{\Delta}_Q \setminus \hat{\Delta}_P$ onto $\mathfrak{a}_Q^P$, it follows that

$$\lim_{\lambda\to\lambda_0} \tilde{Z}_{G^{\mathrm{prin}},G}(\varphi,\lambda) = \tilde{Z}^M_{M^{\mathrm{prin}},M}(\varphi^P,\lambda_0).$$

This shows that the functions $c_P(\lambda) = e^{-\langle\lambda,T_P\rangle}\tilde{Z}^M_{M^{\mathrm{prin}},M}(\varphi^P,\lambda^P)$ make up a (G,Q)-family, which is a product of a frugal and a cofrugal one. We can rewrite our formula for the principal unipotent contribution tautologically as

$$Z^T_{G^{\mathrm{prin}}}(\varphi,\lambda) = \sum_{P\supset Q} \epsilon_P \tilde{Z}^M_{M^{\mathrm{prin}},M}(\varphi^P,\lambda^P)\hat{\theta}_Q^P(\lambda)^{-1}\theta_P^T(\lambda)^{-1}.$$

Now the regularity of the right-hand side for $\mathrm{Re}\,\lambda$ in a neighbourhood of zero follows from Lemma 1. $\square$

Let us discuss the second method of removing singularities that was mentioned in the proof. Note that the principal part of a meromorphic function on a complex space is not invariantly defined. Thus, we exploit the L-invariant splitting $V = V_P \oplus V_R$ of our prehomogeneous vector space valid for each pair of parabolics P and R containing Q for which Δ_Q is the disjoint union of Δ_Q^P and Δ_Q^R. Although the image of L in $\mathrm{Aut}(V)$ need not split accordingly, that of its centre does, leading to the decomposition $\mathfrak{a}_Q^G = \mathfrak{a}_P^G \oplus \mathfrak{a}_R^G$. Since R is determined by P and Q, we denote λ_R by $\lambda^{P/Q}$, which may serve as an argument for $Z^M_{M^{\mathrm{prin}},M}$, because both $\mathfrak{a}_R^G$ and $\mathfrak{a}_Q^P$ are canonically isomorphic to $\mathfrak{a}_Q/\mathfrak{a}_P$. (This approach is dual to the one applied in the proof of Lemma 6.1 in [A-inv].) We define the second modified distribution as

$$\tilde{Z}_{G^{\mathrm{prin}}}(\varphi,\lambda) = \sum_{P\supset Q} \epsilon_P Z^M_{M^{\mathrm{prin}},M}(\varphi^P,\lambda^{P/Q})\hat{\theta}_P(\lambda)^{-1}$$

(without the additional subscript G), which is also holomorphic for $\mathrm{Re}\,\lambda$ a neighbourhood of zero, because the poles along each singular hyperplane cancel. We have a version of this distribution for the Levi component M' of every parabolic $P' \supset Q$ in the role of G, and by induction we can easily prove the converse relation

$$Z_{G^{\mathrm{prin}},G}(\varphi,\lambda) = \sum_{P'\supset Q} \tilde{Z}^{M'}_{M'^{\mathrm{prin}}}(\varphi^{P'},\lambda^{P'/Q})\hat{\theta}_{P'}(\lambda)^{-1}.$$

Plugging its relative version into the formula for $Z^T_{G^{\mathrm{prin}},P}(\varphi,\lambda)$ and summing over P, we obtain after a change of summation a second formula

$$Z^T_{G^{\mathrm{prin}}}(\varphi,\lambda) = \sum_{P'\supset Q}\sum_{P\supset P'} \epsilon_P \tilde{Z}^{M'}_{M'^{\mathrm{prin}}}(\varphi^{P'},(\lambda^P)^{M\cap P'/M\cap Q})\hat{\theta}^P_{P'}(\lambda)^{-1}\theta^T_P(\lambda)^{-1}.$$

The functions $c_P(\lambda_{P'}) = e^{-\langle\lambda,T_P\rangle}\tilde{Z}^{M'}_{M'^{\mathrm{prin}}}(\varphi^{P'},(\lambda^P)^{M\cap P'/M\cap Q})$ for fixed P' and $\lambda^{P'}$ constitute a (G,P')-family, hence the inner sum is holomorphic in $\lambda_{P'}$ by Lemma 1. It is actually holomorphic in λ, because the family depends holomorphically on $\lambda^{P'}$ in the obvious sense. For $P' = G$, it reduces to $\tilde{Z}_{G^{\mathrm{prin}}}(\varphi,\lambda)$, while the contribution of $P' = Q$ converges to

$$Z^L_{L^{\mathrm{prin}}}(\varphi^Q)\int_{\mathfrak{a}_Q^G}\Gamma'_Q(H,T_Q)\,dH$$

as $\lambda \to 0$ by Lemma 2.2 of [A-inv].

6.3 Explicit Weight Factors

Now let $f \in C_c^\infty(G(F_S)^1)$ for a finite set S of places. As in Sect. 5.1, we have the expansion

$$J^T_{G^{\mathrm{prin}},P}(f,\lambda) = \sum_{[\gamma' N]_S} \zeta_P(S,\gamma',\lambda)J^T_P(\gamma' N,f,\lambda)$$

for any $P \supset Q$, where $[\gamma' N]_S$ runs through the $Q(F_S)$-conjugacy classes in $(G^{\mathrm{prin}}(F_S) \cap Q(F_S))N(F_S)/N(F_S)$. Here $\zeta_P(S, \gamma', \lambda)$ is a certain zeta function associated with the prehomogeneous vector space V^P, and the weighted orbital integral $J_P^T(\gamma' N, f, \lambda)$ is given by

$$\int_{Q^{\gamma' N}(F_S)^1 \backslash G(F_S)^1} e^{-\langle \lambda, H_Q(x) \rangle} \int_{N(F_S)} f(x^{-1}\gamma' n x)\, dn\, \hat{\tau}_P^T(x)\, dx.$$

As in the proof of Lemma 8, we see that

$$J_P^T(\gamma' N, f, \lambda) = \epsilon_P J_M^M(\gamma' N, f^P, \lambda^P) \theta_P^T(\lambda)^{-1},$$

where the superscript M indicates the analogue of the distribution for M in place of G. The latter is holomorphic, hence the zeta function is responsible for the remaining singularities of the product. We remove them by setting

$$\tilde{\zeta}_P(S, \gamma', \lambda) = \zeta_P(S, \gamma', \lambda) \hat{\theta}_Q^P(\lambda).$$

The preimage in $G(F_S)^{\mathrm{prin}} \cap Q(F_S)$ of a $Q(F_S)$-orbit $[\gamma' N]_S$ consists of several $Q(F_S)$-orbits. Using the isomorphism $N_\gamma \backslash N \times N_\gamma \to \gamma N$ as in the proof of Lemma 7, we get

$$J_M^M(\gamma' N, f^P, \lambda^P) = \sum_{[\gamma]_S : [\gamma N]_S = [\gamma' N]_S} J_P(\gamma, f, \lambda),$$

where the functions

$$J_P(\gamma, f, \lambda) = \int_{G^\gamma(F_S)^1 \backslash G(F_S)^1} e^{-\langle \lambda, H_Q^P(x) \rangle} f(x^{-1}\gamma x)\, dx$$

form a cofrugal (G, Q)-family because the weight factors do. In total, we obtain

$$J_{G^{\mathrm{prin}}, P}^T(f, \lambda) = \epsilon_P \sum_{[\gamma]_S} \tilde{\zeta}_P(S, \gamma, \lambda) J_P(\gamma, f, \lambda) \hat{\theta}_Q^P(\lambda)^{-1} \theta_P^T(\lambda)^{-1}.$$

By Lemmas 8 and 7, we have

$$\sum_{[\gamma]_S} \tilde{\zeta}_P(S, \gamma, \lambda) J_P(\gamma, f, \lambda) = \tilde{Z}_{M^{\mathrm{prin}}, M}^M(f_V^P, \lambda^P).$$

We have seen in the proof of Lemma 9 that the zeta integrals with removed singularities on the right-hand side form a cofrugal (G, Q)-family, and we deduce the same property for the functions $\tilde{\zeta}_P(S, \gamma, \lambda)$ with fixed S and γ by choosing f supported in the $G(F_S)$-conjugacy class of γ.

Summing the above formulas for $J^T_{G^{\mathrm{prin}},P}(f,\lambda)$ over $P \supset Q$ and applying Lemma 4 with $d_P(\lambda) = e^{-\langle\lambda,T_P\rangle} J_P(\gamma,f,\lambda)$, we obtain

$$J^T_{G^{\mathrm{prin}}}(f) = \epsilon_Q \sum_{[\gamma]_S} \sum_{P\supset Q} (\tilde{\xi}^P_Q)'(S,\gamma,0) J^T_P(\gamma,f),$$

where $(\tilde{\xi}^P_Q)'(S,\gamma,\lambda)$ is as in Lemma 1 and

$$J^T_P(\gamma,f) = \int_{G^\gamma(F_S)^1\backslash G(F_S)^1} f(x^{-1}\gamma x) w^T_P(x)\,dx$$

with the weight factor

$$w^T_P(x) = \lim_{\lambda\to 0} \sum_{P'\supset P} \epsilon^{P'}_P e^{-\langle\lambda,H^{P'}_P(x)+T_{P'}\rangle} \hat{\theta}^{P'}_P(\lambda)^{-1} \theta_{P'}(\lambda)^{-1}.$$

Applying Lemma 4 again, we get

$$w^T_P(x) = \sum_{P'\supset P} \epsilon^{P'}_P v^{P'}_P(H^{P'}_P(x)) v_{P'}(T_{P'})$$

in terms of the relative versions of the function

$$v_Q(X) = \int_{\mathfrak{a}^G_Q} \Gamma'_Q(H,X)\,dH = \epsilon_Q \int_{\mathfrak{a}^G_Q} \Gamma''_Q(H,X)\,dH,$$

where the equality of the integrals follows from Lemma 2(iii). Since $v_Q(X) = \epsilon_Q v_Q(-X)$, we can also write

$$w^T_P(x) = \sum_{P'\supset P} v^{P'}_P(-H^{P'}_P(x)) v_{P'}(T_{P'}) = v_P(T_P - H_P(x)),$$

where the last equality follows with Lemma 4.

There is an alternative formula. The $(M, Q\cap M)$-family giving rise to $(\tilde{\xi}^P_Q)'(S,\gamma,\lambda)$ depends only on $\gamma N(F_S)$, so we can write

$$J^T_{G^{\mathrm{prin}}}(f) = \epsilon_Q \sum_{P\supset Q} \sum_{[\gamma' N]_S} (\tilde{\xi}^P_Q)'(S,\gamma',0) J^T_P(\gamma' N,f),$$

where $J^T_P(\gamma' N,f)$ is the sum of the $J^T_P(\gamma,f)$ over all $[\gamma]_S$ with $[\gamma N]_S = [\gamma' N]_S$. Recombining the integrals, we get

$$J^T_P(\gamma' N,f) = \epsilon_P \int_{L^{\gamma' N}(F_S)\backslash L(F_S)} f^P(l^{-1}\gamma' l) v_P(H^T_P(l))\,dl.$$

7 Examples

We are going to illustrate the constructions of this paper by some examples, restricting ourselves to subregular unipotent conjugacy classes in low-dimensional split classical groups. Such a group G is up to isogeny either the group $\mathrm{GL}(V)$, where V is an F-vector space, or the subgroup stabilising a symmetric bilinear form b or a symplectic form ω on V. Parabolic subgroups are stabilisers of flags $V_0 \subset \cdots \subset V_r$ in V, which have to be self-dual in the orthogonal and symplectic cases, i.e., $V_i^\perp = V_{r-i}$ for each i. To every conjugacy class of unipotent elements $\gamma = \exp X$ or, equivalently, to every adjoint orbit of nilpotent elements X, one associates a partition of the natural number $\dim V$ (cf. Sect. 5.1 of [CG]). We avoid the orthogonal case, in which both assignments are not quite bijective. Although in the notation $\{V_0, \dots, V_r\}$ for a flag one ought include $V_0 = \{0\}$ and $V_r = V$, we will list only nonzero proper subspaces for brevity, so that G, considered as its own parabolic subgroup, appears as the stabiliser of the empty flag.

For each representative γ, we will present the canonical flag determining the canonical parabolic Q of γ, the corresponding prehomogeneous vector space defined in Theorem 5 and applied in Sect. 5.2, its basic relative invariants as in Theorem 4 and the split torus A_{L_v}/A_G, as mentioned after Hypothesis 5, by means of its faithful action on a subquotient of a suitable flag. We will also describe the poset $\mathcal{P}_\gamma^{\mathrm{infl}}(F)$ defined in Theorem 2(iv) and applied in Corollary 1. If applicable, we will indicate the splitting of $C(F)$ into truncation classes defined in Sect. 4.3 and the refined set $\mathcal{P}_\gamma^{\min\mathrm{infl}}(F)$. For each parabolic $P = MN$ in this set, we will give the group $N^{[\gamma]}$ defined in Hypothesis 1, whose present notation was introduced in Lemma 6.

7.1 General Linear Group of Rank 2

Here $G(F) = \mathrm{GL}(V)$ with $\dim V = 3$, and the subregular unipotent class corresponds to the partition $[2, 1]$. The canonical flag of a representative $\gamma = \exp X$ is $\{V_-, V_+\}$, where

$$V_- = \operatorname{Im} X, \qquad V_+ = \operatorname{Ker} X,$$

and X defines an isomorphism $V/V_+ \to V_-$. The Hasse diagram of these subspaces is shown in Fig. 1. The corresponding prehomogeneous vector space is $\operatorname{Hom}(V/V_+, V_-)$ with any nonzero linear function as basic relative invariant, and A_{L_v}/A_G acts on V_+/V_- by homotheties. The Hasse diagram of the parabolic subgroups P in $\mathcal{P}_\gamma^{\mathrm{infl}}(F)$, or rather their corresponding flags, is shown in Fig. 2.

For each such P with unipotent radical N, the related group $N' = N^{[\gamma]}$ is the unipotent radical of a parabolic P'. Here and below, we encode the assignment $P \mapsto P'$ in the Hasse diagram by an arrow between the corresponding flags. If no arrow starts at a flag, this means that we have $P' = P$ for the corresponding parabolic.

Fig. 1 Subspaces determined by γ

$$\begin{tikzcd} V \arrow[d, dash] \\ V_+ \arrow[d, dash] \\ V_- \arrow[d, dash] \\ \{0\} \end{tikzcd}$$

Fig. 2 Inflating parabolics

$$\begin{tikzcd} & \emptyset & \\ \{V_-\} \arrow[ur] & & \{V_+\} \arrow[ul] \end{tikzcd}$$

7.2 *Symplectic Group of Rank 2*

Here $G(F) = \mathrm{Sp}(V, \omega)$ with $\dim V = 4$, and the subregular unipotent class corresponds to the partition $[2, 2]$. The canonical flag of a representative $\gamma = \exp X$ is $\{V_0\}$, where

$$V_0 = \mathrm{Ker}\, X = \mathrm{Im}\, X.$$

The element X induces an isomorphism $V/V_0 \to V_0$ and defines symmetric bilinear forms b_+ on V/V_0 and b_- on V_0 by

$$b_+(u, v) = \omega(u, Xv) = b_-(Xu, Xv).$$

If b_+ or, equivalently, b_- splits over F into a product of two linear forms, then there are isotropic lines U_+/V_0, W_+/V_0 for b_+ and U_-, W_- for b_-. In this case X determines four additional F-subspaces with the properties

$$XU_+ = U_+^\perp = U_-, \qquad XW_+ = W_+^\perp = W_-.$$

The Hasse diagram of these subspaces is shown in Fig. 3 with the parts shaded that are only present in the split case. The corresponding prehomogeneous vector space is the space $\mathrm{Quad}(V/V_0)$ of quadratic forms on V/V_0 with the discriminant as basic relative invariant. The torus A_{L_ν}/A_G acts as the split special orthogonal group on $V/V_0 \cong V_0$ if $b_\pm$ is split, while it is trivial otherwise.

The class $C(F)$ splits into two truncation classes O and O' containing the elements for which the forms $b_\pm$ are anisotropic resp. split. The Hasse diagram of $\mathcal{P}_\gamma^{\min \mathrm{infl}}(F)$ for γ in O resp. O' is shown in Figs. 4 resp. 5 with the same encoding of the assignment $P \mapsto P'$ as above.

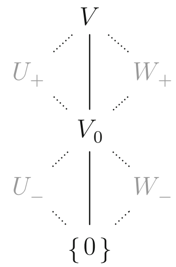

Fig. 3 Subspaces determined by γ

Fig. 4 Inflating parabolics for which O is minimal

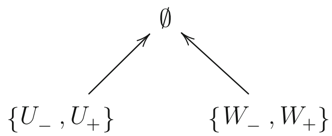

Fig. 5 Inflating parabolics for which O' is minimal

In this case, Lemma 7 is true unconditionally, see [HoWa] for details. The sum $Z_C^T(\varphi, \lambda) = Z_O^T(\varphi, \lambda) + Z_{O'}^T(\varphi, \lambda)$ of zeta integrals was called "adjusted zeta function" in [Yu].

7.3 General Linear Group of Rank 3

Here $G(F) = \mathrm{GL}(V)$ with $\dim V = 4$, and the subregular unipotent class corresponds to the partition $[3, 1]$. The canonical flag of a representative $\gamma = \exp X$ is $\{V_-, V_+\}$, where

$$V_- = \operatorname{Ker} X \cap \operatorname{Im} X = \operatorname{Im} X^2,$$

$$V_+ = \operatorname{Ker} X + \operatorname{Im} X = \operatorname{Ker} X^2.$$

The corresponding prehomogeneous vector space is

$$\mathrm{Hom}(V/V_+, V_+/V_-) \times \mathrm{Hom}(V_+/V_-, V_-),$$

and the value of the basic relative invariant on $\nu = (\nu_1, \nu_2)$ in this space is the composition $\nu_2 \circ \nu_1 \in \mathrm{Hom}(V/V_+, V_-)$. The torus A_{L_ν}/A_G acts by homotheties on $\operatorname{Ker} \nu_2$ stabilising $\operatorname{Im} \nu_1$. Figure 3 shows the Hasse diagram of the pertinent subspaces together with $\operatorname{Ker} X$ and $\operatorname{Im} X$, whose stabilisers also belong to the set $\mathcal{P}_\gamma^{\mathrm{infl}}(F)$ (Fig. 6).

Fig. 6 Subspaces determined by γ

$$
\begin{tikzcd}[arrows=-]
 & V \arrow[d] & \\
 & V_+ \arrow[dl] \arrow[dr] & \\
\operatorname{Ker} X \arrow[dr] & & \operatorname{Im} X \arrow[dl] \\
 & V_- \arrow[d] & \\
 & 0 &
\end{tikzcd}
$$

Fig. 7 Inflating parabolics

$$
\begin{tikzcd}
 & & & \emptyset & & & \\
\{\operatorname{Im} X\} \arrow[urrr] & & \{V_-\} \arrow[ur, dash] & & \{V_+\} \arrow[ul, dash] & & \{\operatorname{Ker} X\} \arrow[ulll] \\
 & \{V_-, \operatorname{Im} X\} \arrow[ul, dash] \arrow[ur] & & \{V_-, V_+\} \arrow[ul, dash] \arrow[ur, dash] & & \{\operatorname{Ker} X, V_+\} \arrow[ul] \arrow[ur, dash] &
\end{tikzcd}
$$

The Hasse diagram of the latter poset appears in Fig. 7 with the same encoding of the assignment $P \mapsto P'$ as above. There is no minimal parabolic contained in all its members, hence working with standard parabolic subgroups is inadequate. The stabilisers of $\operatorname{Ker} X$ and $\operatorname{Im} X$ are the first examples where the prehomogeneous affine space $\gamma N/N'$ is special under $P_{\gamma N}$, although the tangent prehomogeneous vector space $\mathfrak{n}/\mathfrak{n}'$ is not. The zeta integral $Z^T_{C,G}(\varphi, \lambda)$ in this case has not yet been meromorphically continued. It is the first example in which the truncation function χ^T_ν in Hypothesis 5 really depends on ν.

7.4 Symplectic Group of Rank 3

Here $G(F) = \operatorname{Sp}(V, \omega)$ with $\dim V = 6$, and the subregular unipotent class corresponds to the partition $[4, 2]$. The canonical flag of a representative $\gamma = \exp X$ is $\{V_-, V_0, V_+\}$, where

$$
\begin{aligned}
V_+ &= \operatorname{Ker} X^3 = \operatorname{Ker} X^2 + \operatorname{Im} X, \\
V_0 &= \operatorname{Ker} X^2 \cap \operatorname{Im} X = \operatorname{Ker} X + \operatorname{Im} X^2, \\
V_- &= \operatorname{Ker} X \cap \operatorname{Im} X^2 = \operatorname{Im} X^3.
\end{aligned}
$$

The element X induces isomorphisms

$$
X : V_+/V_0 \to V_0/V_-, \qquad X^2 : V/V_+ \to V_-
$$

and defines symmetric bilinear forms b_+ on V_+/V_0, b_- on V_0/V_- by

$$
b_+(u, v) = \omega(u, Xv) = b_-(Xu, Xv).
$$

The nonisotropic lines $\operatorname{Im} X/V_0$ for b_+ and $\operatorname{Ker} X/V_-$ for b_- will also play a role, whence we have included $\operatorname{Ker} X$ and $\operatorname{Im} X$ in the Hasse diagram of subspaces shown in Fig. 8. If b_+ or, equivalently, b_- splits over F into a product of two linear forms, then there are isotropic lines U_+/V_0, W_+/V_0 for b_+ and U_-/V_-, W_-/V_- for b_-. In this case X determines four additional F-subspaces, which are shaded in the diagram, with the properties

$$XU_+ = U_+^\perp = U_-, \qquad XW_+ = W_+^\perp = W_-.$$

In any case, $\operatorname{Hom}(V/V_+, V_+/V_0) \times \operatorname{Quad}(V_+/V_0)$ is the associated prehomogeneous vector space. One basic relative invariant is the discriminant of the quadratic form, the other one is given by composition and takes values in $\operatorname{Quad}(V/V_+)$. The torus A_{L_ν}/A_G is trivial for all ν.

The class $C(F)$ splits into two truncation classes O and O' containing the elements for which the forms $b_\pm$ are anisotropic resp. split. The Hasse diagram of $\mathcal{P}_\gamma^{\min\operatorname{infl}}(F)$ for γ in O resp. O' is shown in Figs. 9 resp. 10.

The class O' is the first example of a truncation class for whose elements γ the group $N' = N^{[\gamma]}$ cannot be chosen as the unipotent radical of a parabolic, hence cannot be encoded by arrows in the diagram. If N is the unipotent radical of the stabiliser of (U_-, U_+), we may set

$$\mathfrak{n}' = \{Z \in \mathfrak{n} \mid ZV \subset U_-,\ ZU_+ = 0\},$$

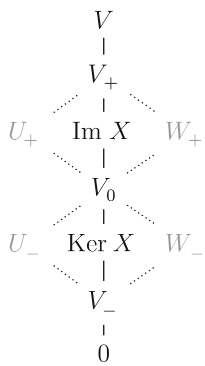

Fig. 8 Subspaces determined by γ

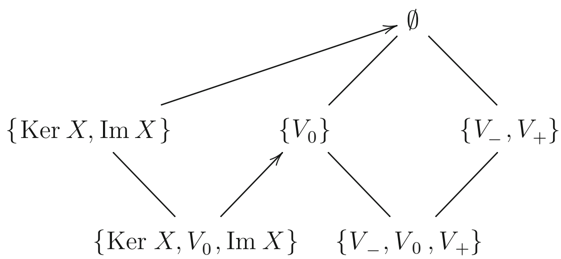

Fig. 9 Inflating parabolics for which O is minimal

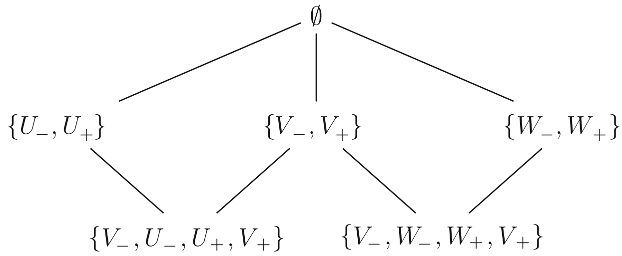

Fig. 10 Inflating parabolics for which O' is minimal

whereas if N is the stabiliser of (V_-, U_-, U_+, V_+), we may set

$$\mathfrak{n}' = \{Z \in \mathfrak{n} \mid ZU_+ \subset V_-,\ ZV_+ \subset U_-\}$$

and similarly with the letter U replaced by W. There are infinitely many N' for a fixed canonical parabolic, which suggests that one should search for another type of canonical subgroup attached to γ.

References

[A-int] J. Arthur, An introduction to the trace formula, in *Harmonic Analysis, the Trace Formula, and Shimura Varieties*, ed. by J. Arthur et al. Clay Mathematics Proceedings, vol. 4 (American Mathematical Society, Providence, 2005)

[A-inv] J. Arthur, The trace formula in invariant form. Ann. Math. **114**, 1–74 (1981)

[A-mix] J. Arthur, On a family of distributions obtained from orbits. Can. J. Math. **38**, 179–214 (1986)

[A-rk1] J. Arthur, The Selberg trace formula for groups of F-rank one. Ann. Math. **100**, 326–385 (1974)

[A-trI] J. Arthur, A trace formula for reductive groups I: terms associated to classes in $G(\mathbf{Q})$. Duke Math. J. **45**, 911–952 (1978)

[Bou] N. Bourbaki, Éléments de mathématique. *Groupes et algèbres de Lie*, chap. VIII (Hermann, Paris, 1968)

[Cha] P.-H. Chaudouard, Sur certaines contributions unipotentes dans la formule des traces d'Arthur (2015). arxiv:1510.02783 [math.RT]

[CG] D.H. Collingwood, W.M. McGovern, *Nilpotent Orbits in Semisimple Lie Algebras* (Van Nostrand Reinhold, New York, 1993)

[FL] T. Finis, E. Lapid, On the continuity of the geometric side of the trace formula (2015). arxiv:1512:08753 [math.NT]

[Fl] Y. Flicker, *The Trace Formula and Base Change for* GL_3. Lecture Notes in Mathematics, vol. 927 (Springer, Berlin, 1982)

[Ho-ind] W. Hoffmann, Induced conjugacy classes, prehomogeneous varieties, and canonical parabolic subgroups (2012). arxiv:1206.3068 [math.GR]

[Ho-rk1] W. Hoffmann, The non-semisimple term in the trace formula for rank one lattices. J. Reine Angew. Math. **379**, 1–21 (1987)

[Ho-est] W. Hoffmann, Geometric estimates for the trace formula. Ann. Glob. Anal. Geom. **34**, 233–261 (2008)
[HoWa] W. Hoffmann, S. Wakatsuki, On the geometric side of the Arthur trace formula for the symplectic group of rank 2 (2013). arxiv:1310.0541 [math.NT]
[Ki] T. Kimura, *Introduction to Prehomogeneous Vector Spaces* (American Mathematical Society, Providence, 2003)
[LS] G. Lusztig, N. Spaltenstein, Induced unipotent classes. J. Lond. Math. Soc. **19**(2), 41–52 (1979)
[Ma] J. Matz, Arthur's trace formula for GL(2) and GL(3) and non-compactly supported test functions. Dissertation, Universität Bonn (2011)
[MoWa] M. Morishita, T. Watanabe, A note on the mean-value theorem for special homogeneous spaces. Nagoya Math. J. **143**, 111–117 (1996)
[Ono] T. Ono, A mean value theorem in adele geometry. J. Math. Soc. Jpn. **20**, 276–288 (1968)
[Sa] F. Sato, Zeta functions in several variables associated with prehomogeneous vector spaces. I: functional equations. Tôhoku Math. J. **34**, 437–483 (1982)
[Sh] T. Shintani, On zeta functions associated with the vector space of quadratic forms. J. Fac. Sci. Univ. Tokyo Sect. I A Math. **22**, 25–65 (1975)
[Sai] H. Saito, Convergence of the zeta functions of prehomogeneous vector spaces. Nagoya Math. J. **170**, 1–31 (2003)
[Sp] T.A. Springer, *Linear Algebraic Groups*, 2nd edn. (Birkhäuser, Boston 1998)
[We] A. Weil, Sur quelques résultats de Siegel. Summa Brasil. Math. **1**, 21–39 (1946)
[Yu] A. Yukie, On the Shintani zeta function for the space of binary quadratic forms. Math. Ann. **292**, 355–374 (1992)

The Local Langlands Conjectures for Non-quasi-split Groups

Tasho Kaletha

Abstract We present different statements of the local Langlands conjectures for non-quasi-split groups that currently exist in the literature and provide an overview of their historic development. Afterwards, we formulate the conjectural multiplicity formula for discrete automorphic representations of non-quasi-split groups.

1 Motivation and Review of the Quasi-Split Case

1.1 The Basic Form of the Local Langlands Conjecture

Let F be a local field of characteristic zero (see Sect. 1.6 for a brief discussion of this assumption) and let G be a connected reductive algebraic group defined over F. A basic problem in representation theory is to classify the irreducible admissible representations of the topological group $G(F)$. The Langlands classification reduces this problem to that of classifying the tempered irreducible admissible representations of $G(F)$, whose set of equivalence classes will be denoted by $\Pi_{\mathrm{temp}}(G)$. In this paper, we will focus exclusively on tempered representations.

The local Langlands conjecture, as outlined, for example, in [Bor79], proposes a partition of this set indexed by arithmetic objects that are closely related to representations of the absolute Galois group Γ of F. More precisely, let W_F be the Weil group of F. Then

$$L_F = \begin{cases} W_F, & F \text{ archimedean} \\ W_F \times SU_2(\mathbb{R}), & F \text{ non-archimedean} \end{cases}$$

is the local Langlands group of F, a variant of the Weil–Deligne group suggested in [LanC, p. 209] and [Kot84, p. 647]. Let $\widehat{G}$ be the connected complex Langlands dual group of G, as defined, for example, in [Bor79, §2] or [Kot84, §1], and let

This research is supported in part by NSF grant DMS-1161489.

T. Kaletha (✉)
University of Michigan Arbor, MI 48109, United States
e-mail: kaletha@umich.edu

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_6

${}^LG = \widehat{G} \rtimes W_F$ be the Weil-form of the L-group of G. Let $\Phi_{\text{temp}}(G)$ be the set of $\widehat{G}$-conjugacy classes of tempered admissible L-homomorphisms $L_F \to {}^LG$. We recall from [Lan83, IV.2], see also [Bor79, §8], that an L-homomorphism is a homomorphism $\phi : L_F \to {}^LG$ that commutes with the projections to W_F of its source and target. It is called admissible if it is continuous and sends elements of W_F to semi-simple elements of ${}^LG$. It is called tempered if its image projects to a bounded subset of $\widehat{G}$.

The basic form of the local Langlands conjecture is the following.

Conjecture A. *1. There exists a map*

$$LL : \Pi_{\text{temp}}(G) \to \Phi_{\text{temp}}(G), \tag{1}$$

with finite fibers $\Pi_\phi(G) = LL^{-1}(\phi)$.

2. The fiber $\Pi_\phi(G)$ is empty if and only if ϕ is not relevant, i.e. its image is contained in a parabolic subgroup of ${}^LG$ that is not relevant for G.

3. If $\phi \in \Phi_{\text{temp}}(G)$ is unramified, then each $\pi \in \Pi_\phi(G)$ is K_π-spherical for some hyperspecial maximal compact subgroup K_π and for every such K there is exactly one K-spherical $\pi \in \Pi_\phi(G)$. The correspondence $\Pi_\phi(G) \leftrightarrow \phi$ is given by the Satake isomorphism.

4. If one element of $\Pi_\phi(G)$ belongs to the essential discrete series, then all elements of $\Pi_\phi(G)$ do, and this is the case if and only if the image of ϕ is not contained in a proper parabolic subgroup of ${}^LG$ (or equivalently in a proper Levi subgroup of ${}^LG$).

5. If $\phi \in \Phi_{\text{temp}}(G)$ is the image of $\phi_M \in \Phi_{\text{temp}}(M)$ for a proper Levi subgroup $M \subset G$, then $\Pi_\phi(G)$ consists of the irreducible constituents of the representations that are parabolically induced from elements of $\Pi_{\phi_M}(M)$.

There are further expected properties, some of which are listed in [Bor79, §10] and are a bit technical to describe here. This basic form of the local Langlands conjecture has the advantage of being relatively easy to state. It is, however, insufficient for most applications. What is needed is the ability to address individual representations of $G(F)$, rather than finite sets of representations. Ideally this would lead to a bijection between the set $\Pi_{\text{temp}}(G)$ and a refinement of the set $\Phi_{\text{temp}}(G)$. Moreover, one needs a link between the classification of representations of reductive groups over local fields and the classification of automorphic representations of reductive groups over number fields. Both of these are provided by the refined local Langlands conjecture.

1.2 The Refined Local Langlands Conjecture for Quasi-Split Groups 1

Formulating the necessary refinement of the local Langlands conjecture is a non-trivial task. We will begin with the case when G is quasi-split, in which a statement has been known for some time.

Given $\phi \in \Phi_{\text{temp}}(G)$, we consider the complex algebraic group $S_\phi = \text{Cent}(\phi(L_F), \widehat{G})$. The arguments of [Kot84, §10] show that S_ϕ° is a reductive group. Let $\bar{S}_\phi = S_\phi/Z(\widehat{G})^\Gamma$. The first refinement of the basic local Langlands conjecture can now be stated as follows.

Conjecture B. *There exists an injective map*

$$\iota : \Pi_\phi \to Irr(\pi_0(\bar{S}_\phi)), \tag{2}$$

which is bijective if F is p-adic.

We have denoted here by Irr the set of equivalence classes of irreducible representations of the finite group $\pi_0(\bar{S}_\phi)$. Various forms of this refinement appear in the works of Langlands and Shelstad, see, for example, [SheC], as well as Lusztig [Lus83].

A further refinement rests on a conjecture of Shahidi stated in [Sha90, §9]. To describe it, recall that a Whittaker datum for G is a $G(F)$-conjugacy class of pairs (B, ψ), where B is a Borel subgroup of G defined over F with unipotent radical U, and ψ is a non-degenerate character $U(F) \to \mathbb{C}^\times$, i.e. a character whose restriction to each simple relative root subgroup of U is non-trivial. When G is adjoint, it has a unique Whittaker datum. In general, there can be more than one Whittaker datum, but there are always only finitely many. Given a Whittaker datum $\mathfrak{w} = (B, \psi)$, an admissible representation π is called $\mathfrak{w}$-generic if $\text{Hom}_{U(F)}(\pi, \psi) \neq 0$. A strong form of Shahidi's conjecture is the following.

Conjecture C. *Each set $\Pi_\phi(G)$ contains a unique $\mathfrak{w}$-generic constituent.*

This allows us to assume, as we shall do from now on, that ι maps the unique $\mathfrak{w}$-generic constituent of $\Pi_\phi(G)$ to the trivial representation of $\pi_0(\bar{S}_\phi)$. It is then more apt to write $\iota_\mathfrak{w}$ instead of just ι. We shall soon introduce another refinement, which will specify $\iota_\mathfrak{w}$ uniquely. One can then ask the question: How does $\iota_\mathfrak{w}$ depend on $\mathfrak{w}$. This dependence can be quantified precisely [KalGe, §4], but we will not go into this here. We will next state a further refinement that ties the sets $\Pi_\phi(G)$ into the stabilization of the Arthur–Selberg trace formula. It also has the effect of ensuring that the map $\iota_\mathfrak{w}$ is unique (provided it exists).

1.3 Endoscopic Transfer of Functions

Before we can state the next refinement of the local Langlands conjecture we must review the notion of endoscopic transfer of functions, and for this we must review the notion of endoscopic data and transfer factors. The notion of endoscopic data was initially introduced in [LS87] and later generalized to the twisted case in [KS99]. We will present the point of view of [KS99], but specialized to the ordinary, i.e. non-twisted, case.

Definition 1. 1. An endoscopic datum is a tuple $\mathfrak{e} = (G^{\mathfrak{e}}, \mathcal{G}^{\mathfrak{e}}, s^{\mathfrak{e}}, \eta^{\mathfrak{e}})$, where $G^{\mathfrak{e}}$ is a quasi-split connected reductive group defined over F, $\mathcal{G}^{\mathfrak{e}}$ is a split extension of $\widehat{G}^{\mathfrak{e}}$ by W_F (but without a chosen splitting), $s^{\mathfrak{e}} \in \hat{G}$ is a semi-simple element, and $\eta^{\mathfrak{e}} : \mathcal{G}^{\mathfrak{e}} \to {}^LG$ is an L-homomorphism that restricts to an isomorphism of complex reductive groups $\hat{G}^{\mathfrak{e}} \to \mathrm{Cent}(s^{\mathfrak{e}}, \hat{G})^{\circ}$ and satisfies the following: There exists $s' \in Z(\hat{G})s^{\mathfrak{e}}$ such that for all $h \in \mathcal{G}^{\mathfrak{e}}$, $s'\eta^{\mathfrak{e}}(h) = \eta^{\mathfrak{e}}(h)s'$.
2. An isomorphism between endoscopic data $\mathfrak{e}_1$ and $\mathfrak{e}_2$ is an element $g \in \widehat{G}$ satisfying $g\eta^{\mathfrak{e}_1}(\mathcal{G}^{\mathfrak{e}_1})g^{-1} = \eta^{\mathfrak{e}_2}(\mathcal{G}^{\mathfrak{e}_2})$ and $gs^{\mathfrak{e}_1}g^{-1} \in Z(\widehat{G}) \cdot s^{\mathfrak{e}_2}$.
3. A z-pair for $\mathfrak{e}$ is a pair $\mathfrak{z} = (G_1^{\mathfrak{e}}, \eta_1^{\mathfrak{e}})$, where $G_1^{\mathfrak{e}}$ is an extension of $G^{\mathfrak{e}}$ by an induced[1] torus with the property that $G_{1,\mathrm{der}}^{\mathfrak{e}}$ is simply connected, and $\eta_1^{\mathfrak{e}} : \mathcal{G}^{\mathfrak{e}} \to {}^LG_1^{\mathfrak{e}}$ is an injective L-homomorphism that restricts to the homomorphism $\widehat{G}^{\mathfrak{e}} \to \widehat{G}_1^{\mathfrak{e}}$ dual to the given projection $G_1^{\mathfrak{e}} \to G^{\mathfrak{e}}$.

We emphasize here that the crucial properties of a z-pair are that the representation theory of $G^{\mathfrak{e}}(F)$ and $G_1^{\mathfrak{e}}(F)$ is very closely related, and that the map $\eta_1^{\mathfrak{e}}$ exists. The latter is a consequence of the simply-connectedness of the derived subgroup of $G_1^{\mathfrak{e}}$.

There are two processes that produce endoscopic data [She83, §4.2], one appearing in the stabilization of the geometric side of the trace formula, and one in the stabilization of the spectral side (or, said differently, in the spectral interpretation of the stable trace formula). These processes naturally produce the extension $\mathcal{G}^{\mathfrak{e}}$. This extension is, however, not always isomorphic to the L-group of $G^{\mathfrak{e}}$. The purpose of the z-pair is to circumvent this technical difficulty. It is shown in [KS99, §2.2] that z-pairs always exist.

In some cases the extension $\mathcal{G}^{\mathfrak{e}}$ is isomorphic to ${}^LG^{\mathfrak{e}}$ and the z-pair becomes superfluous. For example, this is the case when G_{der} is simply connected [Lan79, Proposition 1]. Further examples are the symplecic and special orthogonal groups. It is then more convenient to work with a hybrid notion that combines an endoscopic datum and a z-pair. Moreover, we can replace in the above definition $s^{\mathfrak{e}}$ by s' without changing the isomorphism class of the endoscopic datum. This leads to the following definition.

Definition 2. An extended endoscopic triple is a triple $\mathfrak{e} = (G^{\mathfrak{e}}, s^{\mathfrak{e}}, {}^L\eta^{\mathfrak{e}})$, where $G^{\mathfrak{e}}$ is a quasi-split connected reductive group defined over F, $s^{\mathfrak{e}} \in \widehat{G}$ is a

[1]We remind the reader that an induced torus is a product of tori of the form $\mathrm{Res}_{E/F}\mathbb{G}_m$ for finite extensions E/F.

semi-simple element, and ${}^L\eta^{\mathfrak{e}} : {}^LG^{\mathfrak{e}} \to {}^LG$ is an L-homomorphism that restricts to an isomorphism of complex reductive groups $\widehat{G}^{\mathfrak{e}} \to \mathrm{Cent}(s^{\mathfrak{e}}, \widehat{G})^\circ$ and satisfies $s^{L\eta^{\mathfrak{e}}}(h) = {}^{\mathfrak{e}L\eta^{\mathfrak{e}}}(h)s^{\mathfrak{e}}$.

The relationship between Definitions 1 and 2 is the following: If $(G^{\mathfrak{e}}, s^{\mathfrak{e}}, {}^L\eta^{\mathfrak{e}})$ is an extended endoscopic triple, then $(G^{\mathfrak{e}}, {}^LG^{\mathfrak{e}}, s^{\mathfrak{e}}, {}^L\eta^{\mathfrak{e}}|_{\widehat{G}^{\mathfrak{e}}})$ is an endoscopic datum. Moreover, even though $G^{\mathfrak{e}}$ will generally not have a simply connected derived group, one can take $(G^{\mathfrak{e}}, \mathrm{id})$ as a z-pair for itself.

In this paper we will work with the notion of an extended endoscopic triple. This will allow us to avoid some routine technical discussions. The more general case of an endoscopic datum and a z-pair doesn't bring any substantial changes but comes at the cost of burdening the exposition. Thus we now assume given an extended endoscopic triple $\mathfrak{e}$ for G, as well as a Whittaker datum $\mathfrak{w}$ for G. Associated with these data there is a transfer factor, i.e. a function

$$\Delta[\mathfrak{w}, \mathfrak{e}] : G^{\mathfrak{e}}_{\mathrm{sr}}(F) \times G_{\mathrm{sr}}(F) \to \mathbb{C},$$

where the subscript "sr" means semi-simple and strongly regular (those are the elements whose centralizer is a maximal torus). A variant of the factor Δ was defined in [LS87], then renormalized in [KS99], and slightly modified in [KS12] to make it compatible with a corrected version of the twisted transfer factor. We will review the construction, taking these developments into account. For the readers familiar with these references we note that the factor we are about to describe is the factor denoted by Δ'_λ in [KS12, (5.5.2)], in the case of ordinary endoscopy. In particular, it does not correspond to the relative factor Δ defined in [LS87]. The difference between the two lies in the inversion of the endoscopic element s. We work with this modified factor in order to avoid having to use inverses later when dealing with inner forms.

We first recall the notion of an admissible isomorphism between a maximal torus $S^{\mathfrak{e}}$ of $G^{\mathfrak{e}}$ and a maximal torus S of G. Let (T, B) be a Borel pair of G defined over F and let $(\widehat{T}, \widehat{B})$ be a Γ-stable Borel pair of $\widehat{G}$. Part of the datum of the dual group is an identification $X_*(T) = X^*(\widehat{T})$. The same is true for $G^{\mathfrak{e}}$ and we fix a Borel pair $(T^{\mathfrak{e}}, B^{\mathfrak{e}})$ of $G^{\mathfrak{e}}$ defined over F and a Γ-stable Borel pair $(\widehat{T}^{\mathfrak{e}}, \widehat{B}^{\mathfrak{e}})$ of $\widehat{G}^{\mathfrak{e}}$. The notion of isomorphism of endoscopic data allows us to assume that $\eta^{-1}(\widehat{T}, \widehat{B}) = (\widehat{T}^{\mathfrak{e}}, \widehat{B}^{\mathfrak{e}})$. Then η induces an isomorphism $X^*(\widehat{T}^{\mathfrak{e}}) \to X^*(\widehat{T})$, and this leads to an isomorphism $T^{\mathfrak{e}} \to T$. An isomorphism $S^{\mathfrak{e}} \to S$ is called admissible, if it is the composition of the following kinds of isomorphisms:

- $\mathrm{Ad}(h) : S^{\mathfrak{e}} \to T^{\mathfrak{e}}$ for $h \in G^{\mathfrak{e}}$.
- $\mathrm{Ad}(g) : S \to T$ for $g \in G$.
- The isomorphism $T^{\mathfrak{e}} \to T$.

Let $\gamma \in G^{\mathfrak{e}}_{\mathrm{sr}}(F)$. Let $S^{\mathfrak{e}} \subset G^{\mathfrak{e}}$ be the centralizer of γ, which is a maximal torus of $G^{\mathfrak{e}}$. Let $\delta \in G_{\mathrm{sr}}(F)$ and let $S \subset G$ be its centralizer. The elements γ and δ are called related if there exists an admissible isomorphism $S^{\mathfrak{e}} \to S$ mapping γ to δ. If such an isomorphism exists, it is unique, and will be called $\varphi_{\gamma,\delta}$.

Next, we recall the relationship between pinnings and Whittaker data from [KS99, §5.3]. Extend the Borel pair (T, B) to an F-pinning $(T, B, \{X_\alpha\})$. Here α runs over the set Δ of absolute roots of T in G that are simple relative to B and X_α is a non-zero root vector for α. Each X_α determines a homomorphism $\xi_\alpha : \mathbb{G}_a \to U$ by the rule $d\xi_\alpha(1) = X_\alpha$. Combining all homomorphisms x_α we obtain an isomorphism $\prod_\alpha \mathbb{G}_a \to U/[U, U]$. Composing the inverse of this isomorphism with the summation map $\prod_\alpha \mathbb{G}_a \to \mathbb{G}_a$ we obtain a homomorphism $U \to \mathbb{G}_a$ that is defined over F and hence leads to a homomorphism $U(F) \to F$. Composing the latter with an additive character $\psi_F : F \to \mathbb{C}^\times$ we obtain a character $\psi : U(F) \to \mathbb{C}^\times$ which is generic by construction. Thus (B, ψ) is a Whittaker datum. Since all Whittaker data arise from this construction, we may assume that our choices of pinning and ψ_F were made in such a way that (B, ψ) represents $\mathfrak{w}$.

We can now review the construction of the transfer factor $\Delta[\mathfrak{w}, \mathfrak{e}]$. If γ and δ are not related, we set $\Delta[\mathfrak{w}, \mathfrak{e}](\gamma, \delta) = 0$. Otherwise, it is the product of terms

$$\epsilon_L(V, \psi_F)\Delta_I^{-1}\Delta_{II}\Delta_{III_2}\Delta_{IV},$$

which we will explain now. Note that the term Δ_{III_1} of [LS87] is missing, as it is being subsumed by Δ_I in the quasi-split case. The letter V stands for the degree 0 virtual Galois representation $X^*(T) \otimes \mathbb{C} - X^*(T^{\mathfrak{e}}) \otimes \mathbb{C}$. The term $\epsilon_L(V, \psi)$ is the local L-factor normalized according to [Tat, §3.6]. The term Δ_{IV} is the quotient

$$\frac{|\det(\mathrm{Ad}(\delta) - 1|\mathrm{Lie}(G)/\mathrm{Lie}(S))|^{\frac{1}{2}}}{|\det(\mathrm{Ad}(\gamma) - 1|\mathrm{Lie}(G^{\mathfrak{e}})/\mathrm{Lie}(S^{\mathfrak{e}}))|^{\frac{1}{2}}}.$$

To describe the other terms, we need additional auxiliary data. We fix a set of a-data [LS87, §2.2] for the set $R(S, G)$ of absolute roots of S in G, which is a function

$$R(S, G) \to \overline{F}^\times, \alpha \mapsto a_\alpha$$

satisfying $a_{\sigma\lambda} = \sigma(a_\lambda)$ for $\sigma \in \Gamma$ and $a_{-\lambda} = -a_\lambda$. We also fix a set of χ-data [LS87, §2.5] for $R(S, G)$. To recall what this means, let $\Gamma_\alpha = \mathrm{Stab}(\alpha, \Gamma)$ and $\Gamma_{\pm\alpha} = \mathrm{Stab}(\{\alpha, -\alpha\}, \Gamma)$ for $\alpha \in R(S, G)$. Let F_α and $F_{\pm\alpha}$ be the fixed fields of Γ_α and $\Gamma_{\pm\alpha}$, respectively. Then $F_\alpha/F_{\pm\alpha}$ is an extension of degree 1 or 2. A set of χ-data is a set of characters $\chi_\alpha : F_\alpha^\times \to \mathbb{C}^\times$ for each $\alpha \in R(S, G)$, satisfying the conditions $\chi_{\sigma\alpha} = \chi_\alpha \circ \sigma^{-1}$, $\chi_{-\alpha} = \chi_\alpha^{-1}$, and if $[F_\alpha : F_{\pm\alpha}] = 2$, then $\chi_\alpha|_{F_{\pm\alpha}^\times}$ is non-trivial but trivial on the subgroup of norms from $F_\alpha^\times$.

With these choices, we have

$$\Delta_{II} = \prod_\alpha \chi_\alpha\left(\frac{\alpha(\delta) - 1}{a_\alpha}\right),$$

where the product is taken over the set $[R(S, G) \smallsetminus \varphi_{\gamma,\delta}^{*,-1}(R(S^{\mathfrak{e}}, G^{\mathfrak{e}}))]/\Gamma$.

The term Δ_I involves the so-called splitting invariant [LS87, §2.3] of S. Let $g \in G$ be such that $gTg^{-1} = S$. Write $\Omega(T, G)$ for the absolute Weyl group. For each $\sigma \in \Gamma$ there exists $\omega(\sigma) \in \Omega(T, G)$ such that for all $t \in T$

$$\omega(\sigma)\sigma(t) = g^{-1}\sigma(gtg^{-1})g.$$

Let $\omega(\sigma) = s_{\alpha_1} \dots s_{\alpha_k}$ be a reduced expression and let n_i be the image of $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ under the homomorphism $\mathrm{SL}_2 \to G$ attached to the simple root vector X_{α_i}. Then $n(\sigma) = n_1 \dots n_k$ is independent of the choice of reduced expression. The splitting invariant of S is the class $\lambda \in H^1(\Gamma, S_{\mathrm{sc}})$ of the 1-cocycle

$$\sigma \mapsto \prod \alpha^\vee(a_\alpha) g(n(\sigma)[g^{-1}\sigma(g)]^{-1})g^{-1}.$$

The product runs over the subset $\{\alpha > 0, \sigma^{-1}\alpha < 0\}$ of $R(S, G)$, with positivity being taken with respect to the Borel subgroup gBg^{-1}. The term Δ_I is defined as

$$\langle \lambda, s^{\mathfrak{e}} \rangle$$

where the pairing $\langle -, - \rangle$ is the canonical pairing between $H^1(\Gamma, S_{\mathrm{sc}})$ and $\pi_0([\widehat{S}/Z(\widehat{G})]^\Gamma)$ induced by Tate–Nakayama duality. Here we interpret $s^{\mathfrak{e}}$ as an element of $[Z(\widehat{G}^{\mathfrak{e}})/Z(\widehat{G})]^\Gamma$, embed the latter into $\widehat{S}^{\mathfrak{e}}/Z(\widehat{G})$, and use the admissible isomorphism $\varphi_{\gamma,\delta}$ to transport it to $\widehat{S}/Z(\widehat{G})$.

We turn to the term Δ_{III_2}. The construction in [LS87, §2.6] associates with the fixed χ-data a $\widehat{G}$-conjugacy class of L-embeddings $\xi_G : {}^LS \to {}^LG$. This construction is rather technical and we will not review it here. Via the admissible isomorphism $\varphi_{\gamma,\delta}$, the χ-data can be transferred to $S^{\mathfrak{e}}$ and provides a $\widehat{G}^{\mathfrak{e}}$-conjugacy class of L-embeddings $\xi_{\mathfrak{e}} : {}^LS^{\mathfrak{e}} \to {}^LG^{\mathfrak{e}}$. The admissible isomorphism $\varphi_{\gamma,\delta}$ provides dually an L-isomorphism ${}^L\varphi_{\gamma,\delta} : {}^LS \to {}^LS^{\mathfrak{e}}$. The composition $\xi' = {}^L\eta \circ \xi_{\mathfrak{e}} \circ {}^L\varphi_{\gamma,\delta}$ is then another $\widehat{G}$-conjugacy class of L-embeddings ${}^LS \to {}^LG$. Via conjugation by $\widehat{G}$ we can arrange that ξ_G and ξ' coincide on $\widehat{S}$. Then we have $\xi' = a \cdot \xi_G$ for some $a \in Z^1(W_F, \widehat{S})$. The term Δ_{III_2} is then given by

$$\langle a, \delta \rangle$$

where $\langle -, - \rangle$ is the pairing given by Langlands duality for tori.

We have completed the review of the construction of the transfer factor $\Delta[\mathfrak{w}, \mathfrak{e}]$. We now recall the notion of matching functions from [KS99, §5.5].

Definition 3. Two functions $f^{\mathfrak{w},\mathfrak{e}} \in \mathcal{C}_c^\infty(G^{\mathfrak{e}}(F))$ and $f \in \mathcal{C}_c^\infty(G(F))$ are called matching (or $\Delta[\mathfrak{w}, \mathfrak{e}]$-matching, if we want to emphasize the transfer factor) if for all $\gamma \in G^{\mathfrak{e}}_{\mathrm{sr}}(F)$ we have

$$SO_\gamma(f^{\mathfrak{w},\mathfrak{e}}) = \sum_\delta \Delta[\mathfrak{w}, \mathfrak{e}](\gamma, \delta) O_\delta(f),$$

where δ runs over the set of conjugacy classes in $G_{\mathrm{sr}}(F)$.

We remark that the stable orbital integrals at regular (but possibly not strongly regular) semi-simple elements can be expressed in terms of the stable orbital integrals at strongly regular semi-simple elements by continuity, but one has to be careful with the summation index, see [LS87, §4.3]. The stable orbital integrals at singular elements can be related to the stable orbital integrals at regular elements, see [Kot88, §3].

One of the central pillars of the theory of endoscopy is the following theorem.

Theorem 4. *For each function $f \in \mathcal{C}_c^\infty(G(F))$ there exists a matching function $f^{\mathfrak{w},\mathfrak{e}} \in \mathcal{C}_c^\infty(G^{\mathfrak{e}}(F))$.*

In the case of archimedean F this theorem was proved by Shelstad in [She81] and [She82] in the setting of Schwartz-functions and extends to the setting of smooth compactly supported functions by the results of Bouaziz [Bou]. In the case of non-archimedean F the proof of this theorem involves the work of many authors, in particular Waldspurger [Wal97, Wal06], and Ngo [Ngo10].

1.4 The Refined Local Langlands Conjecture for Quasi-Split Groups 2

With the endoscopic transfer of functions at hand we can state the final refinement of the local Langlands conjecture in the setting of quasi-split groups.

Recall that Conjecture B asserted the existence of a map $\iota_{\mathfrak{w}} : \Pi_\phi(G) \to \mathrm{Irr}(\pi_0(\bar{S}_\phi))$. We can write this map as a pairing

$$\langle -,- \rangle : \Pi_\phi \times \pi_0(\bar{S}_\phi) \to \mathbb{C}, \qquad (\pi, s) \mapsto \mathrm{tr}(\iota_{\mathfrak{w}}(\pi)(s)).$$

When F is p-adic, so that the map $\iota_{\mathfrak{w}}$ is expected to be bijective, we may allow ourselves to call this pairing "perfect". Since $\pi_0(\bar{S}_\phi)$ may be non-abelian the word "perfect" is to be interpreted with care, but its definition is simply the one that is equivalent to saying that the map $\iota_{\mathfrak{w}}$, which can be recovered from $\langle -,- \rangle$, is bijective. Using this pairing we can form, for any $\phi \in \Phi_{\mathrm{temp}}(G)$ and $s \in S_\phi$ the virtual character

$$\Theta_\phi^s = \sum_{\pi \in \Pi_\phi(G)} \langle \pi, s \rangle \Theta_\pi, \tag{3}$$

where Θ_π is the Harish-Chandra character of the admissible representation π. Let now $\mathfrak{e} = (G^{\mathfrak{e}}, s^{\mathfrak{e}}, {}^L\eta^{\mathfrak{e}})$ be an extended endoscopic triple and $\phi^{\mathfrak{e}} \in \Phi_{\mathrm{temp}}(G^{\mathfrak{e}})$. Put $\phi = {}^L\eta^{\mathfrak{e}} \circ \phi^{\mathfrak{e}}$. It is then automatic that $s^{\mathfrak{e}} \in S_\phi$.

Conjecture D. *For any pair of matching functions $f^{\mathfrak{w},\mathfrak{e}}$ and f we have the equality*

$$\Theta^1_{\phi^{\mathfrak{e}}}(f^{\mathfrak{w},\mathfrak{e}}) = \Theta_\phi^{s^{\mathfrak{e}}}(f).$$

Note that this statement implies that the distribution Θ_ϕ^1 is stable.

This is the last refinement of the local Langlands conjecture for quasi-split groups. Notice that the linear independence of the distributions Θ_π, together with the disjointness of the packets $\Pi_\phi(G)$, implies that the map $\iota_{\mathfrak{w}}$ of (2) is unique, provided it exists and satisfies Conjectures B and D. On the other hand, these conjectures do not characterize the assignment $\phi \mapsto \Pi_\phi$. The most obvious case is that of those ϕ for which Π_ϕ is a singleton set. For them the content of the refined conjecture is that the constituent of Π_ϕ is generic with respect to each Whittaker datum and its character is a stable distribution. In the case of quasi-split symplectic and special orthogonal groups, Arthur [Art13] shows that the addition of a supplementary conjecture—twisted endoscopic transfer to GL_n—is sufficient to uniquely characterize the correspondence $\phi \mapsto \Pi_\phi$. For general groups such a unique characterization is sill not known.

From now on we will group these four conjectures under the name "refined local Langlands conjecture." In the archimedean case, this conjecture is known by the work of Shelstad. Many statements were derived in [She81, She82], but with an implicit set of transfer factors instead of the explicitly constructed ones that we have reviewed in the previous section, as those were only developed in [LS87]. The papers [SheT1, SheT2, SheT3] recast the theory using the canonical factors of [LS87] and provide many additional and stronger statements. In particular, the refined local Langlands conjecture is completely known for quasi-split real groups. We note here that Shelstad's work is not limited to the case of quasi-split groups. This will be discussed soon.

In the non-archimedean case, much less is known. On the one hand, there are general results for special kinds of groups. The case of GL_n (in which most of the refinements discussed here do not come to bear) is known by the work of Harris–Taylor [HT01] and Henniart [Hen00]. The book [Art13] proves the refined local Langlands conjecture for quasi-split symplectic and odd special orthogonal groups, and a slightly weaker version of it for even special orthogonal groups. Arthur's strategy has been reiterated in [Mok] to cover the case of quasi-split unitary groups. In these cases, the uniqueness of the generic constituent in Conjecture C is not proved. This uniqueness follows from the works of Moeglin–Waldspurger, Waldspurger, and Beuzart–Plessis, on the Gan–Gross–Prasad conjecture. A short proof can be found in [At15]. On the other hand, there are results about special kinds of representations for general classes of groups. The papers [DR09, KalEC] cover the case of regular depth-zero supercuspidal representations of unramified p-adic groups, while the papers [RY14, KalEp] cover the case of epipelagic representations of tamely ramified groups. Earlier work of Kazhdan–Lusztig [KL87] and Lusztig [Lus95] proves a variant of this conjecture for unipotent representations of split simple adjoint groups, where the representations are not assumed to be tempered and the character identities are not studied.

1.5 Global Motivation for the Refinement

We now take F to be a number field, and G to be a connected reductive group, defined and quasi-split over F. We fix a Borel subgroup $TU = B \subset G$ and generic character $\psi : U(F) \setminus U(\mathbb{A}_F) \to \mathbb{C}^\times$.

We have the stabilization [ArtS1, (0.4)] of the geometric side of the trace formula

$$I^G_{\mathrm{geom}}(f) = \sum \iota(G, G^{\mathfrak{e}}) S^{G^{\mathfrak{e}}}(f^{\mathfrak{e}}).$$

Here the sum runs over isomorphism classes of global elliptic extended endoscopic triples $\mathfrak{e} = (G^{\mathfrak{e}}, s^{\mathfrak{e}}, {}^L\eta^{\mathfrak{e}})$, $S^{G^{\mathfrak{e}}}$ is the so-called stable trace formula for $G^{\mathfrak{e}}$, and $f^{\mathfrak{e}} = (f^{\mathfrak{e}}_v)$ is a function on $G^{\mathfrak{e}}(\mathbb{A})$ such that $f^{\mathfrak{e}}_v$ matches f_v. We note that global extended endoscopic triples are defined in the same way as in the local case in Definition 2, with only one difference: The condition on $\eta^{\mathfrak{e}}$ is that there exists $a \in Z^1(W_F, Z(\widehat{G}))$ whose class is everywhere locally trivial, so that $s^{\mathfrak{e}}\eta^{\mathfrak{e}}(h) = z(\bar{h})\eta^{\mathfrak{e}}(h)s^{\mathfrak{e}}$ for all $h \in \mathcal{G}^{\mathfrak{e}}$, where $\bar{h} \in W_F$ is the projection of h. One checks that $\eta^{\mathfrak{e}}$ provides a Γ-equivariant injection $Z(\widehat{G}) \to Z(\widehat{G}^{\mathfrak{e}})$. The triple $\mathfrak{e}$ is called *elliptic* if this injection restricts to a bijection $Z(\widehat{G})^{\Gamma,\circ} \to Z(\widehat{G}^{\mathfrak{e}})^{\Gamma,\circ}$.

The trace formula is an identity of the form

$$I^G_{\mathrm{spec}}(f) = I^G_{\mathrm{geom}}(f),$$

where the right-hand side is [ArtS1, (0.1)] and the left hand side is [ArtS, (0.2)]. The stabilization of the geometric side has as a formal consequence a stabilization of the spectral side. This allows us to write

$$I^G_{\mathrm{disc}}(f) = \sum \iota(G, G^{\mathfrak{e}}) S^{G^{\mathfrak{e}}}_{\mathrm{disc}}(f^{\mathfrak{e}}).$$

Here I^G_{disc} is the essential part of I^G_{spec}, see [ArtS, (3.5)] or [ArtI, (4.3)]. It contains not only the trace of discrete automorphic representations of $G(\mathbb{A})$, but also some contributions coming from Eisenstein series. This is the part of the trace formula one would like to understand in order to study automorphic representations, and the stabilization identity is meant to shed some light on it.

However, it is a-priori unclear what the spectral content of $S^{G^{\mathfrak{e}}}_{\mathrm{disc}}(f^{\mathfrak{e}})$ is. The key to understanding this content lies in the refined local Langlands correspondence. Namely, just like the central ingredients of $I^G_{\mathrm{disc}}(f)$ are the characters of discrete automorphic representations, the central ingredients of $S^{G^{\mathfrak{e}}}_{\mathrm{disc}}(f^{\mathfrak{e}})$ are the *stable* characters of discrete automorphic L-packets. This is the content of Arthur's "stable multiplicity formula," as stated, for example, in [Art13, Theorem 4.1.2]. However, unlike the case of stable orbital integrals, which are defined unconditionally and in an elementary way, stable characters can only be defined once the refined local Langlands correspondence, or at least Conjectures A and B have been established. Granting these, they are the global analogs of the characters Θ^1_ϕ of Eq. (3) and can be constructed out of these once a suitable notion of global parameters has been introduced, as was done, for example, in [Art13]. A global discrete parameter ϕ provides local parameters $\phi_v : L_{F_v} \to {}^LG$ and the associated stable character is then the product of the characters $\Theta^1_{\phi_v}$ over all v. Moreover, to have a chance at proving the stable multiplicity formula, Conjecture D must also be established.

Another crucial ingredient in the interpretation of the spectral side of the stable trace formula is the multiplicity formula for discrete automorphic representations. Given a global discrete parameter ϕ one obtains from the local parameters $\phi_v : L_{F_v} \to {}^LG$ the packets Π_{ϕ_v}. One also obtains a group $\bar{S}_\phi$ with maps $\bar{S}_\phi \to \bar{S}_{\phi_v}$. For each $\pi = \otimes'_v \pi_v$ with $\pi_v \in \Pi_{\phi_v}$ one considers the formula

$$m(\pi, \phi) = |\pi_0(\bar{S}_\phi)|^{-1} \sum_{x \in \pi_0(\bar{S}_\phi)} \prod_v \langle \pi_v, x \rangle.$$

It is then conjectured that the integer $m(\pi, \phi)$ is the ϕ-contribution of π to the discrete spectrum of G, and that the multiplicity of π in the discrete spectrum is equal to the sum of $m(\pi, \phi)$ over all (equivalence classes of) global parameters ϕ. We will discuss this formula in more detail in Sect. 5, where we will extend it to the case of non-quasi-split groups.

In all of these formulas, the existence of the map $\iota_{\mathfrak{w}} : \Pi_{\phi_v} \to \mathrm{Irr}\,(\pi_0(\bar{S}_{\phi_v}))$, and hence of the pairing $\langle -, - \rangle$, is crucial. There are further formulas which one can obtain, for example the inversion of endoscopic transfer, which allows one to obtain the characters of tempered representations from the stable characters of tempered L-packets. We refer the reader to [SheT3] for a statement of this in the archimedean case, and to [KalGe] for a sample application.

1.6 Remarks on the Characteristic of F

We have assumed throughout this section that F has characteristic zero. While it is believed that most of this material carries over in some form for fields (local or global as appropriate) of positive characteristic, most of the literature assumes that F has characteristic zero. For example, the work [LS87, LS90, KS99] is written with this assumption. The later work [KS12] is written for arbitrary local fields, which suggests that the definition of transfer factors should work in positive characteristic. However, the descent theory of [LS90] is not worked out in this setting. The fundamental lemma is proved in [Ngo10] in positive characteristic and then transfered to characteristic zero in [Wal09]. But the proof of the transfer theorem (Theorem 4) is only done in characteristic zero [Wal97]. Turning to the global situation, the theory of the trace formula, even before stabilization, for general reductive groups over global fields of positive characteristic is not developed. Thus, while most definitions, results, and conjectures, presented here are expected to hold (either in the same form or with some modifications) in positive characteristic, little factual information is actually present.

2 Non-quasi-split Groups: Problems and Approaches

We return now to the case of a local field F of characteristic zero and let G be a connected reductive group defined over F, but not necessarily quasi-split. We would like to formulate a refined local Langlands correspondence for G and to have global applications for it similar to the ones outlined in the last section. We are then met with the following problems

- There is no Whittaker datum, hence no canonical normalization of the transfer factor $\Delta(-,-)$.

The transfer factor $\Delta(-,-)$ is still defined in [LS87, KS99], but only up to a complex scalar. This has the effect that the notion of matching functions is also only defined up to a scalar. The trouble with this is that Conjecture D can no longer be stated in the precise form given above, and this makes the spectral interpretation of the stable trace formula problematic. Even worse, Arthur notices in [Art06, (3.1)] the following.

- Even the non-canonical normalizations of $\Delta(\gamma, \delta)$ are not invariant under automorphisms of endoscopic data.

This is a problem, because in the stabilization identity we are summing over isomorphism classes of endoscopic groups. The problem can be overcome, but it does indicate that something is not quite right.

- There is no good map $\iota : \Pi_\phi \to \mathrm{Irr}(\pi_0(\bar{S}_\phi))$.

The standard example for this comes from the work of Labesse and Langlands [LL79]. We follow here Shelstad's report [SheC]. Let F be p-adic and G the unique inner form of SL_2, so that $G(\mathbb{Q}_p)$ is the group of elements of reduced norm 1 in the unique quaternion algebra over F. We construct a parameter by taking a quadratic extension E/F and a character $\theta : E^\times \to \mathbb{C}^\times$ for which $\theta^{-1} \cdot (\theta \circ \sigma)$ is non-trivial and of order 2, where $\sigma \in \Gamma_{E/F}$ is the non-trivial element. Let $\sigma^\circ \in W_{E/F}$ be a lift of σ. Then

$$\phi(e) = \begin{bmatrix} \theta(e) & 0 \\ 0 & \theta(\sigma(e)) \end{bmatrix}, e \in E^\times, \qquad \phi(\sigma^\circ) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

is a homomorphism $W_{E/F} \to \mathrm{PGL}_2(\mathbb{C})$. One checks that

$$\bar{S}_\phi = S_\phi = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\} \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}.$$

The packet $\Pi_\phi(\mathrm{SL}_2(F))$ has exactly four elements. However, the packet $\Pi_\phi(G)$ has only one element π. Moreover, no character of χ of S_ϕ can be paired with this π so that the endoscopic character identities hold. In fact, in order to have the desired character identities, one must attach to π the function on $\bar{S}_\phi$ given by

$$f(s) = \begin{cases} 2, & s = 1 \\ 0, & s \neq 1 \end{cases}, \tag{4}$$

which is obviously not a character.

This last problem is most severe. Without the pairing between $\pi_0(\bar{S}_\phi)$ and Π_ϕ, we cannot state the global multiplicity formula and we cannot hope for a spectral interpretation of the stable trace formula.

2.1 Shelstad's Work on Real Groups

Despite these problems, we have a very good understanding of the case of real groups thanks to the work of Langlands and Shelstad. Langlands has constructed in [Lan97] the map (1) and has shown that Conjecture A holds. Shelstad has shown [She82, SheT2, SheT3] that once an arbitrary choice of the transfer factor $\Delta(-,-)$ has been fixed, and further choices specific to real groups have been made, there exists an embedding $\iota : \Pi_\phi(G) \to \mathrm{Irr}(\pi_0(\bar{S}_\phi))$, thus verifying Conjecture B, and has moreover shown that the corresponding pairing makes the endoscopic character identities of Conjecture D true. Even more, Shelstad has shown that if one combines the maps ι for multiple groups G, namely those that comprise a so-called K-group, then one obtains a bijection between the disjoint union of the corresponding L-packets and the set $\mathrm{Irr}(\pi_0(\bar{S}_\phi))$. For the notion of K-group we refer the reader to [Art99, §1] and [SheT3, §4], and we note here only that it is unrelated to the Adams–Barbasch–Vogan notion of strong real forms that we will encounter below.

It may be worth pointing out here that the group $\pi_0(\bar{S}_\phi)$ is always an elementary 2-group in the archimedean case, so that $\mathrm{Irr}(\pi_0(\bar{S}_\phi))$ is in fact the Pontryagin dual group of that elementary 2-group. This work uses the results of Harish-Chandra and Knapp–Zuckerman on the classification of discrete series, and more generally of tempered representations, of real semi-simple groups.

2.2 Arthur's Mediating Functions

Turning now to p-adic fields, the example of the inner form of SL_2 shows that we cannot expect to have a result in the p-adic case that is similar to that of Shelstad in the real case, because the virtual characters needed in the formulation of Conjecture D for general groups cannot be obtained from characters of $\pi_0(\bar{S}_\phi)$. In his monograph [ArtU], Arthur proposes to replace the pairing $\langle -,- \rangle$ by a combination of two objects. The first object is called the "spectral transfer factor," and denoted by $\Delta(\phi^\mathfrak{e}, \pi)$. Here again we assume to be given an extended endoscopic triple $\mathfrak{e}$ for G. We moreover assume fixed some arbitrary normalization of the transfer factor Δ, which we now qualify as "geometric," in order to distinguish it from the new "spectral" transfer factor. The spectral transfer factor takes as variables tempered parameters $\phi^{\mathfrak{e}}$ for $G^{\mathfrak{e}}$, as well as tempered representations π of $G(F)$. The role of the spectral transfer factor is to make the identity

$$\Theta^1_{\phi^{\mathfrak{e}}}(f^{\mathfrak{e}}) = \sum_{\pi} \Delta(\phi^{\mathfrak{e}}, \pi)\Theta_\pi(f)$$

true, whenever $f^{\mathfrak{e}}$ and f are matching with respect to the fixed normalization of the geometric transfer factor. Thus in particular the spectral factor depends on the geometric factor. Moreover, the isomorphisms of endoscopic data have to disturb the spectral factor in the same way that they disturb the geometric factor.

The second object is called the "mediating function," and denoted by $\rho(\Delta, s)$. The role of the mediating function is to make the product $\langle\pi, s\rangle = \rho(\Delta, s)\cdot\Delta(\phi^{\mathfrak{e}}, \pi)$ independent of the choice of geometric factor Δ, invariant under isomorphisms of endoscopic data, and a class function on the group $\pi_0(\bar{S}_\phi)$.

In the later paper [Art06], Arthur modifies this proposition to involve not the group $\bar{S}_\phi$, but rather its preimage S_ϕ^{sc} in the simply connected cover of $\widehat{G}$, and demands that $\langle\pi, s\rangle$ is not just a class function, but in fact a character of an irreducible representation of $\pi_0(S_\phi^{\mathrm{sc}})$. This is supported by the observation that the function (4) is indeed the character of the unique 2-dimensional irreducible representation of the quaternion group, which is the group S_ϕ^{sc} in the case of the inner forms of SL_2. Besides this observation, the introduction of the group S_ϕ^{sc} has its roots in Kottwitz's theorem [Kot86, Theorem 1.2] that relates the Galois cohomology set $H^1(\Gamma, G)$ to the Pontryagin dual of the finite abelian group $\pi_0(Z(\widehat{G})^\Gamma)$.

Let us be more precise. It is known that there exists a connected reductive group G^*, defined and quasi-split over F, together with an isomorphism $\xi : G^* \to G$ defined over $\overline{F}$ and having the property that for all $\sigma \in \Gamma_F$ the automorphism $\xi^{-1}\sigma(\xi)$ of G^* is inner. It is furthermore known that G^* is uniquely determined by G. Then G is called an inner form of G^* and $\xi : G^* \to G$ is called an inner twist. The inner twist provides an identification of the dual groups of G^* and G. The function $\sigma \mapsto \xi^{-1}\sigma(\xi)$ is an element of $Z^1(\Gamma, G^*_{\mathrm{ad}})$. Kottwitz's theorem interprets this element as a character $[\xi] : Z(\widehat{G}^*_{\mathrm{sc}})^\Gamma \to \mathbb{C}^\times$. Arthur suggests that one choose an arbitrary extension $\Xi : Z(\widehat{G}^*_{\mathrm{sc}}) \to \mathbb{C}^\times$ of this character. Then, for every $\phi \in \Phi_{\mathrm{temp}}(G^*)$, the L-packet $\Pi_\phi(G)$ should be in (non-canonical) bijection with the set $\mathrm{Irr}(\pi_0(S_\phi^{\mathrm{sc}}), \Xi)$ of irreducible representations of the finite group $\pi_0(S_\phi^{\mathrm{sc}})$ that transform under the image of $Z(\widehat{G}^*_{\mathrm{sc}})$ by the character Ξ. For $\pi \in \Pi_\phi(G)$, the character of the representation of $\pi_0(S_\phi^{\mathrm{sc}})$ corresponding to π via this bijection should be the class function $\langle\pi, -\rangle$. For each choice of geometric transfer factor Δ, we should have the expression $\langle\pi, -\rangle = \rho(\Delta, -)\cdot\Delta(\phi^{\mathfrak{e}}, \pi)$ as above.

This conjecture is stated uniformly for archimedean and non-archimedean local fields. In the archimedean case, this conjecture has been settled by Shelstad in [SheT2, SheT3], using deep information about the representation theory and harmonic analysis of real reductive groups. In the non-archimedean case, the conjecture is open. The main challenges that impede its resolution are that the conjectural objects $\Delta(\phi^{\mathfrak{e}}, \pi)$ and $\rho(\Delta, s)$ make the extension of the refined local Langlands conjecture to non-quasi-split groups less precise and harder to state, and this leads to a weaker grip on them by the trace formula.

2.3 Vogan's Pure Inner Forms

The work of Adams-Barbasch-Vogan [ABV92], introduces the following fundamental idea: When trying to describe L-packets, one should treat all reductive groups in a given inner class together. That is, instead of trying to describe the L-packets of G alone, one should fix the quasi-split inner form G^* of G and then describe the L-packets of all inner forms of G^* (of which G is one) at the same time. Here is a nice numerical example that underscores this idea: For a fixed positive integer n, the real groups $U(p,q)$ with $p+q=n$ constitute an inner class. For any discrete Langlands parameter ϕ one has $|S_\phi| = 2^n$ and $|\bar{S}_\phi| = 2^{n-1}$. On the other hand, one has $|\Pi_\phi(U(p,q))| = \binom{p+q}{q}$. Thus

$$| \sqcup_{p+q=n} \Pi_\phi(U(p,q))| = |S_\phi|.$$

Notice, however, that $U(p,q)$ and $U(q,p)$ are the same inner form of the quasi-split unitary group G^* (and are isomorphic as groups), but in order for the above equation to work out, we must treat them separately. This is not just a numerical quirk. It hints at a fundamental technical difficulty that will be of crucial importance.

In order to describe this difficulty more precisely, we need to recall a bit of Galois cohomology. The set of isomorphism classes of groups G which are inner forms of G^* is in bijection with the image of $H^1(\Gamma_F, G^*_{\mathrm{ad}})$ in $H^1(\Gamma_F, \mathrm{Aut}(G^*))$. However, this is a badly behaved set. Indeed, we can treat GL_n as an inner form of itself either via the identity map or via the isomorphism $g \mapsto g^{-t}$. Those two identifications clearly have different effects on representations. Thus, if we want to parameterize representations, we should treat these cases separately. This leads to considering not just the groups G which are inner forms of G^*, up to isomorphism, but rather inner twists $\xi : G^* \to G$, up to isomorphism. Here, an isomorphism from $\xi_1 : G^* \to G_1$ to $\xi_2 : G^* \to G_2$ is an isomorphism $f : G_1 \to G_2$ defined over F for which $\xi_2^{-1} \circ f \circ \xi_1$ is an inner automorphism of G^*. According to this definition, $f = id$ is not an isomorphism between the two inner twists $id : \mathrm{GL}_n \to \mathrm{GL}_n$ and $(-)^{-t} : \mathrm{GL}_n \to \mathrm{GL}_n$. In fact, we have achieved a rigidification of the problem, which means that we have cut down the automorphism group from $\mathrm{Aut}(G)(F)$ to $\mathrm{Aut}(\xi)$, where $\mathrm{Aut}(\xi)$ works out to be the subgroup of $\mathrm{Aut}(G)$ given by $G_{\mathrm{ad}}(F)$. However, as Vogan points out in [Vog93, §2], this rigidification is not enough. Indeed, we run into problems already with a group as simple as $G^* = \mathrm{SL}_2/\mathbb{R}$. Let $\theta = \mathrm{Ad}\begin{bmatrix} 1 & \\ & -1 \end{bmatrix}$. Then the map $f = \mathrm{id}$:

$G^* \to G^*$ is an isomorphism between the inner twists $id : G^* \to G^*$ and $\theta : G^* \to G^*$. However, θ swaps the constituents of discrete series L-packets (this can be computed explicitly in this example using K-types; it is, however, a general feature that the action of $G_{\mathrm{ad}}(F)$ on $G(F)$ preserves each tempered L-packet Π_ϕ, as one can see from the stability of Θ^1_ϕ and the linear independence of characters). This is a problem because we would like an isomorphism between inner twists to be compatible with the parameterization of L-packets.

This leads Vogan to introduce in [Vog93] the notion of a pure inner twist (in fact, Vogan calls it "pure rational form"), which is a pair (ξ, z) with $\xi : G^* \to G$ inner twist and $z \in Z^1(\Gamma, G^*)$ having the property $\xi^{-1}\sigma(\xi) = \mathrm{Ad}(z(\sigma))$. An isomorphism from (ξ_1, z_1) to (ξ_2, z_2) is now a pair (f, δ) with $f : G_1 \to G_2$ an isomorphism over F, $\delta \in G^*$ and satisfying the identities $\xi_2^{-1} f \xi_1 = \mathrm{Ad}(\delta)$ and $z_1(\sigma) = \delta^{-1} z_2(\sigma)\sigma(\delta)$. One can now check that $\mathrm{Aut}((\xi, z)) = G(F)$, thus an automorphism of (ξ, z) fixes each isomorphism class of representations and each rational conjugacy class of elements. We now finally have a shot of trying to parameterize the disjoint union of L-packets $\Pi_\phi((\xi, z))$, where (ξ, z) runs over the set of isomorphism classes of pure inner twists of a given quasi-split group G^*, and where $\Pi_\phi((\xi, z))$ is the L-packet on the group G that is the target of the pure inner twist $(\xi, z) : G^* \to G$. According to Vogan's formulation of the local Langlands correspondence [Vog93, Conjectures 4.3 and 4.15], there should exist a bijection

$$\iota_{\mathfrak{w}} : \sqcup_{(\xi,z)} \Pi_\phi((\xi, z)) \to \mathrm{Irr}(\pi_0(S_\phi)).$$

Note that we are not using $\bar{S}_\phi = S_\phi / Z(\widehat{G})^\Gamma$ here. In terms of the example with unitary groups, one checks that $U(p, q)$ and $U(q, p)$, despite being the same group, are not isomorphic pure inner twists of the quasi-split unitary group G^*. In fact, the set of isomorphism classes of pure inner twists of G^* is in bijection with $H^1(\Gamma_F, G^*)$. In the case of unitary groups, this set is precisely the set of pairs (p, q) of non-negative integers such that $p + q = n$.

Note furthermore that now, both in the real and in the p-adic case, the map $\iota_{\mathfrak{w}}$ is expected to be a bijection. Thus this generalization of Conjecture B makes it more uniform than its version for quasi-split groups. Moreover, it is still normalized to send the unique $\mathfrak{w}$-generic representation in $\Pi_\phi((id, 1))$ to the trivial representation of $\pi_0(S_\phi)$, i.e. it is compatible with Conjecture C.

The bijection $\iota_{\mathfrak{w}}$ is expected to fit in the following commutative diagram

$$\begin{tikzcd} \bigsqcup\limits_{(\xi,z)} \Pi_\phi((\xi, z)) \arrow[r, "\iota_{\mathfrak{w}}"] \arrow[d] & \mathrm{Irr}(\pi_0(S_\phi)) \arrow[d] \\ H^1(\Gamma, G) \arrow[r] & \pi_0(Z(\widehat{G})^\Gamma)^* \end{tikzcd} \tag{5}$$

The bottom map is Kottwitz's map [Kot86, Theorem 1.2]. The left map sends any constituent of $\Pi_\phi((\xi, z))$ to the class of z. The right map assigns to an irreducible representation of $\pi_0(S_\phi)$ the character by which the group $\pi_0(Z(\widehat{G})^\Gamma)$ acts. When F

is p-adic, the bottom map is a bijection. This means that the set $\Pi_\phi((\xi, z))$, which is an L-packet on the pure inner form G of G^* that is the target of the pure inner twist $(\xi, z) : G^* \to G$, is in bijection with the corresponding fiber of the right map. When F is real, one can obtain a similar statement by considering K-groups.

We have thus seen that Conjectures B and C generalize beautifully to pure inner twists. It was an observation of Kottwitz that Conjecture D also does. The first step is to construct a natural normalization of the geometric transfer factor for a pure inner twist $(\xi, z) : G^* \to G$ and an extended endoscopic triple $\mathfrak{e}$, which we shall call $\Delta[\mathfrak{w}, \mathfrak{e}, z]$. This was carried out in [KalEC, §2] and we will review it here. Let $\gamma \in G^{\mathfrak{e}}_{\mathrm{sr}}(F)$ and $\delta \in G_{\mathrm{sr}}(F)$ be related. Using a theorem of Steinberg one can show that there exists $g \in G^*$ such that $\delta = \xi(g\delta^* g^{-1})$ with $\delta^* \in G^*(F)$. By definition, γ and δ^* are also related, so the value $\Delta[\mathfrak{w}, \mathfrak{e}](\gamma, \delta^*)$ is non-zero. Moreover, $\sigma \mapsto g^{-1}z(\sigma)\sigma(g)$ is a 1-cocycle of Γ in $S = \mathrm{Cent}(\delta, G)$ whose class we call $\mathrm{inv}[z](\delta^*, \delta)$. We then set

$$\Delta[\mathfrak{w}, \mathfrak{e}, z](\gamma, \delta) = \Delta[\mathfrak{w}, \mathfrak{e}](\gamma, \delta^*) \cdot \langle \mathrm{inv}[z](\delta^*, \delta), s^{\mathfrak{e}} \rangle, \tag{6}$$

where $s^{\mathfrak{e}}$ is transported to $\widehat{S}$ via the maps $Z(\widehat{G}^{\mathfrak{e}})^\Gamma \to \widehat{S}^{\mathfrak{e}} \to \widehat{S}$, with $S^{\mathfrak{e}} = \mathrm{Cent}(\gamma, G^{\mathfrak{e}})$ and the second map coming from the admissible isomorphism $\phi_{\gamma,\delta}$. One then has to check that the function $\Delta[\mathfrak{w}, \mathfrak{e}, z]$ is indeed a geometric transfer factor and this is done in [KalEC, Proposition 2.2.2]. With the transfer factor and the bijection $\iota_{\mathfrak{w}}$ in place, we can now state Conjecture D exactly as it was stated in the case of quasi-split groups. We will give the statement of the new versions of Conjectures B, C, and D, together as a new conjecture.

Conjecture E. *Let G^* be a quasi-split connected reductive group defined over F and let $\mathfrak{w}$ be a Whittaker datum for G^*. Let $\phi \in \Phi_{\mathrm{temp}}(G^*)$. For each pure inner twist $(\xi, z) : G^* \to G$ let $\Pi_\phi((\xi, z))$ denote the L-packet $\Pi_\phi(G)$ of Conjecture A. Then there exists a bijection $\iota_{\mathfrak{w}}$ making Diagram 5 commutative and sending the unique $\mathfrak{w}$-generic constituent of $\Pi_\phi((id, 1))$ to the trivial representation of $\pi_0(S_\phi)$. Moreover, if $\mathfrak{e}$ is an extended endoscopic triple for G^* and if $f^{\mathfrak{e}} \in \mathcal{C}^\infty_c(G^{\mathfrak{e}}(F))$ and $f \in \mathcal{C}^\infty_c(G(F))$ are $\Delta[\mathfrak{w}, \mathfrak{e}, z]$-matching functions, then*

$$\Theta^1_{\phi^{\mathfrak{e}}}(f^{\mathfrak{e}}) = e(G) \sum_{\pi \in \Pi_\phi((\xi, z))} \langle \pi, s^{\mathfrak{e}} \rangle \Theta_\pi(f)$$

provided $\phi^{\mathfrak{e}} \in \Phi_{\mathrm{temp}}(G^{\mathfrak{e}})$ is such that $\phi = {}^L\eta^{\mathfrak{e}} \circ \phi^{\mathfrak{e}}$.

Here $e(G) \in \{\pm 1\}$ is the so-called Kottwitz sign of G, defined in [Kot83]. Note that the set $\Pi_\phi((\xi, z))$ does not depend on z (but we still need to include z in the notation for counting purposes, because the same ξ can be equipped with multiple z). The bijection $\iota_{\mathfrak{w}}$, however, does depend on z. We shall specify how later.

This conjecture is very close to the formulation of the local Langlands conjecture given by Vogan in [Vog93], apart from the fact that Vogan does not discuss endoscopic transfer. In the real case, it can be shown using Shelstad's work that this conjecture is true. We refer the reader to [KalR, §5.6] for details. In the p-adic case, its validity has been checked in [DR09, KalEC] for regular depth-zero supercuspidal L-packets. It has also been checked in [KalEp] for the L-packets consisting of epipelagic representations [RY14]. In fact, the latter work is valid in the broader framework of isocrystals with additional structure, which will be discussed next.

The relationship between the statements of Conjectures B and D given here and those suggested by Arthur in [ArtU] is straightforward. One has to replace S_ϕ^{sc} with S_ϕ and demand $\rho(\Delta[\mathfrak{w}, \mathfrak{e}, z], s^{\mathfrak{e}}) = 1$. This specifies the function $\rho(\Delta, s^{\mathfrak{e}})$ uniquely and Arthur's formulation of the conjectures follows from the one given here.

It thus appears that pure inner twists provide a resolution to all problems obstructing a formulation of the refined local Langlands conjecture for general reductive groups. Unfortunately, this is not quite true. The theory is perfect for inner twists whose isomorphism class, which is an element of $H^1(\Gamma_F, G^*_{\mathrm{ad}})$, is in the image of the natural map $H^1(\Gamma_F, G^*) \to H^1(\Gamma_F, G^*_{\mathrm{ad}})$. However, since this map is in general not surjective, not every group G can be described as the target of a pure inner twist $(\xi, z) : G^* \to G$ of a quasi-split group G^*. Basic examples are provided by the groups of units of central simple algebras. These are inner forms of the quasi-split group $G^* = \mathrm{GL}_n$. However, the generalized Hilbert 90 theorem states that $H^1(\Gamma, G^*) = \{1\}$. Thus no non-trivial inner form of G^* can be made pure. There are also other examples, involving inner forms of symplectic and special orthogonal groups.

2.4 Work of Adams, Barbasch, and Vogan

The fact that pure inner forms are not sufficient to describe the refined local Langlands conjecture for all connected reductive groups begs the question of whether there exists a notion that is more general than pure inner forms yet still has the necessary structure as to allow a version of Conjecture E to be stated. In the archimedean case, such a notion is presented by Adams et al. in [ABV92]. It is the notion of a "strong rational form." The set of equivalence classes of strong rational forms contains the set of equivalence classes of pure inner forms. At the same time it is large enough to encompass all inner forms. Moreover, in [ABV92] a bijection

$$\iota : \sqcup_x \Pi_\phi(x)) \to \mathrm{Irr}_{\mathrm{alg}}(\pi_0(\tilde{S}_\phi))$$

is constructed, where $\tilde{S}_\phi$ is the preimage of S_ϕ in the universal covering of $\widehat{G}$. When $\widehat{G}$ is semi-simple, this covering is just $\widehat{G}_{\mathrm{sc}}$, but when $\widehat{G}$ is a torus, this covering is affine space. In general, it is a mix of these two cases.

Thus, the book [ABV92] contains a proof of suitable generalizations of Conjectures B and C. It does not discuss the character identities stated as Conjecture D. The main focus of [ABV92] is in fact the study of how non-tempered representations interface with the conjectures of Langlands and Arthur. This is a fascinating topic that is well beyond the scope of our review.

2.5 Kottwitz's Work on Isocrystals with Additional Structure

The notion of "strong rational forms" introduced by Adams et al. resolved in the archimedean case the problem that pure inner forms are not sufficient to allow a statement of Conjecture E that encompasses all connected reductive groups. It thus became desirable to find an analogous notion in the non-archimedean case. This was formally formulated as a problem in [Vog93, §9], where Vogan lists the desired properties that this conjectural notion should have. The solution in the archimedean case did not suggest in any way whether a solution in the non-archimedean case exists and where it might be found, as the construction of strong rational forms in [ABV92] made crucial use of the fact that $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ has only one non-trivial element.

Led by his and Langlands' work on Shimura varieties, Kottwitz introduced in [Kot85, Kot97] the set $B(G)$ of equivalence classes of isocrystals with G-structure, for any connected reductive group G defined over a non-archimedean local field. The notion of an isocrystal plays a central role in the classification of p-divisible groups. Let F be a p-adic field and F^u its maximal unramified extension, and L its completion. An isocrystal is a finite-dimensional L-vector space V equipped with a Frobenius-semi-linear bijection. According to Kottwitz, an isocrystal with G-structure is a $\otimes$-functor from the category of finite-dimensional representations of the algebraic group G to the category of isocrystals. This can be given a cohomological description. Indeed, the set of isomorphism classes of n-dimensional isocrystals can be identified with $H^1(W_F, \mathrm{GL}_n(\bar{L}))$, and the set of isomorphism classes of isocrystals with G-structure can be identified with $H^1(W_F, G(\bar{L}))$.

Manin has shown that the category of isocrystals is semi-simple and the simple objects are classified by the set $\mathbb{Q}$ of rational numbers. The rational number corresponding to a given simple object is called its slope. A general isocrystal is thus given by a string of rational numbers, called its slope decomposition. The objects of constant slope, i.e. the isotypic objects, are called *basic* isocrystals. Kottwitz generalizes this notion to the case of isocrystals with G-structure. The set $B(F)_{\mathrm{bas}}$ of equivalence classes of basic isocrystals with additional structure is a subset of $B(G)$.

Kottwitz shows that there exists a functorial injection $H^1(\Gamma, G) \to B(G)_{\mathrm{bas}}$. He furthermore shows that each element $b \in B(G)_{\mathrm{bas}}$ leads to an inner form G^b of G. More precisely, one needs to take b to be a representative of the equivalence class given by an element of $B(G)_{\mathrm{bas}}$, and then one obtains an inner twist $\xi : G \to G^b$. We will call the pair (ξ, b) an extended pure inner twist, for a lack of a better name.

The bijection $H^1(\Gamma, G) \to \pi_0(Z(\widehat{G})^\Gamma)^*$ used in Diagram 5 extends to a bijection $B(G)_{\mathrm{bas}} \to X^*(Z(\widehat{G})^\Gamma)$. This allows one to conjecture the existence of a diagram similar to 5, but with $B(G)_{\mathrm{bas}}$ in place of $H^1(\Gamma, G)$. In order to be able to state an analog of Conjecture E, the last missing ingredient is the normalization of the transfer factor. This has been established in [KalI, §2]. We will not review the construction here, as it is quite analogous to the one reviewed in the section on pure inner forms. The analog of Conjecture E in the context of isocrystals is then the following conjecture made by Kottwitz.

Conjecture F. *Let G^* be a quasi-split connected reductive group defined over F, $\mathfrak{w}$ a fixed Whittaker datum for G^*, and $\phi \in \Phi_{\text{temp}}(G^*)$. Let $S_\phi^\natural = S_\phi/[S_\phi \cap [\widehat{G}]_{der}]^\circ$. For each extended pure inner twist $(\xi, b) : G^* \to G$ let $\Pi_\phi((\xi, b))$ denote the L-packet $\Pi_\phi(G)$ provided by Conjecture A. Then there exists a commutative diagram*

$$\begin{tikzcd} \coprod_{(\xi,b)} \Pi_\phi((\xi,b)) \arrow[r, "\iota_{\mathfrak{w}}"] \arrow[d] & Irr(S_\phi^\natural) \arrow[d] \\ B(G^*)_{bas} \arrow[r] & X^*(Z(\widehat{G})^\Gamma) \end{tikzcd} \qquad (7)$$

in which the top arrow is bijective. We have used Irr to denote the set of irreducible algebraic representation of the disconnected reductive group $S_\phi^\natural$. The image of the unique $\mathfrak{w}$-generic constituent of $\Pi_\phi((id, 1))$ is the trivial representation of $S_\phi^\natural$.

Given an extended pure inner twist $(\xi, b) : G^* \to G$ and an extended endoscopic triple $\mathfrak{e}$ for G^*, for any $\Delta[\mathfrak{w}, \mathfrak{e}, b]$-matching functions $f^\mathfrak{e} \in \mathcal{C}_c^\infty(G^\mathfrak{e}(F))$ and $f \in \mathcal{C}_c^\infty(G(F))$ the equality

$$\Theta^1_{\phi^\mathfrak{e}}(f^\mathfrak{e}) = e(G) \sum_{\pi \in \Pi_\phi((\xi,b))} \langle \pi, s^\mathfrak{e} \rangle \Theta_\pi(f)$$

holds, where $\phi^\mathfrak{e} \in \Phi_{\text{temp}}(G^\mathfrak{e})$ is such that $\phi = {}^L\eta^\mathfrak{e} \circ \phi^\mathfrak{e}$.

A version of this conjecture was stated in [Rap95, §5], and later in [KalI, §2.4]. A verification of this conjecture was given in [KalI] for regular depth-zero supercuspidal parameters, and in [KalEp] for epipelagic parameters. Moreover, while we have only considered non-archimedean fields so far, the conjecture also makes sense for archimedean fields thanks to Kottwitz's recent construction [Kot] of $B(G)$ for all local and global fields.

Given this conjecture, there is the following obvious question: How much bigger is $B(G^*)_{\text{bas}}$ than $H^1(\Gamma, G^*)$? Is it enough to treat all reductive groups?

The answer is the following: When $Z(G^*)$ is connected, Kottwitz has shown that the natural map $B(G^*)_{\text{bas}} \to H^1(\Gamma, G^*_{\text{ad}})$ is surjective. In other words, every inner form can be enriched with the datum of an extended pure inner twist. For such groups G^*, Conjecture F provides a framework to treat all their inner forms. Important examples of such groups G^* are the group $\mathrm{GL}(N)$, whose inner forms are the multiplicative groups of central simple algebras of degree N; the unitary groups $U_{E/F}(N)$ associated with quadratic extensions E/F; as well as the similitude groups $GU_{E/F}(N)$, GSp_N, and GO_N.

At the other end of the spectrum are the semi-simple groups. For them, the natural injection $H^1(\Gamma, G^*) \to B(G^*)_{\mathrm{bas}}$ is surjective. Thus the set $B(G^*)_{\mathrm{bas}}$ does not provide any additional inner forms beyond the pure ones, and Conjecture F is the same as Conjecture E. In particular, no inner forms of $\mathrm{SL}(N)$ and $\mathrm{Sp}(N)$ can be reached by either conjecture.

3 The Canonical Galois Gerbe and Its Cohomology

In [LR87], Langlands and Rapoport introduced the notion of a "Galois gerbe." Their motivation is the study of the points on the special fiber of a Shimura variety. In [Kot97], Kottwitz observed that the set $B(G)$ can be described using the cohomology of certain Galois gerbes. This led to the idea that it might be possible to overcome the limitations of the set $B(G)$ discussed in the previous section by using different Galois gerbes.

In this section, we are going to describe the construction of a canonical Galois gerbe over a local field of characteristic zero and discuss its properties. We will see in the next section how this gerbe leads to a generalization of Conjecture E that encompasses all connected reductive groups.

3.1 The Canonical Galois Gerbe

Langlands and Rapoport define [LR87, §2] a Galois gerbe to be an extension of groups

$$1 \to u \to W \to \Gamma \to 1$$

where u is the set of $\overline{F}$-points of an affine algebraic group and Γ is the absolute Galois group of F. Given such a gerbe, one can let it act on $G^*(\overline{F})$ through its map to Γ and consider the cohomology group $H^1(W, G^*)$.

From now on, let F be a local field of characteristic zero. A simple example of a Galois gerbe can be obtained as follows. The relative Weil group of a finite Galois extension E/F is an extension of topological groups

$$1 \to E^\times \to W_{E/F} \to \Gamma_{E/F} \to 1.$$

Pulling back along the natural surjection $\Gamma_F \to \Gamma_{E/F}$ and then pushing out along the natural injection $E^\times \to \bar{F}^\times$ provides a Galois gerbe

$$1 \to \mathbb{G}_m \to \mathcal{E}_{E/F} \to \Gamma_F \to 1.$$

These are called Dieudonne gerbes in [LR87, §2] and are the ones that Kottwitz uses in [Kot97, §8] to provide an alternative description of the set $B(G^*)$. More precisely, Kottwitz shows that if T is an algebraic torus defined over F and split over E, then there is a natural isomorphism

$$H^1_{\mathrm{alg}}(\mathcal{E}_{E/F}, T) \to B(T),$$

where H^1_{alg} is the subgroup of H^1 consisting of the classes of those 1-cocycles whose restriction to $\mathbb{G}_m$ is a homomorphism $\mathbb{G}_m \to T$ of algebraic groups.

One could hope that using more sophisticated Galois gerbes might lead to a cohomology theory that allows an analog of Conjecture E to be stated that applies to all reductive algebraic groups. For this to work, the gerbe needs to have the following properties.

1. It should be naturally associated with any local field F of characteristic zero, so as to provide a uniform statement of the conjecture.
2. In order to have a well-defined cohomology group $H^1(W, G^*)$, the gerbe W needs to be rigid, i.e. have no unnecessary automorphisms. This amounts to the requirement $H^1(\Gamma, u) = 1$.
3. In order to be able to capture all reductive groups, the gerbe W has to have the property that $H^1(W, G^*)$ comes equipped with a natural map $H^1(W, G^*) \to H^1(\Gamma, G^*_{\mathrm{ad}})$ which is *surjective*.
4. In order to be useful for endoscopy, there needs to exist a TateNakayama type isomorphism identifying $H^1(W, G^*)$ with an object definable in terms of $\widehat{G}^*$.

There is of course no a priori reason or even a hint that a Galois gerbe satisfying these conditions should exist. In fact, some experimentation reveals that conditions 2 and 3 seem to pull in opposite directions.

However, it turns out that if one slightly enlarges the scope of consideration, a suitable gerbe does exist. Namely, one has to give up the requirement that u is an affine algebraic group and rather allow it to be a profinite algebraic group, whose $\bar{F}$-points will then carry the natural profinite topology. The pro-finite group u that we are going to consider is the following.

$$u = \varprojlim_{n,E/F} (\mathrm{Res}_{E/F}\, \mu_n)/\mu_n.$$

This is a profinite algebraic group that encodes in a certain way the arithmetic of F. One can show the following [KalR, Theorem 3.1].

Proposition 5. *We have the canonical identification*

$$H^2(\Gamma, u) = \begin{cases} \widehat{\mathbb{Z}}, & F \text{ is non-arch.} \\ \mathbb{Z}/2\mathbb{Z}, & F = \mathbb{R} \end{cases}, \qquad H^1(\Gamma, u) = 1.$$

Here the continuous cohomology groups are taken with respect to the natural topology on $u(\overline{F})$ coming from the inverse limit.

Thus there exists a canonical isomorphism class of extensions of Γ by u, and each extension in this isomorphism class has as its group of automorphisms only the inner automorphisms coming from u. This means that if we take W to be any extension in the canonical isomorphism class and consider the set $H^1(W, G)$, this set will be independent of the choice of W.

However, it turns out that this is not quite the right object to consider. For example, it does not come equipped with a map to $H^1(\Gamma, G_{\mathrm{ad}})$ when G is a connected reductive group. The following slight modification is better suited for our purposes: Define $\mathcal{A}$ to be the category of injections $Z \to G$, where G is an affine algebraic group and Z is a finite central subgroup. For an object $[Z \to G] \in \mathcal{A}$, let $H^1(u \to W, Z \to G)$ be the subset of $H^1(W, G)$ consisting of those classes whose restriction to u takes image in Z. This provides a functor $\mathcal{A} \to \mathsf{Sets}$ and there is an obvious natural transformation $H^1(u \to W, Z \to G) \to H^1(\Gamma, G/Z)$ between functors $\mathcal{A} \to$ Sets. Furthermore, when G is reductive, we have the obvious map $H^1(\Gamma, G/Z) \to H^1(\Gamma, G_{\mathrm{ad}})$.

3.2 Properties of $H^1(u \to W, Z \to G)$

The basic properties of the functor $H^1(u \to W, Z \to G)$ are summarized in the following commutative diagram [KalR, (3.6)]

$$\begin{tikzcd}
1 \arrow[r] & H^1(\Gamma, Z) \arrow[r, "\mathrm{Inf}"] \arrow[d] & H^1(u \to W, Z \to Z) \arrow[r, "\mathrm{Res}"] \arrow[d] & \mathrm{Hom}(u, Z)^\Gamma \arrow[d, equal] & \\
1 \arrow[r] & H^1(\Gamma, G) \arrow[r, "\mathrm{Inf}"] \arrow[d, equal] & H^1(u \to W, Z \to G) \arrow[r, "\mathrm{Res}"] \arrow[d, "a"] & \mathrm{Hom}(u, Z)^\Gamma \arrow[r] \arrow[d, "b"] & * \arrow[d, equal] \\
 & H^1(\Gamma, G) \arrow[r] \arrow[d] & H^1(\Gamma, G/Z) \arrow[r] \arrow[d] \arrow[dl] & H^2(\Gamma, Z) \arrow[r] \arrow[d] & * \\
 & H^1(\Gamma, G/Z(G)) & 1 & 1 &
\end{tikzcd}$$

where $*$ is to be taken as $H^2(\Gamma, G)$ if G is abelian and disregarded otherwise. The three rows are exact, and so is the outer arc (after identifying the two copies of $\mathrm{Hom}(u, Z)^\Gamma$). The middle column is exact, and the map b is surjective. The middle exact sequence is an inflation-restriction-type sequence. By itself it already gives some information about the set $H^1(u \to W, Z \to G)$. First, it shows that $H^1(u \to W, Z \to G)$ contains as a subset $H^1(\Gamma, G)$, thus it faithfully captures the set of equivalence classes of pure inner forms. Second, it tells us that $H^1(u \to W, Z \to G)$ fibers over $\mathrm{Hom}(u, Z)^\Gamma$. One easily sees that the latter is finite, which implies

- $H^1(u \to W, Z \to G)$ is finite.

Using the basic twisting argument in group cohomology, one sees that the fibers of this fibration are of the form $H^1(\Gamma, G^\dagger)$, where $G^\dagger$ runs over suitable inner forms of G. In particular, we obtain the disjoint union decomposition

- $H^1(u \to W, Z \to G) = \bigsqcup H^1(\Gamma, G^\dagger)$.

This allows one to effectively compute $H^1(u \to W, Z \to G)$ using the standard tools of Galois cohomology. One can moreover ask, what is the meaning of $\mathrm{Hom}(u, Z)^\Gamma$. This question is answered by the map b. When Z is split (that is, when $X^*(Z)$ has trivial Γ-action), the map b in the above diagram is bijective. Thus, in a slightly vague sense, the group u represents the functor $Z \mapsto H^2(\Gamma, Z)$ restricted to the category of split finite multiplicative algebraic groups (the group u is itself of course not finite). Note that any continuous homomorphism $u \to Z$ factors through a finite quotient of u and is automatically algebraic, so we can write $\mathrm{Hom}(u, Z)^\Gamma = \mathrm{Hom}_F(u, Z)$. On the larger category of general finite multiplicative algebraic groups, one sees easily that the functor $Z \mapsto H^2(\Gamma, Z)$ is not representable, even in the above more vague sense, as it is not left exact. Nonetheless, the map b is surjective, so we can think of u as coming close to representing that functor. In other words, $H^1(u \to W, Z \to G)$ interpolates between $H^1(\Gamma, G)$ and $H^2(\Gamma, Z)$. Moreover, the surjectivity of b leads to the surjectivity of a. When G is reductive and Z is large enough, the map $H^1(\Gamma, G/Z) \to H^1(\Gamma, G_{\mathrm{ad}})$ is also surjective. For example, this is true as soon as $Z = Z(G_{\mathrm{der}})$. For some purposes it is thus sufficient to fix $Z = Z(G_{\mathrm{der}})$. In general, the flexibility afforded by allowing Z to vary is quite useful. For example, fixing Z would not provide a functorial assignment, and this would make basic operations like parabolic descent unnecessarily complicated.

3.3 Tate–Nakayama-Type Isomorphism

We have thus seen that the Galois gerbe W satisfies the first three of the four required properties listed in Sect. 3.1. The fourth property—the Tate–Nakayama-type isomorphism, is the most crucial. Luckily, the gerbe W satisfies that property too.

To give the precise statement, we let $\mathcal{R} \subset \mathcal{A}$ be the subcategory consisting of those $[Z \to G]$ for which G is connected and reductive. We have the functor

$$\mathcal{R} \to \mathsf{Sets}, \qquad [Z \to G] \mapsto H^1(u \to W, Z \to G).$$

We now define a second functor. Given $[Z \to G] \in \mathcal{R}$, let $\bar{G} = G/Z$. The isogeny $G \to \bar{G}$ provides an isogeny of Langlands dual groups $\widehat{\bar{G}} \to \widehat{G}$. Let $Z(\widehat{\bar{G}})^+$ denote the preimage in $\widehat{\bar{G}}$ of $Z(\widehat{G})^\Gamma$. Then $\pi_0(Z(\widehat{\bar{G}})^+)$ is a finite abelian group and one checks easily that

$$\mathcal{R} \to \mathsf{Sets}, \qquad [Z \to G] \mapsto \mathrm{Hom}(\pi_0(Z(\widehat{\bar{G}})^+), \mathbb{C}^\times)$$

is a functor.

The following theorem, proved in [KalR, §4], contains the precise statement how the gerbe W satisfies the expected property 4 of Sect. 3.1.

Theorem 6. • *There is a unique morphism between the two above functors that extends the Tate–Nakayama isomorphism between the restrictions of these functors to the subcategory consisting of objects* $[1 \to T]$*, where* T *is an algebraic torus, and that lifts a certain natural morphism* $Hom(\pi_0(Z(\widehat{\bar{G}})^+), \mathbb{C}^\times) \to Hom_F(u, Z)$*.*

- *The morphism is an isomorphism between the restrictions of the above functors to the subcategory consisting of objects* $[Z \to T]$*, where* T *is an algebraic torus.*
- *The morphism is an isomorphism between the above functors when* F *is non-archimedean.*
- *The kernel and cokernel of the morphism can be explicitly described when* F *is archimedean.*
- *The morphism restricts to Kottwitz's map on the subcategory of objects* $[1 \to G]$*.*

The fact that the morphism is not an isomorphism when F is archimedean is not surprising. If it were, it would endow each set $H^1(u \to W, Z \to G)$, and in particular each set $H^1(\Gamma, G)$, with the structure of a finite abelian group in a functorial way. However, it is generally not possible to endow $H^1(\Gamma, G)$ with a group structure in such a way that natural maps, like $H^1(\Gamma, G) \to H^1(\Gamma, G_{\mathrm{ad}})$, are group homomorphisms.

When F is p-adic, this theorem does endow the set $H^1(u \to W, Z \to G)$ with the structure of a finite abelian group in a functorial way. It furthermore gives a simple way to effectively compute the set $H^1(u \to W, Z \to G)$. The most important consequences of the theorem for us will, however, be to the theory of endoscopy. More precisely, the theorem will allow us to construct a normalization of the geometric transfer factor and to state a conjecture analogous to Conjecture E that encompasses all connected reductive groups.

4 Local Rigid Inner Forms and Endoscopy

In this section we are going to see how the Galois gerbe W constructed in the previous section leads to a generalization of Conjecture E that encompasses all connected reductive groups. Just like Conjecture E, its statement will be uniform for all local fields of characteristic zero.

We begin with a few simple definitions, essentially modeling those for pure inner forms. Let F be a local field of characteristic zero and let G^* be a quasi-split connected reductive group defined over F.

Definition 7. 1. A rigid inner twist $(\xi, z) : G^* \to G$ is a pair consisting of an inner twist $\xi : G^* \to G$ and an element $z \in Z^1(u \to W, Z \to G^*)$, for some finite central $Z \subset G^*$, such that $\xi^{-1}\sigma(\xi) = \mathrm{Ad}(\bar{z}(\sigma))$, where $\bar{z} \in Z^1(\Gamma, G^*_{\mathrm{ad}})$ is the image of z.

2. Given two rigid inner twists $(\xi_i, z_i) : G^* \to G_i$, $i = 1, 2$, an isomorphism (f, δ) : $(\xi_1, z_1) \to (\xi_2, z_2)$ of rigid inner twists is a pair consisting of an isomorphism $f : G_1 \to G_2$ defined over F and an element $\delta \in G^*$, satisfying the identities $\xi_2^{-1} f \xi_1 = \mathrm{Ad}(\delta)$ and $z_1(w) = \delta^{-1} z_2(w) \sigma_w(\delta)$.

Here σ_w is the image of $w \in W$ in Γ, and $\bar{z}$ is the image of $z \in Z^1(u \to W, Z \to G)$ in $Z^1(\Gamma, G/Z)$. It is again straightforward to check that $\mathrm{Aut}(\xi, z) = G(F)$.

4.1 Refined Endoscopic Data and Canonical Transfer Factors

The fact that the Tate–Nakayama-type isomorphism pairs the cohomology set $H^1(u \to W, Z \to G)$ not with elements of $\widehat{G}$, but rather of $\widehat{\bar{G}}$, leads to the necessity to modify the notion of endoscopic data. The notion of an endoscopic datum was reviewed in Sect. 1.3. Let $\mathfrak{e} = (G^{\mathfrak{e}}, \mathcal{G}^{\mathfrak{e}}, s^{\mathfrak{e}}, \eta^{\mathfrak{e}})$. A refinement of $\mathfrak{e}$ is a tuple $\dot{\mathfrak{e}} = (G^{\mathfrak{e}}, \mathcal{G}^{\mathfrak{e}}, s^{\dot{\mathfrak{e}}}, \eta^{\mathfrak{e}})$. The only difference is the element $s^{\dot{\mathfrak{e}}}$, which should be an element of $\widehat{\bar{G}}$ that lifts $s^{\mathfrak{e}}$. This refinement also suggests a modification of the notion of an isomorphism. Namely, an isomorphism between $\dot{\mathfrak{e}}_1$ and $\dot{\mathfrak{e}}_2$ is now an element $g \in \widehat{G}$ that satisfies two conditions. The first is $g\eta^{\mathfrak{e}_1}(\mathcal{G}^{\mathfrak{e}_1})g^{-1} = \eta^{\mathfrak{e}_2}(\mathcal{G}^{\mathfrak{e}_2})$, which is the same as before. To describe the second, let $H_i = G^{\mathfrak{e}_i}$. We use the canonical embedding $Z(G) \to Z(H_i)$ to form $\bar{H}_i = H_i/Z$. Then $\mathrm{Ad}(g)$ provides an isomorphism $\widehat{\bar{H}_1} \to \widehat{\bar{H}_2}$, which induces an isomorphism $\pi_0(Z(\widehat{\bar{H}_1})^+) \to \pi_0(Z(\widehat{\bar{H}_2})^+)$. The element $s^{\dot{\mathfrak{e}}_i}$ provides an element $\bar{s}^{\dot{\mathfrak{e}}_i} \in \pi_0(Z(\widehat{\bar{H}_i})^+)$ and we require that $\mathrm{Ad}(g)\bar{s}^{\dot{\mathfrak{e}}_1} = \bar{s}^{\dot{\mathfrak{e}}_2}$.

One checks that every endoscopic datum can be refined, and there are only finitely many isomorphism classes of refined endosocpic data that lead to isomorphic unrefined endoscopic data. This allows one to refine sums over isomorphism classes of endoscopic data by sums over isomorphism classes of refined endoscopic data.

One can analogously define the notion of a refined extended endoscopic triple, but we leave this to the reader.

The notion of a refined endoscopic data can be used, together with Theorem 6, to obtain a canonical normalization of the geometric transfer factor. The construction of the factor is essentially the same as the one for pure inner twists given by Eq. (6). Given a rigid inner twist $(\xi, z) : G^* \to G$ and a refined extended endoscopic triple $\dot{\mathfrak{e}}$, let $\gamma \in G^{\mathfrak{e}}(F)$ and $\delta \in G(F)$ be semi-simple strongly regular related elements, and let $\delta^* \in G^*(F)$ and $g \in G^*$ be as in Eq. (6). Then $g^{-1} \cdot z(w) \cdot \sigma_w(g)$ is an element of $Z^1(u \to W, Z \to S)$, whose class we call $\mathrm{inv}[z](\delta^*, \delta)$, and we set

$$\Delta[\mathfrak{w}, \dot{\mathfrak{e}}, z](\gamma, \delta) = \Delta[\mathfrak{w}, \mathfrak{e}](\gamma, \delta^*) \cdot \langle \mathrm{inv}[z](\delta^*, \delta), s^{\dot{\mathfrak{e}}} \rangle, \tag{8}$$

where now the pairing is between $H^1(u \to W, Z \to S)$ and $\pi_0([\widehat{\bar{S}}]^+)$ and is given by the Tate–Nakayama-type isomorphism of Theorem 6.

One can then prove [KalR, §5.3] the following.

Theorem 8. *The function* $\Delta[\mathfrak{w}, \dot{\mathfrak{e}}, z]$ *is indeed a transfer factor. Moreover, it is invariant under all automorphisms of* $\dot{\mathfrak{e}}$.

We see that the notion of refined endoscopic data and their isomorphisms resolves the problem of non-invariance of transfer factors under isomorphism noted by Arthur in [Art06].

4.2 Conjectural Structure of Tempered L-Packets

We are now ready to state the refined local Langlands conjecture for general connected reductive groups. Again we take G^* to be a quasi-split connected reductive group defined over F and we fix a Whittaker datum $\mathfrak{w}$ for it. We fix a finite central subgroup $Z \subset G^*$ and set as before $\bar{G}^* = G^*/Z$. Let $\phi \in \Phi_{\text{temp}}(G^*)$. We are of course interested in the L-packet for ϕ on non-quasi-split groups G that occur as inner forms of G^*. Recall $S_\phi = \text{Cent}(\phi, \widehat{G}^*)$. Set

$$S_\phi^+ = S_\phi \times_{\widehat{G}^*} \widehat{\bar{G}^*},$$

which is simply the preimage of S_ϕ under the isogeny $\widehat{\bar{G}^*} \to \widehat{G}^*$.

Conjecture G. *For each rigid inner twist* $(\xi, z) : G^* \to G$ *with* $z \in Z^1(u \to W, Z \to G^*)$ *let* $\Pi_\phi((\xi, z))$ *denote the L-packet* $\Pi_\phi(G)$ *whose existence is asserted by Conjecture A. Then there exists a commutative diagram*

$$\begin{array}{ccc} \bigsqcup_{(\xi,z)} \Pi_\phi((\xi,z)) & \xrightarrow{\iota_{\mathfrak{w}}} & Irr(\pi_0(S_\phi^+)) \\ \downarrow & & \downarrow \\ H^1(u \to W, Z \to G^*) & \longrightarrow & \pi_0(Z(\widehat{\bar{G}^*})^+)^* \end{array} \tag{9}$$

in which the top arrow is bijective. The image of the unique $\mathfrak{w}$*-generic constituent of* $\Pi_\phi((id, 1))$ *is the trivial representation of* $\pi_0(S_\phi^+)$.

Given a rigid inner twist $(\xi, z) : G^* \to G$ *and a refined endoscopic triple* $\dot{\mathfrak{e}}$ *for* G^*, *for any* $\Delta[\mathfrak{w}, \dot{\mathfrak{e}}, z]$*-matching functions* $f^{\dot{\mathfrak{e}}} \in \mathcal{C}_c^\infty(G^{\mathfrak{e}}(F))$ *and* $f \in \mathcal{C}_c^\infty(G(F))$ *the equality*

$$\Theta^1_{\phi^{\mathfrak{e}}}(f^{\dot{\mathfrak{e}}}) = e(G) \sum_{\pi \in \Pi_\phi((\xi,z))} \langle \pi, s^{\dot{\mathfrak{e}}} \rangle \Theta_\pi(f)$$

holds, where $\phi^{\mathfrak{e}} \in \Phi_{\text{temp}}(G^{\mathfrak{e}})$ *is such that* $\phi = {}^L\eta^{\mathfrak{e}} \circ \phi^{\mathfrak{e}}$, *and* $\langle \pi, - \rangle = tr(\iota_{\mathfrak{w}}(\pi)(-))$.

If we are interested in a particular fixed non-quasi-split group G, then we endow it with the datum of a rigid inner twist $(\xi, z) : G^* \to G$ and consider the fiber over the class of z of the diagram. On the left, this fiber is the L-packet on G (or rather, the K-group of G when F is archimedean), and on the right, this fiber consists of those irreducible representations which transform under $\pi_0(Z(\widehat{\bar{G}^*})^+)$ by the character determined by z.

Note that when G^* is split and semi-simple, the group S_ϕ^+ coincides with the group S_ϕ^{sc} suggested by Arthur in [Art06]. However, when G is a general connected reductive group, in particular a torus, then S_ϕ^+ is quite different, and in fact more closely related to the group used in [ABV92].

We emphasize also that the group $\pi_0(S_\phi^+)$ is in general more complicated than the groups $\pi_0(\bar{S}_\phi)$ or $\pi_0(S_\phi)$. Indeed, in the archimedean case the latter two groups are elementary 2-groups, while the former need not be a 2-group. It is still abelian, however. In the non-archimedean case it is known that the latter two groups may be non-abelian, but the former is non-abelian much more often. Indeed, already in the case of SL_2 the octonian group occurs as the group $\pi_0(S_\phi^+)$ for the parameter discussed in Sect. 2.

We have formulated the endoscopic character identities in Conjecture G only for refined extended endoscopic triples. For a formulation in the slightly more general context of refined endoscopic data and z-pairs, we refer the reader to [KalR, §5.4].

4.3 Results for Real Groups

So far we have not addressed the question of how the rigid inner forms we have defined, when specialized to the case $F = \mathbb{R}$, compare to the strong rational forms defined in [ABV92]. A-priori the two constructions are very different and in fact the construction of rigid inner twists was initially motivated by non-archimedean examples. Nonetheless, we have the following result [KalR, §5.2].

Theorem 9. *There is an equivalence between the category of rigid inner twists of a real reductive group and the category of strong rational forms of that group.*

We will not discuss here the precise definition of these categories and refer the reader to [KalR, §5.2] for their straightforward definition.

Another natural question to ask is: What can be said about Conjecture G when $F = \mathbb{R}$? As we discussed in Sect. 2.1, the structure of tempered L-packets and their endoscopic character identities are very well understood for real groups by the work of Shelstad. A careful study of her arguments leads to the following result [KalR, §5.6].

Theorem 10. *Conjecture G holds when $F = \mathbb{R}$.*

It is easy and instructive to explicitly compute the extension $1 \to u \to W \to \Gamma \to 1$ in the case of $F = \mathbb{R}$. In that case, $u(\mathbb{C}) = u(\mathbb{R})$ is the trivial Γ-module $\widehat{\mathbb{Z}}$

and the class of this extension can be represented by the 2-cocycle ξ determined by $\xi(\sigma,\sigma) = 1$, where $\sigma \in \Gamma$ is the non-trivial element. Recalling that the Weil group of $\mathbb{R}$ is an extension $1 \to \mathbb{C}^\times \to W_{\mathbb{C}/\mathbb{R}} \to \Gamma \to 1$ whose class can be represented by the 2-cocycle c determined by $c(\sigma,\sigma) = -1$, we see that it can be recovered as the pushout of W along the map $\widehat{\mathbb{Z}} \to \widehat{\mathbb{Z}}/2\widehat{\mathbb{Z}} \cong \{\pm 1\} \subset \mathbb{C}^\times$.

This computation shows that the extension $1 \to u \to W \to \Gamma \to 1$ is very closely related to the Weil group $W_{\mathbb{C}/\mathbb{R}}$. While for any finite Galois extension E/F of p-adic fields the relative Weil group $W_{E/F}$ has a similar structure as $W_{\mathbb{C}/\mathbb{R}}$, the absolute Weil group $W_{\overline{F}/F}$ is not an extension of the absolute Galois group Γ, but rather a dense subgroup of it. One can thus think of the extension $1 \to u \to W \to \Gamma \to 1$ as a closer analog for p-adic fields of the absolute Weil group of $\mathbb{R}$.

4.4 Dependence on the Choice of z

In Conjecture G we defined $\Pi_\phi((\xi,z))$ to be the L-packet $\Pi_\phi(G)$, where $(\xi,z) : G^* \to G$ is a rigid inner twist with $z \in Z^1(u \to W, Z \to G^*)$. It is clear from this definition that the set $\Pi_\phi((\xi,z))$ does not depend on z. What does depend on z is the representation of $\pi_0(S_\phi^+)$ that $\iota_{\mathfrak{w}}$ assigns to $\pi \in \Pi_\phi((\xi,z))$, and hence the value $\langle \pi, s^{\dot{e}} \rangle$ that enters the endoscopic character identity. This dependence can be quantified precisely.

Let $\xi : G^* \to G$ be an inner twist and let $z_1, z_2 \in Z^1(u \to W, Z \to G^*)$ be two elements such that (ξ, z_1) and (ξ, z_2) are rigid inner twists. According to Definition 7 and the diagram in Sect. 3.2 we have $z_2 = x z_1$ with $x \in Z^1(u \to W, Z \to Z) = Z^1(W, Z)$.

Let $\widehat{Z}$ denote the kernel of the isogeny $\widehat{\overline{G^*}} \to \widehat{G}^*$. It is shown in [KalRI, §6] that the finite abelian groups $H^1(W,Z)$ and $Z^1(\Gamma,\widehat{Z})$ are in canonical duality. Moreover, this duality is compatible with the duality between $H^1(u \to W, Z \to T)$ and $\pi_0([\widehat{\overline{T}}]^+)$ of Theorem 6.

Consider the map

$$(-d) : S_\phi^+ \to Z^1(\Gamma,\widehat{Z}), \qquad s \mapsto \phi(w_\sigma)s^{-1}\phi(w_\sigma)^{-1}s,$$

where $w_\sigma \in L_F$ is any lift of $\sigma \in \Gamma$. The result is independent of the lift because the finiteness of $\widehat{Z}$ implies $Z^1(L_F,\widehat{Z}) = Z^1(\Gamma,\widehat{Z})$. One can show that $(-d)$ is a group homomorphism. Moreover, since $[S_\phi^+]^\circ \subset \mathrm{Cent}(\phi, \widehat{\overline{G^*}})$, we see that $(-d)$ factors through $\pi_0(S_\phi^+)$. One can then show [KalRI, Lemma 6.2] that if $\pi \in \Pi_\phi(G)$ and if $\langle \pi, s^{\dot{e}} \rangle_1$ and $\langle \pi, s^{\dot{e}} \rangle_2$ are the values of $\mathrm{tr}(\iota_{\mathfrak{w}}(\pi)(s^{\dot{e}}))$ obtained by considering π as an element of $\Pi_\phi((\xi,z_1))$ and $\Pi_\phi((\xi,z_2))$ respectively, then the validity of Conjecture G implies

$$\langle \pi, s^{\dot{e}} \rangle_2 = \langle [x], (-d)s^{\dot{e}} \rangle \langle \pi, s^{\dot{e}} \rangle_1.$$

4.5 Comparison with Isocrystals

Even though Conjecture F cannot be stated for arbitrary connected reductive groups, as we discussed at the end of Sect. 2.5, it is still a very important part of the theory, due to the geometric significance of Kottwitz's theory of isocrystals with additional structure. For example, Conjecture F is the basis of Kottwitz's conjecture [Rap95, Conjecture 5.1] on the realization of the local Langlands correspondence in the cohomology of Rapoport–Zink spaces. Moreover, Fargues and Fontaine [FFC] have recently proved that G-bundles on the Fargues–Fontaine curve are parameterized by the set $B(G)$. Based on that, Fargues [Far] has outlined a geometric approach that would hopefully lead to a proof of Conjecture F. It it therefore desirable to understand the relationship between Conjectures F and G. This relationship is examined in [KalRI].

The simplest qualitative statement that can be made is the following: The validity of Conjecture F for all connected reductive groups with connected center is equivalent to the validity of Conjecture G for all connected reductive groups.

Let us now be more specific. Let G be a connected reductive group. Define

$$H^1(u \to W, Z(G) \to G) = \varinjlim H^1(u \to W, Z \to G)$$

where Z runs over the finite subgroups of $Z(G)$ defined over F. Then there exists a canonical map [KalRI, (3.14)]

$$B(G)_{\mathrm{bas}} \to H^1(u \to W, Z(G) \to G). \tag{10}$$

One can give an explicit formula for the dual of this map. For this, we need some preparation. Let $Z_n \subset Z(G)$ be the preimage in $Z(G)$ of the group of n-torsion points of the torus $Z(G)/Z(G_{\mathrm{der}})$. The Z_n form an exhaustive tower of finite subgroups of $Z(G)$ and we can use this tower to form the above limit. Set $G_n = G/Z_n$. Then $G_n = G_{\mathrm{ad}} \times Z(G_n)$ and $Z(G_n) = Z(G_1)/Z(G_1)[n]$, where $Z(G_1) = Z(G)/Z(G_{\mathrm{der}})$. Dually we have $\widehat{G}_n = \widehat{G}_{\mathrm{sc}} \times \widehat{C}_n$, where $\widehat{C}_n$ is the torus dual to $Z(G_n)$. Since $Z(G_1)$ is the maximal torus quotient of G, its dual $\widehat{C}_1$ is the maximal normal torus of $\widehat{G}$, i.e. $Z(\widehat{G})^\circ$. It will be convenient to represent $\widehat{C}_n$ as $\widehat{C}_1 = Z(\widehat{G})^\circ$, and then the natural quotient map $\widehat{C}_m \to \widehat{C}_n$ for $n|m$ becomes the m/n-power map $\widehat{C}_1 \to \widehat{C}_1$. Set $\widehat{C}_\infty = \varprojlim \widehat{C}_n$.

Consider the group $Z(\widehat{G}_{\mathrm{sc}}) \times \widehat{C}_\infty$. Elements of it are of the form $(a, (b_n)_n)$, where $a \in Z(\widehat{G}_{\mathrm{sc}})$ and $b_n \in \widehat{C}_1$ is a sequence satisfying $(b_m)^{m/n} = b_n$ for all $n|m$. We have the obvious map

$$Z(\widehat{G}_{\mathrm{sc}}) \times \widehat{C}_\infty \to Z(\widehat{G}), \qquad (a, (b_n)) \mapsto a_{\mathrm{der}} \cdot b_1,$$

where a_{der} is the image in $Z(\widehat{G}_{\mathrm{der}})$ of a. Let $(Z(\widehat{G}_{\mathrm{sc}}) \times \widehat{C}_\infty)^+$ be the subgroup consisting of those elements whose image in $Z(\widehat{G})$ is Γ-fixed. One can show that the duality pairing of Theorem 6 is compatible with the limit and becomes a pairing [KalRI, (3.12)]

$$\pi_0((Z(\widehat{G}_{\mathrm{sc}}) \times \widehat{C}_\infty)^+) \times H^1(u \to W, Z(G) \to G) \to \mathbb{C}^\times.$$

Now consider the map

$$(Z(\widehat{G}_{\mathrm{sc}}) \times \widehat{C}_\infty)^+ \to Z(\widehat{G}), \qquad (a, (b_n)) \mapsto \frac{a_{\mathrm{der}} \cdot b_1}{N_{E/F}(b_{[E:F]})}, \tag{11}$$

where E/F is any finite Galois extension so that Γ_E acts trivially on $Z(\widehat{G})$. The choice of E/F doesn't matter and one can show that the above map factors through $\pi_0((Z(\widehat{G}_{\mathrm{sc}}) \times \widehat{C}_\infty)^+)$ and is the map dual to (10), see [KalRI, Proposition 3.3].

We now turn to the comparison of Conjectures F and G. Assume first that G^* is a quasi-split connected reductive group with connected center. Let $\xi : G^* \to G$ be an inner twist. There exists a representative b of an element of $B(G^*)_{\mathrm{bas}}$ such that (ξ, b) is an extended pure inner twist. Via the map (10) (which also works on the level of cocycles) we obtain from b an element $z \in Z^1(u \to W, Z(G^*) \to G^*)$ so that (ξ, z) is a rigid inner twist. Then one can show [KalRI, §4] that Conjecture F for (ξ, b) is equivalent to Conjecture G for (ξ, z). Not only that, but one can explicitly relate the internal parameterization of the L-packets $\Pi_\phi((\xi, b))$ and $\Pi_\phi((\xi, z))$. This is realized by an explicit bijection

$$\mathrm{Irr}(S_\phi^\natural, b) \to \mathrm{Irr}(\pi_0(S_\phi^+), z),$$

where $\mathrm{Irr}(S_\phi^\natural, b)$ is the subset of those irreducible algebraic representations of $S_\phi^\natural$ which transform under $Z(\widehat{G})$ via the character determined by b, and $\mathrm{Irr}(\pi_0(S_\phi^+), z)$ is defined analogously. This bijection is given as the pull-back of representations under a group homomorphism

$$\pi_0(S_\phi^+) \to S_\phi^\natural$$

that can be defined as follows. We may take as the finite central subgroup $Z \subset G^*$ one of the groups Z_n defined above. Moreover, we can take it so that n is a multiple of the degree $k = [E : F]$ of some finite Galois extension E/F as above. Then $S_\phi^+ \subset \widehat{G}_n = \widehat{G}_{\mathrm{sc}} \times \widehat{C}_n$ and we define the above map to send $(a, b_n) \in S_\phi^+$ to $[a_{\mathrm{der}} \cdot b_n^n] N_{E/F}(b_n^{-\frac{n}{k}})$. In other words, we use the same formula as for (11).

We have thus compared Conjectures F and G for a fixed quasi-split group G^* with connected center. In order to obtain the above qualitative statement, we must now reduce the proof of Conjecture G to the case of groups with connected center. This is possible [KalRI, §5] and involves a construction, called a z-embedding, which embeds the connected reductive group G^* into another connected reductive group $\tilde{G}^*$ whose center is connected and whose endoscopy is comparable. One can then show that Conjecture G for G^* is equivalent to Conjecture G for $\tilde{G}^*$, see [KalRI, §5.2].

4.6 Relationship with Arthur's Formulation

The formulation of the refined local Langlands conjecture due to Arthur, that we briefly discussed in Sect. 2.2, is quite different from Conjecture G. For example, the group S_ϕ^{sc} that Arthur proposes is in general different from $\pi_0(S_\phi^+)$. Nonetheless, it turns out [KalGR, §4.6] that Conjecture G implies a strong form of Arthur's formulation. Let G^* be a quasi-split connected reductive group and let $\xi : G^* \to G$ be an inner twist. From ξ one obtains the 1-cocycle $\sigma \to \xi^{-1}\sigma(\xi)$, an element of $Z^1(\Gamma, G^*_{\mathrm{ad}})$. According to Kottwitz's theorem the class of this element provides a character $[\xi] : Z(\widehat{G}^*_{\mathrm{sc}})^\Gamma \to \mathbb{C}^\times$. Arthur suggests that one should choose an arbitrary extension $\Xi : Z(\widehat{G}^*_{\mathrm{sc}}) \to \mathbb{C}^\times$. Then, for any $\phi \in \Phi_{\mathrm{temp}}(G^*)$ there should be a non-canonical bijection between $\mathrm{Irr}(S_\phi^{\mathrm{sc}}, \Xi)$ and the L-packet $\Pi_\phi(G)$.

In order to relate Conjecture G to Arthur's formulation, it is not enough to choose $z \in Z^1(u \to W, Z \to G)$ so that (ξ, z) becomes a rigid inner twist. Rather, we consider the inner twist $\xi : G^*_{\mathrm{sc}} \to G_{\mathrm{sc}}$ on the level of simply connected covers induced by ξ and fix an element $z_{\mathrm{sc}} \in Z^1(u \to W, Z(G^*_{\mathrm{sc}}) \to G^*_{\mathrm{sc}})$ so that $(\xi, z_{\mathrm{sc}}) : G^*_{\mathrm{sc}} \to G_{\mathrm{sc}}$ becomes a rigid inner twist. According to the duality of Theorem 6, the class of $[z_{\mathrm{sc}}]$ provides a character $Z(\widehat{G}^*_{\mathrm{sc}}) \to \mathbb{C}^\times$ that extends the character $[\xi] : Z(\widehat{G}^*_{\mathrm{sc}})^\Gamma \to \mathbb{C}^\times$. Thus, we see that from our current point of view the choice of extension Ξ of the character $[\xi]$ corresponds to the choice of z_{sc} lifting the cocycle $\sigma \mapsto \xi^{-1}\sigma(\xi)$. In fact, when F is p-adic the class of $[z_{\mathrm{sc}}]$ and the extension Ξ determine each other. When F is real, however, the class $[z_{\mathrm{sc}}]$ is the primary object, because it determines Ξ, but is not determined by it.

The real strength of the new point of view comes from the fact that z_{sc} provides not just the character Ξ, but at the same time a normalization of the Langlands–Shelstad transfer factor Δ, namely $\Delta[\mathfrak{w}, \dot{\mathfrak{e}}, z]$, where $z \in Z^1(u \to W, Z(G^*_{\mathrm{der}}) \to G^*)$ is the image of z_{sc}. In this way it specifies the mediating function $\rho(\Delta, -)$ and the spectral transfer factor $\Delta(\phi^{\mathfrak{e}}, \pi)$. Namely, $\rho(\Delta[\mathfrak{w}, \dot{\mathfrak{e}}, z], s^{\dot{\mathfrak{e}}}) = 1$ and $\Delta(\phi^{\mathfrak{e}}, \pi) = \langle \pi, s^{\dot{\mathfrak{e}}} \rangle$.

Let us now show that the internal parameterization of the L-packet $\Pi_\phi(G)$ given by Conjecture G implies the parameterization expected by Arthur. Let $\bar{G}^* = G^*/Z(G^*_{\mathrm{der}}) = G^*_{\mathrm{ad}} \times Z(G^*)/Z(G^*_{\mathrm{der}})$. Then dually $\widehat{\bar{G}^*} = \widehat{G}^*_{\mathrm{sc}} \times Z(\widehat{G}^*)^\circ$. We have $Z(\widehat{\bar{G}^*}) = Z(\widehat{G}^*_{\mathrm{sc}}) \times Z(\widehat{G}^*)^\circ$ and the subgroup $Z(\widehat{\bar{G}^*})^+$ can be described as the set of pairs (a, z) such that $a_{\mathrm{der}} \cdot z \in Z(\widehat{G}^*)$ is Γ-fixed, where $a_{\mathrm{der}} \in Z(\widehat{G}^*)$ is the image of a. Similarly, the subgroup $S_\phi^+ \subset \widehat{\bar{G}^*}$ can be described as the set of pairs $(a, z) \in \widehat{G}^*_{\mathrm{sc}} \times Z(\widehat{G}^*)^\circ$ with the property that $a_{\mathrm{der}} \cdot z \in S_\phi$, where $a_{\mathrm{der}} \in \widehat{G}^*$ is the image of a. One checks that the map

$$S_\phi^+ \oplus_{Z(\widehat{\bar{G}^*})^+} Z(\widehat{G}^*_{\mathrm{sc}}) \to S_\phi^{\mathrm{sc}}, \qquad ((a, z), x) \mapsto ax$$

is an isomorphism of groups. If $\rho \in \mathrm{Irr}(\pi_0(S_\phi^+), [z])$, then the representation $\rho \otimes [z_{\mathrm{sc}}]$ of $S_\phi^+ \times Z(\widehat{G}_{\mathrm{sc}}^*)$ descends to the quotient $S_\phi^+ \oplus_{Z(\widehat{G}^*)^+} Z(\widehat{G}_{\mathrm{sc}}^*)$ and via the above isomorphism becomes a representation of S_ϕ^{sc}. This gives a bijection

$$\mathrm{Irr}(\pi_0(S_\phi^+), [z]) \to \mathrm{Irr}(\pi_0(S_\phi^{\mathrm{sc}}), \Xi). \tag{12}$$

5 The Automorphic Multiplicity Formula

In Sect. 1.5 we discussed that the internal structure of L-packets is a central ingredient in the multiplicity formula for discrete automorphic representations of quasi-split connected reductive groups defined over number fields. In this section we shall formulate the multiplicity formula for general (i.e., not necessarily quasi-split) connected reductive groups, using the conjectural internal structure of tempered L-packets given by Conjecture G. Since we are only considering tempered L-packets locally, the multiplicity formula will be limited to the everywhere tempered automorphic representations. This restriction is just cosmetic—one can incorporate non-tempered automorphic representations by replacing local L-packets with local Arthur packets in the same way as is done in the quasi-split case.

When one attempts to use the local results of the previous sections to study automorphic representations, one realizes that the local cohomological constructions are by themselves not sufficient. They need to be supplemented by a parallel global cohomological construction that ensures that the local cohomological data at the different places of the global field behave coherently. We shall thus begin this section with a short overview of the necessary results. We will then state the multiplicity formula, beginning first with the case of groups that satisfy the Hasse principle, for which the notation simplifies and the key constructions become more transparent, and treating the general case afterwards.

5.1 The Global Gerbe and Its Cohomology

Let F be a number field, $\overline{F}$ a fixed algebraic closure, and $\Gamma = \mathrm{Gal}(\overline{F}/F)$. For each place v of F let F_v denote the completion, $\overline{F_v}$ a fixed algebraic closure, and $\Gamma_v = \mathrm{Gal}(\overline{F_v}/F_v)$. Fixing an embedding $\overline{F} \to \overline{F_v}$ over F (which we think of as a place $\dot{v}$ of $\overline{F}$ over v) provides a closed embedding $\Gamma_v \to \Gamma$, whose image we call $\Gamma_{\dot{v}}$.

It is shown in [KalGR] that there exists a set of places $\dot{V}$ of $\overline{F}$ lifting the places of F, a pro-finite algebraic group P (depending on $\dot{V}$), and an extension

$$1 \to P \to \mathcal{E} \to \Gamma \to 1$$

with the following properties. For an affine algebraic group G and a finite central subgroup $Z \subset G$, both defined over F, let $H^1(P \to \mathcal{E}, Z \to G) \subset H^1(\mathcal{E}, G)$ be defined analogously to the local set $H^1(u \to W, Z \to G)$ of Sect. 3.1. In fact, let us denote the local set now by $H^1(u_v \to \mathcal{E}_v, Z \to G)$ to emphasize the local field F_v. Then for each $v \in V$ there is a localization map

$$\mathrm{loc}_v : H^1(P \to \mathcal{E}, Z \to G) \to H^1(u_v \to \mathcal{E}_v, Z \to G). \tag{13}$$

This map is functorial in $Z \to G$. Moreover, it is already well defined on the level of 1-cocycles, up to coboundaries of Γ_v valued in Z, that is there is a well-defined map

$$\mathrm{loc}_v : Z^1(P \to \mathcal{E}, Z \to G) \to Z^1(u_v \to \mathcal{E}_v, Z \to G)/B^1(\Gamma_v, Z), \tag{14}$$

that induces (13).

Let now G be connected and reductive. For a fixed $x \in H^1(P \to \mathcal{E}, Z \to G)$, the class $\mathrm{loc}_v(x)$ is trivial for almost all v. Thus we have the total localization map

$$H^1(P \to \mathcal{E}, Z \to G) \to \coprod_v H^1(u_v \to \mathcal{E}_v, Z \to G), \tag{15}$$

where we have used the coproduct sign to denote the subset of the product consisting of tuples almost all of whose entries are trivial. One can show that the kernel of this map coincides with the kernel of the usual total localization map

$$H^1(\Gamma, G) \to \coprod_v H^1(\Gamma_v, G).$$

One can also characterize the image of the total localization map (15). This is based on the duality between $H^1(u_v \to \mathcal{E}_v, Z \to G)$ and $\pi_0(Z(\widehat{\bar{G}})^{+_v})$ from Theorem 6, as well as an analogous global duality [KalGR, §3.7]. Recall that $\bar{G} = G/Z$ and that $Z(\widehat{\bar{G}})^{+_v}$ is the subgroup of $Z(\widehat{\bar{G}})$ consisting of those elements whose image in $Z(\widehat{G})$ is Γ_v-fixed. In the same way we define $Z(\widehat{\bar{G}})^+$, where we now demand that the image in $Z(\widehat{G})$ is Γ-fixed. The obvious inclusions $Z(\widehat{\bar{G}})^+ \to Z(\widehat{\bar{G}})^{+_v}$ lead on the level of characters to the summation map

$$\bigoplus_v \pi_0(Z(\widehat{\bar{G}})^{+_v})^* \to \pi_0(Z(\widehat{\bar{G}})^+)^*.$$

Then the image of (15) is the kernel of the composition

$$\coprod_v H^1(u_v \to \mathcal{E}_v, Z \to G) \to \bigoplus_v \pi_0(Z(\widehat{\bar{G}})^{+_v})^* \to \pi_0(Z(\widehat{\bar{G}})^+)^*. \tag{16}$$

Finally, we remark that when Z is sufficiently large (for example, when it contains $Z(G_{\mathrm{der}})$) then the natural map $H^1(P \to \mathcal{E}, Z \to G) \to H^1(\Gamma, G_{\mathrm{ad}})$ is surjective.

5.2 Global Parameters

It is conjectured [Kot84, §12] that there exists a topological group L_F, called the Langlands group of the global field F, which is an extension of the Weil group W_F by a compact group, such that the irreducible complex n-dimensional representations of L_F parameterize the cuspidal automorphic representations of GL_n/F. For each place v of F there should exist an embedding $L_{F_v} \to L_F$, well defined up to conjugation in L_F. We shall admit the existence of this group in order to have a clean formulation of global parameters. In the case of classical groups the use of L_F can be avoided using Arthur's formal parameters, see [Art13, §1.4].

Let G^* be a quasi-split connected reductive group defined over F and let $\xi : G^* \to G$ be an inner twist. A discrete generic global parameter is a continuous semi-simple L-homomorphism $\phi : L_F \to {}^LG^*$ with bounded projection to $\widehat{G}^*$, whose image is not contained in a proper parabolic subgroup of ${}^LG^*$. Given such ϕ and a place v of F, let ϕ_v be the restriction of ϕ to L_{F_v}, a tempered (but usually not discrete) local parameter. Define the adelic L-packet $\Pi_\phi(G, \xi)$ as

$$\Pi_\phi(G, \xi) = \{\pi = \otimes'_v \pi_v | \pi_v \in \Pi_{\phi_v}(G),\ \pi_v \text{ is unramified for a.a. } v\}$$

where the local L-packet $\Pi_{\phi_v}(G)$ is the one from Conjecture A. Note that we are using ξ to identify $\widehat{G}^*$ with $\widehat{G}$.

The question we want to answer in the following sections is this: Which elements $\pi \in \Pi_\phi(G, \xi)$ are discrete automorphic representations and what is their multiplicity in the discrete spectrum? More precisely, let $\chi : Z(G)(\mathbb{A}) \to \mathbb{C}^\times$ denote the central character of π. The locally compact topological group $G(\mathbb{A})$ is unimodular. We endow $G(\mathbb{A})$ with a Haar measure and the discrete group $G(F)$ with the counting measure and obtain a $G(\mathbb{A})$-invariant measure on the quotient space $G(F) \backslash G(\mathbb{A})$. Denote by $L^2_\chi(G(F) \backslash G(\mathbb{A}))$ the space of those square-integrable functions on the quotient $G(F)\backslash G(\mathbb{A})$ that satisfy $f(zg) = \chi(z)f(g)$ for $z \in Z(G)(\mathbb{A})$. The question we want to answer is this: What is the multiplicity of π as a closed subrepresentation of this space?

The answer to this question will be given in terms of objects that depend on G, ξ, and π. However, the construction of these objects will use the global cohomology set $H^1(P \to \mathcal{E}, Z \to G^*)$. In preparation for this, we define a global rigid inner twist $(\xi, z) : G^* \to G$ to consist of an inner twist $\xi : G^* \to G$ and $z \in Z^1(P \to \mathcal{E}, Z \to G^*)$, where $Z \subset G^*$ is a finite central subgroup defined over F, so that the image of z in $Z^1(\Gamma, G^*_{\mathrm{ad}})$ equals $z_{\mathrm{ad}}(\sigma) = \xi^{-1}\sigma(\xi)$.

5.3 Groups That Satisfy the Hasse Principle

Let G be a connected reductive group defined over F. Recall that G is said to satisfy the Hasse principle if the total localization map

$$H^1(\Gamma, G) \to \coprod_v H^1(\Gamma_v, G)$$

is injective. This is always true if G is semi-simple and either simply connected or adjoint, see [PR94, Theorems 6.6, 6.22]. Other groups that are known to satisfy the Hasse principle are unitary groups and special orthogonal groups. It was shown by Kottwitz [Kot84, §4] that G satisfies the Hasse principle if and only if the restriction map

$$H^1(\Gamma, Z(\widehat{G})) \to \bigoplus_v H^1(\Gamma_v, Z(\widehat{G}))$$

is injective.

We assume now that G satisfies the Hasse principle. Let G^* be the unique quasi-split inner form of G and let $\xi : G^* \to G$ be an inner twist. Let $z_{\mathrm{ad}}(\sigma) \in Z^1(\Gamma, G^*_{\mathrm{ad}})$ be given by $z_{\mathrm{ad}}(\sigma) = \xi^{-1}\sigma(\xi)$. Fix $z \in Z^1(P \to \mathcal{E}, Z(G^*_{\mathrm{der}}) \to G^*)$ lifting z_{ad}. For every place v let $z_v \in Z^1(u_v \to \mathcal{E}_v, Z(G^*_{\mathrm{der}}) \to G^*)$ be the localization of z, well defined up to $B^1(\Gamma_v, Z(G^*_{\mathrm{der}}))$. Then $(\xi, z_v) : G^* \to G$ is a (local) rigid inner twist.

Let $\phi : L_F \to {}^LG^*$ be a discrete generic global parameter. For such ϕ, the centralizer $S_\phi = \mathrm{Cent}(\phi, \widehat{G}^*)$ is finite modulo $Z(\widehat{G}^*)^\Gamma$. For any place v of F we have the tempered local parameter $\phi_v = \phi|_{L_{F_v}}$ and $S_\phi \subset S_{\phi_v}$. Let $\pi \in \Pi_\phi(G, \xi)$ and let χ be its central character. We interpret π_v as an element of $\Pi_{\phi_v}((\xi_v, z_v))$ and obtain from Conjecture G the class function $\langle \pi_v, - \rangle$ on $\pi_0(S^+_{\phi_v})$. Let S^+_ϕ be the preimage in $\widehat{\bar{G}^*}$ of S_ϕ and let $\langle \pi, - \rangle$ be the product over all places v of the pull-back to $\pi_0(S^+_\phi)$ of $\langle \pi_v, - \rangle$. It is a consequence [KalGR, Proposition 4.2] of the description (16) of the image of (15) that this class function descends to the quotient $\pi_0(\bar{S}_\phi) := \pi_0(S^+_\phi / Z(\widehat{\bar{G}^*})^+) = \pi_0(S_\phi / Z(\widehat{G}^*)^\Gamma)$ and is moreover independent of the choice of z. It is the character of a finite-dimensional representation of $\pi_0(\bar{S}_\phi)$.

Conjecture H. *The natural number*

$$\sum_\phi |\pi_0(\bar{S}_\phi)|^{-1} \sum_{x \in \pi_0(\bar{S}_\phi)} \langle \pi, x \rangle,$$

where ϕ runs over the $\widehat{G}$-conjugacy classes of discrete generic global parameters satisfying $\pi_v \in \Pi_{\phi_v}(G)$, is the multiplicity of π in $L^2_\chi(G(F) \setminus G(\mathbb{A}))$.

This conjecture is essentially the one from [Kot84, §12]. The only addition here is that we have explicitly realized the global pairing $\langle \pi, - \rangle$ as a product of normalized local pairings $\langle \pi_v, - \rangle$ with the help of the local and global Galois gerbes, and we have built in the simplifications implied by the Hasse principle.

In order to apply the stable trace formula to the study of this conjecture one needs to have a coherent local normalization of the geometric transfer factors. Let $\mathfrak{e} = (G^\mathfrak{e}, s^\mathfrak{e}, {}^L\eta^\mathfrak{e})$ be a global elliptic extended endoscopic triple. Due to the validity of the Hasse principle for G we may and will assume that $s^\mathfrak{e} \in Z(\widehat{G}^\mathfrak{e})^\Gamma$. To this triple, Kottwitz and Shelstad associate [KS99, §7.3] a canonical adelic transfer factor

$$\Delta_{\mathbb{A}} : G^\mathfrak{e}_{\mathrm{sr}}(\mathbb{A}) \times G_{\mathrm{sr}}(\mathbb{A}) \to \mathbb{C}.$$

Note, however, that the original definition needs a correction, as explained in [KS12]. We assume henceforth that $\Delta_{\mathbb{A}}$ is the corrected global factor corresponding to the local factors Δ' of [KS12, §5.4].

Choose a lift $s^{\dot{\mathfrak{e}}} \in \widehat{\bar{G}^{\mathfrak{e}}}$ of $s^{\mathfrak{e}}$. For each place v of F, $\dot{\mathfrak{e}}_v = (G^{\mathfrak{e}}, s^{\dot{\mathfrak{e}}}, {}^L\eta^{\mathfrak{e}})$ is a refined local extended endoscopic triple and we have the normalized transfer factor $\Delta[\mathfrak{w}, \dot{\mathfrak{e}}_v, \xi_v, z_v]$.

Theorem 11 ([KalGR, Proposition 4.1]). *For $\delta \in G_{sr}(\mathbb{A})$ and $\gamma \in G^{\mathfrak{e}}_{sr}(\mathbb{A})$ one has*

$$\Delta_{\mathbb{A}}(\gamma, \delta) = \prod_v \Delta[\mathfrak{w}, \dot{\mathfrak{e}}_v, \xi_v, z_v](\gamma_v, \delta_v).$$

5.4 General Groups

We shall now explain how to modify Conjecture H and Theorem 11 in the case when G does not satisfy the Hasse principle. In order to handle this case, it is not enough to choose $z \in Z^1(P \to \mathcal{E}, Z(G^*_{\mathrm{der}}) \to G^*)$ lifting z_{ad}. Instead, we consider the inner twist $\xi : G^*_{\mathrm{sc}} \to G_{\mathrm{sc}}$ on the level of the simply connected covers of the derived subgroups. Let $z_{\mathrm{sc}} \in Z^1(P \to \mathcal{E}, Z(G^*_{\mathrm{sc}}) \to G^*_{\mathrm{sc}})$ lift z_{ad}. Let $z_{\mathrm{sc},v} \in Z^1(u_v \to \mathcal{E}_v, Z(G^*_{\mathrm{sc}}) \to G^*_{\mathrm{sc}})$ denote the localization of z_{sc}, well defined up to $B^1(\Gamma_v, Z(G^*_{\mathrm{sc}}))$. Let $z_v \in Z^1(u_v \to \mathcal{E}_v, Z(G^*_{\mathrm{der}}) \to G^*)$ be the image of $z_{\mathrm{sc},v}$.

Let $\phi : L_F \to {}^LG^*$ be a discrete generic global parameter. The group $\bar{S}_\phi = S_\phi/Z(\widehat{G}^*)^\Gamma$ that we used when G satisfied the Hasse principle is now not adequate any more. The reason is that two global parameters ϕ_1 and ϕ_2 are considered equivalent not only when they are $\widehat{G}^*$-conjugate, but when there exists $a \in Z^1(L_F, Z(\widehat{G}^*))$ whose class is everywhere locally trivial, and $g \in \widehat{G}^*$, so that $\phi_2(x) = a(x) \cdot g^{-1}\phi_1(x)g$, see [Kot84, §10]. Then the group of self-equivalences S_ϕ of a global parameter ϕ is defined to consist of those $g \in \widehat{G}^*$ for which $x \mapsto g^{-1}\phi(x)g\phi(x)^{-1}$ takes values in $Z(\widehat{G}^*)$ (then it is a 1-cocycle for formal reasons) and its class is everywhere locally trivial. This group contains not just $Z(\widehat{G}^*)^\Gamma$, but all of $Z(\widehat{G}^*)$, and we set $\bar{S}_\phi = S_\phi/Z(\widehat{G}^*)$.

As before we have for each $\pi \in \Pi_\phi(G, \xi)$ the local representation π_v as an element of $\Pi_{\phi_v}((\xi_v, z_v))$ and hence the class function $\langle \pi_v, - \rangle$ on $\pi_0(S^+_{\phi_v})$. We want to produce from these class functions a class function on $\pi_0(\bar{S}_\phi)$. Let $x \in \bar{S}_\phi$. Choose a lift $x_{\mathrm{sc}} \in \widehat{G}^*_{\mathrm{sc}}$ and let x_{der} be its image in $\widehat{G}^*_{\mathrm{der}}$. For each place v there exists $y_v \in Z(\widehat{G}^*)$ so that $x_{\mathrm{der}}y_v \in S_{\phi_v}$. Write $y_v = y'_v y''_v$ with $y'_v \in Z(\widehat{G}^*_{\mathrm{der}})$ and $y''_v \in Z(\widehat{G}^*)^\circ$ and choose a lift $\dot{y}'_v \in Z(\widehat{G}^*_{\mathrm{sc}})$. Since $\bar{G}^* = G^*/Z(G^*_{\mathrm{der}})$ we have $\widehat{\bar{G}^*} = \widehat{G}^*_{\mathrm{sc}} \times Z(\widehat{G}^*)^\circ$. Then $(x_{\mathrm{sc}}\dot{y}'_v, y''_v) \in S^+_{\phi_v}$. The reason we had to choose z_{sc} is that now the class $[z_{\mathrm{sc},v}] \in H^1(u_v \to \mathcal{E}_v, Z(G^*_{\mathrm{sc}}) \to G^*_{\mathrm{sc}})$ becomes a character of $Z(\widehat{G}^*_{\mathrm{sc}})$, which we can evaluate on $\dot{y}'_v$. It can be shown [KalGR, Proposition 4.2] that the product $\langle \pi, x \rangle = \prod_v \langle [z_{\mathrm{sc},v}], \dot{y}'_v \rangle^{-1} \langle \pi_v, (x_{\mathrm{sc}}\dot{y}'_v, y''_v) \rangle$ is a class function on $\pi_0(\bar{S}_\phi)$ that is independent of the choices of z_{sc}, x_{sc}, $\dot{y}'_v$, and y''_v, and is the character of a finite-dimensional representation. We note here that each individual factor $\langle [z_{\mathrm{sc},v}], \dot{y}'_v\rangle^{-1}\langle \pi_v, (x_{\mathrm{sc}}\dot{y}'_v, y''_v)\rangle$, as a function of x_{sc}, is the character of an irreducible representation of the finite group $\pi_0(S^{\mathrm{sc}}_{\phi_v})$ discussed in Sect. 4.6. In fact, it is precisely the character of $\pi_0(S^{\mathrm{sc}}_{\phi_v})$ that is the image of the character $\langle \pi_v, -\rangle$ under the map (12).

Conjecture I. *The natural number*

$$\sum_{\phi} |\pi_0(\bar{S}_\phi)|^{-1} \sum_{x \in \pi_0(\bar{S}_\phi)} \langle \pi, x\rangle,$$

where ϕ runs over the equivalence classes of discrete generic global parameters satisfying $\pi_v \in \Pi_{\phi_v}(G)$, is the multiplicity of π in $L^2_\chi(G(F) \setminus G(\mathbb{A}))$.

A similar procedure is necessary in order to decompose the canonical adelic transfer factor $\Delta_{\mathbb{A}}$ into a product of normalized local transfer factors. Let $\mathfrak{e} = (G^{\mathfrak{e}}, s^{\mathfrak{e}}, {}^L\eta^{\mathfrak{e}})$ be a global elliptic extended endoscopic triple. Choose a lift $s_{\mathrm{sc}} \in \widehat{G}^*_{\mathrm{sc}}$ of the image of $s^{\mathfrak{e}}$ in $\widehat{G}^*_{\mathrm{ad}}$, and let $s_{\mathrm{der}} \in \widehat{G}^*_{\mathrm{der}}$ be the image of s_{sc}. For each place v there is $y_v \in Z(\widehat{G}^*)$ so that $s_{\mathrm{der}}y_v \in Z(\widehat{G}^{\mathfrak{e}})^{\Gamma_v}$. Here we have identified $\widehat{G}^{\mathfrak{e}}$ as a subgroup of $\widehat{G}^*$ via ${}^L\eta^{\mathfrak{e}}$. Write $y_v = y'_v y''_v$ with $y'_v \in Z(\widehat{G}^*_{\mathrm{der}})$ and $y''_v \in Z(\widehat{G}^*)^\circ$ and choose a lift $\dot{y}'_v \in Z(\widehat{G}^*_{\mathrm{sc}})$. Then $(s_{\mathrm{sc}}\dot{y}'_v, y''_v) \in Z(\widehat{\bar{G}^{\mathfrak{e}}})^{+_v}$, so $\dot{\mathfrak{e}}_v = (\widehat{G}^{\mathfrak{e}}, (s_{\mathrm{sc}}\dot{y}'_v, y''_v), {}^L\eta^{\mathfrak{e}})$ is a refined local extended endoscopic triple.

Theorem 12 ([KalGR, Proposition 4.1]). *For $\delta \in G_{sr}(\mathbb{A})$ and $\gamma \in G^{\mathfrak{e}}_{sr}(\mathbb{A})$ one has*

$$\Delta_{\mathbb{A}}(\gamma, \delta) = \prod_v \langle [z_{sc,v}], \dot{y}'_v\rangle^{-1} \Delta[\mathfrak{w}, \dot{\mathfrak{e}}_v, \xi_v, z_v](\gamma_v, \delta_v).$$

5.5 Known Cases

There are a few cases in which Conjecture H has been established. In [KMSW] this conjecture is verified for pure inner forms of unitary groups. In [Taib] this conjecture has been verified in the following setting. One considers non-quasi-split symplectic and orthogonal groups G for which there exists a finite set S of real places such that at $v \in S$ the real group $G(F_v)$ has discrete series, and for $v \notin S$ the local group $G \times F_v$ is quasi-split. For those groups, Taïbi studies the subspace $L^2_{\mathrm{disc}}(G(F)\backslash G(\mathbb{A}))^{S-\mathrm{alg.reg.}}$ of discrete automorphic representations whose infinitesimal character at each place $v \in S$ is regular algebraic and shows that Conjecture H is valid for this subspace.

References

[ABV92] J. Adams, D. Barbasch, D.A. Vogan Jr., *The Langlands Classification and Irreducible Characters for Real Reductive Groups*. Progress in Mathematics, vol. 104 (Birkhäuser, Boston, MA, 1992)

[Art06] J. Arthur, A note on L-packets. Pure Appl. Math. Q. **2**(1), 199–217 (2006). Special Issue: In honor of John H. Coates. Part 1

[Art13] J. Arthur, *The Endoscopic Classification of Representations: Orthogonal and symplectic groups*. American Mathematical Society Colloquium Publications, vol. 61 (American Mathematical Society, Providence, RI, 2013)

[Art99] J. Arthur, On the transfer of distributions: weighted orbital integrals. Duke Math. J. **99**(2), 209–283 (1999)

[ArtI] J. Arthur, The invariant trace formula. II. Global theory. J. Am. Math. Soc. **1**(3), 501–554 (1988)

[ArtS] J. Arthur, A stable trace formula. II. Global descent. Invent. Math. **143**(1), 157–220 (2001)

[ArtS1] J. Arthur, A stable trace formula. I. General expansions. J. Inst. Math. Jussieu **1**(2), 175–277 (2002)

[ArtU] J. Arthur, Unipotent automorphic representations: conjectures. Astérisque **171–172**, 13–71 (1989). Orbites unipotentes et représentations, II

[At15] H. Atobe, On the uniqueness of generic representations in an L-packet (2015). Preprint, arXiv:1511.08897

[Bor79] A. Borel, Automorphic L-functions, in *Automorphic Forms, Representations and L-functions*. Proceedings of Symposia in Pure Mathematics, Oregon State University, Corvallis, OR, 1977, Part 2. Proceedings of Symposia in Pure Mathematics, vol. XXXIII (American Mathematical Society, Providence, RI, 1979), pp. 27–61

[Bou] A. Bouaziz, Intégrales orbitales sur les groupes de Lie réductifs. Ann. Sci. École Norm. Sup. (4) **27**(5), 573–609 (1994)

[DR09] S. DeBacker, M. Reeder, Depth-zero supercuspidal L-packets and their stability. Ann. Math. (2) **169**(3), 795–901 (2009)

[Far] L. Fargues, Geometrization of the local langlands correspondence: an overview (2016). arXiv:1602.00999

[FFC] L. Fargues, J.-M. Fontaine, Courbes et fibrés vectoriels en théorie de Hodge p-adique (2011). Preprint

[Hen00] G. Henniart, Une preuve simple des conjectures de Langlands pour $\mathrm{GL}(n)$ sur un corps p-adique. Invent. Math. **139**(2), 439–455 (2000)

[HT01] M. Harris, R. Taylor, *The Geometry and Cohomology of Some Simple Shimura Varieties*. Annals of Mathematics Studies, vol. 151 (Princeton University Press, Princeton, NJ, 2001). With an appendix by Vladimir G. Berkovich

[KalEC] T. Kaletha, Endoscopic character identities for depth-zero supercuspidal L-packets. Duke Math. J. **158**(2), 161–224 (2011).

[KalEp] T. Kaletha, Epipelagic L-packets and rectifying characters. Invent. Math. **202**(1), 1–89 (2015)

[KalGe] T. Kaletha, Genericity and contragredience in the local Langlands correspondence. Algebra Number Theory **7**(10), 2447–2474 (2013)

[KalGR] T. Kaletha, Global rigid inner forms and multiplicities of discrete automorphic representations (2015). arXiv:1501.01667

[KalI] T. Kaletha, Supercuspidal L-packets via isocrystals. Am. J. Math. **136**(1), 203–239 (2014)

[KalR] T. Kaletha, Rigid inner forms of real and p-adic groups. Ann. Math. **184**(2), 559–632 (2016)

[KalRI] T. Kaletha, Rigid inner forms vs isocrystals. J. Eur. Math. Soc. (to appear)

[KL87] D. Kazhdan, G. Lusztig, Proof of the Deligne-Langlands conjecture for Hecke algebras. Invent. Math. **87**(1), 153–215 (1987)

[KMSW] T. Kaletha, A. Minguez, S.W. Shin, P.-J. White, Endoscopic classification of representations: inner forms of unitary groups (2014). arXiv:1409.3731

[Kot] R.E. Kottwitz, $B(G)$ for all local and global fields (2014). arXiv:1401.5728
[Kot83] R.E. Kottwitz, Sign changes in harmonic analysis on reductive groups. Trans. Am. Math. Soc. **278**(1), 289–297 (1983)
[Kot84] R.E. Kottwitz, Stable trace formula: cuspidal tempered terms. Duke Math. J. **51**(3), 611–650 (1984)
[Kot85] R.E. Kottwitz, Isocrystals with additional structure. Compos. Math. **56**(2), 201–220 (1985)
[Kot86] R.E. Kottwitz, Stable trace formula: elliptic singular terms. Math. Ann. **275**(3), 365–399 (1986)
[Kot88] R.E. Kottwitz, Tamagawa numbers. Ann. Math. (2) **127**(3), 629–646 (1988)
[Kot97] R.E. Kottwitz, Isocrystals with additional structure. II. Compos. Math. **109**(3), 255–339 (1997)
[KS12] R.E. Kottwitz, D. Shelstad, On splitting invariants and sign conventions in endoscopic transfer (2012). arXiv:1201.5658
[KS99] R.E. Kottwitz, D. Shelstad, Foundations of twisted endoscopy. Astérisque **255**, vi+190 (1999)
[LanC] R.P. Langlands, Automorphic representations, Shimura varieties, and motives. Ein Märchen, in *Automorphic Forms, Representations and L-functions*. Proceedings of Symposia in Pure Mathematics, Oregon State University, Corvallis, OR, 1977, Part 2. Proceedings of Symposia in Pure Mathematics, vol. XXXIII (American Mathematical Society, Providence, RI, 1979), pp. 205–246
[Lan79] R.P. Langlands, Stable conjugacy: definitions and lemmas. Can. J. Math. **31**(4), 700–725 (1979)
[Lan83] R.P. Langlands, Les débuts d'une formule des traces stable, in *Publications Mathématiques de l'Université Paris VII* (Mathematical Publications of the University of Paris VII), vol. 13 (Université de Paris VII, U.E.R. de Mathématiques, Paris, 1983)
[Lan97] R.P. Langlands, Representations of abelian algebraic groups. Pac. J. Math. (Special Issue) 231–250 (1997). Olga Taussky-Todd: in memoriam
[LL79] J.-P. Labesse, R.P. Langlands, L-indistinguishability for SL(2). Can. J. Math. **31**(4), 726–785 (1979)
[LR87] R.P. Langlands, M. Rapoport, Shimuravarietäten und Gerben. J. Reine Angew. Math. **378**, 113–220 (1987)
[LS87] R.P. Langlands, D. Shelstad, On the definition of transfer factors. Math. Ann. **278**(1–4), 219–271 (1987)
[LS90] R. Langlands, D. Shelstad, Descent for transfer factors, in *The Grothendieck Festschrift*, vol. II. Progress in Mathematics, vol. 87 (Birkhäuser, Boston, MA, 1990), pp. 485–563
[Lus83] G. Lusztig, Some examples of square integrable representations of semisimple p-adic groups. Trans. Am. Math. Soc. **277**(2), 623–653 (1983)
[Lus95] G. Lusztig, Classification of unipotent representations of simple p-adic groups. Int. Math. Res. Not. **1995**(11), 517–589 (1995)
[Mok] C.P. Mok, Endoscopic classification of representations of quasi-split unitary groups. Mem. Am. Math. Soc. **235**(1108) (2015)
[Ngo10] B.C. Ngô, Le lemme fondamental pour les algèbres de Lie. Publ. Math. Inst. Hautes Études Sci. **111**, 1–169 (2010)
[PR94] V. Platonov, A. Rapinchuk, *Algebraic Groups and Number Theory*. Pure and Applied Mathematics, vol. 139 (Academic, Boston, MA, 1994). Translated from the 1991 Russian original by Rachel Rowen
[Rap95] M. Rapoport, Non-Archimedean period domains, in *Proceedings of the International Congress of Mathematicians, vol. 1, 2 (Zürich, 1994)* (Birkhäuser, Basel, 1995), pp. 423–434
[RY14] M. Reeder, J.-K. Yu, Epipelagic representations and invariant theory. J. Am. Math. Soc. **27**(2), 437–477 (2014)
[Sha90] F. Shahidi, A proof of Langlands' conjecture on Plancherel measures; complementary series for p-adic groups. Ann. Math. (2) **132**(2), 273–330 (1990)
[She81] D. Shelstad, Embeddings of L-groups. Can. J. Math. **33**(3), 513–558 (1981)

[She82] D. Shelstad, *L*-indistinguishability for real groups. Math. Ann. **259**(3), 385–430 (1982)
[She83] D. Shelstad, Orbital integrals, endoscopic groups and *L*-indistinguishability for real groups, in *Conference on Automorphic Theory (Dijon, 1981)*. Publications Mathématiques de l'Université Paris VII, vol. 15 (University of Paris VII, Paris, 1983), pp. 135–219
[SheC] D. Shelstad, Notes on *L*-indistinguishability (based on a lecture of R. P. Langlands), in *Automorphic Forms, Representations and L-Functions*. Proceedings of Symposia in Pure Mathematics, Oregon State University, Corvallis, OR, 1977, Part 2. Proceedings of Symposia in Pure Mathematics, vol. XXXIII (American Mathematical Society, Providence, RI, 1979), pp. 193–203
[SheT1] D. Shelstad, Tempered endoscopy for real groups. I. Geometric transfer with canonical factors, in *Representation Theory of Real Reductive Lie Groups*. Contemporary Mathematics, vol. 472 (American Mathematical Society, Providence, RI, 2008), pp. 215–246
[SheT2] D. Shelstad, Tempered endoscopy for real groups. II. Spectral transfer factors, in *Automorphic Forms and the Langlands Program*. Advanced Lectures in Mathematics, vol. 9 (International Press, Somerville, MA, 2010), pp. 236–276
[SheT3] D. Shelstad, Tempered endoscopy for real groups. III. Inversion of transfer and *L*-packet structure. Representation Theory **12**, 369–402 (2008)
[Taib] O. Taïbi, Arthur's multiplicity formula for certain inner forms of special orthogonal and symplectic groups. J. Eur. Math. Soc. (2015), to appear
[Tat] J. Tate, Number theoretic background, in *Automorphic Forms, Representations and L-functions*. Proceedings of Symposia in Pure Mathematics, Oregon State University, Corvallis, OR, 1977. Part 2. Proceedings of Symposia in Pure Mathematics, XXXIII (American Mathematical Society, Providence, RI, 1979), pp. 3–26
[Vog93] D.A. Vogan Jr., The local Langlands conjecture, in *Representation Theory of Groups and Algebras*. Contemporary Mathematics, vol. 145 (American Mathematical Society, Providence, RI, 1993), pp. 305–379
[Wal06] J.-L. Waldspurger, Endoscopie et changement de caractéristique. J. Inst. Math. Jussieu **5**(3), 423–525 (2006)
[Wal09] J.-L. Waldspurger, Endoscopie et changement de caractéristique: intégrales orbitales pondérées. Ann. Inst. Fourier (Grenoble) **59**(5), 1753–1818 (2009)
[Wal97] J.-L. Waldspurger, Le lemme fondamental implique le transfert. Compos. Math. **105**(2), 153–236 (1997)

Asymptotics and Local Constancy of Characters of p-adic Groups

Ju-Lee Kim, Sug Woo Shin, and Nicolas Templier

Abstract In this paper we study quantitative aspects of trace characters Θ_π of reductive p-adic groups when the representation π varies. Our approach is based on the local constancy of characters and we survey some other related results. We formulate a conjecture on the behavior of Θ_π relative to the formal degree of π, which we are able to prove in the case where π is a tame supercuspidal. The proof builds on J.-K. Yu's construction and the structure of Moy–Prasad subgroups.

1 Introduction

For an admissible representation π of a p-adic reductive group G, its trace character distribution is defined by

$$\langle \Theta_\pi, f \rangle = \operatorname{tr} \pi(f), \quad f \in \mathcal{C}_c(G).$$

Harish-Chandra showed that it is represented by a locally integrable function on G still denoted by Θ_π, which moreover is locally constant on the open subset of regular elements.

Our goal in this paper is to initiate a quantitative theory of trace characters Θ_π when the representation π varies. One motivation is towards a better understanding of the spectral side of the trace formula where one would like to control the global

J.-L. Kim
Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
e-mail: julee@math.mit.edu

S.W. Shin (✉)
Department of Mathematics, UC Berkeley, Berkeley, CA 94720, USA

Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul 130-722, Republic of Korea
e-mail: sug.woo.shin@berkeley.edu

N. Templier
Department of Mathematics, Cornell University, Ithaca, NY 14853, USA
e-mail: templier@math.cornell.edu

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_7

behavior of characters [KST]. Another motivation comes from the Weyl character formula. For a finite dimensional representation σ of a compact Lie group and a regular element γ,

$$D(\gamma)^{\frac{1}{2}}|\mathrm{tr}\,\sigma(\gamma)| \leq |W|,$$

where W is the Weyl group and $D(\gamma)$ is the Weyl discriminant which appears in the denominator of the character formula. More generally the Harish-Chandra formula for characters of discrete series yields similar estimates for real reductive groups, see Sect. 5.2 below.

If π is a square-integrable representation of G, we denote by $\deg(\pi)$ its formal degree. Let γ be a fixed regular semisimple element. The central conjecture we would like to propose in this paper (Conjecture 4.1) is essentially that $\frac{\Theta_\pi(\gamma)}{\deg(\pi)}$ converges to zero as $\deg(\pi)$ grows.

It is nowadays possible to study such a question thanks to recent progress in constructing supercuspidal representations and computing their trace characters, see notably [ADSS11, AS09] and the references there.

The main result of this paper (Theorems 4.2 and 4.18, with the latter improved as in Sect. 4.6 below) verifies our conjecture for the tame supercuspidal representations π constructed by J.-K. Yu for topologically unipotent elements γ when the residual characteristic of the base field is large enough (in an effective manner). In such a setup we establish that for some constants $A, \kappa > 0$ depending on the group G,

$$\frac{D(\gamma)^A|\Theta_\pi(\gamma)|}{\deg(\pi)^{1-\kappa}} \tag{1}$$

is bounded both as a function of γ topologically unipotent and as π varies over the set of irreducible supercuspidal representations of G.

Yu's construction gives tame supercuspidal representations $\pi = \text{c-ind}_J^G \rho$ as compactly induced from an explicit open compact-modulo-center subgroup J given in terms of a sequence of tamely ramified twisted Levi subgroups (whose definition is recalled in Sect. 2.3 below). The main theorem of Yu [Yu01] is that the induction is irreducible, and therefore is supercuspidal. This may be summarized by the inclusions,

$$\mathrm{Irr}^{\mathrm{Yu}}(G) \subset \mathrm{Irr}^{\text{c-ind}}(G) \subset \mathrm{Irr}^{\mathrm{sc}}(G),$$

where $\mathrm{Irr}^{\mathrm{sc}}(G)$ consists of all irreducible supercuspidal representations (up to isomorphism), and the first two subsets are given by Yu's construction and by compact induction from open compact-modulo-center subgroups, respectively. The formal degree $\deg(\pi)$ is proportional to $\dim(\rho)/\mathrm{vol}(J)$. Moreover the first-named author [Kim07] has shown that if the residue characteristic is large enough, then Yu's construction exhausts all supercuspidals, i.e. the above inclusions are equalities. This means that our result (1) is true for *all* supercuspidal representations in that case.

One important ingredient in proving our main result is using the local constancy of characters. For a given regular semisimple element γ, if Θ_π is constant on γK for a (small) open compact subgroup K of G, then

$$\Theta_\pi(\gamma) = \frac{1}{\mathrm{vol}(K)}\langle \Theta_\pi, 1_{\gamma K}\rangle = \mathrm{trace}(\pi(\gamma)|V_\pi^K) \tag{2}$$

where $1_{\gamma K}$ is the characteristic function of γK. The results of Adler and Korman [AK07] and Meyer and Solleveld [MS12] determine the size of K, which depends on the (Moy–Prasad) depth of π and the singular depth of γ (see Definition 3.3 below). For our main result, as we vary π such that the formal degree of π increases (equivalently, the depth of π increases), we choose K appropriately to be able to approximate the size of $\Theta_\pi(\gamma)$. Write G_x for the parahoric subgroup of G associated with x. The fact that $\pi = \text{c-ind}_J^G \rho$, via Mackey's formula, allows us to bound $|\Theta_\pi(\gamma)|/\deg(\pi)$ in terms of the number of fixed points of γ (which may be assumed to lie in G_x) acting on $(G_x \cap gJg^{-1})\backslash G_x$ by right translation for various $g \in G$. To bound the cardinality of the fixed points we prove quite a few numerical inequalities as Yu's data vary by a systematic study of Moy–Prasad subgroups in Yu's construction.

The celebrated regularity theorem of Harish-Chandra [HC70] says that $D(\gamma)^{\frac{1}{2}}\Theta_\pi(\gamma)$ is locally bounded as a function of γ and similarly for any G-invariant admissible distribution. It implies that Θ_π is given by a locally integrable function on G and moreover there is a germ expansion [HC99] when γ approaches a non-regular element. In comparison our result concerning (1) is much less precise but at the same time we also allow π to vary.

The local constancy (2) is used similarly in [KL13a, KL13b] to compute the characters of unipotent representations at very regular elements. In such situation the depth of π is sufficiently larger than the singular depth of γ, and the size of K is determined by the depth of π. Another application of the local constancy of trace characters is [MS12] which considers trace characters of representations π in positive characteristic different from p. Among other results they show that the trace character Θ_π exists as a function essentially as a consequence of the formula (2).

1.1 Notation and Conventions

Let p be a prime. Let k be a finite extension of $\mathbb{Q}_p$. Denote by q the cardinality of the residue field of k. For any tamely ramified finite extension E of k, let ν denote the valuation on E which coincides with the valuation of $\mathbb{Q}_p$ when restricted. Let $\mathcal{O}_E$ and $\mathfrak{p}_E$ be the ring of integers in E and the prime ideal of $\mathcal{O}_E$ respectively. We fix an additive character Ω_k of k, with conductor $\mathfrak{p}_k$.

Let $\mathbf{G}$ be a connected reductive group over k, whose Lie algebra is denoted $\mathbf{g}$. Let r_G be the difference between the absolute rank of $\mathbf{G}$ (the dimension of any maximal torus in $\mathbf{G}$) and the dimension of the center $\mathbf{Z_G}$ of G. Write G and $\mathfrak{g}$ for $\mathbf{G}(k)$ and $\mathbf{g}(k)$, respectively. The linear dual of $\mathfrak{g}$ is denoted by $\mathfrak{g}^*$. Denote the set of regular semisimple elements in G by G_{reg}.

Throughout the paper, by a unipotent subgroup, we mean the unipotent subgroup given by the unipotent radical of a parabolic subgroup.

For a subset S of a group H and an element $g \in H$, we write S^g or ${}^{g^{-1}}S$ for $g^{-1}Sg$. Similarly if $g, h \in H$, we write h^g or ${}^{g^{-1}}h$ for $g^{-1}hg$. If S is a subgroup of H and ξ is a representation of S, denote by ξ^g or ${}^{g^{-1}}\xi$ the representation of $S^g = {}^{g^{-1}}S$ given by $\xi^g(s) = {}^{g^{-1}}\xi(s) = \xi(gsg^{-1})$, $s \in S$.

2 Minimal K-types and Yu's Construction of Supercuspidal Representations

In this section we review the construction of supercuspidal representations of a p-adic reductive group from the so-called generic data due to Jiu-Kang Yu and recall a result by the first author that his construction exhausts all supercuspidal representations provided the residue characteristic of the base field is sufficiently large. The construction yields a supercuspidal representation concretely as a compactly induced representation, and this will be an important input in the next section.

2.1 *Moy–Prasad Filtrations*

For a tamely ramified extension E of k, denote by $\mathcal{B}(\mathbf{G}, E)$ (resp. $\mathcal{B}^{\mathrm{red}}(G)$) be the extended (resp. reduced) building of $\mathbf{G}$ over E. When $E = k$, we write $\mathcal{B}(G)$ (resp. $\mathcal{B}^{\mathrm{red}}(G)$) for $\mathcal{B}(\mathbf{G}, k)$ (resp. $\mathcal{B}^{\mathrm{red}}(\mathbf{G}, k)$) for simplicity. If $\mathbf{T}$ is a maximal E-split k-torus, let $\mathcal{A}(\mathbf{T}, \mathbf{G}, E)$ denote the apartment associated with $\mathbf{T}$ in $\mathcal{B}(\mathbf{G}, E)$. When $E = k$, write $\mathcal{A}(T)$ for the same apartment. It is known that for any tamely ramified Galois extension E' of E, $\mathcal{A}(\mathbf{T}, \mathbf{G}, E)$ can be identified with the set of all $\mathrm{Gal}(E'/E)$-fixed points in $\mathcal{A}(\mathbf{T}, \mathbf{G}, E')$. Likewise, $\mathcal{B}(\mathbf{G}, E)$ can be embedded into $\mathcal{B}(\mathbf{G}, E')$ and its image is equal to the set of the Galois fixed points in $\mathcal{B}(\mathbf{G}, E')$ [Rou77, Pra01].

For $(x, r) \in \mathcal{B}(\mathbf{G}, E) \times \mathbb{R}$, there is a filtration lattice $\mathfrak{g}(E)_{x,r}$ and a subgroup $\mathbf{G}(E)_{x,r}$ if $r \geq 0$ defined by Moy and Prasad [MP94]. We assume that the valuation is normalized such that for a tamely ramified Galois extension E' of E and $x \in \mathcal{B}(\mathbf{G}, E) \subset \mathcal{B}(\mathbf{G}, E')$, we have

$$\mathfrak{g}(E)_{x,r} = \mathfrak{g}(E')_{x,r} \cap \mathfrak{g}(E).$$

If $r > 0$, we also have

$$\mathbf{G}(E)_{x,r} = \mathbf{G}(E')_{x,r} \cap \mathbf{G}(E).$$

For simplicity, we put $\mathfrak{g}_{x,r} := \mathfrak{g}(k)_{x,r}$ and $G_{x,r} := \mathbf{G}(k)_{x,r}$, etc. We will also use the following notation. Let $r \in \mathbb{R}$ and $x \in \mathcal{B}(G)$:

1. $\mathfrak{g}_{x,r^+} := \cup_{s>r}\mathfrak{g}_{x,s}$, and if $r \geq 0$, $G_{x,r^+} := \cup_{s>r}G_{x,s}$;
2. $\mathfrak{g}^*_{x,r} := \{\chi \in \mathfrak{g}^* \mid \chi(\mathfrak{g}_{x,(-r)^+}) \subset \mathfrak{p}_k\}$;
3. $\mathfrak{g}_r := \cup_{y\in\mathcal{B}(G)}\mathfrak{g}_{y,r}$ and $\mathfrak{g}_{r^+} := \cup_{s>r}\mathfrak{g}_s$;
4. $G_r := \cup_{y\in\mathcal{B}(G)}G_{y,r}$ and $G_{r^+} := \cup_{s>r}G_s$ for $r \geq 0$.
5. For any facet $F \subset \mathcal{B}(G)$, let $G_F := G_{x,0}$ for some $x \in F$. Let $[F]$ be the image of F in $\mathcal{B}^{\mathrm{red}}(G)$. Then, let $G_{[F]}$ denote the stabilizer of $[F]$ in G. Note that $G_F \subset G_{[F]}$. Similarly, $G_{[x]}$ is the stabilizer of $[x] \in \mathcal{B}^{\mathrm{red}}(G)$ in G. However, G_x will denote $G_{x,0}$, the parahoric subgroup associated with x.

2.2 *Unrefined Minimal* K*-types and Good Cosets*

For simplicity, as in [MP94], we assume that there is a natural isomorphism ι : $G_{x,r}/G_{x,r^+} \longrightarrow \mathfrak{g}_{x,r}/\mathfrak{g}_{x,r^+}$ when $r > 0$. By Yu [Yu01, (2.4)], such an isomorphism exists whenever $\mathbf{G}$ splits over a tamely ramified extension of k (see also [Adl98, §1.6]).

Definition 2.1. An *unrefined minimal* K*-type* (or *minimal* K*-type*) is a pair $(G_{x,\varrho}, \chi)$, where $x \in \mathcal{B}(G)$, ϱ is a nonnegative real number, χ is a representation of $G_{x,\varrho}$ trivial on G_{x,ϱ^+} and

(i) if $\varrho = 0$, χ is an irreducible cuspidal representation of $G_x/G_{x,0^+}$ inflated to G_x,
(ii) if $\varrho > 0$, then χ is a nondegenerate character of $G_{x,\varrho}/G_{x,\varrho^+}$.

The ϱ in the above definition is called the *depth* of the minimal K-type $(G_{x,\varrho}, \chi)$. Recall that a coset $X+\mathfrak{g}^*_{x,(-\varrho)^+}$ in $\mathfrak{g}^*$ is nondegenerate if $X+\mathfrak{g}^*_{x,(-\varrho)^+}$ does not contain any nilpotent element. If a character χ of $G_{x,\varrho}$ is *represented* by $X + \mathfrak{g}^*_{x,(-\varrho)^+}$, i.e. $\chi(g) = \Omega_k(X'(\iota(g)))$ with $X' \in X + \mathfrak{g}^*_{x,(-\varrho)^+}$, a character χ of $G_{x,\varrho}$ is *nondegenerate* if $X + \mathfrak{g}^*_{x,(-\varrho)^+}$ is nondegenerate.

Definition 2.2. Two minimal K-types $(G_{x,\varrho}, \chi)$ and $(G_{x',\varrho'}, \chi')$ are said to be *associates* if they have the same depth $\varrho = \varrho'$, and

(i) if $\varrho = 0$, there exists $g \in G$ such that $G_x \cap G_{gx'}$ surjects onto both $G_x/G_{x,0^+}$ and $G_{gx'}/G_{gx',0^+}$, and χ is isomorphic to ${}^g\chi'$,
(ii) if $\varrho > 0$, the G-orbit of the coset which realizes χ intersects the coset which realizes χ'.

We also recall the definition of good cosets. In Sect. 3, we will prove some facts concerning good K-types. The following is a minor modification of the definition in [AK00] (see also [KM03, §2.4]).

Definition 2.3. (i) Let $\mathbf{T} \subset \mathbf{G}$ be a maximal k-torus which splits over a tamely ramified extension E of k. Let $\Phi(\mathbf{T}, E)$ be the set of E-roots of $\mathbf{T}$. Then, $X \in \mathfrak{t}$ is a *good element of depth r* if $X \in \mathfrak{t}_r \backslash \mathfrak{t}_{r^+}$ and for any $\alpha \in \Phi(\mathbf{T}, E)$, $\nu(d\alpha(X)) = r$ or ∞.

(ii) Let $r < 0$ and $x \in \mathcal{B}(G)$. A coset $\mathcal{S}$ in $\mathfrak{g}_{x,r}/\mathfrak{g}_{x,r^+}$ is *good* if there is a good element $X \in \mathfrak{g}$ of depth r such that $\mathcal{S} = X + \mathfrak{g}_{x,r^+}$ and $x \in \mathcal{B}(\mathbf{C_G}(X), k)$.

(iii) A minimal K-type $(G_{x,\varrho}, \chi)$ with $\varrho > 0$ is *good* if the associated dual coset is good.

2.3 Generic G-datum

Yu's construction of supercuspidal representations starts with a *generic G-datum*, which consists of five components. Recall $\mathbf{G}' \subset \mathbf{G}$ is a *tamely ramified twisted Levi subgroup* if $\mathbf{G}'(E)$ is a Levi subgroup of $\mathbf{G}(E)$ for a tamely ramified extension E of k.

Definition 2.4. A *generic G-datum* is a quintuple $\Sigma = (\vec{\mathbf{G}}, x, \vec{r}, \vec{\phi}, \rho)$ satisfying the following:

D1. $\vec{\mathbf{G}} = (\mathbf{G}^0 \subsetneq \mathbf{G}^1 \subsetneq \cdots \subsetneq \mathbf{G}^d = \mathbf{G})$ is a tamely ramified twisted Levi sequence such that $\mathbf{Z}_{\mathbf{G}^0}/\mathbf{Z}_{\mathbf{G}}$ is anisotropic.

D2. $x \in \mathcal{B}(\mathbf{G}^0, k)$.

D3. $\vec{r} = (r_0, r_1, \cdots, r_{d-1}, r_d)$ is a sequence of positive real numbers with $0 < r_0 < \cdots < r_{d-2} < r_{d-1} \le r_d$ if $d > 0$, and $0 \le r_0$ if $d = 0$.

D4. $\vec{\phi} = (\phi_0, \cdots, \phi_d)$ is a sequence of quasi-characters, where ϕ_i is a generic quasi-character of G^i (see [Yu01, §9] for the definition of generic quasi-characters); ϕ_i is trivial on G^i_{x,r_i^+}, but non-trivial on G^i_{x,r_i} for $0 \le i \le d-1$. If $r_{d-1} < r_d$, then ϕ_d is trivial on G^d_{x,r_d^+} and nontrivial on G^d_{x,r_d}, and otherwise if $r_{d-1} = r_d$, then $\phi_d = 1$.

D5. ρ is an irreducible representation of $G^0_{[x]}$, the stabilizer in G^0 of the image $[x]$ of x in the reduced building of $\mathbf{G}^0$, such that $\rho|G^0_{x,0+}$ is isotrivial and $c\text{-Ind}_{G^0_{[x]}}^{G^0} \rho$ is irreducible and supercuspidal.

Remark 2.5. (i) By (6.6) and (6.8) of [MP96], **D**5 is equivalent to the condition that G^0_x is a maximal parahoric subgroup in G^0 and $\rho|G^0_x$ induces a cuspidal representation of $G^0_x/G^0_{x,0+}$.

(ii) Recall from [Yu01] that there is a canonical sequence of embeddings $\mathcal{B}(\mathbf{G}^0, k) \hookrightarrow \mathcal{B}(\mathbf{G}^1, k) \hookrightarrow \cdots \hookrightarrow \mathcal{B}(\mathbf{G}^d, k)$. Hence, x can be regarded as a point of each $\mathcal{B}(\mathbf{G}^i, k)$.

(iii) There is a finite number of pairs $(\vec{\mathbf{G}}, x)$ up to G-conjugacy, which arise in a generic G-datum: By §1.2 in [KY11], there are finitely many choices for $\vec{\mathbf{G}}$ up to G-conjugacy. In particular, there are finitely many choices for $\mathbf{G}^0$, and for each $\mathbf{G}^0$ the number of vertices in $\mathcal{B}(G^0)$ is finite up to G^0-conjugacy.

2.4 Construction of J_Σ

Let $\Sigma = (\vec{\mathbf{G}}, x, \vec{r}, \vec{\phi}, \rho)$ be a generic G-datum. Set $s_i := r_i/2$ for each i. Associated with $\vec{\mathbf{G}}$, x and $\vec{r}$, we define the following open compact subgroups.

1. $K^0 := G^0_{[x]}$; $K^0_+ := G^0_{x,0+}$.
2. $K^i := G^0_{[x]} G^1_{x,s_0} \cdots G^i_{x,s_{i-1}}$; $K^i_+ := G^0_{x,0+} G^1_{x,s_0^+} \cdots G^i_{x,s_{i-1}^+}$ for $1 \le i \le d$.
3. $J^i := (\mathbf{G}^{i-1}, \mathbf{G}^i)(k)_{x,(r_{i-1},s_{i-1})}$; $J^i_+ := (\mathbf{G}^{i-1}, \mathbf{G}^i)(k)_{x,(r_{i-1},s_{i-1}^+)}$ in the notation of Yu [Yu01, §1].

For $i > 0$, J^i is a normal subgroup of K^i and we have $K^{i-1}J^i = K^i$ (semi-direct product). Similarly J^i_+ is a normal subgroup of K^i_+ and $K^{i-1}_+ J^i_+ = K^i_+$. Finally let $J_\Sigma := K^d$ and $J_+ := K^d_+$, and also $s_\Sigma := s_{d-1}$ and $r_\Sigma := r_{d-1}$. When there is no confusion, we will drop the subscript Σ and simply write J, r, s, etc.

2.5 Construction of ρ_Σ

One can define the character $\hat{\phi}_i$ of $K^0 G^i_x G_{x,s_i^+}$ extending ϕ_i of $K^0 G^i_x \subset G^i$. For $0 \le i < d$, there exists by the Stone–von Neumann theorem a representation $\tilde{\phi}_i$ of $K^i \ltimes J^{i+1}$ such that $\tilde{\phi}_i | J^{i+1}$ is $\hat{\phi}_i | J^{i+1}_+$-isotypical and $\tilde{\phi}_i | K^i_+$ is isotrivial.

Let $\inf(\phi_i)$ denote the inflation of $\phi_i | K^i$ to $K^i \ltimes J^{i+1}$. Then $\inf(\phi_i) \otimes \tilde{\phi}_i$ factors through a map

$$K^i \ltimes J^{i+1} \longrightarrow K^i J^{i+1} = K^{i+1}.$$

Let κ_i denote the corresponding representation of K^{i+1}. Then it can be extended trivially to K^d, and we denote the extended representation again by κ_i (in fact κ_i could be further extended to the semi-direct product $K^{i+1} G_{x,s_{i+1}} \supset K^d$ by making it trivial on $G_{x,s_{i+1}}$). Similarly we extend ρ from $G^0_{[x]}$ to a representation of K^d and denote this extended representation again by ρ. Define a representation κ and ρ_Σ of K^d as follows:

$$\begin{aligned} \kappa \ \ &:= \kappa_0 \otimes \cdots \otimes \kappa_{d-1} \otimes (\phi_d | K^d), \\ \rho_\Sigma &:= \rho \otimes \kappa. \end{aligned} \tag{3}$$

Note that κ is defined only from $(\vec{\mathbf{G}}, x, \vec{r}, \vec{\phi})$ independently of ρ.

Remark 2.6. One may construct κ_i as follows: set $J^i_1 := G^i_{x,0+} G_{x,s_i}$ and $J^i_2 := G^i_{x,0+} G_{x,s_i^+}$. Write also $\hat{\phi}_i$ for the restriction of $\hat{\phi}_i$ to J^i_2. Then, one can extend $\hat{\phi}_i$

to J_1^i via Heisenberg representation and to $G^i_{[y]}G_{x,s_i}$ by Weil representation upon fixing a special isomorphism (see [Yu01] for details):

$$\begin{array}{ccc} J_2^i \to & J_1^i & \to G^i_{[y]}G_{x,s_i} \\ \hat{\phi}_i & \rho_{\hat{\phi}_i} & \omega_{\hat{\phi}_i}. \end{array}$$

Note that we have inclusions $J_\Sigma = K^d \subset G^i_{[x]}G_{x,s_i^+} \subset G_{[x]}$, and we have $\kappa_i \simeq \omega_{\hat{\phi}_i}|K^d$.

Theorem 2.7 (Yu). *$\pi_\Sigma = c\text{-Ind}_{J_\Sigma}^G \rho_\Sigma$ is irreducible and thus supercuspidal.*

Remark 2.8. Let Σ be a generic G-datum. If G is semisimple, comparing Moy–Prasad minimal K-types and Yu's constructions, we observe the following:

(i) The depth of π_Σ is given $r_\Sigma = r_d = r_{d-1}$. (Even if G is not semisimple, the depth is r_d, cf. [Yu01, Remark 3.6], but it may not equal r_{d-1}.)
(ii) $(G_{x,r_{d-1}}, \phi_{d-1})$ is a good minimal K-type of π_Σ in the sense of Kim and Murnaghan [KM03].

2.6 Supercuspidal Representations Via Compact Induction

Denote by $\mathrm{Irr}(G)$ the set of (isomorphism classes of) irreducible smooth representations of G. Fix a Haar measure on G. Write $\mathrm{Irr}^2(G)$ (resp. $\mathrm{Irr}^{\mathrm{sc}}(G)$) for the subset of square-integrable (resp. supercuspidal) members. For each $\pi \in \mathrm{Irr}^2(G)$ let $\deg(\pi)$ denote the formal degree of π. For each $\pi \in \mathrm{Irr}(G)$, Θ_π is the Harish-Chandra character, which is in $L^1_{\mathrm{loc}}(G)$ and locally constant on G_{reg}.

Define $\mathrm{Irr}^{\mathrm{Yu}}(G)$ to be the subset of $\mathrm{Irr}^{\mathrm{sc}}(G)$ consisting of all supercuspidal representations which are constructed by Yu, namely of the form π_Σ as above. Write $\mathrm{Irr}^{\text{c-ind}}(G)$ for the set of $\pi \in \mathrm{Irr}^{\mathrm{sc}}(G)$ which are compactly induced, meaning that there exist an open compact-mod-center subgroup $J \subset G$ and an irreducible admissible representation ρ of J such that $\pi \simeq \text{c-ind}_J^G(\rho)$. We have that

$$\mathrm{Irr}^{\mathrm{Yu}}(G) \subset \mathrm{Irr}^{\text{c-ind}}(G) \subset \mathrm{Irr}^{\mathrm{sc}}(G).$$

The first inclusion is a consequence of Yu's theorem (Theorem 2.7) and generally strict. A folklore conjecture asserts that the second inclusion is always an equality. It has been verified through the theory of types for GL_n and SL_n by Bushnell–Kutzko, for inner forms of GL_n by Broussous and Sécherre–Stevens, and for p-adic classical groups by Stevens when $p \neq 2$ [BK93, BK94, Bro98, SS08, Ste08]. In general, according to the main result of Kim [Kim07], there exists a lower bound p_0 depending only on the absolute ramification index of k and the absolute root datum of G, such that both inclusions are equalities if $p \geq p_0$. Precisely this is true for every prime p such that the hypotheses (Hk), (HB), (HGT), and (H$\mathcal{N}$) of [Kim07, §3.4] are satisfied.

2.7 Hypotheses

The above hypotheses will be assumed in a variety of our results in the next two sections. We will clearly state when the hypotheses are needed. As they are too lengthy to copy here, the reader interested in the details is referred to [Kim07, §3.4]. For our purpose it suffices to recall the nature of those hypotheses: (Hk) is about the existence of filtration preserving exponential map, (HB) is to identify $\mathfrak{g}$ and its linear dual $\mathfrak{g}^*$, (HGT) is about the abundance of good elements, and (H$\mathcal{N}$) is regarding nilpotent orbits.

2.8 Formal Degree

Recall that $\deg(\pi)$ denotes the formal degree of π.

Lemma 2.9. *Let Σ be a generic G-datum. Then*

(i) $\deg(\pi_\Sigma) = \dim(\rho_\Sigma)/\mathrm{vol}_{G/Z_G}(J_\Sigma/Z_G)$.

(ii) $\frac{1}{\mathrm{vol}_{G/Z_G}(J_\Sigma/Z_G)} \le \deg(\pi_\Sigma) \le \frac{q^{\dim(\mathfrak{g})}}{\mathrm{vol}_{G/Z_G}(J_\Sigma/Z_G)}$.

Proof. Assertion (i) is easily deduced from the defining equality for $\deg(\pi_\Sigma)$:

$$\deg(\pi_\Sigma)\int_{G/Z_G} \Theta_{\rho_\Sigma}(g)\overline{\Theta_{\rho_\Sigma}(g)}dg = \dim \rho_\Sigma,$$

cf. [BH96, Theorem A.14].

(ii) Let ρ and κ be as in (3). One sees from the construction of supercuspidal representations that $\dim(\rho) \le [G_x^0 : G_{x,0+}^0]$, and the dimension formula for finite Heisenberg representations yields

$$\begin{aligned}\dim(\kappa_i) = [J^{i+1} : J_+^{i+1}]^{\frac{1}{2}} &= [(\mathfrak{g}^i, \mathfrak{g}^{i+1})_{x,(r_i,s_i)} : (\mathfrak{g}^i, \mathfrak{g}^{i+1})_{x,(r_i,s_i^+)}]^{\frac{1}{2}} \\ &\le [\mathfrak{g}^{i+1}(\mathbb{F}_q) : \mathfrak{g}^i(\mathbb{F}_q))]^{\frac{1}{2}}, \\ \dim(\kappa) = \prod_{i=i}^{d-1} \dim(\kappa_i) &\le [\mathfrak{g}(\mathbb{F}_q) : \mathfrak{g}^{\mathbf{0}}(\mathbb{F}_q))]^{\frac{1}{2}}\end{aligned}$$

Hence,

$$1 \le \dim(\rho_\Sigma) = \dim(\rho)\dim(\kappa) \le [G_x^0 : G_{x,0+}^0][\mathfrak{g}(\mathbb{F}_q) : \mathfrak{g}^{\mathbf{0}}(\mathbb{F}_q))]^{\frac{1}{2}} \le q^{\dim(\mathfrak{g})}$$

$\square$

There is no exact formula yet known for the formal degree $\deg(\pi_\Sigma)$ of tame supercuspidals, or equivalently for $\mathrm{vol}_{G/Z_G}(J_\Sigma/Z_G)$ and $\dim(\rho_\Sigma)$, which is also an indication of the difficulty in computing the trace character $\Theta_{\pi_\Sigma}(\gamma)$ in this generality since $\deg(\pi_\Sigma)$ appears as the first term in the local character expansion. In this direction a well-known conjecture of Hiraga–Ichino–Ikeda [HII08] expresses $\deg(\pi_\Sigma)$ in terms of the Langlands parameter conjecturally attached to π_Σ.

For our purpose it is sufficient to know that $\frac{1}{r_\Sigma}\log_q \mathrm{vol}_{G/Z_G}(J_\Sigma/Z_G)$ is bounded above and below as Σ varies by constants depending only on G, which follows from Lemma 2.9. Below we shall use similarly that $\frac{1}{r_\Sigma}\log_q \mathrm{vol}_{G/Z_G}(L_{s_\Sigma})$ is bounded above and below, where $L_{s_\Sigma} := G^{d-1}_{[x]} G_{s_\Sigma}$. Note that $J_\Sigma \subset L_{s_\Sigma}$.

3 Preliminary Lemmas on Moy–Prasad Subgroups

In this subsection, we prove technical lemmas that we need to prove the main theorem. We keep the notation from the previous section.

3.1 Lemmas on π_Σ and γ

Recall from [KM03] that when (HB) and (HGT) are valid, any irreducible smooth representation (τ, V_τ) contains a good minimal K-type. The following lemma analyzes other possible minimal K-types occurring in (τ, V_τ).

Lemma 3.1. *Suppose (HB) and (HGT) are valid. Let (V_τ, τ) be an irreducible smooth representation of G of positive depth ϱ. Let $(\chi, G_{x,\varrho})$ be a good minimal K-type of τ represented by $X + \mathfrak{g}_{x,(-\varrho)+}$ where $X \in \mathfrak{g}_{x,(-\varrho)}$ is a good element of depth $(-\varrho)$. Let G' be the connected component of the centralizer of X in G.*

(1) Fix an embedding $\mathcal{B}(G') \hookrightarrow \mathcal{B}(G)$ (such an embedding can be chosen by [Lan00, Theorem 2.2.1]) and let C'_x be a facet of maximal dimension in $\mathcal{B}(G')$ containing x in its closure $\overline{C}'_x$. There exists a facet of maximal dimension C_x in $\mathcal{B}(G)$ such that $x \in \overline{C}_x \cap \overline{C}'_x$ and $C'_x \cap \overline{C}_x$ is of maximal dimension in $\mathcal{B}(G')$.

(2) Let $y \in \mathcal{B}(G)$ and suppose that $V_\tau^{G_{y,\varrho+}} \neq 0$. As a representation of $G_{y,\varrho}$, $V_\tau^{G_{y,\varrho+}}$ is a sum of characters χ''s which are represented by ${}^h(X + \eta') + \mathfrak{g}_{y,(-\varrho)+} \subset \mathfrak{g}_{y,(-\varrho)}$ for some $\eta' \in \mathfrak{g}'_{(-\varrho)+}$ and $h \in G_{[y]} S_x$ for some compact mod center set S_x. Moreover, one can choose S_x in a way depending only on x, G'.

Remark 3.2. Note that $x \in \mathcal{B}(G')$ by [KM03, Theorem 2.3.1].

Proof. (1) Let $V := \cup_C \overline{C}$ where the union runs over the set of facets of maximal dimension $C \subset \mathcal{B}(G)$ with $x \in \overline{C}$. Let V° be the interior of V. Then, $x \in V^\circ$ and

V° is open in $\mathcal{B}(G)$, hence, $C'_x \cap V^\circ \neq \emptyset$ and is open in $\mathcal{B}(G')$. Therefore, at least one of $\overline{C} \cap C'_x$ with $x \in \overline{C}$ contains an open set in $\mathcal{B}(G')$. Set C_x to be one of such facets.

(2) Since the action of $G_{y,\varrho}$ on $V_\tau^{G_{y,\varrho+}}$ factors through the finite abelian quotient $G_{y,\varrho}/G_{y,\varrho^+}$, we see that $V_\tau^{G_{y,\varrho+}}$ decomposes as a direct sum of characters of $G_{y,\varrho}$. Let χ' be a $G_{y,\varrho}$ subrepresentation of $V_\tau^{G_{y,\varrho+}}$. Then, $(\chi', G_{y,\varrho})$ is also a minimal K-type of τ. Let $X' + \mathfrak{g}_{y,(-\varrho)^+} \subset \mathfrak{g}_{y,(-\varrho)}$ be the dual cosets representing χ'. Then, $(X + \mathfrak{g}_{x,(-\varrho)^+}) \cap {}^G(X' + \mathfrak{g}_{y,(-\varrho)^+}) \neq \emptyset$. Since $(X + \mathfrak{g}_{x,(-\varrho)^+}) = {}^{G_{x,0+}}(X + \mathfrak{g}'_{x,(-\varrho)^+})$, there are $\eta \in \mathfrak{g}'_{x,(-\varrho)^+}$ and $g \in G$ such that $X + \eta \in {}^{g^{-1}}(X' + \mathfrak{g}_{y,(-\varrho)^+}) \subset \mathfrak{g}_{g^{-1}y,-\varrho}$. By [KM03, Lemma 2.3.3], $g^{-1}y \in \mathcal{B}(G')$.

To choose S_x, let $\mathcal{A}(T)$ be an apartment in $\mathcal{B}(G)$ such that $C_x \cup C'_x \subset \mathcal{A}(T)$. For each alcove $C \subset \mathcal{A}(T)$ with $\overline{C} \cap \overline{C}'_x \neq \emptyset$, choose $w_C \in N_G(T)$ such that $C = w_C C_x$. Now, set

$$S_x := \{\delta \cdot w_C^{-1} \mid C \text{ is an alcove in } \mathcal{B}(G) \text{ with } \overline{C} \cap \overline{C}'_x \neq \emptyset,\ \delta \in G_{[C_x]}\}.$$

We claim that there is $g' \in G'$ such that $gg' \in G_{[y]}S_x$. Let $g' \in G'$ such that $(gg')^{-1}y \in \overline{C}'_x$. Then, there is $g''^{-1} \in S_x$ such that $g''y = (gg')^{-1}y$. Hence, $gg'g'' \in G_{[y]}$ and $gg' \in G_{[y]}S_x$. Then one can take $h = gg'$ and $\eta' = g'^{-1}\eta g'$ since ${}^h(X + \eta') \equiv {}^g({}^{g'}X + \eta) \equiv {}^g(X + \eta) \equiv X' \pmod{\mathfrak{g}_{y,(-\varrho)^+}}$. By construction S_x depends only on x and G'. □

Definition 3.3. Let $\gamma \in G_{\mathrm{reg}}$. Let $\mathbf{T}^\gamma$ be the unique maximal torus containing γ, and $\Phi := \Phi(\mathbf{T}^\gamma)$ the set of absolute roots of $\mathbf{T}^\gamma$. Let $\Phi^+ := \Phi^+(\mathbf{T}^\gamma)$ be the set of positive roots.

(i) Define the *singular depth* $\mathrm{sd}_\alpha(\gamma)$ of γ in the direction of $\alpha \in \Phi$ as

$$\mathrm{sd}_\alpha(\gamma) := \nu(\alpha(\gamma) - 1).$$

and the *singular depth* $\mathrm{sd}(\gamma)$ of γ as

$$\mathrm{sd}(\gamma) := \max_{\alpha\in\Phi} \mathrm{sd}_\alpha(\gamma).$$

When γ is not regular, see [AK07, §4] for definition. In [MS12, §4.2], $\mathrm{sd}(\gamma)$ is defined as $\max_{\alpha\in\Phi^+} \mathrm{sd}_\alpha(\gamma)$. When γ is compact, both definitions coincide.

(ii) Recall that the height of $\alpha \in \Phi^+(\mathbf{T}^\gamma)$ is defined inductively as follows:

- $\mathrm{ht}(\alpha) = 1$ if $\alpha \in \Phi^+$ is simple;
- $\mathrm{ht}(\alpha + \beta) = \mathrm{ht}(\alpha) + \mathrm{ht}(\beta)$ if $\alpha, \beta, \alpha + \beta \in \Phi^+$.

Define the *height h_G of* Φ as $\max_{\alpha\in\Phi^+} \mathrm{ht}(\alpha)$. Note that the height of Φ depends only on G.

Lemma 3.4. *Suppose $\gamma \in G_{\mathrm{reg}} \cap T_0^\gamma$ splits over a tamely ramified extension. Suppose $z \in \mathcal{A}(T^\gamma)$, and ${}^g\gamma \in G_z$ for $g \in G$. Then, ${}^gT^\gamma_{h_\gamma^+} \subset G_z$, where $h_\gamma := h_G \cdot \mathrm{sd}(\gamma)$.*

Proof. This is a reformulation of Meyer and Solleveld [MS12, Lemma 4.3]. More precisely, ${}^g\gamma \in G_z$ is equivalent to $z \in \mathcal{B}(G)^{{}^g\gamma}$, hence $z \in \mathcal{B}(G)^{{}^g(\gamma T_{h_\gamma^+})}$ by *loc. cit.* , which in turn implies that ${}^g(\gamma T_{h_\gamma^+}) \subset G_z$ and ${}^uT_{h_\gamma^+} \subset G_z$. □

Lemma 3.5. *Suppose $\gamma \in G_{\mathrm{reg}} \cap G_0$ splits over a tamely ramified extension E. Let $\mathbf{T}^\gamma \subset \mathbf{G}^0 \subsetneq \mathbf{G}^1 \subsetneq \cdots \subsetneq \mathbf{G}^d = \mathbf{G}$ be an E-twisted tamely ramified Levi sequence, $z \in \mathcal{A}(\mathbf{T}^\gamma, E) \subset \mathcal{B}(\mathbf{G}^0, E)$ and $a_i \in \mathbb{R}$ with $0 \le a_1 \le \cdots \le a_d$. Set $K_E = \mathbf{G}^0(E)_z\mathbf{G}^1(E)_{z,a_1}\mathbf{G}^2(E)_{z,a_2}\cdots\mathbf{G}^d(E)_{z,a_d}$. Suppose $g \in G$ such that ${}^g\gamma \in K_E$. Then, we have ${}^g\mathbf{T}^\gamma(E)_{A+} \subset K_E$ where $A = h_\gamma + a_d$ and h_γ is as in Lemma 3.4.*

Proof. Without loss of generality, we may assume $E = k$. Let $O \in \mathcal{A}(T^\gamma)$ defined by $\alpha(O) = 0$ for all $\alpha \in \Phi$. For $\alpha \in \Phi$, let U_α be the root subgroup associated with α. We fix the pinning $x_\alpha : k \to U_\alpha$. Define $U_{\alpha,r}$ to be the image $\{u \in k \mid \nu(u) \ge r\}$ under the isomorphism $x_\alpha : k \to U_\alpha$. Let $C \subset \mathcal{A}(T^\gamma)$ be the facet of maximal dimension with $O \in \overline{C}$. One may assume that $z \in \overline{C}$ (by conjugation by an element of $N_G(T^\gamma)$ if necessary). Then, G_C is an Iwahori subgroup and we have the Bruhat decomposition $G = G_C N_G(T^\gamma) G_C$, cf. [Tit79, 3.3.1]. Let $w \in N_G(T^\gamma)$ with $g \in G_C w G_C$. Let $A_w = \{\alpha \in \Phi(T^\gamma) \mid U_{w\alpha} \cap G_C \subsetneq w(U_\alpha \cap G_C)\}$. Write $U_w = \prod_{\alpha \in A_w} U_\alpha$. Note that U_w may not necessarily be a group. We prove the lemma through steps (1)–(8) below.

(1) Let $w \in N_G(T^\gamma)$. Then, there is a Borel subgroup B containing $wU_w w^{-1}$.

Proof. Each chamber D containing O in its closure defines an open cone $\mathcal{C}_D$ in $\mathcal{A}(T^\gamma)$ and we have $\mathcal{A}(T^\gamma) = \cup_{O \in \overline{D}} \overline{\mathcal{C}_D}$ is the union is over the chambers D with $O \in \overline{D}$. Recall that each $\mathcal{C}_D$ defines a Borel subgroup B_D. If $\mathcal{C}_D$ contains wC, one can take $B = B_D$. □

For D as in (1), let Φ_D^+ be the set of positive T^γ-roots associated with $\mathcal{C}_D$ and write $\mathrm{ht}_D = \mathrm{ht}_{\Phi_D^+}$ for simplicity.

(2) Write $B_D = T^\gamma U$ and let $\overline{U}$ be the opposite unipotent subgroup. Then, $gw^{-1} = t \cdot \overline{u} \cdot u$ for $t \in T_0^\gamma$, $\overline{u} \in G_C \cap \overline{U}$ and $u \in U$.

Proof. We have that G_C has an Iwahori decomposition with respect to B: $G_C = (G_C \cap T^\gamma)(G_C \cap \overline{U})(G_C \cap U)$. Then, the above follows from $gw^{-1} \in G_C w G_C w^{-1} \subset G_C w U_w w^{-1} \subset (G_C \cap T^\gamma)(G_C \cap \overline{U})(G_C \cap U)U \subset (G_C \cap T^\gamma)(G_C \cap \overline{U})U$. □

(3) Since $t \in T_0^\gamma \subset K_E$, we may assume without loss of generality that $t = 1$ or $gw^{-1} = \overline{u}u$.

(4) We have $u \in G_{x+\mathrm{sd}(\gamma)\mathrm{ht}_D} \cap U$.

Proof. Observe that ${}^w\gamma \in K_E$ and $\mathrm{sd}({}^w\gamma) = \mathrm{sd}(\gamma)$. Observe also that ${}^u({}^w\gamma) \in G_C \subset G_z$ since $\overline{u} \in G_z$ and $K_E \subset G_z$. Since ${}^u({}^w\gamma) \in G_C \subset G_z$ we can apply [MS12, Proposition 4.2] to deduce that $u \in G_{z+\mathrm{sd}(\gamma)\mathrm{ht}_D} \cap U$. ☐

(5) For $\gamma' \in T^\gamma_{A+}$, we have (i) $(u, {}^w(\gamma\gamma')) \equiv (u, {}^w\gamma) \pmod{G_{z+\mathrm{sd}(\gamma)\mathrm{ht}_D, A+} \cap U}$, and (ii) $(u, {}^w(\gamma\gamma')) \equiv (u, {}^w\gamma) \pmod{G_{z,a_d} \cap U}$.

Proof. For $\gamma' \in T^\gamma_{A+} \subset G_{z+\mathrm{sd}(\gamma)\mathrm{ht}_D, A+}$, the commutators $(u, {}^w(\gamma\gamma'))$ and $(u, {}^w\gamma)$ are in $G_{z+\mathrm{sd}(\gamma)\mathrm{ht}_D} \cap U$ and also in the same coset mod $G_{z+\mathrm{sd}(\gamma)\mathrm{ht}_D, A+} \cap U$. Hence, (i) follows.

The assertion (ii) follows from (i). Indeed, we note that

$$G_{z+\mathrm{sd}(\gamma)\mathrm{ht}_D, A+} \cap U = \prod_{\alpha \in \Phi_D^+} U_{\alpha, -\alpha(z) - \mathrm{sd}(\gamma)\mathrm{ht}_D(\alpha) + A+}$$

is contained in $\prod_{\alpha \in \Phi_D^+} U_{\alpha, -\alpha(z)+s+}$ which is itself contained in $G_{z,s} \cap U$. ☐

(6) For any $\gamma' \in T^\gamma_{A+}$, we have ${}^w\gamma\overline{u}^{-1}{}^w\gamma^{-1} \equiv ({}^w(\gamma\gamma'))\overline{u}^{-1}({}^w(\gamma\gamma'))^{-1} \pmod{G_{z,a_d}}$.

(7) For any $\gamma' \in T^\gamma_{A+}$, we have $g\gamma g^{-1}{}^w\gamma^{-1} \equiv g(\gamma\gamma')g^{-1}{}^w(\gamma\gamma')^{-1} \pmod{G_{z,a_d}}$

Proof. Write $g\gamma g^{-1}{}^w\gamma^{-1} = \overline{u}(u, {}^w\gamma){}^w\gamma\overline{u}^{-1}{}^w\gamma^{-1}$ and

$$g(\gamma\gamma')g^{-1}{}^w(\gamma\gamma')^{-1} = \overline{u}(u, {}^w(\gamma\gamma')){}^w(\gamma\gamma')\overline{u}^{-1}{}^w(\gamma\gamma')^{-1}.$$

Then, (7) follows from (5) and (6). ☐

(8) ${}^gT^\gamma_{A+} \subset K_E$.

Proof. Note that ${}^w\gamma, {}^w(\gamma\gamma') \in K_E$. Hence, $g\gamma g^{-1} \equiv g(\gamma\gamma')g^{-1} \pmod{K_E}$ by (6). Since $g\gamma g^{-1} \in K_E$, we have $gT^\gamma_{A+}g^{-1} \subset K_E$. ☐

The proof of Lemma 3.5 is now complete. ☐

Lemma 3.6. *In the same situation as in Lemma 3.5, suppose in addition that $h_G = 1$. Then, we have ${}^g\mathbf{T}^\gamma(E)_{\mathrm{sd}(\gamma)+} \subset K_E$.*

Proof. Since $h_G = 1$, we have $d = 0$ or $d = 1$. If $d = 0$, the assertion is precisely Lemma 3.4. We assume $d = 1$ from now. Then, $K_E = T^\gamma_0 G_{z,a_1}$.

As before, we may assume $E = k$. The assertions (1)–(8) below refer to those in the above proof of Lemma 3.5. Following the proof of Lemma 3.5, write $\Phi_D = \{\pm\alpha\}$. Let $gw^{-1} = \overline{u}u$ for $u \in U_\alpha$ and $\overline{u} \in U_{-\alpha}$ as in (3). Under the isomorphism via $x_\alpha : k \to U_\alpha$, we will use the same notation for u and $x_\alpha^{-1}(u)$. Then, we can write $x_\alpha(u) = u$. Similarly, $x_{-\alpha}(\overline{u}) = \overline{u}$. Let $\alpha^\vee : k^\times \to T^\gamma$ be the coroot of α. It is enough to prove that $g(\gamma\gamma')g^{-1}\,{}^w(\gamma\gamma')^{-1} \in K_E$ for any $\gamma' \in T^\gamma_{\mathrm{sd}(\gamma)+}$.

We have $(u, {}^w(\gamma\gamma')) = x_\alpha((1 - \alpha({}^w(\gamma\gamma')))u)$ and $(\overline{u}, {}^w(\gamma\gamma')) = x_{-\alpha}((1 - \alpha({}^w(\gamma\gamma')^{-1}))\overline{u})$. For simplicity, write $u_{\gamma'} = (1 - \alpha({}^w(\gamma\gamma')))u = x_\alpha(u_{\gamma'})$ and $\overline{u}_{\gamma'} = (1 - \alpha({}^w(\gamma\gamma')^{-1}))\overline{u} = x_{-\alpha}(\overline{u}_{\gamma'})$. Since $x_\alpha(u_{\gamma'}) \in U_\alpha \cap G_z$ by (4) and

Lemma 3.4 and $x_{-\alpha}(\overline{u}) \in U_{-\alpha} \cap G_z$, we have

$$\nu(u_{\gamma'}) \geq -\alpha(z), \quad \nu(\overline{u}) \geq \alpha(z).$$

Similarly as in (7), we calculate $g(\gamma\gamma')g^{-1}\,{}^{w}(\gamma\gamma')^{-1}$, but, now explicitly using the Chevalley basis. Then,

$$\begin{aligned} &g(\gamma\gamma')g^{-1w}(\gamma\gamma')^{-1} = \overline{u}(u, {}^{w}(\gamma\gamma'))^{w}(\gamma\gamma')\overline{u}^{-1w}(\gamma\gamma')^{-1} \\ &= x_{-\alpha}(\overline{u})x_{\alpha}(u_{\gamma'})x_{-\alpha}(-\overline{u})x_{-\alpha}(\overline{u})\,\mathrm{Ad}({}^{w}(\gamma\gamma'))(x_{-\alpha}(-\overline{u})) \\ &= x_{\alpha}(u_{\gamma'}(1+\overline{u}u_{\gamma'})^{-1})\alpha^{\vee}((1+\overline{u}u_{\gamma'})^{-1})x_{-\alpha}(-\overline{u}^2u_{\gamma'}(1+\overline{u}u_{\gamma'})^{-1})x_{-\alpha}(\overline{u}_{\gamma'}) \end{aligned}$$

When $\gamma' = 1$, we have

$$g\gamma g^{-1w}\gamma^{-1} = x_{\alpha}(u_1(1+\overline{u}u_1)^{-1})\alpha^{\vee}((1+\overline{u}u_1)^{-1})x_{-\alpha}(-\overline{u}^2u_1(1+\overline{u}u_1)^{-1})x_{-\alpha}(\overline{u}_1).$$

Since $g\gamma g^{-1w}\gamma^{-1} \in K_E$, we have $\nu(1+\overline{u}u_{-1}) = 0$ and $\nu(u_1) = \nu(u_1(1+\overline{u}u_1)^{-1}) \geq -\alpha(z) + a_1$. Combining this with $\nu(\overline{u}) \geq \alpha(z)$, we have $\nu(\overline{u}^2u_1(1+\overline{u}u_1)^{-1}) \geq \alpha(z) + a_1$. Note that

(*i*) $\nu(1+\overline{u}u_{\gamma'}) = \nu(1+\overline{u}u_1)$ and $\nu(u_1) = \nu(u_{\gamma'})$;
(*ii*) $\nu(\overline{u}^2u_1(1+\overline{u}u_1)^{-1}) = \nu(\overline{u}^2u_{\gamma'}(1+\overline{u}u_{\gamma'})^{-1}) \geq \alpha(z) + a_1$;
(*iii*) $\nu(\overline{u}_1) = \nu(\overline{u}_{\gamma'}) \geq \alpha(z) + a_1$. The last inequality follows from $x_{-\alpha}(\overline{u}_1) \in K_E$.

From (*i*)–(*iii*), we have $g(\gamma\gamma')g^{-1w}(\gamma\gamma')^{-1} \in K_E$ and conclude that ${}^{g}T^{\gamma}_{\mathrm{sd}(\gamma)+} \subset K_E$. □

Proposition 3.7. *Recall the subgroup J_Σ from Sect. 2.4. Let $\gamma \in G_{\mathrm{reg}}$, $g \in G$, and suppose that $\gamma \in H_{\Sigma,g} := {}^{g}J_\Sigma \cap G_x$. Let*

$$A_{\gamma,\Sigma} := \begin{cases} h_G \cdot \mathrm{sd}(\gamma) + s_{d-1} & \text{if } h_G > 1, \\ \mathrm{sd}(\gamma) & \text{if } h_G = 1. \end{cases}$$

Then, we have

(i) $T^{\gamma}_{A_{\gamma,\Sigma}+} \subset H_{\Sigma,g}$,
(ii) $\sharp\left((T^\gamma \cap G_x)H_{\Sigma,g}/\,H_{\Sigma,g}\right) \leq q^{r_G(A_{\gamma,\Sigma}+1)}$.

Proof. For simplicity of notation, we write A for $A_{\gamma,\Sigma}$.

(i) By Lemma 3.4, $T^{\gamma}_{A+} \subset G_x$. Since $T^{\gamma}_{A+} = \mathbf{T}^{\gamma}(E)_{A+} \cap G_{0+} \subset {}^{g}K_E \cap G_{0+}$ from Lemmas 3.5 and 3.6, we have $T^{\gamma}_{A+} \subset {}^{g}J_\Sigma \cap G_{0+}$. Hence, $T^{\gamma}_{A+} \subset H_{\Sigma,g}$.
(ii) Note that $G_x \cap T^\gamma \subset T^\gamma_0$. Then, we have

$$\sharp\left((T^\gamma \cap G_x)H_{\Sigma,g})/\,H_{\Sigma,g}\right) \leq \sharp\left(T^\gamma_0/\,T^\gamma_{A+}\right) \leq q^{r_G(A+1)}.$$

□

3.2 Inverse Image Under Conjugation

In this subsection we prove a lemma to control the volume change of an open compact subgroup under the conjugation map as we will need the result in Proposition 4.15 below. For regular semisimple elements $g \in G$ and $X \in \mathfrak{g}$, denote by $\mathfrak{g}_g$ and $\mathfrak{g}_X$ the centralizer of g and X in $\mathfrak{g}$, respectively. Define the Weyl discriminant as

$$D(g) := |\det(\mathrm{Ad}(g)|\mathfrak{g}/\mathfrak{g}_g)|, \quad D(X) := |\det(\mathrm{ad}(X)|\mathfrak{g}/\mathfrak{g}_X)|.$$

Recall that the rank of G, to be denoted r_G, is the dimension of (any) maximal torus in G. Define $\psi(r_G)$ to be the maximal $d \in \mathbb{Z}_{\geq 0}$ such that $\phi(d) \leq r_G$, where ϕ is the Euler phi-function. Put $N(G) := \max(\psi(r_G), \dim G)$. Under the assumption that $p > N(G) + 1$, recall from [Wal08, Appendix B, Proposition B] that there is a homeomorphism

$$\exp : \mathfrak{g}_{0+} \xrightarrow{\sim} G_{0+}.$$

Under the hypothesis (Hk) this exp map is filtration preserving, in particular, it preserves the ratio of volumes.

Lemma 3.8. *Suppose $p > N(G) + 1$ and (Hk) is valid. Let $x \in \mathcal{B}(\mathbf{G})$. Let $\gamma \in G_x \cap G_{0+}$ and suppose that γ is regular semisimple (so that $D(\gamma) \neq 0$). Consider the conjugation map $\psi_\gamma : G \to G$ given by $\delta \mapsto \delta\gamma\delta^{-1}$. For each open compact subgroup $H \subset G_x$ containing γ, we have*

$$\frac{\mathrm{vol}_{G/Z}(\psi_\gamma^{-1}(H) \cap G_x)}{\mathrm{vol}_{G/Z}(H)} \leq \sharp\left((G_x \cap T^\gamma)H/H\right) \cdot D(\gamma)^{-1} \cdot \sharp\left(H/(H \cap G_{x,0+})\right).$$

Proof. It suffices to prove that the left-hand side is bounded by $\sharp\,(G_x \cap T^\gamma/H \cap T^\gamma)\cdot D(\gamma)^{-1}$ under the additional assumption that $H \subset G_x \cap G_{0+}$. Indeed, in general, one only needs to also count the contribution on $H/(H \cap G_{x,0+})$, which is bounded by its cardinality.

Put $Y := \exp^{-1}(\gamma)$ and $\mathfrak{h} := \exp^{-1}(H) \subset \mathfrak{g}_x \cap \mathfrak{g}_{0+}$. Note that $\mathfrak{h}$ is an $\mathcal{O}_k$-lattice in $\mathfrak{g}$ since H is an open compact subgroup. Write $C_Y : \mathfrak{g} \to \mathfrak{g}$ for the map $X \mapsto [X, Y]$, whose restriction to $\mathfrak{g}_{0+}$ is going to be denoted by $C_{Y,0+}$. Define $C_\gamma : G_{0+} \to G_{0+}\gamma^{-1}$ by $\delta \mapsto \delta\gamma\delta^{-1}\gamma^{-1}$, which is the composition of ψ_γ with the right multiplication by γ^{-1}. Since $\gamma \in H$ we have $C_\gamma^{-1}(H) = \psi_\gamma^{-1}(H) \subset G_{0+}$. Via the exponential map, $C_\gamma : \psi_\gamma^{-1}(H) \cap G_x \to H$ corresponds to

$$C_{Y,0+} : \exp^{-1}(C_\gamma^{-1}(H) \cap G_x) = C_{Y,0+}^{-1}(\mathfrak{h} \cap \mathfrak{g}_x) \to \mathfrak{h}.$$

Since the exponential map preserves the ratio of volumes,

$$\frac{\mathrm{vol}_{G/Z}(\psi_\gamma^{-1}(H)\cap G_x)}{\mathrm{vol}_{G/Z}(H)} = \frac{\mathrm{vol}_{\mathfrak{g}/\mathfrak{z}}(C_{Y,0+}^{-1}(\mathfrak{h})\cap \mathfrak{g}_x)}{\mathrm{vol}_{\mathfrak{g}/\mathfrak{z}}(\mathfrak{h})} \leq \frac{\mathrm{vol}_{\mathfrak{g}/\mathfrak{z}}(C_Y^{-1}(\mathfrak{h})\cap \mathfrak{g}_x))}{\mathrm{vol}_{\mathfrak{g}/\mathfrak{z}}(\mathfrak{h})}$$
$$= [C_Y^{-1}(\mathfrak{h})\cap \mathfrak{g}_x : \mathfrak{h}].$$

We will be done if we show that $[C_Y^{-1}(\mathfrak{h})\cap \mathfrak{g}_x : \mathfrak{h}] \leq D(Y)^{-1}[\mathfrak{g}_Y\cap\mathfrak{g}_x : \mathfrak{g}_Y\cap\mathfrak{h}]$. Write pr for the projection map $\mathfrak{g}_x \to \mathfrak{g}_x/(\mathfrak{g}_x\cap\mathfrak{g}_Y)$. Consider the commutative diagram

$$\begin{tikzcd}
\mathfrak{g}_Y\cap\mathfrak{g}_x \arrow[r] \arrow[d, hook] & \mathfrak{g}_Y\cap\mathfrak{h} \arrow[r, hook] \arrow[d, hook] & \mathfrak{g}_Y\cap\mathfrak{g}_x \arrow[d, hook] \\
C_Y^{-1}(\mathfrak{h})\cap\mathfrak{g}_x \arrow[r, "C_Y"] \arrow[d, hook] & \mathfrak{h} \arrow[r, hook] \arrow[d, hook] & \mathfrak{g}_x \arrow[d, "\mathrm{pr}"] \\
\mathrm{pr}(C_Y^{-1}(\mathfrak{h})\cap\mathfrak{g}_x) \arrow[r, "\overline{C}_Y"] & \mathrm{pr}(\mathfrak{h}) \arrow[r, hook] & \mathfrak{g}_x/(\mathfrak{g}_Y\cap\mathfrak{g}_x)\ \subset\ \mathfrak{g}/\mathfrak{g}_Y
\end{tikzcd}$$

Denote by the preimage of $\mathrm{pr}(\mathfrak{h})$ in $\mathrm{pr}(C_Y^{-1}(\mathfrak{h})$ (resp. $C_Y^{-1}(\mathfrak{h})\cap\mathfrak{g}_x$) by L_2 and L_1. Then L_2 and L_1 are $\mathcal{O}_k$-lattices in $\mathfrak{g}/\mathfrak{g}_Y$ and $\mathfrak{g}$, respectively, where $\mathcal{O}_k$ is the ring of integers in k. So we have

$$[C_Y^{-1}(\mathfrak{h})\cap\mathfrak{g}_x : \mathfrak{h}] \leq [L_1 : \mathfrak{h}] = [\mathfrak{g}_Y\cap\mathfrak{g}_x : \mathfrak{g}_Y\cap\mathfrak{h}][L_2 : \mathrm{pr}(\mathfrak{h})].$$

Since $\overline{C}_Y = \mathrm{ad}(Y)$ as k-linear isomorphisms on $\mathfrak{g}/\mathfrak{g}_Y$, we see that $[L_2 : \mathrm{pr}(\mathfrak{h})] = D(Y)^{-1}$, completing the proof. □

Combining Proposition 3.7 and Lemma 3.8, we have the following:

Corollary 3.9. *We keep the situation and notation from Proposition 3.7. Then, we have*

$$\frac{\mathrm{vol}_{G/Z}(\psi_\gamma^{-1}(H_{\Sigma,g})\cap G_x)}{\mathrm{vol}_{G/Z}(H_{\Sigma,g})} \leq D(\gamma)^{-1}q^{\dim(G)+r_G(A_{\gamma,\Sigma}+1)}$$

3.3 Intersection of L_s with a Maximal Unipotent Subgroup

For later use, we study the intersections of L_s with unipotent subgroups in this subsection. Consider a tamely ramified twisted Levi sequence $(\mathbf{G}', \mathbf{G})$. Let $\mathbf{T}$ (resp. $\mathbf{T}'$) be a maximally k-split maximal torus of $\mathbf{G}$ (resp. $\mathbf{G}'$) such that $\mathbf{T}'^s \subset \mathbf{T}^s$ where $\mathbf{T}^s$ and $\mathbf{T}'^s$ are the k-split components of $\mathbf{T}$ and $\mathbf{T}'$, respectively. Set $M := Z_G^\circ(T^s)$ (resp. $M' := Z_G^\circ(T'^s)$), a minimal Levi subgroup of G containing T (resp. T'). Note that $M \subset M'$. Fix a parabolic subgroup MN of G (resp. $M'N'$), where N (resp. N') is the unipotent subgroup such that $N' \subset N$.

Lemma 3.10. *We keep the notation from above. Let $x \in \mathcal{A}(T')$.*

(i) *For any $a, a' \in \mathbb{R}$ with $a > a' > 0$,*

$$[(G_{x,a'} \cap N') : ((G'_{x,a'} G_{x,a}) \cap N')]^2 \leq q^{\dim_k(N)} [G_{x,a'} : (G'_{x,a'} G_{x,a})].$$

(ii)

$$[(G_x \cap N') : ((G'_x G_{x,0+}) \cap N')]^2 \leq q^{\dim_k(N)} [G_x : (G'_{x',0} G_{x,0+})].$$

(iii)

$$[(G_x \cap N') : ((G'_x G_{x,a}) \cap N')]^2 \leq q^{2\dim_k(N)} [G_x : (G'_x G_{x,a})].$$

Proof. (i) Both $\mathcal{J} := G_{x,a'}$ and $\mathcal{J}' := (G'_{x,a'} G_{x,a})$ are decomposable with respect to M' and N', that is, $\mathcal{J} = (\mathcal{J} \cap \overline{N}') \cdot (\mathcal{J} \cap M') \cdot (\mathcal{J} \cap N')$, etc. Write $Y_X := Y \cap X$ for any $X, Y \subset G$. Then, $[\mathcal{J} : \mathcal{J}'] = [\mathcal{J}_{\overline{N}'} : \mathcal{J}'_{\overline{N}'}] \cdot [\mathcal{J}_{M'} : \mathcal{J}'_{M'}] \cdot [\mathcal{J}_{N'} : \mathcal{J}'_{N'}]$ and we have

$$\begin{aligned}
[\mathcal{J} : \mathcal{J}'] &= [\mathcal{J}_{M'} : \mathcal{J}'_{M'}][\mathcal{J}_{N'} : \mathcal{J}'_{N'}][\mathcal{J}_{\overline{N}'} : \mathcal{J}'_{\overline{N}'}] \geq [\mathcal{J}_{N'} : \mathcal{J}'_{N'}][\mathcal{J}_{\overline{N}'} : \mathcal{J}'_{\overline{N}'}] \\
&\geq \frac{1}{q^{\dim_k(N')}} [\mathcal{J}_{N'} : \mathcal{J}'_{N'}]^2 \geq \frac{1}{q^{\dim_k(N)}} [\mathcal{J}_{N'} : \mathcal{J}'_{N'}]^2.
\end{aligned}$$

(ii) This follows from

$$\begin{aligned}
[G_x : G'_x G_{x+}] &\geq [(G_x)_{N'} G_{x,0+} : (G'_x G_{x+})_{N'} G_{x,0+}] \cdot [(G_x)_{\overline{N}'} G_{x,0+} : (G'_x G_{x+})_{\overline{N}'} G_{x,0+}] \\
&= [(G_x)_{N'} G_{x,0+} : (G'_x G_{x+})_{N'} G_{x,0+}] \cdot [(G_x)_{\overline{N}'} : (G'_x G_{x+})_{\overline{N}'}] \\
&\geq \frac{1}{q^{\dim_k(N)}} [(G_x)_{N'} : (G'_x G_{x,0+})_{N'}]^2.
\end{aligned}$$

(iii) We have

$$\begin{aligned}
[G_x : (G'_x G_{x,a})] &= [G_x : (G'_x G_{x,0+})][(G'_x G_{x,0+}) : (G'_x G_{x,a})] \\
&= [G_x : (G'_x G_{x,0+})][G_{x,0+} : (G'_{x,0+} G_{x,a})].
\end{aligned}$$

Combining (i) and (ii), we have

$$\begin{aligned}
[G_x : (G'_x G_{x,a})] &\geq \frac{1}{q^{2\dim_k(N)}} [(G_x)_{N'} : (G'_x G_{x,0+})_{N'}]^2 [(G_{x,0+})_{N'} : (G'_{x,0+} G_{x,a})_{N'}]^2 \\
&= \frac{1}{q^{2\dim_k(N)}} [(G_x)_{N'} : (G'_x G_{x,a})_{N'}]^2.
\end{aligned}$$

□

4 Asymptotic Behavior of Supercuspidal Characters

4.1 Main Theorem

Conjecture 4.1. *Consider the set of* π *in* $\mathrm{Irr}^{\mathrm{sc}}(G)$ *such that the central character of* π *is unitary. For each fixed* $\gamma \in G_{\mathrm{reg}}$*,*

$$\frac{\Theta_\pi(\gamma)}{\deg(\pi)} \to 0 \quad \textit{as} \ \deg(\pi) \to \infty;$$

namely for each $\epsilon > 0$ *there exists* $d_\epsilon > 0$ *such that* $|\Theta_\pi(\gamma)/\deg(\pi)| < \epsilon$ *whenever* $\deg(\pi) > d_\epsilon$.

Our main theorem in the qualitative form is a partial confirmation of the conjecture under the hypotheses discussed in Sect. 2.7 and above Lemma 3.8.

Theorem 4.2. *Suppose that* (Hk), (HB)*, and* (HGT) *are valid. Then, Conjecture 4.1 holds true if* γ *and* π *are restricted to the sets* G_{0+} *and* $\mathrm{Irr}^{\mathrm{Yu}}(G)$*, respectively.*

The proof is postponed to Sect. 4.5 below. Actually we will establish a rather explicit upper bound on $|\Theta_\pi(\gamma)/\deg(\pi)|$, which will lead to a quantitative strengthening of the above theorem. See Theorem 4.18 below.

The central character should be unitary in the conjecture; it would be a problem if each π is twisted by arbitrary (non-unitary) unramified characters of G. However the assumption plays no role in the theorem since every unramified character is trivial on G_{0+}.

We are cautious to restrict the conjecture to supercuspidal representations as our result does not extend beyond the supercuspidal case. However it is natural to ask whether the conjecture is still true for discrete series representations. As a small piece of psychological evidence we verify the analogue of Conjecture 4.1 for discrete series of real groups on elliptic regular elements in Sect. 5.2 below.

4.2 Reductions

Let π be as in Theorem 4.2 associated with a generic G-datum Σ. In proving Theorem 4.2, we may assume that each π (and hence Σ) is associated with a fixed orbit of $(\vec{\mathbf{G}}, x)$ since there are only a finite number of $(\vec{\mathbf{G}}, x)$ up to conjugacy. Let $\mathbf{T}^\gamma$ denote the unique maximal torus containing γ, and pick any $y \in \mathcal{A}(T^\gamma) = \mathcal{B}(T^\gamma)$. We may assume the following without loss of generality.

1. $\phi_d = 1$: since $\Theta_{\overline{\phi}_d \otimes \pi}(\gamma) = \overline{\phi}_d(\gamma)\Theta_\pi(\gamma)$ and $\deg(\pi) = \deg(\overline{\phi}_d \otimes \pi)$, it is enough to verify the theorem only for the generic G-data such that $\phi_d = 1$. Note that the depth of π is given by r_{d-1} if $\phi_d = 1$.

2. $r_\Sigma = r_{d-1} \geq 1$: Lemma 2.9 implies that $\deg(\pi) \to \infty$ is equivalent to $\mathrm{vol}_{G/Z}(J_\Sigma) \to 0$ as $i \to \infty$, which is in turn equivalent to $r_\Sigma \to \infty$. Hence, we may assume $r_\Sigma \geq 1$ without loss of generality.
3. $x \in \overline{C}_y$ where C_y is a facet of maximal dimension in $\mathcal{B}(G)$ containing y in its closure: this is due to that the G-orbit of x in $\mathcal{B}(G)$ intersects with $\overline{C}_y$ nontrivially.

The following is a consequence of (3):

3.′ $\gamma \in G_x$ since $\gamma \in T^{\gamma}_{0+} \subset G_{y,0+} \subset G_x$. Moreover, $G_{y,r} \subset G_x$ for any $r > 0$ and $G_{y,r+} \supset G_{x,(r+1)+} \supset G_{x,r_\circ^+}$. where $r_\circ := \lceil r+1 \rceil \in \mathbb{Z}$.

4.3 Mackey's Theorem for Compact Induction

Besides the local constancy of characters, we are going to need the classical Mackey's theorem in the context of compactly induced representations.

Lemma 4.3. *Let $J \subset G$ be an open compact mod center subgroup and $H \subset G$ a closed subgroup. Let (J,ρ) and (H,τ) be smooth representations such that* $\dim \rho < \infty$. *Then*

$$\mathrm{Hom}_G(c\text{-ind}_J^G \rho, \mathrm{Ind}_H^G \tau) \simeq \bigoplus_{g \in H\backslash G/J} \mathrm{Hom}_{J\cap H^g}(\rho, \tau^g).$$

In fact it is a canonical isomorphism. A natural map will be constructed in the proof below.

Proof. Since the details are in [Kut77], where a more general result is proved, we content ourselves with outlining the argument. Let $S(\rho,\tau)$ denote the space of functions $s : G \to \mathrm{End}_{\mathbb{C}}(\rho,\tau)$ such that $s(hgj) = \tau(h)s(g)\rho(j)$ for all $h \in H$, $g \in G$, and $j \in J$. For each $g \in H\backslash G/J$ define $S_x(\rho,\tau)$ to be the subspace of $s \in S(\rho,\tau)$ such that $\mathrm{Supp}\, s \subset HxJ$. Clearly $S(\rho,\tau) = \oplus_{g\in H\backslash G/J} S_g(\rho,\tau)$. For each $v \in \rho$ we associate $f_v \in S(\rho,\tau)$ such that $f_v(j) = \rho(j)v$ if $j \in J$ and $f_v(j) = 0$ if $j \notin J$. We define a map

$$\mathrm{Hom}_G(\text{c-ind}_J^G \rho, \mathrm{Ind}_H^G \tau) \to S(\rho,\tau), \quad \phi \mapsto s_\phi$$

such that $s_\phi(g)$ sends $v \in \rho$ to $(\phi(f_v))(g)$. We also have a map

$$S_g(\rho,\tau) \to \mathrm{Hom}_{J\cap H^g}(\rho,\tau^g), \quad s \mapsto s(g).$$

(It is readily checked that $s(g) \in \mathrm{Hom}_{J\cap H^g}(\rho,\tau^g)$.) It is routine to check that the two displayed maps are isomorphisms. □

The following corollary was observed in [Nev13, Lemma 4.1].

Corollary 4.4. *In the setup of Lemma 4.3, further assume that* $c\text{-}\mathrm{ind}_J^G \rho$ *is admissible and that* H *is an open compact subgroup. Then*

$$\left(c\text{-}\mathrm{ind}_J^G \rho\right)|_H \simeq \bigoplus_{g \in J\backslash G/H} c\text{-}\mathrm{ind}_{H\cap J^g}^H \rho^g.$$

Proof. Let (H, τ) be an admissible representation. Then by Frobenius reciprocity and the preceding lemma,

$$\mathrm{Hom}_H(\text{c-ind}_J^G \rho, \tau) \simeq \mathrm{Hom}_G(\text{c-ind}_J^G \rho, \mathrm{Ind}_H^G \tau) \simeq \bigoplus_{g \in H\backslash G/J} \mathrm{Hom}_{J\cap H^g}(\rho, \tau^g).$$

Further, through conjugation by g and Frobenius reciprocity, the summand is isomorphic to

$$\mathrm{Hom}_{J^{g^{-1}}\cap H}(\rho^{g^{-1}}, \tau) \simeq \mathrm{Hom}_H\left(\text{c-ind}_{J^{g^{-1}}\cap H}^H \rho^{g^{-1}}, \tau\right).$$

The proof is finished by replacing g by g^{-1} in the sum and applying Yoneda's lemma. ◻

4.4 Main Estimates

This subsection establishes the main estimates towards of the proof of Theorem 4.2. Let Σ be a generic G-datum associated with π. That is, $\pi \simeq \pi_\Sigma$. In this entire subsection we keep conditions (1), (2), and (3) of Sect. 4.2.

Lemma 2.9 implies that $\deg(\pi) \to \infty$ is equivalent to $\mathrm{vol}_{G/Z}(J_\Sigma) \to 0$, which is in turn equivalent to $r_\Sigma \to \infty$. Henceforth we will often drop the subscript Σ when the context is clear.

Lemma 4.5. *Let* Σ, π, *and* γ *be as above. For simplicity, we write* r *for* r_Σ. *Suppose*

$$\mathrm{sd}(\gamma) \le \frac{r}{2}.$$

Then we have

$$\Theta_\pi(\gamma) = \mathrm{Tr}\left(\pi(\gamma)\Big|V_\pi^{G_{y,r+}}\right). \tag{4}$$

for any $y \in \mathcal{A}(T^\gamma) = \mathcal{B}(T^\gamma)$. *Here* $V_\pi^{G_{y,r+}}$ *is the space of* $G_{y,r+}$*-invariants in* V_π.

Remark 4.6. Recall by Sect. 4.2(2), the depth of π_Σ is $r = r_\Sigma$. Note that $\gamma \in G_{[y]}$. Hence, γ normalizes $G_{y,r+}$, and the right-hand side of the formula in (4) is well defined.

Proof. Given a subset $X \subset G$, let ch_X denote the characteristic function of X. By Adler and Korman [AK07, Corollary 12.9] (see also [MS12]) Θ_π is constant on $\gamma G_{y,r+} \subset {}^{G_{y,0+}}\left(\gamma T^\gamma_{r+}\right)$.[1] Thus we have

$$\begin{aligned}\Theta_\pi(\gamma) &= \tfrac{1}{\mathrm{vol}_G(G_{y,r+})}\int_G \Theta_\pi(g)\mathrm{ch}_{\gamma G_{y,r+}}\,dg\\ &= \tfrac{1}{\mathrm{vol}_G(G_{y,r+})}\mathrm{Tr}(\pi(\mathrm{ch}_{\gamma G_{y,r+}})) = \mathrm{Tr}\left(\pi(\gamma)\pi\left(\tfrac{\mathrm{ch}_{G_{y,r+}}}{\mathrm{vol}_G(G_{y,r+})}\right)\right)\\ &= \mathrm{Tr}\left(\pi(\gamma)|V_\pi^{G_{y,r+}}\right).\end{aligned}$$

The last equality follows from the fact that $\pi\left(\frac{\mathrm{ch}_{G_{y,r+}}}{\mathrm{vol}_G(G_{y,r+})}\right)$ is the projection of V_π onto $V_\pi^{G_{y,r+}}$. □

Our aim is to prove Proposition 4.15 below using Lemma 4.5. Recall $y \in \mathcal{A}(T^\gamma)$ is fixed. If $V_\pi^{G_{y,r+}} = 0$, we have $\Theta_\pi(\gamma) = 0$. Hence, from now on, we assume that $V_\pi^{G_{y,r+}} \neq 0$ without loss of generality. In the following series of lemmas, we first describe the space $V_\pi^{G_{y,r+}}$. The following result is originally due to Jacquet [Jac71].

Lemma 4.7. *Let J be an open compact mod center subgroup of G and ρ an irreducible representation of J such that $\pi = c\text{-}\mathrm{ind}_J^G\rho$ is irreducible (thus supercuspidal). Then, for any nontrivial unipotent subgroup N of G, we have $V_\rho^{N\cap J} = 0$.*

Proof. Applying Frobenius reciprocity and Lemma 4.3 with $H = N$,

$$0 = \mathrm{Hom}_N(\pi, 1_N) = \mathrm{Hom}_G(\text{c-ind}_J^G\rho, \mathrm{Ind}_N^G 1_N) \simeq \oplus_{g\in N\backslash G/J}\mathrm{Hom}_{J\cap N^g}(\rho, 1_{N^g}).$$

□

Let $J := J_\Sigma$ and $\rho := \rho_\Sigma$. We deduce from Corollary 4.4 that

$$\mathrm{Res}_{G_x}\text{c-ind}_J^G\rho \simeq \bigoplus_{g\in G_x\backslash G/J}\mathrm{Ind}_{G_x\cap {}^gJ}^{G_x}{}^g\rho. \tag{5}$$

Definition and Remark 4.8. (1) Define

$$\mathcal{X}'_\Sigma := \left\{g\in G \mid G_{g^{-1}x,r_\circ^+}\cap N \supset G_x\cap N \text{ for some unipotent subgroup } N \neq \{1\}\right\}$$

[1] Since γ is regular, the summation in [AK07, Corollary 12.9] runs over no nilpotent elements other than 0. So the corollary tells us that $\Theta_\pi(\gamma')$ is equal to a constant c_0 for all γ' in the G-conjugacy orbit of $\gamma + T_a^\gamma$ for $a > \max(2\mathrm{sd}(\gamma), \rho(\pi))$, where $\rho(\pi)$ denotes the depth of π, which is r. For $y \in \mathcal{A}(T^\gamma)$, we have $G_{y,a}$ contained in the G-orbit of $\gamma + T_a$. In our case $\max(2\mathrm{sd}(\gamma), \rho(\pi)) = r$, thus Θ_π is indeed constant on $\gamma G_{y,r+}$.

and

$$\mathcal{X}_\Sigma := G - \mathcal{X}'_\Sigma.$$

We observe that

(a) $G_x \cap N = G_{[x]} \cap N \supset J \cap N$ for any unipotent subgroup N, and
(b) $\mathcal{X}'_\Sigma$, $\mathcal{X}_\Sigma$ are left and right $G_{[x]}$-invariant.

(2) Suppose r is sufficiently large so that $G_{y,r+} \subset G_x$. Set

$$\mathcal{X}^\circ_\Sigma := \left\{ g \in G \,\middle|\, \left(\mathrm{Ind}_{G_x \cap {}^gJ}^{G_x} {}^g\rho\right)^{G_{y,r+}} \neq 0 \right\}.$$

In this case, from Lemma 4.5 and the above definition it is clear that

$$\Theta_\pi(\gamma) = \sum_{g \in G_x \backslash \mathcal{X}^\circ_\Sigma / J} \mathrm{Tr}\left(\pi(\gamma) \,\middle|\, \left(\mathrm{Ind}_{G_x \cap {}^gJ}^{G_x} {}^g\rho\right)^{G_{y,r+}}\right). \tag{6}$$

Note that $\mathcal{X}^\circ_\Sigma$ is right J-invariant, and also left G_x-invariant.

Lemma 4.9. *If $g \in \mathcal{X}'_\Sigma$, then the space $\mathrm{Ind}_{G_x \cap {}^gJ}^{G_x} {}^g\rho$ has no nonzero $G_{y,r+}$-invariant vector. That is, $\mathcal{X}^\circ_\Sigma \cap \mathcal{X}'_\Sigma = \emptyset$, or equivalently $\mathcal{X}^\circ_\Sigma \subset \mathcal{X}_\Sigma$.*

Proof. Recall that $G_{y,r} \subset G_x$ (see Sect. 4.2) and hence $G_{y,r+} \subset G_x$. Another application of Mackey's formula yields

$$\mathrm{Res}_{G_{y,r+}} \mathrm{Ind}_{G_x \cap {}^gJ}^{G_x} {}^g\rho \simeq \bigoplus_{h \in G_{y,r+} \backslash G_x / G_x \cap {}^gJ} \mathrm{Ind}_{G_{y,r+} \cap {}^{hg}J}^{G_{y,r+}} {}^{hg}\rho.$$

(This is derived from the formula in representation theory of finite groups since $[G_x : G_x \cap gJg^{-1}]$ and $\dim \rho$ are finite. We do not need Corollary 4.4.) By Frobenius reciprocity and conjugation by hg, we obtain

$$\mathrm{Hom}_{G_{y,r+}}\left(1, \mathrm{Ind}_{G_{y,r+} \cap {}^{hg}J}^{G_{y,r+}} {}^{hg}\rho\right) \simeq \mathrm{Hom}_{G_{y,r+} \cap {}^{hg}J}\left(1, {}^{hg}\rho\right) \simeq \mathrm{Hom}_{G_{g^{-1}h^{-1}y,r+} \cap J}(1, \rho).$$

Since $G_{y,r+} \supset G_{x,r_\circ^+}$ and $h \in G_x$, we have $G_{h^{-1}y,r+} \supset G_{h^{-1}x,r_\circ^+} = G_{x,r_\circ^+}$ and thus $G_{g^{-1}h^{-1}y,r+} \supset G_{g^{-1}x,r_\circ^+}$. It suffices to verify that the last Hom space is zero. If $g \in \mathcal{X}'_\Sigma$, then $G_{g^{-1}x,r_\circ^+} \cap N \supset J \cap N$ for some N and thus $G_{g^{-1}h^{-1}y,r+} \cap J \supset J \cap N$. This and Lemma 4.7 imply that the Hom space indeed vanishes. □

For simplicity, we will write $\mathcal{X}^\circ$, $\mathcal{X}$, and $\mathcal{X}'$ for $\mathcal{X}^\circ_\Sigma$, $\mathcal{X}_\Sigma$, and $\mathcal{X}'_\Sigma$ when the context is clear. For the purpose of our character computation, it is natural to estimate the cardinality of $G_x \backslash \mathcal{X}^\circ_\Sigma / J$ in view of (6). Instead we bound the size of $G_x \backslash \mathcal{X}_\Sigma / J$, which is larger by the preceding lemma but easier to control. To this end we begin by setting up some notation for the Cartan and Iwahori decompositions.

Notation 4.10. Let $\mathbf{T}^0 \subset \mathbf{G}^0$ and $\mathbf{T} \subset \mathbf{G}$ be maximal and maximally k-split tori such that $x \in \mathcal{A}(\mathbf{T}^0, k) \subset \mathcal{A}(T) := \mathcal{A}(\mathbf{T}, \mathbf{G}, k) \subset \mathcal{B}(G)$. Let C be a facet of maximal dimension in $\mathcal{A}(T)$ with $x \in \overline{C}$. Let Δ be the set of simple T-roots associated with C, and N_Δ the maximal unipotent subgroup with simple roots Δ. Let $G_C \subset G_x$ be the Iwahori subgroup fixing C so that $G = G_C N_G(T) G_C$ and G_z be a special maximal parahoric subgroup with $z \in \overline{C}$ such that $G_C \subset G_z$. Note that elements in $W := N_G(T)/C_G(T)$ can be lifted to elements in G_z. Let $T^- := \{t \in T \mid tUt^{-1} \subset U$ for any open subgroup U in $N_\Delta\}$ so that we have a Cartan decomposition $G = G_z T^- G_z$.

Lemma 4.11. *Let* $T^-(r_\circ) := \{t \in T^- \mid 1 \le |\alpha(t^{-1})| \le q^{r_\circ+2},\ \alpha \in \Delta\}$. *Then*

$$\mathcal{X}^\circ \subset \mathcal{X} \subset G_z T^-(r_\circ) G_z \tag{7}$$

Proof. The first inclusion is from Lemma 4.9. For the second inclusion, for $v \in \mathcal{B}(G)$ and $a \in \mathbb{R}_{\ge 0}$, let

$$\mathcal{X}'(v,a) = \left\{g \in G \mid G_{g^{-1}v,a^+} \cap N \supset G_v \cap N \text{ for some unipotent subgroup } N \ne \{1\}\right\}$$
$$\mathcal{X}(v,a) = G - \mathcal{X}'(v,a).$$

It is enough to show that

$$(G - G_z T^-(r_\circ) G_z) \subset \mathcal{X}'(z, r_\circ + 1) \subset \mathcal{X}'(x, r_\circ) = \mathcal{X}'. \tag{8}$$

For the second inclusion in (8), let $g \in \mathcal{X}'(z, r_\circ + 1)$. Since $x, z \in \overline{C}$ and $G_{g^{-1}z,(r_\circ+1)^+} \cap N \supset G_z \cap N$, we have $G_{g^{-1}x,r_\circ^+} \supset G_{g^{-1}z,(r_\circ+1)^+} \supset G_{g^{-1}z,(r_\circ+1)^+} \cap N \supset G_z \cap N \supset G_x \cap N$. Hence, $g \in \mathcal{X}'(x, r_\circ)$.

For the first inclusion in (8), let $g = g_1 t' g_2 \in G - G_z T^-(r_\circ) G_z$ with $g_i \in G_z$ and $t' \in T - T^-(r_\circ)$. Then, there is $\alpha \in \Delta$ such that $|\alpha(t'^{-1})| > q^{r_\circ+2}$. Let N_α be the maximal unipotent subgroup associated with α. Then,

$$G_{t'^{-1}z,(r_\circ+1)^+} \cap N_\alpha \supset G_z \cap N_\alpha.$$

Since we have $G_{g^{-1}z,(r_\circ+1)^+} = G_{g_2^{-1}t'^{-1}z,(r_\circ+1)^+} = {}^{g_2^{-1}}G_{t'^{-1}z,(r_\circ+1)^+}$ and $G_{g_2 z} = G_z$, we have

$$G_{g^{-1}z,(r_\circ+1)^+} \cap {}^{g_2^{-1}}N_\alpha \supset G_z \cap {}^{g_2^{-1}}N_\alpha, \quad \text{thus} \quad G_{t'^{-1}z,(r_\circ+1)^+} \cap N_\alpha \supset G_z \cap N_\alpha.$$

We conclude that $g \in \mathcal{X}'(z, r_\circ + 1)$. □

To give another description of $\mathcal{X}^\circ$, we define compact mod center sets $S_{x,y} \subset \mathcal{S}_{x,y}$ as follows:

$$S_{x,y} := G_{[y]} S_x G_{y,0^+}, \qquad \mathcal{S}_{x,y} := G_x S_{x,y} G_x.$$

In particular the quotient $(Z_G G_x)\backslash\mathcal{S}_{x,y}$ is finite. Recall that S_x is the set constructed in the proof of Lemma 3.1.

Lemma 4.12. *Suppose* (HB) *and* (HGT) *are valid. Then, for any double coset* $G_x gJ \subset \mathcal{X}^\circ$*, we have*

$$G_x g G_x \cap S_{x,y}\, G^{d-1} G_{x,0+} \neq \emptyset.$$

Proof. Since $G_{y,r} \subset G_x$ by Sect. 4.2, Mackey's formula gives us, as in the proof of Lemma 4.9, that

$$\mathrm{Res}_{G_{y,r}} \mathrm{Ind}_{G_x \cap {}^g J}^{G_x} {}^g\rho \simeq \bigoplus_{\ell \in G_{y,r}\backslash G_x / G_x \cap {}^g J} \mathrm{Ind}_{G_{y,r} \cap {}^{\ell g}J}^{G_{y,r}} {}^{\ell g}\rho,$$

Since $g \in \mathcal{X}^\circ$ there is $\ell \in G_x$ such that $\left(\mathrm{Ind}_{G_{y,r}\cap {}^{\ell g}J}^{G_{y,r}} {}^{\ell g}\rho\right)^{G_{y,r+}} \neq 0$. By replacing g with $\ell^{-1}g$ if necessary, we may assume $\left(\mathrm{Ind}_{G_{y,r}\cap {}^{g}J}^{G_{y,r}} {}^{g}\rho\right)^{G_{y,r+}} \neq 0$. Let $X \in \mathfrak{z}_{\mathfrak{g}^{d-1}}$ be a good element representing ϕ_{d-1}. Let $(G_{y,r}, \phi)$ be a minimal K-type appearing in $(\mathrm{Ind}_{G_{y,r}\cap {}^{g}J}^{G_{y,r}} {}^{g}\rho)^{G_{y,r+}}$. Then, by Lemma 3.1, there are $h \in G_{[y]}S_x$ and $\eta \in \mathfrak{g}^{d-1}_{(-r)+}$ such that ϕ is represented by ${}^h(X+\eta)$. On the other hand, $G_{gx,r} \subset {}^gJ$ and ${}^g\rho|G_{gx,r}$ is a self-direct sum of ${}^g\phi_{d-1}$. Therefore ϕ is a $(G_{y,r} \cap G_{gx,r})$-subrepresentation of such a self-direct sum. This means that $\phi = {}^g\phi_{d-1}$ on $G_{y,r} \cap G_{gx,r}$. Equivalently, $\left({}^h(X+\eta) + \mathfrak{g}_{y,(-r)+}\right) \cap {}^g(X + \mathfrak{g}_{x,(-r)+}) \neq \emptyset$ in terms of dual cosets. By [KM03, Corollary 2.3.5], this is in turn equivalent to ${}^{hG_{h^{-1}x,0+}}(X + \eta + \mathfrak{g}^{d-1}_{h^{-1}y,(-r)+}) \cap {}^{gG_{x,0+}}(X + \mathfrak{g}^{d-1}_{x,(-r)+}) \neq \emptyset$. This implies $h^{-1}g \in G_{h^{-1}y,0+} C_G(X) G_{x,0+}$ by [KM03, Lemma 2.3.6]. Hence, $g \in G_{y,0+} h C_G(X) G_{x,0+}$. It follows that $h^{-1}g \in G_{y,0+} G^{d-1} G_{x,0+}$. Hence, $g \in G_{[y]} S_x G_{y,0+} G^{d-1} G_{x,0+} \subset S_{x,y} G^{d-1} G_{x,0+}$. □

Observe that $G_z T^-(r_\circ) G_z = G_C W T^-(r_\circ) W G_C$, and $G_C \subset G_x$. Combining these with Lemmas 4.11 and 4.12,

$$\mathcal{X}^\circ \subset (\mathcal{S}_{x,y}\, G^{d-1} G_x) \cap (G_x W T^-(r_\circ) W G_x) \tag{9}$$

when (HB) and (HGT) are valid. We note that $\mathcal{S}_{x,y}$ depends only on $(G, G' = G^{d-1}, x)$ and y.

Proposition 4.13. *The double coset space* $G_x\backslash\mathcal{X}^\circ/J$ *is finite. More precisely, setting* $L_s := G^{d-1}_{[x]} G_{x,s}$*, we have*

$$|G_x\backslash\mathcal{X}^\circ/J| \leq C_{x,G'} \cdot \sharp(W)^2 (r_\circ + 3)^{r_G} [L_s : J] \cdot \mathrm{vol}_{G/Z_G}(L_s)^{-\frac{1}{2}}, \tag{10}$$

for some constant $C_{x,G'} > 0$ *(which may be chosen explicitly; see Lemma 4.14 below).*

Proof. Note that each $T^{-}(r_\circ)/(C_G(T)_0Z_G)$ is finite where $C_G(T)_0$ is the maximal parahoric subgroup of $C_G(T)$ and $\sharp\,(T^{-}(r_\circ)/(C_G(T)_0Z_G)) \le (r_\circ + 3)^{r_G}$. Hence, $\sharp\left(G_z\backslash\mathcal{X}/(G_{[z]})\right) \le (r_\circ+3)^{r_G}$ by (7). Since $G_CN_G(T)G_C = G_xN_G(T)G_x = G_zT^{-}G_z$ and $N_G(T)/(C_G(T)_0Z) = W(T^{-}/(C_G(T)_0Z))W$, we have

$$\sharp(G_x\backslash\mathcal{X}^\circ/G_{[x]}) \le (\sharp W)^2(r_\circ+3)^{r_G}.$$

Since $|G_x\backslash\mathcal{X}^\circ/J| = \sum_{x\in G_x\backslash\mathcal{X}^\circ/G_{[x]}} \left|G_x\backslash G_xgG_{[x]}/J\right|$, the proof of (10) is completed by the following lemma:

Lemma 4.14. *If $G_xgJ \subset \mathcal{X}^\circ$, then*

$$\left|G_x\backslash G_xgG_{[x]}/J\right| \le C_{x,G'}\cdot[L_s:J]\,\mathrm{vol}_{G/Z_G}(L_s)^{-\frac{1}{2}}$$

where $C_{x,G'} := \mathrm{vol}_{G/Z_G}(G_{[x]})^{1/2}\cdot\sharp\left(Z_GG_x\backslash\mathcal{S}_{x,y}\right)\cdot q^{(2\dim_k(G)+\dim_k(N))}\,\sharp\left(G_{[x]}/(Z_GG_x)\right)$.

Proof. By (9), there is a $w\in G^{d-1}$ such that $G_xgG_x \subset \mathcal{S}_{x,y}wG_x$. Then

$$\left|G_x\backslash G_xgG_{[x]}/J\right| \le \left|G_x\backslash\mathcal{S}_{x,y}wG_{[x]}/J\right| \le [L_s:J]\cdot\sharp\big((Z_GG_x\backslash\mathcal{S}_{x,y}\big)\cdot\sharp\big(G_xwG_{[x]}/L_s\big)\,.$$

Let $\mathbf{T}$ be as before. Let $\mathbf{T}^{d-1}$ be a maximal and maximally k-split torus of $\mathbf{G}^{d-1}$ such that the k-split components T_k^0 and T_k^{d-1} of T^0 and T^{d-1}, respectively, satisfy that $T_k^0 \subset T_k^{d-1} \subset T$ and $x\in\mathcal{A}(\mathbf{T}^0)\subset\mathcal{A}(\mathbf{T}^{d-1})\subset\mathcal{A}(\mathbf{T})$. By the Iwahori decomposition of G^{d-1} one may write $w = u_1w_0u_2$ with $u_1,u_2\in G_x^{d-1}$ and $w_0\in N_{G^{d-1}}(T^{d-1})$. Replacing w with w_0 if necessary, one may assume that $w\in N_{G^{d-1}}(T^{d-1})$ since this doesn't change $\mathcal{S}_{x,y}wG_x$.

It is enough to show that there is a unipotent subgroup U such that

$$\left|G_x\backslash G_xwG_{[x]}/L_s\right| \le q^{2\dim(G)}\cdot\sharp\left(G_{[x]}/(Z_GG_x)\right)\cdot[(U\cap G_{[x]}):(U\cap L_s)].$$

Indeed if the inequality is true, since $G_{[x]}\cap U = G_x\cap U$, Lemma 3.10 applied to U implies that

$$\begin{aligned}[(U\cap G_{[x]}):(U\cap L_s)] &\le q^{\dim_k N}[G_{[x]}:L_s]^{1/2}\\ &\le \mathrm{vol}_{G/Z_G}(G_{[x]})^{1/2}q^{\dim_k N}\mathrm{vol}_{G/Z_G}(L_s)^{-1/2},\end{aligned}$$

ending the proof. (As x is fixed, $C_0 = \mathrm{vol}_{G/Z_G}(G_{[x]})^{1/2}$ depends only on the Haar measure of G.)

It remains to find a desired U. Let $\mathbf{M} = C_{\mathbf{G}}(\mathbf{T}_k^{d-1})$ (resp. $\mathbf{M}^{d-1} := C_{\mathbf{G}^{d-1}}(\mathbf{T}_k^{d-1})$) be the minimal Levi subgroup of $\mathbf{G}$ (resp. $\mathbf{G}^{d-1}$) containing $\mathbf{T}^{d-1}$. Write $w := w_0t$ with w_0 in a maximal parahoric subgroup of G^{d-1} containing G_x^{d-1} and $t\in T_0^{d-1}$. Let $\mathbf{L}$, $\mathbf{U}$, and $\overline{\mathbf{U}}$ associated with t as in [Del76] (so that our $\mathbf{L}$, $\mathbf{U}$, and $\overline{\mathbf{U}}$ are his M_g, U_g^+, U_g^- for $g=t$). That is, $\mathbf{L} = \{\ell\in G \mid \{{}^{t^n}\ell\}_{n\in\mathbb{Z}}$ is bounded$\}$, $\mathbf{U} = \{u\in G \mid {}^{t^n}u\to\infty$ as $n\to\infty\}$ and $\overline{\mathbf{U}} = \{\overline{u}\in G \mid {}^{t^n}\overline{u}\to 1$ as $n\to\infty\}$. Then, U and

$\overline{U}$ are opposite unipotent subgroups with respect to the Levi subgroup L. Note that $\mathbf{M} \subset \mathbf{L}$. Now, since $G_{x,0+}$ and $G^{d-1}_{x,0+} G_{x,s}$ are decomposible with respect to $L, U, \overline{U}$, we have

$$\begin{aligned}\left|G_x\backslash G_x w G_{[x]}/L_s\right| &\le q^{\dim_k(G)} \sharp \left(G_{[x]}/(Z_G G_x)\right) \left|G_x\backslash G_x w G_{x,0+}/(G_{x,0+} \cap L_s)\right| \\ &\le q^{2\dim_k(G)} \sharp \left(G_{[x]}/(Z_G G_x)\right) [G_{x,0+} \cap U : (G^{d-1}_{x,0+} G_{x,s}) \cap U] \\ &\le q^{2\dim_k(G)} \sharp \left(G_{[x]}/(Z_G G_x)\right) [G_x \cap U : L_s \cap U].\end{aligned}$$

□

4.5 Proof of the Main Theorems

This subsection is devoted to the proof of Theorem 4.2 and its quantitative version in Theorem 4.18. Combining the estimates of the previous subsection, we have the following:

Proposition 4.15. *Suppose* (HB), (HGT)*, and* (Hk) *are valid. Let* $\pi = \pi_\Sigma$. *Let* $\gamma \in G_{0+} \cap G_{\mathrm{reg}}$ *with* $\mathrm{sd}(\gamma) \le \frac{r}{2}$. *Then, we have*

$$\left|\frac{\Theta_\pi(\gamma)}{\deg(\pi)}\right| \le C_1 \cdot (\sharp W)^2 \cdot q^{\dim(G)+r_G(A_{\gamma,\Sigma}+1)} \cdot D(\gamma)^{-1} \cdot (r+4)^{r_G} \mathrm{vol}_{G/Z}(L_s)^{\frac{1}{2}}, \qquad (11)$$

where $C_1 = \max\{C_{x,G'}\}$ *where the maximum runs over the finitely many G-orbits of* (x, G') *and* $C_{x,G'}$ *is the constant as in Lemma 4.14.*

Proof. Without loss of generality, we may reduce to cases as in Sect. 4.2. In the following, $\psi : G_x \to G_x$ is the map defined by $\psi(g) = g\gamma g^{-1}$ and for $g \in G_x$, $H_g := G_x \cap {}^gJ$. Our starting point is formula (6) computing $\Theta_\pi(\gamma)$. The summand for each g has a chance to contribute to the trace of γ only if ${}^{g'}\gamma \in G_x \cap {}^gJ$ for some $g' \in G_x$. In the summand for g, decompose $\mathrm{ind}^{G_x}_{G_x \cap {}^gJ} {}^g\rho$ as the direct sum of the spaces of functions supported on exactly one left $G_x \cap {}^gJ$-coset in G_x. The element γ permutes the spaces by translating the functions by γ on the right. So it is easy to see that each space may contribute to the trace of γ on $V_\pi^{G_{y,r+}}$ only if the supporting $G_x \cap {}^gJ$-coset is fixed by γ. Hence we have

$$\begin{aligned}\left|\frac{\Theta_\pi(\gamma)}{\deg(\pi)}\right| &= \frac{\mathrm{vol}_{G/Z_G}(J)}{\dim\rho} \left|\mathrm{Tr}\left(\pi(\gamma)|V_\pi^{G_{y,r+}}\right)\right| \\ &\le \frac{\mathrm{vol}_{G/Z_G}(J)}{\dim\rho} \sum_{\substack{g \in G_x\backslash \mathcal{X}^\circ/J \\ \text{s.t. } {}^{g'}\gamma \in H_g}} \#\mathrm{Fix}(\gamma|(G_x \cap {}^gJ)\backslash G_x) \cdot \dim\rho\end{aligned}$$

$$= \frac{\mathrm{vol}_{G/Z_G}(J)}{\dim\rho} \sum_{\substack{g\in G_x\backslash\mathcal{X}^\circ/J \\ \text{s.t. } {}^{g'}\gamma\in H_g}} [\psi^{-1}(H_g):H_g]\dim\rho$$

$$= \mathrm{vol}_{G/Z_G}(J) \sum_{\substack{g\in G_x\backslash\mathcal{X}^\circ/J \\ \text{s.t. } {}^{g'}\gamma\in H_g}} [\psi^{-1}(H_g):H_g]$$

$$\leq C_1 \cdot q^{\dim(G)+r_G(A_{\gamma,\Sigma}+1)} \cdot D(\gamma)^{-1} \cdot (\sharp W)^2 (r_\circ + 3)^{r_G} [L_s : J]$$

$$\cdot \mathrm{vol}_{G/Z_G}(L_s)^{-\frac{1}{2}} \mathrm{vol}_{G/Z_G}(J)$$

$$= C_1 \cdot q^{\dim(G)+r_G(A_{\gamma,\Sigma}+1)} \cdot D(\gamma)^{-1} \cdot (\sharp W)^2 (r+4)^{r_G} \mathrm{vol}_{G/Z_G}(L_s)^{1/2}.$$

The $\sum$ above runs over $g \in G_x\backslash\mathcal{X}^\circ/J$ such that ${}^{g'}\gamma \in H_g$ for some $g' \in G_x$. The second last inequality follows from Corollary 3.9 and Proposition 4.13. The last equality follows from $[L_s : J]\mathrm{vol}_{G/Z_G}(J) = \mathrm{vol}_{G/Z_G}(L_s)$. □

Proof of Theorem 4.2. Let $\pi_i = \pi_{\Sigma_i}$, $i = 1, 2, \cdots$ be a sequence of supercuspidal representations in $\mathrm{Irr}^{\mathrm{Yu}}(G)$ with $\deg(\pi_i) \to \infty$. It is enough to consider such a sequence since there are countably many isomorphism classes of irreducible supercuspidal representations up to character twists. Recall from Sect. 4.2, we may assume $x_{\Sigma_i} \in \Sigma_i$ is in a fixed G-orbit. Since $\deg(\pi_i) \to \infty$ as $i \to \infty$, we have $r_{\Sigma_i} \to \infty$ as $i \to \infty$ and $r_{\Sigma_i} > 2\mathrm{sd}(\gamma)$ for almost all π_i. Hence, (11) holds for π with i large enough. Note that in (11), when $h_G > 1$, only $(r_{\Sigma_i}+4)^{r_G} \cdot q^{r_G s}\mathrm{vol}_{G/Z_G}(L_{s_{\Sigma_i}})^{\frac{1}{2}}$ varies as π_i varies with i large enough. It suffices to show that this quantity approaches zero as $r_{\Sigma_i} \to \infty$. Since the term $(r+4)^{r_G}$ has a polynomial growth in r while $q^{r_G s}\mathrm{vol}_{G/Z}(L_s)$ decays exponentially as $s = r/2$ tends to infinity by Lemma 4.17. The case $h_G = 1$ is similar, hence we are done. □

Remark 4.16. It is also interesting to discuss the role of the subgroup $G_{y,r+}$ in our proof. This subgroup appears in Lemma 4.5 from the local constancy of the character Θ_π. Of course for the purpose of that lemma any open subgroup $K \subset G_{y,r+}$ would work. However, from the fact that $G_{y,r}$-representations in $V_\pi^{G_{y,r+}}$ are minimal K-types, we acquire another description of $\mathcal{X}^\circ$ as in Lemma 4.12, which is again used to get an estimate in Lemma 4.14.

In the remainder of this section we upgrade Theorem 4.2 to a uniform quantitative statement. From here on we may and will normalize the Haar measure on G/Z such that $\mathrm{vol}(G_{[x]}/Z) = 1$. This is harmless because there are finitely many conjugacy classes of $(\vec{\mathbf{G}}, x)$ as explained in Sect. 4.2.

Lemma 4.17. *(i) If $h_G > 1$ and G has irreducible root datum, there exists a constant $\kappa > 0$ such that for all $\pi = \pi_\Sigma \in \mathrm{Irr}^{\mathrm{Yu}}(G)$,*

$$q^{r_G s}\mathrm{vol}(L_s)^{1/2} \leq q^{\dim G} \cdot \deg(\pi)^{-\kappa}.$$

(ii) *If* $h_G = 1$, *there exists a constant* $\kappa > 0$ *such that for all* $\pi = \pi_\Sigma \in \mathrm{Irr}^{\mathrm{Yu}}(G)$*,*

$$\mathrm{vol}(L_s)^{1/2} \le q^{\dim G} \cdot \deg(\pi)^{-\kappa}.$$

(iii) *In general, suppose G is reductive. There exists a constant* $\kappa > 0$ *such that for all* $\pi = \pi_\Sigma \in \mathrm{Irr}^{\mathrm{Yu}}(G)$*,*

$$q^{r_G s}\mathrm{vol}(L_s)^{1/2} \le q^{\dim G} \cdot \deg(\pi)^{-\kappa}.$$

Proof. (i) Observe that

$$\begin{aligned}
\mathrm{vol}(J/Z)^{-1} &= [G_{[x]} : J] \le [ZG_x : ZG_{x,s}] = [G_x : (Z \cap G_x)G_{x,s}] \le q^{(\dim G - \dim Z)s},\\
\mathrm{vol}(L_s/Z)^{-1} &= [G_{[x]} : L_s] \ge [G_{[x]} : G'_{[x]}G_{x,s}] \ge [G_{x,0+} : G'_{x,0+}G_{x,s}]\\
&\ge q^{(\dim G - \dim G')(s-1)},
\end{aligned}$$

with $G' = G^{d-1}$. Recall that $\deg(\pi) \le q^{\dim G}\mathrm{vol}(J/Z)^{-1}$. Take

$$\kappa := \min_{G' \subsetneq G} \frac{\dim G - \dim G' - 2r_G}{2(\dim G - \dim Z)},$$

where G' runs over the set of proper tamely ramified twisted Levi subgroups of G. If we know $\kappa > 0$ then the lemma follows from the following chain of inequalities:

$$\begin{aligned}
\deg(\pi)^\kappa &\le q^{\kappa \dim G} q^{(\dim G - \dim G' - 2r_G)s/2} \le q^{\dim G/2} q^{(\dim G - \dim G' - 2r_G)s/2}\\
&\le q^{\dim G} \cdot q^{-r_G s} \cdot \mathrm{vol}(L_s/Z)^{-1/2}.
\end{aligned}$$

It remains to show that $\kappa > 0$. Since $\dim \mathbf{G} = \dim_k G$, it is enough to show that $\dim \mathbf{G} - \dim \mathbf{M} - 2r_G > 0$ when $\mathbf{M}$ is a proper Levi subgroup which arises in a supercuspidal datum. This can be seen as follows. If $\mathbf{G}$ of type other than A, the inequality holds for any proper Levi subgroup $\mathbf{M}$. If $\mathbf{G}$ is of type A_n, then $\kappa > 0$ unless $\mathbf{M}$ is of type A_{n-1}. However such a Levi subgroup does not arise as part of supercuspidal datum when $n \ge 2$, and the assumption $h_G > 1$ excludes the case $n = 1$.

(ii) In this case, we can take $\kappa = \min_{G' \subsetneq G} \frac{\dim G - \dim G'}{2(\dim G - \dim Z)}$. It is clear that $\kappa > 0$ and the rest of the proof works as in (i).

(iii) This follows from (i) and (ii). □

Since $J \subset L_s$, the above proof implies the lower bound $\mathrm{vol}(J/Z)^{-1} \ge q^{(\dim G - \dim G')(s-1)}$. Combined with Lemma 2.9, this yields $\deg(\pi) \ge q^{(\dim G - \dim G')(s-1)}$. The following theorem is an improvement of Theorem 4.2.

Theorem 4.18. *Assume hypotheses* (HB)*,* (HGT)*, and* (Hk)*. There exist constants* $A, \kappa, C > 0$ *depending only on* G *such that the following holds. For every* $\gamma \in G_{0+} \cap G_{\mathrm{reg}}$ *and* $\pi \in \mathrm{Irr}^{\mathrm{Yu}}(G)$ *such that* $\mathrm{sd}(\gamma) \le r/2$*,*

$$D(\gamma)^A |\Theta_\pi(\gamma)| \leq C \cdot \deg(\pi)^{1-\kappa}. \tag{12}$$

Proof. Let $\pi := \pi_\Sigma \in \mathrm{Irr}^{\mathrm{Yu}}(G)$ and recall that $r = r_\Sigma$ is the depth of π. Since $\gamma \in G_0$ we have $\nu(1 - \alpha(\gamma)) \geq 0$ for all $\alpha \in \Phi(\mathbf{T}^\gamma)$. In view of Definition 3.3, we have

$$D(\gamma) = \prod_{\alpha \in \Phi(\mathbf{T}^\gamma)} |1 - \alpha(\gamma)| \leq q^{-\mathrm{sd}(\gamma)} \leq 1.$$

Proposition 4.15 yields a bound of the form

$$D(\gamma) \left| \frac{\Theta_\pi(\gamma)}{\deg(\pi)} \right| \leq C \cdot q^{r_G A_{\gamma,\Sigma}} \mathrm{vol}(L_s)^{1/2},$$

where $C \in \mathbb{R}_{>0}$ is a constant depending only on G. Consider the case that $h_G > 1$. Recall that $A_{\gamma,\Sigma} = h_G \mathrm{sd}(\gamma) + s$. Let us take $A := r_G h_G + 1$. Then

$$D(\gamma)^A \left| \frac{\Theta_\pi(\gamma)}{\deg(\pi)} \right| \leq q^{-r_G h_G \mathrm{sd}(\gamma)} D(\gamma) \left| \frac{\Theta_\pi(\gamma)}{\deg(\pi)} \right| \leq C \cdot q^{r_G \cdot s} \mathrm{vol}(L_s)^{1/2}.$$

Then, by Lemma 4.17(i), we have

$$D(\gamma)^A |\Theta_\pi(\gamma)| \leq C_0 \cdot \deg(\pi)^{1-\kappa}$$

with $C_0 := Cq^{\dim G}$ and the same κ as in that lemma, completing the proof when $h_G > 1$.

In the remaining case $h_G = 1$, we have $A_{\gamma,\Sigma} = \mathrm{sd}(\gamma)$. Take $A := r_G + 1$. Then

$$D(\gamma)^A \left| \frac{\Theta_\pi(\gamma)}{\deg(\pi)} \right| \leq q^{-r_G \mathrm{sd}(\gamma)} D(\gamma) \left| \frac{\Theta_\pi(\gamma)}{\deg(\pi)} \right| \leq C \cdot \mathrm{vol}(L_s)^{1/2}.$$

The proof is finished by applying Lemma 4.17(ii) and taking $C_0 = Cq^{\dim G}$ again. □

4.6 *On the Assumption* $\mathrm{sd}(\gamma) \leq r/2$

Theorem 4.18 above remains valid without the assumption $\mathrm{sd}(\gamma) \leq r/2$. Indeed the uniform estimate (12) holds in the range $r < 2\mathrm{sd}(\gamma)$ by a different argument that we now explain.[2]

[2]A priori we are proving the bound (12) in two disjoint regions with two different values of (A, κ); call them (A_1, κ_1) and (A_2, κ_2). When we say that Theorem 4.18 is valid without the assumption $\mathrm{sd}(\gamma) \leq r/2$, it means that there's a single choice of (A, κ) that works in both regions. This is immediate because $\gamma \in G_{0+}$, in which case it follows that $D(\gamma) \ll 1$. So it, enough to take $A = \max(A_1, A_2)$ and $\kappa = \min(\kappa_1, \kappa_2)$, possibly at the expense of increasing the constant C in (12).

It suffices to produce a polynomial upper-bound on the trace character $|\Theta_\pi(\gamma)|$ in the range $r < 2\mathrm{sd}(\gamma)$. By [AK07, Corollary 12.9] and [MS12] (as explained in the proof of Lemma 4.5) the character Θ_π is constant on $\gamma G_{y,t}$ if $t = 2\mathrm{sd}(\gamma) + 1$. The analogue of (4) holds, hence $|\Theta_\pi(\gamma)| \le \dim V_\pi^{G_{y,t}}$. It is sufficient to show under the same hypotheses (HB), (HGT), and (Hk) as above that for all $\pi \in \mathrm{Irr}^{\mathrm{Yu}}(G)$, $y \in \mathcal{B}(G)$, and $t \ge 1$,

$$\dim V_\pi^{G_{y,t}} \le q^{Bt} \tag{13}$$

for some constant $B > 0$ depending only on G. Indeed $D(\gamma) \le q^{-\mathrm{sd}(\gamma)}$ and $2s = r < t = 2\mathrm{sd}(\gamma) + 1$ so (12) follows with $A = 2B$ and $\kappa = 1$.

Similarly as in Definition 4.8 we introduce the subset $\mathcal{X}^\circ \subset G$. Mackey's decomposition implies that $\dim V_\pi^{G_{y,t}} \le \dim \rho \cdot [G_x : G_{y,t}] \cdot |G_x \backslash \mathcal{X}^\circ / J|$. The dimension $\dim \rho$ is uniformly bounded by Lemma 2.9, the term $[G_x : G_{y,t}]$ is polynomially bounded (see Lemma 4.17), as well as the third term by Proposition 4.13. This establishes (13).

5 Miscellanies

5.1 *Trace Characters Versus Orbital Integrals of Matrix Coefficients*

The bulk of the proof above was to establish a power saving as $\deg(\pi) \to \infty$. In fact we can develop two distinct approaches, the first of which is taken in this paper.

1. We have handled the trace characters $\Theta_\pi(g)$ using first their local constancy [SS97]. The proof then was by a uniform estimate on the irreducible factors of the restriction of $\text{c-ind}_J^G \rho$ to a suitable subgroup.
2. The other approach developed in [KST] is via the orbital integral $O_\gamma(\phi_\pi)$ of a matrix coefficient ϕ_π which can be written explicitly from Yu's construction. The proof is via a careful analysis of the conjugation by γ on J and uses notably a recent general decomposition theorem of Adler–Spice [AS08].

For γ regular *and* elliptic semisimple the orbital integral $O_\gamma(\phi_\pi)$ and the trace character $\Theta_\pi(\gamma)$ coincide as follows from the local trace formula of Arthur; in this case, our approaches (1) and (2) produce similar estimates. The approach (1), where γ is regular, is well suited to establish our proposed conjecture, while the approach (2), where γ is elliptic (but not regular), is well suited for application of the trace formula. Indeed the goal of [KST] is to establish properties of families of automorphic representations, similarly to [ST11], as we prescribe varying supercuspidal representations at a given finite set of primes.

5.2 Analogues for Real Groups

We would like to see the implication of Harish-Chandra's work on the analogue of Conjecture 4.1 for real groups. Only in this subsection, let G be a connected reductive group over $\mathbb{R}$. Write A_G for the maximal split torus in the center of G and put $A_{G,\infty} := A_G(\mathbb{R})^0$. Then $G(\mathbb{R})$ has discrete series if and only if G contains an elliptic maximal torus T over $\mathbb{R}$, namely a maximal torus T such that $T(\mathbb{R})/A_{G,\infty}$ is compact. Fix a choice of T and a maximal compact subgroup $K \subset G(\mathbb{R})$ such that $T \subset KA_{G,\infty}$. Let $W_{\mathbb{R}}$ denote the relative Weyl group for $T(\mathbb{R})$ in $G(\mathbb{R})$. Write $\mathfrak{t} := \mathrm{Lie}\, T(\mathbb{R})$ and $\mathfrak{t}^*$ for its linear dual. Set $q(G) := \frac{1}{2}\dim_{\mathbb{R}}(G(\mathbb{R})/K) \in \mathbb{Z}$. Let π be an (irreducible) discrete series of $G(\mathbb{R})$ whose central character is unitary, and denote by $\lambda_\pi \in i\mathfrak{t}^*$ its infinitesimal character. Let γ be a regular element of $T(\mathbb{R})$, which is uniquely written as $\gamma = z \exp H$ for $z \in K \cap Z(G(\mathbb{R}))$ and $H \in \mathrm{Lie}\, T(\mathbb{R})$.

Proposition 5.1. *The real group analogue of Conjecture 4.1 is verified for elliptic regular elements γ and discrete series with unitary central characters.*

Proof. Harish-Chandra's character formula for discrete series on elliptic maximal tori implies that (as usual $D(\gamma)$ is the Weyl discriminant)

$$D(\gamma)^{1/2}\Theta_\pi(\gamma) = (-1)^{q(G)} \sum_{w \in W_{\mathbb{R}}} \mathrm{sgn}(w) e^{\lambda_\pi(H)}.$$

Hence $D(\gamma)^{1/2}|\Theta_\pi(\gamma)| \leq |W_{\mathbb{R}}|$. ◻

Note that we have a much stronger version for part (ii) of the conjecture, allowing $\epsilon = 1$. To verify part (i) when γ is contained in a non-elliptic maximal torus, one can argue similarly by using the character formula due to Martens [Mar75] as far as holomorphic discrete series are concerned. A general approach would be to use a similar character formula as above, which exists but comes with a subtle coefficient in each summand which depends on w and π. The coefficients can be analyzed in two steps: firstly one studies the analogous coefficients for stable discrete series characters (as studied by Herb; also see [Art89, p. 273]), and secondly relates the character of a single discrete series of $G(\mathbb{R})$ to the stable discrete series characters on endoscopic groups of $G(\mathbb{R})$ following the idea of Langlands and Shelstad. For instance, this has been done in [Her83] (also see [Art89, p. 273] for the first step). On the other hand, Herb has another approach avoiding endoscopy in [Her98]. We do not pursue either approach further in this paper as it would take us too far afield.

5.3 Analogues for Finite Groups

Let G be a finite group of Lie type over a finite field with $q \geq 5$ elements. Gluck [Glu95] has shown that if π is a nontrivial irreducible representation of G and γ is a noncentral element, then the trace character satisfies

$$|\chi_\pi(\gamma)| \leq \frac{\dim(\pi)}{\sqrt{q}-1}.$$

The bound has interesting applications, see, e.g., [Glu97, LS05, LS01, LOST10].

5.4 Open Questions

In this subsection we raise the question of the possible upper-bounds on $\Theta_\pi(\gamma)$ in terms of both π and γ. One may ask about the sharpest possible bound. Our main result was a bound of the form (Theorem 4.18)

$$D(\gamma)^A|\Theta_\pi(\gamma)| \leq C\deg(\pi)^\kappa, \tag{14}$$

where C is independent of $\gamma \in G_{0+} \cap G_{\mathrm{reg}}$ and $\pi \in \mathrm{Irr}^{\mathrm{Yu}}(G)$. Slightly more generally we fix a bounded subset $\mathcal{B} \subset G$ and assume in the following that $\gamma \in \mathcal{B} \cap G_{\mathrm{reg}}$.

The most optimistic bound would be that

$$D(\gamma)^{\frac{1}{2}}|\Theta_\pi(\gamma)| \overset{?}{\leq} C, \tag{15}$$

where $\gamma \in \mathcal{B}$ and C depends only on $\mathcal{B}$. In the appendix we shall verify that the estimate (15) is valid for the group $G = \mathrm{SL}_2(k)$. However the analogue of this bound already doesn't hold in higher rank when varying the residue characteristic and π is a Steinberg representation, as we explain in [KST]. It would be interesting to investigate all the counterexamples to (15) in general and find in which cases it holds.

There is a wide range of possibilities between (14) and (15). The exact asymptotic of $|\Theta_\pi(\gamma)|$ lies somewhere in between, and based on Harish-Chandra regularity theorem and our work in [KST] it seems plausible that the exact bound should be $A = \frac{1}{2}$ and κ slightly below 1 depending on G.

Appendix: The Sally–Shalika Character Formula

We study the Sally–Shalika formula [SS68] for characters of admissible representations of $G = \mathrm{SL}_2(k)$, where k is a p-adic field. Our goal is to establish a bound of the form (15) for the character $\Theta_\pi(\gamma)$ that is completely uniform in π. The explicit calculation of character values is crucial for this. It would be interesting to investigate where such a result can hold in general.

Let $Z = \{\pm 1\}$ be the center of G. We follow mostly the notation and convention from [ADSS11]. We assume throughout that $p \geq 2e + 3$ where e is the absolute ramification degree of k. We let $\theta \in \{\varepsilon, \varepsilon\varpi, \varpi\}$, with ϖ is a uniformizer of $\mathcal{O}_k$ and ε is any fixed element of $\mathcal{O}_k^\times \backslash (\mathcal{O}_k^\times)^2$. Let $k_\theta := k(\sqrt{\theta})$ and $k_\theta^1 \subset k_\theta^\times$ be the subgroup of elements of norm one. We extend the valuation from $k^\times$ to $k_\theta^\times$. Note that there is a unique non-trivial quadratic character $\varphi_\varepsilon : k_\varepsilon^1 \to \{\pm 1\}$.

Inside G we let $T^\theta \simeq k_\theta^1$ be the associated maximal elliptic tori given by the matrices $\left(\begin{smallmatrix} a & b \\ b\theta & a \end{smallmatrix}\right)$ where $a + b\theta \in k_\theta^1$. As θ ranges in $\{\varepsilon, \varepsilon\varpi, \varpi\}$ this describes the stable conjugacy classes of elliptic tori (abstractly k_θ is the splitting field of T^θ). There is a finer classification of G-conjugacy classes: there are two unramified conjugacy classes of unramified elliptic tori, denoted $T^\varepsilon = T^{\varepsilon,1}$ and $T^{\varepsilon,\varpi}$ while for ramified elliptic tori, the answer depends on whether -1 is a square in the residue field. If -1 is not a square, then besides $T^\varpi = T^{\varpi,1}$ and $T^{\varepsilon\varpi} = T^{\varepsilon\varpi,1}$ there are two additional G-conjugacy classes denoted $T^{\varpi,\varepsilon}$ and $T^{\varepsilon\varpi,\varepsilon}$.

The torus filtration is as described in [ADSS11, §3.2], namely an element $1 + x \in k_\theta^1 \simeq T^\theta$ with $v(x) > 0$ has depth equal to $v(x)$. In particular we have $D(\gamma) = q^{-2d_+(\gamma)}$ for all regular semisimple $\gamma \in G$ where $d_+(\gamma) := \max_{z\in Z} d(z\gamma)$ is the maximal depth.

Every supercuspidal representation of G is of the form $\pi = \pi^\pm(T, \varphi)$ where $(T, \varphi, \pm)$ is a supercuspidal parameter. Here T is an elliptic tori up to G-conjugation and φ is a quasi-character of T. The depth of π is equal to the depth of φ which is the smallest $r \geq 0$ such that φ is trivial on T_{r^+}.

Let dg be the Haar measure on $G/Z(G)$ is as in [ADSS11, §6], thus $\mathrm{vol}(\mathrm{SL}(2, \mathcal{O}_k)) = \frac{q^2-1}{q^{\frac{1}{2}}}$. The formal degree is by construction $\deg(\pi) = \frac{\dim(\rho)}{\mathrm{vol}(J)}$. By a theorem of Harish-Chandra $\deg(\pi)$ is proportional to the constant term $c_0(\pi)$ in the expansion of Θ_π near the identity. Here we find $\deg(\pi) = c \cdot c_0(\pi)$ where $c := -\frac{2q^{\frac{1}{2}}}{q+1}$.

The Sally–Shalika formula is an exact formula for the character $\Theta_\pi(\gamma)$ for any regular noncentral semisimple element $\gamma \in G$. Here we shall give a direct consequence tailored to our purpose of studying of the asymptotic behavior of characters.

Proposition A.1. *If $\pi = \pi(T^\varepsilon, \varphi)$ has depth r, then the following holds:*

$$D(\gamma)^{\frac{1}{2}} |\Theta_\pi(\gamma)| = \begin{cases} \left|\varphi(\gamma) + \varphi(\gamma^{-1})\right|, & \gamma \in T^\varepsilon \backslash Z T^\varepsilon_{r^+} \\ 1 \pm \deg(\pi) D(\gamma)^{\frac{1}{2}}, & \gamma \in T^{\varepsilon,\eta}_{r^+},\ \eta \in \{1, \varpi\} \\ 1 - \deg(\pi) D(\gamma)^{\frac{1}{2}}, & \gamma \in A_{r^+} \\ \deg(\pi) D(\gamma)^{\frac{1}{2}}, & \textit{o/w if } \gamma \in G_{r^+}. \end{cases}$$

The character vanishes in the other cases, namely if $\gamma \notin G_{r^+} \cup T^\varepsilon$. The formal degree is $\deg(\pi) = q^r$.

Proof. This is [ADSS11, §14]. Note that in their notation the quasi-character φ is denoted ψ there; the additive character Ω_k is denoted Λ there.

Since the Gauss sum $H(\Lambda', k_\varepsilon)$ is unramified, we have that it is equal to $(-1)^{r+1}$ according to [ADSS11, Lemma 4.2].

The assertion on the formal degree is [ADSS11, Remark 10.16] since $c_0(\pi) = -q^r$. □

Proposition A.2. *If $\pi = \pi(\varpi, \varphi)$ has depth r, then the following holds:*

$$D(\gamma)^{\frac{1}{2}}|\Theta_\pi(\gamma)| \leq \begin{cases} 2, & \gamma \in T^\theta \backslash ZT_r^\theta \\ 1.5, & \gamma \in T_r^{\varpi,\eta} \backslash T_{r+}^{\varpi,\eta},\ \eta \in \{1, \varpi\} \\ 1, & \gamma \in T_r^{\varepsilon\varpi,\eta} \backslash T_{r+}^{\varepsilon\varpi,\eta},\ \eta \in \{1, \varpi\} \\ 1 + \deg(\pi) D(\gamma)^{\frac{1}{2}}, & \gamma \in T_{r+}^{\varpi,\eta} \cup A_{r+} \\ \deg(\pi) D(\gamma)^{\frac{1}{2}}, & \text{o/w if } \gamma \in G_{r+}. \end{cases}$$

The character vanishes in the other cases, namely if $\gamma \notin G_{r+} \cup T^\theta$. The formal degree is $\deg(\pi) = \frac{1}{2}(q+1)q^{r-\frac{1}{2}}$.

Proof. Again this is [ADSS11, §14] where it is shown that $c_0(\pi) = -\frac{1}{2}(q+1)q^{r-\frac{1}{2}}$.

The ramified Gauss sum $H(\Lambda', k_\varpi)$ is a fourth root of unity according to [ADSS11, Lemma 4.2]. In the second case we have the inequality $\leq 1 + |A|$ where the exponential sum is

$$A := \frac{1}{2\sqrt{q}} \sum_{\substack{x \in (k_\varpi^1)_{r:r+} \\ x \neq \gamma^{\pm 1}}} \operatorname{sgn}_\varpi(\operatorname{tr}(\gamma - x))\varphi(x).$$

Here $k_\theta^1 \subset k_\theta^\times$ is the subgroup of elements of norm 1, and $(k_\varpi^1)_{r:r+}$ denotes [ADSS11, §5.1] the quotient group $(k_\varpi^1)_r/(k_\varpi^1)_{r+}$. This is an additive group that can be described by writing explicitly $x = 1 + \alpha^{2r}X$ where $X \in \mathcal{O}/(\varpi)$. We have $\operatorname{tr}(x) = 2 + \operatorname{tr}(\alpha^{2r})X$, and similarly we shall write $\gamma = 1 + \alpha^{2r}Y$.

Since $\operatorname{sgn}_\varpi$ is the quadratic character attached to ϖ, we are left with $\chi(X - Y)$ where χ is the Legendre symbol on $\mathcal{O}/(\varpi)$. The character φ has conductor r, thus $X \mapsto \varphi(1 + \alpha^{2r}X)$ is a non-trivial additive character. Finally the exponential sum A is a unit times a Gauss sum, thus $|A| = \frac{1}{2}$.

In the third case the character is equal to an exponential sum which can be handled similarly. □

We finally consider the remaining four "exceptional" supercuspidal representations. They all have depth zero.

Proposition A.3. *Suppose that π is an exceptional supercuspidal representation induced from T^ε. Then the following holds:*

$$2D(\gamma)^{\frac{1}{2}}|\Theta_\pi(\gamma)| \leq 1 + D(\gamma)^{\frac{1}{2}}, \quad \gamma \in T^\varepsilon \backslash ZT_{0+}^\varepsilon \cup A_{0+} \cup T_{0+},$$

where T is any of the elliptic tori, and the character vanishes otherwise. The formal degree is $\deg(\pi) = \frac{1}{2}$. *If π is induced from $T^{\varepsilon,\varpi}$, the same formula holds with T^{ε} replaced by $T^{\varepsilon,\varpi}$.*

Remark A.4. The behavior $D(\gamma) \to \infty$ is qualitatively different than for the other "ordinary" supercuspidals.

Proof. This follows from [ADSS11, §9, §15]. The passage from T^{ε} to $T^{\varepsilon,\varpi}$ is explained in [ADSS11, Remark 9.8]. □

Corollary A.5. *For all supercuspidal representations π of* $\mathrm{SL}(2,k)$ *and all regular semisimple γ, the following holds:*

$$D(\gamma)^{\frac{1}{2}}|\Theta_{\pi}(\gamma)| \leq 2 + D(\gamma)^{\frac{1}{2}}.$$

Proof. This follows by combining the Propositions A.1, A.2 and A.3. In the last three cases of Proposition A.1 we need to observe that $\gamma \in G_{r+}$ which is equivalent to $d(\gamma) > r$. This implies $d_{+}(\gamma) > r$ and thus $D(\gamma) < q^{-2r}$. Therefore $\deg(\pi)D(\gamma)^{\frac{1}{2}} < 1$.

Similarly in the last two cases of Proposition A.2 we have that $\gamma \in G_{r+}$ and in view of the normalization of the valuation this implies that $D(\gamma) < q^{-2r-1}$. Therefore $\deg(\pi)D(\gamma)^{\frac{1}{2}} \leq \frac{1}{2}$ which concludes the claim. □

Acknowledgements We would like to thank Julius Witte for his comments and also for pointing out a mistake in Lemma 3.5 in an earlier version. We would like to thank Anne-Marie Aubert and the referee for their helpful comments.

Ju-Lee Kim was partially supported by NSF grants, Sug Woo Shin was partially supported by NSF grant DMS-1162250/1449558 and a Sloan Fellowship. Nicolas Templier was partially supported by NSF grant DMS-1200684/1454893.

References

[Adl98] J.D. Adler, Refined anisotropic K-types and supercuspidal representations. Pac. J. Math. **185**(1), 1–32 (1998)

[AS09] J.D. Adler, L. Spice, Supercuspidal characters of reductive p-adic groups. Am. J. Math. **131**(4), 1137–1210 (2009)

[AS08] J.D. Adler, L. Spice, Good product expansions for tame elements of p-adic groups. Int. Math. Res. Pap. (1), 95 (2008). Article ID rpn 003

[ADSS11] J.D. Adler, S. DeBacker, P.J. Sally Jr., L. Spice, Supercuspidal characters of SL_2 over a p-adic field, in *Harmonic Analysis on Reductive, p-Adic Groups*. Contemporary Mathematics, vol. 543 (American Mathematical Society, Providence, RI, 2011), pp. 19–69

[AK00] J.D. Adler, A. Roche, An intertwining result for p-adic groups. Can. J. Math. **52**(3), 449–467 (2000)

[AK07] J.D. Adler, J. Korman, The local character expansion near a tame, semisimple element. Am. J. Math. **129**(2), 381–403 (2007)

[Art89] J. Arthur, The L^2-Lefschetz numbers of Hecke operators. Invent. Math. **97**(2), 257–290 (1989)

[KM03] J.-L. Kim, F. Murnaghan, Character expansions and unrefined minimal K-types. Am. J. Math. **125**(6), 1199–1234 (2003)

[BH96] C.J. Bushnell, G. Henniart, Local tame lifting for GL(N). I. Simple characters. Inst. Hautes Études Sci. Publ. Math. **83**, 105–233 (1996)

[BK93] C.J. Bushnell, P.C. Kutzko, *The Admissible Dual of* GL(N) *Via Compact Open Subgroups*. Annals of Mathematics Studies, vol. 129 (Princeton University Press, Princeton, NJ, 1993)

[BK94] C.J. Bushnell, P.C. Kutzko, The admissible dual of SL(N). II. Proc. Lond. Math. Soc. (3) **68**(2), 317–379 (1994)

[Bro98] P. Broussous, Extension du formalisme de Bushnell et Kutzko au cas d'une algèbre à division. Proc. Lond. Math. Soc. (3) **77**(2), 292–326 (1998)

[Del76] P. Deligne, Le support du caractère d'une représentation supercuspidale. C. R. Acad. Sci. Paris Sér. A-B **283**(4), Aii, A155–A157 (1976)

[Glu95] D. Gluck, Sharper character value estimates for groups of Lie type. J. Algebra **174**(1), 229–266 (1995)

[Glu97] D. Gluck, Characters and random walks on finite classical groups. Adv. Math. **129**(1), 46–72 (1997)

[HC70] Harish-Chandra, *Harmonic Analysis on Reductive p-Adic Groups*. Lecture Notes in Mathematics, vol. 162 (Springer, Berlin, 1970). Notes by G. van Dijk

[HC99] Harish-Chandra, *Admissible Invariant Distributions on Reductive p-Adic Groups*. University Lecture Series, vol. 16 (American Mathematical Society, Providence, RI, 1999). Preface and notes by Stephen DeBacker and Paul J. Sally Jr

[Her83] R.A. Herb, Discrete series characters and Fourier inversion on semisimple real Lie groups. Trans. Am. Math. Soc. **277**(1), 241–262 (1983)

[Her98] R.A. Herb, Discrete series characters and two-structures. Trans. Am. Math. Soc. **350**(8), 3341–3369 (1998)

[HII08] K. Hiraga, A. Ichino, T. Ikeda, Formal degrees and adjoint γ-factors. J. Am. Math. Soc. **21**(1), 283–304 (2008)

[Jac71] H. Jacquet, Représentations des groupes linéaires p-adiques, in *Theory of Group Representations and Fourier Analysis (Centro Internaz. Mat. Estivo (C.I.M.E.), II Ciclo, Montecatini Terme, 1970)* (Edizioni Cremonese, Rome, 1971), pp. 119–220

[Kim07] J.-L. Kim, Supercuspidal representations: an exhaustion theorem. J. Am. Math. Soc. **20**(2), 273–320 (electronic) (2007)

[KL13a] J.-L. Kim, G. Lusztig, On the characters of unipotent representations of a semisimple p-adic group. Representation Theory **17**, 426–441 (2013)

[KL13b] J.-L. Kim, G. Lusztig, On the Steinberg character of a semisimple p-adic group. Pac. J. Math. **265**(2), 499–509 (2013)

[KST] J.-L. Kim, S.W. Shin, N. Templier, Asymptotic behavior of supercuspidal representations and Sato-Tate equidistribution for families. Preprint

[KY11] J.-L. Kim, J.-K. Yu, Twisted Levi sequences and explicit types on Sp_4, in *Harmonic Analysis on Reductive, p-Adic Groups*. Contemporary Mathematics, vol. 543 (American Mathematical Society, Providence, RI, 2011), pp. 135–154

[Kut77] P.C. Kutzko, Mackey's theorem for nonunitary representations. Proc. Am. Math. Soc. **64**(1), 173–175 (1977)

[Lan00] E. Landvogt, Some functorial properties of the Bruhat-Tits building. J. Reine Angew. Math. **518**, 213–241 (2000)

[LS01] M.W. Liebeck A. Shalev, Diameters of finite simple groups: sharp bounds and applications. Ann. Math. (2) **154**(2), 383–406 (2001)

[LS05] M.W. Liebeck, A. Shalev, Fuchsian groups, finite simple groups and representation varieties. Invent. Math. **159**(2), 317–367 (2005)

[Mar75] S. Martens, The characters of the holomorphic discrete series. Proc. Natl. Acad. Sci. U.S.A. **72**(9), 3275–3276 (1975)

[MP94] A. Moy, G. Prasad, Unrefined minimal K-types for p-adic groups. Invent. Math. **116**(1–3), 393–408 (1994)

[MP96] A. Moy, G. Prasad, Jacquet functors and unrefined minimal K-types. Comment. Math. Helv. **71**(1), 98–121 (1996)

[MS12] R. Meyer, M. Solleveld, Characters and growth of admissible representations of reductive p-adic groups. J. Inst. Math. Jussieu **11**(2), 289–331 (2012)

[Nev13] M. Nevins, Branching rules for supercuspidal representations of $\mathrm{SL}_2(k)$, for k a p-adic field. J. Algebra **377**, 204–231 (2013)

[LOST10] M.W. Liebeck, E.A. O'Brien, A. Shalev, P.H. Tiep, The Ore conjecture. J. Eur. Math. Soc. **12**(4), 939–1008 (2010)

[Pra01] G. Prasad, Galois-fixed points in the Bruhat-Tits building of a reductive group. Bull. Soc. Math. Fr. **129**(2), 169–174 (2001)

[Rou77] G. Rousseau, Immeubles des groupes réducitifs sur les corps locaux. Thèse de doctorat, Publications Mathématiques d'Orsay, No. 221-77.68. U.E.R. Mathématique, Université Paris XI, Orsay (1977)

[SS08] V. Sécherre, S. Stevens, Représentations lisses de $\mathrm{GL}_m(D)$. IV. Représentations supercuspidales. J. Inst. Math. Jussieu **7**(3), 527–574 (2008)

[SS97] P. Schneider, U. Stuhler, Representation theory and sheaves on the Bruhat-Tits building. Inst. Hautes Études Sci. Publ. Math. **85**, 97–191 (1997)

[SS68] P.J. Sally Jr., J.A. Shalika, Characters of the discrete series of representations of SL(2) over a local field. Proc. Natl. Acad. Sci. U.S.A. **61**, 1231–1237 (1968)

[ST11] S.W. Shin, N. Templier, Sato-Tate theorem for families and low-lying zeros of automorphic L-functions. Invent. Math. **203**(1), 1–177 (2016). Appendix A by Robert Kottwitz, and Appendix B by Raf Cluckers, Julia Gordon and Immanuel Halupczok

[Ste08] S. Stevens, The supercuspidal representations of p-adic classical groups. Invent. Math. **172**(2), 289–352 (2008)

[Tit79] J. Tits, Reductive groups over local fields, in *Automorphic Forms, Representations and L-Functions*. Proceedings of Symposia in Pure Mathematics, Oregon State University, Corvallis, OR, 1977, Part 1. Proceedings of Symposia in Pure Mathematics, vol. XXXIII (American Mathematical Society, Providence, RI, 1979), pp. 29–69

[Wal08] J.-L. Waldspurger, L'endoscopie tordue n'est pas si tordue. Mem. Am. Math. Soc. **194**(908), x+261 (2008)

[Yu01] J.-K. Yu, Construction of tame supercuspidal representations. J. Am. Math. Soc. **14**(3), 579–622 (electronic) (2001)

Endoscopy and Cohomology of a Quasi-Split $U(4)$

Simon Marshall

Abstract We prove asymptotic upper bounds for the L^2 Betti numbers of the locally symmetric spaces associated with a quasi-split $U(4)$. These manifolds are 8-dimensional, and we prove bounds in degrees 2 and 3, with the behavior in the other degrees being well understood. In degree 3, we conjecture that these bounds are sharp. Our main tool is the endoscopic classification of automorphic representations of $U(N)$ by Mok.

1 Introduction

Let E be an imaginary quadratic field. Let $N \geq 1$, let $U(N)$ be the quasi-split unitary group of degree N with respect to $E/\mathbb{Q}$, and let G be an inner form of $U(N)$. Let $\Gamma \subset G(\mathbb{Q})$ be an arithmetic congruence lattice, and for $n \geq 1$ let $\Gamma(n)$ be the corresponding principal congruence subgroup of Γ. Let K_∞ be a maximal compact subgroup of $G(\mathbb{R})$. Let $Y(n) = \Gamma(n)\backslash G(\mathbb{R})/K_\infty$, which is a complex orbifold (or manifold if n is large enough). We let $H^i_{(2)}(Y(n))$ be the L^2 cohomology groups of $Y(n)$. By Borel and Casselman [BC], $H^i_{(2)}(Y(n))$ is equal to the space of square-integrable harmonic i-forms on $Y(n)$, and we shall identify it with this space from now on. Note that $H^i_{(2)}(Y(n)) = H^i(Y(n))$ when $Y(n)$ is compact. We set $h^i_{(2)}(Y(n)) = \dim H^i_{(2)}(Y(n))$. This article is interested in how $h^i_{(2)}(Y(n))$ grow with n, specifically in the case when $G = U(4)$.

We let $V(n) = |\Gamma : \Gamma(n)|$, which is asymptotically equal to the volume of $Y(n)$. The standard bound that we wish to improve over is $h^i_{(2)}(Y(n)) \ll V(n)$. This follows from the equality of $h^i_{(2)}$ with an ordinary Betti number if Γ is cocompact, and otherwise from the noncompact version of Matsushima's formula in [BG, Proposition 5.6] which expresses $h^i_{(2)}(Y(n))$ in terms of automorphic representations, together with Savin's bound [Sa] for the multiplicity of a representation in the cuspidal spectrum and Langlands' theory of Eisenstein series.

S. Marshall (✉)
Department of Mathematics, University of Wisconsin–Madison, 480 Lincoln Drive, Madison, WI 53706, USA
e-mail: marshall@math.wisc.edu

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_8

The basic principle that we shall use to bound $h^i_{(2)}(Y(n))$ is the fact that, if i is not half the dimension of $Y(n)$, the archimedean automorphic forms that contribute to $h^i_{(2)}(Y(n))$ must be nontempered. In the case where Γ is cocompact, one may combine this principle with the trace formula and asymptotics of matrix coefficients to prove a bound of the form $h^i_{(2)}(Y(n)) \ll V(n)^{1-\delta}$ for some $\delta > 0$. In [SX], Sarnak and Xue suggest the optimal bound that one should be able to prove in this way using only the archimedean trace formula. In the case when $N = 3$ and Γ is cocompact (which implies that $Y(n)$ have real dimension 4), they predict that $h^1_{(2)}(Y(n)) \ll_\epsilon V(n)^{1/2+\epsilon}$, while they prove that $h^1_{(2)}(Y(n)) \ll_\epsilon V(n)^{7/12+\epsilon}$.

There is a deeper way in which one may exploit nontemperedness to prove bounds for cohomology. In [Mo] Mok, following Arthur [Art], classifies the automorphic spectrum of $U(N)$ in terms of conjugate self-dual cusp forms on GL_M/E for $M \leq N$. One of the implicit features of this classification is that if a representation π on $U(N)$ is sufficiently nontempered at one place, then it must be built up from cusp forms on groups GL_M/E with M strictly less than N—in other words, π comes from a smaller group. We have been interested in deriving quantitative results from this qualitative feature of the classification. In [Ma], we used this (more precisely, the complete solution of endoscopy for $U(3)$ by Rogawski in [Ro]) to prove that $h^1_{(2)}(Y(n)) \ll_\epsilon V(n)^{3/8+\epsilon}$ when $N = 3$ and G is arbitrary, strengthening the bound of Sarnak and Xue. Moreover, we proved that this bound is sharp. In this article, we partially extend this result to the case $G = U(4)$. Note that in this case, the real dimension of $Y(n)$ is 8.

Theorem 1.1. *If $G = U(4)$ and $i = 2$ or 3, and n is only divisible by primes that split in E, we have $h^i_{(2)}(Y(n)) \ll_\epsilon V(n)^{8/15+\epsilon}$.*

See Theorem 3.1 for a precise statement. We expect Theorem 1.1 to be sharp in the case $i = 3$, but when $i = 2$ we expect the true order of growth to be $V(n)^{2/5+\epsilon}$ for reasons discussed below. Note that we have $h^1_{(2)}(Y(n)) = 0$ for all n, by combining the noncompact Matsushima formula of Borel and Garland [BG, Proposition 5.6] with the vanishing theorems of, e.g., §10.1 of Borel and Wallach [BW]. The results of Savin [Sa] also imply that $h^4_{(2)}(Y(n)) \gg V(n)$.

1.1 Outline of Proof

To describe the method of proof of Theorem 1.1 in more detail, we begin by outlining the classification of Arthur and Mok. We define an Arthur parameter for $U(N)$ to be a formal linear combination $\psi = \nu(n_1) \boxtimes \mu_1 \boxplus \ldots \boxplus \nu(n_l) \boxtimes \mu_l$, where $\nu(k)$ denotes the unique irreducible (complex-algebraic) representation of $SL(2, \mathbb{C})$ of dimension k, and μ_i is a conjugate self-dual cusp form on GL_{m_i}/E, subject to certain conditions including that $N = \sum n_i m_i$. To each ψ, there is associated a packet Π_ψ of representations of $U(N)(\mathbb{A})$, certain of which occur in the automorphic spectrum. Moreover, the entire automorphic spectrum is obtained in this way. If we combine this classification with the noncompact case of Matsushima's formula, we have

$$h^i_{(2)}(Y(n)) \leq \sum_{\psi} \sum_{\pi \in \Pi_\psi} h^i(\mathfrak{g}, K; \pi_\infty) \dim \pi_f^{K(n)}. \tag{1}$$

Here, $\mathfrak{g}$ is the Lie algebra of $U(N)(\mathbb{R})$, K is a maximal compact subgroup of $U(N)(\mathbb{R})$, we let $H^i(\mathfrak{g}, K; \pi_\infty)$ denote $(\mathfrak{g}, K)$ cohomology, and $h^i(\mathfrak{g}, K; \pi_\infty) = \dim H^i(\mathfrak{g}, K; \pi_\infty)$. As mentioned above, if i is not the middle degree, then those ψ contributing to the sum must be *non-generic*, i.e. one of the representations of $SL(2, \mathbb{C})$ must be nontrivial.

We deduce Theorem 1.1 from (1) in two steps.

Step 1: Bound $\sum_{\pi_f \in \Pi_{\psi,f}} \dim \pi_f^{K(n)}$ for each ψ, where $\Pi_{\psi,f}$ denotes the finite part of the packet Π_ψ.

Step 2: Sum the resulting bounds over those ψ that contribute to cohomology in the required degree.

We begin step 1 by writing $\Pi_{\psi,f} = \otimes_p \Pi_{\psi,p}$, so that we must bound $\sum_{\pi_p \in \Pi_{\psi,p}} \dim \pi_p^{K(n)}$ for each p. When p is split in E, $\Pi_{\psi,p}$ is an explicitly described singleton, and it is easy to do this directly. When p is nonsplit, we use the trace identities appearing in the definition of $\Pi_{\psi,p}$ [Mo, Theorem 3.2.1]. By writing $\dim \pi_p^{K(n)}$ as a trace, these allow us to relate $\sum_{\pi_p \in \Pi_{\psi,p}} \dim \pi_p^{K(n)}$ to objects like $\dim \mu_i^{K'(n)}$, where μ_i is one of the cusp forms appearing in ψ and $K'(n)$ is a suitable congruence subgroup of $GL(m_i)$.

As an example, one type of packet that contributes to (1) when $N = 4$ is those of the form $\psi = \nu(2) \boxtimes \mu$, where μ is a cusp form on GL_2/E. After carrying out step 1 in this case, we obtain

$$\sum_{\pi_f \in \Pi_{\psi,f}} \dim \pi_f^{K(n)} \ll n^{5+\epsilon} \sum_{\pi'_f \in \Pi(\mu)_f} \dim \pi_f'^{K'(n)} \tag{2}$$

where $\Pi(\mu)$ is the packet on $U(2)$ corresponding to μ, and $K'(n)$ is the standard principal congruence subgroup of level n on $U(2)$. Step 2 is bounding the right-hand side of (2). We do this by observing that if Π_ψ contains a cohomological representation, and $\pi' \in \Pi(\mu)$ as in (2), then there are only finitely many possibilities for the infinitesimal character of π'_∞, and hence of π'_∞ itself. We may therefore bound the right-hand side of (2) in terms of the multiplicities of archimedean representations on $U(2)$, and these may be bounded by the results of Savin.

The reason we do not expect Theorem 1.1 to be sharp when $i = 2$ is that the main contribution to $h^2_{(2)}$ comes from parameters of the form $\nu(2) \boxtimes \mu$ with μ on GL_2. (Note that this relies on the Adams–Johnson conjectures on the structure of cohomological Arthur packets, which have now been proved by Arancibia, Moeglin, and Renard [AMR].) We do not have sharp bounds for the contribution from these parameters, because we do not have sharp bounds for the dimensions of spaces of K-fixed vectors in Speh representations induced from $GL_2 \times GL_2$ on GL_4. To be more precise, if π is such a Speh representation of $GL(4, \mathbb{Q}_p)$, we require a bound for $\dim \pi^{K(p^k)}$, where $K(p^k)$ is the usual principal congruence subgroup, that is uniform in both k and π. In particular, this is more difficult than knowing the Kirillov dimension of these representations.

We have restricted to levels that are split in E because of an issue with the twisted fundamental lemma, which is used in step 1 in the case of inert primes. Allowing level in this argument would require an extension of the twisted FL, which states that the twisted transfer takes the characteristic functions of principal congruence subgroups to functions of the same type. This would follow from the twisted FL for Lie algebras, which is not known at this time. However, it should be possible to prove it by following Waldspurger's proof for groups in [Wa].

The tools used in the proof should extend to a general $U(N)$ with a little extra work. However, because the recipe for the degrees of cohomology on $U(N)$ to which an Arthur parameter can contribute is complicated, the result this would give for cohomology growth would not be as strong.

2 The Endoscopic Classification for $U(N)$

In this section we describe the endoscopic classification for the quasi-split group $U(N)$ by Mok. Because of the large amount of notation that must be introduced to do this in full, we shall often omit details that are not directly relevant to the proof of Theorem 1.1.

2.1 Number Fields

Throughout this section, F will denote a local or global field of characteristic 0, and E will denote a quadratic étale F-algebra. We will assume that E is a quadratic extension of F unless specified otherwise. The conjugation of E over F will be denoted by c. We set $\Gamma_F = \mathrm{Gal}(\overline{F}/F)$. The Weil groups of F and E will be denoted by W_F and W_E, respectively. If F is local, we let L_F denote its local Langlands group, which is given by W_F if F is archimedean and $W_F \times SU(2)$ otherwise. If F is global, the adeles of F and E will be denoted by $\mathbb{A}$ and $\mathbb{A}_E$. If F is local (resp. global), χ will denote a character of $E^\times$ (resp. $\mathbb{A}_E^\times/E^\times$) whose restriction to $F^\times$ (resp. $\mathbb{A}^\times$) is the quadratic character associated with E/F by class field theory. We will often think of χ as a character of W_E.

2.2 Algebraic Groups

For any $N \geq 1$, we let $U(N)$ denote the quasi-split unitary group over F with respect to E/F, whose group of F-points is

$$U(N)(F) = \{g \in GL(N,E) \mid {}^t c(g) J g = J\}$$

where

$$J = \begin{pmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{pmatrix}.$$

In the case when $E = F \times F$, we have

$$U(N)(F) = \{(g_1, g_2) \in GL(N,F) \times GL(N,F) \mid g_2 = J {}^t g_1^{-1} J^{-1}\}.$$

Projection onto the first and second factors defines isomorphisms $\iota_1, \iota_2 : U(N)(F) \simeq GL(N,F)$, and we have $\iota_2 \circ \iota_1^{-1} : g \mapsto J {}^t g^{-1} J^{-1}$.

We define $G(N) = \mathrm{Res}_{E/F} GL(N)$. We let θ denote the automorphism of $G(N)$ whose action on F-points is given by

$$\theta(g) = \Phi_N {}^t c(g)^{-1} \Phi_N^{-1} \quad \text{for} \quad g \in G(N)(F) \simeq GL(N,E),$$

where

$$\Phi_N = \begin{pmatrix} & & & 1 \\ & & -1 & \\ & \cdot^{\cdot^{\cdot}} & & \\ (-1)^{N-1} & & & \end{pmatrix}.$$

We define $\widetilde{G}^+(N) = G(N) \rtimes \langle\theta\rangle$, and let $\widetilde{G}(N)$ denote the $G(N)$-bitorsor $G(N) \rtimes \theta$. We will denote these groups by $U_{E/F}(N)$, $G_{E/F}(N)$, etc. when we want to explicate the dependence on the extension E/F.

Our discussion in this section will implicitly require choosing Haar measures on the F-points of these groups when F is local, in particular when discussing transfers of functions and character relations. We may do this in an arbitrary way, subject only to the condition that the Haar measures assign mass 1 to a hyperspecial maximal compact subgroup when one exists. This condition allows us to state the fundamental lemma without the introduction of any constant factors.

2.3 L-groups and Embeddings

If G is a connected reductive algebraic group over F, the L-group ${}^LG$ is an extension $\widehat{G} \rtimes W_F$, where $\widehat{G}$ is the complex dual group of G. If G_1 and G_2 are two such groups, an L-morphism ${}^LG_1 \to {}^LG_2$ is a map that reduces to the identity map on W_F. An L-embedding is an injective L-morphism. In this paper we shall only need to consider ${}^LG$ when G is a product of the groups $U(N)$, $GL(N)$, and $G(N)$. Because the L-group of $G_1 \times G_2$ is the fiber product of ${}^LG_1$ and ${}^LG_2$ over W_F, it suffices to specify ${}^LG$ when G is one of these groups. We have ${}^LGL(N) = GL(N,\mathbb{C}) \times W_F$. We have ${}^LU(N) = GL(N,\mathbb{C}) \rtimes W_F$, where W_F acts through its quotient $\mathrm{Gal}(E/F)$ via the automorphism

$$g \mapsto \Phi_N {}^tg^{-1} \Phi_N^{-1}.$$

We have ${}^LG(N) = (GL(N,\mathbb{C}) \times GL(N,\mathbb{C})) \rtimes W_F$, where W_F acts through $\mathrm{Gal}(E/F)$ by switching the two factors. We let $\hat{\theta}$ denote the automorphism of $\widehat{G(N)}$ given by $\hat{\theta}(x,y) = (\Phi_N {}^ty^{-1}\Phi_N^{-1}, \Phi_N {}^tx^{-1}\Phi_N^{-1})$.

We define the L-embedding $\xi_\kappa : {}^LU(N) \to {}^LG(N)$ for $\kappa = \pm 1$ as follows. (Note we will often abbreviate ± 1 to simply $\pm$.) We define $\chi_+ = 1$ and $\chi_- = \chi$, and we choose $w_c \in W_F \setminus W_E$. We define ξ_κ by the following formulae.

$$\begin{aligned} g \rtimes 1 &\mapsto (g, {}^tg^{-1}) \rtimes 1 \text{ for } g \in GL(N,\mathbb{C}) \\ I \rtimes \sigma &\mapsto (\chi_\kappa(\sigma)I, \chi_\kappa^{-1}(\sigma)I) \rtimes \sigma \text{ for } \sigma \in W_E \\ I \rtimes w_c &\mapsto (\kappa\Phi_N, \Phi_N^{-1}) \rtimes w_c. \end{aligned}$$

Note that the conjugacy class of $\xi_\pm$ is independent of the choice of w_c.

2.4 Endoscopic Data

In the cases we consider in this paper, it suffices to work with a simplified notion of endoscopic datum that we now describe. See [KS] for the general definition. Let G^0 be a connected reductive group over F, and let θ be a semisimple automorphism of G^0. Let G be the G^0-bitorsor $G^0 \rtimes \theta$. We let $\hat{\theta}$ be the automorphism of $\widehat{G}^0$ that is dual to θ and preserves a fixed Γ_F-splitting of $\widehat{G}_0$. We shall only need to consider the cases where θ is trivial or G is the torsor $\widetilde{G}(N)$ defined in Sect. 2.2, in which cases the dual automorphism $\hat{\theta}$ is the one given in Sect. 2.3.

We let $\widehat{G} = \widehat{G}^0 \rtimes \hat{\theta}$. An endoscopic datum for G is a triple (G', s, ξ') satisfying the following conditions.

- $s \in \widehat{G}$ is semi-simple.
- G' is a quasi-split connected reductive group over F.

- $\xi' : {}^L G' \to {}^L G^0$ is an L-embedding.
- The restriction of ξ' to $\widehat{G}'$ is an isomorphism $\widehat{G}' \simeq \mathrm{Cent}(s, \widehat{G}^0)^0$.
- We have $\mathrm{Ad}(s) \circ \xi' = a \cdot \xi'$, where $a : W_F \to Z(\widehat{G}^0)$ is a 1-cocycle that is cohomologically trivial if F is local, and is everywhere locally trivial if F is global.

We refer to [KS, Sect. 2.1] for the definition of equivalence of endoscopic data. We will often omit the data s and ξ' if they are not immediately relevant. We say that an endoscopic datum is elliptic if we have

$$(Z(\widehat{G}')^{\Gamma_F})^0 \subset Z(\widehat{G}^0)^{\hat{\theta},\Gamma_F}.$$

We denote the set of equivalence classes of endoscopic data for G by $\mathcal{E}(G)$, and the subset of elliptic data by $\mathcal{E}_{\mathrm{ell}}(G)$. We set $\mathcal{E}(\widetilde{G}(N)) = \widetilde{\mathcal{E}}(N)$. From now on, we shall only use the notation $\mathcal{E}(G)$ when G is a group, i.e. when θ is trivial.

There is a subset $\widetilde{\mathcal{E}}_{\mathrm{sim}}(N) \subset \widetilde{\mathcal{E}}_{\mathrm{ell}}(N)$, called the set of simple endoscopic data, that consists of the elements $(U(N), \xi_+)$ and $(U(N), \xi_-)$ where $\xi_\pm$ are the embeddings of Sect. 2.3.

2.5 Transfer of Functions

From now until the end of Sect. 2.7, we assume that F is local. If G is an F-group, we denote $C_0^\infty(G)$ by $\mathcal{H}(G)$. We denote $C_0^\infty(\widetilde{G}(N))$ by $\widetilde{\mathcal{H}}(N)$. If G is a connected reductive group over F and $(G', \xi') \in \mathcal{E}(G)$, there is a correspondence between $\mathcal{H}(G)$ and $\mathcal{H}(G')$ known as the endoscopic transfer. More precisely, there is a nonempty subset of $\mathcal{H}(G')$ associated with any $f \in \mathcal{H}(G)$, and we let $f^{(G',\xi')}$ (which we will often abbreviate to $f^{G'}$) denote a choice of function from it. We say that $f^{(G',\xi')}$ is an endoscopic transfer of f to G'. The transfer is defined using orbital integrals on G and G' in a way that we do not need to make explicit in this paper. Its construction is primarily due to Shelstad in the real case, and Waldspurger [Wa3] in the p-adic case (assuming the fundamental lemma). See [Art, Sect. 2.1] for more details.

We shall require the fundamental lemma, due to Laumon and Ngô [LN, Ngo], Hales [Ha], Waldspurger [Wa2], and others. This states that if the local field F is p-adic, all data are unramified, and K and K' are hyperspecial maximal compact subgroups of G and G', then the characteristic functions 1_K and $1_{K'}$ correspond under endoscopic transfer.

There is a similar transfer in the twisted case. If $(G, \xi) \in \widetilde{\mathcal{E}}(N)$, this associates a function $f^{(G,\xi)} \in \mathcal{H}(G)$ with a function $f \in \widetilde{\mathcal{H}}(N)$. There is a twisted fundamental lemma, derived by Waldspurger in [Wa] from the untwisted case and his nonstandard variant, which states that the characteristic functions of hyperspecial maximal compact subgroups are associated by transfer if F is p-adic and all data are unramified.

2.6 Local Parameters

Let G be a connected reductive algebraic group over F. A Langlands parameter for G is an admissible homomorphism

$$\phi : L_F \to {}^L G.$$

We let $\Phi(G)$ denote the set of Langlands parameters up to conjugacy by $\widehat{G}$. An Arthur parameter for G is an admissible homomorphism

$$\psi : L_F \times SL(2, \mathbb{C}) \to {}^L G$$

such that the image of L_F in $\widehat{G}$ is bounded. We let $\Psi(G)$ denote the set of Arthur parameters modulo conjugacy by $\widehat{G}$, and let $\Psi^+(G)$ denote the set of parameters obtained by dropping this boundedness condition.

If $\psi \in \Psi^+(G)$, we define the following groups, which control the character identities for the local Arthur packet associated with ψ.

$$\begin{aligned} S_\psi &= \mathrm{Cent}(\mathrm{Im}\psi, \widehat{G}), \\ \overline{S}_\psi &= S_\psi / Z(\widehat{G})^{\Gamma_F}, \\ \mathcal{S}_\psi &= \pi_0(\overline{S}_\psi). \end{aligned}$$

In all cases we consider, we will have $\mathcal{S}_\psi \simeq (\mathbb{Z}/2\mathbb{Z})^r$ for some r. We also define

$$s_\psi = \psi\left(1, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}\right),$$

which is a central semi-simple element of S_ψ.

2.6.1 Endoscopic Data Associated with Arthur Parameters

There is a correspondence between pairs (G', ψ') with $G' \in \mathcal{E}(G)$ and $\psi' \in \Psi(G')$, and pairs (ψ, s) with $\psi \in \Psi(G)$ and s a semi-simple element of $\overline{S}_\psi$. (Note that we place a stronger equivalence relation on G' here than the usual equivalence of endoscopic data; see [Mo, Sect. 3.2] for details.) In one direction, this correspondence associates with a pair (G', ψ') (where G' is an abbreviation of (G', s', ξ')) the pair (ψ, s), where $\psi = \xi' \circ \psi'$ and s is the image of s' in $\overline{S}_\psi = S_\psi / Z(\widehat{G})^{\Gamma_F}$.

Conversely, suppose we have a pair (ψ, s). Let s' be any lift of s to S_ψ. We set $\widehat{G}' = \mathrm{Cent}(s', \widehat{G})^0$. Because $\psi(W_F)$ commutes with s' it normalizes $\widehat{G}'$, and this action allows us to define an L-group ${}^L G'$. We may combine $\psi|_{W_F}$ and the inclusion

$\widehat{G}' \subset \widehat{G}$ to obtain an L-embedding $\xi' : {}^L G' \to {}^L G$, which gives an endoscopic datum (G', s', ξ'). Because ψ factors through $\xi'({}^L G')$, this gives an L-parameter $\psi' \in \Psi(G')$.

2.6.2 Base Change Maps

We now discuss the map from parameters of $U(N)$ to parameters of $G(N)$ given by $\xi_\pm$. We first note that there is an isomorphism $\Phi(G(N)) \simeq \Phi(GL(N,E))$, which is given explicitly in [Mo, Sect. 2.2], and corresponds to the fact that both sets parametrize representations of $G(N)(F) \simeq GL(N,E)$. If $\phi \in \Phi(U(N))$, the parameter in $\Phi(GL(N,E))$ corresponding to $\xi_\pm \circ \phi$ under this isomorphism is just $\phi|_{L_E} \otimes \chi_\pm$. In particular, in the case of ξ_+ the parameter is just obtained by restriction to L_E (this is usually known as the standard base change map).

2.6.3 Parities of Local Parameters

One may characterize the image of $\Phi(U(N))$ in $\Phi(G(N))$ under $\xi_\pm$. We say that an admissible homomorphism $\rho : L_E \to GL(N,\mathbb{C})$ is conjugate self-dual if $\rho^c \simeq \rho^\vee$, where $\rho^c(\sigma) = \rho(w_c^{-1}\sigma w_c)$ for $\sigma \in L_E$ and $w_c \in W_F \setminus W_E$. There is a notion of parity for a conjugate self-dual representation [Mo, Sect. 2.2], which is analogous to a self-dual representation being either orthogonal (even) or symplectic (odd). We have the following characterization of the image of $\xi_\pm$ on parameters.

Lemma 2.1. *For $\kappa = \pm 1$, the image of*

$$\xi_\kappa : \Phi(U(N)) \to \Phi(G(N)) \simeq \Phi(GL(N,E))$$

is given by the parameters in $\Phi(GL(N,E))$ that are conjugate self-dual with parity $\kappa(-1)^N$.

2.7 Local Arthur Packets

In Sect. 2.5, Theorem 2.5.1, and Theorem 3.2.1 of [Mo], Mok associates a packet Π_ψ of representations of $U(N)$ with any $\psi \in \Psi^+(U(N))$. We recall some of the key features of this construction in the case when $\psi \in \Psi(U(N))$, which is all we shall need in this paper. The first step is to associate with any $\psi^N \in \widetilde{\Psi}(N)$ an irreducible unitary representation of $G(N)$, denoted π_{ψ^N}. We have the Langlands parameter ϕ_{ψ^N} associated with ψ^N, given by

$$\phi_{\psi^N}(\sigma) = \psi^N\left(\sigma, \begin{pmatrix} |\sigma|^{1/2} & 0 \\ 0 & |\sigma|^{-1/2} \end{pmatrix}\right), \quad \sigma \in L_F. \tag{3}$$

Let ρ_{ψ^N} be the standard representation of $G(N)$ associated with ϕ_{ψ^N}, and let π_{ψ^N} be its Langlands quotient. π_{ψ^N} is an irreducible admissible conjugate self-dual representation of $G(N) \simeq GL(N, E)$, and in [Mo, Sect. 3.2] Mok defines a canonical extension of π_{ψ^N} to $\widetilde{G}(N)^+$, denoted $\tilde{\pi}_{\psi^N}$. Mok defines a linear form on $\widetilde{\mathcal{H}}(N)$ by

$$\tilde{f} \mapsto \tilde{f}^N(\psi^N), \quad \tilde{f} \in \widetilde{\mathcal{H}}(N)$$
$$\tilde{f}^N(\psi^N) = \operatorname{tr} \tilde{\pi}_{\psi^N}(\tilde{f}).$$

If $G \in \mathcal{E}(U(N))$ and $\psi \in \Psi(G)$, Mok defines a linear form

$$f \mapsto f^G(\psi), \quad f \in \mathcal{H}(G). \tag{4}$$

In the case $G = U(N)$, Mok characterizes $f^G(\psi)$ as a transfer of the linear form $\tilde{f}^N(\xi \circ \psi)$ for $\xi = \xi_\pm$.

Proposition 2.2 (Theorem 3.2.1(a) of Mok [Mo]). *Let $G = U(N)$, and let $\psi \in \Psi(G)$. For either of the embeddings $\xi_\pm$, we have*

$$\tilde{f}^G(\psi) = \tilde{f}^N(\xi_\pm \circ \psi), \quad \tilde{f} \in \widetilde{\mathcal{H}}(N),$$

where $\tilde{f}^G(\psi)$ denotes the evaluation of the linear form $f^G(\psi)$ on the transfer of $\tilde{f}$ to $\mathcal{H}(G)$ associated with $\xi_\pm$.

Proposition 2.2 in fact gives a definition of $f^G(\psi)$ when $G = U(N)$, because both transfer mappings $\widetilde{\mathcal{H}}(N) \to \mathcal{H}(U(N))$ associated with $\xi_\pm$ are surjective by Mok [Mo, Proposition 3.1.1(b)]. As a general $G \in \mathcal{E}(U(N))$ is a product of the groups $U(M)$ and $G(M)$, and the definition of $f^G(\psi)$ is easy for $G(M)$ because it is a general linear group, this can be used to define $f^G(\psi)$ for all G. We will only need to consider the case where G is a product of two unitary groups in this paper.

We shall use the following character identities, which relate the linear forms $f^G(\psi)$ to traces of irreducible representations of $U(N)$.

Proposition 2.3 (Theorem 3.2.1(b) of Mok [Mo]). *Let $\psi \in \Psi(U(N))$. There exists a finite multi-set Π_ψ whose elements are irreducible admissible representations of $U(N)$, and a mapping*

$$\Pi_\psi \to \widehat{\mathcal{S}}_\psi$$
$$\pi \mapsto \langle \cdot, \pi \rangle$$

with the following property. If $s \in S_\psi$, and (G', ψ') is the element of $\mathcal{E}(U(N))$ corresponding to (ψ, s) as in Sect. 2.6.1, then we have

$$f^{G'}(\psi') = \sum_{\pi \in \Pi_\psi} \langle s_\psi s, \pi \rangle \operatorname{tr} \pi(f), \quad f \in \mathcal{H}(U(N)).$$

Here we have identified $s_\psi s$ with its image in $\mathcal{S}_\psi$, and $f^{G'}(\psi')$ denotes the evaluation of the linear form $f^{G'}(\psi')$ on the transfer of f to $\mathcal{H}(G')$.

The multiset Π_ψ is referred to as the Arthur packet associated with ψ. Note that if we set $s = 1$ in Proposition 2.3, then we obtain an expression for $f^{U(N)}(\psi)$ in terms of traces of the representations in Π_ψ. Because we always have $s_\psi^2 = 1$, if we set $s = s_\psi$ we obtain

$$f^{G'}(\psi') = \sum_{\pi \in \Pi_\psi} \operatorname{tr} \pi(f), \quad f \in \mathcal{H}(U(N)).$$

We will use this to bound $\sum_{\pi \in \Pi_\psi} \dim \pi^K$ for various compact open subgroups K of $U(N)(F)$.

2.8 Global Parameters

We now discuss the global version of the constructions of Sects. 2.6 and 2.7. For the rest of Sect. 2 we assume that F is global. The main difficulty in adapting these constructions is that we do not have a global analogue of the Langlands group L_F. However, if L_F existed, its irreducible N-dimensional representations would correspond to cusp forms on GL_N. Therefore, instead of considering representations of $L_F \times SL(2, \mathbb{C})$, Mok considers formal linear combinations of products of GL_N cusp forms with representations of $SL(2, \mathbb{C})$, and parametrizes the spectrum of $U(N)$ using these.

For $n \geq 1$, we let $\nu(n)$ denote the unique irreducible (complex-) algebraic representation of $SL(2, \mathbb{C})$ of dimension n. We let $\Psi_{\mathrm{sim}}(N)$ denote the set of simple global Arthur parameters, which are formal expressions $\psi^N = \mu \boxtimes \nu$ where μ is a unitary cuspidal automorphic representation of $GL(m, \mathbb{A}_E)$ and $\nu = \nu(n)$ for some n, and $N = mn$. We let $\Psi(N)$ denote the set of global Arthur parameters, which are formal expressions

$$\psi^N = \psi_1^{N_1} \boxplus \cdots \boxplus \psi_r^{N_r}$$

with $\psi_i^{N_i} \in \Psi_{\mathrm{sim}}(N_i)$ and $N_1 + \cdots + N_r = N$. If $\psi^N = \mu \boxtimes \nu \in \Psi_{\mathrm{sim}}(N)$, we define its conjugate dual to be $\psi^{N,*} = \mu^* \boxtimes \nu$, where μ^* is the conjugate dual representation to μ, and say that ψ^N is conjugate self-dual if $\psi^N = \psi^{N,*}$. We denote the set of conjugate self-dual parameters in $\Psi_{\mathrm{sim}}(N)$ by $\widetilde{\Psi}_{\mathrm{sim}}(N)$. We extend these notions to $\Psi(N)$ by defining the conjugate dual of

$$\psi^N = \psi_1^{N_1} \boxplus \cdots \boxplus \psi_r^{N_r} \in \Psi(N)$$

to be

$$\psi^{N,*} = \psi_1^{N_1,*} \boxplus \cdots \boxplus \psi_r^{N_r,*}.$$

We denote the set of conjugate self-dual parameters in $\Psi(N)$ by $\widetilde{\Psi}(N)$. Note that requiring $\psi^N \in \Psi(N)$ to be conjugate self-dual is not the same as requiring that $\psi_i^{N_i} = \psi_i^{N_i,*}$ for all i, as we are free to rearrange the terms. We say that

$$\psi^N = \psi_1^{N_1} \boxplus \cdots \boxplus \psi_r^{N_r} \in \widetilde{\Psi}(N)$$

is elliptic if the $\psi_i^{N_i}$ are distinct and $\psi_i^{N_i} = \psi_i^{N_i,*}$ for all i, and denote the set of elliptic parameters by $\widetilde{\Psi}_{\text{ell}}(N)$. We denote the set of generic parameters, that is those for which all the representations ν are trivial, by $\Phi(N)$, and define $\widetilde{\Phi}_*(N) = \widetilde{\Psi}_*(N) \cap \Phi(N)$. It follows that we have chains of parameters

$$\widetilde{\Psi}_{\text{sim}}(N) \subseteq \widetilde{\Psi}_{\text{ell}}(N) \subseteq \widetilde{\Psi}(N), \quad \text{and}$$
$$\widetilde{\Phi}_{\text{sim}}(N) \subseteq \widetilde{\Phi}_{\text{ell}}(N) \subseteq \widetilde{\Phi}(N).$$

To any parameter $\psi^N \in \widetilde{\Psi}(N)$, Mok [Mo, Sect. 2.4] associates a group $\mathcal{L}_{\psi^N}$ that is an extension of W_F by a complex algebraic group, and an L-homomorphism $\tilde{\psi}^N : \mathcal{L}_{\psi^N} \times SL(2,\mathbb{C}) \to {}^L G(N)$. We will not recall the definition of these objects, and give a qualitative description of them instead. If we think of ψ^N as corresponding to a hypothetical representation of $L_F \times SL(2,\mathbb{C})$, $\mathcal{L}_{\psi^N}$ would contain the image of this representation. Because of this, we will use $\mathcal{L}_{\psi^N}$ and $\tilde{\psi}^N$ to define what it means for ψ^N to factor through the maps $\xi_\pm : {}^L U(N) \to {}^L G(N)$, and thus give a parameter for $U(N)$.

If $(U(N), \xi_\pm) \in \widetilde{\mathcal{E}}_{\text{sim}}(N)$, we define $\Psi(U(N), \xi_\pm)$ to be the set of pairs $\psi = (\psi^N, \tilde{\psi})$, where $\psi^N \in \widetilde{\Psi}(N)$ and

$$\tilde{\psi} : \mathcal{L}_{\psi^N} \times SL(2,\mathbb{C}) \to {}^L U(N)$$

is an L-homomorphism such that $\tilde{\psi}^N = \xi_\pm \circ \tilde{\psi}$. If $\psi = (\psi^N, \tilde{\psi}) \in \Psi(U(N), \xi_\pm)$, we set $\mathcal{L}_\psi = \mathcal{L}_{\psi^N}$.

2.8.1 Parities of Global Parameters

If $\phi^N \in \widetilde{\Phi}_{\text{sim}}(N)$ is associated with a conjugate self-dual cusp form μ, Theorem 2.4.2 of Mok [Mo] states that there is a unique base change map ξ_κ with $\kappa = \pm$ such that μ is the weak base change of a representation of $U(N)$ under ξ_κ. Following Mok, we refer to $\kappa(-1)^{N-1}$ as the parity of ϕ^N and μ. We may extend this definition to $\psi^N = \mu \boxtimes \nu \in \widetilde{\Psi}_{\text{sim}}(N)$ as follows: if we assume that μ is a base change under ξ_δ, we define $\kappa = \delta(-1)^{N-m-n+1}$, and define $\kappa(-1)^{N-1}$ to be the parity of ψ^N. It follows from these definitions that the parity of $\mu \boxtimes \nu$ is the product of the parities of μ and ν, where the parity of $\nu(n)$ is defined to be opposite to the parity of n (corresponding to the fact that $\nu(n)$ is orthogonal if n is odd and symplectic if n is even).

This is compatible with the notion of parity discussed in Sect. 2.6.3. In particular, if $\psi \in \widetilde{\Psi}_{\mathrm{sim}}(N)$ has invariant κ, then the L-homomorphism $\mathcal{L}_\psi \times SL(2,\mathbb{C}) \to {}^L G(N)$ factors through ξ_κ. In particular, if $\psi^N \in \widetilde{\Phi}_{\mathrm{sim}}(N)$, then $\mathcal{L}_{\psi^N} = {}^L U(N)$ and $\tilde{\psi}^N$ is the product of ξ_κ with the trivial map on $SL(2,\mathbb{C})$. We will also see in Sect. 2.9 that if v is nonsplit in E, the localization $\psi_v^N : L_{F_v} \times SL(2,\mathbb{C}) \to {}^L G_{E_v/F_v}(N)$ of ψ^N factors through the local base change map $\xi_{\pm,v}$.

2.8.2 Square-Integrable Parameters

We define $\Psi_2(U(N),\xi_\pm)$ to be the subset of $\Psi(U(N),\xi_\pm)$ for which $\psi^N \in \widetilde{\Psi}_{\mathrm{ell}}(N)$. This is known as the set of square-integrable parameters of $U(N)$ with respect to $\xi_\pm$, because these are the parameters that give the discrete automorphic spectrum of $U(N)$. In concrete terms, a parameter $\psi^N \in \widetilde{\Psi}_{\mathrm{ell}}(N)$ can be extended to $\psi = (\psi^N, \tilde{\psi}) \in \Psi_2(U(N),\xi_\kappa)$ if and only if $\psi^N = \psi_1^{N_1} \boxplus \ldots \boxplus \psi_l^{N_l}$ with the parameters $\psi_i^{N_i} \in \widetilde{\Psi}_{\mathrm{sim}}(N_i)$ all having parity $\kappa(-1)^{N-1}$. More concretely, if $\psi_i^{N_i} = \phi_i \boxtimes \nu(n_i)$ with $\phi_i \in \widetilde{\Phi}_{\mathrm{sim}}(m_i)$, we require that $\delta_i(-1)^{m_i+n_i} = \kappa(-1)^N$ for all i, where δ_i is such that the cusp form μ_i associated with ϕ_i is a weak base change from $U(m_i)$ under ξ_{δ_i}.

2.9 Localization of Parameters

Having introduced global and local versions of our parameters, we now discuss the localization maps taking the former to the latter. We let v be a place of F, and let $E_v = E \otimes_F F_v$, $U(N)_v = U_{E_v/F_v}$, and $G(N)_v = G_{E_v/F_v}(N)$.

We first assume that v does not split in E. Consider a simple generic parameter $\phi^N \in \Phi_{\mathrm{sim}}(N)$. As ϕ^N corresponds to a cusp form μ on $GL(N,\mathbb{A}_E)$, we may consider the local factor μ_v, which is an irreducible unitary representation of $GL(N,E_v)$. By the local Langlands correspondence for $GL(N)$ by Harris-Taylor [HT] and Henniart [Hen], and the isomorphism $\Phi(GL(N,E_v)) \simeq \Phi(G(N)_v)$ of Sect. 2.6.2, μ_v corresponds to a local Langlands parameter $\phi_v^N \in \Phi_v(N) := \Phi(G(N)_v)$. This gives the localization map from $\Phi_{\mathrm{sim}}(N)$ to $\Phi_v(N)$, which takes $\widetilde{\Phi}_{\mathrm{sim}}(N)$ to $\widetilde{\Phi}_v(N)$. This may be naturally extended to a map $\psi^N \mapsto \psi_v^N$ from $\Psi(N)$ to $\Psi_v^+(N)$ that takes $\widetilde{\Psi}(N)$ to $\widetilde{\Psi}_v^+(N)$.

Now consider a parameter $\psi = (\psi^N, \tilde{\psi}) \in \Psi(U(N),\xi_\kappa)$. By Mok [Mo, Corollary 2.4.11], the localization ψ_v^N factors through the embedding $\xi_{\kappa,v} : {}^L U(N)_v \to {}^L G(N)_v$. This allows us to define $\psi_v \in \Psi(U(N)_v)$ by requiring that $\xi_{\kappa,v} \circ \psi_v = \psi_v^N$.

We now assume that v splits in E, and write $v = w\overline{w}$. As in Sect. 2.2 we have isomorphisms $\iota_w : U(N)_v \to GL(N, E_w)$ and $\iota_{\overline{w}} : U(N)_v \to GL(N, E_{\overline{w}})$ corresponding to the projections of E_v to E_w and $E_{\overline{w}}$. If $\psi = (\psi^N, \tilde{\psi}) \in \Psi(U(N), \xi_\kappa)$, we may think of the localisations ψ_w^N and $\psi_{\overline{w}}^N$ as elements of $\Psi^+(GL(N, E_w))$ and $\Psi^+(GL(N, E_{\overline{w}}))$. When $\psi_w^N \in \Psi(GL(N, E_w))$, we may define $\pi_{\psi_w^N}$ to be the representation associated with $\phi_{\psi_w^N}$ by local Langlands, where $\phi_{\psi_w^N}$ is as in (3). The definition of $\pi_{\psi_w^N}$ for $\psi_w^N \in \Psi^+(GL(N, E_w))$ is given in [Mo, Sect. 2.4], and will not be needed in this paper because the GL_N cusp forms we consider are known to satisfy the Ramanujan conjectures.

The conjugate self-duality of ψ^N implies that $\pi_{\psi_w^N} = (\pi_{\psi_{\overline{w}}^N})^\vee$, and $\iota_w \circ \iota_{\overline{w}}^{-1}$ is the automorphism $g \mapsto J^t g^{-1} J^{-1}$ of Sect. 2.2 (under the identification $E_w = E_{\overline{w}} = F_v$). Therefore the pullback of $\pi_{\psi_w^N}$ via ι_w is isomorphic to the pullback of $\pi_{\psi_{\overline{w}}^N}$ via $\iota_{\overline{w}}$. We denote this representation of $U(N)_v$ by π_{ψ_v}. We define $\psi_v \in \Psi^+(U(N)_v)$ to be the parameter obtained by composing $\psi_w^N : L_{F_v} \times SL(2, \mathbb{C}) \simeq L_{E_w} \times SL(2, \mathbb{C}) \to {}^L GL(N, E_w)$ with the isomorphism ${}^L\iota_w : {}^L GL(N, E_w) \to {}^L U(N)_v$ induced by ι_w. We define $\Pi_{\psi_v} = \{\pi_{\psi_v}\}$ to be the local Arthur packet associated with ψ_v.

2.10 The Global Classification

We may now state the global classification theorem. For any ψ in the set of global parameters $\Psi_2(U(N), \xi_\pm)$, we have the localizations ψ_v and the local Arthur packets Π_{ψ_v} associated with ψ_v in Sects. 2.7 and 2.9. We define the global Arthur packet Π_ψ to be the restricted direct product of the Π_{ψ_v}, in the sense that it contains those $\otimes_v \pi_v \in \otimes_v \Pi_{\psi_v}$ such that the (global analogue of the) character $\langle \cdot, \pi_v \rangle$ is trivial for almost all v. We will write $\Pi_\psi = \otimes_v \Pi_{\psi_v}$ by slight abuse of notation. In [Mo, Sect. 2.5], Mok defines a subset $\Pi_\psi(\epsilon_\psi) \subset \Pi_\psi$ in terms of symplectic root numbers and the pairings in Proposition 2.3, which we do not need to make explicit. The classification is as follows.

Theorem 2.4. *For $\kappa = \pm 1$, we have a $U(N)(\mathbb{A})$-module decomposition of the discrete automorphic spectrum of $U(N)$:*

$$L^2_{\mathrm{disc}}(U(N)(F) \backslash U(N)(\mathbb{A})) = \sum_{\psi \in \Psi_2(U(N), \xi_\kappa)} \sum_{\pi \in \Pi_\psi(\epsilon_\psi)} \pi.$$

Mok's proof of Theorem 2.4 builds on work by many authors, notably Arthur, who classified the discrete spectrum of quasi-split symplectic and orthogonal groups in [Art], and Moeglin and Waldspurger, who proved the stabilization of the twisted trace formula. Theorem 2.4 is being extended to general forms of unitary groups by Kaletha et al. in [KMSW] and its projected sequels. In joint work with Shin, we hope to show that this extension of Theorem 2.4 implies strong (and conjecturally sharp) upper bounds for cohomology growth on arithmetic manifolds associated with $U(n, 1)$ for any n.

3 Application of the Global Classification

In this section, we rephrase Theorem 1.1 in terms of Arthur packets by applying the results of Sect. 2 to the manifolds $Y(n)$.

3.1 Notation

Let E be an imaginary quadratic field with ring of integers $\mathcal{O}$. We apply the notation of Sect. 2 to the extension $E/\mathbb{Q}$. We denote places of $\mathbb{Q}$ and E by v and w, respectively. We recall the character χ of $E^\times \backslash \mathbb{A}_E^\times$ whose restriction to $\mathbb{A}^\times$ is the character associated with $E/\mathbb{Q}$ by class field theory. We let S_f be a finite set of finite places of $\mathbb{Q}$ that contains all finite places at which E is ramified, and all finite places that are divisible by a place of E at which χ is ramified.

If G is an algebraic group over $\mathbb{Q}$ or $\mathbb{Q}_v$, we denote $G(\mathbb{Q}_v)$ by G_v, and likewise for groups over E. For any $N \geq 1$ we let $\widetilde{G}(N)_v = G(N)_v \rtimes \theta$, and $\widetilde{\mathcal{H}}_v(N) = C_0^\infty(\widetilde{G}(N)_v)$. We fix Haar measures on $U(N)_v$ and $\widetilde{G}(N)_v$ for all $N \geq 1$ and all v, subject to the condition that these measures assign volume 1 to a hyperspecial maximal compact when v is finite and the groups are unramified. All traces and twisted traces will be defined with respect to these measures.

We shall identify the infinitesimal character of an irreducible admissible representation of $U(N)_\infty$ and $GL(N, \mathbb{C})$ with a point in $\mathbb{C}^N/S_N$ and $(\mathbb{C}^N/S_N) \times (\mathbb{C}^N/S_N)$ respectively, where S_N is the symmetric group.

We choose a compact open subgroup $K = \prod_p K_p \subset U(4)(\mathbb{A}_f)$, subject to the condition that $K_p = U(4)(\mathbb{Z}_p)$ for $p \notin S_f$. For any $n \geq 1$ that is relatively prime to S_f, we define $K_p(n)$ to be the subgroup of K_p consisting of elements congruent to 1 modulo n when $p \notin S_f$, and $K_p(n) = K_p$ otherwise, and define $K(n) = \prod_p K_p(n)$.

We let K_∞ be the standard maximal compact subgroup of $U(4)_\infty$. For any $n \geq 1$ that is relatively prime to S_f, we define $Y(n) = U(4)(\mathbb{Q})\backslash U(4)(\mathbb{A})/K_\infty K(n)$. For any $0 \leq i \leq 8$, we let $h^i_{(2)}(Y(n))$ denote the dimension of the space of square integrable harmonic i-forms on $Y(n)$.

3.2 Reduction of Theorem 1.1 to Arthur Packets

The precise form of Theorem 1.1 we shall prove is the following.

Theorem 3.1. *If $i = 2, 3$, and n is relatively prime to S_f and divisible only by primes that split in E, we have $h^i_{(2)}(Y(n)) \ll n^9$.*

The implied constant depends only on K, and we shall ignore the dependence of implied constants on K for the rest of the paper. By considering the action of the center on the connected components of $Y(n)$, Theorem 3.1 implies that the connected component $Y^0(n)$ of the identity satisfies $h^i_{(2)}(Y^0(n)) \ll_\epsilon n^{8+\epsilon}$. This implies Theorem 1.1 when combined with the asymptotic $\mathrm{Vol}(Y^0(n)) = n^{15+o(1)}$.

We shall only prove Theorem 3.1 in the case $i = 3$, as the case $i = 2$ is identical. We begin by applying the extension of Matsushima's formula to noncompact quotients [BG, Proposition 5.6], which gives

$$h^3_{(2)}(Y(n)) = \sum_{\pi \in L^2_{\mathrm{disc}}(U(4)(\mathbb{Q})\backslash U(4)(\mathbb{A}))} h^3(\mathfrak{g}, K; \pi_\infty) \dim \pi_f^{K(n)}. \tag{5}$$

If we combine this with Theorem 2.4, we obtain

$$h^3_{(2)}(Y(n)) \le \sum_{\psi \in \Psi_2(U(4),\xi_+)} \sum_{\pi \in \Pi_\psi} h^3(\mathfrak{g}, K; \pi_\infty) \dim \pi_f^{K(n)}. \tag{6}$$

It follows from the proof of the Adams–Johnson conjectures in [AMR], or Proposition 13.4 of Bergeron et al. [BMM], that if $\pi \in \Pi_\psi$ satisfies $h^3(\mathfrak{g}, K; \pi_\infty) \neq 0$, then ψ is not generic. It follows that ψ^N must be of one of the following types.

1. $\nu(2) \boxtimes \phi_1^N \boxplus \phi_2^N$, $\phi_i^N \in \widetilde{\Phi}_{\mathrm{ell}}(i)$.
2. $\nu(2) \boxtimes \phi^N$, $\phi^N \in \widetilde{\Phi}_{\mathrm{ell}}(2)$.
3. $\nu(3) \boxtimes \phi_1^N \boxplus \phi_2^N$, $\phi_i^N \in \widetilde{\Phi}(1)$.
4. $\nu(4) \boxtimes \phi^N$, $\phi^N \in \widetilde{\Phi}(1)$.

We bound the contribution of parameters of types (1) and (2) in Sects. 4 and 5, respectively. It follows from the description of the packets Π_ψ at split places that all representations contained in packets of type (4) must be characters, and these make a contribution of $\ll_\epsilon n^{1+\epsilon}$ to $h^3_{(2)}(Y(n))$. We shall also omit the case of parameters of type (3); it may be proven that they make a contribution of $\ll_\epsilon n^{5+\epsilon}$ using the same methods as in Sect. 5.

4 The Case $\psi^N = \nu(2) \boxtimes \phi_1^N \boxplus \phi_2^N$

Let $h^3_{(2)}(Y(n))^\star$ denote the contribution to $h^3_{(2)}(Y(n))$ from parameters of the form $\nu(2) \boxtimes \phi_1^N \boxplus \phi_2^N$, which by (6) satisfies

$$h^3_{(2)}(Y(n))^\star \le \sum_{\substack{\psi \in \Psi_2(U(4),\xi_+) \\ \psi^N = \nu(2)\boxtimes\phi_1^N \boxplus \phi_2^N}} \sum_{\pi \in \Pi_\psi} h^3(\mathfrak{g}, K; \pi_\infty) \dim \pi_f^{K(n)}. \tag{7}$$

We assume that the sum is restricted to those ϕ_2^N lying in $\widetilde{\Phi}_{\mathrm{sim}}(2)$ until the end of Sect. 4.2, and describe how to treat composite ϕ_2^N in Sect. 4.3. We note that $\psi \in \Psi_2(U(4), \xi_+)$ implies that ϕ_1^N and ϕ_2^N must be even and odd, respectively. The main result of this section is the following.

Proposition 4.1. *We have the bound* $h^3_{(2)}(Y(n))^\star \ll n^9$.

For $i = 1, 2$, we let $K_i = \prod_p K_{i,p}$ be a compact open subgroup of $U(i)(\mathbb{A}_f)$ such that $K_{i,p} = U(i)(\mathbb{Z}_p)$ for all $p \notin S_f$, and let $\widetilde{K}_i = \prod_w \widetilde{K}_{i,w}$ be a compact open subgroup of $GL(i, \mathbb{A}_{E,f})$ such that $\widetilde{K}_{i,w} = GL(i, \mathcal{O}_w)$ for all $w|p \notin S_f$. We define $\widetilde{K} \subset GL(4, \mathbb{A}_{E,f})$ in a similar way. The groups $K_{2,p}$ and $\widetilde{K}_{1,w}$ for $w|p \in S_f$ will be specified in the proof of Proposition 4.2, and the groups $K_{1,p}$, $\widetilde{K}_{2,w}$, and $\widetilde{K}_w$ for $w|p \in S_f$ may be chosen arbitrarily. We define congruence subgroups $K_*(n)$ of these groups for n relatively prime to S_f in the usual way, and recall that n will only be divisible by primes that split in E.

We let $\widetilde{P}$ be the standard parabolic subgroup of $GL(4, E)$ with Levi $\widetilde{L} = GL(2, E) \times GL(2, E)$, and let P be the corresponding standard parabolic subgroup of $U(4)$.

4.1 Controlling a Single Parameter

We first bound the contribution from a single Arthur parameter to $h^3_{(2)}(Y(n))^\star$. We therefore fix $\phi_i^N \in \widetilde{\Phi}_{\mathrm{sim}}(i)$ for $i = 1, 2$ with ϕ_1^N even and ϕ_2^N odd, and let $\psi \in \Psi(U(4), \xi_+)$ be the unique parameter with $\psi^N = \nu(2) \boxtimes \phi_1^N \boxplus \phi_2^N$. We let ϕ_i^N correspond to a conjugate self-dual cuspidal automorphic representation μ_i of $GL(i, \mathbb{A}_E)$. We assume that μ_i are tempered at all places. This assumption is not necessary, but simplifies the proof of Proposition 4.2 and will be proven to hold for all parameters that contribute to cohomology.

We define $\psi_1^N = \nu(2) \boxtimes \phi_1^N$ and $\psi_2^N = \phi_2^N$, and for $i = 1, 2$ we let $\psi_i \in \Psi(U(2), \xi_+)$ be the corresponding unitary parameters. We shall prove the following bound for the finite part of the contribution of Π_ψ to $h^3_{(2)}(Y(n))^\star$.

Proposition 4.2. *There is a choice of* $\widetilde{K}_{1,w}$ *for* $w|p$, $p \in S_f$, *and* $K_{2,p}$ *for* $p \in S_f$, *depending only on* K, *such that*

$$\sum_{\pi_f \in \Pi_{\psi,f}} \dim \pi_f^{K(n)} \ll [K : (K \cap P(\mathbb{A}_f))K(n)] \dim \mu_1^{\widetilde{K}_1(n)} \sum_{\pi'_f \in \Pi_{\psi_2,f}} \dim \pi_f'^{K_2(n)},$$

where $\Pi_{\psi,f} = \otimes_p \Pi_{\psi_p}$ *is the finite part of* Π_ψ, *and likewise for* Π_{ψ_2}.

The proposition will follow from the factorization of $\Pi_{\psi,f}$, and the series of lemmas below.

Lemma 4.3. *Let* $p \notin S_f$ *be nonsplit in* E, *and let* $w|p$. *We have*

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} = \dim \mu_{1,w}^{\widetilde{K}_{1,w}} \sum_{\pi'_p \in \Pi_{\psi_{2,p}}} \dim \pi_p'^{K_{2,p}}.$$

Proof. We have

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} = \sum_{\pi_p \in \Pi_{\psi_p}} \mathrm{tr}(\pi_p(1_{K_p})),$$

and we may manipulate the right-hand side using the local character identities of Propositions 2.2 and 2.3. Let $(G', \xi') \in \mathcal{E}_{\mathrm{ell}}(U(4)_p)$ be the unique endoscopic datum with $G' = U(2)_p \times U(2)_p$, and let $\psi'_p = \psi_{1,p} \times \psi_{2,p} \in \Psi(G')$. It may be seen that (G', ψ'_p) is the pair associated with (ψ_p, s_ψ) by the correspondence of Sect. 2.6.1. We recall the distribution $f \mapsto f^{G'}(\psi'_p)$ on $\mathcal{H}(G')$ associated with ψ'_p in (4). Applying Proposition 2.3 with $s = s_{\psi_p}$, and the fundamental lemma for the group $G' \in \mathcal{E}(U(4)_p)$, gives

$$\sum_{\pi_p \in \Pi_{\psi_p}} \mathrm{tr}(\pi_p(1_{K_p})) = (1_{K_{2,p}} \times 1_{K_{2,p}})^{G'}(\psi'_p).$$

Because $\psi'_p = \psi_{1,p} \times \psi_{2,p}$, the factorization property of the linear form $f^{G'}(\psi'_p)$ allows us to write this as

$$\sum_{\pi_p \in \Pi_{\psi_p}} \mathrm{tr}(\pi_p(1_{K_p})) = 1_{K_{2,p}}^{U(2)}(\psi_{1,p}) 1_{K_{2,p}}^{U(2)}(\psi_{2,p}),$$

where $f \mapsto f^{U(2)}(\psi_{i,p})$ are the distributions on $\mathcal{H}(U(2)_p)$ associated with $\psi_{i,p}$. Because $s_{\psi_{i,p}} = e$ for $i = 1, 2$, we may express $1_{K_{2,p}}^{U(2)}(\psi_{i,p})$ in terms of traces of representations by applying Proposition 2.3 with $s = e$, which gives

$$1_{K_{2,p}}^{U(2)}(\psi_{i,p}) = \sum_{\pi'_p \in \Pi_{\psi_{i,p}}} \mathrm{tr}(\pi'_p(1_{K_{2,p}})) = \sum_{\pi'_p \in \Pi_{\psi_{i,p}}} \dim \pi_p'^{K_{2,p}}. \tag{8}$$

This gives the required expression for $1_{K_{2,p}}^{U(2)}(\psi_{2,p})$.

We evaluate $1_{K_{2,p}}^{U(2)}(\psi_{1,p})$ by applying Proposition 2.2 with the embedding ξ chosen to be $\xi_+ : {}^L U(2)_p \to {}^L G(2)_p$. If we restrict the map

$$\xi_+ \circ \psi_{1,p} : L_{\mathbb{Q}_p} \times SL(2, \mathbb{C}) \to {}^L G(2)_p$$

to $L_{E_w} \times SL(2, \mathbb{C})$, it is equivalent to

$$\begin{aligned} \xi_+ \circ \psi_{1,p} : L_{E_w} \times SL(2, \mathbb{C}) &\to GL(2, \mathbb{C}) \\ \sigma \times A &\mapsto \phi_{1,w}^N(\sigma) A. \end{aligned}$$

It follows that the representation of $G(2)_p \simeq GL(2, E_w)$ associated with $\xi_+ \circ \psi_{1,p}$ is equal to $\mu_{1,w} \circ \det$. We denote the canonical extension of this representation to

$\widetilde{G}^+(2)_p$ by $\tilde{\pi}_1$. If we identify $\widetilde{K}_{2,w}$ with a subgroup of $G(2)_p$, the twisted fundamental lemma implies that we may take $\tilde{f} = 1_{\widetilde{K}_{2,w} \rtimes \theta} \in \widetilde{\mathcal{H}}_p(2)$ in Proposition 2.2 to obtain

$$1_{K_{2,p}}^{U(2)}(\psi_{1,p}) = \mathrm{tr}(\tilde{\pi}_1(1_{\widetilde{K}_{2,w} \rtimes \theta})).$$

Because $\theta^2 = 1$, we have

$$\mathrm{tr}(\tilde{\pi}_1(1_{\widetilde{K}_{2,w} \rtimes \theta})) = \pm \dim \tilde{\pi}_1^{\widetilde{K}_{2,w}} = \pm \dim \mu_{1,w}^{\widetilde{K}_{1,w}}.$$

Applying Eq. (8) with $i = 1$ implies that $1_{K_{2,p}}^{U(2)}(\psi_{1,p}) \geq 0$, which means that we must take the positive sign. This completes the proof. □

Lemma 4.4. *Let $p \notin S_f$ be split in E, and let $w|p$. Let $\Pi_{\psi_p} = \{\pi_p\}$, and $\Pi_{\psi_{2,p}} = \{\pi'_p\}$. We have*

$$\dim \pi_p^{K_p(n)} = [K_p : (K_p \cap P_p)K_p(n)] \dim \mu_{1,w}^{\widetilde{K}_{1,w}(n)} \dim \pi_p'^{K_{2,p}(n)}.$$

Proof. Under the identification $U(4)_p \simeq GL(4, E_w)$, the discussion of Sect. 2.9 implies that π_p is isomorphic to the representation induced from the representation $(\mu_{1,w} \circ \det) \otimes \mu_{2,w}$ of $\widetilde{P}_w$. The restriction of π_p to K_p is isomorphic to the induction of $(\mu_{1,w} \circ \det) \otimes \mu_{2,w}$ from $\widetilde{P}_w \cap \widetilde{K}_w$ to $\widetilde{K}_w$. Because $\widetilde{K}_w(n) \cap \widetilde{L}_w = \widetilde{K}_{2,w}(n) \times \widetilde{K}_{2,w}(n)$, and $\dim(\mu_{1,w} \circ \det)^{\widetilde{K}_{2,w}(n)} = \dim \mu_{1,w}^{\widetilde{K}_{1,w}(n)}$, we have

$$\dim \pi_p^{K_p(n)} = [\widetilde{K}_w : (\widetilde{K}_w \cap \widetilde{P}_w)\widetilde{K}_w(n)] \dim \mu_{1,w}^{\widetilde{K}_{1,w}(n)} \dim \pi_p'^{K_{2,p}(n)}$$

which is equivalent to the lemma. □

Lemma 4.5. *Let $p \in S_f$, and let $w|p$. There is a choice of $\widetilde{K}_{1,w}$ and $K_{2,p}$, depending only on K_p, such that*

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} \ll \dim \mu_{1,w}^{\widetilde{K}_{1,w}} \sum_{\pi'_p \in \Pi_{\psi_{2,p}}} \dim \pi_p'^{K_{2,p}}.$$

Proof. If p is split, this follows from the explicit description of Π_{ψ_p} as in Lemma 4.4. Assume that p is nonsplit, and continue to use the notation of Lemma 4.3. Let $\widetilde{1}_{K_p} \in \mathcal{H}(G')$ be a transfer of 1_{K_p} to G'. Reasoning as in the proof of Lemma 4.3 gives

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} = \mathrm{vol}(K_p)^{-1} \widetilde{1}_{K_p}^{G'}(\psi'_p),$$

where $\mathrm{vol}(K_p)$ denotes the volume of K_p with respect to our chosen Haar measure on $U(4)_p$. We may write $\widetilde{1}_{K_p} = \sum f_{i,1} \times f_{i,2}$ for $f_{i,j} \in \mathcal{H}(U(2)_p)$, and the factorization property of $f^{G'}(\psi'_p)$ gives

$$\widetilde{1}_{K_p}^{G'}(\psi'_p) = \sum_i (f_{i,1} \times f_{i,2})^{G'}(\psi'_p)$$
$$= \sum_i f_{i,1}^{U(2)}(\psi_{1,p}) f_{i,2}^{U(2)}(\psi_{2,p}).$$

Applying Proposition 2.3 with $s = e$ gives

$$f_{i,2}^{U(2)}(\psi_{2,p}) = \sum_{\pi'_p \in \Pi_{\psi_{2,p}}} \mathrm{tr}(\pi'_p(f_{i,2}))$$
$$\leq C(f_{i,2}) \sum_{\pi'_p \in \Pi_{\psi_{2,p}}} \dim \pi_p'^{K_{2,p}}$$

if $K_{2,p}$ is chosen so that $f_{i,2}$ is bi-invariant under $K_{2,p}$ for all i. Likewise, applying Proposition 2.2 and the definition of $\tilde{\pi}_{\psi_{1,p}^N}$ shows that $f_{i,1}^{U(2)}(\psi_{1,p}) \leq C(f_{i,1}) \dim \mu_{1,w}^{\widetilde{K}_{1,w}}$ if $\widetilde{K}_{1,w}$ is chosen sufficiently small depending on $f_{i,1}$. As the collection of functions $f_{i,j}$ depended only on K_p, so do $\widetilde{K}_{1,w}$ and $K_{2,p}$, and the constant factors. □

4.2 Summing Over Parameters

We now use Proposition 4.2 to control the contribution to $h^3_{(2)}(Y(n))^\star$ from all ψ.

Lemma 4.6. *Let $\psi \in \Psi(U(4), \xi_+)$, and suppose that $\psi^N = \nu(2) \boxtimes \phi_1^N \boxplus \phi_2^N$ with $\phi_i^N \in \widetilde{\Phi}_{\mathrm{sim}}(i)$. If $\pi \in \Pi_{\psi_\infty}$ satisfies $H^*(\mathfrak{g}, K; \pi) \neq 0$, then we have*

$$\phi_{1,\infty}^N : z \mapsto (z/\overline{z})^{\alpha'}$$
$$\phi_{2,\infty}^N : z \mapsto \begin{pmatrix} (z/\overline{z})^{\alpha_1} & \\ & (z/\overline{z})^{\alpha_2} \end{pmatrix}$$

with $\alpha' \in \{1, 0, -1\}$, $\alpha_i \in \{3/2, 1/2, -1/2, -3/2\}$, and $\alpha_1 \neq \alpha_2$.

Proof. We write

$$\phi_{1,\infty}^N : z \mapsto z^{\alpha'} \overline{z}^{\beta'}$$
$$\phi_{2,\infty}^N : z \mapsto \begin{pmatrix} z^{\alpha_1} \overline{z}^{\beta_1} & \\ & z^{\alpha_2} \overline{z}^{\beta_2} \end{pmatrix}$$

with $\alpha' - \beta', \alpha_i - \beta_i \in \mathbb{Z}$. If we let ϕ_{ψ_∞} be the Langlands parameter associated with ψ_∞ as in (3), any $\pi \in \Pi_{\psi_\infty}$ has the same infinitesimal character as the representations in the L-packet of ϕ_{ψ_∞}, which is $(\alpha' + 1/2, \alpha' - 1/2, \alpha_1, \alpha_2) \in \mathbb{C}^4/S_4$ (see, for instance, [Vo, Proposition 7.4]). If π is to have cohomology, it must have the same infinitesimal character as the trivial representation, so that $\{\alpha' + 1/2, \alpha' - 1/2, \alpha_1, \alpha_2\} = \{3/2, 1/2, -1/2, -3/2\}$. This implies that $\alpha' \in \{1, 0, -1\}$ and $\alpha_i \in \{3/2, 1/2, -1/2, -3/2\}$ with $\alpha_1 \neq \alpha_2$. Because μ_1 is a character we have $\alpha' = -\beta'$, and because μ_2 is a cusp form on $GL(2, E)$ we have $|\alpha_i + \beta_i| < 1/2$ so that $\alpha_i = -\beta_i$. This completes the proof. □

For $i = 1, 2$, we define $\Phi_{\text{rel}}(i) \subset \widetilde{\Phi}_{\text{sim}}(i)$ to be the set of parameters ϕ_i^N such that $\phi_{i,\infty}^N$ satisfies the relevant constraints of Lemma 4.6. If $\phi_2^N \in \Phi_{\text{rel}}(2)$ is associated with a cuspidal representation μ, it follows that μ is regular algebraic, conjugate self-dual, and cuspidal, and hence tempered at all places by Theorem 1.2 of Caraiani [Ca].

Lemma 4.6 and Eq. (7) imply that

$$\begin{aligned} h^3_{(2)}(Y(n))^\star &\ll \sum_{\substack{\psi^N = \nu(2)\boxtimes\phi_1^N \boxplus \phi_2^N \\ \phi_i^N \in \Phi_{\text{rel}}(i)}} \sum_{\pi\in\Pi_\psi} \dim \pi_f^{K(n)} \\ &= \sum_{\substack{\psi^N = \nu(2)\boxtimes\phi_1^N \boxplus \phi_2^N \\ \phi_i^N \in \Phi_{\text{rel}}(i)}} \#(\Pi_{\psi_\infty}) \sum_{\pi_f\in\Pi_{\psi,f}} \dim \pi_f^{K(n)}. \end{aligned}$$

We may ignore the factor $\#(\Pi_{\psi_\infty})$ because there are only finitely many possibilities for ψ_∞. Applying Proposition 4.2 to the right-hand side gives

$$h^3_{(2)}(Y(n))^\star \ll [K : (K\cap P(\mathbb{A}_f))K(n)] \sum_{\phi_1^N\in\Phi_{\text{rel}}(1)} \dim \mu_1^{\widetilde{K}_1(n)} \sum_{\phi_2^N\in\Phi_{\text{rel}}(2)} \sum_{\pi'_f\in\Pi_{\psi_2,f}} \dim \pi_f'^{K_2(n)},$$

where μ_1 is the automorphic character associated with ϕ_1^N. We may enlarge the sum from $\Pi_{\psi_2,f}$ to Π_{ψ_2}, which gives

$$h^3_{(2)}(Y(n))^\star \ll [K : (K\cap P(\mathbb{A}_f))K(n)] \sum_{\phi_1^N\in\Phi_{\text{rel}}(1)} \dim \mu_1^{\widetilde{K}_1(n)} \sum_{\phi_2^N\in\Phi_{\text{rel}}(2)} \sum_{\pi'\in\Pi_{\psi_2}} \dim \pi_f'^{K_2(n)}. \tag{9}$$

Lemma 4.6 implies that there are only three possibilities for $\mu_{1,\infty}$, and therefore

$$\sum_{\phi_1^N\in\Phi_{\text{rel}}(1)} \dim \mu_1^{\widetilde{K}_1(n)} \ll [K_1 : K_1(n)]. \tag{10}$$

There is a finite set Ξ_∞ of representations of $U(2)_\infty$ such that if $\phi_2^N \in \Phi_{\text{rel}}(2)$ and $\pi' \in \Pi_{\psi_2}$, then $\pi'_\infty \in \Xi_\infty$. Moreover, because ψ_2 is a simple generic parameter, we have $\Pi_\psi(\epsilon_\psi) = \Pi_\psi$ and so every $\pi' \in \Pi_{\psi_2}$ occurs in $L^2_{\mathrm{disc}}(U(2)(\mathbb{Q})\backslash U(2)(\mathbb{A}))$ with multiplicity one. We define $X(n) = U(2)(\mathbb{Q})\backslash U(2)(\mathbb{A})/K_2(n)$, and let $m(\pi_\infty, X(n))$ denote the multiplicity with which a representation π_∞ occurs in $L^2_{\mathrm{disc}}(X(n))$. We have

$$\begin{aligned}\sum_{\phi_2^N \in \Phi_{\mathrm{rel}}(2)} \sum_{\pi' \in \Pi_{\psi_2}} \dim \pi_f'^{K_2(n)} &\leq \sum_{\substack{\pi' \in L^2_{\mathrm{disc}}(U(2)(\mathbb{Q})\backslash U(2)(\mathbb{A})) \\ \pi'_\infty \in \Xi_\infty}} \dim \pi_f'^{K_2(n)} \\ &= \sum_{\pi_\infty \in \Xi_\infty} m(\pi_\infty, X(n)) \\ &\ll [K_2 : K_2(n)]. \end{aligned} \tag{11}$$

Combining (9)–(11) gives

$$h^3_{(2)}(Y(n))^\star \ll [K : (K \cap P(\mathbb{A}_f))K(n)][K_2 : K_2(n)][K_1 : K_1(n)].$$

Applying the formula for the order of $GL(N)$ over a finite field completes the proof.

4.3 The Case of ϕ_2^N Composite

We now briefly explain how to bound the contribution to $h^3_{(2)}(Y(n))^\star$ from parameters with $\phi_2^N = \phi_{21}^N \boxplus \phi_{22}^N$, where $\phi_{2i}^N \in \widetilde{\Phi}(1)$. We let ϕ_{2i}^N correspond to a conjugate self-dual character μ_{2i} on $GL(1, \mathbb{A}_E)$. Let P_2 be the standard Borel subgroup of $U(2)$. We may prove the following analogue of Proposition 4.2.

Proposition 4.7. *There is a choice of $\widetilde{K}_{1,w}$ for $w|p$, $p \in S_f$, depending only on K, such that*

$$\begin{aligned}\sum_{\pi_f \in \Pi_{\psi,f}} \dim \pi_f^{K(n)} \ll [K : (K \cap P(\mathbb{A}_f))K(n)][K_2 : (K_2 \cap P_2(\mathbb{A}_f))K_2(n)] \\ \dim \mu_1^{\widetilde{K}_1(n)} \dim \mu_{21}^{\widetilde{K}_1(n)} \dim \mu_{22}^{\widetilde{K}_1(n)}. \end{aligned} \tag{12}$$

The proof follows the same lines, by using the explicit description of π_{ψ_p} when p is split and the character identities of Propositions 2.2 and 2.3 when p is inert. There are $\ll n^3$ choices for the three characters, and the coset factors in (12) make a contribution of $\ll_\epsilon n^{5+\epsilon}$. Therefore the contribution to cohomology of parameters of this type is bounded by $\ll_\epsilon n^{8+\epsilon}$ as required.

5 The Case $\psi^N = \nu(2) \boxtimes \phi^N$

We now define $h^3_{(2)}(Y(n))^\star$ to be the contribution to $h^3_{(2)}(Y(n))$ from parameters of the form $\nu(2) \boxtimes \phi^N$. As in Sect. 4, we assume that $\phi^N \in \widetilde{\Phi}_{\rm sim}(2)$ until the end of Sect. 5.2, and describe how to treat composite ϕ^N in Sect. 5.3. We note that $\psi \in \Psi_2(U(4), \xi_+)$ implies that ϕ^N must be even. The main result of the section is the following.

Proposition 5.1. *We have the bound $h^3_{(2)}(Y(n))^\star \ll n^9$.*

We define compact open subgroups $K' = \prod_p K'_p \subset U(2)(\mathbb{A}_f)$, $\widetilde{K}' = \prod_w \widetilde{K}'_w \subset GL(2, \mathbb{A}_{E,f})$, and $\widetilde{K} = \prod_w \widetilde{K}_w \subset GL(4, \mathbb{A}_{E,f})$. We assume that $K'_p = U(2)(\mathbb{Z}_p)$ for all $p \notin S_f$, and likewise for the other groups. The local components of these groups for $w|p \in S_f$ will be specified in the proof of Proposition 5.2. We define congruence subgroups $K'(n)$, etc. of these groups for n relatively prime to S_f in the usual way, and recall that n will only be divisible by primes that split in E.

We let $\widetilde{P}$ be the standard parabolic subgroup of $GL(4, E)$ with Levi $\widetilde{L} = GL(2, E) \times GL(2, E)$, and let P be the corresponding standard parabolic subgroup of $U(4)$. We let P' be the standard Borel subgroup of $U(2)$.

5.1 Controlling a Single Parameter

We fix an even parameter $\phi^N \in \widetilde{\Phi}_{\rm sim}(2)$, and let $\psi \in \Phi(U(4), \xi_+)$ be the unique parameter with $\psi^N = \nu(2) \boxtimes \phi^N$. We let ϕ^N correspond to a conjugate self-dual cuspidal automorphic representation μ of $GL(2, \mathbb{A}_E)$. We assume that μ is tempered at all places; as before, this is done only for simplicity. We let $\psi' \in \Psi(U(2), \xi_-)$ be the unique parameter with $\psi'^N = \phi^N$. We shall prove the following bound for the finite part of the contribution of Π_ψ to $h^3_{(2)}(Y(n))^\star$.

Proposition 5.2. *There is a choice of K'_p for $p \in S_f$, depending only on K, such that*

$$\sum_{\pi_f \in \Pi_{\psi f}} \dim \pi_f^{K(n)} \ll [K' : (K' \cap P'(\mathbb{A}_f))K'(n)][K : (K \cap P(\mathbb{A}_f))K(n)] \sum_{\pi'_f \in \Pi_{\psi' f}} \dim \pi_f'^{K'(n)}. \tag{13}$$

We begin the proof of Proposition 5.2 with Lemma 5.4 and Corollary 5.5 below, which control the left-hand side of (13) in terms of μ.

Lemma 5.3. *Let $p \notin S_f$ be split in E, and let $w|p$. Let $\Pi_{\psi_p} = \{\pi_p\}$. We have*

$$\dim \pi_p^{K_p(n)} \leq [K_p : (K_p \cap P_p)K_p(n)](\dim \mu_w^{\widetilde{K}'_w(n)})^2. \tag{14}$$

Proof. Under the identification $U(4)_p \simeq GL(4, E_w)$, π_p is the Langlands quotient of the representation ρ_{ψ_w} of $GL_4(E_w)$ induced from the representation $\mu_w(x_1)|\det(x_1)|^{1/2} \otimes \mu_w(x_2)|\det(x_2)|^{-1/2}$ of $\widetilde{P}_w$. We have

$$\dim \pi_p^{K_p(n)} \leq \dim \rho_{\psi_w}^{\widetilde{K}_w(n)}.$$

The restriction of ρ_{ψ_w} to $\widetilde{K}_w$ is isomorphic to the induction of $\mu_w(x_1) \times \mu_w(x_2)$ from $\widetilde{K}_w \cap \widetilde{P}_w$ to $\widetilde{K}_w$. We see that

$$\begin{aligned}\dim \rho_{\psi_w}^{\widetilde{K}_w(n)} &= [\widetilde{K}_w : (\widetilde{K}_w \cap \widetilde{P}_w)\widetilde{K}_w(n)] \dim(\mu_w \times \mu_w)^{\widetilde{L}_w \cap \widetilde{K}_w(n)} \\ &= [\widetilde{K}_w : (\widetilde{K}_w \cap \widetilde{P}_w)\widetilde{K}_w(n)](\dim \mu_w^{\widetilde{K}'_w(n)})^2,\end{aligned}$$

which is equivalent to the lemma. □

We remove the square on the right-hand side of (14) using the following lemma.

Lemma 5.4. *If $p \notin S_f$ is split and $w|p$, we have*

$$\dim \mu_w^{\widetilde{K}'_w(n)} \leq [\widetilde{K}'_w : (\widetilde{K}'_w \cap \widetilde{P}'_w)\widetilde{K}'_w(n)] = [K'_p : (K'_p \cap P'_p)K'_p(n)].$$

Proof. If μ_w is a principal series representation or a twist of Steinberg, this is immediate. If μ_w is supercuspidal, this follows by examining the construction of supercuspidal representations given in §7.A of Gelbart [Ge]. □

Corollary 5.5. *Let $p \notin S_f$ be split in E, and let $w|p$. Let $\Pi_{\psi_p} = \{\pi_p\}$. We have*

$$\dim \pi_p^{K_p(n)} \leq [K_p : (K_p \cap P_p)K_p(n)][K'_p : (K'_p \cap P'_p)K'_p(n)] \dim \mu_w^{\widetilde{K}'_w(n)}.$$

Lemma 5.6. *Let $p \notin S_f$ be nonsplit in E, and let $w|p$. We have*

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} \leq \dim \mu_w^{\widetilde{K}'_w}.$$

Proof. Identify $\widetilde{K}_w$ with a subgroup of $G(4)_p$. The twisted fundamental lemma implies that the functions 1_{K_p} and $1_{\widetilde{K}_w \rtimes \theta}$ are related by transfer. Applying Proposition 2.3 with $s = e$ gives

$$1_{K_p}^{U(4)}(\psi_p) = \sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p},$$

and combining this with Proposition 2.2 and the twisted fundamental lemma gives

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} = \mathrm{tr}(\tilde{\pi}_{\psi_p}(1_{\widetilde{K}_w \rtimes \theta})).$$

The twisted trace $\mathrm{tr}(\tilde{\pi}_{\psi_p}(1_{\widetilde{K}_w \rtimes \theta}))$ is equal to the trace of $\tilde{\pi}_{\psi_p}(\theta)$ on $\pi_{\psi_p}^{\widetilde{K}_w}$, so we have

$$\mathrm{tr}(\tilde{\pi}_{\psi_p}(1_{\widetilde{K}_w \rtimes \theta})) \leq \dim \pi_{\psi_p}^{\widetilde{K}_w}.$$

Under the identification $G(4)_p \simeq GL(4, E_w)$, π_{ψ_p} is the Langlands quotient of the representation ρ_{ψ_w} induced from $\mu_w(x_1)|\det(x_1)|^{1/2} \otimes \mu_w(x_2)|\det(x_2)|^{-1/2}$. We therefore have

$$\dim \pi_{\psi_p}^{\widetilde{K}_w} \leq \dim \rho_{\psi_w}^{\widetilde{K}_w} \leq \dim \mu_w^{\widetilde{K}'_w},$$

and the result follows. □

Lemma 5.7. *Let $p \in S_f$, and let $w|p$. There is a choice of $\widetilde{K}'_w$, depending only on K, such that*

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} \ll \dim \mu_w^{\widetilde{K}'_w}.$$

Proof. Suppose that p is nonsplit. By Mok [Mo, Proposition 3.1.1(b)], we may choose a function $\widetilde{1}_{K_p} \in \widetilde{\mathcal{H}}_p(4)$ corresponding to 1_{K_p} under twisted transfer. Reasoning as in Lemma 5.6 gives

$$\sum_{\pi_p \in \Pi_{\psi_p}} \dim \pi_p^{K_p} = \mathrm{vol}(K_p)^{-1} \mathrm{tr}(\tilde{\pi}_{\psi_p}(\widetilde{1}_{K_p})),$$

where $\mathrm{vol}(K_p)$ denotes the volume of K_p with respect to our choice of Haar measure on $U(4)_p$. If we choose $\widetilde{K}_w \subset GL(4, E_w) \simeq G(4)_p$ to be a compact open subgroup such that $\widetilde{1}_{K_p}$ is bi-invariant under $\widetilde{K}_w$, we have

$$\mathrm{tr}(\tilde{\pi}_{\psi_p}(\widetilde{1}_{K_p})) \ll \dim \pi_{\psi_p}^{\widetilde{K}_w}.$$

Under the identification $G(4)_p \simeq GL(4, E_w)$, π_{ψ_p} is the Langlands quotient of the representation ρ_{ψ_w} induced from $\mu_w(x_1)|\det(x_1)|^{1/2} \otimes \mu_w(x_2)|\det(x_2)|^{-1/2}$. Choose $\widetilde{K}'_w$ so that the product $\widetilde{K}'_w \times \widetilde{K}'_w$ is contained in $\widetilde{K}_w$. We then have

$$\dim \pi_{\psi_p}^{\widetilde{K}_w} \leq \dim \rho_{\psi_w}^{\widetilde{K}_w} \ll (\dim \mu_w^{\widetilde{K}'_w})^2.$$

Bounding $\dim \mu_w^{\widetilde{K}'_w}$ by a constant depending on $\widetilde{K}'_w$, and hence K_p, completes the proof for p nonsplit. The proof in the split case follows in exactly the same way using the explicit description of π_p. □

Let $S_{E/\mathbb{Q}}$ be a set of finite places of E that contains exactly one place above every finite place of $\mathbb{Q}$. Combining Corollary 5.5, Lemma 5.6, and Lemma 5.7 gives

$$\sum_{\pi\in\Pi_{\psi,f}} \dim \pi_f^{K(n)} \ll [K' : (K'\cap P'(\mathbb{A}_f))K'(n)][K : (K\cap P(\mathbb{A}_f))K(n)] \prod_{w\in S_{E/\mathbb{Q}}} \dim \mu_w^{\widetilde{K}'_w(n)}.$$

Proposition 5.2 now follows from the lemma below.

Lemma 5.8. *There is a choice of K'_p for $p \in S_f$, depending only on K, such that*

$$\prod_{w\in S_{E/\mathbb{Q}}} \dim \mu_w^{\widetilde{K}'_w(n)} \ll \sum_{\pi'_f\in\Pi_{\psi',f}} \dim \pi_f'^{K'(n)}.$$

Proof. We may factorize the right-hand side as

$$\sum_{\pi'_f\in\Pi_{\psi',f}} \dim \pi_f'^{K'(n)} = \prod_p \sum_{\pi'_p\in\Pi_{\psi'_p}} \dim \pi_p'^{K'_p(n)}.$$

Let p be an arbitrary prime, and $w|p$. It suffices to show that

$$\dim \mu_w^{\widetilde{K}'_w(n)} \le \sum_{\pi'_p\in\Pi_{\psi'_p}} \dim \pi_p'^{K'_p(n)} \tag{15}$$

if $p \notin S_f$, and that if $p \in S_f$ the same inequality holds with a constant factor depending only on $\widetilde{K}'$, and hence K.

If p is split, then $\Pi_{\psi'_p}$ contains a single representation that is isomorphic to $\mu_w \otimes \chi_w^{-1}$ under the identification $U(2)_p \simeq GL(2, E_w)$, and (15) is immediate.

Suppose that $p \notin S_f$ is nonsplit. The definition of ψ'_p implies that if $\xi_- : {}^LU(2)_p \to {}^LG(2)_p$, the representation of $G(2)_p \simeq GL(2, E_w)$ associated with $\xi_- \circ \psi'_p \in \Psi_p(2)$ is μ_w. We let $\tilde{\mu}_w$ denote the canonical extension of μ_w to a representation of $\widetilde{G}^+(2)_p$, and identify $\widetilde{K}'_w$ with a subgroup of $G(2)_p$. Proposition 2.2 and the twisted fundamental lemma give

$$\mathrm{tr}(\tilde{\mu}_w(1_{\widetilde{K}'_w\rtimes\theta})) = \sum_{\pi'_p\in\Pi_{\psi'_p}} \mathrm{tr}(\pi'_p(1_{K'_p})) = \sum_{\pi'_p\in\Pi_{\psi'_p}} \dim \pi_p'^{K'_p}. \tag{16}$$

The left-hand side of (16) is equal to the trace of $\tilde{\mu}_w(\theta)$ on $\mu_w^{\widetilde{K}'_w}$. If $\dim \mu_w^{\widetilde{K}'_w} = 0$, then both sides of (16) are 0, and (15) holds. If $\dim \mu_w^{\widetilde{K}'_w} = 1$, then $\theta^2 = 1$ implies that $\mathrm{tr}(\tilde{\mu}_w(1_{\widetilde{K}'_w\rtimes\theta})) = \pm 1$. Positivity implies that we must take the plus sign so that (15) also holds.

Suppose that $p \in S_f$ is nonsplit, and suppose that the left-hand side of (15) is nonzero. Up to twist, there are only finitely many possibilities for μ_w that are supercuspidal or Steinberg, and we may deal with these cases by simply choosing K'_p so that (15) is true in each case. If μ_w is induced from a unitary character of the Borel, then $\Pi_{\psi'_p}$ is described explicitly in §11.4 of Rogawski [Ro] and (15) follows easily from this description. □

5.2 Summing Over Parameters

We define $\Phi_{\mathrm{rel}} \subset \widetilde{\Phi}_{\mathrm{sim}}(2)$ to be the set of even parameters ϕ^N such that ϕ^N_∞ is given by

$$\phi^N_\infty : z \mapsto \begin{pmatrix} z/\overline{z} & \\ & \overline{z}/z \end{pmatrix}.$$

It may be shown in the same way as Lemma 4.6 that if $\psi \in \Psi(U(4), \xi_+)$ satisfies $\psi^N = \nu(2) \boxtimes \phi^N$ with $\phi^N \in \widetilde{\Phi}_{\mathrm{sim}}(2)$, and $\pi \in \Pi_{\psi_\infty}$ satisfies $H^*(\mathfrak{g}, K; \pi) \neq 0$, then $\phi^N \in \Phi_{\mathrm{rel}}$. If $\phi^N \in \Phi_{\mathrm{rel}}$ corresponds to the cusp form μ, and χ_∞ is given by $\chi_\infty(z) = (z/\overline{z})^{1/2+t}$ with $t \in \mathbb{Z}$, then $\mu_\infty \times \chi_\infty$ has infinitesimal character $(3/2+t, -1/2+t; -3/2-t, 1/2-t) \in (\mathbb{C}^2/S_2) \times (\mathbb{C}^2/S_2)$. Theorem 1.2 of Caraiani [Ca] then implies that μ is tempered at all places. It follows from this discussion that

$$h^3_{(2)}(Y(n))^\star \ll \sum_{\substack{\psi^N = \nu(2)\boxtimes\phi^N \\ \phi^N \in \Phi_{\mathrm{rel}}}} \sum_{\pi \in \Pi_\psi} \dim \pi_f^{K(n)}. \tag{17}$$

Applying Proposition 5.2 to the sum on the right-hand side (and ignoring the factors $\#(\Pi_{\psi_\infty})$ as in Sect. 4.2) gives

$$\begin{aligned} h^3_{(2)}(Y(n))^\star \ll [K' : (K' \cap P'(\mathbb{A}_f))K'(n)][K : (K \cap P(\mathbb{A}_f))K(n)] \\ \times \sum_{\substack{\psi' \in \Psi(U(2), \xi_-) \\ \psi'^N \in \Phi_{\mathrm{rel}}}} \sum_{\pi' \in \Pi_{\psi'}} \dim \pi_f'^{K'(n)}. \end{aligned} \tag{18}$$

The restriction on the infinitesimal characters of parameters in Φ_{rel} implies that there is a finite set of representations Ξ_∞ of $U(2)_\infty$ such that if $\psi'^N \in \Phi_{\mathrm{rel}}$, then all the representations in $\Pi_{\psi'_\infty}$ are in Ξ_∞. Because Φ_{rel} consists of simple generic parameters we have $\Pi_{\psi'} = \Pi_{\psi'}(\epsilon_{\psi'})$, and so every $\pi' \in \Pi_{\psi'}$ occurs in $L^2_{\mathrm{disc}}(U(2)(\mathbb{Q})\backslash U(2)(\mathbb{A}))$ with multiplicity one. If we define $X(n) =$

$U(2)(\mathbb{Q})\backslash U(2)(\mathbb{A})/K'(n)$, and let $m(\pi_\infty, X(n))$ denote the multiplicity as in Sect. 4.2, this gives

$$\begin{aligned}\sum_{\substack{\psi' \in \Psi(U(2),\xi_-) \\ \psi'^N \in \Phi_{\mathrm{rel}}}} \sum_{\pi' \in \Pi_{\psi'}} \dim \pi_f'^{K'(n)} &\leq \sum_{\substack{\pi' \in L^2_{\mathrm{disc}}(U(2)(\mathbb{Q})\backslash U(2)(\mathbb{A})) \\ \pi'_\infty \in \Xi_\infty}} \dim \pi_f'^{K'(n)} \\ &= \sum_{\pi_\infty \in \Xi_\infty} m(\pi_\infty, X(n)) \\ &\ll [K' : K'(n)]. \end{aligned} \tag{19}$$

Combining (17)–(19) gives

$$h^3_{(2)}(Y(n))^\star \ll [K' : (K' \cap P'(\mathbb{A}_f))K'(n)][K : (K \cap P(\mathbb{A}_f))K(n)][K' : K'(n)],$$

and applying the formula for the order of $GL(N)$ over a finite field completes the proof.

5.3 The Case of Composite ϕ^N

We now suppose that $\phi^N = \phi_1^N \boxplus \phi_2^N$, where $\phi_i^N \in \widetilde{\Phi}(1)$ correspond to conjugate self-dual characters μ_i. We may prove the following analogue of Proposition 5.2.

Proposition 5.9. *There is a choice of $\widetilde{K}_{1,w}$ for $w|p \in S_f$, depending only on K, such that*

$$\sum_{\pi_f \in \Pi_{\psi,f}} \dim \pi_f^{K(n)} \ll [K : (K \cap P(\mathbb{A}_f))K(n)] \dim \mu_1^{\widetilde{K}_1(n)} \dim \mu_2^{\widetilde{K}_1(n)}.$$

Unlike Proposition 5.2, this bound is sharp. The reason for this is that the representation π_{ψ_p} for split p is equivalent to the induction of $(\mu_{1,w} \circ \det(x_1))|\det(x_1)|^{1/2} \otimes (\mu_{2,w} \circ \det(x_2))|\det(x_2)|^{-1/2}$ from $\widetilde{P}_w$ to $GL(4, E_w)$, and it is easy to give a sharp bound for the dimension of invariants under $\widetilde{K}_w(n)$, unlike the Speh representations considered in Lemma 5.3. We obtain a bound of $n^{6+\epsilon}$ for the contribution of these parameters to $h^3_{(2)}(Y(n))^\star$.

Acknowledgements This material is based upon work supported by the National Science Foundation under Grant No. DMS-1201321 and DMS-1509331.

References

[AMR] N. Arancibia, C. Moeglin, D. Renard, Paquets d'Arthur des groupes classiques et unitaires. Preprint, arXiv:1507.01432

[Art] J. Arthur, The endoscopic classification of representations: orthogonal and symplectic groups. AMS Colloq. Publ. **61** (2013)

[BC] A. Borel, W. Casselman, L^2-cohomology of locally symmetric manifolds of finite volume. Duke Math. J. **50**, 625–647 (1983)

[BG] A. Borel, H. Garland, Laplacian and the discrete spectrum of an arithmetic group. Am. J. Math. **105**(2), 309–335 (1983)

[BMM] N. Bergeron, J. Milson, C. Moeglin, The Hodge conjecture and arithmetic quotients of complex balls. Acta Math. **216**(1), 1–125 (2016)

[BW] A. Borel, N. Wallach, Continuous cohomology, discrete subgroups, and representations of reductive groups. AMS Math. Surv. Monogr. **67** (1999)

[Ca] A. Caraiani, Local-global compatibility and the action of monodromy on nearby cycles. Duke Math. J. **161**(12), 2311–2413 (2012)

[Ge] S. Gelbart, *Automorphic Forms on Adele Groups*. Annals of Mathematics Studies, vol. 83 (Princeton University Press, Princeton, NJ, 1975)

[Ha] T. Hales, On the fundamental lemma for standard endoscopy: reduction to unit elements. Can. J. Math. **47**(5), 974–994 (1995)

[Hen] G. Henniart, Une preuve simple des conjectures de Langlands pour $GL(n)$ sur un corps p-adique (French, with English summary). Invent. Math. **139**(2), 439–455 (2000). doi:10.1007/s002220050012. MR1738446 (2001e:11052)

[HT] M. Harris, R. Taylor, *The Geometry and Cohomology of Some Simple Shimura Varieties*. Annals of Mathematics Studies, vol. 151 (Princeton University Press, Princeton, NJ, 2001). With an appendix by Vladimir G. Berkovich. MR1876802(2002m:11050)

[KMSW] T. Kaletha, A. Minguez, S.W. Shin, P.-J. White, Endoscopic classification of representations: inner forms of unitary groups. Preprint, arXiv:1409.3731

[KS] R. Kottwitz, D. Shelstad, Foundations of twisted endoscopy (English, with English and French summaries). Astérisque **255**, vi+190 (1999). MR1687096 (2000k:22024)

[LN] G. Laumon, B.C. Ngô, Le lemme fondamental pour les groupes unitaires (French, with English summary). Ann. Math. (2) **168**(2), 477–573 (2008). doi:10.4007/annals.2008.168.477. MR2434884 (2009i:22022)

[Ma] S. Marshall, Endoscopy and cohomology growth on $U(3)$. Compos. Math. **150**, 903–910 (2014)

[Mo] C.-P. Mok, Endoscopic classification of representation of quasi-split unitary groups. Mem. Am. Math. Soc. **235**(1108), (2015)

[Ngo] B.C. Ngô, Le lemme fondamental pourles algébres de Lie (French). Publ. Math. Inst. Hautes Études Sci. **111**, 1–169 (2010). doi:10.1007/s10240-010-0026-7. MR2653248 (2011h:22011)

[Ro] J. Rogawski, *Automorphic Representations of Unitary Groups in Three Variables*. Annals of Mathematics Studies, vol. 123 (Princeton University Press, Princeton, NJ, 1990)

[Sa] G. Savin, Limit multiplicities of cusp forms. Invent. Math. **95**, 149–159 (1989)

[SX] P. Sarnak, X. Xue, Bounds for multiplicities of automorphic representations. Duke Math. J. **64**, 207–227 (1991)

[Vo] D. Vogan, The local Langlands conjecture, in *Representation Theory of Groups and Algebras*, ed. by J. Adams. Contemporary Mathematics, vol. 145 (American Mathematical Society, Providence, RI, 1993), pp. 305–379

[Wa] J.-L. Waldspurger, L'endoscopie tordue n'est pas si tordue. Mem. AMS **908** (2008)

[Wa2] J.-L. Waldspurger, Endoscopie et changement de caractéristique. J. Inst. Math. Jussieu **5**(3), 423–525 (2006)

[Wa3] J.-L. Waldspurger, Le lemme fondamental implique le transfert. Compos. Math. **105**(2), 153–236 (1997)

Distribution of Hecke Eigenvalues for GL(n)

Jasmin Matz

Abstract The purpose of this survey is to briefly summarize and explain the results of Matz (Weyl's law for Hecke operators on GL(n) over imaginary quadratic number fields, 2013, arXiv:1310.6525) and joint work with Templier (Sato-Tate equidistribution for the family of Hecke-Maass forms on SL(n, $\mathbf{Z}$), arXiv:1505.07285) about the asymptotic distribution of eigenvalues of Hecke operators on cusp forms for GL(n). We also to sketch some motivation and potential extensions of our results.

1 Introduction

Let F be a number field with ring of adeles $\mathbb{A}_F$, and let $n \geq 2$ be an integer. Let $G = \mathrm{GL}(n)$, and let $G(\mathbb{A}_F)^1 := \{g \in G(\mathbb{A}_F) \mid |\det g|_{\mathbb{A}_F} = 1\}$ where $|\cdot|_{\mathbb{A}_F}$ denotes the adelic absolute value on $\mathbb{A}_F^\times$. One is interested in the spectral decomposition of the space $L^2(G(F)\backslash G(\mathbb{A}_F)^1)$ under the right regular representation of $G(\mathbb{A}_F)$. Under $G(\mathbb{A}_F)$ the space $L^2(G(F)\backslash G(\mathbb{A}_F)^1)$ decomposes into invariant subspaces as

$$L^2(G(F)\backslash G(\mathbb{A}_F)^1) = L^2_{\mathrm{cusp}}(G(F)\backslash G(\mathbb{A}_F)^1) \oplus L^2_{\mathrm{res}}(G(F)\backslash G(\mathbb{A}_F)^1) \\ \oplus L^2_{\mathrm{cts}}(G(F)\backslash G(\mathbb{A}_F)^1),$$

where L^2_{cusp} (resp. L^2_{res}, resp. L^2_{cts}) denotes the cuspidal (resp. residual, resp. continuous) part of L^2 under the right regular representation of $G(\mathbb{A}_F)^1$. The cuspidal part is the most fundamental one in the sense that the residual and continuous parts can be described in terms of Eisenstein series and their residues attached to cuspidal representations on Levi subgroups of G [Lan76, MW95]. It is therefore of importance to understand the spectral properties of the space of cusp forms. One of the most basic questions is to asymptotically count the number of Laplace eigenfunctions of bounded eigenvalue for the locally symmetric spaces $G(F)\backslash G(\mathbb{A}_F)^1/K$ where $K = \mathbf{K}_\infty \cdot K_f \subseteq \mathbf{K}$ is a finite index subgroup of a fixed maximal compact subgroup $\mathbf{K} \subseteq G(\mathbb{A}_F)$. The Weyl law answers this question in many cases.

J. Matz (✉)
Mathematisches Institut, Universität Leipzig, Postfach 100920, 04009 Leipzig, Germany
e-mail: matz@math.uni-leipzig.de

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_9

Suppose $F = \mathbb{Q}$ for the rest of the introduction. Then $\mathrm{O}(n)$ is a maximal compact subgroup of $G(F_\infty) = G(\mathbb{R})$, and $G(\mathbb{A}_\mathbb{Q})^1/\mathrm{O}(n) \simeq \mathrm{SL}_n(\mathbb{R})/\,\mathrm{SO}(n) =: X$. Let Δ be the Laplacian on $L^2(X)$, and let $\Gamma \subseteq \mathrm{SL}_n(\mathbb{R})$ be an arithmetic congruence subgroup. Weyl's law in its most basic form counts the number of eigenvalues of Δ in $L^2_{\mathrm{cusp}}(\Gamma\backslash X)$. More precisely, let $0 \le \mu_1 \le \mu_2 \le \ldots$ be the cuspidal eigenvalues of Δ (with multiplicities). Then

$$\#\{i \mid \mu_i^2 \le Y\} \sim c \operatorname{vol}(\Gamma\backslash X) Y^d \tag{1}$$

as $Y \to \infty$ for $c > 0$ a constant depending only on n, and $d = \dim_\mathbb{R} X$. This was proven by Selberg for $n = 2$ [Sel56], by Miller for $n = 3$ [Mil01], and by Müller for general n [Mül07]. The Weyl law also holds for more general groups, cf. [DKV79, LV07]. This in particular proves the existence of infinitely many cusp forms in $L^2(\Gamma\backslash X)$ but stills gives only crude information on the spectral properties of $\Gamma\backslash X$. Apart from Δ there are many more naturally occurring operators on $L^2_{\mathrm{cusp}}(\Gamma\backslash X)$, and one can study the distribution of their (joint) eigenvalues as well.

Let $\mathcal{D}(X)$ be the algebra of $\mathrm{SL}_n(\mathbb{R})$-invariant differential operators on X. It is isomorphic to the Weyl group invariants $Z(\mathfrak{sl}_n(\mathbb{C}))^W \simeq \mathfrak{a}_\mathbb{C}^W$ of the center $Z(\mathfrak{sl}_n(\mathbb{C}))$ of the universal enveloping algebra of the Lie algebra $\mathfrak{sl}_n(\mathbb{C})$ of $\mathrm{SL}_n(\mathbb{C})$. Here $\mathfrak{a}$ is the Lie algebra of the maximal diagonal torus in $\mathrm{SL}_n(\mathbb{R})$ which can be identified with all vectors $(X_1, \ldots, X_n) \in \mathbb{R}^n$ such that $\sum_i X_i = 0$, and $\mathfrak{a}_\mathbb{C}$ is its complexification. If $\pi \subseteq L^2_{\mathrm{cusp}}(X)$ is an irreducible component, elements of $\mathcal{D}(X)$ act by a scalar on π by Schur's Lemma so that π defines a character $\lambda_\pi : \mathfrak{a}_\mathbb{C}^W \longrightarrow \mathbb{C}$ (the infinitesimal character), that is, λ_π is a W-invariant element in the dual space $\mathfrak{a}_\mathbb{C}^*$ of $\mathfrak{a}_\mathbb{C}$. In generalization of (1) one can ask how the λ_π distribute if one takes larger and larger subsets of $\mathfrak{a}_\mathbb{C}^*$. This question was answered by Lapid and Müller [LM09], see also below.

Apart from the algebra of differential operators, there is a second family of operators acting on $L^2_{\mathrm{cusp}}(\Gamma\backslash X)$, namely, the algebra of Hecke operators. The Hecke algebra is commutative and preserves the eigenspaces of $\mathcal{D}(X)$. Suppose $\{T_n\}_{n\in\mathbb{N}}$ is a family of Hecke operators, and let $\psi_1, \psi_2, \ldots$ be a joint eigenbasis for $L^2_{\mathrm{cusp}}(\Gamma\backslash X)$ for $\mathcal{D}(X)$ and $\{T_n\}_{n\in\mathbb{N}}$. For every i let $\lambda_i \in \mathfrak{a}_\mathbb{C}^*/W$ be the infinitesimal character of the irreducible representation generated by ψ_i, and let $a_i(n) \in [-\|T_n\|, \|T_n\|]$ be the eigenvalue of ψ_i under T_n. Here $\|T_n\|$ denotes the operator norm of T_n. Then

$$\Lambda_i := (\lambda_i, a_i(1), a_i(2), \ldots)$$

defines a point in the space

$$\mathcal{A} := \mathfrak{a}_\mathbb{C}^*/W \times \prod_{n\in\mathbb{N}} [-\|T_n\|, \|T_n\|],$$

and one can ask how these Λ_i distribute in $\mathcal{A}$ (with respect to the chosen ordering of the basis). This question was studied in [Sar87] for $n = 2$, and in [ST15] for groups G for which $G(\mathbb{R})$ has discrete series, cf. also [SST16].

2 Results

2.1 Notation

Recall that $n \geq 2$ and $G = \mathrm{GL}(n)$ over a fixed number field F. $\mathbb{A}_F$ denotes the ring of adeles of F, and $\mathbb{A}_{F,f}$ the finite part of $\mathbb{A}_{F,f}$. Let $\mathcal{O}_F$ be the ring of integers of F. If v is a non-archimedean place of F, we write $\mathcal{O}_{F_v}$ for the ring of integers in F_v, $\mathfrak{q}_v \subseteq \mathcal{O}_{F_v}$ for the maximal ideal in $\mathcal{O}_{F_v}$, $\varpi_v \in \mathfrak{q}_v$ for a fixed uniformizing element of F_v, and q_v for the cardinality of the residue field at v. Let $T_0 \subseteq G$ be the maximal torus consisting of diagonal matrices, and let $P_0 = T_0 U_0$ be the usual minimal parabolic subgroup of upper triangular matrices with U_0 the unipotent radical of P_0. We write $Z \subseteq G$ for the center of G. We also identify $Z(F)$ with a subgroup of the finite part $Z(\mathbb{A}_{F,f}) \subseteq G(\mathbb{A}_{F,f})$.

We fix the usual maximal compact subgroup $\mathbf{K} \subseteq G(\mathbb{A}_F)$, $\mathbf{K} = \prod_v \mathbf{K}_v$, with

$$\mathbf{K}_v = \begin{cases} \mathrm{O}(n) & \text{if } v \text{ is a real place,} \\ \mathrm{U}(n) & \text{if } v \text{ is a complex place,} \\ G(\mathcal{O}_{F_v}) & \text{if } v \text{ is non-archimedean.} \end{cases}$$

For a non-archimedean place v and an integer $m \geq 0$ let

$$\mathbf{K}_v(\mathfrak{q}_v^m) = \ker\left(\mathbf{K}_v \longrightarrow G(\mathcal{O}_{F_v}/\mathfrak{q}_v^m)\right)$$

be the principal congruence subgroup of level $\mathfrak{q}_v^m$. If $\mathfrak{a} \subseteq \mathcal{O}_F$ is an ideal with prime factorization $\mathfrak{a} = \prod_{v<\infty} \mathfrak{q}_v^{m_v}$, we put

$$\mathbf{K}_f(\mathfrak{a}) = \prod_{v<\infty} \mathbf{K}_v(\mathfrak{q}_v^{m_v}),$$

and $\mathbf{K}(\mathfrak{a}) = \mathbf{K}_\infty \cdot \mathbf{K}_f(\mathfrak{a})$ with $\mathbf{K}_\infty = \prod_{v|\infty} \mathbf{K}_v \subseteq G(F_\infty) = \prod_{v|\infty} G(F_v)$. If $F = \mathbb{Q}$ and $N \in \mathbb{Z}_{\geq 1}$, we also write $\mathbf{K}_f(N) = \mathbf{K}_f(N\mathbb{Z})$ and $\mathbf{K}(N) = \mathbf{K}(N\mathbb{Z})$.

Let $\Pi_{\mathrm{cusp}}(G(\mathbb{A}_F)^1)$ denote the set of irreducible cuspidal automorphic representations of $G(\mathbb{A}_F)^1$, and for $\pi = \pi_\infty \cdot \pi_f \in \Pi_{\mathrm{cusp}}(G(\mathbb{A}_F)^1)$ let $\lambda_{\pi_\infty} \in \mathfrak{a}_{\mathbb{C}}^*/W$ denote the infinitesimal character of π_∞. If convenient, we identify π with its representation space so that we may write $\dim \pi^K$ for the dimension of the subspace of K-fixed vectors in the representation space of π.

For simplicity of the statements below we choose the Haar measure on $G(F_v)$ such that it gives $\mathbf{K}_v$ volume 1 for every v. We then take the product measure on $G(\mathbb{A}_F)$, and fix the measure on $G(\mathbb{A}_F)^1$ via the exact sequence

$$1 \longrightarrow G(\mathbb{A}_F)^1 \hookrightarrow G(\mathbb{A}_F) \longrightarrow \mathbb{R}_{>0} \longrightarrow 1$$

where we take the usual multiplicative Lebesgue measure on $\mathbb{R}_{>0}$. The maximal compact subgroup $\mathbf{K}$ then has volume 1 with respect to the measure on $G(\mathbb{A}_F)^1$. If $\Xi \subseteq G(\mathbb{A}_F)$ (resp. $\Xi \subseteq G(\mathbb{A}_F)^1$, resp. $\Xi \subseteq G(\mathbb{A}_{F,f})$) is a measurable subset, we write $\mathrm{vol}(\Xi)$ for the volume on Ξ with respect to the measure on $G(\mathbb{A}_F)$ (resp. $G(\mathbb{A}_F)^1$, resp. $G(\mathbb{A}_{F,f})$). For different choices of measures one might need to adjust some of the constants below accordingly.

2.2 Weyl Law with Remainder Term for $\mathrm{SL}_n(\mathbb{R})$

Let $F = \mathbb{Q}$, and let $K = \mathbf{K}_\infty \cdot K_f$ with $K_f \subseteq \mathbf{K}_f$ a finite index subgroup such $K_f \subseteq \mathbf{K}_f(N)$ for some $N \geq 3$. This last requirement ensures that K_f does not have any non-trivial element of finite order.

Lapid and Müller [LM09] proved a refined version of the Weyl law for $G(F)\backslash G(\mathbb{A}_F)^1/K$: If $\Omega \subseteq i\mathfrak{a}^*$ is a W-invariant bounded domain with piecewise C^2-boundary, then

$$\sum_{\substack{\pi\in\Pi_{\mathrm{cusp}}(G(\mathbb{A}_\mathbb{Q})^1):\\ \lambda_{\pi_\infty}\in t\Omega}} \dim \pi^K = \frac{\mathrm{vol}(G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q})^1/K_f)}{|W|}\int_{t\Omega} \mathbf{c}(\lambda)^{-2}\,d\lambda + O(t^{d-1}(\log t)^{\max\{3,n\}}), \tag{2}$$

as $t \to \infty$, where $\mathbf{c}(\lambda)$ denotes the Harish-Chandra $\mathbf{c}$-function for $\mathrm{SL}_n(\mathbb{R})$ so that $\mathbf{c}(\lambda)^{-2}d\lambda$ is the spherical Plancherel measure for $\mathrm{SL}_n(\mathbb{R})$. In more classical terms this gives the asymptotic distribution (with weight factor $\dim \pi^{K(N)}$) of the infinitesimal characters of cusp forms on $\Gamma(N)\backslash X$, $N \geq 3$. This is because the quotient $G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q})^1/K(N)$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z})^\times$-copies of $\Gamma(N)\backslash X$ for $\Gamma(N) = \{\gamma \in \mathrm{SL}_n(\mathbb{Z}) \mid \gamma \equiv 1 \mod N\}$ the principal congruence subgroup of level N. Taking Ω to be the unit ball in $i\mathfrak{a}^*$, one recovers the usual Weyl law (1) together with an upper bound for the error term.

Let $B_t(0)$ denote the ball of radius t in $\mathfrak{a}^*_\mathbb{C}$. According to [LM09] one also has

$$\sum_{\substack{\pi\in\Pi_{\mathrm{disc}}(G(\mathbb{A}_\mathbb{Q})^1):\\ \lambda_{\pi_\infty}\in B_t(0)\backslash i\mathfrak{a}^*}} \dim \pi^K = O(t^{d-2}) \tag{3}$$

i.e., the number of non-tempered $\pi \in \Pi_{\mathrm{cusp}}(G(\mathbb{A}_\mathbb{Q})^1)$ (which are supposed to be non-existent according to the generalized archimedean Ramanujan Conjecture) is at most of lower order than the number of tempered representations.

2.3 Traces of Hecke Operators

We now turn to the main results of [Mat, MT]. Let F be an imaginary quadratic number field (that is, a quadratic field extension of $\mathbb{Q}$ with one complex place) or $F = \mathbb{Q}$. The first case is covered in [Mat] while the second case is the subject of [MT].

2.3.1 Hecke Algebra

For every non-archimedean place v of F consider the spherical Hecke algebra $\mathcal{H}_v = C_c^\infty(G(F_v) /\!/ \mathbf{K}_v)$ of locally constant, compactly supported bi-$\mathbf{K}_v$-invariant functions. This is a commutative $\mathbb{C}$-algebra under convolution for which the characteristic function of $\mathbf{K}_v$ is the unit element. For $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$ let $\tau_{v,\lambda} \in \mathcal{H}_v$ denote the characteristic function of the double coset $\mathbf{K}_v \varpi_v^\lambda \mathbf{K}_v$, where

$$\varpi_v^\lambda := \begin{pmatrix} \varpi_v^{\lambda_1} & & \\ & \ddots & \\ & & \varpi_v^{\lambda_n} \end{pmatrix}.$$

The set of functions $\{\tau_\lambda \mid \lambda \in \mathbb{Z}^n,\ \lambda_1 \geq \ldots \geq \lambda_n\}$ generates $\mathcal{H}_v$ as a $\mathbb{C}$-algebra. We write $\mathbb{Z}^{n,+}$ for the set of tuples $(\lambda_1, \dots, \lambda_n) \in \mathbb{Z}^n$ with $\lambda_1 \geq \ldots \geq \lambda_n$. If $\lambda \in \mathbb{Z}^n$, we write $\|\lambda\| = (\sum \lambda_i^2)^{1/2}$ for the usual Euclidean norm of λ. If $\kappa \geq 0$, we let $\mathcal{H}_v^{\leq \kappa}$ be the sub-vector-space of $\mathcal{H}_v$ generated (as a vector space over $\mathbb{C}$) by the functions $\tau_{v,\lambda}$ with $\|\lambda\| \leq \kappa$. If S is a finite set of non-archimedean places, we put $\mathcal{H}_S = \prod_{v \in S} \mathcal{H}_v$, and $\mathcal{H}_S^{\leq \boldsymbol{\kappa}} = \prod_{v \in S} \mathcal{H}_v^{\leq \kappa_v}$ if $\boldsymbol{\kappa} = (\kappa_v)_{v \in S}$ is a sequence of non-negative numbers. If $\tau_S \in \mathcal{H}_S$, we also identify τ_S with a function $\tau \in C_c^\infty(G(\mathbb{A}_{F,f}) /\!/ \mathbf{K}_f)$ by putting $\tau = \tau_S \cdot \mathbf{1}_{\mathbf{K}^{S \cup S_\infty}}$ where S_∞ is the set of archimedean places of F, and $\mathbf{1}_{\mathbf{K}^{S \cup S_\infty}} : G(\mathbb{A}_F^{S \cup S_\infty}) \longrightarrow \mathbb{C}$ the characteristic function of $\mathbf{K}^{S \cup S_\infty} = \prod_{v \notin S \cup S_\infty} \mathbf{K}_v$.

If $\boldsymbol{\kappa} = (\kappa_v)_{v \in S}$ is a sequence of non-negative numbers, we set

$$\Pi_\kappa = \prod_{v \in S} q_v^{\kappa_v}.$$

This number provides an upper bound for the "degrees" (that is, L^1-norms) of the Hecke operators in $\mathcal{H}_S^{\leq \boldsymbol{\kappa}}$: There exists $a > 0$ such that for every $\tau_S \in \mathcal{H}_S^{\boldsymbol{\kappa}}$ with $|\tau_S| \leq 1$ we have $\|\tau\|_{L^1(G(\mathbb{A}_{F,f}))} = \|\tau_S\|_{L^1(G(F_S))} \leq \Pi_{\boldsymbol{\kappa}}^a$.

2.3.2 Distribution of Traces of Hecke Operators

Let $\mathcal{F} = \{\pi \in \Pi_{\mathrm{cusp}}(G(\mathbb{A}_F)^1) \mid \pi^{\mathbf{K}} \neq 0\}$ be the spectral set of all everywhere unramified cuspidal representations with trivial $\mathbf{K}_\infty$-type, cf. [SST16]. Let $\Omega \subseteq i\mathfrak{a}^*$ be as before. We use the infinitesimal character and the domain Ω to put an order on the set $\mathcal{F}$: For $t > 0$ let

$$\mathcal{F}(t) = \mathcal{F}_\Omega(t) = \{\pi \in \mathcal{F} \mid \lambda_{\pi_\infty} \in t\Omega\}.$$

According to the generalized archimedean Ramanujan conjecture, every element of $\mathcal{F}$ should eventually appear in $\mathcal{F}(t)$ for t sufficiently large if Ω is "thick enough," that is, if Ω is such that $\bigcup_{t>0} t\Omega = i\mathfrak{a}^*$. In any case, the estimate (3) from [LM09] shows (for $F = \mathbb{Q}$; but one can show that a similar statement is true for F imaginary quadratic) that one does not miss "too many" elements.

Theorem 2.1. *(i) As $t \to \infty$ we have*

$$|\mathcal{F}(t)| \sim |\mathcal{O}_F^\times| \frac{\operatorname{vol}(G(F)\backslash G(\mathbb{A}_F)^1/\mathbf{K}_f)}{|W|} \int_{t\Omega} \mathbf{c}(\lambda)^{-2}\, d\lambda$$

in the sense that the difference of the left and right-hand side tends to 0 as $t \to \infty$. Here $|\mathcal{O}_F^\times|$ is the number of multiplicative units in $\mathcal{O}_F$.

(ii) There exist constants $a, b, \delta > 0$ (depending only on n, Ω, and F) such that the following holds: For every finite set of non-archimedean places S_0, every sequence of non-negative numbers $\boldsymbol{\kappa} = (\kappa_v)_{v\in S_0}$ and every $\tau_{S_0} \in \mathcal{H}_{S_0}^{\le\boldsymbol{\kappa}}$ with $|\tau| \le 1$ we have

$$\lim_{t\to\infty} |\mathcal{F}(t)|^{-1} \sum_{\pi\in\mathcal{F}(t)} \operatorname{tr} \pi_f(\tau) = \lim_{t\to\infty} |\mathcal{F}(t)|^{-1} \sum_{\pi\in\mathcal{F}(t)} \operatorname{tr} \pi_{S_0}(\tau_{S_0}) = \sum_{z\in Z(F)/Z(\mathcal{O}_F)} \tau(z), \tag{4}$$

and

$$\left| |\mathcal{F}(t)|^{-1} \sum_{\pi\in\mathcal{F}(t)} \operatorname{tr} \pi_f(\tau) - \sum_{z\in Z(F)/Z(\mathcal{O}_F)} \tau(z) \right| \le a \Pi_{\boldsymbol{\kappa}}^b t^{-\delta} \tag{5}$$

for every $t \ge 1$.

Remark 2.2. (i) The number $|\mathcal{O}_F^\times|$ is finite by our assumption that F has only one archimedean place.

(ii) Taking $S_0 = \emptyset$ so that τ is the characteristic function of $\mathbf{K}_f$, the second part of Theorem 2.1 also gives an upper bound for the remainder term of the asymptotic of the first part. Hence for $F = \mathbb{Q}$ we obtain the analogue of [LM09] but for the full modular group $\Gamma = \mathrm{SL}_n(\mathbb{Z})$ (which was excluded in [LM09] for technical reasons)—however with a slightly worse error term.

(iii) Taking $\tau_{S_0} = \prod_{v\in S_0} \tau_{v,\lambda_v}$ in the above theorem, we see that the main term, that is the right-hand side of (4), vanishes for many sequences of λ_v. More precisely, the main term vanishes unless

$$\lambda_{v,1} = \ldots = \lambda_{v,n} \tag{6}$$

for every $v < \infty$. In this situation, τ corresponds to an ideal $\mathfrak{a} \subseteq \mathcal{O}_F$ defined by $\mathfrak{a} = \prod_{v\in S_0} \mathfrak{q}_v^{\lambda_{v,1}}$ and $\tau(z) \neq 0$ if and only if z (identified with an element in $F^\times$) generates $\mathfrak{a}$ so that $\mathfrak{a}$ needs to be principal. Hence if for every v (6) is satisfied and if the sequence of $\lambda_{v,1}$, $v \in S_0$, corresponds to a principal ideal, we get

$$\lim_{t\to\infty} |\mathcal{F}(t)|^{-1} \sum_{\pi\in\mathcal{F}(t)} \operatorname{tr}\pi_f(\tau) = 1,$$

and the left-hand side vanishes in all other cases. In general, any $\tau_{S_0} \in \mathcal{H}_{S_0}^{\kappa}$ is a linear combination of characteristic functions of double cosets so that this consideration can be applied to an arbitrary τ_{S_0}.

If $F = \mathbb{Q}$, we can reformulate the above result in more measure theoretic terms, namely in terms of measures on the unitary dual of $\mathrm{PGL}_n(\mathbb{Q}_S)$. (For $F \neq \mathbb{Q}$ one can make a similar reformulation but one has to be more careful with central characters.) Let $H = \mathrm{PGL}_n$. Let $\mathcal{F}^H = \{\pi_0 \in \Pi_{\mathrm{cusp}}(H(\mathbb{A}_\mathbb{Q})) \mid \pi^{\mathbf{K}^H} \neq 0\}$ for $\mathbf{K}^H = \mathbf{K} \cap H(\mathbb{A}_\mathbb{Q})$ the usual maximal compact subgroup of $H(\mathbb{A}_\mathbb{Q})$, and put $\mathcal{F}^H(t) = \{\pi_0 \in \mathcal{F}^H \mid \lambda_{\pi_\infty} \in t\Omega\}$. The sets $\mathcal{F}$ and $\mathcal{F}^H$ as well as $\mathcal{F}(t)$ and $\mathcal{F}^H(t)$ can be canonically identified with each other since every $\pi \in \mathcal{F}$ has trivial central character so that it can be identified with an element of $\mathcal{F}^H$. Hence if $\tau = \tau_{S_0} \otimes \mathbf{1}_{\mathbf{K}^{S_0}} \in C_c^\infty(G(\mathbb{A}_{\mathbb{Q},f}))$ is bi-$\mathbf{K}_f$-invariant,

$$\operatorname{tr}\pi_f(\tau) = \int_{G(\mathbb{A}_{\mathbb{Q},f})} \tau(x)\varphi(x)\,dx$$

where φ is a normalized spherical matrix coefficient for π_f. This equals

$$\int_{Z(\mathbb{A}_{\mathbb{Q},f})\backslash G(\mathbb{A}_{\mathbb{Q},f})} \int_{Z(\mathbb{A}_{\mathbb{Q},f})} \tau(zg)\,dz\,\varphi(g)\,dg = \int_{Z(\mathbb{A}_{\mathbb{Q},f})\backslash G(\mathbb{A}_{\mathbb{Q},f})} \sum_{\gamma\in Z(\mathbb{Q})/Z(\mathbb{Z})} \tau(\gamma g)\varphi(g)\,dg$$

Hence the above equals

$$\int_{Z(\mathbb{A}_{\mathbb{Q},f})\backslash G(\mathbb{A}_{\mathbb{Q},f})} \left(\sum_{\gamma\in Z(\mathbb{Q})/Z(\mathbb{Z})} \tau(\gamma g) \right) \varphi(g)\,dg = \operatorname{tr}\pi_f(\tilde{\tau})$$

where

$$\tilde{\tau}(x) = \sum_{\gamma\in Z(\mathbb{Q})/Z(\mathbb{Z})} \tau(\gamma x) = \sum_{\gamma\in Z(\mathbb{Z}[S_0^{-1}])/Z(\mathbb{Z})} \tau_{S_0}(\gamma x) = \widetilde{\tau_{S_0}}(x)$$

with $\mathbb{Z}[S_0^{-1}] = \mathbb{Z}[p^{-1} \mid p \in S_0]$. In particular,

$$\sum_{\pi\in\mathcal{F}(t)} \operatorname{tr}\pi_{S_0}(\tau_{S_0}) = \sum_{\pi\in\mathcal{F}^H(t)} \operatorname{tr}\pi_{S_0}(\widetilde{\tau_{S_0}}).$$

Since $|\mathcal{F}(t)| = |\mathcal{F}^H(t)|$ we get by Theorem 2.1

$$\lim_{t\to\infty} \left|\mathcal{F}^H(t)\right|^{-1} \sum_{\pi\in\mathcal{F}^H(t)} \operatorname{tr} \pi_{S_0}(\widetilde{\tau_{S_0}}) = \widetilde{\tau_{S_0}}(1). \tag{7}$$

Each $\pi_0 \in \mathcal{F}^H$ defines a point in

$$\mathcal{A}^{\mathrm{ur}} := \prod_{v<\infty} \widehat{H(\mathbb{Q}_v)}^{\mathrm{ur}},$$

as well as its projection to the S_0-component

$$\mathcal{A}^{\mathrm{ur}}_{S_0} := \prod_{v\in S_0} \widehat{H(\mathbb{Q}_v)}^{\mathrm{ur}}$$

where $\widehat{H(\mathbb{Q}_v)}^{\mathrm{ur}}$ denotes the unramified unitary dual of $H(\mathbb{Q}_v)$. Hence we can ask how the set $\mathcal{F}^H(t)$, considered as a subset of $\mathcal{A}^{\mathrm{ur}}$ or $\mathcal{A}^{\mathrm{ur}}_{S_0}$, distributes in $\mathcal{A}^{\mathrm{ur}}$ or $\mathcal{A}^{\mathrm{ur}}_{S_0}$. For $n = 2$ this question was studied in [Sar87] for $F = \mathbb{Q}$, and in [IR10] for F imaginary quadratic; for groups with discrete series at ∞, this question was studied in [Ser97, CDF97, Shi12, ST15].

For $\pi_v \in \widehat{H(\mathbb{Q}_v)}^{\mathrm{ur}}$ let δ_{π_v} denote the Dirac measure supported at π_v, and let $\delta_{\pi_{S_0}} = \prod_{v\in S_0} \delta_{\pi_v}$. Put

$$\mu^{S_0}_{\mathrm{count},t} = \left|\mathcal{F}^H(t)\right|^{-1} \sum_{\pi_0\in\mathcal{F}^H(t)} \delta_{\pi_{S_0}}.$$

For each $v < \infty$ we also have the spherical Plancherel measure $\mu_{\mathrm{Pl},v}$ on $\widehat{H(\mathbb{Q}_v)}^{\mathrm{ur}}$. Let $\mu_{\mathrm{Pl},S_0} = \prod_{v\in S_0} \mu_{\mathrm{Pl},v}$. Then (7) says that

$$\mu_{\mathrm{count},t}(\widehat{\widetilde{\tau_{S_0}}}) \longrightarrow \mu^{S_0}_{\mathrm{Pl}}(\widehat{\widetilde{\tau_{S_0}}})$$

as $t \to \infty$ for every $\tau_{S_0} \in \mathcal{H}_{S_0}$, and it also gives an upper bound for the error term.

Here $\widehat{\widetilde{\tau_{S_0}}} = \prod_{v\in S_0} \widehat{\widetilde{\tau_v}}$ with $\widehat{\widetilde{\tau_v}}$ defined by

$$\widehat{\widetilde{\tau_v}}(\pi_v) = \operatorname{tr} \pi_v(\widetilde{\tau_v})$$

for every tempered $\pi_v \in \widehat{H(\mathbb{Q}_v)}^{\mathrm{ur}}$. By Sauvageot's density principle [Sau97] (cf. also [Shi12, ST15, FLM15]) this is enough to prove that $\mu^{S_0}_{\mathrm{count,t}} \longrightarrow \mu^{S_0}_{\mathrm{Pl}}$ since the bi-$\mathbf{K}^H_{S_0}$-invariant functions on $H(F_{S_0})$ are contained in the image of $\mathcal{H}_{S_0}$ under the map $\tau_{S_0} \mapsto \widetilde{\tau_{S_0}}$.

2.3.3 Standard L-Functions

The above theorem gives information on the coefficients of L-functions attached to unramified cuspidal representations: If $\pi \in \mathcal{F}$, there is a standard L-function $L(s,\pi)$ associated with π for $\Re s$ sufficiently large. The L-function can be written as a Dirichlet series

$$L(s,\pi) = \sum_{\mathfrak{a}\subseteq\mathcal{O}_F} A_{\mathfrak{a}}(\pi)\mathbb{N}(\mathfrak{a})^{-s}$$

for suitable coefficients $A_{\mathfrak{a}}(\pi) \in \mathbb{C}$, where the sum runs over all integral ideals in $\mathcal{O}_F$, and $\mathbb{N}(\mathfrak{a}) = |\mathcal{O}_F/\mathfrak{a}|$ denotes the norm of the ideal $\mathfrak{a}$. Moreover, for each $\mathfrak{a}$ there exists an element in the Hecke algebra $\tau_{S_0} \in \mathcal{H}_{S_0}$ (with S_0 the set of places dividing $\mathfrak{a}$) such that $A_{\mathfrak{a}}(\pi) = \operatorname{tr}\pi_{S_0}(\tau_{S_0})$ for all $\pi \in \mathcal{F}$. More precisely, this τ is a linear combination of those $\prod_{v<\infty}\tau_{\lambda_v}$ with $\mathbb{N}(\mathfrak{a}) = \prod_{v<\infty} q_v^{\sum_i \lambda_{v,i}}$ and $\lambda_{v,1} \geq \ldots \geq \lambda_{v,n} \geq 0$. In the case of $F = \mathbb{Q}$ and $\mathfrak{a}$ a principal ideal $N\mathbb{Z}$, then $\tau = T_N$ is the usual Hecke operator attached to N [Gol06, §9].

Then the above theorem implies that there exist $a, b, \delta > 0$ such that for every ideal $\mathfrak{a} \subseteq \mathcal{O}_F$ we have

$$\lim_{t\to\infty} |\mathcal{F}(t)|^{-1} \sum_{\pi\in\mathcal{F}(t)} A_{\mathfrak{a}}(\pi) = \begin{cases} 1 & \text{if } \mathfrak{a} = \mathfrak{b}^n \text{ for some principal ideal } \mathfrak{b} \subseteq \mathcal{O}_F, \\ 0 & \text{else,} \end{cases}$$

and further,

$$\left| |\mathcal{F}(t)|^{-1} \sum_{\pi\in\mathcal{F}(t)} A_{\mathfrak{a}}(\pi) - \delta_n(\mathfrak{a}) \right| \leq a\mathbb{N}(\mathfrak{a})^b t^{-\delta}, \quad t \geq 1,$$

where $\delta_n(\mathfrak{a}) = 1$ if $\mathfrak{a}$ is the nth power of some principal ideal in $\mathcal{O}_F$, and $\delta_n(\mathfrak{a}) = 0$ otherwise.

Using Hecke relations, one can similarly compute the asymptotics for higher moments $\sum_{\pi\in\mathcal{F}(t)} A_{\mathfrak{a}}(\pi)^k$ for any $k \in \mathbb{Z}_{\geq 0}$.

2.4 The Relevance of the Error Term

Since much work needs to be invested to prove the estimate (5), we want to indicate briefly a motivation for it: As explained above, the traces of Hecke operators are closely related to standard L-functions of automorphic representations. Our spectral set of representations $\mathcal{F}(t)$ defines a family of L-functions $L(s,\pi)$, $\pi \in \mathcal{F}(t)$. There has been much recent interest in the distribution of low-lying zeros of families of L-functions, cf. [KS99, ILS00, ST15, SST16]. More precisely, one is interested in the k-level densities

$$|\mathcal{F}(t)|^{-1} \sum_{\pi \in \mathcal{F}(t)} \sum_{\gamma_{j_1}^\pi, \ldots, \gamma_{j_k}^\pi} \Phi\left(\frac{\gamma_{j_1}^\pi \log t}{2\pi}, \ldots, \frac{\gamma_{j_k}^\pi \log t}{2\pi}\right), \tag{8}$$

where Φ is a Schwartz–Bruhat function on $\mathbb{R}^k$ whose Fourier transform has compact support, and the $\rho_{j_1}^\pi = \frac{1}{2} + i\gamma_{j_1}^\pi, \ldots, \rho_{j_k}^\pi = \frac{1}{2} + i\gamma_{j_k}^\pi$ run over all pairwise different k-tuples of zeros of $L(s, \pi)$. Since we do not assume GRH, the γ_j^π may happen to be complex, and we identify Φ with its holomorphic extension to $\mathbb{C}^k$. (Similar expressions can be studied for other families of L-functions of course.)

It is conjectured that the low-lying zeros of families of L-functions are distributed according to certain symmetry types associated with the families (cf. [SST16, Conjecture 2]). This means that for any Schwartz–Bruhat function Φ the limit of (8) as $t \to \infty$ is supposed to equal

$$\int_{\mathbb{R}^k} \Phi(x) W(x)\, dx \tag{9}$$

where $W(x)$ is a certain density attached to the conjectured symmetry type.

One can attack this problem by using the explicit formula for L-functions (cf. [ST15]). To control unwanted terms in the explicit formula one then uses the estimate (5) among other things. In particular, one can show that for the family of L-functions attached to $\mathcal{F}(t)$ the expression (8) approaches (9) as $t \to \infty$ for any Schwartz–Bruhat functions Φ whose Fourier transform has sufficiently small support, see [MT]. The quality of the estimate (5) controls the allowed size of the support of the Fourier transform of Φ.

There is another application of our results, see [MT, Corollaries 1.6, 1.7], namely, we can give a bound towards the p-adic Ramanujan conjecture on average (see [LM09] for an average bound towards the archimedean Ramanujan conjecture). If $\pi \in \mathcal{F}$, then for every finite prime p we can identify π_p with its Satake parameter in $\alpha_\pi(p) = \operatorname{diag}(\alpha_\pi^{(1)}(p), \ldots, \alpha_\pi^{(n)}(p)) \in T_0(\mathbb{C})/W$. The p-adic Ramanujan conjecture asserts that in fact $\alpha_\pi(p) \in T_0(\mathbb{C})^1/W$ for all $\pi \in \mathcal{F}$ and all finite primes p where $T_0(\mathbb{C})^1$ denotes the group of all complex diagonal matrices with entries of absolute value 1. From our results we can now deduce the following: For $\theta, t > 0$ define

$$R(p, t, \theta) = |\{\pi \in \mathcal{F}(t) \mid \max_{1 \le j \le n} \log_p |\alpha_\pi^{(j)}(p)| > \theta\}|.$$

Hence the p-adic Ramanujan conjecture asserts that $R(p, t, \theta) = 0$ for every $\theta > 0$. Note that it is known that $R(p, t, \theta) = 0$ whenever $\theta > \frac{1}{2} - \frac{1}{n^2+1}$ by Luo et al. [LRS99]. Then we can deduce, on the one hand, that there are constants $c, \omega > 0$ such that for all $t \ge 1$, all $\theta > 0$ and all finite primes p we have

$$R(p, t, \theta) \le C t^{d - c\theta + \frac{\omega}{\log p}}$$

for some $C > 0$, a constant which depends on p and θ. On the other hand we can show that if we are given a finite set S_0 of finite primes, then for every $\theta > 0$ there exists a constant $\rho > 0$ such that

$$R(p, t, \theta) \leq C' t^{d-\rho}$$

for all $t \geq 1$. Here $C' > 0$ is again a constant depending only on S_0 and θ.

3 Idea of Proof

The main tool for proving Theorem 2.1 is the Arthur–Selberg trace formula for the group $G = \mathrm{GL}(n)$ over F (again, $F = \mathbb{Q}$ or F is imaginary quadratic in this section). It is a common approach to use various kinds of trace formulae to prove the Weyl law in its different forms, cf. [Sel56, DKV79, Mil01, Mül07, LV07, LM09, Mül16]. In fact, one motivation for Selberg to develop the trace formula was to prove the Weyl law for locally symmetric spaces $\Gamma \backslash \mathrm{SL}_2(\mathbb{R}) / \mathrm{SO}(2)$ for $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ an arithmetic congruence subgroup.

Recall that the Arthur–Selberg trace formula is an identity of distributions

$$J_{\mathrm{geom}}(f) = J_{\mathrm{spec}}(f)$$

of the so-called geometric and spectral side on the space of smooth, compactly supported test functions $f \in C_c^\infty(G(\mathbb{A}_F)^1)$, cf. [Art05]. The main strategy is then as follows: For an appropriate choice of test function (or rather a family of test function—see below for details), it is not too hard to show that $\sum_{\pi:\, \lambda_{\pi_\infty} \in t\Omega} \operatorname{tr} \pi_f(\tau)$ is the main part of the spectral side (or rather of some integral over $t\Omega$ of the spectral side) as $t \to \infty$. Similarly, it can be shown that $|\mathcal{F}(t)|^{-1} \sum_{z \in Z(F)/Z(\mathcal{O}_F)} \tau(z)$ is the main part of the (integral over $t\Omega$ of the) geometric side. The main difficulty is to obtain an upper bound for the error term, and in particular, to prove its effectiveness in τ. This is achieved by analyzing the remaining parts of the geometric and spectral side of the trace formula.

Finding good upper bounds for the remaining parts of the geometric side of the trace formula is the most difficult part. Bounding the remaining parts on the spectral side is very similar to the proof in [LM09], and we will not go into further details. Many of the problems on the geometric side which we need to consider do not appear in the treatment of the geometric side in [LM09]. This is because in [LM09] the non-archimedean test function is fixed in contrast to the fact that we want to vary our S_0 and τ_{S_0}. In fact, in [LM09] it can be achieved that only the unipotent part of the geometric side of the trace formula remains to study (see also below for a short reminder of the coarse expansion of $J_{\mathrm{geom}}(f)$).

To explain the proof in some more detail we first need to explain our choice of test functions.

3.1 Test Functions

The family of test functions used in our proof is constructed with the spectral side in mind: It is of the form $F^{\mu,\tau} = (f_\infty^\mu \cdot \tau)_{|G(\mathbb{A}_F)^1}$ for a suitable family of bi-$\mathbf{K}_\infty$-invariant functions $f_\infty^\mu \in C_c^\infty(G(F_\infty)^1 /\!/ \mathbf{K}_\infty)$ depending on the spectral parameter $\mu \in \mathfrak{a}_\mathbb{C}^*$. The choice of the non-archimedean part of the test function suggests itself from what we want to get from the cuspidal part of the trace formula, and it is the same as in [LM09]. More precisely, it is chosen such that $\operatorname{tr} \pi_\infty(f_\infty^\mu)$ only contributes if λ_{π_∞} is very close to μ. In particular, the integral

$$\int_{t\Omega} \sum_{\pi \in \mathcal{F}} \operatorname{tr} \pi(F^{\mu,\tau})\, d\mu \tag{10}$$

basically captures only those $\pi \in \mathcal{F}$ with $\lambda_{\pi_\infty} \in t\Omega$, that is, it equals

$$\sum_{\pi \in \mathcal{F}:\ \lambda_{\pi_\infty} \in t\Omega} \operatorname{tr} \pi_f(\tau)$$

up to an error term which can be estimated, cf. also [LM09].

The family f_∞^μ is constructed following the ideas of [DKV79]. By the Paley–Wiener Theorem the diagram

$$\begin{tikzcd} C_c^\infty(G(F_\infty)^1 /\!/ \mathbf{K}_\infty) \arrow[r, "\mathcal{H}"] \arrow[dr, "\mathcal{A}"'] & \mathcal{P}(\mathfrak{a}_\mathbb{C}^*)^W \\ & C_c^\infty(\mathfrak{a}_\mathbb{C})^W \arrow[u, "\mathcal{F}"'] \end{tikzcd}$$

is commutative and all maps are isomorphisms. Here:

- $\mathcal{P}(\mathfrak{a}_\mathbb{C}^*)^W$ is the space of Weyl group invariant Paley–Wiener functions on $\mathfrak{a}_\mathbb{C}^*$,
- $\mathcal{H}$ denotes the spherical Fourier transform (= Harish–Chandra transform),
- $\mathcal{A}$ is the Abel transform, and
- $\mathcal{F}$ the Fourier transform.

Hence the inverses $\mathcal{A}^{-1}$ and $\mathcal{H}^{-1}$ are well defined. If $h \in C_c^\infty(\mathfrak{a}_\mathbb{C})^W$ and $\mu \in \mathfrak{a}_\mathbb{C}^*$, we put $h_\mu(X) := h(X)e^{-\langle \mu, X\rangle}$ where $\langle \cdot, \cdot \rangle$ denotes the pairing on $\mathfrak{a}_\mathbb{C}^* \times \mathfrak{a}_\mathbb{C}$. One then fixes an appropriate choice of $h \in C_c^\infty(\mathfrak{a}_\mathbb{C})^W$ as in [DKV79] (cf. [LM09]) and puts $f_\infty^\mu := \mathcal{A}^{-1}(h_\mu)$. More precisely,

$$f_\infty^\mu(g) = |W|^{-1} \int_{i\mathfrak{a}^*} \mathcal{F}(h_\mu)(\lambda)\phi_\lambda(g)\mathbf{c}(\lambda)^{-2}\, d\lambda, \tag{11}$$

where

$$\phi_\lambda(g) = \int_{\mathbf{K}_\infty} e^{\langle \lambda+\rho, H_0(kg)\rangle}\, dk$$

is the elementary spherical function of parameter λ, and $\mathbf{c}(\lambda)$ denotes the Harish-Chandra $\mathbf{c}$-function for $G(F_\infty)$.

3.2 Expansions of the Geometric Side

The starting point for the analysis of $J_{\mathrm{geom}}(F^{\mu,\tau})$ is its coarse expansion, see [Art78], [Art05, §10]: Two elements $g_1, g_2 \in G(F)$ are called geometrically equivalent if their semisimple parts (in the Jordan decomposition) are conjugate in $G(F)$. Since $G = \mathrm{GL}(n)$, this amounts to saying that g_1 and g_2 have the same characteristic polynomial. $G(F)$ then decomposes into a disjoint union of geometric equivalence classes under this relation, and we write $\mathcal{O}$ for the set of all these equivalence classes.

Example 3.1. The variety of unipotent elements $\mathfrak{o}_{\mathrm{unip}}$ in $G(F)$ constitutes one of the equivalence classes in $\mathcal{O}$. Similarly, for any central element $\gamma \in Z(F)$, the geometric equivalence class generated by γ equals $\gamma\mathfrak{o}_{\mathrm{unip}}$.

Arthur shows that there exist distributions $J_{\mathfrak{o}} : C_c^\infty(G(\mathbb{A}_F)^1) \longrightarrow \mathbb{C}$, $\mathfrak{o} \in \mathcal{O}$, such that

$$J_{\mathrm{geom}}(f) = \sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}(f),$$

see [Art05, §10]. For a fixed compactly supported test function, all but finitely many $J_{\mathfrak{o}}(f)$ vanish so that the coarse expansion is in fact a finite sum. More precisely, the distribution $J_{\mathfrak{o}}$ has support in

$$\bigcup_{\gamma\in\mathfrak{o}} \mathrm{Ad}\, G(\mathbb{A}_F)\cdot\gamma,$$

where $\mathrm{Ad}\, G(\mathbb{A}_{\mathbb{F}})\cdot\gamma$ is the $G(\mathbb{A}_F)$-conjugacy class of γ.

Each of the distributions $J_{\mathfrak{o}}$ has a finer expansion (cf. [Art05, §19]): Let $\mathfrak{o} \in \mathcal{O}$ and let S be a sufficiently large set of places of F depending on f and $\mathfrak{o}$ as explained in [Art86, §7]. In particular, S must contain the archimedean place of F, and it has to be so large that f can be written as $f_S \otimes \mathbf{1}_{\mathbf{K}^S}$ with $f_S \in C_c^\infty(G(F_S)^1)$ and $\mathbf{1}_{\mathbf{K}^S} \in C_c^\infty(G(\mathbb{A}_F^S))$ the characteristic function of $\mathbf{K}^S = \prod_{v\notin S}\mathbf{K}_v$. Then

$$J_{\mathfrak{o}}(f) = \sum_M \frac{|W^M|}{|W^G|}\sum_\gamma a^M(S,\gamma) J_M^G(\gamma, f_S), \tag{12}$$

where

- M runs over all F-Levi subgroups of G containing the maximal torus T_0 of diagonal matrices,
- W^M denotes the Weyl group of the pair (T_0, M),
- γ runs over a (arbitrary) set of representatives for the $M(F)$-conjugacy classes in $M(F) \cap \mathfrak{o}$,
- $a^M(S, \gamma) \in \mathbb{C}$ are certain "global" coefficient that are independent of f,
- $J_M^G(f_S, \gamma)$ are certain S-adic weighted orbital integrals, and
- $a^M(S, \gamma) \in \mathbb{C}$ and $J_M^G(f_S, \gamma)$ depend only on the $M(F)$-conjugacy class of γ.

Since there are only finitely many $M(F)$-conjugacy classes in $M(F) \cap \mathfrak{o}$, this fine expansion of $J_\mathfrak{o}(f)$ is a finite sum. One should note that the sum over γ in general needs to be taken over a set of representatives for a certain equivalence relation on $M(F) \cap \mathfrak{o}$ that depends on S. It is a special feature of $G = \mathrm{GL}(n)$ that this equivalence relation reduces to conjugacy and thus is independent of S.

Using our family of test functions $F^{\mu,\tau}$ in the geometric side of the trace formula and integrating over $\mu \in t\Omega$ (hence mirroring the integral (10) on the geometric side of the trace formula), we need to consider for each $\mathfrak{o} \in \mathcal{O}$ the sum-integral

$$\sum_M \frac{|W^M|}{|W^G|} \sum_\gamma a^M(S, \gamma) \int_{t\Omega} J_M^G(f_\infty^\mu \cdot \tau_{S\setminus\{\infty\}}, \gamma)\, d\mu.$$

The pairs $(M, \gamma) \in \{G\} \times Z(F)$ are exactly those which contribute to the main term: If $M = G$ and $\gamma \in Z(F)$, one has

$$a^G(S, \gamma) = \mathrm{vol}(G(F)\backslash G(\mathbb{A}_F)^1)$$

and

$$\int_{t\Omega} J_M^G(f_\infty^\mu \cdot \tau_{S\setminus\{\infty\}}, \gamma)\, d\mu = \tau(z) \int_{t\Omega} f_\infty^\mu(1)\, d\mu.$$

Using Plancherel inversion, one can show that this last integral equals $|W|^{-1} \int_{t\Omega} \mathbf{c}(\lambda)^{-2}\, d\lambda$ up to a contribution to the error term (see [LM09]).

Hence it remains to show that the rest of the geometric side only contributes to the error term in (5). The remaining main steps in the proof of Theorem 2.1 in [Mat, MT] are therefore as follows:

1. Find the (finitely many) classes $\mathfrak{o} \in \mathcal{O}$ for which $J_\mathfrak{o}(F^{\mu,\tau}) \neq 0$, and keep track of how they depend on τ.
2. Find a sufficiently large set of places S such that the fine expansion (12) holds for any $\mathfrak{o}$ from step (1). Keep track of the dependence of S on τ.
3. For any pair $(M, \gamma) \notin \{G\} \times Z(F)$ with $\gamma \in \mathfrak{o} \cap M(F)$ find an upper bound for $a^M(S, \gamma)$ for any $\mathfrak{o}$ from step (1) and S from step (2). Keep track of the dependence on τ.

4. For any pair $(M, \gamma) \notin \{G\} \times Z(F)$ with $\gamma \in \mathfrak{o} \cap M(F)$ find an upper bound for the integral $|\int_{t\Omega} J_M^G(F_S^{\mu,\tau}, \gamma)\, d\mu|$ for any $\mathfrak{o}$ from step (1) and S from step (2). Keep track of the dependence on τ.

We will not comment any further on steps (1) and (2) but explain the relevance and main difficulties in the last two steps.

3.3 Global Coefficients

The global coefficients $a^M(S, \gamma)$ are in general only understood in some special cases, although there has been some recent progress [CL, Cha]. If γ is semisimple, $a^M(S, \gamma)$ is independent of S and equals

$$a^M(S, \gamma) = \mathrm{vol}(M_\gamma(F)\backslash M_\gamma(\mathbb{A}_F)^1),$$

where $M_\gamma(F)$ is the centralizer of γ in $M(F)$ [Art86, Theorem 8.2]. If γ is not semisimple, exact expressions for $a^M(S, \gamma)$ are only known in a few low-rank examples [JL70, Fli82, HW]. For $\mathrm{GL}(n)$ there exists at least an upper bound which is sufficiently good to prove the error estimate in (5) [Mat15]: There exist $a, b > 0$ depending only on n and the degree of F over $\mathbb{Q}$ such that

$$a^M(S, \gamma) \le a D_F^b \sum_{\substack{(s_v)_{v\in S} \in \mathbb{Z}_{\ge 0}^{|S|}:\\ \sum_v s_v \le n-1}} \prod_{v \in S\setminus S_\infty} \left| \frac{\zeta_{F_v}^{(s_v)}(1)}{\zeta_{F_v}(1)} \right|,$$

where $\zeta_{F_v}(s) = (1 - q_v^{-s})^{-1}$ denotes the local Dedekind zeta function, and $\zeta_{F_v}^{(s_v)}(s)$ its s_vth derivative. For certain types of γ the upper bound for $a^M(\gamma, S)$ has recently been improved in [Cha].

3.4 Erratum to [Mat15]

There is a mistake in the volume formula for $G(F)\backslash G(\mathbb{A}_F)^1$ as stated in [Mat15] which has some effect on the formulation of a conjecture in that paper.

In fact, in the normalization of measures in [Mat15] the adelic quotient $G(F)\backslash G(\mathbb{A}_F)^1$ has volume

$$\mathrm{vol}(G(F)\backslash G(\mathbb{A}_F)^1) = D_F^{\frac{n(n-1)}{4}} \operatorname*{res}_{s=1} \zeta_F(s) \prod_{k=2}^{n} \zeta_F(k),$$

where ζ_F is the Dedekind zeta function of F, and if $n = 1$, the empty product is interpreted as 1. This formula was incorrectly stated in [Mat15] where the factor $D_F^{\frac{n(n-1)}{4}}$ was missing on the right-hand side. This does not have any effect on the statement or proof of the results of Matz [Mat15]. However, the statement of the first part of Matz [Mat15, Conjecture 1.3] needs to be modified by the obvious power of the discriminant of D_F.

More precisely, the inequality (4) in [Mat15, Conjecture 1.3] should read

$$\left|a^M(\mathcal{V}, S)\right| \le C D_F^{N_M+\kappa} \sum_{\substack{s_v \in \mathbb{Z}_{\ge 0}, v \in S_{\text{fin}}: \\ \sum s_v \le \eta}} \prod_{v \in S_{\text{fin}}} \left| \frac{\zeta_{F,v}^{(s_v)}(1)}{\zeta_{F,v}(1)} \right|,$$

see [Mat15] for the missing notation. Here the number N_M is defined as follows: There is a partition $(n_1, \dots, n_r)$ of n such that M is over F isomorphic to $\mathrm{GL}(n_1) \times \dots \times \mathrm{GL}(n_r)$. We then define $N_M = \sum_{i=1}^r n_i(n_i - 1)/4$.

3.5 Weighted Orbital Integrals

To attack step (4), one first needs to better understand the weighted orbital integrals. The first step is to reduce the S-adic integral $J_M^G(\gamma, f_S)$ to a linear combination of products of v-adic integrals for $v \in S$. This can be done by using Arthur's splitting formula for weighted orbital integrals [Art88, § 9]. It reduces step (4) basically to two different problems, namely, to bound for every Levi $L \supseteq M$

- the archimedean integral:

$$\left| \int_{t\Omega} J_M^L(\gamma, f_\infty^\mu)\, d\mu \right|, \tag{13}$$

- the non-archimedean integrals $\left|J_M^L(\gamma, \tau_v)\right|$ for $v \in S \backslash \{\infty\}$.

For the non-archimedean integrals, it was shown in [Mat] by using explicit computations on the Bruhat–Tits building as in [ST15, § 7] combined with bounds for unweighted orbital integrals [ST15, § 7, Appendix B] that there exist $a, b, c > 0$ depending only on n and the global field F such for any non-archimedean v, any $\kappa_v \ge 0$, and any $\tau_v \in \mathcal{H}_v^{\le \kappa_v}$, $|\tau_v| \le 1$, we have

$$\left|J_M^L(\gamma, \tau_v)\right| \le q_v^{a+b\kappa_v} \Delta_v^-(\gamma)^c$$

where

$$\Delta_v^-(\gamma) := \prod_\alpha \max\{1, |1 - \alpha(\tilde{\gamma})|_{F(\gamma)}^{-1}\}$$

with α running over all positive roots of (T_0, G), $F(\gamma)/F$ the splitting field of γ, and $\tilde{\gamma} \in T_0(F(\gamma))$ a diagonal matrix having the same eigenvalues as γ (in $F(\gamma)$). Note that $\Delta_v^-(\gamma)$ is well defined since the entries of $\tilde{\gamma}$ are unique up to permutation (i.e., $\tilde{\gamma}$ is unique up to conjugation by Weyl group elements).

To estimate (13), we require a good pointwise upper bound for the elementary spherical functions ϕ_λ. This task is significantly easier if F is imaginary quadratic than if $F = \mathbb{Q}$. In the former case $F_\infty = \mathbb{C}$, the elementary spherical functions ϕ_λ for $\mathrm{GL}_n(\mathbb{C})$ are well understood and can be expressed as rational functions in $e^{\langle \lambda, H_0(\cdot)\rangle}$ and $e^{\langle \rho, H_0(\cdot)\rangle}$. In the latter case, $F_\infty = \mathbb{R}$, the elementary spherical functions for $\mathrm{GL}_n(\mathbb{R})$ can only be expressed as integrals, but not as rational functions of elementary functions as in the complex case. It is not easy to obtain a non-trivial estimate for these functions which is effective in the spectral parameter as well as the group parameter. Recently a sufficiently good upper bound for these spherical functions was proven in [BP, MT]. There were several preceding upper bounds for spherical functions, cf. [DKV83, Mar], but they always required at least one of the variables (the spherical parameter or the group element) to stay in a bounded set and away from the singular set.

4 Further Directions

4.1 Improving the Error Term

As explained in Sect. 2.4, the effective dependence of the bound (5) on τ makes Theorem 2.1 applicable in proving certain conjectures about low-lying zeros of families of L-functions. It would be desirable to improve the bound (5) or at least to control the constant b in terms of n as this would lead to a better understanding of how large the support of the Fourier transform of the test function Φ in (8) may be.

The main obstacles when trying to give an upper bound for b are bounding the non-archimedean weighted orbital integrals $J_M^L(\gamma, \tau_v)$, and bounding the global coefficients $a^M(\gamma, S)$. In principle, the upper bounds for both quantities can be at least made effective in n, but with our types of proofs only very crude bounds would arise. Recent work [Cha] gives good bounds for the global coefficients in some special cases.

4.2 General Number Fields

The purpose of this section is to formulate the analogue of our main theorem over a general number field [see (14)], and to explain what points then need to be changed in the proof of the theorem. In particular, the construction of the archimedean test function needs to be modified.

Suppose F is a number field of degree $d = [F : \mathbb{Q}]$ with r_1 real and r_2 complex places so that $d = r_1 + 2r_2$. For each $v|\infty$ let $\mathfrak{a}_{v,0}^* = X(T_0/\mathbb{Q})_{\mathbb{Q}} \otimes \mathbb{R} \simeq \mathbb{R}^n$ where $X(T_0/\mathbb{Q})_{\mathbb{Q}}$ denotes the group of rational characters $T_0 \longrightarrow \mathrm{GL}(1)$ for T_0 considered as a group over $\mathbb{Q}$, and $\mathfrak{a}_{v,0} = \mathrm{Hom}_{\mathbb{R}}(\mathfrak{a}_{v,0}^*, \mathbb{R})$. Similarly, let $\mathfrak{a}_0^* = X(\mathrm{Res}_{F/\mathbb{Q}} T_0)_{\mathbb{Q}} \otimes \mathbb{R}$ and $\mathfrak{a}_0 = \mathrm{Hom}_{\mathbb{R}}(\mathfrak{a}_0^*, \mathbb{R})$ with $X(\mathrm{Res}_{F/\mathbb{Q}} T_0)_{\mathbb{Q}}$ the group of $\mathbb{Q}$-characters of T_0 as a group over F. Then

$$\mathfrak{a}_0 \simeq \bigoplus_{v|\infty} \mathfrak{a}_{v,0}, \text{ and } \mathfrak{a}_0^* \simeq \bigoplus_{v|\infty} \mathfrak{a}_{v,0}^*.$$

We define $\mathfrak{a}_{v,G}$, $\mathfrak{a}_{v,G}^*$, $\mathfrak{a}_G$, and $\mathfrak{a}_G^*$ similarly with G in place of T_0. We let $\mathfrak{a}_v$, $\mathfrak{a}_v^*$, $\mathfrak{a}$, and $\mathfrak{a}^*$ be the spaces such that $\mathfrak{a}_{v,0} = \mathfrak{a}_v \oplus \mathfrak{a}_{v,G}$, $\mathfrak{a}_{v,0}^* = \mathfrak{a}_v^* \oplus \mathfrak{a}_{v,G}^*$, and so on. Let

$$\mathfrak{a}^\infty = \{X = (X_i)_{1 \le i \le n(r_1+r_2)} \in \mathfrak{a}_0 \mid \sum_i X_i = 0\},$$

and

$$\mathfrak{a}^{\infty,*} = \{\lambda = (\lambda_i)_{1 \le i \le n(r_1+r_2)} \in \mathfrak{a}_0^* \mid \sum_i \lambda_i = 0\}.$$

Further, let

$$\mathfrak{a}_G^\infty = \mathfrak{a}_G \cap \mathfrak{a}^\infty, \text{ and } \mathfrak{a}_G^{\infty,*} = \mathfrak{a}_G^* \cap \mathfrak{a}^{\infty,*}.$$

If $\pi \in \Pi_{\mathrm{cusp}}(G(\mathbb{A}_F)^1)$, the infinitesimal character λ_{π_∞} is now an element in $\mathfrak{a}_{\mathbb{C}}^{\infty,*}$. It has a unique decomposition

$$\lambda_{\pi_\infty} = \lambda_{\xi_{\pi_\infty}} + \sum_{v|\infty} \lambda_{\pi_v'} \in \mathfrak{a}_{G,\mathbb{C}}^{\infty,*} \oplus \bigoplus_{v|\infty} \mathfrak{a}_{v,\mathbb{C}}^*$$

where $\lambda_{\xi_{\pi_\infty}}$ corresponds to the central character ξ_{π_∞} of π_∞, $\pi_\infty' = \xi_{\pi_\infty}^{-1} \pi_\infty$, and $\pi_\infty' = \prod_{v|\infty} \pi_v'$.

Let $\Omega \subseteq \mathfrak{a}_{\mathbb{C}}^{\infty,*}$ be a nice bounded set. For simplicity we assume that Ω is of the form

$$\Omega = \Omega_Z \oplus \bigoplus_{v|\infty} \Omega_v$$

for suitable nice bounded subsets $\Omega_Z \subseteq \mathfrak{a}_{G,\mathbb{C}}^{\infty,*}$, and $\Omega_v \subseteq \mathfrak{a}_{v,\mathbb{C}}^*$.

For each $v|\infty$ let $f_v^{\mu_v}$, $\mu_v \in \mathfrak{a}_{v,\mathbb{C}}^*$ be constructed as before. Let $f_Z : Z(F_\infty)/\mathbb{R}_{>0}(Z(F_\infty) \cap \mathbf{K}_\infty) \longrightarrow \mathbb{C}$ be a compactly supported function with $f_Z(1) = 1$, and let $\widehat{f_Z}$ denote its Fourier transform on $\mathfrak{a}_{G,\mathbb{C}}^{\infty,*}$. For $\mu_Z \in \mathfrak{a}_{G,\mathbb{C}}^{\infty,*}$ let $f_Z^{\mu_Z}$ be such that $\widehat{f_Z^{\mu_Z}}(\lambda_Z) = \widehat{f_Z}(\lambda_Z - \mu_Z)$. We then define

$$f_\infty^\mu(g) = f_Z^{\mu_Z}(z) \prod_{v|\infty} f_v^{\mu_v}(g_v')$$

for $g \in G(F_\infty)^1$ with z its central component in $Z(F_\infty) \cap G(F_\infty)^1$, $g' = z^{-1}g = \prod_{v|\infty} g'_v$, and $\mu = \mu_Z + \sum_{v|\infty} \mu_v \in \mathfrak{a}_{G,\mathbb{C}}^{\infty,*} \oplus \bigoplus_{v|\infty} \mathfrak{a}_{v,\mathbb{C}}^* = \mathfrak{a}_{\mathbb{C}}^{\infty,*}$.

This choice of test function allows us to essentially reduce the analysis of the trace formula to the previously considered cases for $F = \mathbb{Q}$ or F imaginary quadratic. In particular, the integral over the cuspidal part of the spectral side of the trace formula for the test function $f_\infty^\mu \cdot \tau$

$$\int_{t\Omega} J_{\mathrm{cusp}}(f_\infty^\mu \cdot \tau)\, d\mu$$

should equal, up to an error term,

$$\sum_{\pi \in \mathcal{F}(t)} \operatorname{tr} \pi_f(\tau),$$

where $\mathcal{F} = \{\pi \in \Pi_{\mathrm{cusp}}(G(\mathbb{A}_F)^1) \mid \pi^{\mathbf{K}} \neq 0\}$ and $\mathcal{F}(t) = \{\pi \in \mathcal{F} \mid \lambda_{\pi_\infty} \in t\Omega\}$ as before. On the other hand, if $z \in Z(F)$, then it should follow similarly as in the other cases that up to a negligible error term we have

$$\sum_{z \in Z(F)} \int_{t\Omega} f_\infty^\mu \cdot \tau(z)\, d\mu = \Lambda(t) \sum_{z \in Z(F_1)} \tau(z)$$

with

$$\Lambda(t) = \mathrm{vol}(G(F)\backslash G(\mathbb{A}_F)^1/\mathbf{K}_f) \prod_{v|\infty} |W|^{-1} \int_{t\Omega_v} \mathbf{c}_v(\lambda)^{-2}\, d\lambda$$

where $\mathbf{c}_v$ denotes the Harish-Chandra $\mathbf{c}$-function for $G(F_v)^1$, and F_1 the set of all elements in $F^\times$ which lie in the kernel of the composite map

$$F^\times \longrightarrow \mathfrak{a}_G \longrightarrow \mathfrak{a}_G^\infty.$$

Here the first map is given by $x \mapsto (\log |x_v|_v)_{v|\infty}$, and the second map is the orthogonal projection onto $\mathfrak{a}_G^\infty$. That $z \in Z(F)\backslash Z(F_1)$ only contribute to the error term can be seen by Fourier inversion and integration by parts. The remaining parts of the trace formula again should only contribute to the error term.

Hence the final statement is expected to be

$$\lim_{t \to \infty} \Lambda(t)^{-1} \sum_{\pi \in \mathcal{F}(t)} \operatorname{tr} \pi_{S_0}(\tau_{S_0}) = \sum_{z \in Z(F_1)} \tau(z). \tag{14}$$

4.3 General Level

In the previous section we only considered the family $\mathcal{F}$ of everywhere unramified cuspidal representations. If $K_f \subseteq \mathbf{K}_f$ is a finite index subgroup, one can more generally consider the family

$$\mathcal{F}_K = \{\pi \in \Pi_{\mathrm{cusp}}(G(\mathbb{A}_F)^1) \mid \pi^K \neq 0\}$$

of cuspidal representations having a K-fixed vector for $K := \mathbf{K}_\infty \cdot K_f$. We can accordingly put $\mathcal{F}_K(t) = \{\pi \in \mathcal{F}_K \mid \lambda_{\pi_\infty} \in t\Omega\}$. For $F = \mathbb{Q}$ and K_f contained in $K_f(N)$ for some $N \geq 3$, the Weyl law was proven in [LM09] as explained above. However, the dependence of the estimate of the error term on K_f was left unspecified in [LM09]. It might be interesting to make this dependence explicit as this might also allow to study families of representations with varying level.

Remark 4.1. The method of using the trace formula to prove the Weyl law has the disadvantage that one counts the representations in $\mathcal{F}_K(t)$ with a certain weight factor, namely the dimension $\dim \pi^K$ of the K-fixed space of π. If $K = \mathbf{K}$ is the maximal compact subgroup, then by multiplicity-one one has $\dim \pi^{\mathbf{K}} = 1$ for every $\pi \in \mathcal{F}(t) = \mathcal{F}_{\mathbf{K}}(t)$ so that in this case one indeed counts the number of representations in $\mathcal{F}_{\mathbf{K}}(t)$. It would be interesting to see whether one can count the number of $\pi \in \mathcal{F}_K(t)$ of conductor K, or at least the number of newforms over $\pi \in \mathcal{F}_K(t)$.

In [Mat] the upper bound in the error term was made effective in K_f if the ground field F is imaginary quadratic (the same can be done for $F = \mathbb{Q}$). More precisely, we prove the following in [Mat]: Let $K_f \subseteq \mathbf{K}_f$ be a finite index subgroup and put $K = \mathbf{K}_\infty \cdot K_f$. The Weyl law then becomes (cf. (2) from [LM09] for $F = \mathbb{Q}$)

$$\Lambda_K(t) := \sum_{\pi \in \mathcal{F}_K(t)} \dim \pi^K \sim |Z(F) \cap K_f| \frac{\mathrm{vol}(G(F)\backslash G(\mathbb{A}_F)^1/K)}{|W|} \int_{t\Omega} \mathbf{c}(\lambda)^{-2}\, d\lambda \tag{15}$$

as $t \to \infty$. (Recall that $\mathbf{c}$ denotes the Harish-Chandra $\mathbf{c}$-function on $G(F_\infty)$, that is, here it is the $\mathbf{c}$-function for $\mathrm{GL}_n(\mathbb{C})$.) Moreover, there exist constants $a, b, c, \delta > 0$ depending only on n, F, and Ω such that the following holds: Let $\Xi \subseteq G(\mathbb{A}_{F,f})$ be an open compact subset which is bi-K_f-invariant (that is, $k_1 \Xi k_2 = \Xi$ for all $k_1, k_2 \in K_f$), and let $\tau_\Xi \in C_c^\infty(G(\mathbb{A}_{F,f}))$ be the characteristic function of Ξ normalized by $\mathrm{vol}(K_f)^{-1}$. Then

$$\lim_{t\to\infty} \Lambda_K(t)^{-1} \sum_{\pi \in \mathcal{F}_K(t)} \mathrm{tr}\, \pi_f(\tau_\Xi) = \sum_{z \in Z(F)/Z(F)\cap K_f} \tau_\Xi(z) = \left|(Z(F) \cap \Xi)/(Z(F) \cap K_f)\right|,$$

and

$$\left| \Lambda_K(t)^{-1} \sum_{\pi \in \mathcal{F}_K(t)} \operatorname{tr} \pi_f(\tau_\Xi) - \left|(Z(F) \cap \Xi)/(Z(F) \cap K_f)\right| \right| \le a[\mathbf{K} : K]^b \operatorname{vol}(\Xi)^c t^{-\delta} \tag{16}$$

for every $t \geq 1$.

Remark 4.2. (i) Taking $\Xi = K_f$, (16) also provides an upper bound for the error term in (15).

(ii) If $K_f = \mathbf{K}_f$, the upper bound for the remainder term in (16) is the same a in Theorem 2.1: In the situation of Theorem 2.1 we may assume that $\tau_{S_0} = \prod_{v \in S_0} \tau_v$ with $\tau_v \in \mathcal{H}_v^{\leq \kappa_v}$ the characteristic function of $\Xi_v := \mathbf{K}_v \varpi_v^{\lambda_v} \mathbf{K}_v$ for suitable λ_v with $\|\lambda_v\| \leq \kappa_v$. But then the volume of $\Xi = \prod_{v<\infty} \Xi_v$ (which equals the degree of the Hecke operator τ) is $\leq \Pi_\kappa^a$ for some $a > 0$ depending only on n and F.

4.4 General $\mathbf{K}_\infty$-type

So far we only considered representations with trivial $\mathbf{K}_\infty$-type, that is, π such that $\pi_\infty^{\mathbf{K}_\infty} \neq 0$. Suppose σ is an irreducible unitary representation of $\mathbf{K}_\infty$ with representation space V_σ. One can consider $\pi \in \Pi_{\text{cusp}}(G(\mathbb{A}_F)^1)$ which have $\mathbf{K}_\infty$-type σ, that is, for which σ occurs in the decomposition of the restriction of π_∞ to $\mathbf{K}_\infty$ into irreducibles. For $F = \mathbb{Q}$ and $K_f \subseteq \mathbf{K}_f$ a finite index subgroup, the main term of the Weyl law for representations with K_f-fixed vector and $\mathbf{K}_\infty$-type σ (with $\sigma(-1) = id$ if $-1 \in K_f$) was proven in [Mül07]. More precisely, taking $\Omega = B_1(0)$ the unit ball in $i\mathfrak{a}^*$ [Mül07] proves that as $t \to \infty$

$$\sum_{\pi \in \mathcal{F}_K(t)} \dim \pi_f^{K_f} \dim (\mathcal{H}_{\pi_\infty} \otimes V_\sigma)^{\mathbf{K}_\infty} \sim \frac{\delta_{K_f} \dim \sigma}{\operatorname{vol}(K_f)} \frac{\operatorname{vol}(G(\mathbb{Q})\backslash G(\mathbb{A}_\mathbb{Q})^1)}{(4\pi)^{d/2}\Gamma(d/2+1)} t^d$$

where $\mathcal{H}_{\pi_\infty}$ denotes the representation space of π_∞, and δ_{K_f} equals 1 or 2 depending on whether $-1 \notin K_f$ or $-1 \in K_f$.

The method of proof of [Mül07] is not applicable if one wants to obtain a bound on the error term. It might, however, be possible to modify the proof of [LM09, Mat, MT] to incorporate more general $\mathbf{K}_\infty$-types. Already in [MT] we use a particular non-trivial $\mathbf{K}_\infty$-type to obtain odd Maass forms. In general, however, one major obstacle in carrying this approach over to arbitrary σ is that the inversion formula (11) for the spherical Harish-Chandra transform is in general not valid. For certain $\mathbf{K}_\infty$-types it still holds (cf. [HS94, Chap. I, § 5]), but in general one needs to take into account the residues arising in the proof of the Paley–Wiener theorem when changing the contour of certain integrals [Del82, Art83, Shi94].

Suppose for simplicity that σ is one-dimensional, and consider for $\lambda \in i\mathfrak{a}_{\mathbb{C}}^*$ the elementary σ-spherical function

$$\Phi_{\sigma,\lambda}(g) = \int_{\mathbf{K}_\infty} e^{\langle \lambda+\rho, H_0(kg)\rangle} \sigma(k^{-1}\kappa(kg))\, dk, \; g \in G(F_\infty),$$

where $\kappa(kg)$ denotes the $\mathbf{K}_\infty$-component of kg in its Iwasawa decomposition $kg = tuk_1 \in T_0(\mathbb{R})U_0(\mathbb{R})\mathbf{K}_\infty$. Then $\Phi_{\sigma,\lambda}(g) \in \operatorname{End} V_\sigma$, and $\Phi_{\sigma,\lambda}$ satisfies the invariance properties

$$\Phi_{\sigma,\lambda}(k_1 g k_2) = \sigma(k_1 k_2)\Phi_{\sigma,\lambda}(g)$$

for all $k_1, k_2 \in \mathbf{K}_\infty$, $g \in G(F_\infty)$. The Harish-Chandra transform gives a map

$$f \mapsto \mathcal{H}(f)(\lambda) := \int_{G(F_\infty)^1} f(g)\Phi_{\sigma^{-1},\lambda}(g)\, dg, \;\; \lambda \in \mathfrak{a}_{\mathbb{C}}^*$$

for $f \in C_c^\infty(G(F_\infty)^1, \sigma)$, the space of all $f \in C_c^\infty(G(F_\infty)^1)$ satisfying $f(k_1 g k_2) = \sigma(k_1 k_2) f(g)$ for all $k_1, k_2 \in \mathbf{K}_\infty$ and $g \in G(F_\infty)$. The resulting function is a holomorphic function on $\mathfrak{a}_{\mathbb{C}}^*$. However, the inversion formula (11) for $\mathcal{H}$ is only valid for certain σ.

References

[Art05] J. Arthur, An introduction to the trace formula, in *Harmonic Analysis, the Trace Formula, and Shimura Varieties*. Clay Mathematics Proceedings, vol. 4 (American Mathematical Society, Providence, RI, 2005), pp. 1–263

[Art78] J. Arthur, A trace formula for reductive groups. I. Terms associated to classes in $G(\mathbf{Q})$. Duke Math. J. **45**(4), 911–952 (1978)

[Art83] J. Arthur, A Paley-Wiener theorem for real reductive groups. Acta Math. **150**(1), 1–89 (1983)

[Art86] J. Arthur, On a family of distributions obtained from orbits. Can. J. Math. **38**(1), 179–214 (1986)

[Art88] J. Arthur, The invariant trace formula. I. Local theory. J. Am. Math. Soc. **1**(2), 323–383 (1988)

[BP] V. Blomer, A. Pohl, The sup-norm problem for the Siegel modular space of rank two (2014). arXiv:1402.4635v2

[Cha] P.-H. Chaudouard, Sur la contribution unipotente dans la formule des traces d'Arthur pour les groupes généraux linéaires (2014). arXiv:1411.3005

[CDF97] J.B. Conrey, W. Duke, D.W. Farmer, The distribution of the eigenvalues of Hecke operators. Acta Arith. **78**(4), 405–409 (1997)

[CL] P.-H. Chaudouard, G. Laumon, Un théoreme du support pour la fibration de Hitchin (2014). arXiv:1407.3562

[Del82] P. Delorme, Théorème de type Paley-Wiener pour les groupes de Lie semi-simples réels avec une seule classe de conjugaison de sous groupes de Cartan. J. Funct. Anal. **47**(1), 26–63 (1982)

[DKV79] J.J. Duistermaat, J.A.C. Kolk, V.S. Varadarajan, Spectra of compact locally symmetric manifolds of negative curvature. Invent. Math. **52**(1), 27–93 (1979).

[DKV83] J.J. Duistermaat, J.A.C. Kolk, V.S. Varadarajan, Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups. Compos. Math. **49**(3), 309–398 (1983).

[Fli82] Y.Z. Flicker, *The Trace Formula and Base Change for* GL(3). Lecture Notes in Mathematics, vol. 927 (Springer, Berlin, 1982)

[FLM15] T. Finis, E. Lapid, W. Müller, Limit multiplicities for principal congruence subgroups of $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$. J. Inst. Math. Jussieu **14**(03), 589–638 (2015)

[Gol06] D. Goldfeld, *Automorphic Forms and L-Functions for the Group* $\mathrm{GL}(n,\mathbb{R})$. Cambridge Studies in Advanced Mathematics, vol. 99 (Cambridge University Press, Cambridge, 2006)

[HS94] G. Heckman, H. Schlichtkrull, *Harmonic Analysis and Special Functions on Symmetric Spaces*. Perspectives in Mathematics, vol. 16 (Academic, San Diego, CA, 1994)

[HW] W. Hoffmann, S. Wakatsuki, On the geometric side of the arthur trace formula for the symplectic group of rank 2 (2013). arXiv:1310.0541

[ILS00] H. Iwaniec, W. Luo, P. Sarnak, Low lying zeros of families of *L*-functions. Inst. Hautes Études Sci. Publ. Math. **2000**(91), 55–131 (2001)

[IR10] Ö. Imamoḡlu, N. Raulf, On the behaviour of eigenvalues of Hecke operators. Math. Res. Lett. **17**(1), 51–67 (2010).

[JL70] H. Jacquet, R.P. Langlands, *Automorphic Forms on* GL(2). Lecture Notes in Mathematics, vol. 114 (Springer, Berlin, 1970).

[KS99] N. Katz, P. Sarnak, *Random Matrices, Frobenius Eigenvalues, and Monodromy*, vol. 45 (American Mathematical Society, Providence, RI, 1999)

[Lan76] R.P. Langlands, *On the Functional Equations Satisfied by Eisenstein Series*. Lecture Notes in Mathematics, vol. 544 (Springer, Berlin, 1976).

[LM09] E. Lapid, W. Müller, Spectral asymptotics for arithmetic quotients of $\mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$. Duke Math. J. **149**(1), 117–155 (2009)

[LRS99] W. Luo, Z. Rudnick, P. Sarnak, On the generalized Ramanujan conjecture for $\mathrm{GL}(n)$, in *Proceedings of Symposia in Pure Mathematics*, vol. 66 (American Mathematical Society, Providence, RI, 1998, 1999), pp. 301–310

[LV07] E. Lindenstrauss, A. Venkatesh, Existence and Weyl's law for spherical cusp forms. Geom. Funct. Anal. **17**(1), 220–251 (2007)

[Mar] S. Marshall, L^p norms of higher rank eigenfunctions and bounds for spherical functions. J. Eur. Math. Soc. **18**(7), 1437–1493 (2016)

[Mat] J. Matz, Weyl's law for Hecke operators on $\mathrm{GL}(n)$ over imaginary quadratic number fields (2013). arXiv:1310.6525

[Mat15] J. Matz, Bounds for global coefficients in the fine geometric expansion of Arthur's trace formula for $\mathrm{GL}(n)$. Israel J. Math. **205**(1), 337–396 (2015)

[Mil01] S.D. Miller, On the existence and temperedness of cusp forms for $\mathrm{SL}_3(\mathbb{Z})$. J. Reine Angew. Math. **533**, 127–169 (2001)

[MT] J. Matz, N. Templier, Sato-Tate equidistribution for the family of Hecke-Maass forms on $\mathrm{SL}(n,\mathbf{Z})$. arXiv:1505.07285

[Mül07] W. Müller, Weyl's law for the cuspidal spectrum of SL_n. Ann. Math. (2) **165**(1), 275–333 (2007)

[Mül16] W. Müller, Asymptotics of automorphic spectra and the trace formula (2016). These proceedings

[MW95] C. Mœglin, J.-L. Waldspurger, *Spectral Decomposition and Eisenstein Series*. Cambridge Tracts in Mathematics, vol. 113 (Cambridge University Press, Cambridge, 1995). Une paraphrase de l'Écriture (A paraphrase of Scripture)

[Sar87] P. Sarnak, Statistical properties of eigenvalues of the Hecke operators, in *Analytic Number Theory and Diophantine Problems (Stillwater, OK, 1984)*, vol. 70 (Birkhäuser, Boston, MA, 1987), pp. 321–331

[Sau97] F. Sauvageot, Principe de densité pour les groupes réductifs. Compos. Math. **108**(02), 151–184 (1997)

[Sel56] 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.

[Ser97] J.-P. Serre, Répartition asymptotique des valeurs propres de l'opérateur de Hecke T_p. J. Am. Math. Soc. **10**(1), 75–102 (1997)

[Shi12] S.W. Shin, Automorphic Plancherel density theorem. Isr. J. Math. **192**(1), 83–120 (2012)

[Shi94] N. Shimeno, The Plancherel formula for spherical functions with a one-dimensional K-type on a simply connected simple Lie group of Hermitian type. J. Funct. Anal. **121**(2), 330–388 (1994)

[SST16] P. Sarnak, S.-W. Shin, N. Templier, Families of L-functions and their symmetry (2014). These proceedings

[ST15] S.W. Shin, N. Templier, Sato–Tate theorem for families and low-lying zeros of automorphic L-functions. Invent. Math. **203**(1), 1–177 (2016)

Zeta Functions for the Adjoint Action of GL(n) and Density of Residues of Dedekind Zeta Functions

Jasmin Matz

Abstract We define zeta functions for the adjoint action of GL_n on its Lie algebra and study their analytic properties. For $n \leq 3$ we are able to fully analyse these functions. If $n = 2$, we recover the Shintani zeta function for the prehomogeneous vector space of binary quadratic forms. Our construction naturally yields a regularisation, which is necessary to improve the analytic properties of these zeta function, in particular for the analytic continuation if $n \geq 3$.

We further obtain upper and lower bounds on the mean value $X^{-\frac{5}{2}} \sum_E \mathrm{res}_{s=1} \zeta_E(s)$ as $X \to \infty$, where E runs over totally real cubic number fields whose second successive minimum of the trace form on its ring of integers is bounded by X. To prove the upper bound we use our new zeta function for GL_3. These asymptotic bounds are a first step towards a generalisation of density results obtained by Datskovsky in case of quadratic field extensions.

1 Introduction

The purpose of this paper is twofold. First of all, we want to provide another point of view for the construction of the Shintani zeta function $Z(s, \Psi)$ associated with the space of binary quadratic forms [Shi75, Yuk92], and we want to generalise this approach to higher dimensions, namely, to the action of $\mathrm{GL}_1 \times \mathrm{GL}_n$ on the Lie algebra $\mathfrak{gl}_n$. The analytic properties of the zeta function $Z(s, \Psi)$ are unsatisfactory but it can be "adjusted" (cf. [Yuk92, Dat96]) to satisfy a simple functional equation with only finitely many poles. The advantage of our approach is that a suitable modification naturally emerges (for $Z(s, \Psi)$ as well as for the higher dimensional case).

The second purpose of this paper is to make a first step towards the generalisation of a result from [Dat96] to higher dimensions: We prove upper and lower bounds on the density of residues of Dedekind zeta functions attached to totally real cubic number fields. For the upper bound we use our new zeta function for $n = 3$.

J. Matz (✉)
Mathematisches Institut, Universität Leipzig, Postfach 100920, 04009 Leipzig, Germany
e-mail: matz@math.uni-leipzig.de

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_10

Our construction of the zeta functions has also close connections with the Arthur–Selberg trace formula for the group $\mathrm{GL}(n)$. We shall comment further on this below and in Sect. 6.1.

An interesting class of zeta functions, namely the Shintani zeta functions, can be constructed from prehomogeneous vector spaces, cf. [SS74, Shi75, Yuk92, Kim03]. One fundamental example of a prehomogeneous vector space is the space of binary quadratic forms with rational coefficients together with the group $\mathrm{GL}_1 \times \mathrm{GL}_2$ acting on this space by multiplication by scalars and by changing basis, respectively. One can associate a zeta function $Z(s, \Psi)$ with this space as in [Shi75, Yuk92]. There are two natural generalisations of this space to higher dimensions corresponding to different interpretations: From the point of view of quadratic forms, the obvious generalisation is to consider $\mathrm{GL}_1 \times \mathrm{GL}_n$ acting on quadratic forms in n variables. This is again a prehomogeneous vector space which was studied in [Shi75, Suz79], for example.

On the other hand, we can equally well identify the space of binary quadratic forms with the Lie algebra $\mathfrak{sl}_2$ of SL_2 so that the action of GL_2 becomes the adjoint representation on $\mathfrak{sl}_2$. From this point of view, it is more natural to generalise to higher dimensions by considering the action of $\mathrm{GL}_1 \times \mathrm{GL}_n$ on $\mathfrak{sl}_n$ (or $\mathfrak{gl}_n = \mathfrak{sl}_n \oplus \mathfrak{gl}_1$) by letting GL_1 act by multiplication by scalars and GL_n by the adjoint action. This is the point of view we take in this paper. However, this is not a prehomogeneous vector space for $n \geq 3$ so that the general theory of Shintani zeta functions does not apply.

One reason to study such zeta functions is that in many cases the Dirichlet coefficients of these functions contain information on certain arithmetic quantities which can often be studied with Tauberian theory, see, for example, [Shi75, DW88, WY92, Dat96, Bha05, Bha10, TT13, BST13]. For example, the Shintani zeta function $Z(s, \Phi)$ introduced above can be used to deduce density theorems for class numbers of binary quadratic forms as well as for residues of Dedekind zeta functions for quadratic number fields, cf. [Shi75, Dat96].

We shall see that one can find the residues of Dedekind zeta functions of certain field extensions over $\mathbb{Q}$ in the Dirichlet coefficients of the zeta function we are going to define. Although the underlying structure of our space is not prehomogeneous in general, we can still extract some information from our zeta function, at least in the cubic case.

The paper consists of two main parts. The second part applies the results from the first part but is otherwise independent from it.

We are now going to describe our results in some more detail and try to give a guide for reading the paper. Let $n \geq 2$, $G = \mathrm{GL}_n$, and let $\mathfrak{g} = \mathfrak{gl}_n$ be the Lie algebra of G. Then G acts on $\mathfrak{g}$ by the adjoint action Ad. Let $\mathbb{A}$ denote the ring of adeles of $\mathbb{Q}$, $|\cdot|_{\mathbb{A}}$ the usual absolute norm on $\mathbb{A}^\times$, and $G(\mathbb{A})^1 = \{g \in G(\mathbb{A}) \mid |\det g|_{\mathbb{A}} = 1\}$.

Part 1

1.1 Definition of the Zeta Function

We generalise the zeta function $Z(s, \Psi)$ (we will recall the definition of the Shintani zeta function in Sect. 6.3) to higher dimensions by defining the *main* (or unregularised) zeta function for G by

$$\Xi_{\text{main}}(s, \Phi) = \int_0^\infty \lambda^{n(s+\frac{n-1}{2})} \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \sum_{[X]} \Phi(\lambda \operatorname{Ad} x^{-1} X)\, dx\, d^\times \lambda$$

for $s \in \mathbb{C}$, $\Re s \gg 0$, and $\Phi : \mathfrak{g}(\mathbb{A}) \longrightarrow \mathbb{C}$ a Schwartz–Bruhat function. Here the sum inside the integral runs over all $\operatorname{Ad} G(\mathbb{Q})$-equivalence classes $[X]$ of regular elliptic elements in $\mathfrak{g}(\mathbb{Q})$, that is, elements which have an irreducible characteristic polynomial over $\mathbb{Q}$. This defines a holomorphic function for $\Re s > \frac{n+1}{2}$, cf. Theorem 1.1 below. For $n = 2$ the function $\Xi_{\text{main}}(\cdot, \Phi)$ basically coincides with the (unmodified) Shintani zeta function $Z(s, \Psi)$ from [Shi75, Yuk92, Dat96] (cf. Sect. 6.3 and [Mat11]) for Ψ constructed from Φ in a certain way.

To study $\Xi_{\text{main}}(\cdot, \Phi)$ one needs to regularise it in a suitable way. For $n = 2$ a regularisation is needed to obtain a "nice" functional equation and only finitely many poles (cf. [Yuk92, Dat96] for $Z(s, \Psi)$), but for higher dimensions, the regularisation appears to be even more essential: Already for $n = 3$, it seems that $\Xi_{\text{main}}(\cdot, \Phi)$ cannot be continued to all of $\mathbb{C}$, cf. [Mat11, IV.iii]. Our method of regularisation is different from the previously used methods for $Z(s, \Phi)$: In [Yuk92, Dat96] smoothed Eisenstein series were used to cut off diverging integrals. In contrast to this we use a more geometric truncation process that is analogous to the one employed by Arthur for his trace formula in the group case; cf. also [Lev99] for a similar truncation for the Shintani zeta function attached to the space of binary quartic forms. To perform this truncation we use Chaudouard's trace formula for $\mathfrak{g}$ (= truncated summation formula) from [Cha02]: Let $\mathcal{O}$ denote the set of geometric equivalence classes on $\mathfrak{g}(\mathbb{Q})$. This set corresponds bijectively to $\operatorname{Ad} G(\mathbb{Q})$-orbits of semisimple elements, cf. Sect. 2.4. Let $\mathfrak{n} \in \mathcal{O}$ be the nilpotent variety in $\mathfrak{g}$. One can attach to every $\mathfrak{o} \in \mathcal{O}$ and to every truncation parameter T in the coroot space $\mathfrak{a}$ of G a distribution $J_{\mathfrak{o}}^T$ on the space of Schwartz–Bruhat functions $\Phi : \mathfrak{g}(\mathbb{A}) \longrightarrow \mathbb{C}$, cf. Sect. 2.7. They are defined similar to Arthur's distributions on the space of test functions on a reductive algebraic group appearing in Arthur's trace formula. We now define the regularised zeta function $\Xi^T(s, \Phi)$ as follows: If $\lambda \in \mathbb{R}_{>0}$, set $\Phi_\lambda(x) = \Phi(\lambda x)$. Then

$$\Xi^T(s, \Phi) := \int_0^\infty \lambda^{n(s+\frac{n-1}{2})} \sum_{\mathfrak{o}\in\mathcal{O},\ \mathfrak{o}\neq\mathfrak{n}} J_{\mathfrak{o}}^T(\Phi_\lambda)\, d^\times \lambda, \tag{1}$$

provided this integral converges. For later applications in the second part of the paper we need to extend this definition to a certain class $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$, $0 < \nu \le \infty$, of not necessarily smooth test functions, cf. Sect. 2.5. This extension to non-smooth functions is important for later applications in Part 2. The function $\Xi_{\mathrm{main}}(\cdot, \Phi)$ corresponds to the partial sum over such $\mathfrak{o} \in \mathcal{O}$ which are attached to orbits of regular elliptic elements in the definition of $\Xi^T(s, \Phi)$.

1.2 Relation to Arthur's Trace Formula and Automorphic L-Functions

The function $\Xi^T(\cdot, \Phi)$ is closely related to Arthur's trace formula for G as $\Xi^T(\cdot, \Phi)$ "contains" the geometric side of Arthur's trace formula for a certain non-standard test function, cf. Sect. 6. $\Xi^T(s, \Phi)$ therefore also "contains" the spectral side of Arthur's trace formula. The discrete spectrum, contributing to the spectral side, therefore also contributes to $\Xi^T(s, \Phi)$. Choosing a suitable non-standard test function the contribution from the discrete spectrum to $\Xi^T(s, \Phi)$ in fact equals

$$\sum_{\pi} L^*(s, \pi)$$

where the sum runs over all unramified automorphic representations of $G(\mathbb{A})^1$ appearing in $L^2_{\mathrm{disc}}(G(\mathbb{Q})\backslash G(\mathbb{A})^1)$, and $L^*(s, \pi)$ is the standard L-function of π with a suitable completion at ∞. This follows from the construction in [GJ72]. Such sums of L-functions play a central role in the theory of "Beyond Endoscopy" (cf. [Lan04]). That these sums show up as a "part" of our zeta functions reflects the fact that the Lie algebra $\mathfrak{g}$ is Vinberg's universal monoid for $G = \mathrm{GL}(n)$, cf. [Ngô14].

1.3 Analytic Properties of $\Xi^T(s, \Phi)$

Our first main result is the following:

Theorem 1.1 (cf. Theorem 3.4). *Let $n \geq 2$. There exists $\nu \in (0, \infty)$ depending only on n such that for every $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ the following holds:*

(i) If T is sufficiently regular, the integral defining $\Xi^T(s, \Phi)$ converges absolutely and locally uniformly for $\Re s > \frac{n+1}{2}$. In particular, $\Xi^T(s, \Phi)$ is holomorphic in this half plane.

(ii) $\Xi^T(s, \Phi)$ is a polynomial in T of degree at most $\dim \mathfrak{a} = n-1$ and can be defined for every $T \in \mathfrak{a}$. Then for every T the function $\Xi^T(s, \Phi)$ is holomorphic in $\Re s > \frac{n+1}{2}$.

Here $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ for $\nu \in (0, \infty]$ is a generalisation of the space of Schwartz–Bruhat functions on $\mathfrak{g}(\mathbb{A})$, see Sect. 2.5 for the definition. In fact, if $\nu = \infty$,

$\mathcal{S}(\mathfrak{g}(\mathbb{A})) = \mathcal{S}^{\infty}(\mathfrak{g}(\mathbb{A}))$ is equal to the usual space of Schwartz–Bruhat functions. If ν is finite, elements of $\mathcal{S}^{\nu}(\mathfrak{g}(\mathbb{A}))$ are in general only differentiable up to order ν and satisfy the same conditions of a Schwartz–Bruhat function but only up to this order.

In this way we get a well-defined family $\Xi^T(s, \Phi)$ of zeta functions indexed by the parameter $T \in \mathfrak{a}$ and varying continuously with T. By the nature of our construction this family depends on an initial choice of minimal parabolic subgroup in G. We can, however, choose a zeta function in this family which is independent of this choice: Taking $T = 0$, the function $\Xi^0(s, \Phi)$ does not depend on the fixed minimal parabolic subgroup but only on the fixed maximal compact subgroup and maximal split torus (cf. [Art81, Lemma 1.1]).

One of the standard methods to get the meromorphic continuation and functional equation of zeta functions is to use the Poisson summation formula. In our context, Chaudouard's trace formula takes the place of the Poisson summation formula, and the main obstruction to obtain the meromorphic continuation and the functional equation for $\Xi^T(s, \Phi)$ is to understand the nilpotent contribution $J_{\mathfrak{n}}^T(\Phi_\lambda)$. Restricting to $n \leq 3$, we are able to analyse the nilpotent distribution $J_{\mathfrak{n}}^T(\Phi_\lambda)$ completely (see Sects. 4 and 5), obtaining our main result of Part 1:

Theorem 1.2 (cf. Theorem 5.7). *Let $G = \mathrm{GL}_n$ with $n \leq 3$, and let $R > n$ be given. Then there exists $\nu \in (0, \infty)$ such that for every $\Phi \in \mathcal{S}^{\nu}(\mathfrak{g}(\mathbb{A}))$ and $T \in \mathfrak{a}$ the following holds.*

(i) $\Xi^T(s, \Phi)$ *has a meromorphic continuation to all $s \in \mathbb{C}$ with $\Re s > -R$, and satisfies for such s the functional equation*

$$\Xi^T(s, \Phi) = \Xi^T(1 - s, \hat{\Phi}).$$

(ii) The poles of $\Xi^T(s, \Phi)$ in $\Re s > -R$ are parametrised by the nilpotent orbits $\mathcal{N} \subseteq \mathfrak{g}(\mathbb{Q})$. More precisely, its poles occur exactly at the points

$$s_{\mathcal{N}}^- = \frac{1-n}{2} + \frac{\dim \mathcal{N}}{2n} \quad \textit{and} \quad s_{\mathcal{N}}^+ = \frac{1+n}{2} - \frac{\dim \mathcal{N}}{2n}$$

and are of order at most $\dim \mathfrak{a} + 1 = n$. In particular, the furthermost right and furthermost left pole in this region are both simple, correspond to $\mathcal{N} = 0$, and are located at the points $s_0^+ = \frac{1+n}{2}$ and $s_0^- = \frac{1-n}{2}$, respectively. The residues at these poles are given by

$$\operatorname*{res}_{s=s_0^-} \Xi^T(s, \Phi) = \mathrm{vol}(G(\mathbb{Q})\backslash G(\mathbb{A})^1)\Phi(0), \quad \textit{and}$$

$$\operatorname*{res}_{s=s_0^+} \Xi^T(s, \Phi) = \mathrm{vol}(G(\mathbb{Q})\backslash G(\mathbb{A})^1) \int_{\mathfrak{g}(\mathbb{A})} \Phi(X)\, dX.$$

Remarks 1.3.

(i) The inconvenient way in which the first part of the theorem is stated is due to the fact that the space of functions $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ is not closed under Fourier transform if $\nu < \infty$, cf. the definition of the space $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ in Sect. 2.5.
(ii) If $\nu = \infty$, then Φ is a Schwartz–Bruhat function and $\Xi^T(s, \Phi)$ can be meromorphically continued to all of $\mathbb{C}$.
(iii) For applications in the second part of the paper we indeed need to be able to choose test functions in $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ with $\nu < \infty$. This is because for the proof of Theorem 1.5 we use test functions which are not smooth but only differentiable up to a certain order.
(iv) Similar results as Theorem 1.1 and Theorem 1.2 hold if we replace GL_n by SL_n and $\mathfrak{gl}_n$ by $\mathfrak{sl}_n$. The region of absolute convergence then has to be adjusted to $\Re s > (n+1)/2 - 1/n$, and the locations of the poles have to be adjusted accordingly.

Chaudouard's trace formula is valid for any reductive group. In principle, it should be possible to define the zeta function $\Xi^T(s, \Phi)$ as in (1) for G an arbitrary connected reductive group acting on its Lie algebra. At least Theorem 1.1 should stay true for more general $\mathbb{Q}$-split reductive groups; we restricted to GL_n mainly to make it not more technical as it already is. The main obstruction for extending Theorem 1.2 to $n > 3$ (or more general groups) lies in understanding the nilpotent distribution $J_{\mathfrak{n}}^T(\Phi_\lambda)$, cf. Sects. 4, 5, and Appendix 1. For $n = 2, 3$ the structure of the decomposition of the nilpotent variety into nilpotent orbits gives rise to the functional equation and the position of the poles of the zeta function. This is expected to be the case also for $n > 3$. For $n > 3$ there is a different approach to obtain the meromorphic continuation of $\Xi^T(s, \Phi)$ than we present here, cf. Remark 5.4 and Example 5.10. The advantage of our approach is that it also gives the full principal parts of the Laurent expansion at all poles from the knowledge of certain polynomials (in T) $J_{\mathcal{N}}^T(\Phi)$ attached to the nilpotent orbits $\mathcal{N}$.

One could take this approach even further, by considering a general rational representation of the group instead of its adjoint representation. In [Lev01] equivalence classes $\mathfrak{o}$ and corresponding distributions $J_{\mathfrak{o}}^T(\Phi)$ are defined for such a representation, and also a kind of "trace formula" is proved for this situation. For the Shintani zeta function of binary quartic forms such an approach has been carried out in [Lev99].

For $G = \mathrm{GL}_2$ and $G = \mathrm{GL}_3$, we can show that $\Xi_{\mathrm{main}}(s, \Phi)$ is indeed the main part of $\Xi^T(s, \Phi)$ in the following sense:

Proposition 1.4 (cf. Corollaries 7.3 and 7.5). *If $G = \mathrm{GL}_2$ or $G = \mathrm{GL}_3$, then $\Xi^T(s, \Phi) - \Xi_{main}(s, \Phi)$ continues holomorphically at least to $\Re s > \frac{n}{2}$. In particular, the furthermost right pole of $\Xi^T(s, \Phi)$ and $\Xi_{main}(s, \Phi)$ coincide and have the same residue.*

A similar result should of course also hold for $n > 3$. This result will become important in Part 2, where we will use the analytic properties of $\Xi_{\mathrm{main}}(s, \Phi)$ to apply a Tauberian theorem in order to obtain information on the Dirichlet coefficients of $\Xi_{\mathrm{main}}(s, \Phi)$ for $n = 3$ in which case they are related to geometric properties of cubic number fields.

Part 2

1.4 Density of Residues

A main application of the Shintani zeta function $Z(s, \Psi)$, which is attached to the space of binary quadratic forms, is to prove the mean value behaviour of the class numbers of binary quadratic forms [Shi75]. From our point of view, another closely related density result obtained from $Z(s, \Psi)$ is of more interest to us: Datskovsky [Dat96] proved that if S is a finite set of prime places of $\mathbb{Q}$ including the Archimedean place, and $r_S = (r_v)_{v\in S}$ is a fixed signature for quadratic number fields, then as $X \to \infty$ one has

$$\sum_L \operatorname*{res}_{s=1} \zeta_L(s) = \alpha(r_S)X, \tag{2}$$

where L runs over all quadratic fields of signature r_S at the places in S and absolute discriminant D_L less than or equal to X, and $\alpha(r_S)$ is a suitable non-zero constant. Here $\zeta_L(s)$ is the Dedekind zeta function attached to L. As a first step towards generalising this, we prove upper and lower bounds for the densities of residues of Dedekind zeta functions of totally real cubic number fields.

Suppose E is a totally real number field of degree n with ring of integers $\mathcal{O}_E \subseteq E$. We denote by $Q_E : \mathcal{O}_E/\mathbb{Z} \longrightarrow \mathbb{R}$ the positive definite quadratic form $Q_E(\xi) = \mathrm{tr}_{E/\mathbb{Q}}\,\xi^2 - \frac{1}{n}(\mathrm{tr}_{E/\mathbb{Q}}\,\xi)^2$ for $\xi \in \mathcal{O}_E/\mathbb{Z}$, where $\mathrm{tr}_{E/\mathbb{Q}} : E \longrightarrow \mathbb{Q}$ denotes the field trace of $E/\mathbb{Q}$. We denote the successive minima of Q_E on $\mathcal{O}_E/\mathbb{Z}$ by $m_1(E) \leq m_2(E) \leq \ldots \leq m_{n-1}(E)$. If $n = 2$, then $m_1(L) = D_L/2$ for every quadratic field L so that the sum in (2) runs over all quadratic fields with $m_1(E) \leq X/2$. Our main result of Part 2 is the following:

Theorem 1.5 (cf. Theorem 10.1). *We have*

$$\limsup_{X\to\infty} X^{-\frac{5}{2}} \sum_{E:\, m_1(E)\leq X} \operatorname*{res}_{s=1} \zeta_E(s) < \infty \tag{3}$$

where the sum extends over all totally real cubic number fields E for which the first successive minimum $m_1(E)$ is bounded by X.

We complement the above upper bound (3) with the following result:

Proposition 1.6 (cf. Proposition 10.3). *For every $\varepsilon > 0$, we have*

$$\liminf_{X\to\infty} X^{-\frac{5}{2}+\varepsilon} \sum_{E:\, m_1(E)\leq X} \operatorname*{res}_{s=1} \zeta_E(s) = \infty,$$

where the sum extends over totally real cubic number fields E.

This is a first step towards a generalisation of (2) to the cubic case with $S = \{\infty\}$ and the Archimedean signature of totally real cubic number fields. As in the quadratic case, one expects that in fact the limit of the left-hand side in (3) exists and is non-zero:

Conjecture 1.7. There exists a constant $\alpha_3 > 0$ such that as $X \to \infty$

$$\sum_{E\colon m_1(E)\leq X} \operatorname*{res}_{s=1} \zeta_E(s) \sim \alpha_3 X^{\frac{5}{2}},$$

where the sum extends over all totally real cubic number fields E for which the first successive minimum $m_1(E)$ is bounded by X.

The strategies to prove Theorem 1.5 and Proposition 1.6 are quite different from each other: For the first result we use a suitable sequence of test functions and apply a Tauberian Theorem to $\Xi_{\mathrm{main}}(s,\Phi)$ to obtain an asymptotic for the density of certain orbital integrals in Proposition 9.2. These orbital integrals are basically products of $\operatorname{res}_{s=1}\zeta_E(s)$ and a certain quantity $c(\xi,\Phi_f)$, $\xi \in E$, obtained from the non-Archimedean part Φ_f of the test function Φ. For an appropriate Φ_f we can show that $c(\xi,\Phi_f) \geq 1$ for every relevant ξ so that Theorem 1.5 is a direct consequence of Proposition 9.2. To prove Proposition 1.6, on the other hand, we use a different approach (independent of our results for $\Xi^T(s,\Phi)$): We basically show that there are sufficiently many irreducible cubic polynomials.

In principle, we would like to deduce the full conjectured asymptotic from Proposition 9.2, that is, from the properties of $\Xi_{\mathrm{main}}(s,\Phi)$. This would indeed follow if we would be allowed to replace the coefficients $c(\xi,\Phi_f)$ by 1. In Appendix 2 we give a sequence of test functions $(\Phi_f^{\mathfrak{m}})_{\mathfrak{m}}$ for which $c(\xi,\Phi_f^{\mathfrak{m}}) \to 1$. However, a certain uniformity of the convergence with respect to $Q_E(\xi)$ is needed to prove Conjecture 1.7. We were not able to do this so far.

Our methods can at least heuristically be applied to GL_n for every $n \geq 2$. In particular, the first pole of $\Xi^T(s,\Phi)$ as well as $\Xi_{\mathrm{main}}(s,\Phi)$ for GL_n should be at $s = \frac{n+1}{2}$. This suggests

Conjecture 1.8. For every $n \geq 3$ there exists $\alpha_n > 0$ such that as $X \to \infty$

$$\sum_{E\colon m_1(E)\leq X} \operatorname*{res}_{s=1} \zeta_E(s) \sim \alpha_n X^{\frac{n(n+1)-2}{4}},$$

where the sum extends over all totally real n-dimensional number fields E for which the first successive minimum $m_1(E)$ is bounded by X.

Ordering fields with respect to the first successive minimum of Q_E (in contrast to the discriminant) is also related to a conjecture of Ellenberg–Venkatesh, cf. [EV06, Remark 3.3]: Basically they conjecture that $X^{-\frac{n(n+1)-2}{4}} \sum_{E\colon m_1(E)\leq X} 1$ has a non-zero limit as $X \to \infty$ where E runs over n-dimensional number fields. As remarked in [EV06], it is possible to show a "weak form" of this asymptotic under a strong hypothesis on the existence of sufficiently many squarefree polynomials. If one could prove an n-dimensional analogue of Proposition 9.2 and make the passage from $c(\xi, \Phi_f)$ to 1 work (e.g., with a sequence of test function as $(\Phi_f^{\mathfrak{m}})_{\mathfrak{m}}$), this should lead to another approach to (a slightly weaker form of) the conjecture of Ellenberg–Venkatesh.

This second part of the paper is organised as follows: In Sect. 8 we first recall and prove some properties of orbital integrals, before stating and proving an asymptotic for the mean value of certain orbital integrals in Sect. 9, cf. Proposition 9.2. Our main result Theorem 1.5 in Sect. 10 will then be an easy consequence of Proposition 9.2 together with results in Sect. 8. Finally, we will prove Proposition 1.6 at the end of Sect. 10.

2 Notation and General Conventions

2.1 General Notation

We fix notation, mainly following [Cha02, Art05]:

- $\mathbb{A}$ denotes the ring of adeles of $\mathbb{Q}$. If v is a place of $\mathbb{Q}$, $\mathbb{Q}_v$ denotes the completion of $\mathbb{Q}$ at v, $|\cdot|_v$ is the usual v-adic norm on $\mathbb{Q}_v$ so that if $v = p$ is a non-Archimedean place, we have $|p|_p = p^{-1}$. Then $|\cdot|_{\mathbb{A}}$ denotes the norm on $\mathbb{A}^\times$ given by the product of the $|\cdot|_v$'s. If it is clear from the context, we may also write $|\cdot|$ for $|\cdot|_{\mathbb{A}}$ or $|\cdot|_v$.
- $n \geq 2$ is an integer, and G denotes GL_n as a group defined over $\mathbb{Q}$ with Lie algebra $\mathfrak{g} = \mathfrak{gl}_n$. $\mathbf{1}_n \in G$ denotes the identity element.
- $P_0 = T_0 U_0$ is the minimal parabolic subgroup of upper triangular matrices with T_0 the torus of diagonal elements and U_0 its unipotent radical of upper triangular matrices. If $P \supseteq T_0$ is a $\mathbb{Q}$-defined parabolic subgroup with Levi component $M = M_P \supseteq T_0$, then $\mathcal{F}(M)$ denotes the set of ($\mathbb{Q}$-defined) parabolic subgroups containing M, and $\mathcal{P}(M) \subseteq \mathcal{F}(M)$ the subset of parabolic subgroups with Levi component M. For $P \in \mathcal{F}(T_0)$ with Levi decomposition $P = M_P U_P$, we denote by $\mathfrak{p} = \mathfrak{m}_P + \mathfrak{u}_P$ the corresponding decomposition of the Lie algebra. For $P_1, P_2 \in \mathcal{F}(T_0)$ with $P_1 \subseteq P_2$, put $\mathfrak{u}_{P_1}^{P_2} := \mathfrak{u}_1^2 := \mathfrak{u}_{P_1} \cap \mathfrak{m}_{P_2}$ and $\overline{\mathfrak{u}}_{P_1}^{P_2} := \overline{\mathfrak{u}}_{P_1} \cap \mathfrak{m}_{P_2} := \mathfrak{u}_{\bar{P}_1} \cap \mathfrak{m}_{P_1}$ for $\bar{P}_1 \in \mathcal{P}(M_{P_1})$ the opposite parabolic subgroup. $A_M \subseteq M(\mathbb{R})$ denotes the identity component of the split component of the center in $M(\mathbb{A})$. We usually identify the groups $M(\mathbb{A})^1$ and $A_M \backslash M(\mathbb{A})$.
- $P \in \mathcal{F}(T_0)$ is called standard if $P_0 \subseteq P$ and we write $\mathcal{F}_{\mathrm{std}} \subseteq \mathcal{F}(T_0)$ for the set of standard parabolic subgroups.
- $\mathfrak{a}_P^*$ is the root space, i.e. the $\mathbb{R}$-vector space spanned by all rational characters $M_P \longrightarrow \mathrm{GL}_1$, and $\mathfrak{a}_P = \mathfrak{a}_{M_P} = \mathrm{Hom}_{\mathbb{R}}(\mathfrak{a}_P^*, \mathbb{R})$ is the coroot-space. Σ_P denotes the set of reduced roots of the pair (A_{M_P}, U_P)

  We denote by $\Delta_{P_1}^{P_2} = \Delta_1^2$ the set of simple roots and by $\Sigma_{P_1}^{P_2} = \Sigma_1^2$ the set of all positive roots of the action of $A_1 = A_{P_1}$ on $U_1 \cap M_2$. If $\alpha \in \Delta_1^2$, then $\alpha^\vee$ denotes the corresponding coroot. Similarly, $\widehat{\Delta}_{P_1}^{P_2} = \widehat{\Delta}_1^2$ is the set of simple weights, and if $\varpi \in \widehat{\Delta}_1^2$, then $\varpi^\vee$ denotes the corresponding coweight. If $\alpha \in \Delta_1^2$, we denote by $\varpi_\alpha \in \widehat{\Delta}_1^2$ the weight such that $\varpi_\alpha(\beta^\vee) = \delta_{\alpha\beta}$ for all $\beta \in \Delta_1^2$ (here $\delta_{\alpha\beta}$ is the Kronecker δ).

- If $a \in A_P$ and $\lambda \in \mathfrak{a}_P^*$, write $\lambda(a) = e^{\lambda(H_P(a))}$. For $P_1 \subseteq P_2$, let

$$A_{P_1}^{P_2} = A_1^2 = \{a \in A_{P_1} \mid \forall \alpha \in \Delta_{P_2} : \alpha(a) = 1\} \simeq A_{P_1}/A_{P_2},$$

  and $\mathfrak{a}_{P_1}^{P_2} = \log A_1^2 \subseteq \mathfrak{a}_{P_1}$. Set $\mathfrak{a} = \mathfrak{a}_0^G$. For $M \subseteq G$ let $M(\mathbb{A})^1$ be the intersection of the kernels of all rational characters $M(\mathbb{A}) \longrightarrow \mathbb{C}$. Let $\mathfrak{a}_0^+ = \{H \in \mathfrak{a}_0 \mid \forall \alpha \in \Delta_0 : \alpha(H) > 0\}$ be the positive chamber in $\mathfrak{a}_0$ with respect to our fixed minimal parabolic subgroup. Similarly, we define $\mathfrak{a}^+$. Denote by $\rho_1^2 = \rho_{P_1}^{P_2} \in \mathfrak{a}_0^+$ the unique element in $\mathfrak{a}_0^+$ such that the modulus function of $M_{P_1}(\mathbb{A})$ on $\mathfrak{u}_{P_1}^{P_2}(\mathbb{A})$ satisfies $\delta_1^2(m) := \delta_{P_1}^{P_2}(m) := |\det \operatorname{Ad} m_{|\mathfrak{u}_{P_1}^{P_2}(\mathbb{A})}| = e^{2\rho_1^2(H_0(m))}$ for all $m \in M_{P_1}(\mathbb{A})$, and we write $\rho_1 = \rho_{P_1} = \rho_{P_1}^G$ and $\delta_0 = \delta_{P_0}^G$.
- Let $H_P = H_{M_P} : G(\mathbb{A}) = M(\mathbb{A})U(\mathbb{A})\mathbf{K} \longrightarrow \mathfrak{a}_P$ be the map characterised by $H_P(muk) = H_P(m)$ and $H_P(\exp H) = H$ for all $H \in \mathfrak{a}_P$.
- We denote by $\Phi(A_0, M_R)$ the set of weights of A_0 with respect to M_R so that $\Phi(A_0, M_R) = \Sigma_0^R \cup \{0\} \cup \left(-\Sigma_0^R\right)$. Then we have a direct sum decomposition $\mathfrak{g} = \bigoplus_{\beta \in \Phi(A_0, M_R)} \mathfrak{g}_\beta$ for $\mathfrak{g}_\beta$ the eigenspace of β in $\mathfrak{g}$. We take the usual vector norm $\|\cdot\|_{\mathbb{A}} = \|\cdot\|$ on $\mathfrak{g}(\mathbb{A})$ obtained by identifying $\mathfrak{g}(\mathbb{A})$ with $\mathbb{A}^{n^2}$ via the matrix coordinates. Then if $X \in \mathfrak{g}(\mathbb{A})$, $X = \sum_{\beta \in \Phi(A_0, M_R)} X_\beta$ with $X_\beta \in \mathfrak{g}_\beta(\mathbb{A})$, then $\|X\| = \sum_{\beta \in \Phi(A_0, M_R)} \|X_\beta\|$.
- If $M = T_0$, we write $\mathcal{F} = \mathcal{F}(T_0)$, $H_0 = H_{M_0}$, $\mathfrak{a}_0 = \mathfrak{a}_{M_0}$, etc., and further put $\mathfrak{a} = \mathfrak{a}_0^G$ and $\mathfrak{a}^+ = \left(\mathfrak{a}_0^G\right)^+$.

2.2 Characteristic Functions

Let $P_1, P_2, P \in \mathcal{F}$ be parabolic subgroups with $P_1 \subseteq P_2$. We define the following functions (cf. [Art78]):

- $\hat{\tau}_{P_1}^{P_2} = \hat{\tau}_1^2 : \mathfrak{a}_0 \longrightarrow \mathbb{C}$ is the characteristic function of the set

$$\{H \in \mathfrak{a}_0 \mid \forall \varpi \in \hat{\Delta}_1^2 : \varpi(H) > 0\}.$$

  If $P_2 = G$, we also write $\hat{\tau}_{P_1} = \hat{\tau}_1 = \hat{\tau}_1^G$.
- $\tau_{P_1}^{P_2} = \tau_1^2 : \mathfrak{a}_0 \longrightarrow \mathbb{C}$ is the characteristic function of the set

$$\{H \in \mathfrak{a}_0 \mid \forall \alpha \in \Delta_1^2 : \alpha(H) > 0\}.$$

  If $P_2 = G$, we also write $\tau_{P_1} = \tau_1 = \tau_1^G$.

- $\sigma_{P_1}^{P_2} = \sigma_1^2 : \mathfrak{a}_0 \longrightarrow \mathbb{C}$ is the characteristic function of the set

$$\{H \in \mathfrak{a}_0 \mid \forall \alpha \in \Delta_1^2 : \alpha(H) > 0;\ \forall \alpha \in \Delta_1 \backslash \Delta_1^2 : \alpha(H) \leq 0;\ \forall \varpi \in \hat{\Delta}_2 : \varpi(H) > 0\}.$$

Remark 2.1. The function σ_1^2 is related to τ_1^2 and $\hat{\tau}_1^2$ by $\sigma_1^2 = \sum_{R:\ P_2 \subseteq R} (-1)^{\dim \mathfrak{a}_2^R} \tau_1^R \hat{\tau}_R$.

- $T \in \mathfrak{a}^+$ is called *sufficiently regular* if $d(T) := \min_{\alpha \in \Delta_0} \alpha(T)$ is sufficiently large, i.e., if T is sufficiently far away from the walls of the positive Weyl chamber (cf. [Art78]). We fix a small number $\delta > 0$ such that the set of sufficiently regular $T \in \mathfrak{a}$ satisfying $d(T) > \delta \|T\|$ is a non-empty open cone in $\mathfrak{a}^+$.
- For sufficiently regular $T \in \mathfrak{a}^+$ the function $F^P(\cdot, T) : G(\mathbb{A}) \longrightarrow \mathbb{C}$ is defined as the characteristic function of all $x = umk \in G(\mathbb{A}) = U(\mathbb{A})M(\mathbb{A})\mathbf{K}$, $P = MU$, satisfying

$$\varpi(H_0(\mu m) - T) \leq 0$$

  for all $\mu \in M(\mathbb{Q})$ and $\varpi \in \widehat{\Delta}_0^M$. If $P = G$, we sometimes write $F(\cdot, T) = F^G(\cdot, T)$.
- If $T \in \mathfrak{a}^+$ is sufficiently regular, [Art78, Lemma 6.4] gives for every $x \in G(\mathbb{A})$ the identity

$$\sum_{R:\ P_0 \subseteq R \subseteq P} \sum_{\delta \in R(\mathbb{Q}) \backslash P(\mathbb{Q})} F^R(\delta x, T) \tau_R^P(H_0(\delta x) - T) = 1.$$

2.3 Measures

We fix the following maximal compact subgroups: If v is a non-Archimedean place, then $\mathbf{K}_v = G(\mathbb{Z}_v)$. If v is Archimedean, we take $\mathbf{K}_\infty = \mathrm{O}(n)$. Globally, we take $\mathbf{K} = \prod_{v \leq \infty} \mathbf{K}_v \subseteq G(\mathbb{A})^1$. Up to normalisation there exists a unique Haar measure on $\mathbf{K}_v$, and we normalise it by $\mathrm{vol}(\mathbf{K}_v) = 1$ for every v, and then take the product measure on $\mathbf{K}$. We further choose measures as follows:

- $\mathbb{Q}_v$ and $\mathbb{Q}_v^\times$, $v < \infty$: normalised by $\mathrm{vol}(\mathbb{Z}_v) = 1 = \mathrm{vol}(\mathbb{Z}_v^\times)$.
- $\mathbb{R}$, $\mathbb{R}^\times$, $\mathbb{R}_{>0}$, A_G, A_0: usual Lebesgue measures.
- $\mathbb{C}$, $\mathbb{C}^\times$: twice the usual Lebesgue measure.
- $\mathbb{A}$ and $\mathbb{A}^\times$: product measures.
- $\mathbb{A}^1 = \{a \in \mathbb{A}^\times \mid |a|_{\mathbb{A}} = 1\}$: measure induced by the exact sequence $1 \longrightarrow \mathbb{A}^1 \hookrightarrow \mathbb{A}^\times \xrightarrow{|\cdot|_{\mathbb{A}}} \mathbb{R}_{>0} \longrightarrow 1$.
- V finite dimensional $\mathbb{Q}$-vector space with fixed basis: take the measures induced from $\mathbb{A}$ (resp. $\mathbb{Q}_v$) on $V(\mathbb{A})$ (resp. $V(\mathbb{Q}_v)$) via the isomorphism $V(\mathbb{A}) \simeq \mathbb{A}^{\dim V}$ (respectively, $V(\mathbb{Q}_v) \simeq \mathbb{Q}_v^{\dim V}$) with respect to this basis. This in particular defines measures on $U_0(\mathbb{A})$ and $U_0(\mathbb{Q}_v)$ if we take the canonical bases corresponding to the root coordinates.

- $T_0(\mathbb{A})$ and $T_0(\mathbb{Q}_v)$: measures induced from $\mathbb{A}^\times$ and $\mathbb{Q}_v^\times$ via the isomorphism $T_0(\mathbb{A}) \simeq (\mathbb{A}^\times)^{n-1}$ (respectively, $T_0(\mathbb{Q}_v) \simeq (\mathbb{Q}_v^\times)^{n-1}$) provided by the diagonal coordinates.
- $G(\mathbb{A})$ and $G(\mathbb{Q}_v)$: compatible with the Iwasawa decomposition $G(\mathbb{A}) = T_0(\mathbb{A})U_0(\mathbb{A})\mathbf{K}$ (resp. $G(\mathbb{Q}_v) = T_0(\mathbb{Q}_v)U_0(\mathbb{Q}_v)\mathbf{K}_v$) such that for every integrable function f on $G(\mathbb{A})$ we have

$$\begin{aligned}\int_{G(\mathbb{A})} f(g)dg &= \int_{T_0(\mathbb{A})}\int_{U_0(\mathbb{A})}\int_{\mathbf{K}} f(tuk)\,dk\,du\,dt\\ &= \int_{T_0(\mathbb{A})}\int_{U_0(\mathbb{A})}\int_{\mathbf{K}} \delta_0(t)^{-1}f(utk)\,dk\,du\,dt\end{aligned}$$

  (similarly for the local case).
- $G(\mathbb{A})^1$: measure induced by the exact sequence $1 \longrightarrow G(\mathbb{A})^1 \hookrightarrow G(\mathbb{A}) \xrightarrow{|\det(\cdot)|_{\mathbb{A}}} \mathbb{R}_{>0} \longrightarrow 1$.
- Levi subgroup $M \supseteq T_0$: compatible with previous cases using that M is isomorphic to a direct product of general linear groups.
- parabolic subgroups $P \in \mathcal{F}(T_0)$: compatible with previous cases by using Iwasawa decomposition for P.

2.4 Equivalence Classes

Let $\mathfrak{g}(\mathbb{Q})_{\mathrm{ss}}$ (resp. $G(\mathbb{Q})_{\mathrm{ss}}$) denote the set of semisimple elements in $\mathfrak{g}(\mathbb{Q})$ (resp. $G(\mathbb{Q})$). We define an equivalence relation on $\mathfrak{g}(\mathbb{Q})$ as follows: Let $X, Y \in \mathfrak{g}(\mathbb{Q})$ and write $X = X_s + X_n$, $Y = Y_s + Y_n$ for the Jordan decomposition with $X_s, Y_s \in \mathfrak{g}(\mathbb{Q})_{\mathrm{ss}}$ semisimple and $X_n \in \mathfrak{g}_{X_s}(\mathbb{Q})$, $Y_n \in \mathfrak{g}_{Y_s}(\mathbb{Q})$ nilpotent, where $\mathfrak{g}_{X_s} = \{Y \in \mathfrak{g} \mid [X_s, Y] = 0\}$ is the centraliser of X_s in $\mathfrak{g}$. We call X and Y equivalent if and only if there exists $\delta \in G(\mathbb{Q})$ such that $Y_s = \operatorname{Ad}\delta^{-1}X_s$. We denote the set of equivalence classes in $\mathfrak{g}(\mathbb{Q})$ by $\mathcal{O}$.

Let $\mathfrak{n} \subseteq \mathfrak{g}(\mathbb{Q})$ denote the set of nilpotent elements. Then $\mathfrak{n} \in \mathcal{O}$ constitutes exactly one equivalence class (corresponding to the orbit of $X_s = 0$), and decomposes into finitely many nilpotent orbits under the adjoint action of $G(\mathbb{Q})$. On the other hand, if $\mathfrak{o} \in \mathcal{O}$ corresponds to the orbit of a regular semisimple element X_s (i.e., the eigenvalues of X_s (in an algebraic closure of $\mathbb{Q}$) are pairwise different), then $\mathfrak{o}$ is in fact equal to the orbit of X_s.

2.5 Test Functions

Let $\mathfrak{b}$ denote the Lie algebra of either one of the standard parabolic subgroups of G, of one of their unipotent radicals, or of one of their Levi components. We fix the standard vector norm $\|\cdot\|$ on $\mathfrak{b}(\mathbb{R})$ by identifying $\mathfrak{b}(\mathbb{R}) \simeq \mathbb{R}^{\dim\mathfrak{b}}$ via the usual matrix coordinates. Let $\mathcal{U}(\mathfrak{b})$ denote the universal enveloping algebra of the complexification $\mathfrak{b}(\mathbb{C})$. For every $\nu \in [0, \infty)$ we fix a basis $\mathcal{B}_\nu = \mathcal{B}_{\mathfrak{b},\nu}$ of the finite dimensional $\mathbb{C}$-vector space $\mathcal{U}(\mathfrak{b})_{\leq \nu}$ of elements in $\mathcal{U}(\mathfrak{b})$ of degree $\leq \nu$. For a real number $a \geq 0$ and a non-negative integer $b \leq \nu$ we define seminorms $\|\cdot\|_{a,b}$ on the spaces $C^\nu(\mathfrak{b}(\mathbb{R}))$ by setting for $f \in C^\nu(\mathfrak{b}(\mathbb{R}))$

$$\|f\|_{a,b} := \sup_{x \in \mathfrak{b}(\mathbb{R})} \left((1 + \|x\|)^a \sum_{X \in \mathcal{B}_b} \left|(Xf)(x)\right| \right)$$

with $(Xf)(x) = \left[\frac{d}{dt} f(xe^{tX})\right]_{t=0}$. We put

$$\mathcal{S}^\nu(\mathfrak{b}(\mathbb{R})) := \{f \in C^\nu(\mathfrak{b}(\mathbb{R})) \mid \forall a < \infty,\ b \leq \nu :\ \|f\|_{a,b} < \infty\},$$

and

$$\mathcal{S}(\mathfrak{b}(\mathbb{R})) = \mathcal{S}^\infty(\mathfrak{b}(\mathbb{R})) := \{f \in C^\infty(\mathfrak{b}(\mathbb{R})) \mid \forall a < \infty,\ b < \infty :\ \|f\|_{a,b} < \infty\}.$$

Then $\mathcal{S}(\mathfrak{b}(\mathbb{R}))$ is the usual space of Schwartz functions on $\mathfrak{b}(\mathbb{R})$. Dualy to $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$ we define for $\nu \leq \infty$

$$\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R})) := \{f \in C^\infty(\mathfrak{b}(\mathbb{R})) \mid \forall a \leq \nu,\ b < \infty :\ \|f\|_{a,b} < \infty\}$$

so that $\mathcal{S}_\infty(\mathfrak{b}(\mathbb{R})) = \mathcal{S}^\infty(\mathfrak{b}(\mathbb{R})) = \mathcal{S}(\mathfrak{b}(\mathbb{R}))$. We define the spaces $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{A}))$ and $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{A}))$ similarly, namely, $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{A})) = \mathcal{S}^\nu(\mathfrak{b}(\mathbb{R})) \otimes \mathcal{S}(\mathfrak{b}(\mathbb{A}_f))$ and $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{A})) = \mathcal{S}_\nu(\mathfrak{b}(\mathbb{R})) \otimes \mathcal{S}(\mathfrak{b}(\mathbb{A}_f))$ where $\mathcal{S}(\mathfrak{b}(\mathbb{A}_f)) = \bigotimes'_{p<\infty} \mathcal{S}(\mathfrak{b}(\mathbb{Q}_p))$ is the usual space of Schwartz Bruhat functions, that is, $\mathcal{S}(\mathfrak{b}(\mathbb{Q}_p))$ is the space of smooth and compactly supported functions $\Phi_p : \mathfrak{b}(\mathbb{Q}_p) \longrightarrow \mathbb{C}$ and the restricted tensor product is taken with respect to the functions Φ_p^0, the characteristic function of $\mathfrak{b}(\mathbb{Z}_p)$. In particular, $\mathcal{S}(\mathfrak{b}(\mathbb{A}))$ is the usual space of Schwartz–Bruhat functions on $\mathfrak{b}(\mathbb{A})$.

The topology induced by the set of seminorms $\|\cdot\|_{a,b}$, $a < \infty$, $b \leq \nu$ (resp. $a \leq \nu$, $b < \infty$) makes $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$) into a Frechet space. We define another family $\|\cdot\|_{a,b,1}$ of seminorms on $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$) with a, b in the same range as before except that $a \leq \nu - \dim \mathfrak{b} - 1$ in the case of $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$ by

$$\|f\|_{a,b,1} = \int_{\mathfrak{b}(\mathbb{R})} (1 + \|x\|)^a \sum_{X \in \mathcal{B}_b} \left|(Xf)(x)\right| dx.$$

Then these seminorms are continuous with respect to the topology induced by the $\|\cdot\|_{a',b'}$. (The words "seminorm" and "continuous seminorm" will be used synonymously.)

Remark 2.2. For our later estimates when we need the seminorms defined above we usually fix the non-Archimedean part of the test function, and only prove that we can find an upper bound in terms of seminorms on $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$ and $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$. With a little more care one could make the upper bounds stronger in the sense that they could be stated in terms of seminorms on the whole space $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{A}))$ and $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{A}))$.

We fix a non-degenerate invariant bilinear form $\langle\cdot,\cdot\rangle : \mathfrak{g}(\mathbb{A}) \times \mathfrak{g}(\mathbb{A}) \longrightarrow \mathbb{A}$ by setting $\langle X, Y\rangle = \mathrm{tr}(XY)$ for $X, Y \in \mathfrak{g}(\mathbb{A})$. Let $\psi : \mathbb{Q}\backslash\mathbb{A} \longrightarrow \mathbb{C}$ be the non-trivial character constructed in [Lan94, XIV, § 1]. We define the Fourier transform

$$\widehat{\ } : \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A})) \longrightarrow \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A})), \quad \widehat{\Phi}(Y) := \int_{\mathfrak{g}(\mathbb{A})} \Phi(X)\psi(\langle X, Y\rangle)\, dX$$

with respect to this bilinear form.

Remark 2.3. It is clear that if $\infty \geq \nu' \geq \nu > 0$, then $\mathcal{S}^{\nu'}(\mathfrak{g}(\mathbb{A})) \subseteq \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ so that every statement holding for $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ for some $\nu \in (0,\infty]$ also holds for each $\Phi' \in \mathcal{S}^{\nu'}(\mathfrak{g}(\mathbb{A}))$ for every $\nu' \geq \nu$.

2.6 Siegel Sets

If $T \in \mathfrak{a}$, let $A_0^G(T)$ denote the set of all $a \in A_0^G$ with $\alpha(H_0(a) - T) > 0$ for all $\alpha \in \Delta_0$. Reduction theory proves the existence of $T_1 \in -\mathfrak{a}^+$ such that

$$G(\mathbb{A})^1 = G(\mathbb{Q})P_0(\mathbb{A})^1 A_0^G(T_1)\mathbf{K}.$$

We fix such a T_1 from now on and write

$$\mathcal{C}_{T_1} = \{g = pk \in P_0(\mathbb{A})\mathbf{K} \mid \forall\alpha \in \Delta_0 : \alpha(H_0(a) - T_1) > 0\}.$$

If then $f : G(\mathbb{Q})\backslash G(\mathbb{A})^1 \longrightarrow \mathbb{R}_{\geq 0}$ is measurable, we have

$$\begin{aligned}\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} f(g)\, dg &\leq \int_{A_G P_0(\mathbb{Q})\backslash \mathcal{C}_{T_1}} f(g)\, dg \\ &= \int_{\mathbf{K}} \int_{U_0(\mathbb{Q})\backslash U_0(\mathbb{A})} \int_{T_0(\mathbb{Q})\backslash T_0(\mathbb{A})^1} \int_{A_0^G} \delta_0(a)^{-1} \tau_0^G(H_0(a) - T_1) f(uatk)\, da\, dt\, du\, dk.\end{aligned} \tag{4}$$

2.7 Distributions Associated with Equivalence Classes

For $\mathfrak{o} \in \mathcal{O}$ and sufficiently regular $T \in \mathfrak{a}^+$ define for $x \in G(\mathbb{A})$ and integrable $\Phi : \mathfrak{g}(\mathbb{A}) \longrightarrow \mathbb{C}$ (cf. [Cha02])

$$K_{P,\mathfrak{o}}(x, \Phi) = \int_{\mathfrak{u}_P(\mathbb{A})} \sum_{X \in \mathfrak{m}_P(\mathbb{Q}) \cap \mathfrak{o}} \Phi(\mathrm{Ad}\, x^{-1}(X + U))\, dU,$$

$$k_{\mathfrak{o}}^T(x,\Phi) = \sum_{P\in\mathcal{F}_{\mathrm{std}}} (-1)^{\dim A_P/A_G} \sum_{\delta\in P(\mathbb{Q})\backslash G(\mathbb{Q})} \hat{\tau}_P(H_0(\delta x)-T)K_{P,\mathfrak{o}}(\delta x,\Phi), \text{ and}$$

$$J_{\mathfrak{o}}^T(\Phi) = \int_{A_G G(\mathbb{Q})\backslash G(\mathbb{A})} k_{\mathfrak{o}}^T(x,\Phi)\,dx$$

provided the sum-integrals converge.

Part 1. The Zeta Function

3 The Trace Formula for Lie Algebras and Convergence of Distributions

Let us recall some of the main results from [Cha02].

Theorem 3.1 ([Cha02], Théoreme 3.1, Théoreme 4.5). *For all $\Phi \in \mathcal{S}(\mathfrak{g}(\mathbb{A}))$ and sufficiently regular $T \in \mathfrak{a}^+$ we have*

$$\int_{A_G G(\mathbb{Q})\backslash G(\mathbb{A})} \sum_{\mathfrak{o}\in\mathcal{O}} |k_{\mathfrak{o}}^T(x,\Phi)|\,dx < \infty. \tag{5}$$

and

$$\sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(\Phi) = \sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(\hat{\Phi}). \tag{6}$$

The distributions $J_{\mathfrak{o}}^T(\Phi)$ and $\sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(\Phi)$ are polynomials in T of degree at most $\dim\mathfrak{a}$.

The Poisson summation like identity (6) is what we refer to as Chaudouard's trace formula for the Lie algebra $\mathfrak{g}$.

Remark 3.2.

(i) Since the distributions in the theorem are polynomials in T for T varying in a non-empty open cone of $\mathfrak{a}$, they can be defined at any point $T \in \mathfrak{a}$, with (6) then being valid for all $T \in \mathfrak{a}$.
(ii) The results in [Cha02] hold for arbitrary connected reductive groups G.
(iii) Equation (5) holds for every $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A})) \cup \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A}))$, and (6) holds for every $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ if $\nu > 0$ is sufficiently large in a sense depending only on n, cf. also the proof of Lemma 3.7 below.

For $\Phi : \mathfrak{g}(\mathbb{A}) \longrightarrow \mathbb{C}$, $\lambda \in (0,\infty)$, and $x \in \mathfrak{g}(\mathbb{A})$ put

$$\Phi_\lambda(x) := \Phi(\lambda x).$$

For fixed λ, $\Phi_\lambda \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ if $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$, and $\Phi_\lambda \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A}))$ if $\Phi \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A}))$. Hence (6) becomes

$$\sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(\Phi_\lambda) = \lambda^{-n^2} \sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(\hat{\Phi}_{\lambda^{-1}})$$

if ν is sufficiently large. Let $\mathcal{O}_* := \mathcal{O}\backslash\{\mathfrak{n}\}$, and for sufficiently regular $T \in \mathfrak{a}^+$ set $J_*^T = \sum_{\mathfrak{o}\in\mathcal{O}_*} J_{\mathfrak{o}}^T$.

Definition 3.3. We define the regularised zeta function by

$$\Xi^T(s,\Phi) = \int_0^\infty \lambda^{n(s+\frac{n-1}{2})} J_*^T(\Phi_\lambda)\, d^\times\lambda$$

provided this sum-integral converges.

Theorem 3.4. *There exists $\nu > 0$ depending only on n such that for all $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ the following holds:*

(i) If T is sufficiently regular, the function

$$\Xi^{T,+}(s,\Phi) = \int_1^\infty \lambda^{n(s+\frac{n-1}{2})} J_*^T(\Phi_\lambda)\, d^\times\lambda$$

is absolutely and locally uniformly convergent for all $s \in \mathbb{C}$ and hence entire.

(ii) If T is sufficiently regular, the integral defining $\Xi^T(s,\Phi)$ and also

$$\Xi_{\mathfrak{o}}^T(s,\Phi) := \int_0^\infty \lambda^{n(s+\frac{n-1}{2})} J_{\mathfrak{o}}^T(\Phi_\lambda)\, d^\times\lambda, \quad \mathfrak{o} \in \mathcal{O}_*,$$

are well defined and absolutely and locally uniformly convergent for $s \in \mathbb{C}$ with $\Re s > \frac{n+1}{2}$ (and hence holomorphic there). Moreover,

$$\Xi^T(s,\Phi) = \sum_{\mathfrak{o}\in\mathcal{O}_*} \Xi_{\mathfrak{o}}^T(s,\Phi).$$

(iii) The distributions $\Xi^{T,+}(s,\Phi)$, $\Xi_{\mathfrak{o}}^T(s,\Phi)$, and $\Xi^T(s,\Phi)$ are polynomials in T of degree at most $\dim\mathfrak{a} = n-1$. The coefficients of these polynomials are holomorphic functions in s for s ranging in the regions indicated above.

We need the analogue results for test functions $\Phi \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A}))$:

Theorem 3.5.

(i) Let $R > n$ be arbitrary. Then there exists $\nu > 0$ such that for all $\Phi \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A}))$ and all sufficiently regular T, the function

$$\Xi^{T,+}(s,\Phi) = \int_1^\infty \lambda^{n(s+\frac{n-1}{2})} J_*^T(\Phi_\lambda)\, d^\times\lambda$$

is absolutely and locally uniformly convergent for all $s \in \mathbb{C}$ with $\Re s < R$.

(ii) *With R, ν, Φ, and T as before, the integral defining $\Xi^T(s,\Phi)$ and also*

$$\Xi_{\mathfrak{o}}^T(s,\Phi) := \int_0^\infty \lambda^{n(s+\frac{n-1}{2})} J_{\mathfrak{o}}^T(\Phi_\lambda)\, d^\times\lambda, \quad \mathfrak{o}\in\mathcal{O}_*,$$

are well defined and absolutely and locally uniformly convergent for $s\in\mathbb{C}$ with $R>\Re s>\frac{n+1}{2}$ (and hence holomorphic there). Moreover,

$$\Xi^T(s,\Phi) = \sum_{\mathfrak{o}\in\mathcal{O}_*} \Xi_{\mathfrak{o}}^T(s,\Phi).$$

(iii) *The distributions $\Xi^{T,+}(s,\Phi)$, $\Xi_{\mathfrak{o}}^T(s,\Phi)$, and $\Xi^T(s,\Phi)$ are polynomials in T of degree at most $\dim\mathfrak{a} = n-1$. The coefficients of these polynomials are holomorphic functions in s for s ranging in the regions indicated above.*

Remark 3.6. The distributions in the theorems can again be defined at every point $T\in\mathfrak{a}$ by taking the value of the polynomial at this point. Their analytic properties as stated in the theorems stay valid for every $T\in\mathfrak{a}$.

Both theorems are immediate consequences of the following lemma.

Lemma 3.7. *Let $T\in\mathfrak{a}^+$ be sufficiently regular and let $\Phi_f\in\mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$. If Φ_∞ is a function on $\mathfrak{g}(\mathbb{R})$ we write $\Phi=\Phi_\infty\cdot\Phi_f$ in the following.*

(i) *There exists an integer $\nu>0$ (depending on n) such that the following holds.*

  (a) *For every $N\in\mathbb{N}$ there exists a seminorm μ_N on the space $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ such that*

$$\int_{A_G G(\mathbb{Q})\backslash G(\mathbb{A})} \sum_{\mathfrak{o}\in\mathcal{O}_*} |k_{\mathfrak{o}}^T(x,\Phi_\lambda)|\, dx \le \mu_N(\Phi_\infty)\lambda^{-N} \tag{7}$$

  for all $\lambda\in[1,\infty)$ and $\Phi_\infty\in\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$.

  (b) *There exists a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$) such that*

$$\int_{A_G G(\mathbb{Q})\backslash G(\mathbb{A})} \sum_{\mathfrak{o}\in\mathcal{O}_*} |k_{\mathfrak{o}}^T(x,\Phi_\lambda)|\, dx \le \mu(\Phi_\infty)\lambda^{-n^2} \tag{8}$$

  for all $\lambda\in(0,1]$ and $\Phi_\infty\in\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\Phi_\infty\in\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$).

(ii) *If $N\in\mathbb{N}$, then there exists an integer $\nu>0$ and a seminorm μ_N on the space $\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$, both depending only on n and N, such that*

$$\int_{A_G G(\mathbb{Q})\backslash G(\mathbb{A})} \sum_{\mathfrak{o}\in\mathcal{O}_*} |k_{\mathfrak{o}}^T(x,\Phi_\lambda)|\, dx \le \mu_N(\Phi_\infty)\lambda^{-N} \tag{9}$$

for all $\lambda\in[1,\infty)$ and $\Phi_\infty\in\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$.

We will prove this lemma in Sect. 3.2 below, but first deduce Theorem 3.4 from it (The proof of Theorem 3.5 is analogous and we omit it here.)

Proof of Theorem 3.4.

(i) By Lemma 3.7 we have for N arbitrarily large and every $\lambda \geq 1$,

$$|\lambda^{n(s+\frac{n-1}{2})} J_*^T(\Phi_\lambda)| \leq \mu_N(\Phi_\infty)\lambda^{n(\Re s+\frac{n-1}{2})}\lambda^{-N},$$

which is of course integrable over $\lambda \in [1, \infty)$ if N is chosen sufficiently large.

(ii) We split the integral defining $\Xi^T(s, \Phi)$ into one integral over $\lambda \in (0, 1]$ and one over $\lambda \in [1, \infty)$. By the first part of the proposition the second integral defines a holomorphic function on all of $\mathbb{C}$. For the first integral we have $|J_*^T(\Phi_\lambda)| \leq \mu(\Phi_\infty)\lambda^{-n^2}$ for all $\lambda \leq 1$ by Lemma 3.7 so that

$$\int_0^1 |\lambda^{n(s+\frac{n-1}{2})} J_*^T(\Phi_\lambda)|\, d^\times\lambda \leq \mu(\Phi_\infty) \int_0^1 \lambda^{n(\Re s+\frac{n-1}{2})}\lambda^{-n^2}\, d^\times\lambda,$$

which is finite if $\Re s > \frac{n+1}{2}$, and hence proving the second part of the proposition.

(iii) By Theorem 3.1 $J_{\mathfrak{o}}^T(\Phi)$ and $J_*^T(\Phi)$ are polynomials of degree at most $\dim \mathfrak{a}$ in T. The assertion thus follows from the previous parts of the proposition.

$\square$

3.1 Auxiliary Results

To prove Lemma 3.7, we need some preparation. Let $P_1, P_2, R \in \mathcal{F}_{\mathrm{std}}$ be standard parabolic subgroups with $P_1 \subseteq R \subseteq P_2$, and write $P_i = M_i U_i$ for their Levi decomposition. We define

$$\tilde{\mathfrak{m}}_1^2 = \tilde{\mathfrak{m}}_{P_1}^{P_2} = \mathfrak{m}_2 \backslash \Big(\bigcup_{\substack{Q \in \mathcal{F}: \\ P_1 \subseteq Q \subsetneq P_2}} \mathfrak{m}_2 \cap \mathfrak{q} \Big).$$

Note that $0 \notin \tilde{\mathfrak{m}}_1^2(\mathbb{Q})$ unless $P_1 = P_2$. Moreover, $\tilde{\mathfrak{m}}_1^2 = \mathfrak{m}_1$ if and only if $P_1 = P_2$. Similarly, put

$$\bar{\mathfrak{u}}_1^{2\prime} = \bar{\mathfrak{u}}_{P_1}^{P_2\prime} = \bar{\mathfrak{u}}_1^2 \backslash \Big(\bigcup_{\substack{Q \in \mathcal{F}: \\ P_1 \subseteq Q \subsetneq P_2}} \bar{\mathfrak{u}}_1^Q \Big) = \bar{\mathfrak{u}}_1^2 \backslash \Big(\bigcup_{\substack{Q \in \mathcal{F}: \\ P_1 \subseteq Q \subsetneq P_2}} \bar{\mathfrak{u}}_{P_1} \cap \mathfrak{m}_Q \Big),$$

and define $\mathfrak{u}_1^{2\prime}$ with $\mathfrak{u}_1^Q$ in place of $\bar{\mathfrak{u}}_1^Q$ analogously. Note that $0 \notin \bar{\mathfrak{u}}_1^{2\prime}(\mathbb{Q})$ unless $P_1 = P_2$.

Definition 3.8.

(i) If $S \subseteq \Sigma_1^2$ is a subset, we say that S has *property* $\Pi(P_1, R, P_2)$ if for every $\alpha \in \Delta_1^2 \backslash \Delta_1^R$ there exists $\beta \in S$ such that $\varpi_\alpha(\beta^\vee) > 0$. In particular, $S = \emptyset$ has property $\Pi(P_1, R, P_2)$.

(ii) If $S \subseteq \Sigma_1^2$ has property $\Pi(P_1, R, P_2)$, we define $\overline{\mathfrak{u}}'_S \subseteq \overline{\mathfrak{u}}_R^2$ as the set consisting of all $Y = \sum_{\beta \in \Sigma_1^2} Y_{-\beta} \in \overline{\mathfrak{u}}_R^2$ with $Y_{-\beta} \neq 0$ for $\beta \in S$ and $Y_{-\beta} = 0$ for $\beta \notin S$. Here $Y_{-\beta}$ denotes the component of Y in the $(-\beta)$-eigenspace of the decomposition of $\overline{\mathfrak{u}}_R^2$ with respect to $-\Sigma_1^2$. In particular, $\overline{\mathfrak{u}}'_\emptyset = \emptyset$ unless $R = P_1$ in which case $\overline{\mathfrak{u}}'_\emptyset = \overline{\mathfrak{u}}_1^{1\prime} = \{0\}$.

(iii) If $S \subseteq \Sigma_1^R$ has property $\Pi(P_1, P_1, R)$, let $\mathfrak{m}_{R,S} \subseteq \mathfrak{m}_R$ consist of all $Y \in \mathfrak{m}_R$ such that $Y_{-\beta} \neq 0$ for all $\beta \in S$ and $Y_{-\beta} = 0$ for all $\beta \in \Sigma_1^R \backslash S$. Here $Y_{-\beta}$ denotes the component of Y in the $(-\beta)$-eigenspace of the decomposition of $\mathfrak{m}_R$ with respect to $\Phi(A_1, M_R)$.

Lemma 3.9. *Write* $\mathfrak{m}_2 = \bigoplus_{\beta \in \Phi(A_1, M_2)} \mathfrak{m}_\beta$ *with* $\mathfrak{m}_\beta$ *the eigenspace for* β *in* $\mathfrak{m}_2$*, and if* $X \in \mathfrak{m}_2(\mathbb{Q})$*, let* $X_\beta \in \mathfrak{m}_\beta(\mathbb{Q})$ *be its* β*-component so that* $X = \sum_{\beta \in \Phi(A_1, M_2)} X_\beta$*. Then:*

(i) For every $Y \in \overline{\mathfrak{u}}_R^{2\prime}(\mathbb{Q})$*, there exists a subset* $S \subseteq \Sigma_1^2$ *with property* $\Pi(P_1, R, P_2)$ *such that* $Y_{-\beta} = 0$ *for all* $\beta \in \Sigma_1^2 \backslash S$ *and* $Y_{-\beta} \neq 0$ *for all* $\beta \in S$*. In particular,*

$$\overline{\mathfrak{u}}_R^{2\prime} = \bigsqcup_{S \subseteq \Sigma_1^2} \overline{\mathfrak{u}}'_S$$

where the disjoint union is over all subsets $S \subseteq \Sigma_1^2$ *having property* $\Pi(P_1, R, P_2)$*.*

(ii) If $P_1 \subsetneq R$ *and* $X \in \tilde{\mathfrak{m}}_{P_1}^R(\mathbb{Q})$*, there exists a non-empty subset* $S \subseteq \Sigma_1^R$ *with property* $\Pi(P_1, P_1, R)$ *such that* $X_{-\beta} \neq 0$ *for every* $\beta \in S$*. In particular,*

$$\tilde{\mathfrak{m}}_{P_1}^R \subseteq \bigsqcup_{S \subseteq \Sigma_1^R} \mathfrak{m}_{R,S}.$$

where the disjoint union is over all non-empty subsets $S \subseteq \Sigma_1^R$ *having property* $\Pi(P_1, P_1, R)$*.*

Proof.

(i) Let $Y \in \overline{\mathfrak{u}}_R^2$. Let the set $S \subseteq \Sigma_1^2$ be defined to consist exactly of those $\beta \in \Sigma_1^2$ with $Y_{-\beta} \neq 0$. S has property $\Pi(P_1, R, P_2)$: For that suppose that instead there exists $\alpha \in \Delta_1^2 \backslash \Delta_1^R$ such that for all $\beta \in S$ we have $\varpi_\alpha(\beta^\vee) \leq 0$. Now every β is a non-negative linear combination of elements in Δ_1^2 so that $\varpi_\alpha(\beta^\vee) \leq 0$ implies $\varpi_\alpha(\beta^\vee) = 0$. But this implies that $\beta \in \Sigma_1^Q$ for some parabolic subgroup $Q \subsetneq P_2$, $R \subseteq Q$. Hence $Y \in \overline{\mathfrak{u}}_R^Q(\mathbb{Q})$ in contradiction to $Y \in \overline{\mathfrak{u}}_R^{2\prime}(\mathbb{Q})$ so that our set S must have property $\Pi(P_1, R, P_2)$.

(ii) This follows from the definitions. $\square$

Lemma 3.10. *Suppose $R \subsetneq P_2$. If $m > \dim \mathfrak{u}_R^2$, then there exist constants $c > 0$ and $k_\alpha \geq 0$ for every $\alpha \in \Delta_1^2$ such that*

- *$k_\alpha > 0$ for all $\alpha \in \Delta_1^2 \backslash \Delta_1^R$, and*
- *for all $a \in A_1^G = A_{P_1}^G$, we have*

$$\sum_{Y \in \overline{\mathfrak{u}}_R^{2\prime}(\frac{1}{N}\mathbb{Z})} || \operatorname{Ad} a^{-1} Y||^{-m} \leq c \prod_{\alpha \in \Delta_1^2} e^{-k_\alpha \alpha(H_0(a))}.$$

Proof. This is a slightly refined version of [Art78, pp. 946–947] in that we give a sufficient lower bound for the exponent m. Suppose first that $m > 0$ is sufficiently large. We shall later see that $m > \dim \mathfrak{u}_R^2$ suffices.

Consider non-empty subsets $S \subseteq \Sigma_R^2$ with property $\Pi(P_1, R, P_2)$. By Lemma 3.9(i) the set $\overline{\mathfrak{u}}_R^{2\prime}(\frac{1}{N}\mathbb{Z})$ is the direct sum over such sets S of $\overline{\mathfrak{u}}_S'(\frac{1}{N}\mathbb{Z})$. For $\beta \in \Sigma_R^2$ let $\{E_{-\beta,i}\}_{i=1,\ldots,d_{-\beta}}$, $d_{-\beta} := \dim \mathfrak{u}_{-\beta}$, be a basis for the eigenspace $\mathfrak{u}_{-\beta}$ of $-\beta$ in $\overline{\mathfrak{u}}_R^2$, which is orthogonal with respect to the norm $\|\cdot\|$, i.e. $\|\sum_i b_i E_{-\beta,i}\| = \sum_i |b_i|$ for all $b_1, \ldots, b_{d_{-\beta}} \in \mathbb{R}$. Thus, if $Y \in \overline{\mathfrak{u}}_S(\frac{1}{N}\mathbb{Z})$, we can uniquely write $Y = \sum_{\beta \in S} \sum_{i=1}^{d_{-\beta}} Y_{-\beta,i} E_{-\beta,i}$, and get for every $a \in A_1^G$ that

$$\| \operatorname{Ad} a^{-1} Y\| = \sum_{\beta \in S} e^{2\beta(H_0(a))} \sum_{i=1}^{d_{-\beta}} \|Y_{-\beta,i}\|.$$

Let $\mathcal{R} = (R_\beta)_{\beta \in S}$ be a tuple of non-empty subsets $R_\beta \subseteq \{1, \ldots, d_{-\beta}\}$, and define

$$\overline{\mathfrak{u}}_{S,\mathcal{R}}' = \{Y \in \overline{\mathfrak{u}}_S' \mid Y_{-\beta,i} \neq 0 \Leftrightarrow \beta \in S \text{ and } i \in R_\beta\}.$$

Clearly, $\overline{\mathfrak{u}}_S' = \bigsqcup_{\mathcal{R}=(R_\beta)_{\beta \in S}} \overline{\mathfrak{u}}_{S,\mathcal{R}}'$ with the disjoint union being over all tuples $\mathcal{R}$ as before. As there are only finitely many such tuples $\mathcal{R}$, it suffices to consider the sum over $Y \in \overline{\mathfrak{u}}_{S,\mathcal{R}}'(\frac{1}{N})$ for one of the tuples $\mathcal{R}$.

Then, since $0 \notin \overline{\mathfrak{u}}_S'$ because of $R \subsetneq P_2$,

$$\begin{aligned}\sum_{Y \in \overline{\mathfrak{u}}_{S,\mathcal{R}}'(\frac{1}{N}\mathbb{Z})} \| \operatorname{Ad} a^{-1} Y\|^{-m} &= \sum_{Y \in \overline{\mathfrak{u}}_{S,\mathcal{R}}'(\frac{1}{N}\mathbb{Z})} \left(\sum_{\beta \in S} \sum_{i \in R_\beta} e^{\beta(H_0(a))} \|Y_{-\beta,i}\| \right)^{-m} \\ &\leq \prod_{\beta \in S} \prod_{i \in R_\beta} \sum_{Y_{-\beta,i} \in \frac{1}{N}\mathbb{Z} \backslash \{0\}} \left(e^{\beta(H_0(a))} \|Y_{-\beta,i}\| \right)^{-\frac{m}{r}},\end{aligned}$$

where $r := \sum_{\beta \in S} |R_\beta| \leq \dim \mathfrak{u}_R^2$. This last product equals

$$\left(\sum_{X \in \frac{1}{N}\mathbb{Z} \backslash \{0\}} \|X\|^{-\frac{m}{r}} \right)^r \prod_{\beta \in S} \prod_{i \in R_\beta} e^{-m\beta(H_0(a))/r} = \left(\sum_{X \in \frac{1}{N}\mathbb{Z} \backslash \{0\}} \|X\|^{-\frac{m}{r}} \right)^r \prod_{\beta \in S} e^{-m|R_\beta|\beta(H_0(a))/r}.$$

The sum $\sum_{X\in\frac{1}{N}\mathbb{Z}\setminus\{0\}}\|X\|^{-\frac{m}{r}}$ is finite if $m>r$, so it is in particular finite if $m>\dim\mathfrak{u}_R^2\geq r$, which gives our lower bound on m. Now every β is a non-negative linear combination of roots in Δ_1^2 so that the above product equals

$$\left(\sum_{X\in\frac{1}{N}\mathbb{Z}\setminus\{0\}}\|X\|^{-\frac{m}{r}}\right)^r\prod_{\alpha\in\Delta_1^2}e^{-k_{\alpha,S,\mathcal{R}}\alpha(H_0(a))}$$

for suitable constants $k_{\alpha,S,\mathcal{R}}\geq 0$. Since S has property $\Pi(P_1,R,P_2)$, there exists for every $\alpha\in\Delta_1^2\backslash\Delta_1^R$ some $\beta\in S$ such that α occurs non-trivially in β. Hence, since $|R_\beta|>0$ for every $\beta\in S$, the corresponding coefficient satisfies $k_{\alpha,S,\mathcal{R}}>0$ if $\alpha\in\Delta_1^2\backslash\Delta_1^R$, which finishes the proof. □

Lemma 3.11. *For $\beta\in\Phi(A_1,\mathfrak{m}_R)=\Phi(A_1^R,\mathfrak{m}_R)=:\Phi_1^R$, denote by $\mathfrak{m}_\beta\subseteq\mathfrak{m}_R$ the eigenspace of β in $\mathfrak{m}_R$ so that $\mathfrak{m}_R=\bigoplus_{\beta\in\Phi_1^R}\mathfrak{m}_\beta$. Put $A_1^R(T_1)=\{a\in A_1^R\mid\forall\alpha\in\Delta_1^R:\ \alpha(H_{P_1}(a)-T_1)>0\}$, let $k>1$ be given, and let $\nu>k+n^2$. Let $N>0$ be a positive integer.*

Then for every $\alpha\in\Delta_0^R$ there exists a constant $k_\alpha\geq 0$, and for every $\beta\in\Phi_0^R$ a seminorm μ_β on $\mathcal{S}^1(\mathfrak{m}_\beta(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{m}_\beta(\mathbb{R}))$) such that the following holds:

- *$k_\alpha>0$ for every $\alpha\in\Delta_0^R\backslash\Delta_0^1$, and*
- *for all $\lambda>0$, all $\varphi_\beta\in\mathcal{S}^1(\mathfrak{m}_\beta(\mathbb{R}))$ (resp. $\varphi_\beta\in\mathcal{S}_\nu(\mathfrak{m}_\beta(\mathbb{R}))$), and all $a\in A_1^R(T_1)$ we have*

$$\delta_0^R(a)^{-1}\sum_{\substack{X\in\tilde{\mathfrak{m}}_{P_1}^R(\frac{1}{N}\mathbb{Z})\\ X\notin\mathfrak{n}}}\prod_{\beta\in\Phi_0^R}\varphi_\beta(\lambda\beta(a)^{-1}X_\beta)$$
$$\leq\begin{cases}\lambda^{-\dim\mathfrak{m}_R}\mu(\varphi)\prod_{\alpha\in\Delta_0^R\backslash\Delta_0^1}e^{-k_\alpha\alpha(H_0(a))} & \text{if } \lambda\leq 1,\\ \lambda^{-k}\mu(\varphi)\prod_{\alpha\in\Delta_0^R\backslash\Delta_0^1}e^{-k_\alpha\alpha(H_0(a))} & \text{if } \lambda\geq 1,\end{cases}\tag{10}$$

where $\mu(\varphi):=\prod_{\beta\in\Phi_1^R}\mu_\beta(\varphi_\beta)$.

Proof. Suppose first that $R\neq P_1$. The left-hand side of (10) can by Lemma 3.9(ii) be bounded by a sum over non-empty subsets $S\subseteq\Sigma_1^R$ with property $\Pi(P_1,P_1,R)$ of the terms

$$\left(\prod_{\beta\in S}\sum_{X_{-\beta}\in\mathfrak{m}_{-\beta}(\frac{1}{N}\mathbb{Z})\setminus\{0\}}\varphi_{-\beta}(\lambda\beta(a)X_{-\beta})\right)\left(\prod_{\beta\in\Phi_0^1\cup\Sigma_1^R}\sum_{X_\beta\in\mathfrak{m}_\beta(\frac{1}{N}\mathbb{Z})}\varphi_\beta(\lambda\beta(a)^{-1}X_\beta)\right).$$

Recall that if V is a finite dimensional vector space and $\Lambda\subseteq V(\mathbb{R})$ some lattice, then for every $r>1$ there exists a seminorm μ_r on $\mathcal{S}_{r+\dim V}(V(\mathbb{R}))$ such that for all $s>0$ and all $\Psi\in\mathcal{S}_{r+\dim V}(V(\mathbb{R}))\cup\mathcal{S}^1(V(\mathbb{R}))$ we have

$$\sum_{X=(X_1,\ldots,X_{\dim V})\in\Lambda,X_1\neq 0}|\Psi(sX)|\leq\mu_r(\Psi)\sup\{1,s^{-1}\}^{\dim V}\sup\{1,s\}^{-r},$$

see, e.g., [Wri85, pp. 510–511]. (Note that in [Wri85] this estimate was only proved for $\Psi \in \mathcal{S}(V(\mathbb{R}))$, but it is clear from the proof there that one only needs a polynomial decay of Ψ up to a certain power and no differentiability at all.) In particular, after possibly changing the seminorm in a way depending only on $\dim V$, we get

$$\sum_{X\in\Lambda} |\Psi(sX)| \leq \begin{cases} \mu_r(\Psi)s^{-\dim V} & \text{if } s \leq 1, \\ \mu_r(\Psi) & \text{if } s \geq 1, \quad \text{and} \end{cases} \tag{11}$$

$$\sum_{X\in\Lambda, X\neq 0} |\Psi(sX)| \leq \begin{cases} \mu_r(\Psi)s^{-\dim V} & \text{if } s \leq 1, \\ \mu_r(\Psi)s^{-r} & \text{if } s \geq 1. \end{cases} \tag{12}$$

From this it follows that for every $\beta \in \{0\} \cup \Sigma_1^R$ there exists a seminorm μ_β on $\mathcal{S}_{\dim \mathfrak{m}_\beta+1}(\mathfrak{m}_\beta(\mathbb{R}))$ (resp. $\mathcal{S}^1(\mathfrak{m}_\beta(\mathbb{R}))$) such that for all $\lambda > 0$ and all $a \in A_1^R(T_1)$ we have

$$\sum_{X_\beta\in\mathfrak{m}_\beta(\frac{1}{N}\mathbb{Z})} \varphi_\beta(\lambda\beta(a)^{-1}X_\beta) \leq \begin{cases} \mu_\beta(\varphi_\beta)\beta(a)^{\dim \mathfrak{m}_\beta}(\lambda^{-1}+1)^{\dim \mathfrak{m}_\beta} & \text{if } \beta \geq 0, \\ \mu_\beta(\varphi_\beta)(\lambda^{-1}+1)^{\dim \mathfrak{m}_\beta} & \text{if } \beta < 0. \end{cases}$$

For this inequality also recall that $a \in A_0^R(T_1)$ implies that $\beta(a)$ is uniformly bounded from below if $\beta > 0$. Hence for all $\lambda > 0$ and $a \in A_1^R(T_1)$,

$$\begin{aligned} \prod_{\beta\in\{0\}\cup\Sigma_1^R} \sum_{X_\beta\in\mathfrak{m}_\beta(\frac{1}{N}\mathbb{Z})} \varphi_\beta(\lambda\beta(a)^{-1}X_\beta) &\leq \delta_0^R(a)(\lambda^{-1}+1)^{\dim \mathfrak{p}_1} \prod_{\beta\in\{0\}\cup\Sigma_1^R} \mu_\beta(\varphi_\beta) \\ &\leq \begin{cases} c\delta_0^R(a)\prod_{\beta\in\Phi_0^1\cup\Sigma_1^R}\mu_\beta(\varphi_\beta) & \text{if } \lambda \geq 1, \\ c\delta_0^R(a)\lambda^{-\dim \mathfrak{p}_1}\prod_{\beta\in\{0\}\cup\Sigma_1^R}\mu_\beta(\varphi_\beta) & \text{if } \lambda < 1, \end{cases} \end{aligned}$$

where $c > 0$ is some constant.

Similarly, for every $\beta \in S$ and every $k > 1$, there is a seminorm $\mu_{-\beta,k}$ on $\mathcal{S}_{k+\dim \mathfrak{m}_{-\beta}}(\mathfrak{m}_{-\beta}(\mathbb{R}))$ (resp. $\mathcal{S}^1(\mathfrak{m}_{-\beta}(\mathbb{R}))$) such that for all $\lambda > 0$ and all $a \in A_0^R(T_1)$ we have

$$\sum_{X_{-\beta}\in\mathfrak{m}_{-\beta}(\frac{1}{N}\mathbb{Z})\setminus\{0\}} \varphi_{-\beta}(\lambda\beta(a)X_{-\beta}) \leq \begin{cases} \mu_{-\beta,k}(\varphi_{-\beta})\lambda^{-k}\beta(a)^{-k} & \text{if } \lambda \geq 1, \\ \mu_{-\beta,k}(\varphi_{-\beta})(\lambda\beta(a))^{-\dim \mathfrak{m}_{-\beta}} & \text{if } \lambda < 1. \end{cases}$$

Hence,

$$\prod_{\beta\in S} \sum_{X_{-\beta}\in\mathfrak{m}_{-\beta}(\frac{1}{N}\mathbb{Z})\setminus\{0\}} \varphi_{-\beta}(\lambda\beta(a)X_{-\beta}) \leq \begin{cases} \lambda^{-k\sum_{\beta\in S}\dim \mathfrak{m}_{-\beta}}\mu_{S,k}(\varphi)\prod_{\beta\in S}\beta(a)^{-k} & \text{if } \lambda \geq 1, \\ \lambda^{-\sum_{\beta\in S}\dim \mathfrak{m}_{-\beta}}\mu_{S,k}(\varphi)\prod_{\beta\in S}\beta(a)^{-\dim \mathfrak{m}_{-\beta}} & \text{if } \lambda < 1, \end{cases}$$

where $\mu_{S,k}(\varphi) := \prod_{\beta\in S}\mu_{-\beta,k}(\varphi_{-\beta})$. Now every $\beta \in S$ can be written as $\beta = \sum_{\alpha\in\Delta_0^R} b_{\beta,\alpha}\alpha$ for $b_{\beta,\alpha} \geq 0$ suitable constants so that $\sum_{\beta\in S}\beta = \sum_{\alpha\in\Delta_0^R} B_\alpha\alpha$ with $B_\alpha := \sum_{\beta\in S} b_{\beta,\alpha}$. Since S has property $\Pi(P_1, P_1, R)$, we have $B_\alpha > 0$ if $\alpha \in \Delta_0^R\backslash\Delta_0^1$ so that

$$\prod_{\beta\in S}\beta(a)^{-k} \leq c \prod_{\alpha\in\Delta_0^R\backslash\Delta_0^1} e^{-kB_\alpha\alpha(H_0(a))}$$

for a suitable constant $c > 0$. Multiplying the above estimates gives the assertion if $R \neq P_1$. If $R = P_1$, we simply use the estimate for the sum over $X \in \Lambda$, $X \neq 0$, given in (11) and (12). □

Remark 3.12. Under the same assumptions and with the same notation as in the previous lemma, it actually follows that for a suitable seminorm μ, we have for every $\lambda \in (0, 1]$

$$\delta_0^R(a)^{-1} \sum_{X\in\tilde{\mathfrak{m}}_{P_1}^R(\frac{1}{N}\mathbb{Z})\cap\mathfrak{n}} \prod_{\beta\in\Phi_1^R} \varphi_\beta(\lambda\beta(a)^{-1}X_\beta) \leq \lambda^{-\dim\mathfrak{m}_R+1}\mu(\varphi) \prod_{\alpha\in\Delta_0^R\backslash\Delta_0^1} e^{-k_\alpha\alpha(H_0(a))}, \tag{13}$$

since if X is nilpotent, $\operatorname{tr} X = 0$. Hence in the proof the sum over $X_0 \in \mathfrak{m}_0(\frac{1}{N}\mathbb{Z})$ can be restricted to the vector subspace of traceless matrices which has codimension 1. Of course, similar versions of this inequality hold if we intersect $\mathfrak{m}_0$ with other vector subspaces of positive codimension.

Lemma 3.13. *Suppose we are given positive numbers $m_\alpha > 0$ for each $\alpha \in \Delta_1^2$. Then for every sufficiently regular $T \in \mathfrak{a}^+$ we have*

$$\int_{A_1^G} \sigma_1^2(H_0(a) - T) \prod_{\alpha\in\Delta_1^2} e^{-m_\alpha\alpha(H_0(a))}\, da < \infty. \tag{14}$$

Proof. This is essentially contained in [Cha02, p. 365] (cf. also [Art78, p. 947]), but we need to find a sufficient lower bound for the m_α. We can write the integral in (14) as

$$\int_{\mathfrak{a}_1^G} \sigma_1^2(H - T) \prod_{\alpha\in\Delta_1^2} e^{-m_\alpha\alpha(H)}\, dH.$$

If $H \in \mathfrak{a}_{P_1}^G$, we decompose it as $H = H_1 + H_2$ with uniquely determined $H_1 \in \mathfrak{a}_1^2$ and $H_2 \in \mathfrak{a}_2^G$. Then $\sigma_1^2(H - T) \neq 0$ implies $t_\alpha := \alpha(H) = \alpha(H_1) > \alpha(T)$ for all $\alpha \in \Delta_1^2$, and also the existence of a constant $c > 0$ (independent of H) such that

$$\|H_2\| \leq c(1 + \sum_{\alpha\in\Delta_1^2} t_\alpha) \leq c \prod_{\alpha\in\Delta_1^2}(1 + t_\alpha)$$

(cf. [Art78, Corollary 6.2]). Hence the volume in $\mathfrak{a}_2^G$ of all contributing H_2 is bounded by a polynomial in the t_α for $\alpha \in \Delta_1^2$ so that there exists some $c > 0$ such that the above integral is bounded by

$$c \prod_{\alpha\in\Delta_1^2} \int_{\alpha(T)}^{\infty} (1+t_\alpha)^k e^{-m_\alpha t_\alpha}\, dt_\alpha.$$

Since $m_\alpha > 0$ for all $\alpha \in \Delta_1^2$, this implies the assertion. $\square$

Let $\mathfrak{b} \subseteq \mathfrak{g}$ be a subspace as in Sect. 2.5, and let S be a set of roots acting on $\mathfrak{b}$ such that we have a direct decomposition $\mathfrak{b} = \bigoplus_{\beta\in S} \mathfrak{b}_\beta$. Let $\|\cdot\|$ denote a norm on $\mathfrak{b}(\mathbb{A})$ compatible with this direct sum decomposition (i.e., if $B = \sum_\beta B_\beta \in \mathfrak{b}(\mathbb{A})$, $B_\beta \in \mathfrak{b}_\beta(\mathbb{A})$, then $\|B\| = \sum_\beta \|B_\beta\|$).

Lemma 3.14. *Let $\nu > 0$ be a sufficiently large integer (with "sufficiently large" depending on n) and let $\Phi_f \in \mathcal{S}(\mathbb{A}_f)$. Then for every $Y \in \mathcal{U}(\mathfrak{b})_{\le\nu}$, there exists a constant $c_Y > 0$ such that the following holds: For every $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$ (resp. $\Phi_\infty \in \mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$) there are functions $\varphi_\beta = \varphi_{\beta,\infty}\cdot\varphi_{\beta,f} \in \mathcal{S}^\infty(\mathfrak{b}_\beta(\mathbb{A}))) = \mathcal{S}(\mathfrak{b}_\beta(\mathbb{A}))$ (resp. $\varphi_\beta = \varphi_{\beta,\infty}\cdot\varphi_{\beta,f} \in \mathcal{S}_\nu(\mathfrak{b}_\beta(\mathbb{A}))$), $\beta \in S$, such that (with $\Phi = \Phi_\infty \cdot \Phi_f$)*

(i) $\varphi_\beta \ge 0$ *for all* β.
(ii) $|\Phi(B)| \le \prod_{\beta\in S} \varphi_\beta(B_\beta)$ *for all* $B = \sum_{\beta\in S} B_\beta \in \mathfrak{b}(\mathbb{A})$.
(iii) For every tuple $(Y_\beta)_{\beta\in S} \in \bigoplus_{\beta\in S} \mathcal{U}(\mathfrak{b}_\beta)$ *of degree* $\sum_{\beta\in S} \deg Y_\beta \le \nu$ *we have*

$$\prod_{\beta\in S} \|Y_\beta \varphi_{\beta,\infty}\|_{L^1(\mathfrak{b}_\beta(\mathbb{R}))} \le c \sum_{X\in\mathcal{B}_{\mathfrak{b},\nu}} \|X\Phi_\infty\|_{L^1(\mathfrak{b}(\mathbb{R}))}$$

and

$$\prod_{\beta\in S} \sup_{B_\beta\in\mathfrak{b}_\beta(\mathbb{R})} |Y_\beta\varphi_{\beta,\infty}(B_\beta)| \le c\|\Phi_\infty\|_{0,\nu}$$

for $c = \max(\{c_{Y_\beta}\}_\beta \cup \{\prod_{\beta\in S} c_{Y_\beta}\})$.

Proof. We basically follow the proof of [FL11b, Lemma 3.4]. Since the set of functions $\bigotimes_\beta \mathcal{S}(\mathfrak{b}_\beta(\mathbb{A}_f))$ is dense in $\mathcal{S}(\mathfrak{b}(\mathbb{A}_f))$ and Φ_f is fixed, it suffices to treat the Archimedean part. As in the proof of [FL11b, Lemma 3.3] it follows that there exists a constant $c_\infty > 0$ such that for any $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$ (resp. $\Phi_\infty \in \mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$) and any $X \in \mathfrak{b}(\mathbb{R})$ we have

$$|\Phi_\infty(X)| \le c_\infty \sum_{Y_1\in\mathcal{B}_{\mathfrak{b},\nu_0}} \int_{\mathfrak{b}([-1,1])} |(Y_1\Phi_\infty)(X+Z)|\, dZ$$

where $\nu_0 > 0$ is a suitable constant which can be chosen to depend only on n. If ν is sufficiently large, the right-hand side is well defined. We now choose a non-negative, smooth, and compactly supported function φ on $\mathfrak{b}(\mathbb{R})$ such that $\varphi \geq c_\infty$ on $\mathfrak{b}([-1,1])$. Put

$$\tilde{\Phi}_\infty := \sum_{Y_1 \in \mathcal{B}_{\mathfrak{b},\nu_0}} |Y_1 \Phi_\infty| * \varphi.$$

Note that the convolution with φ maps $\mathcal{S}^\nu(\mathfrak{b}(\mathbb{R}))$ to $\mathcal{S}(\mathfrak{b}(\mathbb{R}))$ and $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$ to $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$. Hence $\tilde{\Phi}_\infty \in \mathcal{S}(\mathfrak{b}(\mathbb{R}))$ (resp. $\tilde{\Phi}_\infty \in \mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$) and $\tilde{\Phi}$ is non-negative and $\tilde{\Phi}_\infty(B) \geq |\Phi_\infty(B)|$ for all $B \in \mathfrak{b}(\mathbb{R})$. Moreover,

$$\|Y\tilde{\Phi}_\infty\|_{L^1(\mathfrak{b}(\mathbb{R}))} \leq \|Y\varphi_\infty\|_{L^1(\mathfrak{b}(\mathbb{R}))} \sum_{Y_1 \in \mathcal{B}_{\mathfrak{b},\nu_0}} \|Y_1 \Phi_\infty\|_{L^1(\mathfrak{b}(\mathbb{R}))}$$

as well as

$$\sup_{B \in \mathfrak{b}(\mathbb{R})} |Y\tilde{\Phi}_\infty(B)| \leq \|Y\varphi_\infty\|_{L^1(\mathfrak{b}(\mathbb{R}))} \sum_{Y_1 \in \mathcal{B}_{\mathfrak{b},\nu_0}} \sup_{B \in \mathfrak{b}(\mathbb{R})} |Y_1 \Phi_\infty(B)| \leq c_1 \|Y\varphi_\infty\|_{L^1(\mathfrak{b}(\mathbb{R}))} \|\Phi_\infty\|_{0,\nu}$$

for all $Y \in \mathcal{U}(\mathfrak{b}(\mathbb{R}))$ where $c_1 > 0$ is a suitable constant depending only on the chosen bases $\mathcal{B}_{\mathfrak{b},\nu_0}$ and $\mathcal{B}_{\mathfrak{b},\nu}$. Since the set of functions $\bigotimes_\beta \mathcal{S}(\mathfrak{b}_\beta(\mathbb{R}))$ (resp. $\bigotimes_\beta \mathcal{S}_\nu(\mathfrak{b}_\beta(\mathbb{R}))$) is dense in $\mathcal{S}(\mathfrak{b}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{b}(\mathbb{R}))$), we can replace $\tilde{\Phi}_\infty$ by a product of suitable functions $\varphi_{\beta,\infty}$ on the root spaces $\mathfrak{b}_\beta(\mathbb{R})$ without changing any of above properties. □

3.2 *Proof of Lemma 3.7*

Proof of Lemma 3.7. We basically follow the proof of [Cha02, Théorème 3.1], but we need to keep track of the central variable λ the whole time.

(i) Let $\lambda \geq 1$. For $\mathfrak{o} \in \mathcal{O}_*$, the truncated kernel $k_{\mathfrak{o}}^T(x,\Phi)$ can be written as a sum over standard parabolic subgroups P_1, P_2 with $P_1 \subseteq P_2$ of

$$k_{\mathfrak{o}}^T(x,\Phi) = \sum_{\substack{P_1,P,P_2:\\ P_1 \subseteq P \subseteq P_2}} \sum_{\delta \in P_1(\mathbb{Q})\backslash G(\mathbb{Q})} (-1)^{\dim A_P/A_G} F^{P_1}(\delta x, T) \sigma_{P_1}^{P_2}(H_0(\delta x) - T) K_{P,\mathfrak{o}}(\delta x, \Phi), \tag{15}$$

provided the right-hand side converges, cf. [Cha02, Lemma 2.8]. Hence the left-hand side of (7) can be bounded from above by a sum over parabolic subgroups P_1, P_2 with $P_1 \subseteq P_2$, and over $\mathfrak{o} \in \mathcal{O}_*$ of

$$\int_{A_G P_1(\mathbb{Q})\backslash G(\mathbb{A})} F^{P_1}(x,T) \sigma_{P_1}^{P_2}(H_0(x) - T) \cdot$$
$$\left| \sum_{P:\, P_1 \subseteq P \subseteq P_2} (-1)^{\dim A_P/A_G} \sum_{X \in \mathfrak{m}_P(\mathbb{Q}) \cap \mathfrak{o}} \int_{\mathfrak{u}_P(\mathbb{A})} \Phi(\lambda \operatorname{Ad} x^{-1}(X+U))\, dU \right| dx,$$

cf. [Cha02, pp. 360–361]. This can be replaced by the sum over P_1, R, P_2 with $P_1 \subseteq R \subseteq P_2$, and over $\mathfrak{o} \in \mathcal{O}_*$ of

$$\int_{A_G P_1(\mathbb{Q})\backslash G(\mathbb{A})} F^{P_1}(x,T)\sigma_{P_1}^{P_2}(H_0(x)-T)\cdot$$
$$\sum_{X\in\tilde{m}_{P_1}^R(\mathbb{Q})\cap\mathfrak{o}} \left| \sum_{P:\ R\subseteq P\subseteq P_2} (-1)^{\dim A_P/A_G} \sum_{Y\in\mathfrak{u}_R^P(\mathbb{Q})} \int_{\mathfrak{u}_P(\mathbb{A})} \Phi(\lambda \operatorname{Ad} x^{-1}(X+Y+U))dU \right| dx. \tag{16}$$

We can decompose

$$A_G P_1(\mathbb{Q})\backslash G(\mathbb{A}) = U_1(\mathbb{Q})\backslash U_1(\mathbb{A}) \times M_1(\mathbb{Q})\backslash M_1(\mathbb{A})^1 \times A_1^G \times \mathbf{K}$$

and write $x \in A_G P_1(\mathbb{Q})\backslash G(\mathbb{A})$ accordingly as $x = umak$. Then $F^{P_1}(x,T) = F^{P_1}(m,T)$. Following the arguments on [Cha02, p. 361], we can replace Φ by $\int_\Gamma \Phi(\operatorname{Ad} g^{-1}\cdot)\,dg \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ for a suitable compact subset $\Gamma \subseteq G(\mathbb{A})^1$ (depending on T), and consider instead of the integral above the sum over P_1, R, P_2 with $P_1 \subseteq R \subseteq P_2$, and $\mathfrak{o} \in \mathcal{O}_*$ of

$$\int_{A_1^G} e^{-2\rho_0(H_0(a))}\sigma_{P_1}^{P_2}(H_0(a)-T)\cdot$$
$$\sum_{X\in\tilde{m}_{P_1}^R(\mathbb{Q})\cap\mathfrak{o}} \left| \sum_{P:\ R\subseteq P\subseteq P_2} (-1)^{\dim A_P/A_G} \sum_{Y\in\mathfrak{u}_R^P(\mathbb{Q})} \int_{\mathfrak{u}_P(\mathbb{A})} \Phi(\lambda \operatorname{Ad} a^{-1}(X+Y+U))\,dU \right| da. \tag{17}$$

We now distinguish the cases $R = P_2$ and $R \subsetneq P_2$. For $R = P_2$, (17) equals the sum over $P_1 \subseteq P_2$ of

$$\int_{A_1^G} e^{-2\rho_0(H_0(a))}\sigma_{P_1}^{P_2}(H_0(a)-T) \sum_{\substack{X\in\tilde{\mathfrak{m}}_{P_1}^{P_2}(\mathbb{Q})\\ X\notin\mathfrak{n}}} \left| \int_{\mathfrak{u}_{P_2}(\mathbb{A})} \Phi(\lambda \operatorname{Ad} a^{-1}(X+U))\,dU \right| da$$
$$= \lambda^{-\dim \mathfrak{u}_{P_2}} \int_{A_1^G} e^{-2(\rho_0-\rho_2)(H_0(a))}\sigma_{P_1}^{P_2}(H_0(a)-T) \sum_{\substack{X\in\tilde{\mathfrak{m}}_{P_1}^{P_2}(\mathbb{Q})\\ X\notin\mathfrak{n}}} \left|\Psi_{P_2}(\lambda \operatorname{Ad} a^{-1}X)\right| da, \tag{18}$$

where $\Psi_{P_2}(Y) := \int_{\mathfrak{u}_{P_2}(\mathbb{A})} \Phi(Y+U)\,dU \in \mathcal{S}^\nu(\mathfrak{m}_2(\mathbb{A}))$.

For $R \subsetneq P_2$, we apply Poisson summation with respect to the sum over Y. In the resulting alternating sum many terms cancel out as explained in [Cha02, pp. 362–363]. So the sum over $R \subsetneq P_2$ of (17) can be bounded by the sum over $P_1, R, P_2, P_1 \subseteq R \subsetneq P_2$, of

$$\int_{A_1^G} e^{-2\rho_0(H_0(a))} \sigma_{P_1}^{P_2}(H_0(a) - T) \cdot$$
$$\sum_{\substack{X \in \tilde{\mathfrak{m}}_{P_1}^R(\mathbb{Q}) \\ X \notin \mathfrak{n}}} \left| \sum_{\bar{Y} \in \overline{\mathfrak{u}}_R^{P_2\prime}(\mathbb{Q})} \int_{\mathfrak{u}_R(\mathbb{A})} \Phi(\lambda \operatorname{Ad} a^{-1}(X + U)) \psi(\langle U, \bar{Y} \rangle) \, dU \right| da. \tag{19}$$

For our purposes, we can replace Φ by Lemma 3.14 by the product $\Psi_{\mathfrak{m}_R} \Psi_{\mathfrak{u}_R}$ with $\Psi_{\mathfrak{m}_R} \in \mathcal{S}(\mathfrak{m}_R(\mathbb{A}))$, $\Psi_{\mathfrak{u}_R} \in \mathcal{S}(\mathfrak{u}_R(\mathbb{A}))$, $\Psi_{\mathfrak{m}_R}, \Psi_{\mathfrak{u}_R} \geq 0$, satisfying the inequalities of Lemma 3.14.

Changing variables, we may consider instead of (19) the integral

$$\lambda^{-\dim \mathfrak{u}_R} \int_{A_1^G} e^{-2(\rho_0 - \rho_R)(H_0(a))} \sigma_{P_1}^{P_2}(H_0(a) - T) \cdot$$
$$\sum_{\substack{X \in \tilde{\mathfrak{m}}_{P_1}^R(\mathbb{Q}) \\ X \notin \mathfrak{n}}} \Psi_{\mathfrak{m}_R}(\lambda \operatorname{Ad} a^{-1} X) \cdot \sum_{\bar{Y} \in \overline{\mathfrak{u}}_R^{P_2\prime}(\mathbb{Q})} \int_{\mathfrak{u}_R(\mathbb{A})} \Psi_{\mathfrak{u}_R}(U) \psi(\langle U, \lambda^{-1} \operatorname{Ad} a^{-1} \bar{Y} \rangle) \, dU \, da. \tag{20}$$

The compact support of Φ at the finite places implies the existence of $N \in \mathbb{N}$ such that all contributing $\bar{Y}$ and X must have coordinates in $\frac{1}{N}\mathbb{Z}$. Let $m \geq 0$ be a sufficiently large even integer. By standard estimates for Schwartz–Bruhat functions,

$$\begin{aligned} &\left| \int_{\mathfrak{u}_R(\mathbb{A})} \Psi_{\mathfrak{u}_R}(U) \psi(\langle U, \lambda^{-1} \operatorname{Ad} a^{-1} \bar{Y} \rangle) \right| dU \\ &\quad \leq \|\lambda^{-1} \operatorname{Ad} a^{-1} \bar{Y}\|^{-m} \sum_{D \in \mathcal{B}_{m/2}} \int_{\mathfrak{u}_R(\mathbb{A})} |(D\Psi_{\mathfrak{u}_R})(U)| \, dU \\ &\quad =: \|\lambda^{-1} \operatorname{Ad} a^{-1} \bar{Y}\|^{-m} \mu_{\mathfrak{u}_R}^m(\Psi_{\mathfrak{u}_R}). \end{aligned}$$

This last sum over the set of differential operators defines the seminorm $\mu_{\mathfrak{u}_R}^m$ on $\mathcal{S}(\mathfrak{u}_R(\mathbb{A}))$ which is continuous when restricted to $\mathcal{S}(\mathfrak{u}_R(\mathbb{R}))$ for fixed non-Archimedean test function. Hence (20) is bounded by

$$\lambda^{m-\dim \mathfrak{u}_R}\mu_{\mathfrak{u}_R^m}(\Psi_{\mathfrak{u}_R})\int_{A_1^G} e^{-2(\rho_0-\rho_R)(H_0(a'a))}\sigma_{P_1}^{P_2}(H_0(a)-T)\cdot$$
$$\Bigg(\sum_{\bar{Y}\in\bar{\mathfrak{u}}_R^{P_2\prime}(\frac{1}{N}\mathbb{Z})}\|\operatorname{Ad}a^{-1}\bar{Y}\|^{-m}\Bigg)\Bigg(\sum_{\substack{X\in\tilde{\mathfrak{m}}_{P_1}^R(\frac{1}{N}\mathbb{Z})\\ X\notin\mathfrak{n}}}\left|\Psi_{\mathfrak{m}_R}(\lambda\operatorname{Ad}a^{-1}X)\right|\Bigg)\,da. \tag{21}$$

Write $\mathfrak{m}_R=\bigoplus_{\beta\in\Phi_1^R}\mathfrak{m}_{R,\beta}$ for the eigenspace decomposition of $\mathfrak{m}_R$ with respect to $\Phi_1^R=\Phi(A_1,M_R)$. In particular, $\mathfrak{m}_{R,0}=\mathfrak{m}_1$. By Lemma 3.14 there are $\varphi_\beta\in\mathcal{S}(\mathfrak{m}_{R,\beta}(\mathbb{A}))$, $\varphi_\beta\geq 0$, such that $|\Psi_{\mathfrak{m}_R}(Z)|\leq\prod_{\beta\in\Phi(A_0,M_R)}\varphi_\beta(Z_\beta)$ for all $Z=\sum_\beta Z_\beta\in\mathfrak{m}_R(\mathbb{A})=\bigoplus_\beta\mathfrak{m}_{R,\beta}$, and such that they satisfy the estimates of Lemma 3.14. With this, (21) is bounded by

$$\lambda^{m-\dim \mathfrak{u}_R}\mu_{\mathfrak{u}_R}(\Psi_{\mathfrak{u}_R})\int_{A_1^G} e^{-2(\rho_0-\rho_R)(H_0(a))}\sigma_{P_1}^{P_2}(H_0(a)-T)\cdot$$
$$\Bigg(\sum_{\bar{Y}\in\bar{\mathfrak{u}}_R^{P_2\prime}(\frac{1}{N}\mathbb{Z})}\|\operatorname{Ad}a^{-1}\bar{Y}\|^{-m}\Bigg)\Bigg(\sum_{\substack{X\in\tilde{\mathfrak{m}}_{P_1}^R(\frac{1}{N}\mathbb{Z})\\ X\notin\mathfrak{n}}}\prod_{\beta\in\Phi_1^R}\varphi_\beta(\lambda\beta(a)^{-1}X_\beta)\Bigg)\,da. \tag{22}$$

If $m>\dim\mathfrak{u}_R^{P_2}$, then by Lemma 3.10 there are $c_1>0$, and real numbers $k_\alpha\geq 0$ for $\alpha\in\Delta_0^2$ with $k_\alpha>0$ whenever $\alpha\in\Delta_0^2\backslash\Delta_0^R$, such that

$$\sum_{\bar{Y}\in\bar{\mathfrak{u}}_R^{P_2\prime}(\frac{1}{N}\mathbb{Z})}\|\operatorname{Ad}a^{-1}\bar{Y}\|^{-m}\leq c_1\prod_{\alpha\in\Delta_0^2}e^{-k_\alpha\alpha(H_0(a))}. \tag{23}$$

Setting $\sum_{\bar{Y}\in\bar{\mathfrak{u}}_R^{P_2\prime}(\frac{1}{N}\mathbb{Z})}\|\operatorname{Ad}a^{-1}\bar{Y}\|^{-m}:=1$ and $m=0$ in the case $P_2=R$, we can consider the cases $P_2=R$ and $R\subsetneq P_2$ together.

To the second product in (22) we apply Lemma 3.11. This we are allowed to, since $\sigma_{P_1}^{P_2}(H_0(a)-T)\neq 0$ implies that $a\in A_1^2(T_1)$. Thus (21) is bounded by a finite sum of terms of the form

$$c'\lambda^{-N+m-\dim\mathfrak{u}_R}\int_{A_1^G}\sigma_{P_1}^{P_2}(H_1(a)-T)\prod_{\alpha\in\Delta_0^2\backslash\Delta_0^1}e^{-l_\alpha\alpha(H_0(a))}\,da \tag{24}$$

for all $\lambda\geq 1$ and all $N>0$, where $l_\alpha>0$ and $c'>0$ are constants depending only on N. By Lemma 3.13 the second integral is finite. Thus (7) is proven.

(ii) Now assume that $\lambda\in(0,1]$. We essentially argue as above, but have to change the upper bounds for the two products occurring in the integral (22). We apply Lemma 3.10 to bound the left-hand side of (23) again by the same quantity as before. To bound the last term in the integral in (22), we use Lemma 3.11 giving for this term an upper bound of

$$\lambda^{-\dim \mathfrak{m}_R} \delta_0^R(a) \prod_{\alpha \in \Delta_0^R \setminus \Delta_0^1} e^{-k_\alpha \alpha(H_0(a))}$$

times the value of some seminorm applied to the φ_μ's. Hence (21) is bounded by the product of the value of a seminorm (depending on m) applied to Φ_∞ with

$$\lambda^{m-\dim \mathfrak{u}_R - \dim \mathfrak{m}_R}$$

with $m > \dim \mathfrak{u}_R^{P_2}$ arbitrary if $R \neq P_2$ and $m = 0$ if $R = P_2$, and

$$\int_{A_1^G} \sigma_{P_1}^{P_2}(H_0(a) - T) \prod_{\tilde{\alpha} \in \Delta_1^2} e^{-l'_\alpha \tilde{\alpha}(H_0(a))}\, da$$

for suitable $l'_\alpha > 0$. Since (for $P_2 = R$ as well as $R \neq P_2$)

$$\dim \mathfrak{u}_R - m + \dim \mathfrak{m}_R \leq \dim \mathfrak{g} = n^2 \tag{25}$$

the assertion (8) follows again from Lemma 3.13.

(iii) It is clear from the proof of the first part of the lemma that if ν is sufficiently large with respect to N, then the analogue assertion holds for $\Phi \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A}))$ instead of $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$.

□

Remark 3.15. In (25) we have $\dim \mathfrak{u}_R + \dim \mathfrak{m}_R \leq \dim \mathfrak{g} - 1$ unless $R = P_2 = G$, and if $R \subsetneq G$ we have $\dim \mathfrak{u}_R + \dim \mathfrak{m}_R \leq \dim \mathfrak{g} - 2$ unless $n = 2$.

4 Nilpotent Auxiliary Distributions

Recall that $\mathfrak{n} \subseteq \mathfrak{g}(\mathbb{Q})$ denotes the set of nilpotent elements. Under the action of $G(\mathbb{Q})$ it decomposes into finitely many orbits which we denote by $\mathcal{N} \subseteq \mathfrak{n}$. If $\mathcal{N} \neq 0$ and $X_0 \in \mathcal{N}$, X_0 can be embedded into an $\mathfrak{sl}_2$-triple $\{X_0, Y_{X_0}, H_{X_0}\} \subseteq \mathfrak{g}$ with H_{X_0} semisimple and Y_{X_0} nilpotent. The element H_{X_0} defines a grading on $\mathfrak{g}$, $\mathfrak{g} = \bigoplus_{i \in \mathbb{Z}} \mathfrak{g}_i$ with $\mathfrak{g}_i = \{X \in \mathfrak{g} \mid [H_{X_0}, X] = iX\}$ and $X_0 \in \mathfrak{g}_2$. We set $\mathfrak{p}_{X_0} = \bigoplus_{i \geq 0} \mathfrak{g}_i$, which is the associated Jacobson–Morozov parabolic subalgebra, $\mathfrak{u}_{X_0} = \bigoplus_{i>0} \mathfrak{g}_i$, $\mathfrak{u}_{X_0}^{>j} = \bigoplus_{i>j} \mathfrak{g}_i$, and $\mathfrak{u}_{X_0}^{\geq j} = \bigoplus_{i \geq j} \mathfrak{g}_i$. Correspondingly, let $P_{X_0} = M_{X_0} U_{X_0} \subseteq G$ be the Jacobson–Morozov parabolic subgroup with Lie algebra $\mathfrak{p}_{X_0}$ and Levi part M_{X_0} with Lie algebra $\mathfrak{m}_{X_0} = \mathfrak{g}_0$, and unipotent radical U_{X_0} with Lie algebra $\mathfrak{u}_{X_0}$. The representative X_0 of $\mathcal{N}$ can be chosen such that P_{X_0} is a standard parabolic subgroup and $H_{X_0} \in \mathfrak{a}_{M_{X_0}}$. If $\mathcal{N} = 0$, then $X_0 = 0$, and we set $H_{X_0} = 0$, $P_{X_0} = G$.

We have a decomposition

$$\mathcal{N} = \bigcup_{\delta \in P_{X_0}(\mathbb{Q}) \backslash G(\mathbb{Q})} \operatorname{Ad} \delta^{-1} \cdot \mathfrak{u}_{X_0}^{\geq 2}(\mathbb{Q})$$

(disjoint union), and the action of M_{X_0} on $\mathfrak{u}_{X_0}^2 = \mathfrak{g}_2$ defines a prehomogeneous vector space, i.e. the orbit $V_0 := \operatorname{Ad} M_{X_0} X_0 \subseteq \mathfrak{g}_2$ is open and dense. We shall write

$$\mathfrak{u}_{\mathcal{N}}^{2,\mathrm{reg}}(\mathbb{Q}) = \mathcal{N} \cap \mathfrak{g}_2(\mathbb{Q})$$

or just $\mathfrak{u}^{2,\mathrm{reg}}(\mathbb{Q})$ if $\mathcal{N}$ is clear from the context.

Let $C_{M_{X_0}}(X_0) = \{m \in M_{X_0} \mid \operatorname{Ad} m^{-1} X_0 = X_0\}$ be the stabiliser of X_0 under the action of M_{X_0}, and $C_{U_{X_0}}(X_0) = \{u \in U_{X_0} \mid \operatorname{Ad} u^{-1} X_0 = X_0\}$ the stabiliser of X_0 in U_{X_0}. If there is no danger of confusion, we drop the subscript X_0 and write $H = H_{X_0}$, $P = P_{X_0}$, etc.

Note that for every $\lambda \in \mathbb{R}_{>0}$ we have

$$\operatorname{Ad}(\eta_{X_0,\lambda}) X_0 = \lambda X_0, \quad \text{where} \quad \eta_{X_0,\lambda} := e^{\frac{\log \lambda}{2} H_{X_0}} \in Z^{M_{X_0}}(\mathbb{A})$$

where $Z^{M_{X_0}}$=center of M_{X_0}.

Remark 4.1. If $X \in \mathfrak{u}^{2,\mathrm{reg}}(\mathbb{Q})$ and if $\{X, H_X, Y_X\}$ is the associated $\mathfrak{sl}_2$-triple, then $H_X = H_{X_0}$.

Example 4.2. For the cases $n = 2$ and $n = 3$ we list our choice of Jacobson–Morozov parabolics and their relevant properties:

- $n = 2$. There are two nilpotent orbits, the trivial orbit $\mathcal{N}_{\mathrm{triv}}$ and the regular one $\mathcal{N}_{\mathrm{reg}}$:

$\mathcal{N}$	X_0	H_{X_0}	P_{X_0}	$C_U(X_0)$	$\mathfrak{u}^{2,\mathrm{reg}}$
$\mathcal{N}_{\mathrm{triv}}$	0	0	G	$\{\mathbf{1}_2\}$	$\{0\}$
$\mathcal{N}_{\mathrm{reg}}$	$\begin{pmatrix} 0 & x_0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$	P_0	U_0	$\left\{\begin{pmatrix} 0 & x \\ 0 & 0 \end{pmatrix} \mid x \neq 0\right\}$

  where $x_0 \in \mathbb{Q}$ is any non-zero element.
- $n = 3$. There are three nilpotent orbits, the trivial orbit $\mathcal{N}_{\mathrm{triv}}$, the minimal (=subregular) one $\mathcal{N}_{\min}$, and the regular one $\mathcal{N}_{\mathrm{reg}}$: where $x_0, y_0 \in \mathbb{Q}$ are any non-zero elements.

In all of these examples we fix measures on $C_U(X_0, \mathbb{A})$ and $C_M(X_0, \mathbb{A})$ in the obvious way.

$\mathcal{N}$	X_0	H_{X_0}	P_{X_0}	$C_U(X_0)$	$\mathfrak{u}^{2,\mathrm{reg}}$
$\mathcal{N}_{\mathrm{triv}}$	0	0	G	$\{\mathbf{1}_2\}$	$\{0\}$
$\mathcal{N}_{\min}$	$\begin{pmatrix} 0 & 0 & x_0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & & \\ & 0 & \\ & & -1 \end{pmatrix}$	P_0	U_0	$\left\{\begin{pmatrix} 0 & 0 & x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mid x \neq 0\right\}$
$\mathcal{N}_{\mathrm{reg}}$	$\begin{pmatrix} 0 & x_0 & 0 \\ 0 & 0 & y_0 \\ 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & & \\ & 0 & \\ & & -2 \end{pmatrix}$	P_0	$\left\{\begin{pmatrix} 1 & x_1 & x_2 \\ 0 & 1 & x_3 \\ 0 & 0 & 1 \end{pmatrix} \mid x_1 = \frac{x_0}{y_0} x_3\right\}$	$\left\{\begin{pmatrix} 0 & x & 0 \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} \mid x \cdot y \neq 0\right\}$

The following is a slight variant of [RR72, Theorem 1].

Lemma 4.3. *There exists a constant $c > 0$ such that for every function $f : \overline{V_0(\mathbb{A})} = \mathfrak{g}_2(\mathbb{A}) \longrightarrow \mathbb{C}$, which is integrable and for which all occurring integrals are finite, we have*

$$\int_{C_M(X_0,\mathbb{A})\backslash M(\mathbb{A})} f(\operatorname{Ad} m^{-1} X_0)\delta_{U^{\le 2}}(m)^{-1} da = c \int_{V_0(\mathbb{A})} \varphi(X) f(X) dX,$$

where $\varphi : \mathfrak{g}_2(\mathbb{A}) \longrightarrow \mathbb{C}$ is defined as follows: Let $Z_1, \ldots, Z_r$ be a basis of $\mathfrak{g}_1$, and $Z_1', \ldots, Z_r'$ a basis of $\mathfrak{g}_{-1}$, which are dual to each other with respect to the Killing form. For $X \in \mathfrak{g}_2$ write $[X, Z_i'] = \sum c_{ji}(X) Z_j$, and set $\varphi(X) = |\det(c_{ij}(X))_{i,j}|^{\frac{1}{2}}$.

Example 4.4. For $\mathcal{N}$ the trivial or regular orbit from Example 4.2, we have $\mathfrak{g}_1 = 0 = \mathfrak{g}_{-1}$ so that $\varphi(X) \equiv 1$. If $n = 3$ and $\mathcal{N} = \mathcal{N}_{\min}$, then $\mathfrak{g}_1 = \{\begin{pmatrix} 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & 0 \end{pmatrix}\}$ and $\varphi(\begin{pmatrix} 0 & 0 & x \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}) = |x|$.

Proof of Lemma 4.3. Let $X \in \mathfrak{g}_2(\mathbb{A})$ and $m \in M(\mathbb{A})$. Then φ transforms according to [RR72, Lemma 2] via $\varphi(\operatorname{Ad} mX) = |\det \operatorname{Ad} m_{|\mathfrak{g}_1}| \varphi(X) = \delta_{\mathfrak{g}_1}(m)\varphi(X)$. Let

$$\Lambda_1(f) = \int_{C_M(X_0,\mathbb{A})\backslash M(\mathbb{A})} f(\operatorname{Ad} m^{-1} X_0)\delta_{U^{\le 2}}(m)^{-1} dm, \quad \text{and}$$

$$\Lambda_2(f) = \int_{V_0(\mathbb{A})} \varphi(X) f(X) dX.$$

Let $m_0 \in M(\mathbb{A})$ and put $f^{m_0}(X) = f(\operatorname{Ad} m_0^{-1} X)$. Then

$$\begin{aligned} \Lambda_1(f^{m_0}) &= \int_{C_M(X_0,\mathbb{A})\backslash M(\mathbb{A})} f(\operatorname{Ad} m_0{}^{-1} \operatorname{Ad} m^{-1} X_0)\delta_{U^{\le 2}}(m)^{-1} dm \\ &= \int_{C_M(X_0,\mathbb{A})\backslash M(\mathbb{A})} f(\operatorname{Ad}(m m_0)^{-1} X_0)\delta_{U^{\le 2}}(m)^{-1} dm = \delta_{U^{\le 2}}(m_0)\Lambda_1(f), \end{aligned}$$

and, using the above transformation property of φ,

$$\Lambda_2(f^{m_0}) = \int_{V_0(\mathbb{A})} \varphi(X) f(\operatorname{Ad} m_0^{-1} X) dX = \delta_{\mathfrak{g}_2}(m_0) \int_{V_0(\mathbb{A})} \varphi(\operatorname{Ad} m_0 X) f(X) dX$$

$$= \delta_{\mathfrak{g}_2}(m_0)\delta_{\mathfrak{g}_1}(m_0) \int_{V_0(\mathbb{A})} \varphi(X) f(X) dX = \delta_{U^{\leq 2}}(m_0) \int_{V_0(\mathbb{A})} \varphi(X) f(X) dX.$$

□

We need to attach certain auxiliary distributions to the nilpotent orbit $\mathcal{N}$, namely,

$$j_{\mathcal{N}}^T, \tilde{j}_{\mathcal{N}}^T : \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A})) \cup \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A})) \longrightarrow \mathbb{C}$$

(ν sufficiently large as in Lemma 3.7). The first distribution is defined by (for the definition of $\tilde{j}_{\mathcal{N}}^T$ see Definition 4.6 below)

$$j_{\mathcal{N}}^T(\Phi) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} F(x,T) \sum_{\gamma \in \mathcal{N}} \Phi(\operatorname{Ad} x^{-1}\gamma) dx.$$

This integral is absolutely convergent since the integral with the sum over $\gamma \in \mathcal{N}$ replaced by $\gamma \in \mathfrak{g}(\mathbb{Q})$ is already absolutely convergent (cf. [Cha02, Art85]).

Proposition 4.5. *There exists $\nu > 0$ depending only on n such that the following holds. For every nilpotent orbit $\mathcal{N}$ there is a distribution $J_{\mathcal{N}}^T : \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A})) \cup \mathcal{S}_\nu(\mathfrak{g}(\mathbb{A})) \longrightarrow \mathbb{C}$ such that*

$$J_{\mathfrak{n}}^T(\Phi) = \sum_{\mathcal{N}} J_{\mathcal{N}}^T(\Phi).$$

Moreover, $J_{\mathcal{N}}^T(\Phi)$ is a polynomial in T of degree at most $\dim \mathfrak{a}$*, and there exist $c > 0$ and for fixed $\Phi_f \in \mathcal{S}(\mathbb{A}_f))$ a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$ such that*

$$\left|J_{\mathcal{N}}^T(\Phi) - j_{\mathcal{N}}^T(\Phi)\right|$$
$$= \left|J_{\mathcal{N}}^T(\Phi) - \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} F(x,T) \sum_{\gamma \in \mathcal{N}} \Phi(\operatorname{Ad} x^{-1}\gamma) dx\right| \leq \mu(\Phi_\infty) e^{-c\|T\|}$$

for all sufficiently regular $T \in \mathfrak{a}^+$ with $d(T) \geq \delta\|T\|$.

Proof. The assertion is the analogue to [Art85, Theorem 4.2] where it is stated for smooth compactly supported functions on the group $G(\mathbb{A})$. Large parts of the proof of [Art85, Theorem 4.2] carry over to our situation, we have, however, to take into account that our test function is not compactly supported anymore. We define an auxiliary function similar as in [Art85]: Let $\mathcal{N}$ be a nilpotent orbit and let $\varepsilon > 0$ be given. Let $q_1, \ldots, q_r$ be polynomials on $\mathfrak{g}$ such that $\overline{\mathcal{N}} = \{X \in \mathfrak{g} \mid q_1(X) = \ldots = q_r(X) = 0\}$. We can choose $q_1, \ldots, q_r$ with coefficients in $\mathbb{Q}$. Let $\rho_\infty : \mathbb{R} \longrightarrow \mathbb{R}$ be a non-negative smooth function with support in $[-1, 1]$ which identically equals 1 on $[-1/2, 1/2]$ and such that $0 \leq \rho_\infty \leq 1$. Define

$$\Phi^\varepsilon_{\mathcal{N}}(X) = \Phi(X)\rho_\infty(\varepsilon^{-1}|q_1(X)|_\infty) \cdot \ldots \cdot \rho_\infty(\varepsilon^{-1}|q_r(X)|_\infty)$$

so that $\Phi^\varepsilon_{\mathcal{N}} = \Phi$ in a neighbourhood of $\overline{\mathcal{N}}$. It follows from the proof of [Art85, Theorem 4.2] that it suffices to show the analogue of [Art85, Lemma 4.1], namely that

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} F(x, T) \sum_{X \in \mathfrak{n}\backslash\overline{\mathcal{N}}} |\Phi^\varepsilon_{\mathcal{N}}(\operatorname{Ad} x^{-1}X)|dx \leq \mu(\Phi_\infty)\varepsilon^a(1 + \|T\|)^{\dim \mathfrak{a}} \tag{26}$$

for a suitable seminorm μ, and a suitable number $a > 0$. Hence, using (4), we need to bound (after integrating Φ over a compact subset)

$$\int_{A_0^G(T_1)} \delta_0(a)^{-1}F(a, T) \sum_{X \in \mathfrak{n}\backslash\overline{\mathcal{N}}} |\Phi^\varepsilon_{\mathcal{N}}(\operatorname{Ad} a^{-1}X)|da.$$

It suffices to take the sum over $X \in \mathfrak{g}(\mathbb{Q})\backslash\overline{\mathcal{N}}$. Moreover, since Φ is smooth and compactly supported at the non-Archimedean places, there exists $N > 0$ such that we can take the sum instead over points with entries in $\frac{1}{N}\mathbb{Z}$ and replace Φ by its Archimedean part Φ_∞ as Φ_f stays fixed. For $R > 0$ define a function $\Phi_R(X) := \Phi_\infty(X)\rho_\infty(R^{-1}\|X_\infty\|)$ so that the support of Φ_R is compact and contained in $\{X \in \mathfrak{g}(\mathbb{R}) \mid \|X_\infty\| \leq R\}$, and $\Phi_R(X) = \Phi_\infty(X)$ if $\|X_\infty\| \leq R/2$. Moreover, if $D \in \mathcal{U}(\mathfrak{g})$ denotes an element of degree $k \leq \nu$, then there exists a constant $c_D > 0$ depending only on D and ρ_∞ such that

$$\|D\Phi_R\|_{L^1(\mathfrak{g}(\mathbb{R}))} \leq c_D \sum_{Y \in \mathcal{B}_{\mathfrak{g},\nu}} \|Y\Phi_\infty\|_{L^1(\mathfrak{g}(\mathbb{R}))} = c_D\|\Phi_\infty\|_{0,\nu,1}.$$

It follows from the proof of [Art85, Lemma 4.1] that there exist constants $r, a_0, c > 0$ depending only on n such that if ν is sufficiently large (in a sense depending only on n),

$$\begin{aligned}\int_{A_0^G(T_1)} \delta_0(a)^{-1}F(a, T) &\sum_{X \in \mathfrak{g}(\frac{1}{N}\mathbb{Z})\backslash\overline{\mathcal{N}}} |(\Phi_R)^\varepsilon_{\mathcal{N}}(\operatorname{Ad} a^{-1}X)|\, da \\ &\leq cR^{a_0}\|\Phi_\infty\|_{0,\nu,1}\varepsilon^r(1 + \|T\|)^{\dim \mathfrak{a}}\end{aligned}$$

for every $R \geq 1$, since the support of Φ_R is compact and contained in the ball of radius R around $0 \in \mathfrak{g}(\mathbb{R})$. In particular, if $1 \leq R_1 \leq R_2$, we get

$$\begin{aligned}\int_{A_0^G(T_1)} \delta_0(a)^{-1}F(a, T) &\sum_{X \in \mathfrak{g}(\frac{1}{N}\mathbb{Z})\backslash\overline{\mathcal{N}}} |\left(\Phi_{R_1} - \Phi_{R_2}\right)^\varepsilon_{\mathcal{N}}(\operatorname{Ad} a^{-1}X)|\, da \\ &\leq cR_2^{a_0}\mu_\nu^{R_1}(\Phi_\infty)\varepsilon^r(1 + \|T\|)^{\dim \mathfrak{a}},\end{aligned}$$

where

$$\mu_\nu^{R_1}(\Phi_\infty) := \sum_{Y \in \mathcal{B}_{\mathfrak{g},\nu}} \int_{\mathfrak{g}(\mathbb{R})\setminus B_{R_1}} |(Y\Phi_\infty)(X)| dX$$

for $B_{R_1} := \{X \in \mathfrak{g}(\mathbb{R}) \mid \|X\| < R_1\}$. Let $M \in \mathbb{Z}_{>0}$ and suppose $\nu > M$. Since $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R})) \cup \mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$, there exists $C_M > 0$ and $k_M, l_M \geq 0$ such that (possibly enlarging ν accordingly)

$$\mu_\nu^{R_1}(\Phi_\infty) \leq C_M R_1^{-N} \|\Phi_\infty\|_{k_M,l_M,1}.$$

Fix $M > a_0$. By definition $|\Phi_\infty - \Phi_{2^i}| \leq 2\sum_{j\geq i-1} |\Phi_{2^{j+2}} - \Phi_{2^j}|$ so that for every $i > 0$, we get

$$\begin{aligned}
&\int_{A_0^G(T_1)} \delta_0(a)^{-1} F(a,T) \sum_{X \in \mathfrak{g}(\frac{1}{N}\mathbb{Z})\setminus \mathcal{N}} |\big(\Phi_\infty - \Phi_{2^i}\big)^{\varepsilon}_{\mathcal{N},\nu}(\operatorname{Ad} a^{-1}X)| \, da \\
&\leq c_M \sum_{j\geq i-1} 2^{(a_0-M)j} \|\Phi_\infty\|_{k_M,l_M,1} \varepsilon^r (1+\|T\|)^{\dim \mathfrak{a}} \\
&= c'_M 2^{(a_0-M)(i-1)} \|\Phi_\infty\|_{k_M,l_M,1} \varepsilon^r (1+\|T\|)^{\dim \mathfrak{a}}
\end{aligned}$$

for $c_M, c'_M > 0$ suitable constants. Hence if we fix an arbitrary integer $i > 0$, we get

$$\begin{aligned}
&\int_{A_0^G(T_1)} \delta_0(a)^{-1} F(a,T) \sum_{X \in \mathfrak{g}(\frac{1}{N}\mathbb{Z})\setminus \mathcal{N}} |\Phi^{\varepsilon}_{\mathcal{N}}(\operatorname{Ad} a^{-1}X)| \, da \\
&\qquad \leq c\big(2^{(a_0-M)(i-1)} \|\Phi_\infty\|_{k_M,l_M,1} + 2^{ia_0} \|\Phi_\infty\|_{0,\nu}\big) \varepsilon^r (1+\|T\|)^{\dim \mathfrak{a}}
\end{aligned}$$

for a suitable constant $c > 0$ proving the inequality (26). Taking $M = a_0 + 1$ (which only depends on n) and $\nu > a_0 + 1$ also proves the assertion about the existence of ν. □

Definition 4.6. If $T \in \mathfrak{a}^+$ is sufficiently regular, we set

$$\tilde{j}^T_{\mathcal{N}}(\Phi) = \int_{A_G M(\mathbb{Q})\setminus M(\mathbb{A})} \int_{\mathfrak{u}^{>2}(\mathbb{A})} \tilde{F}^M(m,T) \sum_{\gamma \in \mathfrak{u}_{\mathcal{N}}^{2,\mathrm{reg}}(\mathbb{Q})} \delta_U(m)^{-1} \Phi(\operatorname{Ad} m^{-1}(\gamma+X)) \, dX \, dm$$

where the truncation function $\tilde{F}^M(\cdot, T) : G(\mathbb{A}) \longrightarrow \mathbb{C}$ is defined as the characteristic function of the set of all $x \in G(\mathbb{A})$ of the form $x = umk$, $m \in M(\mathbb{A})$, $u \in U(\mathbb{A})$, $k \in \mathbf{K}$, satisfying

$$\forall \varpi \in \widehat{\Delta}_0 \ \forall \gamma \in M(\mathbb{Q}) : \ \varpi(H_0(\gamma m) - T) \leq 0.$$

Note that $\tilde{F}^M(umk, T) = \tilde{F}^M(m,T) = F^M(m,T)\hat{\tau}_P(T - H_0(m))$.

From now on we assume $n \le 3$. However, the main result of the following, namely Theorem 5.7, should basically stay true for general n, cf. Remark 5.9.

We shall further assume that any test function Φ is invariant by $\mathrm{Ad}\, k$ for all $k \in \mathbf{K}$. In particular, the function $\tilde{F}^M(\cdot, T)$ simplifies in our situation. If $\mathcal{N} = \mathcal{N}_{\mathrm{triv}} = 0$ is the trivial orbit, we have $\tilde{F}^M(m, T) = \tilde{F}^G(m, T) = F(m, T)$ and

$$\tilde{j}^T_{\mathcal{N}_{\mathrm{triv}}}(\Phi) = \mathrm{vol}^T(G(\mathbb{Q})\backslash G(\mathbb{A})^1)\Phi(0)$$

where $\mathrm{vol}^T(G(\mathbb{Q})\backslash G(\mathbb{A})^1) = \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} F(x, T)\, dx$ denotes the volume of the truncated quotient which satisfies

$$\lim_{T:\, d(T)\to\infty} \mathrm{vol}^T(G(\mathbb{Q})\backslash G(\mathbb{A})^1) = \mathrm{vol}(G(\mathbb{Q})\backslash G(\mathbb{A})^1).$$

On the other hand, the Jacobson–Morozov parabolic subgroup for any other occurring nilpotent orbit equals $P_0 = T_0 U_0$ so that $\tilde{F}^M(m, T) = \hat{\tau}_0(T - H_0(m))$ in all these cases. Hence in all the non-trivial cases we have

$$\begin{aligned}\tilde{j}^T_{\mathcal{N}}(\Phi) &= \int_{A_G T_0(\mathbb{Q})\backslash T_0(\mathbb{A})} \delta_0(t)^{-1}\hat{\tau}_0(T - H_0(t)) \int_{\mathfrak{u}^{>2}(\mathbb{A})} \sum_{\gamma\in\mathfrak{u}^{2,\mathrm{reg}}_{\mathcal{N}}(\mathbb{Q})} \Phi(\mathrm{Ad}\, t^{-1}(\gamma + X))\, dX\, dt \\ &= \int_{A_0^G} \delta_0(a)^{-1}\hat{\tau}_0(T - H_0(a)) \int_{\mathfrak{u}^{>2}(\mathbb{A})} \sum_{\gamma\in\mathfrak{u}^{2,\mathrm{reg}}_{\mathcal{N}}(\mathbb{Q})} \Phi(\mathrm{Ad}\, a^{-1}(\gamma + X))\, dX\, da\end{aligned}$$

where we used the invariance of Φ under $\mathrm{Ad}\, k$, $k \in \mathbf{K}$, for the equality. This expression is defined for any $T \in \mathfrak{a}$ so that we do not need to assume that T is sufficiently regular.

Lemma 4.7. *Let* $\nu > 0$ *and* $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$. *There exists a seminorm* μ *on* $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ *(resp.* $\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$*) such that for all* $\lambda > 0$ *and all* $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ *(resp.* $\Phi_\infty \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$*) we have (with* $\Phi = \Phi_\infty \cdot \Phi_f$*)*

$$\begin{aligned}\int_{A_G T_0(\mathbb{Q})\backslash T_0(\mathbb{A})} \delta_0(t)^{-1}\hat{\tau}_0(T - H_0(t)) \int_{\mathfrak{u}^{>2}(\mathbb{A})} \sum_{\gamma\in\mathfrak{u}^{2,reg}_{\mathcal{N}}(\mathbb{Q})} |\Phi(\mathrm{Ad}\, t^{-1}(\gamma + X))|\, dX\, dt \\ \le \lambda^{-\dim\mathcal{N}/2}(1 + |\log\lambda|)\mu(\Phi_\infty). \qquad (27)\end{aligned}$$

Moreover, we have

$$\tilde{j}^T_{\mathcal{N}}(\Phi_\lambda) = \lambda^{-\dim\mathcal{N}/2}\tilde{j}^{T-\frac{\log\lambda}{2}H}_{\mathcal{N}}(\Phi)$$

for $H = H_{X_0}$ *the semisimple element of the* $\mathfrak{sl}_2$*-triple attached to* $\mathcal{N}$ *as in Example 4.4.*

Proof. For $\mathcal{N} = 0$ there is nothing to show so that we assume that $\mathcal{N}$ is not the trivial orbit. Without loss of generality we may assume that $\Phi \geq 0$. Let $\Psi : \mathfrak{g}_2(\mathbb{A}) \longrightarrow \mathbb{C}$ be defined by $\Psi(\gamma) = \int_{\mathfrak{u}^{>2}(\mathbb{A})} \Phi(\gamma + X)\, dX$. Then the left-hand side of (27) equals

$$\lambda^{-\dim \mathfrak{u}^{>2}} \int_{A_G T_0(\mathbb{Q})\backslash T_0(\mathbb{A})} \delta_{U^{\leq 2}}(t)^{-1} \hat{\tau}_0(T - H_0(t)) \sum_{\gamma \in \mathfrak{u}^{2,\mathrm{reg}}_{\mathcal{N}}(\mathbb{Q})} \Psi(\lambda \operatorname{Ad} t^{-1}\gamma)\, dt$$

$$= \lambda^{-\dim \mathfrak{u}^{>2}} \int_{C_{T_0}(X_0,\mathbb{A})\backslash T_0(\mathbb{A})} \int_{A_G C_{T_0}(X_0,\mathbb{Q})\backslash C_{T_0}(X_0,\mathbb{A})} \delta_{U^{\leq 2}}(ts)^{-1} \hat{\tau}_0(T - H_0(ts)) \Psi(\lambda \operatorname{Ad} t^{-1} X_0)\, ds\, dt.$$

Now $\delta_{U^{\leq 2}}(ts)^{-1} = \delta_{U^1}(s)^{-1}\delta_{U^{\leq 2}}(t)^{-1}$. Let

$$t(X_0, \cdot) : V_0(\mathbb{A}) \longrightarrow C_{T_0}(X_0, \mathbb{A})\backslash T_0(\mathbb{A})$$

be the inverse of the map $t \mapsto X = \operatorname{Ad} t^{-1} X_0$. Then by Lemma 4.3 the above equals a constant multiple of

$$\lambda^{-\dim \mathfrak{u}^{>2}} \int_{V_0(\mathbb{A})} \varphi(X)\Psi(\lambda X) \int_{A_G C_{T_0}(X_0,\mathbb{Q})\backslash C_{T_0}(X_0,\mathbb{A})} \delta_{U^1}(s)^{-1} \hat{\tau}_0(T - H_0(t(X_0, X)s))\, ds\, dX.$$

Changing λX to Y we obtain

$$\lambda^{-\delta(\mathcal{N})} \int_{V_0(\mathbb{A})} \varphi(Y)\Psi(Y) \int_{A_G C_{T_0}(X_0,\mathbb{Q})\backslash C_{T_0}(X_0,\mathbb{A})} \delta_{U^1}(s)^{-1} \hat{\tau}_0(T - H_0(t(X_0, \lambda^{-1}Y)s))\, ds\, dY$$

where

$$\delta(\mathcal{N}) = \dim \mathfrak{u}^{>2} + \dim V_0 + \frac{1}{2} \dim \mathfrak{g}_1 = \dim \mathcal{N}/2. \tag{28}$$

If $\mathcal{N} = \mathcal{N}_{\mathrm{reg}}$ is the regular orbit, we have $\mathfrak{g}_1 = 0$ and $A_G C_{T_0}(X_0, \mathbb{Q})\backslash C_{T_0}(X_0, \mathbb{A}) = Z(\mathbb{Q})\backslash Z(\mathbb{A})^1$ so that we can bound the inner integral trivially by $\operatorname{vol}(\mathbb{Q}^\times\backslash\mathbb{A}^1) = 1$. Hence in this case we get the upper bound

$$\lambda^{-\dim \mathcal{N}/2} \int_{V_0(\mathbb{A})} \varphi(Y)\Psi(Y)\, dY = \lambda^{-\dim \mathcal{N}/2} \int_{V_0(\mathbb{A})} \Psi(Y)\, dY$$

which is bound by $\lambda^{-\dim \mathcal{N}/2}$ times some seminorm of Φ.

We are left with the case $n = 3$ and $\mathcal{N} = \mathcal{N}_{\min}$ the minimal orbit. In that case every $s \in Z(\mathbb{A}) C_{T_0}(X_0, \mathbb{Q})\backslash C_{T_0}(X_0, \mathbb{A})$ is of the form $s = \operatorname{diag}(a, a^{-2}, a)$ for $a \in \mathbb{Q}^\times\backslash\mathbb{A}^\times$, and $t \in C_{T_0}(X_0, \mathbb{A})\backslash T_0(\mathbb{A})$ can be represented by $t = \operatorname{diag}(b, 1, b^{-1})$, $b \in \mathbb{A}^\times$. In particular, $\delta_{U^1}(s) = 1$. Multiplying Y by λ^{-1} we get

$$H_0(t(X_0, \lambda^{-1}Y)s) = H_0(t(X_0, Y)) + \frac{1}{2} \log \lambda \cdot H + \log |a| \cdot (1, -2, 1),$$

for $H = H_{X_0}$ the semisimple element of the $\mathfrak{sl}_2$-triple we fixed above. Plugging this into the integral, we obtain an upper bounded of the form of a product of $\lambda^{-\delta(\mathcal{N})}(1 + |\log \lambda|)$ times some seminorm in Ψ (depending on T). Using this expression and converting the changes of variables back we also get the second claim. $\square$

5 Nilpotent Distributions, Continuation of $\Xi^T(s, \Phi)$ and Functional Equation

In this section we proof the main Theorem 5.7. We continue to assume that $n \leq 3$ but we shall comment on possible generalisations to $n > 3$ at the appropriate places.

Proposition 4.5 implies that to understand the nilpotent distribution $J_{\mathfrak{n}}^T$ it suffices to study the distributions $J_{\mathcal{N}}^T$ or $j_{\mathcal{N}}^T$. The homogeneity property of the distributions $\tilde{j}_{\mathcal{N}}^T(\Phi_\lambda)$ from Lemma 4.7 will also give a certain homogeneity of the distributions $J_{\mathcal{N}}^T(\Phi_\lambda)$. To prove this, we first need to show that $j_{\mathcal{N}}^T$ can be approximated by $\tilde{j}_{\mathcal{N}}^T$:

Proposition 5.1. *Let $\nu > 0$ be as in Lemma 3.7 and let $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$. There exists a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$), and $\varepsilon > 0$ such that for all $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\Phi_\infty \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$) we have (with $\Phi = \Phi_\infty \cdot \Phi_f$)*

$$\left| j_{\mathcal{N}}^T(\Phi) - \tilde{j}_{\mathcal{N}}^T(\Phi) \right| \leq \mu(\Phi_\infty) e^{-\varepsilon \|T\|} \tag{29}$$

for all sufficiently regular $T \in \mathfrak{a}^+$ with $d(T) \geq \delta \|T\|$.

We postpone the proof of this proposition to Appendix 1.

Remark 5.2. For $n > 3$ this proposition is expected to stay valid at least for certain types of nilpotent orbits. For the regular (or "regular by blocks") nilpotent orbits, see also [CL15] for related results (in the function field case).

Corollary 5.3. *Let $I \subseteq \mathbb{R}_{>0}$ be a compact interval and let ν be as before. Let $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$. Then:*

(i) There exists a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$) and a constant $\varepsilon > 0$ such that for all $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\Phi_\infty \mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$) and all sufficiently regular $T \in \mathfrak{a}^+$ with $d(T) \geq \delta \|T\|$ we have

$$\left| J_{\mathcal{N}}^T(\Phi) - \tilde{j}_{\mathcal{N}}^T(\Phi) \right| \leq \mu(\Phi_\infty) e^{-\varepsilon \|T\|}$$

for every nilpotent orbit $\mathcal{N}$.

(ii) For every $T \in \mathfrak{a}^+$ such that T and $T - \frac{\log \lambda}{2} H_{X_0}$ are sufficiently regular for all $\lambda \in I$, we have

$$J_{\mathcal{N}}^T(\Phi_\lambda) = \lambda^{-\delta(\mathcal{N})} J_{\mathcal{N}}^{T - \frac{\log \lambda}{2} H_{X_0}}(\Phi) \tag{30}$$

for every nilpotent orbit $\mathcal{N}$, all $\lambda \in I$, and $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\Phi_\infty \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$). Recall that $\delta(\mathcal{N})$ was defined in (28).

(iii) As a polynomial, $J_{\mathcal{N}}^{T}(\Phi_\lambda)$ can be defined at every point $T \in \mathfrak{a}$, and (30) holds for all $T \in \mathfrak{a}$ and $\lambda \in \mathbb{R}_{>0}$.

Remark 5.4. The homogeneity property of $J_{\mathcal{N}}^{T}(\Phi_\lambda)$ from the second part of the corollary determines not only the location and order of the poles of $\Xi^T(s, \Phi)$ (see below). One can also read off the principal parts of the Laurent expansions at the poles from the coefficients of the polynomial $J_{\mathcal{N}}^{T}(\Phi)$, cf. Example 5.10. As we shall explain below, there is a way to prove the analytic continuation of $\Xi^T(s, \Phi)$ for general n which also correctly determines the location of the poles. This method, however, does not easily give the correct order of the poles or the principal parts of the Laurent expansions.

Proof of Corollary 5.3.

(i) This is a direct consequence of Propositions 4.5 and 5.1.

(ii) By the first part we have for every $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp. $\Phi_\infty \in \mathcal{S}_\nu(\mathfrak{g}(\mathbb{R}))$) and every $\lambda \in I$ that

$$|J_{\mathcal{N}}^{T}(\Phi_\lambda) - \tilde{j}_{\mathcal{N}}^{T}(\Phi_\lambda)| \leq \mu(\Phi_{\infty,\lambda})e^{-\varepsilon\|T\|}$$

for every sufficiently regular $T \in \mathfrak{a}^+$ with $d(T) \geq \delta\|T\|$. Since I is compact and $\mu(\Phi_{\infty,\lambda})$ varies continuously in λ, $C_I := \max_{\lambda\in I}\mu(\Phi_{\infty,\lambda})$ exists and is finite. Similarly, we have

$$|\lambda^{-\delta(\mathcal{N})}J_{\mathcal{N}}^{T-\frac{\log\lambda}{2}H}(\Phi) - \lambda^{-\delta(\mathcal{N})}\tilde{j}_{\mathcal{N}}^{T-\frac{\log\lambda}{2}H}(\Phi)| \leq \lambda^{-\delta(\mathcal{N})}\mu(\Phi_\infty)e^{-\varepsilon\|T-\frac{\log\lambda}{2}H\|}$$

for all $T \in \mathfrak{a}$ with $d(T - \frac{\log\lambda}{2}H) \geq \delta\|T - \frac{\log\lambda}{2}H\|$ and $T - \frac{\log\lambda}{2}H$ sufficiently regular. As

$$\tilde{j}_{\mathcal{N}}^{T}(\Phi_\lambda) = \lambda^{-\delta(\mathcal{N})}\tilde{j}_{\mathcal{N}}^{T-\frac{\log\lambda}{2}H}(\Phi),$$

we therefore get with $\lambda_I := \min_{\lambda\in I}\lambda$ that

$$|J_{\mathcal{N}}^{T}(\Phi_\lambda) - \lambda^{-\delta(\mathcal{N})}J_{\mathcal{N}}^{T-\frac{\log\lambda}{2}H}(\Phi)| \leq \max\{C_I, \lambda_I^{-\delta(\mathcal{N})}\mu(\Phi_\infty)\}e^{-\varepsilon\|T-\frac{\log\lambda}{2}H\|} \tag{31}$$

for all $T \in \mathfrak{a}$ with $d(T - \frac{\log\lambda}{2}H) \geq \delta\|T - \frac{\log\lambda}{2}H\|$ and $d(T) \geq \delta\|T\|$ if both T as well as $T - \frac{\log\lambda}{2}H$ are sufficiently regular.

The set of $T \in \mathfrak{a}^+$ satisfying both inequalities is an open cone in $\mathfrak{a}^+$ so that $J_{\mathcal{N}}^{T}(\Phi_\lambda)$—being a polynomial in T—is uniquely determined by this estimate. Thus the left-hand side of (31) must identically vanish and the second part of the corollary follows.

(iii) As a polynomial, $J_{\mathcal{N}}^{T}(\Phi_\lambda)$ can be defined at every point $T \in \mathfrak{a}$ with (30) holding for all $\lambda \in I$. Since $I \subseteq \mathbb{R}_{>0}$ is arbitrary, (30) holds for all $\lambda \in \mathbb{R}_{>0}$. □

The next two corollaries are obvious from our previous results so that we omit their proofs.

Corollary 5.5. *Let $T \in \mathfrak{a}$ be arbitrary, and let $\mathcal{N}$ be a nilpotent orbit. Let $\nu > 0$ be as before, and let $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$.*

(i) The function $J_{\mathcal{N}}^{T,-}(s, \Phi)$ defined by

$$J_{\mathcal{N}}^{T,-}(s, \Phi) = \int_0^1 \lambda^{n(s+\frac{n-1}{2})} J_{\mathcal{N}}^T(\Phi_\lambda) d^\times \lambda$$

converges absolutely and locally uniformly for $\Re s > \frac{1-n}{2} + \frac{1}{n}\delta(\mathcal{N})$. It defines a holomorphic function in this half plane and has a meromorphic continuation to all $s \in \mathbb{C}$ with only pole at $\frac{1-n}{2} + \frac{1}{n}\delta(\mathcal{N}) = \frac{1-n}{2} + \frac{\dim \mathcal{N}}{2n}$, which is of order at most $\dim \mathfrak{a} + 1$.

(ii) The function $J_{\mathcal{N}}^{T,+}(1-s, \Phi)$ defined by

$$J_{\mathcal{N}}^{T,+}(1-s, \hat{\Phi}) = \int_0^1 \lambda^{n(s+\frac{n-1}{2})} \lambda^{-n^2} J_{\mathcal{N}}^T(\hat{\Phi}_{\lambda^{-1}}) d^\times \lambda$$

converges absolutely and locally uniformly for $\Re s > \frac{n+1}{2} - \frac{1}{n}\delta(\mathcal{N})$. It defines a holomorphic function in this half plane and has a meromorphic continuation to all $s \in \mathbb{C}$ with only pole at $\frac{n+1}{2} - \frac{1}{n}\delta(\mathcal{N}) = \frac{n+1}{2} - \frac{\dim \mathcal{N}}{2n}$, which is of order at most $\dim \mathfrak{a} + 1$.

Corollary 5.6. *Let $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ and put*

$$I_{\mathcal{N}}^T(s, \Phi) = J_{\mathcal{N}}^{T,+}(1-s, \hat{\Phi}) - J_{\mathcal{N}}^{T,-}(s, \Phi).$$

Then for every $T \in \mathfrak{a}$, $I_{\mathcal{N}}^T(s, \Phi)$ has a meromorphic continuation to all $s \in \mathbb{C}$ and satisfies the functional equation

$$I_{\mathcal{N}}^T(s, \Phi) = I_{\mathcal{N}}^T(1-s, \hat{\Phi}).$$

Its only poles are at

$$\frac{1-n}{2} + \frac{\dim \mathcal{N}}{2n} \quad \textit{and} \quad \frac{n+1}{2} - \frac{\dim \mathcal{N}}{2n},$$

which are both of order at most $\dim \mathfrak{a} + 1$.

Our main theorem is now an easy consequence of the previous results.

Theorem 5.7. *Let $G = \mathrm{GL}_n$ with $n \leq 3$, and let $R > n$ be given. Then there exists $\nu < \infty$ such that for every $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ and $T \in \mathfrak{a}$ the following holds.*

(i) $\Xi^T(s, \Phi)$ is holomorphic for all $s \in \mathbb{C}$ with $\Re s > \frac{n+1}{2}$. It equals a polynomial in T of degree at most $\dim \mathfrak{a} = n - 1$.

(ii) $\Xi^T(s,\Phi)$ *has a meromorphic continuation to all* $s \in \mathbb{C}$ *with* $\Re s > -R$*, and satisfies for such s the functional equation*

$$\Xi^T(s,\Phi) = \Xi^T(1-s,\hat{\Phi}).$$

(iii) The poles of $\Xi^T(s,\Phi)$ *in* $\Re s > -R$ *are parametrised by the nilpotent orbits* $\mathcal{N} \subseteq \mathfrak{n}$*. More precisely, its poles occur exactly at the points*

$$s_{\mathcal{N}}^- = \frac{1-n}{2} + \frac{\dim \mathcal{N}}{2n} \quad \textit{and} \quad s_{\mathcal{N}}^+ = \frac{1+n}{2} - \frac{\dim \mathcal{N}}{2n}$$

and are of order at most $\dim \mathfrak{a} + 1 = n$*. In particular, the furthermost right and furthermost left pole in this region are both simple, correspond to* $\mathcal{N} = 0$*, and are located at the points* $s_0^+ = \frac{1+n}{2}$ *and* $s_0^- = \frac{1-n}{2}$*, respectively. The residues at these poles are given by*

$$\operatorname*{res}_{s=s_0^-} \Xi^T(s,\Phi) = \operatorname{vol}(A_G G(\mathbb{Q})\backslash G(\mathbb{A}))\Phi(0), \ \textit{and}$$

$$\operatorname*{res}_{s=s_0^+} \Xi^T(s,\Phi) = \operatorname{vol}(A_G G(\mathbb{Q})\backslash G(\mathbb{A}))\hat{\Phi}(0).$$

Remark 5.8. If we take $\Phi \in \mathcal{S}^\infty(\mathfrak{g}(\mathbb{A})) = \mathcal{S}(\mathfrak{g}(\mathbb{A}))$, then $\Xi^T(s,\Phi)$ has a meromorphic continuation to all of $\mathbb{C}$.

Proof. We only prove the theorem for $\nu = \infty$. The other case works similar by using the analogue results from the previous sections for $\nu < \infty$ instead and we omit the details for notational reasons. For every $\lambda \in (0,\infty)$ and every $T \in \mathfrak{a}$ Chaudouard's trace formula gives

$$J_*^T(\Phi_\lambda) = \lambda^{-n^2} J_*^T(\hat{\Phi}_{\lambda^{-1}}) + \lambda^{-n^2} J_{\mathfrak{n}}^T(\hat{\Phi}_{\lambda^{-1}}) - J_{\mathfrak{n}}^T(\Phi_\lambda).$$

Define

$$I_{\mathfrak{n}}^T(s,\Phi) = \int_0^1 \lambda^{n(s+\frac{n-1}{2})} \left(\lambda^{-n^2} J_{\mathfrak{n}}^T(\hat{\Phi}_{\lambda^{-1}}) - J_{\mathfrak{n}}^T(\Phi_\lambda)\right) d^\times\lambda$$

which converges for $\Re s > \frac{n+1}{2}$ and defines a holomorphic function there. By Corollary 5.5, we may split $I_{\mathfrak{n}}^T(s,\Phi)$ into a sum $\sum_{\mathcal{N}} I_{\mathcal{N}}^T(s,\Phi)$. Hence for $s \in \mathbb{C}$ with $\Re s > \frac{n+1}{2}$ we get

$$\Xi^T(s,\Phi) = \Xi^{T,+}(s,\Phi) + \Xi^{T,+}(1-s,\hat{\Phi}) + I_{\mathfrak{n}}^T(s,\Phi).$$

The assertions then follow from Theorems 3.4, 3.5, and Corollary 5.6. □

Remark 5.9. For more general $n > 3$ we cannot (yet) prove the homogeneity property of $J_{\mathcal{N}}^T(\Phi_\lambda)$ from Corollary 5.3, but one could try to use another approach to prove results analogous to Theorem 5.7 for $n > 3$. More precisely, using Arthur's fine geometric expansion (or rather its analogue for the Lie algebra) we have (cf. [Art85] in the group case)

$$J_{\mathfrak{n}}^T(\Phi_\lambda) = \sum_{(M,X)} a^M(X,S) J_M^T(X, \Phi_\lambda),$$

where M runs over all Levi subgroups $M \supseteq T_0$ and $X \in \mathfrak{m}(\mathbb{Q})$ over a set of representatives of the nilpotent $M(\mathbb{Q})$-orbits in $\mathfrak{m}(\mathbb{Q})$. The $a^M(X,S)$ are certain global coefficients and $J_M^T(X, \Phi_\lambda)$ certain weighted orbital integrals. The set S is a suitable sufficiently large finite set of places (depending on the support of Φ_λ, but independent of λ).

The global coefficients are in general not well understood (but see [Cha14] for some recent progress), but they are independent of the test function (as long as the set S can be kept fixed) and is therefore irrelevant for our purposes. Note that for $G = \mathrm{GL}_n$ every nilpotent orbit is a Richardson orbit. Hence the weighted orbital integrals can be written as

$$J_M^T(X, \Phi_\lambda) = \int_{\mathfrak{u}_X(\mathbb{Q}_S)} \Phi(\lambda Y) w_M(T, Y)\, dY,$$

where $\mathfrak{p}_X = \mathfrak{m}_X + \mathfrak{u}_X$ is a standard parabolic subalgebra of G such that the orbit $\mathcal{N}$ of X under $G(\mathbb{Q}_S)$ intersects $\mathfrak{u}_X(\mathbb{Q}_S)$ in a dense open subset, and $w_M(T, Y)$ is a certain weight function. Note that for $M = G$ we get the unweighted integral

$$J_G^T(X, \Phi_\lambda) = \int_{\mathfrak{u}_X(\mathbb{Q}_S)} \Phi(\lambda Y)\, dY,$$

which is independent of T and homogeneous of degree $-\dim \mathfrak{u}_X = -\dim \mathcal{N}/2$ in λ, that is, $J_G^T(X, \Phi_\lambda) = \lambda^{-\dim \mathcal{N}/2} J_G^T(X, \Phi)$.

Following Arthur's construction (in the group case) of the weight function $w_M(T, X)$ in [Art88], it should be possible to show that

$$w_M(T, \lambda^{-1} X) = \sum_{i=0}^{n-1} w_{M,i}(T, X)(\log \lambda)^i$$

for $w_{M,i}(T, X)$ suitable weight functions of the same type as $w_M(T, X)$. This would be enough to infer (along the same lines as above) the meromorphic continuation and functional equation of $\Xi^T(s, \Phi)$, and the location of the poles (which are the same as before). It would also give an upper bound on the order of the poles, namely the poles are of order at most n.

However, as pointed out before, this approach does not give the full principal parts of the Laurent expansions at the poles. For this a very good understanding of the weight functions would be necessary.

Example 5.10. We compute the Laurent expansions at the poles of $\Xi^T(s,\Phi)$ for GL_2. The truncation parameter T is in this case of the form $T=(T_1,-T_1)$. For T_1 sufficiently large we have (see [Gel96] for the unipotent contribution to the trace formula for GL_2)

$$J_{\mathfrak{n}}^T(\Phi)=\mathrm{vol}(G(\mathbb{Q})\backslash G(\mathbb{A})^1)\Phi(0)+\frac{d}{ds}\left[(s-1)\int_{\mathbb{A}^\times}|a|^s\int_{\mathbf{K}}\Phi(\mathrm{Ad}\,k^{-1}X(a))\,dk\,d^\times a\right]_{|s=1}$$
$$+2T_1\int_{\mathbb{A}}\int_{\mathbf{K}}\Phi(\mathrm{Ad}\,k^{-1}X(a))\,dk\,da$$

where $X(a)=\left(\begin{smallmatrix}0&a\\0&0\end{smallmatrix}\right)$. Splitting this according to the nilpotent orbits we get $J_{\mathcal{N}_{\mathrm{triv}}}^T(\Phi)=\mathrm{vol}(G(\mathbb{Q})\backslash G(\mathbb{A})^1)\Phi(0)$ and $J_{\mathcal{N}_{\mathrm{reg}}}^T(\Phi)=J_{\mathfrak{n}}^T(\Phi)-J_{\mathcal{N}_{\mathrm{triv}}}^T(\Phi)$. Write $\varphi(a)=\int_{\mathbf{K}}\Phi(\mathrm{Ad}\,k^{-1}X(a))\,dk$ so that $\varphi_\lambda(a):=\varphi(\lambda a)=\int_{\mathbf{K}}\Phi_\lambda(\mathrm{Ad}\,k^{-1}X(a))\,dk$. Then

$$\begin{aligned}J_{\mathcal{N}_{\mathrm{reg}}}^T(\Phi_\lambda)&=\frac{d}{ds}\left[(s-1)\int_{\mathbb{A}^\times}|a|^s\varphi(\lambda a)\,d^\times a\right]_{|s=1}+2T_1\int_{\mathbb{A}}\varphi(\lambda a)\,da\\
&=\lambda^{-1}\frac{d}{ds}\left[(s-1)\int_{\mathbb{A}^\times}|a|^s\varphi(a)\,d^\times a\right]_{|s=1}\\
&\quad+\lambda^{-1}\log\lambda^{-1}\lim_{s\to1}\left[(s-1)\int_{\mathbb{A}^\times}|a|^s\varphi(a)\,d^\times a\right]\\
&\quad+2\lambda^{-1}T_1\int_{\mathbb{A}}\varphi(a)\,da\\
&=\lambda^{-1}\frac{d}{ds}\left[(s-1)\int_{\mathbb{A}^\times}|a|^s\varphi(a)\,d^\times a\right]_{|s=1}+\lambda^{-1}(2T_1-\log\lambda)\int_{\mathbb{A}}\varphi(a)\,da\\
&=\lambda^{-1}J_{\mathcal{N}_{\mathrm{reg}}}^{T-\frac12\log\lambda H_{X_0}}(\Phi).\end{aligned}$$

Computing $I_{\mathcal{N}_{\mathrm{reg}}}^T(s,\Phi)$ we get

$$I_{\mathcal{N}_{\mathrm{reg}}}^T(s,\Phi)=-\frac{1}{4(s-1)^2}\hat{\tilde{\varphi}}(0)+\frac{1}{2(s-1)}J_{\mathcal{N}_{\mathrm{reg}}}^T(\widehat{\Phi})+\frac{1}{4s^2}\hat{\varphi}(0)-\frac{1}{2s}J_{\mathcal{N}_{\mathrm{reg}}}^T(\Phi)$$

where $\tilde{\varphi}$ denotes the function obtained from $\widehat{\Phi}$ analogous to φ obtained from Φ. This gives the principal parts of the Laurent expansions of $\Xi^T(s,\Phi)$ at $s=1$ and at $s=0$.

In particular, $\Xi^T(s,\Phi)$ has poles of second order at $s=1$ and $s=0$, and poles of simple order (from the trivial orbit) at $s=3/2$ and $s=-1/2$.

Similarly the distributions $I_{\mathcal{N}}^T(s,\Phi)$ can be computed for $n=3$ by using the expression for the unipotent distribution (in the group case) from [Fli82] (cf. also [Mat11]).

6 Connections to Arthur's Trace Formula and Shintani Zeta Function

The purpose of this section is to explain some connections of our zeta functions to other previously mentioned topics, namely the Shintani zeta function for binary quadratic forms, Arthur's trace formula, and automorphic zeta functions.

We first define the *main* part of the zeta functions. Let $n \geq 2$ be arbitrary. Recall that $X \in \mathfrak{g}(\mathbb{Q})_{\mathrm{ss}}$ (resp. $\gamma \in G(\mathbb{Q})_{\mathrm{ss}}$) is called *regular* if its eigenvalues (over some algebraic closure of $\mathbb{Q}$) are pairwise different, and that $X \in \mathfrak{g}(\mathbb{Q})_{\mathrm{ss}}$ (resp. $\gamma \in G(\mathbb{Q})_{\mathrm{ss}}$) is called *regular elliptic* if X (resp. γ) is regular and if the commutator subgroup G_X (resp., G_γ) is not contained in any proper parabolic subgroup of G. Note that an element $X \in \mathfrak{g}(\mathbb{Q})$ (resp. $\gamma \in G(\mathbb{Q})$) is regular elliptic if and only if its eigenvalues are pairwise distinct and some (and hence any) of them generates an n-dimensional field extension over $\mathbb{Q}$.

Let $\mathcal{O}_{\mathrm{reg}}$ denote the set of equivalence classes attached to the orbits of regular elements in $\mathfrak{g}(\mathbb{Q})$, and $\mathcal{O}_{\mathrm{er}}$ the set of classes attached to orbits of elliptic regular elements in $\mathfrak{g}(\mathbb{Q})$. Further, write $\mathcal{O}'_{\mathrm{reg}} = \mathcal{O}_{\mathrm{reg}} \backslash \mathcal{O}_{\mathrm{er}}$. We define the "main part" of Ξ^T as

$$\Xi^T_{\mathrm{main}}(s, \Phi) = \int_0^\infty \lambda^{n(s+\frac{n-1}{2})} \sum_{\mathfrak{o} \in \mathcal{O}_{\mathrm{er}}} J^T_{\mathfrak{o}}(\Phi_\lambda)\, d^\times \lambda .$$

By Theorem 3.4 this defines a holomorphic function for $\Re s > \frac{n+1}{2}$. In the next section we will see that, at least for $G = \mathrm{GL}_n$ and $n \leq 3$, this function is indeed the main part of $\Xi^T(s, \Phi)$ in the sense that it is responsible for the rightmost pole.

Note that in fact $\Xi^T_{\mathrm{main}}(s, \Phi)$, or more generally each of the distributions $J^T_{\mathfrak{o}}(\Phi_\lambda)$, $\mathfrak{o} \in \mathcal{O}_{\mathrm{er}}$, is independent of T: If $\mathfrak{o} \in \mathcal{O}_{\mathrm{er}}$, then $k^T_{\mathfrak{o}}(x, \Phi) = K_{G,\mathfrak{o}}(x, \Phi) = \sum_{X \in \mathfrak{o}} \Phi(\mathrm{Ad}\, x^{-1} X)$. So we also write $\Xi_{\mathrm{main}}(s, \Phi) = \Xi^T_{\mathrm{main}}(s, \Phi)$. The distribution $J_{\mathfrak{o}}(\Phi) = J^T_{\mathfrak{o}}(\Phi)$ can also be expressed as an orbital integral: Let $X \in \mathfrak{o}$ so that the centraliser G_X of X in G is reductive. We fix a Haar measure on $G_X(\mathbb{A})$. Denoting the quotient measure on $G_X(\mathbb{A}) \backslash G(\mathbb{A})$ again by dg, we then get

$$J_{\mathfrak{o}}(\Phi) = J^T_{\mathfrak{o}}(\Phi) = \mathrm{vol}(G_X(\mathbb{Q}) \backslash G_X(\mathbb{A})^1) \int_{G_X(\mathbb{A}) \backslash G(\mathbb{A})} \Phi(\mathrm{Ad}\, g^{-1} X)\, dg$$

(cf. [Cha02, § 5]).

6.1 *Relation to Arthur's Trace Formula*

Let $G = \mathrm{GL}_n$, and let $\mathcal{O}^G$ denote the set of geometric equivalence classes in the group $G(\mathbb{Q})$ as defined by Arthur (usually denoted by $\mathcal{O}$). To distinguish them from the equivalence classes we defined here on the set $\mathfrak{g}(\mathbb{Q})$, we shall write $\mathcal{O}^{\mathfrak{g}} = \mathcal{O}$

if necessary. Let $\mathcal{O}_{\mathrm{er}}^{\mathfrak{g}}$ (resp. $\mathcal{O}_{\mathrm{er}}^{G}$) denote the set of equivalence classes attached to orbits of elliptic regular elements $X \in \mathfrak{g}(\mathbb{Q})$ (resp. $\gamma \in G(\mathbb{Q})$). We have a canonical inclusion $G = \mathrm{GL}_n \hookrightarrow \mathfrak{g}$ of G-varieties preserving the semisimple elements $G(\mathbb{Q})_{\mathrm{ss}} \hookrightarrow \mathfrak{g}(\mathbb{Q})_{\mathrm{ss}}$. This is of course a special feature of GL_n and does not apply to general reductive groups. If $\gamma_s \in G(\mathbb{Q})_{\mathrm{ss}}$ and $\mathfrak{o}^G \in \mathcal{O}^G$ is the equivalence class attached to γ_s, it is straightforward that $\mathfrak{o}^G \in \mathcal{O}^{\mathfrak{g}}$ is also the equivalence class attached to γ_s viewed as an element in $\mathfrak{g}(\mathbb{Q})_{\mathrm{ss}}$. This gives an inclusion $\mathcal{O}^G \hookrightarrow \mathcal{O}^{\mathfrak{g}}$ and we view $\mathcal{O}^G$ as a subset of $\mathcal{O}^{\mathfrak{g}}$. Moreover, $\mathcal{O}_{\mathrm{er}}^{\mathfrak{g}} = \mathcal{O}_{\mathrm{er}}^{G}$.

Arthur's trace formula is an identity

$$J_{\mathrm{geom}}^{G,T}(f) = J_{\mathrm{spec}}^{G,T}(f)$$

of the so-called geometric and spectral distribution on a space of suitable test functions f on $G(\mathbb{A})$[1]. The geometric side allows a coarse geometric expansion given by $J_{\mathrm{geom}}^{G,T}(f) = \sum_{\mathfrak{o}\in\mathcal{O}^G} J_{\mathfrak{o}}^{G,T}(f)$ for $T \in \mathfrak{a}$ and $J_{\mathfrak{o}}^{G,T}$ a certain distribution attached to $\mathfrak{o}$, cf. [Art05] (usually $J_{\mathfrak{o}}^{G,T}$ is denoted by $J_{\mathfrak{o}}^{T}$).

Let $\Phi \in \mathcal{S}(\mathfrak{g}(\mathbb{A}))$. For $s \in \mathbb{C}$ with $\Re s > (n+1)/2$ define a smooth function $f_s : G(\mathbb{A}) \longrightarrow \mathbb{C}$ by

$$f_s(g) = \int_0^\infty \lambda^{n(s+\frac{n-1}{2})}\Phi(\lambda g)d\lambda.$$

By [FLM11] f_s may be used as a test function for a certain expansion of the spectral side of Arthur's trace formula if $\Re s > (n+1)/2$. If $n \leq 3$ and $\Re s > (n+1)/2$, there are also expansion, of the geometric side of the trace formula which converge absolutely with f_s as a test function, see [FL11a, Mat11]. Also, the regular elliptic (or more generally semisimple) part of the trace formula converges absolutely for such f_s and any n by [FL11b]. In particular,

$$\Xi_{\mathrm{main}}(s,\Phi) = \sum_{\mathfrak{o}\in\mathcal{O}_{\mathrm{er}}^G} J_{\mathfrak{o}}^{G}(f_s),$$

defines a holomorphic function in $\Re s > (n+1)/2$ for every n. Here the sum is the regular elliptic part of Arthur's trace formula (again, $J_{\mathfrak{o}}^{G,T}(f_s) = J_{\mathfrak{o}}^{G}(f_s)$ is independent of T). Also $J_{\mathrm{geom}}^{G,T}(f_s) = J_{\mathrm{spec}}^{G,T}(f_s)$ defines a holomorphic function in $\Re s > (n+1)/2$ for arbitrary n.

For $n = 2, 3$ one could try to use the geometric side $J_{\mathrm{geom}}^{G,T}(f_s)$ as a regularisation for $\Xi^T(s,\Phi)$ and Arthur's trace formula as a replacement for the Poisson summation formula. However, the geometric (or, equivalently, spectral) side of Arthur's trace formula seems to be "too small" in the sense that the function arising from the continuous spectrum on the spectral side might have no meromorphic continuation to all of $\mathbb{C}$ in general. It is quite possible that $J_{\mathrm{geom}}^{G,T}(f_s)$ (and also $\Xi_{\mathrm{main}}(s,\Phi)$) cannot be meromorphically continued to all of $\mathbb{C}$, cf. [Mat11, IV.iii]. This is one reason why it seems more natural to study $\Xi_{\mathrm{main}}(s,\Phi)$ in the context of the trace formula for $\mathfrak{g}$ instead of G.

For example, if $n = 2$, the spectral side of the trace formula has infinitely many poles coming from the contribution of the continuous spectrum. More precisely, the infinitely many poles come from the intertwining operators. For example, in the unramified case, this contribution actually equals (cf. [Mat11])

$$\frac{1}{2\pi i}\int_{i\mathbb{R}} r(\sigma)^{-1} r'(\sigma)\,\mathrm{tr}\, I(\sigma, f_s)\, d\sigma,$$

where $r(\sigma) = \frac{\zeta^*(1-2\sigma)}{\zeta^*(1+2\sigma)}$ with ζ^* the completed Riemann zeta function is the normalising factor of the intertwining operator, and $I(\sigma, \cdot)$ denotes the representation parabolically induced from the trivial representation on the diagonal torus twisted with the unitary character attached to the parameter σ.

6.2 *Automorphic L-Functions*

Even though $J^{G,T}_{\mathrm{geom}}(f_s)$ does not provide the right regularisation for $\Xi_{\mathrm{main}}(s, \Phi)$, $J^{G,T}_{\mathrm{geom}}(f_s)$ can be viewed as a "piece" of $\Xi^T(s, \Phi)$. Hence also the spectral side $J^{G,T}_{\mathrm{spec}}(f_s)$ is a piece of $\Xi^T(s, \Phi)$. The spectral side with the test function f_s contains a particularly interesting part. Suppose that the function Φ is bi-**K**-invariant. Then the cuspidal part of the spectral side $J^{G,T}_{\mathrm{spec}}(f_s)$ contributes a sum of zeta functions

$$\sum_{\pi} Z(s, \Phi, \pi),$$

where the sum runs over all unramified cuspidal automorphic representations π of $G(\mathbb{A})^1$, and $Z(s, \Phi, \pi)$ is a certain zeta function as defined in [GJ72]. By the theory of Godement–Jacquet [GJ72] the ideal generated by all the $Z(s, \Phi, \pi)$, $\Phi \in \mathcal{S}(\mathfrak{g}(\mathbb{A}))$ is generated by the completed automorphic L-function $L^*(s, \pi)$ attached to π. This actually reflects the fact that the Lie algebra $\mathfrak{g}$ is Vinberg's universal monoid for GL(n), and the f_s are non-standard test functions, cf. [Ngô14, Sak14].

If we choose Φ such that f_s has cuspidal image (that is, $J^{G,T}_{\mathrm{spec}}(f_s) = J^{G,T}_{\mathrm{cusp}}(f_s)$ is just the cuspidal contribution), then

$$\sum_{\pi} Z(s, \Phi, \pi) = J^{G,T}_{\mathrm{geom}}(f_s)$$

is independent of T, and it follows from [GJ72] that $J^{G,T}_{\mathrm{geom}}(f_s)$ has a continuation to an entire function and satisfies the functional equation under $s \leftrightarrow 1-s$ and $\Phi \leftrightarrow \widehat{\Phi}$.

6.3 Connection to the Shintani Zeta Function in the Quadratic Case

The purpose of this section is to explain the connection between the Shintani zeta function [Shi75, Yuk92, Dat96] and the main part of the zeta function $\Xi_{\text{main}}(s, \Phi)$ for GL_2, or, equivalently, the regular elliptic part of Arthur's trace formula for GL_2, cf. also [Lap02]. We shortly review some notation and results from [Dat96] and [Yuk92].

Let $G = \mathrm{GL}_2$. The Shintani zeta function studies the action of G on the three dimensional space of binary quadratic forms with rational coefficients. The space of such forms will be denoted by V, its rational points by $V_{\mathbb{Q}}$ and $V_{\mathbb{A}} = V_{\mathbb{Q}} \otimes \mathbb{A}$. Let $\mathcal{S}(V_{\mathbb{A}})$ be the space of Schwartz–Bruhat functions on $V_{\mathbb{A}}$. If $X = (X_1, X_2, X_3) \in V_{\mathbb{A}}$ is a binary quadratic form, $X(u, v) = X_1 u^2 + X_2 uv + X_3 v^2$, the action of G is given by $g \cdot X(u, v) = X((u, v)g^t)$. Explicitly this is given by

$$X \mapsto \begin{pmatrix} a^2 & 2ac & c^2 \\ ab & ad + bc & cd \\ b^2 & 2bd & d^2 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G(\mathbb{A})$. GL_1 acts by multiplication on the coefficients of V. Let $H = \mathrm{GL}_1 \times \mathrm{GL}_2$ so that H acts on V. We denote this action by $h \cdot X$. We view H as embedded in $\mathrm{GL}(V) \simeq \mathrm{GL}_3$ so that we can write $\det h$ for $h = (a, g) \in H$ which equals $\det h = a \det(g)$.

Note that there is an isomorphism $\mathfrak{g} \simeq V \oplus A^1$ with A^1 the one-dimensional affine space over $\mathbb{Q}$. Under this isomorphism the adjoint action of G on $\mathfrak{g}$ splits into the action of the subgroup $H_G = \{(\det g^{-1}, g) \in H \mid g \in G\} \subseteq H$ on V plus the identity on A^1. In particular, under the projection $\mathfrak{g} \longrightarrow A^1$, $g \mapsto \operatorname{tr} g$, each fibre is isomorphic to V and is invariant under the action of G. For $X \in V$ let $\gamma_X \in \mathfrak{g}$ be the unique element in the fibre above $0 \in A^1$ defined by the above isomorphism. The measure on $V_{\mathbb{A}}$ is the natural one obtained from the identification $V_{\mathbb{A}} \simeq \mathbb{A}^3$ given via the coefficients of the quadratic form. For the inner form $[\cdot, \cdot]$ on $V_{\mathbb{A}}$ we adopt the convention from [Dat96] by defining $[X, Y] = X_1 Y_3 - \frac{1}{2} X_2 Y_2 + X_3 Y_1$. Let $\psi = \bigotimes_v \psi_v : \mathbb{A} \longrightarrow \mathbb{C}^\times$ be the previously fixed non-trivial character. Then $\widehat{\Psi}(Y) = \int_{V_{\mathbb{A}}} \Psi(X) \psi([X, Y])\, dX$ denotes the Fourier transform with respect to ψ. If $\Phi \in \mathcal{S}(\mathfrak{g}(\mathbb{A}))$, we use the same character to define the Fourier transform of Φ on the space of all 2×2 matrices by $\widehat{\Phi}(x) = \int_{\mathfrak{g}(\mathbb{A})} \Phi(y) \psi(\operatorname{tr}(xy)) dy$. Note that if $a \in \mathbb{A}$, $X \in V_{\mathbb{A}}$, then

$$\widehat{\Phi}(a + \gamma_X) = \int_{\mathfrak{g}(\mathbb{A})} \Phi(y) \psi(\operatorname{tr}((a + \gamma_X) y))\, dy = -\int_{\mathbb{A}} \int_{V_{\mathbb{A}}} \Phi(b + \gamma_Y) \psi(2ab) \psi([X, Y])\, dY\, db$$

For a binary quadratic form $X \in V_{\mathbb{Q}}$ we denote the splitting field of X over $\mathbb{Q}$ by $F(X)$, and write $P(X) = X_2^2 - 4X_1X_3$ for the discriminant of the form X. Clearly, $[F(X) : \mathbb{Q}] \leq 2$ and $[F(X) : \mathbb{Q}] = 2$ if and only if $P(X)$ is not a square in $\mathbb{Q}$.

Let $V''_{\mathbb{Q}} = \{X \in V_{\mathbb{Q}} | [F(X) : \mathbb{Q}] = 2\}$. Then V with the above action of H is a prehomogeneous vector space with relative invariant P. In particular the action of $H(\mathbb{R})$ on $V_{\mathbb{R}}$ has only finitely many orbits.

For $\Psi \in \mathcal{S}(V_{\mathbb{A}})$ and $s \in \mathbb{C}$, $\Re s > 3/2$, the Shintani zeta-function (with trivial central character) is defined by [Dat96, Yuk92]

$$Z(\Psi, s) = \int_{H(\mathbb{Q})\backslash H(\mathbb{A})} |\det h|^{2s} \sum_{X \in V''_{\mathbb{Q}}} \Psi(h \cdot X)\, dh.$$

This is a special case of a zeta function associated with a prehomogeneous vector space. It can be shown (see [Dat96]) that this zeta functions has a meromorphic continuation to the whole complex plane. In order to get a functional equation an *adjusted* Shintani zeta function is necessary. One can show that this adjusted function also occurs naturally as a part of the geometric side of the trace formula, cf. [Mat11].

Let $G(\mathbb{Q})_{\text{ell}}$ denote the set of all elliptic elements in $G(\mathbb{Q})$, and $G(\mathbb{Q})_{\text{ell, reg}} \subseteq G(\mathbb{Q})_{\text{ell}}$ the subset of all regular elliptic elements. Then $G(\mathbb{Q})_{\text{ell,reg}} = \bigsqcup_{\mathfrak{o} \in \mathcal{O}^G_{\text{er}}} \mathfrak{o} = \bigsqcup_{\mathfrak{o} \in \mathcal{O}^{\mathfrak{g}}_{\text{er}}} \mathfrak{o}$.

Theorem 6.1. *Suppose $\Phi \in \mathcal{S}(\mathfrak{g}(\mathbb{A}))$ is invariant under scalar matrices with scalars in $\widehat{\mathbb{Z}}^{\times}$. Then the main part $\Xi_{main}(s, \Phi)$ of the zeta function for G equals the Shintani zeta function $Z(\Psi, s)$ up to an entire function where*

$$\Psi(X) = \int_{\mathbb{A}} \Phi(a + \gamma_X) da, \quad X \in V_{\mathbb{A}}.$$

In particular, the poles and residues of $\Xi_{main}(s, \Phi)$ and $Z(\Psi, s)$ coincide.

Remark 6.2. Similarly the adjusted Shintani zeta function from [Yuk92] can be basically identified with $\Xi^T(s, \Phi)$ (see [Mat11]).

Proof. Let $\Re s > \frac{3}{2}$. We use the map $\mathfrak{g}(\mathbb{Q}) \longrightarrow \mathbb{Q}$ from above. Note that the intersection of each fibre with $G(\mathbb{Q})_{\text{ell,reg}}$ is isomorphic to $V''_{\mathbb{Q}}$. By definition the function $\Xi_{\text{main}}(s, \Phi)$ equals

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_0^{\infty} \sum_{X \in V''_{\mathbb{Q}}} \sum_{q \in \mathbb{Q}} \Phi(\lambda q \mathbf{1}_2 + \lambda g^{-1} \gamma_X g) \lambda^{2s+1}\, d^{\times}\lambda\, dg.$$

We split the integral over λ in one over $(0, 1]$ and one over $[1, \infty)$. Since Φ is a Schwartz–Bruhat function, the integral

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_1^{\infty} \sum_{X \in V''_{\mathbb{Q}}} \sum_{q \in \mathbb{Q}} \lambda^{2s+1} \Phi(\lambda q \mathbf{1}_2 + \lambda g^{-1} \gamma_X g)\, d^{\times}\lambda\, dg$$

converges absolutely for all $s \in \mathbb{C}$, i. e. defines a holomorphic function on $\mathbb{C}$.

The remaining part of the integral is

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_0^1 \lambda^{2s+1} \sum_{X\in V''_{\mathbb{Q}}} \sum_{q\in\mathbb{Q}} \Phi(\lambda q \mathbf{1}_2 + \lambda g^{-1}\gamma_X g)\, d^\times\lambda\, dg.$$

We apply the Poisson summation formula to the inner sum over q to get

$$\sum_{q\in\mathbb{Q}} \Phi(\lambda z q \mathbf{1}_2 + \lambda g^{-1}\gamma_X g) = \frac{1}{\lambda} \sum_{a\in\mathbb{Q}} \mathcal{F}_1\Phi(\frac{a}{\lambda} + \lambda g^{-1}\gamma_X g),$$

where

$$\mathcal{F}_1\Phi(y + \lambda z g^{-1}\gamma_X g) = \int_{\mathbb{A}} \Phi(q + \lambda z g^{-1}\gamma_X g)\psi(qy)\, dq$$

is the Fourier-transform in the "central" variable, which is again a Schwartz–Bruhat function on $\mathbb{A} \oplus V_{\mathbb{A}} \simeq \mathfrak{g}(\mathbb{A})$. Using this, the integral for $(0, 1]$ equals

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_0^1 \sum_{X\in V''_{\mathbb{Q}}} \lambda^{2s} \sum_{a\in\mathbb{Q}^\times} \mathcal{F}_1\Phi(\frac{a}{\lambda} + \lambda g^{-1}\gamma_X g)\, d^\times\lambda\, dg$$
$$+ \int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_0^1 \lambda^{2s} \sum_{X\in V''_{\mathbb{Q}}} \mathcal{F}_1\Phi(\lambda g^{-1}\gamma_X g)\, d^\times\lambda\, dg.$$

Changing the variable λ to λ^{-1} in the first integral we get

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_1^\infty \lambda^{-2s-1} \sum_{X\in V''_{\mathbb{Q}}} \sum_{a\in\mathbb{Q}^\times} \mathcal{F}_1\Phi(\lambda z a + \lambda^{-1} g^{-1}\gamma_X g)\, d^\times\lambda\, dg,$$

which again converges absolutely for all $s \in \mathbb{C}$. So the analytic behaviour of the regular elliptic contribution is completely determined by

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_0^1 \lambda^{2s} \sum_{X\in V''_{\mathbb{Q}}} \mathcal{F}_1\Phi(\lambda g^{-1}\gamma_X g)\, d^\times\lambda\, dg,$$

and we change nothing of its analytic properties if instead we consider

$$\int_{G(\mathbb{Q})\backslash G(\mathbb{A})^1} \int_0^\infty \lambda^{2s} \sum_{X\in V''_{\mathbb{Q}}} \mathcal{F}_1\Phi(\lambda g^{-1}\gamma_X g)\, d^\times\lambda\, dg,$$

which is exactly the Shintani zeta function $Z(\Psi, s)$ for $\Psi(X) = \mathcal{F}_1\Phi(\gamma_X)$, $X \in V_{\mathbb{A}}$. $\square$

7 Poles of $\Xi_{\mathrm{main}}(s, \Phi)$ for $G = \mathrm{GL}_n$, $n \leq 3$

In this section let $G = \mathrm{GL}_n$ with $n \leq 3$. We assume throughout that $\nu > 0$ is sufficiently large as in Lemma 3.7. The purpose of this section is to show that $\Xi_{\mathrm{main}}(s, \Phi)$ is indeed the main part of $\Xi^T(s, \Phi)$ in the sense that it is responsible for the furthermost right pole of $\Xi^T(s, \Phi)$.

We group the equivalence classes in $\mathcal{O}_*$ into subsets of different type: Let $\mathcal{O}_c \subseteq \mathcal{O}$ denote the set of equivalence classes attached to the orbits of central elements. Hence $\mathfrak{n} \in \mathcal{O}_c$ and for every $\mathfrak{o} \in \mathcal{O}_c$ there exists $a \in \mathbb{Q}$ such that $\mathfrak{o} = a\mathbf{1}_n + \mathfrak{n}$. Write $\mathcal{O}_{c,*} = \mathcal{O}_c \backslash \{\mathfrak{n}\}$. Then if $n = 2$, we get a disjoint union

$$\mathcal{O}_*^{\mathfrak{gl}_2} = \mathcal{O}_{c,*} \cup \mathcal{O}'_{\mathrm{reg}} \cup \mathcal{O}_{\mathrm{er}}.$$

If $n = 3$, there is one type of equivalence classes missing: Let $\mathcal{O}_{(2,1)}$ denote the set of $\mathfrak{o} \in \mathcal{O} = \mathcal{O}^{\mathfrak{gl}_3}$ for which there are $a, b \in \mathbb{Q}$, $a \neq b$, such that every element $X \in \mathfrak{o}$ has a as an eigenvalue with multiplicity 2 and b as an eigenvalue with multiplicity 1. We denote the equivalence class corresponding to a, b by $\mathfrak{o}_{(a,b)}$. Then

$$\mathcal{O}_*^{\mathfrak{gl}_3} = \mathcal{O}_{c,*} \cup \mathcal{O}'_{\mathrm{reg}} \cup \mathcal{O}_{\mathrm{er}} \cup \mathcal{O}_{(2,1)}.$$

If convenient, we will assume without further notice that Φ is invariant under $\mathrm{Ad}\,\mathbf{K}$.

7.1 Contribution from $\mathcal{O}_{c,*}$

We first deal with the contribution from the classes in $\mathcal{O}_{c,*}$.

Proposition 7.1. *Let $T \in \mathfrak{a}^+$ be sufficiently regular and $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$. Then there exists a constant $C > 0$ such that*

$$\left| \sum_{\mathfrak{o} \in \mathcal{O}_{c,*}} J_{\mathfrak{o}}^T(\Phi_\lambda) \right| \leq C\lambda^{-n^2+1} \tag{32}$$

for all $\lambda \in (0, 1]$.

Proof. By the proof of Lemma 3.7, it suffices to estimate the sum over $\mathfrak{o} \in \mathcal{O}'_c$ and standard parabolic subgroups $P_1 \subseteq R \subseteq P_2$ of (16). It further follows from the proof of that lemma and Remark 3.15 that it suffices to find a bound for the case that $R = P_2 = G$ if $G = \mathrm{GL}_3$, and $R = P_2$ if $G = \mathrm{GL}_2$. However, if $G = \mathrm{GL}_2$ and $R = P_2 \subsetneq G$, we can use the estimate given in Remark 3.12 (recall that $\mathfrak{o} = a\mathbf{1}_n + \mathfrak{n}$

for some $a \in \mathbb{Q}\backslash\{0\}$) in the proof of Lemma 3.7 to get the stated upper bound. Hence we are left with $R = P_2 = G$ for $n = 2$ as well as $n = 3$.

$G = \mathrm{GL}_2$: We need to estimate the sum-integrals

$$\int_{A_G P_0(\mathbb{Q})\backslash G(\mathbb{A})} \tau_0^G(H_0(x) - T) \sum_{X \in \mathbb{Q}\mathbf{1}_2, X \neq 0} \sum_{Y \in \mathfrak{n} \cap \tilde{\mathfrak{m}}_0^G(\mathbb{Q})} \left|\Phi(\lambda(X + \mathrm{Ad}\, x^{-1} Y))\right| dx, \quad \text{and} \tag{33}$$

$$\int_{A_G G(\mathbb{Q})\backslash G(\mathbb{A})} F(x, T) \sum_{X \in \mathbb{Q}\mathbf{1}_2, X \neq 0} \sum_{Y \in \mathfrak{n}} \left|\Phi(\lambda(X + \mathrm{Ad}\, x^{-1} Y))\right| dx. \tag{34}$$

We can replace $|\Phi|$ without loss of generality by a product $\Phi_1 \Phi_2$ with $\Phi_1 \in \mathcal{S}(\mathbb{A})$ and $\mathbf{K}$-conjugation invariant $\Phi_2 \in \mathcal{S}(\mathfrak{sl}_2(\mathbb{A}))$ such that $|\Phi(X)| \leq \Phi_1(\mathrm{tr}\, X)\Phi_2(X - \frac{1}{2}\,\mathrm{tr}\, X\, \mathrm{id})$ for all $X \in \mathfrak{g}(\mathbb{A})$ and such that the relevant seminorms of Φ_1 and Φ_2 are bounded from above by seminorms of Φ in the sense of Lemma 3.14. If $Y = (Y_{ij})_{i,j=1,2} \in \mathfrak{n}$, then $Y_{22} = -Y_{11}$, and either $Y_{11} = Y_{21} = Y_{22} = 0$ (such elements do not occur in the sum (33)), or $Y_{21} \neq 0$ and $Y_{12} = -Y_{11}^2/Y_{21}$ so that $Y = \left(\begin{smallmatrix} 1 & Y_{11}/Y_{21} \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 0 \\ Y_{21} & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & -Y_{11}/Y_{21} \\ 0 & 1 \end{smallmatrix}\right)$. Hence (33) can be bounded from above by

$$\begin{aligned} &\int_{A_G T_0(\mathbb{Q})\backslash G(\mathbb{A})} \tau_0^G(H_0(x) - T) \sum_{a \in \mathbb{Q}\backslash\{0\}} \Phi_1(\lambda a) \sum_{Y_0 \in \mathbb{Q}\backslash\{0\}} \Phi_2\big(\lambda\, \mathrm{Ad}\, x^{-1} \begin{pmatrix} 0 & 0 \\ Y_0 & 0 \end{pmatrix}\big)\, dx \\ &\leq C_1 \lambda^{-1} \int_{A_0^G} \delta_0(a)^{-1} \tau_0^G(H_0(a) - T) \sum_{Y_0 \in \mathbb{Q}\backslash\{0\}} \int_{U_0(\mathbb{A})} \Phi_2\big(\lambda\, \mathrm{Ad}(ua)^{-1} \begin{pmatrix} 0 & 0 \\ Y_0 & 0 \end{pmatrix}\big)\, du\, da, \end{aligned}$$

where $C_1 > 0$ is a suitable constant depending on Φ_1. Now if we write $a = \mathrm{diag}(a, a^{-1}) \in A_0^G$, $a \in \mathbb{R}_{>0}$,

$$\begin{aligned} \int_{U_0(\mathbb{A})} \Phi_2(\lambda\, \mathrm{Ad}(ua)^{-1} \begin{pmatrix} 0 & 0 \\ Y_0 & 0 \end{pmatrix}) du &= \int_{\mathbb{A}} \Phi_2\big(\lambda \begin{pmatrix} -uY_0 & -u^2 a^{-2} Y_0 \\ a^2 Y_0 & uY_0 \end{pmatrix}\big)\, du \\ &\leq \varphi(\lambda a^2 Y_0) \int_{\mathbb{A}} \varphi(\lambda u Y_0) \varphi(-\lambda u^2 a^{-2} Y_0)\, du, \end{aligned}$$

where $\varphi \in \mathcal{S}(\mathbb{A})$ is a suitable function related to Φ_2 by Lemma 3.14. We can moreover assume that φ is monotonically decreasing in the sense that if $x, y \in \mathbb{A}$ with $|x| \leq |y|$, then $\varphi(x) \geq \varphi(y)$. If $\tau_0^G(H_0(a) - T) = 1$, i.e., $2 \log a \geq \alpha(T)$ for α the unique simple root, we distinguish the cases $|u| \leq 1$ and $|u| \geq 1$. With this we can bound the last integral by $\varphi(\lambda a^2 Y_0) a^2 \lambda^{-1} C_2)$ for $C_2 > 0$ a suitable constant. Hence (33) is bounded by

$$C_3 \lambda^{-3} \int_{A_0^G} \delta_0(a)^{-1} \tau_0^G(H_0(a) - T)\, da = C_3 \lambda^{-3} e^{-\alpha(T)}/2$$

for a suitable $C_3 > 0$ depending on Φ.

Now for (34) note that $\mathfrak{n}$ is the disjoint union of $\mathfrak{u}_0$ and $\mathfrak{n}\cap\tilde{\mathfrak{m}}_0^G(\mathbb{Q})$. Then (34) is bounded from above by

$$\lambda^{-1}C_1)\Bigg(\int_{A_GP_0(\mathbb{Q})\backslash\mathcal{C}_{T_1}} F(x,T)\sum_{Y\in\mathfrak{u}_0(\mathbb{Q})}|\Phi_2(\lambda\operatorname{Ad}x^{-1}Y)|\,dx \\ +\int_{A_GP_0(\mathbb{Q})\backslash\mathcal{C}_{T_1}} F(x,T)\sum_{Y\in\mathfrak{n}\cap\tilde{\mathfrak{m}}_0^G(\mathbb{Q})}|\Phi_2(\lambda\operatorname{Ad}x^{-1}Y)|\,dx\Bigg)$$

for which the first sum is bounded by

$$\lambda^{-1}C_1\varphi(0)^2\int_{A_0^G(T_1)}\delta_0(a)^{-1}\Bigg(\int_{U_0(\mathbb{Q})\backslash U_0(\mathbb{A})}F(ua,T)\,du\Bigg)\sum_{Y\in\mathbb{Q}}\varphi(\lambda a^{-2}Y)\,da \\ \leq\lambda^{-1}C_1\varphi(0)^2\int_{e^{\alpha(T_1)/2}}^{e^{\alpha(T)/2}}a^{-2}\sum_{Y\in\mathbb{Q}}\varphi(\lambda a^{-2}Y)\,d^\times a.$$

This is bounded by the product of $\lambda^{-2}C_4$ and a linear polynomial in T for some suitable $C_4>0$ depending on Φ. For the second integral recall that $F(uak,T)\leq\hat{\tau}_0(T-H_0(a))=\tau_0^G(T-H_0(a))$ for all $a\in A_0^G(T_1)$. Using similar manipulations as for (33), the second integral is therefore bounded by

$$C_5\lambda^{-3}\int_{e^{\alpha(T_1)/2}}^{e^{\alpha(T)/2}}a^{-2}\,d^\times a,$$

which equals a constant multiple of $C_5\lambda^{-3}e^{-\alpha(T)}$ for some constant $C_5>0$ depending on Φ. Hence the assertion of the proposition is proven for $G=\mathrm{GL}_2$.

$G=\mathrm{GL}_3$**:** For every standard parabolic in $P_1\subseteq G$ we need to estimate the sum-integral

$$\int_{A_GP_1(\mathbb{Q})\backslash G(\mathbb{A})}F^{P_1}(x,T)\tau_{P_1}^G(H_0(x)-T)\sum_{X\in\mathbb{Q}\mathbf{1}_3,X\neq0}\;\sum_{Y\in\mathfrak{n}\cap\tilde{\mathfrak{m}}_{P_1}^G(\mathbb{Q})}\big|\Phi(\lambda(X+\operatorname{Ad}x^{-1}Y))\big|\,dx, \tag{35}$$

or rather, using the same notation and arguments as in the previous case,

$$\int_{A_GP_1(\mathbb{Q})\backslash G(\mathbb{A})}F^{P_1}(x,T)\tau_{P_1}^G(H_0(x)-T)\sum_{Y\in\mathfrak{n}\cap\tilde{\mathfrak{m}}_{P_1}^G(\mathbb{Q})}\big|\Phi_2(\lambda\operatorname{Ad}x^{-1}Y)\big|\,dx,$$

since again $\sum_{X\in\mathbb{Q},X\neq0}\Phi_1(\lambda X)\leq C_1\lambda^{-1}$ for some constant $C_1>0$ depending on Φ_1. First, suppose $P_1=P_0$ is the minimal parabolic subgroup. Then $\tilde{\mathfrak{m}}_{P_0}^G(\mathbb{Q})\cap\mathfrak{n}$ is the disjoint union of the set of those nilpotent $Y=(Y_{ij})_{i,j=1,2,3}$ with $Y_{31}\neq0$ and those with $Y_{31}=0$, but $Y_{21}\neq0\neq Y_{32}$. The elements Y satisfying the second property are contained in the codimension one vector subspace $\{Y \in \mathfrak{n} \mid Y_{31} = 0\}$ of $\mathfrak{n}$ so that by similar arguments as before, an upper bound as asserted holds for this sum. Hence we are left to consider the sum over those $Y \in \mathfrak{n}$ with $Y_{31} \neq 0$. By the same reasoning we may further restrict to those Y with $Y_{31} \neq 0 \neq Y_{21}$. Since Y is nilpotent, for every such Y there exists $u \in U_0(\mathbb{Q})$ such that in the matrix $\operatorname{Ad} uY$ either the second or third column is identically equal to 0. Moreover, the $(2,1)$- and the $(3,1)$-entry in $\operatorname{Ad} uY$ is the same as in Y and a similar analysis as in the case of $G = \mathrm{GL}_2$ for (33) shows that (35) is bounded as asserted.

Next suppose that $P_1 = M_1U_1$ is the maximal standard parabolic subgroup with $M_1 = \mathrm{GL}_2 \times \mathrm{GL}_1 \hookrightarrow \mathrm{GL}_3$ (diagonally embedded). (The other maximal standard parabolic subgroup is treated the same way.) Then

$$A_G P_1(\mathbb{Q})\backslash G(\mathbb{A}) \simeq U_1(\mathbb{Q})\backslash U_1(\mathbb{A}) \times A_G M_1(\mathbb{Q})\backslash M_1(\mathbb{A}) \times \mathbf{K},$$

$F^{P_1}(umk, T) = F^{M_1}(m, T^{M_1})$ for $u \in U_1(\mathbb{A})$, $m \in M_1(\mathbb{A})$, $k \in \mathbf{K}$, and $\tau_{P_1}^G(H_0(umk) - T) = \tau_{P_1}^G(H_0(m) - T)$. Now if $Y \in \mathfrak{n} \cap \tilde{\mathfrak{m}}_{P_1}^G(\mathbb{Q})$, then $(Y_{31}, Y_{32}) \neq (0,0)$, and there exists $u \in U_0(\mathbb{Q})$ such that the second or third column of $\operatorname{Ad} uY$ is identically 0. If there exists $u \in U_1(\mathbb{Q})$ such that the last column of $\operatorname{Ad} uY$ is 0 (note that the $(3,1)$-and $(3,2)$-entries stay unchanged under $\operatorname{Ad} u$), we proceed similar as in the case of GL_2 and the estimation of (33). Otherwise there exists $u \in U_0^{M_1}(\mathbb{Q})$ such that the second column of $\operatorname{Ad} uY$ is 0 and the $(3,1)$-entry stays unchanged. This again leads to an upper bound of the asserted form by using a similar approach as for GL_2 and (34).

Hence we are left with $P_1 = G$. We estimate the corresponding integral again by an integral over a quotient of the Siegel domain $A_G P_0(\mathbb{Q})\backslash \mathcal{C}_{T_1}$. Moreover, $F^G(umk, T) \leq \hat{\tau}_0^G(T - H_0(m))$ for $umk \in U_0(\mathbb{A})T_0(\mathbb{A})\mathbf{K}$. Hence a similar reasoning as for GL_2 and the integral (34) yields an upper bound as asserted.

Taking the estimates for all standard parabolic subgroups P_1 together, the assertion follows now also for GL_3. □

7.2 Contribution from $\mathcal{O}'_{reg}$

Let $\mathfrak{o} \in \mathcal{O}'_{\mathrm{reg}}$ and let $X_1 \in \mathfrak{o}$ be semisimple. Let P_1 be the smallest standard parabolic subgroup such that $X_1 \in \mathfrak{m}_1(\mathbb{Q})$. We may assume that $X_1 \in \mathfrak{o}$ is chosen such that X_1 is not contained in any proper (not necessarily standard) parabolic subalgebra of $\mathfrak{m}_1(\mathbb{Q})$. We may further assume that if $G = \mathrm{GL}_2$, then $M_1 = \mathrm{GL}_1 \times \mathrm{GL}_1 = T_0$ (diagonally embedded into G), or if $G = \mathrm{GL}_3$, then $M_1 = \mathrm{GL}_1 \times \mathrm{GL}_1 \times \mathrm{GL}_1 = T_0$ or $M_1 = \mathrm{GL}_2 \times \mathrm{GL}_1$. Then $G_{X_1} \subseteq M_{1,X_1}$ and $X_1 \in \mathcal{O}_{\mathrm{er}}^{\mathfrak{m}_1}$ so that $A_{M_1} = A_{G_{X_1}}$, where $\mathcal{O}_{\mathrm{er}}^{\mathfrak{m}_1} \subseteq \mathcal{O}^{\mathfrak{m}_1}$ denotes the set of regular elliptic equivalence classes in $\mathfrak{m}_1(\mathbb{Q})$. Let $\mathcal{M} = \{T_0\}$ if $G = \mathrm{GL}_2$, and $\mathcal{M} = \{T_0, \mathrm{GL}_2 \times \mathrm{GL}_1\}$ if $G = \mathrm{GL}_3$. We have a canonical bijection (given by induction of the equivalence classes along the unipotent radical of an arbitrary parabolic subgroup with Levi component M)

$$\bigcup_{M\in\mathcal{M}} \mathcal{O}^{\mathfrak{m}}_{\text{reg, ell}} \longrightarrow \mathcal{O}'_{\text{reg}}. \tag{36}$$

For $\mathfrak{o} \in \mathcal{O}_{\text{reg}}$ the distribution $J^T_{\mathfrak{o}}(\Phi)$ is a weighted orbital integral and equals for sufficiently regular T

$$J^T_{\mathfrak{o}}(\Phi) = \operatorname{vol}(A_{M_1}G_{X_1}(\mathbb{Q})\backslash G_{X_1}(\mathbb{A})) \int_{G_{X_1}(\mathbb{A})\backslash G(\mathbb{A})} \Phi(\operatorname{Ad} x^{-1}X_1)v_1(x,T)\,dx$$

(cf. [Cha02, § 5.2]), where the weight function $v_1(x,T)$ is given by the volume of the convex hull (in $\mathfrak{a}_1^G$) of the projections of the points

$$s^{-1}T - s^{-1}H_P(w_s^{-1}x)$$

where P runs over all standard parabolic subgroups, $s : \mathfrak{a}_1 \longrightarrow \mathfrak{a}_P$ over all isomorphisms obtained by restriction of Weyl group elements, and $w_s \in G(\mathbb{Q})$ is a representative of this Weyl group element. In particular, $v_1(\cdot,T)$ is left $M_1(\mathbb{A})$- and right $\mathbf{K}$-invariant. It is easily seen that this expression for $J^T_{\mathfrak{o}}(\Phi)$ stays true for $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ with ν as in Theorem 5.7.

Proposition 7.2. *Let $T \in \mathfrak{a}^+$ be sufficiently regular and let $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$.*

(i) For $\mathfrak{o} \in \mathcal{O}'_{reg}$ and $X_1 \in \mathfrak{o}$ as before, $v_1(x,T)$ is a polynomial in the variables $\log(q(x,X_1))$ and T with q ranging over a finite collection of polynomials in the coordinate entries of x, X_1 and T.
(ii) There is a constant $C > 0$ depending on Φ such that

$$\sum_{\mathfrak{o}\in\mathcal{O}'_{reg}} \left|J^T_{\mathfrak{o}}(\Phi_\lambda)\right| \leq C\lambda^{-(n^2-\frac{1}{2})}$$

for all $\lambda \in (0,1]$.

Proof.

(i) This is clear from the definition of the weight function.
(ii) Using Iwasawa decomposition, the left $M_1(\mathbb{A})$-, and the right $\mathbf{K}$-invariance of $v_1(\cdot,T)$, we get for every $\mathfrak{o} \in \mathcal{O}'_{\text{reg}}$

$$J^T_{\mathfrak{o}}(\Phi_\lambda) = v^G_{X_1} \int_{M_{1,X_1}(\mathbb{A})\backslash M_1(\mathbb{A})} \int_{U_1(\mathbb{A})} \Phi_\lambda(\operatorname{Ad} u^{-1} \operatorname{Ad} m^{-1}X_1)v_1(u,T)\,du\,dm$$

where we write $v^G_{X_1} = \operatorname{vol}(A_{G_{X_1}}G_{X_1}(\mathbb{Q})\backslash G_{X_1}(\mathbb{A}))$ (note that $v^G_{X_1} = v^{M_1}_{X_1}$). As X_1 and therefore also $\operatorname{Ad} m^{-1}X_1$ is semisimple and regular (X_1 is regular elliptic in $\mathfrak{m}_1$), the map $U_1(\mathbb{A}) \ni u \mapsto U = U(u, \operatorname{Ad} m^{-1}X_1) :=$

$\operatorname{Ad} u^{-1} \operatorname{Ad} m^{-1} X_1 - \operatorname{Ad} m^{-1} X_1 \in \mathfrak{u}_1(\mathbb{A})$ is a diffeomorphism with Jacobian $D(X_1) := \det(\operatorname{ad}(\operatorname{Ad} m^{-1} X_1); \mathfrak{u}_2) = \det(\operatorname{ad} X_1; \mathfrak{u}_2)$. We denote its inverse by $U \mapsto u(U, \operatorname{Ad} m^{-1} X)) \in U_1(\mathbb{A})$. Hence the above integral equals

$$v_{X_1}^G |D(X_1)|_{\mathbb{A}} \int_{M_{1,X_1}(\mathbb{A}) \backslash M_1(\mathbb{A})} \int_{\mathfrak{u}_1(\mathbb{A})} \Phi_\lambda(\operatorname{Ad} m^{-1} X_1 + U) v_1(u(U, \operatorname{Ad} m^{-1} X_1), T)\, dU\, dm$$
$$= v_{X_1}^G \lambda^{-\dim \mathfrak{u}_1} \int_{M_{1,X_1}(\mathbb{A}) \backslash M_1(\mathbb{A})} \int_{\mathfrak{u}_1(\mathbb{A})} \Phi(\lambda \operatorname{Ad} m^{-1} X_1 + U) v_1(u(\lambda^{-1} U, \operatorname{Ad} m^{-1} X_1), T)\, dU\, dm$$

For $Y \in \mathfrak{m}_1(\mathbb{A})$ define

$$\Psi^{M_1}(\lambda, Y) = \int_{\mathfrak{u}_1(\mathbb{A})} \Phi(\lambda Y + U) v_1(u(\lambda^{-1} U, Y), T)\, dU.$$

By the first part of the proposition, we can find a finite collection of polynomials $Q_1, \ldots, Q_m, q_{1,1}, \ldots, q_{1,l_1}, \ldots, q_{m,l_m}$, and integers $k_1, \ldots, k_m \geq 0$ such that

$$|v_1(u(\lambda^{-1} U, Y), T)| \leq \sum_{i=1}^m |\log \lambda|^{k_i} Q_i\big(\log q_{i,1}(U, Y, T), \ldots, \log q_{i,l_i}(U, Y, T)\big)$$

for all $\lambda \in (0, 1]$ and U, Y, and T as before. Then

$$|\Psi^{M_1}(\lambda, Y)| \leq \sum_{i=1}^m |\log \lambda|^{k_i} \tilde{\Psi}_{i,\lambda}^{M_1}(Y),$$

where for $Y \in \mathfrak{m}_1(\mathbb{A})$,

$$\tilde{\Psi}_i^{M_1}(Y) := \int_{\mathfrak{u}_1(\mathbb{A})} \tilde{\Phi}(Y + U) Q_i\big(\log q_{i,1}(U, Y, T), \ldots, \log q_{i,l_i}(U, Y, T)\big)\, dU$$

and $\tilde{\Psi}_{i,\lambda}^{M_1}(Y) := \tilde{\Psi}_i^{M_1}(\lambda Y)$. Here $\tilde{\Phi} \in \mathcal{S}(\mathfrak{g}(\mathbb{A}))$ is a suitable smooth function satisfying the seminorm estimates as in Lemma 3.14 and such that $\tilde{\Phi} \geq |\Phi|$. Then $\tilde{\Psi}_i^{M_1} \in \mathcal{S}(\mathfrak{m}_1(\mathbb{A}))$, and

$$|J_{\mathfrak{o}}^T(\Phi_\lambda)| \leq \lambda^{-\dim \mathfrak{u}_1} \sum_{i=1}^m |\log \lambda|^{k_i} J_{\mathfrak{o}^{\mathfrak{m}_1}}^{M_1, T^{M_1}}(\tilde{\Psi}_i^{M_1}),$$

where $\mathfrak{o}^{\mathfrak{m}_1} \in \mathcal{O}_{\text{reg, ell}}^{\mathfrak{m}_1}$ denotes the inverse image of $\mathfrak{o}$ under the map (36), T^{M_1} is the projection of T onto $\mathfrak{a}_0^{M_1}$, and $J_{\mathfrak{o}^{\mathfrak{m}_1}}^{M_1, T^{M_1}}$ denotes the distribution associated with $\mathfrak{o}^{\mathfrak{m}_1}$ with respect to M_1. Hence by Lemma 3.7 there exist constants $C_M > 0$ for every $M \in \mathcal{M}$ (depending on Φ) such that for every $\lambda \in (0, 1]$ we have

$$\sum_{\mathfrak{o}\in\mathcal{O}_{\mathrm{reg}}} |J_{\mathfrak{o}}^T(\Phi_\lambda)| \le \sum_{M\in\mathcal{M}} \lambda^{-\dim\mathfrak{u}} \sum_{i=1}^m |\log\lambda|^{k_i} \sum_{\mathfrak{o}'\in\mathcal{O}_{\mathrm{reg,ell}}^{\mathfrak{m}}} J_{\mathfrak{o}'}^{M_1,T^{M_1}}(\tilde{\Psi}_i^{M_1})$$
$$\le \sum_{M\in\mathcal{M}} \lambda^{-\dim\mathfrak{u}} \sum_{i=1}^m |\log\lambda|^{k_i} C_M \lambda^{-\dim\mathfrak{m}}$$
$$\le C \sum_{M\in\mathcal{M}} \lambda^{-\dim\mathfrak{p}} \sum_{i=1}^m |\log\lambda|^{k_i},$$

where $C = \max_M C_M$. Since $\dim\mathfrak{p} \le \dim\mathfrak{g} - 1$ for every $M \in \mathcal{M}$, the assertion follows by using some trivial estimate of the form $|\log\lambda|^{k_i} \le c_i\lambda^{-1/2}$, $c_i > 0$ some constant, for the logarithmic terms. □

Above results together with the fact that all distributions are polynomials in T so that above results hold for every $T \in \mathfrak{a}$ and not only sufficiently regular ones, imply the following:

Corollary 7.3. *If $n = 2$, then $\Xi^T(s,\Phi) - \Xi_{main}(s,\Phi)$ can be holomorphically continued at least to $\Re s > \frac{n+1}{2} - \frac{1}{2} = 1$ for every $T \in \mathfrak{a}$.*

7.3 Contribution of the Classes of Type (2, 1)

For $n = 3$, the contribution from the classes in $\mathcal{O}_{(2,1)}$ is still missing. Let $\mathfrak{o}_{(a,b)} \in \mathcal{O}_{(2,1)}$. Then every semisimple element in $\mathfrak{o}_{(a,b)}$ is over $\mathrm{GL}_3(\mathbb{Q})$ conjugate to $X_s = \mathrm{diag}(a,a,b)$ so that $G_{X_s} = \mathrm{GL}_2 \times \mathrm{GL}_1 =: M_2$ (diagonally embedded in GL_3). Let $P_2 = M_2U_2$ denote the standard parabolic subgroup with Levi component M_2. Note that $\mathfrak{o}_{(a,b)} = a\mathbf{1}_3 + \mathfrak{o}_{(0,b-a)}$.

Proposition 7.4. *Let $T \in \mathfrak{a}^+$ be sufficiently regular and let $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$. There exists a constant $C > 0$ depending on Φ such that*

$$\left|\sum_{\mathfrak{o}\in\mathcal{O}_{(2,1)}} \Phi_{\mathfrak{o}}^T(\Phi_\lambda)\right| \le C\lambda^{-n^2+1}$$

for all $\lambda \in (0,1]$.

Proof. We use the same idea as for the central contribution so that we need to consider the sum-integrals

$$\int_{A_G P_1(\mathbb{Q})\backslash G(\mathbb{A})} F^{P_1}(x,T)\tau_{P_1}^G(H_0(x)-T) \sum_{\mathfrak{o}\in\mathcal{O}_{(2,1)}} \sum_{X\in\tilde{\mathfrak{m}}_{P_1}^G(\mathbb{Q})\cap\mathfrak{o}} \left|\Phi(\lambda\mathrm{Ad}x^{-1}X)\right| dx$$

for standard parabolic subgroups $P_1 \subseteq G$.

Suppose first that $P_1 = P_0$ is minimal. Then $X \in \tilde{\mathfrak{m}}_{P_0}^G(\mathbb{Q})$ if and only if $X_{31} \neq 0$ or $X_{31} = 0$ and $X_{21} \neq 0 \neq X_{32}$. The sum-integral restricted to X satisfying the second property $X_{31} = 0$ gives an upper bound as asserted by the same reasons as before. Hence it suffices to consider

$$\int_{A_G P_0(\mathbb{Q})\backslash G(\mathbb{A})} \tau_{P_0}^G(H_0(x) - T) \sum_{\mathfrak{o}\in\mathcal{O}_{(2,1)}} \sum_{X\in\mathfrak{o}: X_{31}\neq 0} \left|\Phi(\lambda \operatorname{Ad} x^{-1} X)\right| dx.$$

As remarked before, we have $\bigcup_{\mathfrak{o}\in\mathcal{O}_{(2,1)}} \mathfrak{o} = \bigcup_{a\in\mathbb{Q}} \left(a\mathbf{1}_3 + \bigcup_{b\in\mathbb{Q}\backslash\{0\}} \mathfrak{o}_{(0,b)}\right)$. If $Y \in \mathfrak{o}_{(0,b)}$ and $Y_{31} \neq 0$, then $\det Y = 0$ and there exists $u \in U_2(\mathbb{Q})$ such that $Z := \operatorname{Ad} uY$ satisfies $Z_{31} = Y_{31} \neq 0$ and $Z_{13} = Z_{23} = Z_{33} = 0$, or there exists $u \in U_0^{M_2}(\mathbb{Q})$ such that $Z := \operatorname{Ad} uY$ satisfies $Z_{31} = Y_{31} \neq 0$ and $Z_{12} = Z_{22} = Z_{32} = 0$. Let $V_{3,1}^i \subseteq \mathfrak{g}(\mathbb{Q})$, $i = 2, 3$, denote the elements $Z \in \mathfrak{g}(\mathbb{Q})$ with $Z_{31} \neq 0$ and $Z_{1i} = Z_{2i} = Z_{3i} = 0$. Then the above integral is bounded by

$$\int_{A_G U_0^{M_2}(\mathbb{Q})\backslash G(\mathbb{A})} \tau_0^G(H_0(x) - T) \sum_{a\in\mathbb{Q}} \sum_{Z\in V_{3,1}^3} \left|\Phi(\lambda \operatorname{Ad} x^{-1}(a\mathbf{1}_3 + Z))\right| dx$$
$$+ \int_{A_G U_2(\mathbb{Q})\backslash G(\mathbb{A})} \tau_0^G(H_0(x) - T) \sum_{a\in\mathbb{Q}} \sum_{Z\in V_{3,1}^2} \left|\Phi(\lambda \operatorname{Ad} x^{-1}(a\mathbf{1}_3 + Z))\right| dx.$$

From this it follows similarly as in the central case that the integral satisfies the asserted upper bound.

The remaining cases $P_0 \subsetneq P_1 \subseteq G$ are combinations of the previous case and the considerations for the central contribution. We omit the details. $\square$

Corollary 7.5. *If $n = 3$, then $\Xi^T(s, \Phi) - \Xi_{main}(s, \Phi)$ can be holomorphically continued at least to $\Re s > \frac{n+1}{2} - \frac{1}{2} = \frac{3}{2}$ for every $T \in \mathfrak{a}$.*

Part 2. Density Results for the Cubic Case

The purpose of this second part of the paper is to give upper and lower bounds (see Theorem 10.1 and Proposition 10.3) for the mean value

$$X^{-\frac{5}{2}} \sum_{E:\, m_1(E)\leq X} \operatorname*{res}_{s=1} \zeta_E(s) \tag{37}$$

as $X \to \infty$, where E runs over all totally real cubic fields and $m_1(E)$ denotes the second successive minimum of the trace form on the ring of integers of E, see below for a definition. For the upper bound we study the main part of the zeta function $\Xi_{\text{main}}(s, \Phi)$ for GL_3 for suitable test functions Φ. As explained above, the distributions $J_{\mathfrak{o}}(\Phi)$ for $\mathfrak{o} \in \mathcal{O}_{\text{er}}$ occurring in the definition of $\Xi_{\text{main}}(s, \Phi)$ are orbital integrals over orbits of regular elliptic elements. Hence in Sect. 8 we first study the local orbital integrals at the non-Archimedean places. In Sect. 9 we define suitable test functions and show an asymptotic for mean values of orbital integrals by using results from Part 1, before finally proving the asymptotic upper and lower bounds for (37) in Sect. 10.

8 Non-Archimedean Orbital Integrals

In this section let $G = \mathrm{GL}_n$ and $\mathfrak{g} = \mathfrak{gl}_n$ with $n \geq 2$ arbitrary. If E is an n-dimensional field extension of $\mathbb{Q}$, let $\mathcal{O}_E$ be the ring of integers of E. For $\gamma \in G(\mathbb{Q})$ let $[\gamma] = \{x^{-1}\gamma x \mid x \in G(\mathbb{Q})\}$ be the conjugacy class of γ in $G(\mathbb{Q})$. As before, let $G(\mathbb{Q})_{\mathrm{er}}$ denote the set of regular elliptic elements in $G(\mathbb{Q})$. Let $\mathcal{F}_n$ be the set of n-dimensional number fields. We get a surjective map from $G(\mathbb{Q})_{\mathrm{er}}$ onto the set of isomorphism classes in $\mathcal{F}_n$ by attaching to $\gamma \in G(\mathbb{Q})_{\mathrm{er}}$ the conjugacy class of the field $\mathbb{Q}(\xi)$ for ξ an (arbitrary) eigenvalue of γ. If $[E] \subseteq \mathcal{F}_n$ is such a conjugacy class and if $\Gamma_{[E]} \subseteq G(\mathbb{Q})_{\mathrm{er}}$ is the inverse image of $[E]$ under this map, then $\Gamma_{[E]}$ is invariant under conjugation by elements of $G(\mathbb{Q})$, and

$$\{\xi \in E \mid \mathbb{Q}(\xi) = E\} \longrightarrow [\gamma_\xi] \in \Gamma_{[E]}/\sim$$

is surjective. Here $\gamma_\xi \in G(\mathbb{Q})$ denotes the companion matrix of the characteristic polynomial of ξ, and the map is $|\operatorname{Aut}(E/\mathbb{Q})|$-to-1.

If K is a finite field extension of $\mathbb{Q}_p$ with ring of integers $\mathcal{O}_K \subseteq K$, we normalise the measures on K and $K^\times$ such that $\mathrm{vol}(\mathcal{O}_K) = 1 = \mathrm{vol}(\mathcal{O}_K^\times)$. If $\theta \in \mathcal{O}_K$ is such that $\{1, \theta, \ldots, \theta^{n-1}\}$ is a basis of K over $\mathbb{Q}_p$, let $\gamma_\theta \in \mathrm{GL}_n(\mathbb{Q}_p)$ denote the companion matrix of θ. Then $G_{\gamma_\theta}(\mathbb{Q}_p)$ is isomorphic to $K^\times$ via the isomorphism induced by $\{1, \theta, \ldots, \theta^{n-1}\}$ and we define the measure on $G_{\gamma_\theta}(\mathbb{Q}_p)$ via this isomorphism. If $\Phi_p \in \mathcal{S}(\mathfrak{g}(\mathbb{Q}_p))$, we define the p-adic orbital integrals

$$I_p(\Phi_p, \theta) = I_p(\Phi_p, \gamma_\theta) = \int_{G_{\gamma_\theta}(\mathbb{Q}_p)\backslash G(\mathbb{Q}_p)} \Phi_p(g^{-1}\gamma_\theta g)\, dg.$$

If $\gamma \in G(\mathbb{Q})_{\mathrm{er}}$, then $G_\gamma(\mathbb{Q}_p)$ is isomorphic to a direct product of $K_1^\times \times \ldots \times K_r^\times$ for suitable finite field extensions $K_1, \ldots, K_r/\mathbb{Q}_p$ and we choose the measure on $G_\gamma(\mathbb{Q}_p)$ such that it is compatible with our choice of measures on $K_1^\times \times \ldots \times K_r^\times$, and put $I_f(\Phi_f, \gamma) = \prod_{p<\infty} I_p(\Phi_p, \gamma)$. Similarly, we define $I_\infty(\Phi_\infty, \gamma)$ (resp., $I(\Phi, \gamma)$) if $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (resp., $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$).

Our aim in this section is to understand the quantities

$$c(\Phi_p, \gamma) = \frac{I_p(\Phi_p, \gamma)}{[\mathcal{O}_{\mathbb{Q}_p[\gamma]} : \mathbb{Z}_p[\gamma]]}, \quad \text{and} \quad c(\Phi_f, \gamma) = \frac{I_f(\Phi_f, \gamma)}{[\mathcal{O}_{\mathbb{Q}[\gamma]} : \mathbb{Z}[\gamma]]},$$

where we denote for a $\mathbb{Q}$- or $\mathbb{Q}_p$-algebra A the ring of integers of A by $\mathcal{O}_A$. If $\xi \in E$ generates E over $\mathbb{Q}$, we set $I_p(\Phi_p, \xi) = I_p(\Phi_p, \gamma_\xi)$, and define $I_f(\Phi_p, \xi)$, $I(\Phi, \xi)$, $c(\Phi_p, \xi)$, $c(\Phi_f, \xi)$ analogously.

For a prime p and $E \in \mathcal{F}_n$ let $E_p = E \otimes_{\mathbb{Q}} \mathbb{Q}_p$. If Φ_p (resp., Φ_f) is supported in $\mathfrak{g}(\mathbb{Z}_p)$ (resp., $\mathfrak{g}(\hat{\mathbb{Z}})$), then the orbital integral $I_p(\Phi_p, \xi)$ (resp., $I_f(\Phi_f, \xi)$) vanishes unless $\xi \in \mathcal{O}_{E_p}$ (resp., $\xi \in \mathcal{O}_E$). We denote by $\Phi_p^0 \in \mathcal{S}(\mathfrak{g}(\mathbb{Q}_p))$ the characteristic function of $\mathfrak{g}(\mathbb{Z}_p)$, and $\Phi_f^0 = \prod_{p<\infty} \Phi_p^0 \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$.

Proposition 8.1. *Let $E \in \mathcal{F}_n$ and $\xi \in \mathcal{O}_E$ be such that $\mathbb{Q}(\xi) = E$. Then*

(i) If $E_p \simeq K_1 \oplus \ldots \oplus K_r$ with $K_i/\mathbb{Q}_p$ field extensions, and if under this isomorphism ξ corresponds to $(\xi_1, \ldots, \xi_r) \in K_1 \oplus \ldots \oplus K_r$, we have

$$c(\Phi_p^0, \xi) = \prod_{i=1}^{r} c(\Phi_p^0, \xi_i),$$

where Φ_p^0 also denotes the characteristic function of $\mathfrak{g}_{n_i}(\mathbb{Z}_p)$, $n_i := [K_i : \mathbb{Q}_p]$, and c_p is defined on the smaller groups similar as before.

(ii) $c(\Phi_p^0, \xi) \geq 1$.

(iii) $c(\Phi_p^0, \xi + a) = c(\Phi_p^0, \xi)$ for every $a \in \mathbb{Z}$. Hence $c(\Phi_f^0, \cdot)$ is a well defined function on $\mathcal{O}_E/\mathbb{Z}$.

Before proving this proposition we need a few auxiliary results and fix some further notation. If ξ as is in the proposition, denote by $P_{p,\xi}$ the standard parabolic subgroup of type $(n_1, \ldots, n_r)$. Then

$$I_p(\Phi_p^0, \xi) = \delta_{P_{p,\xi}}(\mathrm{diag}(\xi_1, \ldots, \xi_r))^{-1/2} \prod_{i=1}^{r} I_p(\Phi_p^0, \xi_i). \tag{38}$$

Let Δ denote the discriminant map for $E \longrightarrow \mathbb{Q}$ as well as for $\mathfrak{g}(\mathbb{Q}) \longrightarrow \mathbb{Q}$ and $F \longrightarrow \mathbb{Q}_p$ for $F/\mathbb{Q}_p$ a finite field extensions of arbitrary degree. If F is either $\mathbb{Q}$ or $\mathbb{Q}_p$ for some prime $p < \infty$, let A be a finite-dimensional semisimple F-algebra, and $R \subseteq A$ an $\mathcal{O}_F$-order. We denote by $\mathrm{Frac}(R)$ the set of fractional ideals of R in A, i.e. the set of all full-rank $\mathcal{O}_F$-lattices $\mathfrak{a} \subseteq A$ such that $R\mathfrak{a} \subseteq \mathfrak{a}$. If $\mathfrak{a} \subseteq A$ is a lattice of full rank, let $\mathcal{M}(\mathfrak{a}) = \{a \in A \mid a\mathfrak{a} \subseteq \mathfrak{a}\}$ be the multiplier of $\mathfrak{a}$. This is an $\mathcal{O}_F$-order in A, in particular $\mathcal{M}(\mathfrak{a}) \subseteq \mathcal{O}_K$ and $\mathfrak{a} \in \mathrm{Frac}(\mathcal{M}(\mathfrak{a}))$. Let

$$\mathrm{Frac}^0(R) = \{\mathfrak{a} \in \mathrm{Frac}(R) \mid \mathcal{M}(\mathfrak{a}) = R\}.$$

Let $P(R) = \{aR \mid a \in A^\times\}$ be the set of all R-principal ideals in A. In general, neither $\mathrm{Frac}(R)$ nor $\mathrm{Frac}^0(R)$ are groups, but they are acted on by $P(R)$ so that we may build the quotients $\mathrm{Frac}(R)/P(R)$ and $\mathrm{Frac}^0(R)/P(R)$, which are both finite.

Lemma 8.2. *Suppose K is a finite field extension of $\mathbb{Q}_p$, and $\theta \in \mathcal{O}_K$ generates K over $\mathbb{Q}_p$, i.e. $K = \mathbb{Q}_p(\theta)$. Then*

$$I_p(\Phi_p^0, \theta) = \sum_{\mathfrak{o} \subseteq \mathcal{O}_K:\ \theta \in \mathfrak{o}} \left| \mathrm{Frac}^0(\mathfrak{o})/P(\mathfrak{o}) \right| \left[\mathcal{O}_K^\times : \mathfrak{o}^\times\right]$$

where $\mathfrak{o}$ runs over all $\mathbb{Z}_p$-orders in $\mathcal{O}_K$ containing θ.

Remark 8.3. If $\mathbb{Z}_p[\theta] = \mathcal{O}_K$, then $I_p(\Phi_p^0, \theta) = 1$.

Proof. We first show

$$\int_{\mathbb{Q}_p^\times \backslash G(\mathbb{Q}_p)} \Phi_p^0(g^{-1}\gamma_\theta g)\, dg = [K^\times : (\mathcal{O}_K^\times \mathbb{Q}_p^\times)] \sum_{\mathfrak{o} \subseteq \mathcal{O}_K:\ \theta \in \mathfrak{o}} \left| \mathrm{Frac}^0(\mathfrak{o})/P(\mathfrak{o}) \right| \left[\mathcal{O}_K^\times : \mathfrak{o}^\times\right]. \tag{39}$$

The set $\{1, \theta, \dots, \theta^{n-1}\}$ forms a basis of K relative to which the matrix γ_θ corresponds to the endomorphism $K \longrightarrow K$ given by multiplication with θ. Moreover, this basis defines a map

$$\Psi : G(\mathbb{Q}_p) = \mathrm{GL}_n(\mathbb{Q}_p) \longrightarrow \mathcal{L}_p = \{L \subseteq K \mid L \text{ is } \mathbb{Z}_p\text{-lattice of full rank}\}.$$

Hence $\Phi_p^0(g^{-1}\gamma_\theta g) \neq 0$ if and only if θ maps the lattice $L_g = g\mathcal{O}_K \subseteq K$ defined by g into itself, i.e. $\theta L_g \subseteq L_g$, or equivalently $\theta \in \mathcal{M}(L_g) \subseteq \mathcal{O}_K$. Hence the integral equals

$$\sum_{\mathfrak{o} \subseteq \mathcal{O}_K:\ \theta \in \mathfrak{o}} \ \sum_{\mathfrak{a} \in \mathrm{Frac}^0(\mathfrak{o})/\mathbb{Q}_p^\times} \mathrm{vol}\left(\Psi^{-1}(\mathfrak{a})\right).$$

Hence we have to compute the volume of $\Psi^{-1}(\mathfrak{a})$ as a subset of $G(\mathbb{Q}_p)$. Now two elements $g_1, g_2 \in G(\mathbb{Q}_p)$ define the same $\mathbb{Z}_p$-lattice if and only if there exists $k \in G(\mathbb{Z}_p) = \mathbf{K}_p$ with $g_2 = g_1 k$. Hence with our normalisation of measures we get $\mathrm{vol}\left(\Psi^{-1}(\mathfrak{a})\right) = 1$. Since

$$\begin{aligned} \left|\mathrm{Frac}^0(\mathfrak{o})/\mathbb{Q}_p^\times\right| = \left|\mathrm{Frac}^0(\mathfrak{o})/(\mathfrak{o}^\times\mathbb{Q}_p^\times)\right| &= \left|\mathrm{Frac}^0(\mathfrak{o})/(\mathcal{O}_K^\times\mathbb{Q}_p^\times)\right| \left[\mathcal{O}_K^\times : \mathfrak{o}^\times\right] \\ &= \left|\mathrm{Frac}^0(\mathfrak{o})/P(\mathfrak{o})\right| \left|K^\times/(\mathcal{O}_K^\times\mathbb{Q}_p^\times)\right| \left[\mathcal{O}_K^\times : \mathfrak{o}^\times\right], \end{aligned}$$

the assertion (39) follows. If the extension $K/\mathbb{Q}_p$ is unramified, $[K^\times : (\mathcal{O}_K^\times \mathbb{Q}_p^\times)] = 1$. In general, $[K^\times : (\mathcal{O}_K^\times \mathbb{Q}_p^\times)] = [K : \mathcal{O}_K^\times \mathbb{Q}_p]$ so that this index equals the ramification index, and we therefore have $[K^\times : (\mathcal{O}_K^\times \mathbb{Q}_p^\times)] = \mathrm{vol}(\mathbb{Q}_p^\times \backslash K^\times) = \mathrm{vol}(\mathbb{Q}_p^\times \backslash G_\theta(\mathbb{Q}_p))$. Hence the assertion of the lemma follows. □

Proof of Proposition 8.1. (i) This follows from (38) and the identity

$$[\mathcal{O}_E : \mathbb{Z}_p[\xi]]^2 \delta_{p,\xi}(\mathrm{diag}(\xi_1, \dots, \xi_r)) = \prod_{i=1}^r [\mathcal{O}_{K_i} : \mathbb{Z}_p[\xi_i]]^2.$$

(ii) By (i) the quotient $c(\Phi_p^0,\xi)$ equals a finite product of terms of the form

$$\frac{[\mathcal{O}_E^\times:\mathbb{Z}_p[\theta]^\times]}{[\mathcal{O}_E:\mathbb{Z}_p[\theta]]}\left|\operatorname{Frac}^0(\mathbb{Z}_p[\theta])/P(\mathbb{Z}_p[\theta])\right|$$
$$+\frac{1}{[\mathcal{O}_E:\mathbb{Z}_p[\theta]]}\sum_{\substack{\mathfrak{o}:\\ \mathbb{Z}_p[\theta]\subsetneq\mathfrak{o}\subseteq\mathcal{O}_E}}\left|\operatorname{Frac}^0(\mathfrak{o})/P(\mathfrak{o})\right|[\mathcal{O}_E^\times:\mathfrak{o}^\times]$$

for $E/\mathbb{Q}_p$ a finite extension generated by $\theta\in E$ with maximal ideal $\mathfrak{p}\subseteq\mathcal{O}_E$ of norm q. Hence it certainly suffices to show $\frac{[\mathcal{O}_E^\times:\mathbb{Z}_p[\theta]^\times]}{[\mathcal{O}_E:\mathbb{Z}_p[\theta]]}\geq 1$, since $\left|\operatorname{Frac}^0(\mathbb{Z}_p[\theta])/P(\mathbb{Z}_p[\theta])\right|\geq 1$ and the rest of the sum is non-negative.

To show this, let $\mathfrak{f}\subseteq\mathbb{Z}_p[\theta]$ denote the conductor of $\mathbb{Z}_p[\theta]$. Then $\mathfrak{p}/\mathfrak{f}\subseteq\mathcal{O}_E/\mathfrak{f}$ is the unique maximal ideal so that $(\mathfrak{p}\cap\mathbb{Z}_p[\theta])/\mathfrak{f}$ is the unique maximal ideal in $\mathbb{Z}_p[\theta]/\mathfrak{f}$. Hence

$$\#(\mathcal{O}_E/\mathfrak{f})^\times=\#(\mathcal{O}_E/\mathfrak{f})-\#(\mathfrak{p}/\mathfrak{f})=\#(\mathcal{O}_E/\mathfrak{f})(1-q^{-1}),\quad\text{and}$$
$$\#(\mathbb{Z}_p[\theta]/\mathfrak{f})^\times=\#(\mathbb{Z}_p[\theta]/\mathfrak{f})(1-(\#(\mathbb{Z}_p[\theta]/(\mathbb{Z}_p[\theta]\cap\mathfrak{p})))^{-1}).$$

But since $\mathbb{Z}_p[\theta]/(\mathbb{Z}_p[\theta]\cap\mathfrak{p})\hookrightarrow\mathcal{O}_E/\mathfrak{p}$ is injective, we altogether get

$$\frac{[\mathcal{O}_E^\times:\mathbb{Z}_p[\theta]^\times]}{[\mathcal{O}_E:\mathbb{Z}_p[\theta]]}=\frac{1-q^{-1}}{1-(\#(\mathbb{Z}_p[\theta]/(\mathbb{Z}_p[\theta]\cap\mathfrak{p})))^{-1}}\geq 1.$$

(iii) This is a direct consequence of the explicit form of the orbital integral from Lemma 8.2. □

9 An Asymptotic for Orbital Integrals

From now let $G=\mathrm{GL}_3$ and $\mathfrak{g}=\mathfrak{gl}_3$. The aim of this section is to prove a density result for orbital integrals, namely Proposition 9.2 below. If $\gamma\in G(\mathbb{Q})_{\mathrm{er}}$, we take the product measure on $G_\gamma(\mathbb{A})=\prod_{p\leq\infty}G_\gamma(\mathbb{Q}_p)$ with local measures as in the previous section. Let $|\cdot|_E:\mathbb{A}_E^\times\longrightarrow\mathbb{R}_{>0}$ denote the adelic norm. Using the exact sequence $1\longrightarrow\mathbb{A}_E^1\hookrightarrow\mathbb{A}_E^\times\xrightarrow{|\cdot|_E}\mathbb{R}_{>0}\longrightarrow 1$, we also fix a measure on $\mathbb{A}_E^1$. With this choice of normalisation of measures we get

$$\mathrm{vol}(\mathbb{R}_{>0}G_{\gamma_\xi}(\mathbb{Q})\backslash G_{\gamma_\xi}(\mathbb{A}))=\mathrm{vol}(E^\times\backslash\mathbb{A}_E^1)=\rho_E|D_E|^{\frac{1}{2}}$$

for every $\xi\in E$ with $\mathbb{Q}(\xi)=E$, where

$$\rho_E=\operatorname*{res}_{s=1}\zeta_E(s).$$

For a cubic field E the set of $\xi \in E$ generating E over $\mathbb{Q}$ is exactly $E\backslash\mathbb{Q}$, as E does not have non-trivial subfields. For $\Phi \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$, we therefore have

$$\Xi_{\text{main}}(s,\Phi) = \sum_{E\in\mathcal{F}_3} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} |D_E|^{\frac{1}{2}} \sum_{\xi\in E\backslash\mathbb{Q}} \int_0^\infty \int_{G_{\gamma_\xi}(\mathbb{A})\backslash G(\mathbb{A})} \lambda^{3s+3} \Phi(\lambda g^{-1}\gamma_\xi g)\, d^\times\lambda\, dg. \tag{40}$$

Let $\mathcal{F}_3^+ \subseteq \mathcal{F}_3$ be the set of all totally real cubic number fields, and $E \in \mathcal{F}_3^+$. Let $Q_E : \mathcal{O}_E/\mathbb{Z} \longrightarrow \mathbb{R}$ be the positive definite quadratic form $Q_E(\xi) = \operatorname{tr}_{E/\mathbb{Q}} \xi^2 - \frac{1}{3}(\operatorname{tr}_{E/\mathbb{Q}} \xi)^2$. We denote its successive minima by $m_1(E) \le m_2(E)$, and its discriminant by $\Delta(Q_E)$. Similarly, $Q : \mathfrak{g}(\mathbb{A}) \longrightarrow \mathbb{A}$ denotes the quadratic form on the matrices given by $Q(x) = \operatorname{tr} x^2 - \frac{1}{3}(\operatorname{tr} x)^2$.

Remark 9.1. We have $3\Delta(Q_E) = D_E$.

Proposition 9.2. *Let $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$ be supported in $\mathfrak{g}(\widehat{\mathbb{Z}})$, and suppose that $c(\Phi_f, \gamma + a) = c(\Phi_f, \gamma)$ for all $\gamma \in G(\mathbb{Q})$ and $a \in \mathbb{Z}$. Then*

$$\sum_{E\in\mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z}, \xi\neq 0\\ Q_E(\xi)\le X}} c(\Phi_f, \xi) = \beta(\Phi_f) X^{\frac{5}{2}} + o(X^{\frac{5}{2}}) \tag{41}$$

for $X \to \infty$, and $\beta(\Phi_f)$ is a certain constant depending on Φ_f with $\beta(\Phi_f^0) \neq 0$.

The proof of this proposition will occupy the rest of this section.

Remark 9.3.

(i) The constraint on the support of Φ_f is not essential, it only changes the lattices in E one has to sum over.
(ii) It is possible to find an analogue of the asymptotic (41) also for fields with a complex place. However, one has to replace Q_E, since Q_E is no longer positive definite if E has a complex place.

9.1 Test Functions

We want to use the analytic properties of $\Xi_{\text{main}}(s,\Phi)$ to prove the proposition, hence our first task is to find test functions which separate the totally real fields from the rest. To this end, we first construct two sequences of test functions at the Archimedean places. Let $\psi_\varepsilon^\pm : \mathbb{R} \to \mathbb{R}_{\ge 0}$ be smooth non-negative functions satisfying

$$\psi_\varepsilon^+(x) = 0 \quad \text{if} \quad x < \frac{\varepsilon}{2}, \quad 0 \le \psi_\varepsilon^+(x) \le 1 \quad \text{if} \quad \frac{\varepsilon}{2} \le x \le \varepsilon, \quad \text{and } \psi_\varepsilon^+(x) = 1 \quad \text{if} \quad x > \varepsilon,$$

$$\psi_\varepsilon^-(x) = 0 \quad \text{if} \quad |x| > \varepsilon, \quad \text{and} \quad 0 \le \psi_\varepsilon^-(x) \le 1, \quad \text{if} \quad |x| \le \varepsilon,$$

and

$$1 \le \psi_\varepsilon^+(x) + \psi_\varepsilon^-(x) \le 2 \qquad \text{if } x > 0.$$

Define functions $\Psi_\varepsilon^\pm : \mathfrak{g}(\mathbb{R}) \longrightarrow \mathbb{R}$ by

$$\Psi_\varepsilon^\pm(x) = \psi_\varepsilon^\pm\left(\frac{\Delta(x - \frac{1}{3}\operatorname{tr} x \mathbf{1}_3)}{|\operatorname{tr} x^2 - \frac{1}{3}(\operatorname{tr} x)^2|^3}\right) = \psi_\varepsilon^\pm\left(\frac{\Delta(x - \frac{1}{3}\operatorname{tr} x \mathbf{1}_3)}{|Q(x)|^3}\right).$$

These functions are well defined and continuous, since $\psi_\varepsilon^\pm$ is compactly supported. Moreover, away from the set of x with $Q(x) = \operatorname{tr} x^2 - \frac{1}{3}(\operatorname{tr} x)^2 = 0$ they are smooth.

For $x \in \mathfrak{g}(\mathbb{R})$ and large $N \in \mathbb{N}$ put

$$\Phi_\infty^{\varepsilon,\pm}(x) = \psi_\varepsilon^\pm\left(\frac{\Delta(x - \frac{1}{3}\operatorname{tr} x \mathbf{1}_3)}{|Q(x)|^3}\right) Q(x)^N e^{-\pi \operatorname{tr} x^t x} = \Psi_\varepsilon^\pm(x) Q(x)^N e^{-\pi \operatorname{tr} x^t x}.$$

For given $\nu \in \mathbb{N}$, we can choose N large enough such that $\Phi_\infty^{\varepsilon,\pm} \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$. The properties of $\Phi_\varepsilon^\pm$ can be summarised as follows.

Lemma 9.4. *For all $x \in \mathfrak{g}(\mathbb{R})$, $g \in G(\mathbb{R})$, and $\lambda \in \mathbb{R}_{>0}$, we have*

(i) $\Phi_\infty^{\varepsilon,\pm}(\operatorname{Ad} g^{-1} x) = \Phi_\varepsilon^\pm(x)$. *In particular, we may write $\Phi_\infty^{\varepsilon,\pm}(\xi) = \Phi_\infty^{\varepsilon,\pm}(\gamma_\xi)$ for every $\xi \in E$ and $E \in \mathcal{F}_3$.*
(ii) $\Phi_\infty^{\varepsilon,\pm}(\lambda x) = \Phi_\infty^{\varepsilon,\pm}(x)$.
(iii) $\Phi_\infty^{\varepsilon,\pm}(x + \lambda \mathbf{1}_3) = \Phi_\infty^{\varepsilon,\pm}(x)$.
(iv) $\Phi_\infty^{\varepsilon,+}(\lambda \operatorname{Ad} g^{-1} x) = 0$ *if x has a non-real eigenvalue.*

If we fix $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$ as in Proposition 9.2, we define test functions $\Phi^{\varepsilon,+} = \Phi_\infty^{\varepsilon,+}\Phi_f$ and $\Phi^{\varepsilon,-} = \Phi_\infty^{\varepsilon,-}\Phi_f$. They implicitly depend on the integer N, and $\Phi^{\varepsilon,\pm} \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{A}))$ with ν depending on N.

By Lemma 9.4(iv) we have $I_\infty(\Phi^{\varepsilon,+}, \gamma) = 0$ if $\gamma \in G(\mathbb{Q})_{\mathrm{er}}$ is not diagonisable over $G(\mathbb{R})$. Hence for the test function $\Phi^{\varepsilon,+}$ only totally real fields contribute to $\Xi_{\mathrm{main}}(s, \Phi^{\varepsilon,+})$, i.e. we get

$$\mathcal{E}_\varepsilon^+(s) := \Xi_{\mathrm{main}}(s, \Phi^{\varepsilon,+}) = \sum_{E \in \mathcal{F}_3^+} \frac{\operatorname{vol}(E^\times \backslash \mathbb{A}_E^1)}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\xi \in \mathcal{O}_E \backslash \mathbb{Z}} [\mathcal{O}_E : \mathbb{Z}[\xi]] c(\Phi_f, \xi) \Psi_\varepsilon^+(\xi) \cdot$$
$$\left(\int_0^\infty \int_{G_{\gamma_\xi}(\mathbb{R}) \backslash G(\mathbb{R})} \lambda^s (\lambda^2 Q_E(\xi))^N e^{-\pi \lambda^2 \operatorname{tr}(x^{-1}\gamma_\xi x)^t (x^{-1}\gamma_\xi x)} \, d^\times\lambda \, dx \right). \qquad (42)$$

Similarly, we set $\mathcal{E}_\varepsilon^-(s) = \Xi_{\mathrm{main}}(s, \Phi^{\varepsilon,-})$.

Remark 9.5. Separating the totally real fields from the rest is more complicated in the cubic than in the quadratic case. This is due to the absence of a prehomogeneous vector space structure so that there are infinitely many orbits under the action of $\mathrm{GL}_1 \times \mathrm{GL}_3$ on $\mathfrak{g}(\mathbb{A})$.

Lemma 9.6. *There exists $N > 0$ such that the following holds. Let Φ_f be as in Proposition 9.2. Then $\mathcal{E}_\varepsilon^+(s)$ is holomorphic for $\Re s > 2$, and has a meromorphic continuation at least in $\Re s > 3/2$ with only singularity at $s = 2$, which is a simple pole. Moreover, for $\Re s > 2$ the function $\mathcal{E}_\varepsilon^+(s)$ equals up to an entire function the series*

$$I_N(s) \sum_{E \in \mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi \in \mathcal{O}_E/\mathbb{Z}: \\ \xi \neq 0}} c(\Phi_f, \xi) \Psi_\varepsilon^+(\xi) Q_E(\xi)^{-\frac{3s-1}{2}}, \tag{43}$$

where for $\Re s > 0$

$$I_N(s) = \frac{1}{\sqrt{3\pi}} \int_0^\infty \lambda^{3s-1+2N} e^{-\pi\lambda^2} \, d^\times \lambda = \frac{1}{\sqrt{3\pi}} \frac{\Gamma(\frac{3s+2N-1}{2})}{2\pi^{\frac{3s+N-1}{2}}}.$$

Proof. The first assertion follows from Theorem 1.2 and Proposition 1.4. Let $E \in \mathcal{F}_3^+$, and consider the map $\mathcal{O}_E \longrightarrow \mathbb{Z} \oplus \mathcal{O}_E/\mathbb{Z}$, $\xi \mapsto (\operatorname{tr}\xi, \xi + \mathbb{Z})$, which is a group isomorphism. As the coefficients $c(\Phi_f, \cdot)$ and the function Ψ_ε^+ are well-defined maps on $\mathcal{O}_E/\mathbb{Z}$, the inner sum for E in (42) equals

$$\sum_{\substack{\xi_0 \in \mathcal{O}_E/\mathbb{Z}: \\ \xi_0 \neq 0}} [\mathcal{O}_E : \mathbb{Z}[\xi_0]] c(\Phi_f, \xi_0) \Psi_\varepsilon^+(\xi_0) \cdot$$
$$\left(\int_0^\infty \int_{G_{\gamma_{\xi_0}}(\mathbb{R}) \backslash G(\mathbb{R})} \lambda^{3s+3+2N} Q_E(\xi_0)^N \sum_{a \in \mathbb{Z}} e^{-\pi\lambda^2 \operatorname{tr}(x^{-1}\gamma_{\xi_0} x)^t (x^{-1}\gamma_{\xi_0} x) - 3\pi\lambda^2 a^2} \, d^\times \lambda \, dx \right).$$

We split the integral over λ in one integral over $\lambda \in [0, 1]$ and one over $\lambda \in [1, \infty)$. The sum over all E of the second integral defines an entire function on all of $\mathbb{C}$ so that we may ignore it. For the sum over the first one we apply Poisson summation to the sum over $a \in \mathbb{Z}$, to obtain

$$\sum_{a \in \mathbb{Z}} e^{-3\pi\lambda^2 a^2} = \sum_{b \in \mathbb{Z}} (3\pi)^{-\frac{1}{2}} \lambda^{-1} e^{-3\pi^{-1}\lambda^{-2} b^2}.$$

Changing variables $\lambda^{-1} \in [0, 1] \leftrightarrow \lambda \in [1, \infty)$, the sum over $b \neq 0$ yields again an entire function which we can ignore. Hence we are left with the term belonging to $b = 0$. We may add the integral over $\lambda \in [1, \infty)$ without changing its analytic behaviour. Thus up to an entire function, $\mathcal{E}_\varepsilon^+(s)$ equals

$$\frac{1}{\sqrt{3\pi}} \sum_{E \in \mathcal{F}_3^+} \frac{\operatorname{vol}(E^\times \backslash \mathbb{A}_E^1)}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi_0 \in \mathcal{O}_E/\mathbb{Z}: \\ \xi_0 \neq 0}} [\mathcal{O}_E : \mathbb{Z}[\xi]] c(\Phi_f, \xi) \Psi_\varepsilon^+(\xi) \cdot$$
$$\left(\int_0^\infty \int_{G_{\gamma_{\xi_0}}(\mathbb{R}) \backslash G(\mathbb{R})} \lambda^{3s+2+2N} Q_E(\xi_0)^N e^{-\pi\lambda^2 \operatorname{tr}(x^{-1}\gamma_{\xi_0} x)^t (x^{-1}\gamma_{\xi_0} x)} \, d^\times \lambda \, dx \right).$$

As E is totally real, for every $\xi_0 \in \mathcal{O}_E/\mathbb{Z}$, the matrix γ_{ξ_0} is over $G(\mathbb{R})$ conjugate to a diagonal matrix (with pairwise distinct eigenvalues) so that

$$\int_{G_{\gamma_{\xi_0}}(\mathbb{R})\backslash G(\mathbb{R})} e^{-\pi\lambda^2 \operatorname{tr}(x^{-1}\gamma_{\xi_0}x)^t(x^{-1}\gamma_{\xi_0}x)}\,dx$$
$$= \Delta(\xi_0)^{-\frac{1}{2}} e^{-\pi\lambda^2 Q_E(\xi_0)} \int_{U_0(\mathbb{R})} e^{-\pi\lambda^2(u_1^2+u_2^2+u_3^2)}\,du = \Delta(\xi_0)^{-\frac{1}{2}} e^{-\pi\lambda^2 Q_E(\xi_0)}\lambda^{-3}.$$

Notice that $\Delta(\xi_0)^{-\frac{1}{2}} = [\mathcal{O}_E : \mathbb{Z}[\xi_0]]^{-1} D_E^{-\frac{1}{2}}$ and $\operatorname{vol}(E^\times\backslash\mathbb{A}_E^1)D_E^{-\frac{1}{2}} = \operatorname{res}_{s=1}\zeta_E(s) = \rho_E$. Hence changing λ to $Q_E(\xi_0)^{\frac{1}{2}}\lambda$, the assertion follows upon defining I_N as described. $\square$

Lemma 9.7. *There exists $N > 0$ such that the following holds. Let Φ_f be as in Proposition 9.2. Then $\mathcal{E}_\varepsilon^-(s)$ is holomorphic for $\Re s > 2$ and continues to a meromorphic function at least in $\Re s > 3/2$ with only pole at $s = 2$ which is simple. Up to an entire function (defined on all of $\mathbb{C}$), $\mathcal{E}_\varepsilon^-(s)$ equals for $\Re s > 2$ the sum of*

$$I_N(s) \sum_{E\in\mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z}:\\ \xi\neq 0}} c(\Phi_f,\xi)\Psi_\varepsilon^-(\xi)Q_E(\xi)^{-\frac{3s-1}{2}}$$

and

$$4\sqrt{\frac{\pi}{3}}\frac{\Gamma(\frac{3s+2l}{2})}{\pi^{\frac{3s+2l}{2}}} \sum_{E\in\mathcal{F}_3\backslash\mathcal{F}_3^+} \rho_E \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z}:\\ \xi\neq 0}} c(\Phi_f,\xi)\Psi_\varepsilon^-(\xi)J_N(\xi,s)Q_E(\xi)^N,$$

where

$$J_N(\xi,s) = \int_1^\infty (Q_E(\xi) + 4(\Im\tilde{\xi})^2\rho^2)^{-\frac{3s+2N}{2}}\,d\rho,$$

and $\tilde{\xi}$ denotes one of the two non-real conjugates of $\xi \in E\backslash\mathbb{Q}$ if $E \in \mathcal{F}_3\backslash\mathcal{F}_3^+$.

Proof. Again, the first assertion is given by Theorem 1.2 and Proposition 1.4. Similarly as in the proof of Lemma 9.6, $\mathcal{E}_\varepsilon^-(s)$ can be written as the sum over all cubic fields $E \in \mathcal{F}$ (now of any signature) of

$$\frac{\operatorname{vol}(E^\times\backslash\mathbb{A}_E^1)}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi_0\in\mathcal{O}_E/\mathbb{Z}:\\ \xi_0\neq 0}} [\mathcal{O}_E : \mathbb{Z}[\xi_0]]c(\Phi_f,\xi_0)\Psi_\varepsilon^-(\xi_0)\cdot$$
$$\left(\int_0^\infty \int_{G_{\gamma_{\xi_0}}(\mathbb{R})\backslash G(\mathbb{R})} \lambda^{3s+3+2N} Q_E(\xi_0)^N \sum_{a\in\mathbb{Z}} e^{-\pi\lambda^2 \operatorname{tr}(x^{-1}\gamma_{\xi_0}x)^t(x^{-1}\gamma_{\xi_0}x)-\frac{\pi}{3}\lambda^2a^2}\,d^\times\lambda\,dx\right).$$

For totally real extensions, the proof of the last lemma tells us that the respective sum essentially (i.e., up to an entire function) equals

$$I_N(s) \sum_{E \in \mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi \in \mathcal{O}_E/\mathbb{Z}: \\ \xi \neq 0}} c(\Phi_f, \xi) \Psi_\varepsilon^-(\xi) Q_E(\xi)^{-\frac{3s-1}{2}},$$

with $I_N(s)$ defined as before.

For $E \in \mathcal{F}_3 \backslash \mathcal{F}_3^+$ and $\xi_0 \in \mathcal{O}_E/\mathbb{Z}$, $\xi_0 \neq 0$, we can follow along the same lines. However, the integral $\int_{G_{\gamma_{\xi_0}}(\mathbb{R}) \backslash G(\mathbb{R})} e^{-\pi\lambda^2 \operatorname{tr}(x^{-1}\gamma_{\xi_0} x)^t (x^{-1}\gamma_{\xi_0} x)} \, dx$ now equals

$$8\pi\lambda^{-2} |\Delta(\xi)|^{-\frac{1}{2}} \int_{2|\Im\tilde{\xi}|}^{\infty} e^{-\pi\lambda^2(Q_E(\xi)+\rho^2)} \, d\rho,$$

where $\tilde{\xi} \in \mathbb{C}$ denotes one of the two non-real conjugates of ξ. Changing $(Q_E(\xi) + \rho^2)^{\frac{1}{2}}\lambda$ to λ, we obtain for the double integral

$$8\pi |\Delta(\xi)|^{-\frac{1}{2}} Q_E(\xi_0)^N \int_0^\infty \lambda^{3s+2N} e^{-\pi\lambda^2} \, d^\times\lambda \int_{2|\Im\tilde{\xi}|}^{\infty} (Q_E(\xi) + \rho^2)^{-\frac{3s+2N}{2}} \, d\rho$$

from which the assertion follows. □

9.2 Dirichlet Series

To study the Dirichlet series obtained in the last section and to finish the proof of Proposition 9.2, we need to define a few more auxiliary functions. $N > 0$ denotes a sufficiently large integer such that Lemmas 9.6 and 9.7 hold. For $t \in \mathbb{C}$ with $\Re t > 5/2$ set

$$\alpha_\varepsilon^\pm(t) = \sum_{E \in \mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi \in \mathcal{O}_E/\mathbb{Z}: \\ \xi \neq 0}} c(\Phi_f, \xi) \Psi_\varepsilon^\pm(\xi) Q_E(\xi)^{-t}, \text{ and}$$

$$A_\varepsilon^\pm(X) = \sum_{E \in \mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi \in \mathcal{O}_E/\mathbb{Z}, \xi \neq 0 \\ Q_E(\xi) \leq X}} c(\Phi_f, \xi) \Psi_\varepsilon^\pm(\xi)$$

(these are both independent of N). Then by Lemmas 9.6 and 9.7 (as $I_N(\frac{2t+1}{3})$ is holomorphic and non-vanishing in all of $\Re t > 7/4$), the series defining $\alpha_\varepsilon^\pm$ converge absolutely in $\Re t > 5/2$ can be meromorphically continued up to $\Re t > 7/4$, and each has in this half plane only one pole which is located at $t = 5/2$, and is simple with residue

$$\rho_\varepsilon(\Phi_f) := \frac{3}{2} I_N(2)^{-1} \operatorname*{res}_{s=2} \mathcal{E}_\varepsilon^\pm(s).$$

The functions are related by the Mellin transformation and its inverse (cf. [MV07, § 5]): We have for $\sigma_0 \gg 0$

$$A_\varepsilon^{\pm}(X) = \frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \alpha_\varepsilon^{\pm}(t)\frac{X^t}{t}\,dt, \text{ and}$$

$$\alpha_\varepsilon^{\pm}(t) = \int_1^\infty X^{-t}\,dA_\varepsilon^{\pm}(X).$$

Further define

$$\gamma_\varepsilon(t) = \sum_{E\in\mathcal{F}_3\setminus\mathcal{F}_3^+} \rho_E \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z}:\\ \xi\neq 0}} c(\Phi_f,\xi)\Psi_\varepsilon^-(\xi)J(\xi,\frac{2t+1}{3})Q_E(\xi)^N, \quad \text{and}$$

$$\begin{aligned} C_\varepsilon(X) &= \frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma_0 + i\infty} \gamma_\varepsilon(t)\frac{X^t}{t}\,dt \\ &= \sum_{E\in\mathcal{F}_3\setminus\mathcal{F}_3^+} \rho_E \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z}:\\ \xi\neq 0}} c(\Phi_f,\xi)\Psi_\varepsilon^-(\xi)Q_E(\xi)^N \\ &\int_1^{b(\xi,X)} (Q_E(\xi)+4(\Im\tilde{\xi})^2\rho^2)^{-N-\frac{1}{2}}\,d\rho, \end{aligned}$$

where

$$b(\xi,X) = \begin{cases} \max\{1, \frac{\sqrt{X-Q_E(\xi)}}{2|\Im\tilde{\xi}|}\} & \text{if } Q_E(\xi)\le X, \\ 1 & \text{if } Q_E(\xi) > X. \end{cases}$$

This definition together with the definition of $\Psi_\varepsilon^-(\xi)$ ensures that for every X, the sum over E and ξ is in fact finite. From the last expression of $C_\varepsilon(X)$, it is clear that if N is even, $C_\varepsilon(X)$ is a non-negative, monotonically increasing function in X.

Proof of Proposition 9.2. We assume that N is even and sufficiently large such that Lemmas 9.6 and 9.7 hold. By definition of Ψ_ε^+ and Ψ_ε^- we have $\Psi_\varepsilon^+(\xi) \le 1 \le \Psi_\varepsilon^+(\xi) + \Psi_\varepsilon^-(\xi)$ for all $\xi \in E$ if E is totally real. Hence for every $X > 0$, we get

$$A_\varepsilon^+(X) \le \sum_{E\in\mathcal{F}_3^+} \frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|} \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z},\xi\neq 0\\ Q_E(\xi)\le X}} c(\Phi_f,\xi) =: \Sigma(X) \le A_\varepsilon^+(X) + A_\varepsilon^-(X). \tag{44}$$

The coefficients $\frac{\rho_E}{|\operatorname{Aut}(E/\mathbb{Q})|}c(\Phi_f,\xi)\Psi_\varepsilon^+(\xi)$ in the Dirichlet series $\alpha_\varepsilon^+(t)$ are non-negative. Hence the properties of $\alpha_\varepsilon^+(t)$ stated above allow us to apply the Wiener–Ikehara Tauberian Theorem [MV07, Corollary 8.7]. This yields the asymptotic

$$A_\varepsilon^+(X) \sim \rho_\varepsilon(\Phi_f)X^{\frac{5}{2}} + o(X^{\frac{5}{2}})$$

as $X \to \infty$. Therefore,

$$\liminf_{X\to\infty} X^{-\frac{5}{2}}\,\Sigma(X) \geq \rho_\varepsilon(\Phi_f)$$

for every $\varepsilon > 0$ so that

$$\liminf_{X\to\infty} X^{-\frac{5}{2}}\,\Sigma(X) \geq \rho_0(\Phi_f),$$

where

$$\rho_0(\Phi_f) = \frac{2\pi^{9/2}\zeta(3)}{\sqrt{3}} \int_{x\in\mathfrak{g}(\mathbb{R}):\ \Delta(x)>0} e^{-\pi\,\mathrm{tr}\,x^t x}\,dx \int_{\mathfrak{g}(\mathbb{A}_f)} \Phi_f(x_f)\,dx_f,$$

since $\rho_\varepsilon(\Phi_f) \to \rho_0(\Phi_f)$ for $\varepsilon \searrow 0$.

To show the reverse inequality, we have to work harder. Consider now the function $\mathcal{E}_\varepsilon^-(\frac{2t+1}{3})$. It has a simple pole at $t = 5/2$, and is holomorphic elsewhere in some half plane $\Re s > 7/4$. As $4\sqrt{3\pi}\frac{\Gamma(t+N+\frac{1}{2})}{\pi^{t+N+\frac{1}{2}}}$ is holomorphic and non-zero in that half plane, the function

$$\begin{aligned}\frac{\pi^{t+N+\frac{1}{2}}}{4\sqrt{3\pi}\,\Gamma(t+N+\frac{1}{2})}\mathcal{E}_\varepsilon^-(\frac{2t+1}{3}) &= \frac{1}{8\sqrt{\pi}}\frac{\Gamma(t+N)}{\Gamma(t+N+\frac{1}{2})}\alpha_\varepsilon^-(t) + \gamma_\varepsilon(t)\\ &= \frac{1}{8\pi}\beta_N(t)\alpha_\varepsilon^-(t) + \gamma_\varepsilon(t)\end{aligned}$$

has the same properties as $\mathcal{E}_\varepsilon^-$ where

$$\beta_N(t) = \int_{\mathbb{R}} (1+x^2)^{-(t+N+\frac{1}{2})}\,dx = 2\int_1^\infty y^{-(t+N+\frac{1}{2})}\,d\sqrt{y-1}.$$

The residue $\rho_\varepsilon^-(\Phi_f)$ at $t = 5/2$ is given by a constant multiple of

$$\int_{\mathfrak{g}(\mathbb{R})} \Phi_\infty^{\varepsilon,-}(x)\,dx \int_{\mathfrak{g}(\mathbb{A}_f)} \Phi_f(x_f)\,dx_f,$$

which tends to 0 as $\varepsilon \searrow 0$.

For $X > 0$ and $\sigma_0 \gg 0$ sufficiently large, let

$$B_N(X) = \frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty} \beta_N(t)\frac{X^t}{t}\,dt, \text{ and}$$

$$AB_{N,\varepsilon}(X) = \frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty} \beta_N(t)\alpha_\varepsilon^-(t)\frac{X^t}{t}\,dt.$$

In particular,

$$B_N(X) = 2\int_1^X y^{-(N-1)}\, d\sqrt{y-1}.$$

From the definitions it is clear that $C_\varepsilon(X) \geq 0$, $B_N(X) \geq 0$, and $AB_{N,\varepsilon}(X) \geq 0$, and the functions are monotonically increasing. Hence an application of the Wiener-Ikehara Theorem gives $\lim_{X\to\infty} X^{-\frac{5}{2}}(AB_{N,\varepsilon}(X) + C_\varepsilon(X)) = \rho_\varepsilon^-(\Phi_f)$, and, as everything is non-negative, $AB_{N,\varepsilon}(X) \leq \rho_\varepsilon^-(\Phi_f)X^{\frac{5}{2}} + R_\varepsilon(X)$, where $R_\varepsilon(X)$ is a suitable error function with $R_\varepsilon(X) \to 0$ as $X \to \infty$. Therefore,

$$X^{\frac{5}{2}}\rho_\varepsilon^-(\Phi_f) + R_\varepsilon(X) \geq \frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma+i\infty} \beta_N(t)\alpha_\varepsilon(t)\frac{X^t}{t}\, dt$$

and the right-hand side can be written as

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma+i\infty} \alpha_\varepsilon^-(t)\left(\int_1^\infty v^{-t} dB_l(v)\right)\frac{X^t}{t}\, dt \\
&\quad = \int_1^\infty \left(\frac{1}{2\pi i}\int_{\sigma_0 - i\infty}^{\sigma+i\infty} \alpha_\varepsilon^-(t)\left(\frac{X}{v}\right)^t \frac{dt}{t}\right) dB_N(X) = \int_1^\infty A_\varepsilon^-(\frac{X}{v})\, dB_N(v).
\end{aligned}$$

As A_ε^- is monotonically increasing, the last integral is bounded from below by

$$\geq \int_2^3 A_\varepsilon^-(\frac{X}{v})\, dB_N(v) \geq A_\varepsilon^-(\frac{X}{3})\int_2^3 dB_N(v) > 0$$

for all $X > 0$. Hence there exists a constant $c > 0$ such that for every $\varepsilon > 0$, we have $\limsup_{X\to\infty} X^{-\frac{5}{2}} A_\varepsilon^-(X) \leq c\rho_\varepsilon^-(\Phi_f)$, and thus

$$\limsup_{X\to\infty} X^{-\frac{5}{2}} A_\varepsilon^-(X) \longrightarrow 0 \qquad \text{for } \varepsilon \searrow 0.$$

Hence

$$\begin{aligned}
\limsup_{X\to\infty} X^{-\frac{5}{2}}\Sigma(X) &= \limsup_{X\to\infty} X^{-\frac{5}{2}} A_\varepsilon^+(X) + \limsup_{X\to\infty} X^{-\frac{5}{2}} A_\varepsilon^-(X) \\
&\leq \rho_\varepsilon^+(\Phi_f) + c\rho_\varepsilon^-(\Phi_f) \longrightarrow \rho_0(\Phi_f)
\end{aligned}$$

for $\varepsilon \searrow 0$, which finishes the proof of the asymptotic. □

10 Bounds for Mean Values of Residues of Dedekind Zeta Functions

We want to use the result from the last section to obtain information on the mean value of residues of Dedekind zeta functions. As $c(\Phi_f^0, \xi) \geq 1$ for all $\xi \in E\backslash\mathbb{Q}$ and all $E \in \mathcal{F}_3$ by Proposition 8.1, an immediate consequence of Proposition 9.2 is the following upper bound.

Theorem 10.1. *There exists $\alpha < \infty$ such that*

$$\limsup_{X\to\infty} X^{-\frac{5}{2}} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\leq X}} \operatorname*{res}_{s=1} \zeta_E(s) \leq \alpha. \tag{45}$$

Remark 10.2. Note that one can take α to be equal to $\beta(\Phi_f^0)$ (from Proposition 9.2), which is explicitly computable. To obtain a better upper bound (which should be optimal, in fact), one can try to use the sequence of test functions from Appendix 2, that is, the optimal α should equal the limit over $\mathfrak{m}$ of $\beta(\Phi_f^{\mathfrak{m}})$.

To complement this upper bound we show the following lower bound.

Proposition 10.3. *We have for every $\varepsilon > 0$, we have*

$$\liminf_{X\to\infty} X^{-\frac{5}{2}+\varepsilon} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\leq X}} \operatorname*{res}_{s=1} \zeta_E(s) = \infty.$$

In fact, Conjecture 1.7 is expected to be true. The proof of this proposition is of a complete different nature than the proof of Theorem 10.1: Basically we will show that there are sufficiently many irreducible cubic polynomials, cf. also the introduction where a relation to [EV06, Remark 3.3] is explained. Ultimately, one hopes that Proposition 10.3 (and even Conjecture 1.7) can also be deduced from Proposition 9.2, cf. Appendix 2 for a sequence of test functions that might be useful. We need the following auxiliary result to prove Proposition 10.3:

Lemma 10.4.

(i) Let $Q : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be a positive definite quadratic form with discriminant $\Delta(Q)$ and first successive minimum $m_1(Q) \geq 1$. Then, as $X \to \infty$, we have

$$\sum_{\substack{\gamma\in\mathbb{Z}^2:\\ Q(\gamma)\leq X}} 1 = \frac{2\pi X}{\sqrt{\Delta(Q)}} + O\left(\sqrt{\frac{m_1(Q)}{\Delta(Q)}} X^{\frac{1}{2}}\right)$$

with implied constant independent of Q.

(ii) For all $\varepsilon > 0$, we have as $X \to \infty$

$$\sum_{\substack{E\in\mathcal{F}_3^+:\\ m_2(E)\leq X}} \rho_E \sum_{\substack{\xi\in\mathcal{O}_E/\mathbb{Z},\xi\neq 0,\\ Q_E(\xi)\leq X}} 1 = O(X^{2+\varepsilon}).$$

Proof.

(i) We need to count all points in $\mathbb{Z}^2$ which are contained in the ellipse $E_X := \{x \in \mathbb{R}^2 \mid Q(x) \le X\}$. By a theorem of Gauss [Coh80, p.161], the number of such points is equal to the area $\frac{2\pi X}{\sqrt{\Delta(Q)}}$ of the ellipse E_X plus some small error term of order $RX^{\frac{1}{2}}$ for R the length of the major axis of the ellipse E_1 and all implicit constants independent of Q. Since $m_1(Q) \ge 1$, it is easily verified that $R \le \sqrt{\frac{m_1(Q)}{\Delta(Q)}}$ finishing the proof of the assertion.

(ii) By Minkowski's second theorem (see, e.g. [Cas97, VIII.4.3]), there are $a_1, a_2 > 0$ such that for all cubic fields E, $a_1 m_1(E) m_2(E) \le D_E \le a_2 m_1(E) m_2(E)$ so that $m_1(E) \le m_2(E) \le X$ implies $c_0 D_E \le m_1(E) m_2(E) \le 16X^2$ for some $c_0 > 0$, and moreover, $m_1(E)/\Delta(Q_E)$ is bounded from above by an absolute constant. Hence there is by (i) some constant $C > 0$ such that

$$\sum_{\substack{\xi \in \mathcal{O}_E/\mathbb{Z},\ \xi \ne 0,\\ Q_E(\xi) \le X}} 1 \le C \frac{X}{\sqrt{\Delta(Q_E)}}$$

for all E with $m_1(E) \le m_2(E) \le X$. By the Brauer–Siegel Theorem [Lan94, XVI, § 4 Theorem 4], there exists for all $\varepsilon > 0$ some number $C_\varepsilon > 0$ such that $\rho_E = \mathrm{res}_{s=1} \zeta_E(s) = 4D_E^{-\frac{1}{2}} h_E R_E \le C_\varepsilon D_E^\varepsilon$ for all totally real cubic fields E. Hence the left-hand side of (ii) equals

$$\sum_{\substack{E \in \mathcal{F}_3^+:\\ m_2(E) \le X}} \rho_E \sum_{\substack{\xi \in \mathcal{O}_E \setminus \mathbb{Z}:\\ Q_E(\xi) \le X}} 1 \le C C_\varepsilon \sqrt{3} \sum_{E:\ m_2(E) \le X} X D_E^{\varepsilon - \frac{1}{2}}.$$

This can be bounded by

$$C C_\varepsilon \sqrt{3} X \sum_{E:\ D_E \le 16X^2} D_E^{\varepsilon - \frac{1}{2}} \le C C_\varepsilon \sqrt{3} X^{1+\varepsilon} \sum_{E:\ D_E \le 16X^2} D_E^{-\frac{1}{2}}.$$

By [DH71, Theorem 1] or [DW88, Theorem I.1], $\sum_{E:\ D_E \le X} 1 = c_0 X + o(X)$ for some $c_0 > 0$ so that

$$C C_\varepsilon \sqrt{3} X^{1+\varepsilon} \sum_{E:\ D_E \le 16X^2} D_E^{-\frac{1}{2}} \le 16 c_0 C C_\varepsilon \sqrt{3} X^{2+\varepsilon} + o(X^{2+\varepsilon})$$

which is the assertion.

□

Proof of Proposition 10.3. It suffices to assume that $\varepsilon \in (0, 1/2)$. We first show that

$$\liminf_{X \to \infty} X^{-\frac{5}{2}+\varepsilon} \sum_{E \in \mathcal{F}_3^+} \sum_{\substack{\xi \in \mathcal{O}_E/\mathbb{Z}, \xi \ne 0,\\ Q_E(\xi) \le X}} \rho_E = \infty \tag{46}$$

for every $\varepsilon > 0$. Let $\varepsilon > 0$. By the Brauer–Siegel Theorem there exists $A_\varepsilon > 0$ such that $\rho_E \geq A_\varepsilon D_E^{-\frac{\varepsilon}{2}}$ for all E. Thus this sum is bounded from below by $A_\varepsilon X^{-\frac{\varepsilon}{2}} \sum_{E\in\mathcal{F}_3^+} N_E(X)$, where $N_E(X) := \left|\{\xi \in \mathcal{O}_E/\mathbb{Z} : \xi \neq 0,\ Q_E(\xi) \leq X\}\right|$. Hence it will certainly suffice to show that there exists $C > 0$ such that

$$\sum_{E\in\mathcal{F}_3^+} N_E(X) \sim CX^{\frac{5}{2}}$$

as $X \to \infty$. The map associating with the pair $E \in \mathcal{F}_3^+$, $\xi \in \mathcal{O}_E/\mathbb{Z}$, $\xi \neq 0$, the characteristic polynomial $T^3 + a_1T + a_0$ of $\xi - \frac{1}{3}\operatorname{tr}\xi\mathbf{1}_3$ is $3-1$ or $1-1$ depending on whether E is Galois or not. As E is totally real, we have $\Delta(\xi - \frac{1}{3}\operatorname{tr}\xi\mathbf{1}_3) = -4a_1^3 - 27a_0^2 > 0$, or equivalently $a_0^2 \leq -\frac{4}{27}a_1^3$. Since $X \geq Q_E(\xi) = -2a_1 > 0$, this implies

$$-\frac{X}{2} \leq a_1 < 0 \quad \text{and} \quad 0 < a_0 \leq \sqrt{-\frac{4}{27}a_1^3} \leq \frac{1}{3\sqrt{6}}X^{\frac{3}{2}}. \tag{47}$$

Hence, ignoring constants, there are $a_1^{\frac{3}{2}}$ many a_0 and

$$\int_1^{X/2} a_1^{\frac{3}{2}}\, da_1 = \frac{1}{10\sqrt{2}}X^{\frac{5}{2}} - \frac{2}{5}$$

many a_1 satisfying all the conditions. On the other hand, any irreducible polynomial with integral coefficients satisfying the inequalities in (47) defines (a conjugacy class of) a cubic field E and ξ as before. Thus we only need to show that the reducible polynomials with coefficients satisfying above constraints do not contribute to $CX^{\frac{5}{2}}$. If $T^3 + a_1T + a_0$ is reducible over $\mathbb{Q}$, we can write it as a product $(T^2 + b_1T + b_0)(T + c)$ with $b_1, b_0, c \in \mathbb{Z}$. Hence $c = -b_1$, $cb_0 = a_0$ and $b_0 - c^2 = a_1$. Hence if we fix a_0 (for which there are at most $O(X^{\frac{3}{2}})$ possibilities), there are at most $O(a_0^\delta) \leq O(X^\delta)$ possibilities for c and b_0 for any $\delta > 0$. Thus there are only $O(X^{\frac{3}{2}+\delta})$ reducible polynomials satisfying above constraints. This finishes the proof of (46).

Now split the sum over E in the following parts: One belonging to $E \in \mathcal{F}_3^+$ such that $m_1(E) > X$, one over E such that $m_1(E) \leq X < m_2(E)$, and the last one over E such that $m_1(E) < m_2(E) \leq X$. For E with $m_1(E) > X$, there are no ξ contributing to the sum in (46) so that the sum on the left-hand side of (46) equals

$$X^{-\frac{5}{2}+\varepsilon} \sum_{\substack{E\in\mathcal{F}_3^+: \\ m_1(E)\leq X<m_2(E)}} \rho_E N_E(X) + X^{-\frac{5}{2}+\varepsilon} \sum_{\substack{E\in\mathcal{F}_3^+: \\ m_1(E)\leq m_2(E)\leq X}} \rho_E N_E(X). \tag{48}$$

By Lemma 10.4(ii), the second sum tends to 0 for $X \to \infty$ provided $\varepsilon < \frac{1}{2}$. Hence the limes inferior of the first part of the sum is not bounded from below as $X \to \infty$ for any $\varepsilon \in (0, 1/2)$. As $m_1(E) \le X < m_2(E)$, every $\xi \in \mathcal{O}_E/\mathbb{Z}$, $\xi \ne 0$, with $Q_E(x) \le X$ is of the form $\xi = n\xi_0$ for some $n \in \mathbb{N}$, and ξ_0 one of the two non-zero primitive vectors in $\mathcal{O}_E/\mathbb{Z}$. Note that $Q_E(\pm\xi_0) = m_1(E)$. Thus

$$\sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\le X<m_2(E)}} \rho_E N_E(X) = \sum_{n\in\mathbb{N}} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\le X<m_2(E)}} \rho_E \sum_{\substack{\xi_0\in(\mathcal{O}_E/\mathbb{Z})_{\mathrm{prim}},\ xi_0\ne0\\ Q_E(\xi_0)\le\frac{X}{n^2}}} 1$$
$$= 2\sum_{n\in\mathbb{N}} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\le\frac{X}{n^2}<m_2(E)}} \rho_E,$$

where $(\mathcal{O}_E/\mathbb{Z})_{\mathrm{prim}}$ denotes the set of primitive vectors in $\mathcal{O}_E/\mathbb{Z}$. Suppose there are $\kappa \in (0, 1/2)$ and $0 < c_0 < \infty$ such that

$$\liminf_{X\to\infty} X^{-\frac{5}{2}+\kappa} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\le X<m_2(E)}} \rho_E = c_0 < \infty.$$

Then

$$X^{-\frac{5}{2}+\kappa} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\le X<m_2(E)}} \rho_E N_E(X) = 2\sum_{n\in\mathbb{N}} n^{-5+2\kappa} (\frac{X}{n^2})^{-\frac{5}{2}+\kappa} \sum_{\substack{E\in\mathcal{F}_3^+:\\ m_1(E)\le\frac{X}{n^2}<m_2(E)}} \rho_E$$

and, for every n, $\liminf_{X\to\infty} (\frac{X}{n^2})^{-\frac{5}{2}+\kappa} \sum_{E\in\mathcal{F}_3^+,\ m_1(E)\le\frac{X}{n^2}<m_2(E)} \rho_E = c_0$ so that the limit inferior of the above is $2c_0\zeta(5-2\kappa)$ in contradiction to the unboundedness of the limit inferior of the first sum in (48) as $X \to \infty$. This finishes the proof of the proposition. $\square$

Acknowledgements The work on Part 1was started while the author stayed at the MPI in Bonn during August/September 2011, and continued during the year 2011/2012 when the author was supported by grant #964-107.6/2007 from the German-Israeli Foundation for Scientific Research and Development and by a Golda Meir Postdoctoral Fellowship. Part 2 was essentially contained in the author's doctoral dissertation [Mat11] which was written under the direction of T. Finis and supported by grant #964-107.6/2007 from the German-Israeli Foundation for Scientific Research and Development. The author would like to thank T. Finis and E. Lapid for communicating their idea for the proof of Proposition 5.1 carried out in Appendix 1.

Appendix 1: Asymptotic Approximation of Truncation Functions

The purpose of this appendix is to prove Proposition 5.1 in the case of a nilpotent orbit $\mathcal{N} \subseteq \mathfrak{n}$ for $G = \mathrm{GL}_n$ and $n \leq 3$.

The Case $n = 2$

There are two nilpotent orbits in $\mathfrak{g}$, namely $\mathcal{N}_{\mathrm{triv}} = 0$ and $\mathcal{N}_{\mathrm{reg}}$. For $\mathcal{N}_{\mathrm{triv}}$ there is nothing to show so that we only consider $\mathcal{N} = \mathcal{N}_{\mathrm{reg}}$. We denote by X_0 an element as in Example 4.2. The associated Jacobson–Morozov parabolic subgroup for X_0 is $P = P_0 = T_0 U_0$, and $C_{U_0}(X_0) = U_0$.

Lemma A.1. *Let ν be as in Lemma 3.7 and let $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$. Then there exists a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ such that for every $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ and nilpotent orbit $\mathcal{N} \subseteq \mathfrak{g}(\mathbb{Q})$, we have*

$$\left| j_{\mathcal{N}}^T(\Phi) - \tilde{j}_{\mathcal{N}}^T(\Phi) \right| \leq \mu(\Phi_\infty) e^{-c_2 \|T\|}$$

for all $T \in \mathfrak{a}^+$ with $d(T) > \|T\|/2$, where $\Phi = \Phi_\infty \cdot \Phi_f$.

Proof. We only consider $\mathcal{N} = \mathcal{N}_{\mathrm{reg}}$. Let $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$. We may assume that Φ is $\mathbf{K}$-conjugation invariant. Then

$$j_{\mathcal{N}_{\mathrm{reg}}}^T(\Phi) = \int_{A_0^G} \delta_0(a)^{-1} \left(\int_{U_0(\mathbb{Q})\backslash U_0(\mathbb{A})} F(ua, T)\, du \right) \sum_{X \in \mathfrak{u}_0(\mathbb{Q}) \cap \mathcal{N}_{\mathrm{reg}}} \Phi(\mathrm{Ad}\, a^{-1} X)\, da.$$

Note that $\tilde{F}^{T_0}(a, T) = \hat{\tau}_0^G(T - H_0(a)) = 0$ implies $F(ua, T) = 0$ for all $u \in U_0(\mathbb{A})$, i.e., $F(ua, T) \leq \tilde{F}^{T_0}(a, T)$ for all u and a. The sum inside the integral is over all $X = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) =: X(x)$ with $x \in \mathbb{Q}$, $x \neq 0$. Let $\varphi(x) = \Phi(X(x))$, $x \in \mathbb{A}$, and write $a = \mathrm{diag}(b, b^{-1})$ with $b \in (0, \infty)$. Then

$$\sum_{X \in \mathfrak{u}_0(\mathbb{Q}) \cap \mathcal{N}_{\mathrm{reg}}} \Phi(\mathrm{Ad}\, a^{-1} X) = \sum_{x \in \mathbb{Q}^\times} \varphi(b^{-2} x)$$

and there exists a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (depending on Φ_f) such that for all $b \in (0, 1)$ we have

$$\left| \sum_{x \in \mathbb{Q}^\times} \varphi(b^{-2} x) \right| \leq \mu(\Phi_\infty) b^3.$$

In particular, for every $b_0 \in (0,1)$ we have

$$0 \leq \int_0^{b_0} b^{-2}\left(\tilde{F}^{T_0}(a,T) - \int_{U_0(\mathbb{Q})\backslash U_0(\mathbb{A})} F(ua,T)\,du\right)\left|\sum_{x\in\mathbb{Q}^\times}\varphi(b^{-2}x)\right| d^\times b$$
$$\leq \mu(\Phi_\infty)\int_0^{b_0} b\,d^\times b = \mu(\Phi_\infty)b_0.$$

Now let $b > b_0$. We want to find an upper bound for the difference

$$\tilde{F}^{T_0}(a,T) - \int_{U_0(\mathbb{Q})\backslash U_0(\mathbb{A})} F(va,T)\,dv. \tag{49}$$

Let α be the unique positive root of (T_0, U_0) and ϖ the corresponding coroot. We may assume that b is such that $\tilde{F}^{T_0}(a,T) = 1$, that is, $b < e^{\varpi(T)}$, as otherwise (49) vanishes. Note that (49) is always non-negative. It equals

$$\begin{aligned}
&\mathrm{vol}\{v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A}) \mid \exists\gamma \in G(\mathbb{Q}):\ \varpi(H_0(\gamma va) - T) > 0\}\\
\leq\ &\mathrm{vol}\{v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A}) \mid \exists u \in U_0(\mathbb{Q}):\ \varpi(H_0(uva) - T) > 0\}\\
&+ \mathrm{vol}\{v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A}) \mid \exists u \in U_0(\mathbb{Q}):\ \varpi(H_0(wuva) - T) > 0\}
\end{aligned}$$

for $w = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0\end{smallmatrix}\right)$ a representative for the non-trivial Weyl group element. Here we used the left $P_0(\mathbb{Q})$-invariance of H_0. Using again the left $U_0(\mathbb{Q})$-invariance and that $\tilde{F}^{T_0}(a,T) = 1$, the volume of the first set is 0 so that we only need to estimate

$$\mathrm{vol}\{v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A}) \mid \exists u \in U_0(\mathbb{Q}):\ \varpi(H_0(wuva) - T) > 0\}. \tag{50}$$

For that write $u = \left(\begin{smallmatrix} 1 & x \\ 0 & 1\end{smallmatrix}\right) \in U_0(\mathbb{Q})$ and $v = \left(\begin{smallmatrix} 1 & y \\ 0 & 1\end{smallmatrix}\right) \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A})$. Then

$$\varpi(H_0(wuva)) = \varpi(wH_0(a)) + \varpi(H_0(wa^{-1}uva)) = -\varpi(H_0(a)) - \log\|(1, b^{-2}(x+y))\|_{\mathbb{A}}$$

for $\|\cdot\|_{\mathbb{A}}$ the adelic vector norm. But $\|(1, b^{-2}(x+y))\|_{\mathbb{A}} \geq 1$ so that

$$\varpi(H_0(wuva) - T) \leq -\varpi(H_0(a)) - \varpi(T) \leq -\log b_0 - \varpi(T)$$

which is ≤ 0 if $d(T) = \varpi(T) \geq -\log b_0$. In particular, the volume (50) vanishes for every $b \geq b_0$ if $d(T) \geq \log b_0$. Choosing $b_0 = e^{-\|T\|/2}$, we therefore get

$$\left|j_{\mathcal{N}}^T(\Phi) - \tilde{j}_{\mathcal{N}}^T(\Phi)\right| \leq \int_0^{e^{-\|T\|/2}} b^{-2}\left(\tilde{F}^{T_0}(a,T) - \int_{U_0(\mathbb{Q})\backslash U_0(\mathbb{A})} F(ua,T)\,du\right)\left|\sum_{x\in\mathbb{Q}^\times}\varphi(b^{-2}x)\right| d^\times b$$
$$\leq \mu(\Phi_\infty)e^{-\|T\|/2}$$

which proves the assertion of the lemma. $\square$

The Case n = 3

There are now three different nilpotent orbits in $\mathfrak{g}$: The trivial orbit $\mathcal{N}_{\text{triv}} = 0$, the minimal orbit $\mathcal{N}_{\text{min}}$, and the regular orbit $\mathcal{N}_{\text{reg}}$. We recall our choices of X_0, and the Jacobson–Morozov parabolic from Example 4.2. The first case again is trivial so that we only consider the other two. In both of these cases the associated Jacobson–Morozov parabolic is the minimal parabolic.

Lemma A.2. *There are $c_1, c_2 > 0$ such that for every $X_0 \in \mathfrak{u}_{\mathcal{N}}^{2,reg}(\mathbb{Q})$, $\mathcal{N} \in \{\mathcal{N}_{min}, \mathcal{N}_{reg}\}$, and $v' \in C_{U_0}(X_0, \mathbb{A})\backslash U_0(\mathbb{A})$ we have*

$$\left|\tilde{F}^{T_0}(t,T) - \int_{C_{U_0}(X_0,\mathbb{Q})\backslash C_{U_0}(X_0,\mathbb{A})} F(vv't,T)\,dv\right| \le c_1 e^{-c_2\|T\|} \tag{51}$$

for all $t \in T_0(\mathbb{A})$ and all sufficiently regular $T \in \mathfrak{a}^+$ with $d(T) > \delta\|T\|$.

Proof. Write $\Delta_0 = \{\alpha_1, \alpha_2\}$ such that $\alpha(\operatorname{diag}(t_1,t_2,t_3)) = |t_1/t_2|$ and $\alpha_2(\operatorname{diag}(t_1,t_2,t_3)) = |t_2/t_3|$. We consider the two nilpotent orbits separately.

$\mathcal{N} = \mathcal{N}_{\mathbf{min}}$

We have $C_{U_0}(X_0) = U_0$ so that $v' = 1$. Let $t \in T_0(\mathbb{A})$. It is clear that $\tilde{F}^{T_0}(t,T) = 0$ again implies that $F(vt,T) = 0$ for all $v \in U_0(\mathbb{A})\backslash U_0(\mathbb{A})$. Hence we again assume that t is such that $\tilde{F}^{T_0}(t,T) = 1$. To estimate the left-hand side of (51) it will therefore suffice to bound the volume of the set

$$\{v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A}) \mid \exists\gamma \in G(\mathbb{Q})\ \exists\varpi \in \widehat{\Delta}_0 :\ \varpi(H_0(\gamma vt) - T) > 0\}.$$

Using Bruhat decomposition for $G(\mathbb{Q})$ and the left $P_0(\mathbb{Q})$-invariance of H_0, it suffices to bound for each $w \in W$ and $\varpi \in \widehat{\Delta}_0$ the volume of the set

$$V_T(w,\varpi,T) = \{v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A}) \mid \exists u \in U_0(\mathbb{Q}) :\ \varpi(H_0(wuvt) - T) > 0\}.$$

Now for $v \in U_0(\mathbb{Q})\backslash U_0(\mathbb{A})$ and $u \in U_0(\mathbb{Q})$ we have $H_0(wuvt) = H_0((wtw^{-1})(wt^{-1}uvt)) = wH_0(t) + H_0(wt^{-1}uvt)$ so that

$$\varpi(H_0(wuvt) - T) > 0 \Leftrightarrow \varpi(H_0(wt^{-1}uvt)) > \varpi(T - wH_0(t)) \Leftrightarrow e^{-\varpi(H_0(wt^{-1}uvt))} < e^{\varpi(wH_0(t) - T)}.$$

Hence vol $V_T(w,\varpi,t)$ equals

$$\operatorname{vol}\left(\{x_1,x_2,x_3 \in [0,1] \mid \exists y_1,y_2,y_3 \in \mathbb{Q} :\ -\varpi(H_0(wt^{-1}\begin{pmatrix} 1 & x_1+y_1 & x_2+y_2 \\ & 1 & x_3+y_3 \\ & & 1 \end{pmatrix} t)) < -\varpi(T - wH_0(t))\}\right)$$

Suppose $u = \begin{pmatrix} 1 & u_1 & u_2 \\ & 1 & u_3 \\ & & 1 \end{pmatrix} \in U_0(\mathbb{A})$. We first want to compute the last two rows of wuw^{-1}, as they can be used to compute $\varpi(H_0(wuw^{-1}))$.

- $w = w_1 = \mathrm{id}$, then the last two columns equal
$$\begin{pmatrix} 0 & 1 & u_3 \\ 0 & 0 & 1 \end{pmatrix}.$$
- $w = w_2$ is the simple reflexion about the root α_1. Then the last two rows equal
$$\begin{pmatrix} u_1 & 1 & u_2 \\ 0 & 0 & 1 \end{pmatrix}$$
- $w = w_3$ is the simple reflexion about the root α_2. Then the last two rows equal
$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & u_3 & 1 \end{pmatrix}$$
- $w = w_4$ is the longest Weyl element. Then the last two rows equal
$$\begin{pmatrix} u_3 & 1 & 0 \\ u_2 & u_1 & 1 \end{pmatrix}$$
- $w = w_5$ is represented by $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Then the last two rows equal
$$\begin{pmatrix} u_2 & 1 & u_1 \\ u_3 & 0 & 1 \end{pmatrix}$$
- $w = w_6 = w_5^{-1}$. Then the last two rows equal
$$\begin{pmatrix} 0 & 1 & 0 \\ u_1 & u_2 & 1 \end{pmatrix}.$$

The case $\varpi = \varpi_2$: Using the above computations, we have for $u := \begin{pmatrix} 1 & x_1+y_1 & x_2+y_2 \\ & 1 & x_3+y_3 \\ & & 1 \end{pmatrix}$

$$e^{-\varpi_2(H_0(w_i t^{-1} ut))} = \begin{cases} \|(0,0,1)\|_{\mathbb{A}} = 1 & \text{if } i \in \{1,2\}, \\ \|(0, e^{-\alpha_2(H_0(t))}(x_3+y_3), 1)\|_{\mathbb{A}} & \text{if } i \in \{3,5\}, \\ \|(e^{-(\alpha_1+\alpha_2)(H_0(t))}(x_2+y_2), e^{-\alpha_1(H_0(t))}(x_1+y_1), 1)\|_{\mathbb{A}} & \text{if } i \in \{4,6\}. \end{cases}$$

Since $\tilde{F}^{T_0}(t,T) \neq 0$, we have $\varpi_2(T - wH_0(t)) = \varpi_2(T - H_0(t)) \leq 0$ so that $\operatorname{vol} V_T(w_1, \varpi_2, t) = \operatorname{vol} V_T(w_2, \varpi_2, t) = 0$.

Now if $w \in \{w_3, w_5\}$ we have $\varpi_2(wH_0(t)) = (\varpi_1 - \varpi_2)(H_0(t))$, and therefore

$$e^{-\varpi_2(H_0(wt^{-1}ut))} < e^{\varpi_2(wH_0(t)-T)} \Leftrightarrow \|(0, e^{-\alpha_2(H_0(t))}(x_3+y_3), 1)\|_{\mathbb{A}} < e^{(\varpi_1-\varpi_2)(H_0(t))-\varpi_2(T)}.$$

Writing out the adelic norm on the left-hand side, this is equivalent to (recall that $x_3 \in [0,1]$)

$$(1 + e^{-2\alpha_2(H_0(t))}(x_3+y_3)^2)^{1/2} \prod_{p<\infty} \max\{1, |y_3|_p\} < e^{(\varpi_1-\varpi_2)(H_0(t))-\varpi_2(T)}.$$

We can write $y_3 = a/b$ with a, b coprime integers. Then $\prod_{p<\infty} \max\{1, |y_3|_p\} = |b|$ so that the above is equivalent to

$$\begin{aligned} 1 + e^{-2\alpha_2(H_0(t))}(x_3+y_3)^2 &< b^{-2} e^{2(\varpi_1-\varpi_2)(H_0(t))-2\varpi_2(T)} \\ \Leftrightarrow \quad (x_3 + \frac{a}{b})^2 &< \left[b^{-2} e^{2(\varpi_1-\varpi_2)(H_0(t))-2\varpi_2(T)} - 1\right] e^{2\alpha_2(H_0(t))}. \end{aligned}$$

If there exists x_3 satisfying this inequality, we must necessarily have $e^{(\varpi_1-\varpi_2)(H_0(t))-\varpi_2(T)} > 1$ and $|b| < e^{(\varpi_1-\varpi_2)(H_0(t))-\varpi_2(T)}$. It moreover suffices to consider $0 \leq a \leq b$, since if for $a > b$ there still exists x_3 as before, then the volume of $V_T(w, \varpi_2, t)$ equals 1. Hence the volume of all $x_3 \in [0,1]$ for which there exists $y_3 \in \mathbb{Q}$ as above is bounded by

$$\sum_{0<b<e^{(\varpi_1-\varpi_2)(H_0(t))-\varpi_2(T)}} \sum_{0\leq a<b} b^{-1} e^{(\varpi_1-\varpi_2)(H_0(t))-\varpi_2(T)} e^{\alpha_2(H_0(t))} \leq e^{2(\varpi_1-\varpi_2)(H_0(t))-2\varpi_2(T)} e^{\alpha_2(H_0(t))}.$$

Note that $2(\varpi_1-\varpi_2)+\alpha_2 = \varpi_1$ so that, since $\varpi_1(H_0(t)) \leq \varpi_1(T)$ by assumption, we get

$$\operatorname{vol} V_T(w, \varpi_2, t) \leq e^{-\alpha_2(T)}$$

for $w \in \{w_3, w_5\}$.

Now if $w \in \{w_4, w_6\}$, we have $\varpi_2(wH_0(t)) = -\varpi_1(H_0(t))$. Therefore,

$$\begin{aligned} & e^{-\varpi_2(H_0(wt^{-1}ut))} < e^{\varpi_2(wH_0(t)-T)} \\ \Leftrightarrow \quad & \|(e^{-(\alpha_1+\alpha_2)(H_0(t))}(x_2+y_2), e^{-\alpha_1(H_0(t))}(x_1+y_1), 1)\|_{\mathbb{A}} < e^{-\varpi_1(H_0(t))-\varpi_2(T)}. \end{aligned}$$

This is equivalent to

$$(1 + e^{-2\alpha_1(H_0(t))}(x_1+y_1)^2 + e^{-2(\alpha_1+\alpha_2)(H_0(t))}(x_2+y_2)^2)^{1/2} \prod_{p<\infty} \max\{1, |y_1|_p, |y_2|_p\} < e^{-\varpi_1(H_0(t))-\varpi_2(T)}.$$

Write $y_i = a_i/b_i$ with a_i, b_i coprime integers. Then $\prod_{p<\infty} \max\{1, |y_1|_p, |y_2|_p\} = \mathrm{lcm}(b_1, b_2) =: b$, and as above it suffices to consider $0 \le a_1, a_2 < b < e^{-\varpi_1(H_0(t))-\varpi_2(T)}$. Hence the volume of $V_T(w, \varpi_2, t)$ is bounded by the sum over all such a_1, a_2, b of the volume of all $x_1, x_2 \in [0, 1]$ satisfying

$$e^{-2\alpha_1(H_0(t))}(x_1 + \frac{a_1}{b})^2 + e^{-2(\alpha_1+\alpha_2)(H_0(t))}(x_2 + \frac{a_2}{b})^2 < b^{-2}e^{-2\varpi_1(H_0(t))-2\varpi_2(T)} - 1$$

so that for $w \in \{w_4, w_6\}$ we have

$$\begin{aligned} \mathrm{vol}\, V_T(w, \varpi_2, t) &\le \sum_{0<b<e^{-\varpi_1(H_0(t))-\varpi_2(T)}} b e^{\alpha_1(H_0(t))} e^{(\alpha_1+\alpha_2)(H_0(t))} e^{-\varpi_1(H_0(t))-\varpi_2(T)} \\ &\le e^{\alpha_1(H_0(t))} e^{(\alpha_1+\alpha_2)(H_0(t))} e^{-3\varpi_1(H_0(t))-3\varpi_2(T)} = e^{-3\varpi_2(T)}. \end{aligned}$$

The case $\varpi = \varpi_1$: Using the same notation as before, we can compute

$$e^{-\varpi_1(H_0(w_i t^{-1} ut))} = $$
$$\begin{cases} \|(0,0,1)\|_{\mathbb{A}} = 1 & \text{if } i \in \{1,3\}, \\ \|(0, 1, e^{-\alpha_1(H_0(t))}(x_1 + y_1))\|_{\mathbb{A}} & \text{if } i \in \{2,6\}, \\ \|(1, e^{-\alpha_2(H_0(t))}(x_3 + y_3), e^{-(\alpha_1+\alpha_2)(H_0(t))}\big((x_1 + y_1)(x_3 + y_3) - (x_2 + y_2)\big))\|_{\mathbb{A}} & \text{if } i \in \{4,5\}. \end{cases}$$

If $w \in \{w_1, w_3\}$ it follows as before that $V_T(w, \varpi_1, t) = 0$. If $w \in \{w_2, w_6\}$, then $\varpi_1(wH_0(t)) = (\varpi_2 - \varpi_1)(H_0(t))$, and it follows as before that $\mathrm{vol}\, V_T(w, \varpi_1, t)$ is bounded from above by

$$e^{2(\varpi_2-\varpi_1)(H_0(t))-2\varpi_1(T)} e^{\alpha_1(t)} \le e^{-\alpha_1(T)}$$

by our assumption on t.

For the last case $w \in \{w_4, w_5\}$ we have $\varpi_1(wH_0(t)) = -\varpi_2(H_0(t))$ so that

$$e^{-\varpi_1(H_0(wt^{-1}ut))} < e^{\varpi_1(wH_0(t)-T)}$$

is equivalent to

$$\|(1, e^{-\alpha_2(H_0(t))}(x_3 + y_3), e^{-(\alpha_1+\alpha_2)(H_0(t))}\big((x_1 + y_1)(x_3 + y_3) - (x_2 + y_2)\big))\|_{\mathbb{A}} < e^{-\varpi_2(H_0(t))-\varpi_1(T)}.$$

It follows similarly as before (we may replace $(x_1 + y_1)(x_3 + y_3) - (x_2 + y_2)$ by $x_2 + y_2$ for our purposes) that the volume $\mathrm{vol}\, V_T(w, \varpi_1, t)$ is bounded by

$$e^{\alpha_2(H_0(t))} e^{(\alpha_1+\alpha_2)(H_0(t))} e^{-3\varpi_2(H_0(t))-3\varpi_1(T)} = e^{-3\varpi_1(T)}$$

finishing the case $\varpi = \varpi_1$.

Taking all computations for $\varpi = \varpi_1, \varpi_2$ together, we obtain

$$\left|\tilde{F}^{T_0}(t,T) - \int_{C_U(X_0,\mathbb{Q})\backslash C_U(X_0,\mathbb{A})} F(vt,T)\,dv\right| \le 2\big(e^{-\alpha_1(T)} + e^{-\alpha_2(T)} + e^{-3\varpi_1(T)} + e^{-3\varpi_2(T)}\big) \le 8e^{-d(T)}$$

for all $t \in T_0(\mathbb{Q})\backslash T_0(\mathbb{A})$. For $d(T) > \delta\|T\|$ the assertion follows.

$\mathcal{N} = \mathcal{N}_{\mathbf{reg}}$

Let $t \in T_0(\mathbb{A})$ be again such that $\tilde{F}^{T_0}(t,T) = 1$. For $X_0 = \begin{pmatrix} 0 & x_0 & 0 \\ 0 & 0 & y_0 \\ 0 & 0 & 0 \end{pmatrix} \in \mathfrak{u}^{2,\mathrm{reg}}_{\mathcal{N}_{\mathrm{reg}}}(\mathbb{Q})$, the Jacobson–Morozov parabolic subgroup is again $P = P_0$, and

$$C_{U_0}(X_0) = \{\begin{pmatrix} 1 & a & b \\ 0 & 1 & \frac{y_0}{x_0}a \\ 0 & 0 & 1 \end{pmatrix}\}.$$

For notational reasons we only consider the case $x_0 = y_0$, the remaining cases being similar. As a complement of $C_{U_0}(X_0) \subseteq U_0$ we choose the subspace $V := \{\begin{pmatrix} 1 & c & 0 \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{pmatrix}\} \subseteq U_0$. Let $v' = v'(c) = \begin{pmatrix} 1 & c & 0 \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{pmatrix} \in V(\mathbb{A})$ be fixed. We want to approximate the sets

$$V_T(\varpi, t, v') = \{v \in C_{U_0}(X_0,\mathbb{Q})\backslash C_{U_0}(X_0,\mathbb{A}) \mid \exists \gamma \in G(\mathbb{Q}) :\ \varpi(H_0(\gamma v v' t) - T) > 0\}$$

for each $\varpi \in \{\varpi_1, \varpi_2\}$. We split this set into disjoint sets $V_T(w, \varpi, t, v')$ for $w \in W$ according to the Bruhat decomposition as before.

The case $\varpi = \varpi_2$: If applicable, we use the same notation as in the case of the minimal orbit, but now write $x_1 = a + c$, $x_3 = a - c$, and $x_2 = b - ac$ with c fixed and $a, b \in \mathbb{Q}\backslash\mathbb{A}$. Hence

$$e^{-\varpi_2(H_0(wt^{-1}yvv't))} = \begin{cases} \|(0,0,1)\|_{\mathbb{A}} = 1 & \text{if } w \in \{w_1, w_2\}, \\ \|(0, e^{-\alpha_2(t)}(a - c + y_3), 1)\|_{\mathbb{A}} & \text{if } w \in \{w_3, w_5\}, \\ \|(e^{-(\alpha_1+\alpha_2)(t)}(b - ac + y_2), e^{-\alpha_1(t)}(a + c + y_1), 1)\|_{\mathbb{A}} & \text{if } w \in \{w_4, w_6\}. \end{cases}$$

The first case $w \in \{w_1, w_2\}$ again leads to $\operatorname{vol} V_T(w, \varpi_2, t, v') = 0$ for every t with $\tilde{F}^{T_0}(t,T) = 1$.

If $w \in \{w_3, w_5\}$, we now choose a fundamental domain for a as $[c, 1 + c]$ so that this case can in fact be treated similar to the minimal orbit. Hence

$$\operatorname{vol} V_T(w, \varpi_2, t, v') \le e^{-\alpha_2(T)}.$$

Similarly, if $w \in \{w_4, w_6\}$ we can choose the fundamental domains for a and b in such a way that we are left with the same type of estimates as in the case of the minimal orbit. Hence

$$\operatorname{vol} V_T(w, \varpi_2, t, v') \le e^{-3\varpi_2(T)}.$$

The case $\varpi = \varpi_1$: As for the minimal orbit, we obtain

$$e^{-\varpi_1(H_0(w_i t^{-1} y v v' t))} =$$

$$\begin{cases} \|(0,0,1)\|_{\mathbb{A}} = 1 & \text{if } i \in \{1,3\}, \\ \|(0,1,e^{-\alpha_1(t)}(a+c+y_1))\|_{\mathbb{A}} & \text{if } i \in \{2,6\}, \\ \|(1,e^{-\alpha_2(t)}(a-c+y_3), e^{-(\alpha_1+\alpha_2)(t)}\big((a+c+y_1)(a-c+y_3)-(b-ac+y_2)\big))\|_{\mathbb{A}} & \text{if } i \in \{4,5\}. \end{cases}$$

Choosing for each w appropriate fundamental domains for a and b, we are left with the same computations and estimates as in the minimal orbit case.

Taking everything together, we again obtain: For the regular unipotent orbit with Jacobson–Morozov parabolic $P = P_0$ we can approximate $\int_{C_U(u_0,\mathbb{Q})\backslash C_U(u_0,\mathbb{A})} F(vt,T)\,dv$ by $\tilde{F}^{T_0}(t,T)$ asymptotically in T, in fact,

$$\left|\tilde{F}^{T_0}(t,T) - \int_{C_U(X_0,\mathbb{Q})\backslash C_U(X_0,\mathbb{A})} F(vt,T)\,dv\right| \le 8e^{-d(T)}$$

for all $t \in T_0(\mathbb{Q})\backslash T_0(\mathbb{A})$. For $d(T) > \delta\|T\|$ the assertion follows. □

Corollary A.3. *Let $\nu > 0$ be as in Lemma 3.7 and let $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$. There exists a seminorm μ on $\mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ (depending on Φ_f) such that for every $\Phi_\infty \in \mathcal{S}^\nu(\mathfrak{g}(\mathbb{R}))$ and every nilpotent orbit $\mathcal{N}$ we have*

$$\left|j^T_{\mathcal{N}}(\Phi) - \tilde{j}^T_{\mathcal{N}}(\Phi)\right| \le \mu(\Phi_\infty)e^{-c_2\|T\|}$$

for every sufficiently regular $T \in \mathfrak{a}^+$ with $d(T) > \delta\|T\|$, where $\Phi = \Phi_\infty \cdot \Phi_f$.

Proof. Again, we only need to consider the non-trivial orbits, and we moreover may assume that Φ is $\mathbf{K}$-conjugation invariant. First consider the regular orbit $\mathcal{N} = \mathcal{N}_{\mathrm{reg}}$. Using the results and notation of Lemma A.2 and proceeding similar as in the $n = 2$-case, we can bound $\left|j^T_{\mathcal{N}}(\Phi) - \tilde{j}^T_{\mathcal{N}}(\Phi)\right|$ from above by

$$\le c_1 e^{-c_2\|T\|} \int_{A_0^G} \delta_0(a)^{-1}\hat{\tau}_0^G(T - H_0(a)) \int_{\mathfrak{u}^{>2}(\mathbb{A})} \sum_{X \in \mathfrak{u}^2(\mathbb{Q})\cap\mathcal{N}} \left|\Phi(\operatorname{Ad} a^{-1}(X+U))\right| dU\,da.$$

The integral again is bounded from above by a seminorm (independent of T) applied to Φ_∞ as in the proof of Lemma 4.7.

Now consider the case $\mathcal{N} = \mathcal{N}_{\min}$. Similar as before, we are left to estimate

$$c_1 e^{-c_2\|T\|} \int_{A_0^G} \delta_{U^{\le 2}}(a)^{-1}\hat{\tau}_0^G(T - H_0(a)) \sum_{X \in \mathfrak{u}^2(\mathbb{Q})\cap\mathcal{N}} \left|\Phi(\operatorname{Ad} a^{-1}X)\right| da$$

$$\le c_1 e^{-c_2\|T\|} \int_0^{e^{\varpi_1(T)}/2} \int_0^{e^{\varpi_2(T)}/2} a_1^{-2}a_2^{-2} \sum_{x \in \mathbb{Q}\backslash\{0\}} \varphi(a_1^{-1}a_2^{-1}x)\,d^\times a_1\,d^\times a_2,$$

for φ a suitable function. If we change one of the variables to a_1a_2, we can analyse the integral similar as before to obtain the assertion. □

Remark A.4. For the regular orbit, one could prove the corollary without the detailed analysis from the previous lemma by using the behaviour of the test function Φ similarly as in the proof in the case of $n = 2$. See also [CL15] for related work on regular orbits.

Appendix 2: A Sequence of Test Functions

In this appendix, we give a sequence of test functions at the non-Archimedean places which might be useful to deduce Conjecture 1.7 from Proposition 9.2.

For a prime p define $\tilde{\Phi}_p : \mathfrak{g}(\mathbb{Q}_p) \longrightarrow \mathbb{C}$ by

$$\tilde{\Phi}_p(x) = \begin{cases} \frac{[\mathcal{O}_{\mathbb{Q}_p[x]}:\mathbb{Z}_p[x]]}{I_p(\Phi_p^0,x)} = c(\Phi_p^0, x)^{-1} & \text{if } \Delta(x) \neq 0, \text{ and } x \in \mathfrak{g}(\mathbb{Z}_p), \\ 0 & \text{else.} \end{cases}$$

Then $\tilde{\Phi}_p$ is locally constant in $\mathfrak{g}(\mathbb{Q}_p)\backslash\{x \in \mathfrak{g}(\mathbb{Q}_p) \mid \Delta(x) = 0\}$, but not on all of $\mathfrak{g}(\mathbb{Q}_p)$. For $x \in \mathfrak{g}(\mathbb{Q}_p)$ with $\tilde{\Phi}_p(x) \neq 0$, we have

$$c(\tilde{\Phi}_p, x) = \frac{1}{[\mathcal{O}_{\mathbb{Q}_p[x]} : \mathbb{Z}_p[x]]} \int_{G_x(\mathbb{Q}_p)\backslash G(\mathbb{Q}_p)} \tilde{\Phi}_p(g^{-1}xg)\, dg = 1$$

so that in fact one would actually like to use $\tilde{\Phi}_f := \prod_{p<\infty} \tilde{\Phi}_p$ as a test function at the Archimedean places, which we are not allowed to do because of $\tilde{\Phi}_f \notin \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$.

However, we can construct a sequence of functions in $\mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$ converging to $\tilde{\Phi}_f$: Let $\Sigma \subseteq \mathfrak{g}(\mathbb{Z}_p)$ denote the set of all $x \in \mathfrak{g}(\mathbb{Z}_p)$ such that $\Delta(x) = 0$. For $m \in \mathbb{N}_0$ define a function $\Phi_p^m : \mathfrak{g}(\mathbb{Q}_p) \longrightarrow \mathbb{C}$ by

$$\Phi_p^m(x) = \begin{cases} 1 & \text{if } x \in \Sigma + p^m\mathfrak{g}(\mathbb{Z}_p), \\ \tilde{\Phi}_p(x) & \text{if } x \notin \Sigma + p^m\mathfrak{g}(\mathbb{Z}_p). \end{cases}$$

In particular, Φ_p^0 coincides with the characteristic function of $\mathfrak{g}(\mathbb{Z}_p)$. By construction $\Phi_p^m \in \mathcal{S}(\mathfrak{g}(\mathbb{Q}_p))$ and Φ_p^m is $\mathbf{K}_p$-invariant. Let $\mathfrak{m} = (m_p)_{p<\infty}$ be a sequence of integers $m_p \in \mathbb{N}_0$ of which almost all are zero. Let $\mathrm{Div}^+(\mathbb{Q})$ denote the set of all such sequences. It has a partial order given by $\mathfrak{m} \geq \mathfrak{m}'$ if and only if $m_p \geq m_p'$ for all primes p. Define the function $\Phi_f^{\mathfrak{m}} : \mathfrak{g}(\mathbb{A}_f) \longrightarrow \mathbb{C}$ by $\Phi_f^{\mathfrak{m}} = \prod_{p<\infty} \Phi_p^{m_p}$. Then $\Phi_f \in \mathcal{S}(\mathfrak{g}(\mathbb{A}_f))$ and it is $\mathbf{K}_p$-invariant.

By definition we have for all $\mathfrak{m}, \mathfrak{m}' \in \mathrm{Div}^+(\mathbb{Q})$ with $\mathfrak{m} \geq \mathfrak{m}'$ and all $x \in \mathfrak{g}(\mathbb{A}_f)$ we have

$$0 \leq \tilde{\Phi}_f(x) \leq \Phi_f^{\mathfrak{m}}(x) \leq \Phi_f^{\mathfrak{m}'}(x) \leq \Phi_f^0(x) \leq 1.$$

Moreover, $\lim_{\mathfrak{m}} \Phi_f^{\mathfrak{m}}(x) = \tilde{\Phi}_f(x)$ for every x. Similarly, the functions $\Phi_p^{m_p}$ are monotonically decreasing with limit function $\tilde{\Phi}_p$ so that $\lim_{m_p \to \infty} \int_{\mathfrak{g}(\mathbb{Q}_p)} \Phi_p^{m_p}(x)\,dx = \int_{\mathfrak{g}(\mathbb{Q}_p)} \tilde{\Phi}_p(x)\,dx$ and

$$\lim_{m_p \to \infty} \int_{G_\gamma(\mathbb{Q}_p)\backslash G(\mathbb{Q}_p)} \Phi_p^{m_p}(g^{-1}\gamma g)\,dx = \int_{G_\gamma(\mathbb{Q}_p)\backslash G(\mathbb{Q}_p)} \tilde{\Phi}_p(g^{-1}\gamma g)\,dx = 1$$

for all regular elliptic γ. The existence of these limits does not suffice to pass from $c(\xi, \Phi_f)$ to 1 in the asymptotic 9.2 which would prove Conjecture 1.7. It would be necessary to show uniformity of the convergence in $Q(\gamma) = \operatorname{tr}\gamma^2 - \frac{1}{3}(\operatorname{tr}\gamma)^2$ and the number of primes for which $m_p \neq 0$.

References

[Art05] J. Arthur, An introduction to the trace formula. Harmonic analysis, the trace formula, and Shimura varieties, in *Clay Mathematics Proceedings*, vol. 4 (American Mathematical Society, Providence, RI, 2005), pp. 1–263.

[Art78] J. Arthur, A trace formula for reductive groups. I. Terms associated to classes in $G(\mathbf{Q})$. Duke Math. J. **45**(4), 911–952 (1978).

[Art81] J. Arthur, The trace formula in invariant form. Ann. Math. (2) **114**(1) 1–74 (1981).

[Art85] J. Arthur, A measure on the unipotent variety. Can. J. Math. **37**(6), 1237–1274 (1985).

[Art88] J. Arthur, The local behaviour of weighted orbital integrals. Duke Math. J. **56**(2), 223–293 (1988).

[Bha05] M. Bhargava, The density of discriminants of quartic rings and fields. Ann. Math. (2) **162**(2), 1031–1063 (2005).

[Bha10] M. Bhargava, The density of discriminants of quintic rings and fields. Ann. of Math. (2) **172**(3), 1559–1591 (2010).

[BST13] M. Bhargava, A. Shankar, J. Tsimerman, On the Davenport-Heilbronn theorems and second order terms. Invent. Math. **193**(2), 439–499 (2013)

[Cas97] J.W.S. Cassels, *An Introduction to the Geometry of Numbers*. Classics in Mathematics (Springer, Berlin, 1997). Corrected reprint of the 1971 edn.

[Cha02] P.-H. Chaudouard, La formule des traces pour les algèbres de Lie. Math. Ann. **322**(2), 347–382 (2002).

[Cha14] P.-H. Chaudouard, Sur la contribution unipotente dans la formule des traces d'arthur pour les groupes généraux linéaires, arXiv preprint, arXiv:1411.3005 (2014)

[CL15] P.-H. Chaudouard, Gérard Laumon, Sur le comptage des fibres de Hitchin nilpotents. J. Inst. Math. Jussieu **15**(1), 91–164 (2016)

[Coh80] H. Cohn, *Advanced Number Theory* (Dover Publications Inc., New York, 1980). Reprint of ıt A second course in number theory, 1962, Dover Books on Advanced Mathematics.

[Dat96] B.A. Datskovsky, On Dirichlet series whose coefficients are class numbers of binary quadratic forms. Nagoya Math. J. **142**, 95–132 (1996)

[DH71] H. Davenport, H. Heilbronn, On the density of discriminants of cubic fields. II. Proc. R. Soc. Lond. Ser. A **322**(1551), 405–420 (1971).
[DW88] B. Datskovsky, D.J. Wright, Density of discriminants of cubic extensions. J. Reine Angew. Math. **386**, 116–138 (1988).
[EV06] J.S. Ellenberg, A. Venkatesh, The number of extensions of a number field with fixed degree and bounded discriminant. Ann. Math. (2) **163**(2), 723–741 (2006).
[FL11a] T. Finis, E. Lapid, On the Arthur-Selberg trace formula for GL(2). Groups Geom. Dyn. **5**(2), 367–391 (2011).
[FL11b] T. Finis, E. Lapid, On the continuity of Arthur's trace formula: the semisimple terms. Compos. Math. **147**(3), 784–802 (2011).
[Fli82] Y.Z. Flicker, *The Trace Formula and Base Change for* GL(3). Lecture Notes in Mathematics, vol. 927 (Springer, Berlin, 1982).
[FLM11] T. Finis, E. Lapid, W. Müller, On the spectral side of Arthur's trace formula—absolute convergence. Ann. Math. (2) **174**(1), 173–195 (2011).
[Gel96] S. Gelbart, *Lectures on the Arthur-Selberg Trace Formula*, vol. 9 (American Mathematical Society, Providence, RI, 1996)
[GJ72] R. Godement, H. Jacquet, *Zeta Functions of Simple Algebras*. Lecture Notes in Mathematics, vol. 260 (Springer, Berlin, 1972).
[Kim03] T. Kimura, *Introduction to Prehomogeneous Vector Spaces*. Translations of Mathematical Monographs, vol. 215 (American Mathematical Society, Providence, RI, 2003). Translated from the 1998 Japanese original by Makoto Nagura and Tsuyoshi Niitani and revised by the author.
[Lan04] R.P. Langlands, *Beyond Endoscopy*. Contributions to Automorphic Forms, Geometry, and Number Theory (Johns Hopkins University Press, Baltimore, MD, 2004), pp. 611–697.
[Lan94] S. Lang, *Algebraic Number Theory*, 2nd edn. Graduate Texts in Mathematics, vol. 110 (Springer, New York, 1994).
[Lap02] E. Lapid, A different look at the trace formula. Lecture notes (2002)
[Lev01] J. Levy, A truncated integral of the Poisson summation formula. Can. J. Math. **53**(1), 122–160 (2001).
[Lev99] J. Levy, Truncated integrals and the Shintani zeta function for the space of binary quartic forms, in *Automorphic Forms, Automorphic Representations, and Arithmetic (Fort Worth, TX, 1996)*. Proceedings of Symposia in Pure Mathematics, vol. 66 (American Mathematical Society, Providence, RI, 1999), pp. 277–299.
[Mat11] J. Matz, Arthur's trace formula for GL(2) and GL(3) and non-compactly supported test functions. Ph.D. thesis, Heinrich-Heine Universität Düsseldorf (2011)
[MV07] H.L. Montgomery, R.C. Vaughan, *Multiplicative Number Theory. I. Classical Theory*. Cambridge Studies in Advanced Mathematics, vol. 97 (Cambridge University Press, Cambridge, 2007).
[Ngô14] B.C. Ngô, On a certain sum of automorphic *L*-functions. Preprint (2014)
[RR72] R.R. Rao, Orbital integrals in reductive groups, Ann. of Math. (2) **96**, 505–510 (1972).
[Sak14] Y. Sakellaridis, Inverse Satake transforms, arXiv preprint, arXiv:1410.2312 (2014)
[Shi75] T. Shintani, On zeta-functions associated with the vector space of quadratic forms. J. Fac. Sci. Univ. Tokyo Sect. I A Math. **22** 25–65 (1975).
[SS74] M. Sato, T. Shintani, On zeta functions associated with prehomogeneous vector spaces. Ann. Math. (2) **100**, 131–170 (1974).
[Suz79] T. Suzuki, On zeta functions associated with quadratic forms of variable coefficients. Nagoya Math. J. **73**, 117–147 (1979).
[TT13] T. Taniguchi, F. Thorne, Secondary terms in counting functions for cubic fields. Duke Math. J. **162**(13), 2451–2508 (2013)
[Wri85] D.J. Wright, The adelic zeta function associated to the space of binary cubic forms. I. Global theory. Math. Ann. **270**(4), 503–534 (1985).
[WY92] D.J. Wright, A. Yukie, Prehomogeneous vector spaces and field extensions. Invent. Math. **110**(2), 283–314 (1992).
[Yuk92] A. Yukie, On the Shintani zeta function for the space of binary quadratic forms. Math. Ann. **292**(2), 355–374 (1992).

Some Results in the Theory of Low-Lying Zeros of Families of L-Functions

Blake Mackall, Steven J. Miller, Christina Rapti, Caroline Turnage-Butterbaugh, and Karl Winsor

Abstract While Random Matrix Theory has successfully modeled the limiting behavior of many quantities of families of L-functions, especially the distributions of zeros and values, the theory frequently cannot see the arithmetic of the family. In some situations this requires an extended theory that inserts arithmetic factors that depend on the family, while in other cases these arithmetic factors result in contributions which vanish in the limit, and are thus not detected. In this chapter we review the general theory associated with one of the most important statistics, the n-level density of zeros near the central point. According to the Katz–Sarnak density conjecture, to each family of L-functions there is a corresponding symmetry group (which is a subset of a classical compact group) such that the behavior of the zeros near the central point as the conductors tend to infinity agrees with the behavior of the eigenvalues near 1 as the matrix size tends to infinity. We show how these calculations are done, emphasizing the techniques, methods, and obstructions to improving the results, by considering in full detail the family of Dirichlet characters with square-free conductors. We then move on and describe how we may associate a symmetry constant with each family, and how to determine the symmetry group of a compound family in terms of the symmetries of the constituents. These calculations allow us to explain the remarkable universality of behavior, where the main terms

B. Mackall • S.J. Miller (✉)
Department of Mathematics & Statistics, Williams College, Williamstown, MA 01267, USA
e-mail: bmackall60@gmail.com; sjm1@williams.edu; Steven.Miller.MC.96@aya.yale.edu

C. Rapti
Department of Mathematics, Bard College, Annandale-on-Hudson, NY 12504, USA
e-mail: cr9060@bard.edu

C. Turnage-Butterbaugh
Department of Mathematics, Duke University, Durham, NC 27708, USA
e-mail: cturnagebutterbaugh@gmail.com

K. Winsor
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
e-mail: krlwnsr@umich.edu

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_11

are independent of the arithmetic, as we see that only the first two moments of the Satake parameters survive to contribute in the limit. Similar to the Central Limit Theorem, the higher moments are only felt in the rate of convergence to the universal behavior. We end by exploring the effect of lower order terms in families of elliptic curves. We present evidence supporting a conjecture that the average second moment in one-parameter families without complex multiplication has, when appropriately viewed, a negative bias, and end with a discussion of the consequences of this bias on the distribution of low-lying zeros, in particular relations between such a bias and the observed excess rank in families.

1 Introduction

The purpose of this chapter is to describe some results, and the methods used to prove them, in the theory of low-lying zeros and the connections between number theory and random matrix theory. There is now an extensive literature on the subject. See, for example the books [Da, Ed, For, Iw, IwKo, KaSa2, Meh, Ti] and the survey articles [BFMT-B, Con, KaSa1, KeSn1, KeSn2, KeSn3], as well as [Ha, FirMil] for popular accounts of the history of the meeting of the two fields.

Briefly, assuming the Generalized Riemann Hypothesis (GRH) the non-trivial zeros of any nice L-function lie on its critical line, and therefore it is possible to investigate the statistics of its normalized zeros. The work of Montgomery and Odlyzko [Mon, Od1, Od2] suggested that zeros of L-functions in the limit are well modeled by eigenvalues of matrix ensembles. Initially the comparison was made between number theory and the Gaussian Unitary Ensemble (GUE) with statistics such as n-level correlations and spacings between zeros; however, these statistics are insensitive to finitely many zeros and in particular miss the behavior at the central point. This is a significant issue, as there are many situations in number theory where these central values are important, such as the Birch and Swinnerton-Dyer conjecture [BS-D1, BS-D2], and these statistics had the same limiting values both for different families of L-functions and different matrix ensembles. The reader unfamiliar with these statistics and results should see the introduction of [AAILMZ, ILS] (or the introduction of any of the dissertations in low-lying zeros!) for more details.

Following the work of Katz–Sarnak [KaSa1, KaSa2] a new statistic was introduced, the n-level density; unlike the earlier statistics, this depends on the family or ensemble being studied. We mostly concentrate on the 1-level density in this paper, though see [Mil1, Mil2] for some important applications of the 2-level density (which we briefly discuss later).

Let ϕ be an even Schwartz test function on $\mathbb{R}$ whose Fourier transform

$$\hat{\phi}(y) = \int_{-\infty}^{\infty} \phi(x) e^{-2\pi i x y} dx \tag{1}$$

has compact support. Let $\mathscr{F}_N$ be a (finite) family of L-functions satisfying GRH. The 1-level density associated with $\mathscr{F}_N$ is defined by

$$D_{1,\mathscr{F}_N}(\phi) = \frac{1}{|\mathscr{F}_N|} \sum_{f \in \mathscr{F}_N} \sum_j \phi\left(\frac{\log c_f}{2\pi} \gamma_f^{(j)}\right), \tag{2}$$

where $\frac{1}{2} + i\gamma_f^{(j)}$ runs through the non-trivial zeros of $L(s,f)$. Here c_f is the "analytic conductor" of f, and gives the natural scale for the low zeros. As ϕ is Schwartz, only low-lying zeros (i.e., zeros within a distance $\ll 1/\log c_f$ of the central point $s = 1/2$) contribute significantly. Thus the 1-level density can help identify the symmetry type of the family.

Based in part on the function-field analysis where $G(\mathscr{F})$ is the monodromy group associated with the family $\mathscr{F}$, Katz and Sarnak [KaSa1, KaSa2] conjectured that for each reasonable irreducible family of L-functions there is an associated symmetry group $G(\mathscr{F})$ (one of the following five: unitary U, symplectic USp, orthogonal O, SO(even), SO(odd)), and that the distribution of critical zeros near $1/2$ mirrors the distribution of eigenvalues near 1. (Similar correspondences hold for other statistics, such as the values of L-functions being well modeled by values of characteristic polynomials; see, for example, [CFKRS].) The five groups have distinguishable 1-level densities.

To evaluate (2), one applies the explicit formula, converting sums over zeros to sums over primes. By [KaSa1], the 1-level densities for the classical compact groups are

$$\begin{aligned} W_{1,\mathrm{SO(even)}}(x) &= K_1(x,x) \\ W_{1,\mathrm{SO(odd)}}(x) &= K_{-1}(x,x) + \delta(x) \\ W_{1,\mathrm{O}}(x) &= \tfrac{1}{2} W_{1,\mathrm{SO(even)}}(x) + \tfrac{1}{2} W_{1,\mathrm{SO(odd)}}(x) \\ W_{1,\mathrm{U}}(x) &= K_0(x,x) \\ W_{1,\mathrm{USp}}(x) &= K_{-1}(x,x), \end{aligned} \tag{3}$$

where $K(y) = \frac{\sin \pi y}{\pi y}$, $K_\epsilon(x,y) = K(x-y) + \epsilon K(x+y)$ for $\epsilon = 0, \pm 1$, and $\delta(x)$ is the Dirac delta functional. It is often more convenient to work with the Fourier transforms of the densities:

$$\begin{aligned} \widehat{W}_{1,\mathrm{SO(even)}}(u) &= \delta(u) + \tfrac{1}{2} I(u) \\ \widehat{W}_{1,\mathrm{SO(odd)}}(u) &= \delta(u) - \tfrac{1}{2} I(u) + 1 \\ \widehat{W}_{1,\mathrm{O}}(u) &= \delta(u) + \tfrac{1}{2} \\ \widehat{W}_{1,\mathrm{U}}(u) &= \delta(u) \\ \widehat{W}_{1,\mathrm{USp}}(u) &= \delta(u) - \tfrac{1}{2} I(u), \end{aligned} \tag{4}$$

where $I(u)$ is the characteristic function of $[-1,1]$. While these five densities are distinguishable for test functions ϕ where the support of $\hat{\phi}$ exceeds $[-1,1]$, the three orthogonal densities are indistinguishable inside this region. While for many families of interest we cannot calculate the 1-level density beyond $[-1,1]$, we are able to uniquely associate a symmetry group by studying the 2-level densities, which are mutually distinguishable for arbitrarily small support (see [Mil1, Mil2]).

Let $\mathscr{F}$ be a family of L-functions, and $\mathscr{F}_N$ the subset with analytic conductors N (or at most N, or of order N). There is now a large body of work supporting the Katz–Sarnak conjecture that the behavior of zeros near the central point $s = 1/2$ in a family of L-functions (as the conductors tend to infinity) agrees with the behavior of eigenvalues near 1 of a classical compact group (unitary, symplectic, or some flavor of orthogonal). Evidence in support of this conjecture has been obtained for many families of L-functions, including Dirichlet characters [Gao, ER-GR, FioMil, HuRud, LevMil, OS1, OS2, Rub], elliptic curves [HuyKeSn, Mil1, Mil2, Yo1], weight k level N cuspidal newforms [ILS, Ro, HuMil, MilMo, RiRo, Ro], Maass forms [AAILMZ, AMil, GolKon], L-functions attached to number fields [FoIw, MilPe, Ya], symmetric powers of GL_2 automorphic representations [Gü], and Rankin–Selberg convolutions of families [DuMil1, DuMil2] to name a few.

Our purpose is to introduce the reader to some of the techniques and issues of the field. Any introduction must by necessity be brief and must sadly omit many interesting and related results. In particular, we do not discuss other models for zeros near the central point, such as the Hybrid Model (see [GoHuKe], where L-functions are modeled by a partial Euler product, which encodes number theory, and a partial Hadamard product, which is believed to be modeled by matrix ensembles), or the L-function Ratios Conjecture [CFZ1, CFZ2, ConSn, ConSn2, FioMil, GJMMNPP, HuyMM, Mil5, Mil7, MilMo]. We also mostly ignore the issues that arise when studying 2-level (or higher) densities (see [HuMil] for a determination of an alternative to the Katz–Sarnak density conjecture which facilitates comparisons between number theory and random matrix theory).

We begin in Sect. 2 by first calculating the 1-level density of various families of Dirichlet L-functions. This simple family is very amenable to analysis. As such, it provides an excellent introduction to the subject and allows one to see the main ideas and techniques without becoming bogged down in technical computations. We thus show the calculations in complete detail in the hopes that doing so will help introduce newcomers to the subject.

We then turn in Sect. 3 to determining the symmetry group of convolutions of L-functions. Recently Shin and Templier [ShTe] determined the symmetry group for many families (see also the article by Sarnak, Shin, and Templier [SaShTe] in this volume); using the work of Dueñez–Miller [DuMil1, DuMil2] we are able to use inputs such as these to find the symmetry group of Rankin–Selberg convolutions, thus reducing the study of compound families to that of simple ones. In the course of our analysis we see the role lower order terms play. This leads to a nice interpretation of the remarkable universality in behavior between number theory and random matrix theory reminiscent of the universality found in the Central Limit Theorem, which we elaborate on in great detail.

We conclude in Sect. 4 with a *very* brief synopsis of some work in progress on lower order terms in families of elliptic curves, and the effect they have on rates of convergence and detecting the arithmetic of the family (which is missed by the main term in the 1-level density).

2 Families of Dirichlet L-Functions

To date, there has been significant success in showing agreement between zeros near the central point in families of L-functions and eigenvalues near 1 of ensembles of classical compact groups. The purpose of this section is to analyze one of the simplest examples, that of Dirichlet L-functions. The advantage of this calculation is that many of the technical difficulties that plague other families are not present, and thus this provides an excellent opportunity to introduce the reader to the subject. Our first result is the following, proved by Hughes and Rudnick [HuRud].

Theorem 2.1 (1-Level Density for Family of Prime Conductors). *Let $\hat{\phi}$ be an even Schwartz function with supp$(\hat{\phi}) \subset [-2,2]$, m a prime, and $\mathscr{F}_m = \{\chi : \chi$ is primitive mod $m\}$. Then*

$$\frac{1}{\mathscr{F}_m} \sum_{\chi \in \mathscr{F}_m} \sum_{\gamma_\chi : L(\frac{1}{2}+i\gamma_\chi,\chi)=0} \phi\left(\gamma_\chi \frac{\log(m/\pi)}{2\pi}\right) = \int_{-\infty}^{\infty} \phi(y)dy + O\left(\frac{1}{\log m}\right). \tag{5}$$

As $m \to \infty$, the above agrees only with the $N \to \infty$ limit of the 1-level density of $N \times N$ unitary matrices.

The argument below is from notes by the second named author written during the completion of his thesis [Mil1].

After proving this agreement between number theory and random matrix theory, there are two natural ways to proceed. The first is to try to extend the support. It turns out that extending the support is related to deep arithmetic questions concerning the distribution of primes in congruence classes, which we emphasize below. While unfortunately at present there are no unconditional results, recently Fiorilli and Miller [FioMil] showed how to extend the support under various standard assumptions. Depending on the strength of the assumed cancelation, their results range from increasing the support up to $(-4,4)$ all the way to showing agreement for any finite support.

The other direction is to remove the restriction that the conductor is prime.

Theorem 2.2 (Dirichlet Characters from Square-Free Numbers). *Let $\mathscr{F}_{N,\text{sq}-\text{free}}$ denote the family of primitive Dirichlet characters arising from odd square-free numbers $m \in [N, 2N]$, and let $\hat{\phi}$ be an even Schwartz function with supp$(\hat{\phi}) \subset [-2,2]$ Denote the conductor of χ by $c(\chi)$. Then*

$$\begin{aligned} &\frac{1}{|\mathscr{F}_{N,\text{sq}-\text{free}}|} \sum_{\chi \in \mathscr{F}_{N,\text{sq}-\text{free}}} \sum_{\gamma_\chi : L(\frac{1}{2}+i\gamma_\chi,\chi)=0} \phi\left(\gamma_\chi \frac{\log(c(\chi)/\pi)}{2\pi}\right) \\ &= \int_{-\infty}^{\infty} \phi(y)dy + O\left(\frac{1}{\log N}\right). \end{aligned} \tag{6}$$

As $N \to \infty$, the above agrees only with the $N \to \infty$ limit of the 1-level density of $N \times N$ unitary matrices.

While the arguments in [FioMil] also apply to general square-free moduli, their approach is different. We prove this result by first generalizing Theorem 2.1 to a conductor with exactly r distinct prime factors, and obtain good estimates on the error terms as a function of r. Theorem 2.2 then follows by controlling how many square-free numbers have r factors, highlighting a common technique in the subject. We elected to show this method of proof precisely because it showcases an important technique in the subject. It is also possible to attack a fixed m directly, which we do in Theorem 2.9.

2.1 *Dirichlet Characters from Prime Conductors*

Before computing the 1-level density of the low-lying zeros of Dirichlet L-functions, as one of the aims of this article is to provide a self-contained introduction to the subject we first quickly review the needed properties of Dirichlet characters and their associated L-functions. After these preliminaries, we use the explicit formula (see for example [ILS, RudSa]) to relate sums of our test function over the zeros to sums of its Fourier transform weighted by Dirichlet characters. We are able to analyze these sums very easily due to the orthogonality relations of Dirichlet characters, and obtain support up to $[-2, 2]$. See [Da, IwKo] for more on Dirichlet characters.

2.1.1 Review of Dirichlet Characters

If m is prime, then $(\mathbb{Z}/m\mathbb{Z})^*$ is cyclic of order $m-1$ with generator g (so any element is of the form g^a for some a). Let $\zeta_{m-1} = e^{2\pi i/(m-1)}$. The principal character χ_0 is

$$\chi_0(k) = \begin{cases} 1 & \text{if } (k, m) = 1 \\ 0 & \text{if } (k, m) > 1. \end{cases} \tag{7}$$

Each of the $m - 2$ primitive characters are determined (because they are multiplicative) once their action on a generator g is specified. As each $\chi : (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{C}^*$, for each χ there exists an l such that $\chi(g) = \zeta_{m-1}^{\ell}$. Hence for each ℓ, $1 \le \ell \le m-2$, we have

$$\chi_\ell(k) = \begin{cases} \zeta_{m-1}^{\ell a} & \text{if } k \equiv g^a \bmod m \\ 0 & \text{if } (k, m) > 0. \end{cases} \tag{8}$$

In most families one is not so fortunate to have such explicit formulas; these facilitate many calculations (such as proving the orthogonality relations for sums over the characters).

Let χ be a primitive character modulo m. Set

$$c(m, \chi) = \sum_{k=0}^{m-1} \chi(k) e^{2\pi i k/m}; \tag{9}$$

$c(m, \chi)$ is a Gauss sum of modulus $\sqrt{m}$. The associated L-function $L(s, \chi)$ (and the completed L-function $\Lambda(s, \chi)$) are given by

$$L(s,\chi) = \prod_p (1 - \chi(p)p^{-s})^{-1}$$

$$\Lambda(s,\chi) = \pi^{-\frac{1}{2}(s+\epsilon)}\Gamma\left(\frac{s+\epsilon}{2}\right)m^{\frac{1}{2}(s+\epsilon)}L(s,\chi), \tag{10}$$

where

$$\epsilon = \begin{cases} 0 & \text{if } \chi(-1) = 1 \\ 1 & \text{if } \chi(-1) = -1 \end{cases}$$

$$\Lambda(s,\chi) = (-i)^{\epsilon}\frac{c(m,\chi)}{\sqrt{m}}\Lambda(1-s,\bar{\chi}). \tag{11}$$

Let ϕ be an even Schwartz function with compact support, say contained in the interval $(-\sigma, \sigma)$, and let χ be a non-trivial primitive Dirichlet character of conductor m. The explicit formula[1] gives

$$\begin{aligned} \sum_{\gamma_\chi} \phi\left(\gamma_\chi \frac{\log(\frac{m}{\pi})}{2\pi}\right) = & \int_{-\infty}^{\infty} \phi(y)dy \\ & - \sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right)[\chi(p) + \overline{\chi}(p)]p^{-1/2} \\ & - \sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)[\chi^2(p) + \overline{\chi}^2(p)]p^{-1} \\ & + O\left(\frac{1}{\log m}\right), \end{aligned} \tag{12}$$

where we are assuming GRH[2] to write the zeros as $\frac{1}{2} + i\gamma_\chi$, $\gamma_{\chi_\chi} \in \mathbb{R}$, and the contribution from the primes to the third and higher powers are absorbed in the big-Oh term.[3] Sometimes it is more convenient to normalize the zeros not by the

[1]The derivation is by doing a contour integral of the logarithmic derivative of the completed L-function times the test function, using the Euler product and shifting contours; see [RudSa] for details.

[2]It is worth noting that these formulas hold without assuming GRH. In that case, however, the zeros no longer lie on a common line and we lose the correspondence with eigenvalues of Hermitian matrices.

[3]A similar absorbtion holds in other families, so long as the Satake parameters satisfy $|\alpha_i(p)| \leq Cp^{\delta}$ for some $\delta < 1/6$.

logarithm of the analytic conductor but rather by something that is the same to first order.[4] Explicitly, for $m \in [N, 2N]$ we have

$$\begin{aligned}\sum_{\gamma_\chi} \phi\left(\gamma_\chi \frac{\log(\frac{N}{\pi})}{2\pi}\right) = \frac{\log(m/\pi)}{\log(N/\pi)} \int_{-\infty}^{\infty} \phi(y)dy \\ - \sum_p \frac{\log p}{\log(N/\pi)} \hat{\phi}\left(\frac{\log p}{\log(N/\pi)}\right) [\chi(p) + \overline{\chi}(p)] p^{-1/2} \\ - \sum_p \frac{\log p}{\log(N/\pi)} \hat{\phi}\left(2\frac{\log p}{\log(N/\pi)}\right) [\chi^2(p) + \overline{\chi}^2(p)] p^{-1} \\ + O\left(\frac{1}{\log N}\right),\end{aligned} \tag{13}$$

and for any subset $\mathscr{N}$ of $[N, 2N]$

$$\frac{1}{|\mathscr{N}|} \sum_{m \in \mathscr{N}} \frac{\log(m/\pi)}{\log(N/\pi)} = 1 + O\left(\frac{1}{\log N}\right). \tag{14}$$

Consider $\mathscr{F}_m$, the family of primitive characters modulo a prime m. There are $m-2$ elements in this family, given by $\{\chi_\ell\}_{1 \le \ell \le m-2}$. As each χ_ℓ is primitive, we may use the Explicit Formula. To determine the 1-level density we must evaluate

$$\begin{aligned}\int_{-\infty}^{\infty} \phi(y)dy - \frac{1}{m-2} \sum_{\substack{\chi(m) \\ \chi \neq \chi_0}} \sum_p \frac{\log p}{\log(m/\pi)} \hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right) [\chi(p) + \overline{\chi}(p)] p^{-1/2} \\ - \frac{1}{m-2} \sum_{\substack{\chi(m) \\ \chi \neq \chi_0}} \sum_p \frac{\log p}{\log(m/\pi)} \hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right) [\chi^2(p) + \overline{\chi}^2(p)] p^{-1} \\ + O\left(\frac{1}{\log m}\right).\end{aligned} \tag{15}$$

Definition 2.3 (First and Second Sums). We call the two sums in (15) the First Sum and the Second Sum (respectively), denoting them by $S_1(m;\phi)$ and $S_2(m;\phi)$.

The Density Conjecture states that the family average should converge to the Unitary Density:

[4] We comment on this in greater length when we consider the family of all characters with square-free modulus. Briefly, a constancy in the conductors allows us to pass certain sums through the test functions to the coefficients. This greatly simplifies the analysis of the 1-level density; unfortunately cross terms arise in the 2-level and higher cases, and the savings vanish (see [Mil1, Mil2]).

$$\lim_{m\to\infty} D_{1,\mathscr{F}_m}(\phi) = \lim_{m\to\infty} \sum_{\substack{\chi(m)\\ \chi\neq\chi_0}} \sum_{\gamma_\chi} \phi\left(\gamma_\chi \frac{\log(\frac{m}{\pi})}{2\pi}\right) = \int_{-\infty}^{\infty} \phi(y)dy. \tag{16}$$

We prove this for $\hat{\phi}$ supported in $[-2,2]$, which establishes Theorem 2.1. We break the proof into two steps. First, we show in Lemmas 2.4 and 2.5 that the first sum does not contribute as $m\to\infty$ for such $\hat{\phi}$, and then complete the proof in Lemma 2.6 by showing the second sum does not contribute for any finite support.

2.1.2 The First Sum $S_1(m;\phi)$

As one of our goals is to see how far we can get with elementary methods, in the lemma below we show that simple estimation of the prime sums allows us to determine the 1-level for support up to $(-2,2)$, and then immediately strengthen it by using the Brun–Titchmarsh Theorem to get it for $[-2,2]$.

Lemma 2.4 (Contribution from $S_1(m;\phi)$). *For* $\mathrm{supp}(\hat{\phi}) \subset (-\sigma,\sigma)$ *and m prime,* $S_1(m;\phi) \ll m^{\sigma/2-1}$*, implying that this term does not contribute to the main term in the 1-level density for* $\sigma < 2$*.*

Proof. We must analyze

$$S_1(m;\phi) = \frac{1}{m-2} \sum_{\substack{\chi(m)\\ \chi\neq\chi_0}} \sum_{p} \frac{\log p}{\log(m/\pi)} \hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right) [\chi(p)+\overline{\chi}(p)]p^{-1/2}. \tag{17}$$

Since the orthogonality of the Dirichlet characters implies

$$\sum_{\chi(m)} \chi(k) = \begin{cases} m-1 & \text{if } k\equiv 1 \bmod m\\ 0 & \text{otherwise,}\end{cases} \tag{18}$$

we have for any prime $p\neq m$ that

$$\sum_{\substack{\chi(m)\\ \chi\neq\chi_0}} \chi(p) = \begin{cases} m-2 & \text{if } p\equiv 1 \bmod m\\ -1 & \text{otherwise.}\end{cases} \tag{19}$$

Let

$$\delta_m(p,1) = \begin{cases} 1 & \text{if } p\equiv 1 \bmod m\\ 0 & \text{otherwise.}\end{cases} \tag{20}$$

The contribution to the sum from $p = m$ is zero; if instead we substitute -1 for $\sum_{\substack{\chi(m)\\ \chi\neq\chi_0}} \chi(m)$, our error is $O\left(1/\sqrt{m}\right)$ and hence negligible relative to the other errors.

We now calculate $S_1(m;\phi)$ with $\hat{\phi}$ an even Schwartz function with support in $(-\sigma,\sigma)$. As the conductors are constant in the family, we may interchange the summations and first average over the family. This allows us to exploit the cancelation in sums of Dirichlet characters.

$$\begin{aligned}
S_1(m;\phi) &= \frac{1}{m-2}\sum_{\substack{\chi(m)\\ \chi\neq\chi_0}}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right)[\chi(p)+\overline{\chi}(p)]p^{-1/2} \\
&= \frac{1}{m-2}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right)\sum_{\substack{\chi(m)\\ \chi\neq\chi_0}}[\chi(p)+\overline{\chi}(p)]p^{-1/2} \\
&= \frac{2}{m-2}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right)p^{-1/2}(-1+(m-1)\delta_m(p,1)) \\
&\ll \frac{1}{m}\sum_{p=2}^{m^\sigma} p^{-1/2} + \sum_{\substack{p=1\\ p\equiv 1(m)}}^{m^\sigma} p^{-1/2} \\
&\ll \frac{1}{m}\sum_{k=1}^{m^\sigma} k^{-1/2} + \sum_{\substack{k=m+1\\ k\equiv 1(m)}}^{m^\sigma} k^{-1/2} \\
&\ll \frac{1}{m}\sum_{k=1}^{m^\sigma} k^{-1/2} + \frac{1}{m}\sum_{k=1}^{m^\sigma} k^{-1/2} \ll \frac{1}{m}m^{\sigma/2}.
\end{aligned} \tag{21}$$

Notice that we had to be careful with the estimates of the sum over primes congruent to 1 modulo m. Each residue class modulo m has approximately the same sum, with the difference between two classes bounded by the first term of whichever class has the smallest element. As our numbers k are of the form $\ell m+1$ for $\ell \in \{1,2,3,\dots\}$, the class $k\equiv 1(m)$ has the smallest sum of the m classes. Thus if we add all the classes modulo m and divide by m, we increase the sum, justifying the above arguments.

Hence $S_1(m;\phi) = \frac{1}{m}m^{\sigma/2}$, implying that there is no contribution from the first sum if $\sigma < 2$. □

The next lemma illustrates a common theme in the subject: additional arithmetic information translates to increased support (and vice-versa).

Lemma 2.5. *For* $\mathrm{supp}(\hat{\phi}) \subset [-2,2]$ *and m prime,* $S_1(m;\phi) \ll 1/\log m$, *implying that this term does not contribute to the main term in the 1-level density.*

Proof. Following [HuRud] we use the Brun–Titchmarsh Theorem to improve our bound for the prime sums in (21) when $\sigma = 2$. Revisiting that calculation, we find

$$S_1(m;\phi) \ll \frac{1}{m\log m}\sum_{p=1}^{m^2}\frac{\log p}{\sqrt{p}} + \frac{1}{\log m}\sum_{\substack{p=1\\ p\equiv 1(m)}}^{m^2}\frac{\log p}{\sqrt{p}}. \tag{22}$$

The Brun–Titchmarsh theorem (see [HuRud, MonVa]) states that if $x > 2m$ and $(a, m) = 1$ then

$$\pi(x; m, a) \ := \ \#\{p \leq x : p \equiv a(m)\} \ < \ \frac{2x}{\phi(m)\log(x/m)}. \tag{23}$$

We can trivially bound the contribution from the primes in (22) less than $2q$ by the arguments from Lemma 2.4, and for the remaining we argue as in [HuRud]. The two sums are handled similarly. For example, for the second prime sum we have

$$\frac{1}{\log m}\sum_{\substack{p>2m \\ p\equiv 1(m)}}^{m^2}\frac{\log p}{p^{-1/2}} \ll \frac{1}{\log m}\int_{2m}^{m^2}\frac{\log x}{\sqrt{x}}\frac{1}{m}\frac{dx}{\log(x/m)} \ll \frac{1}{\log m}, \tag{24}$$

proving that this term does not contribute when $\sigma = 2$. The first prime sum in (22) follows analogously, completing the proof. □

2.1.3 The Second Sum $S_2(m; \phi)$

Lemma 2.6 (Contribution from $S_2(m; \phi)$**).** *For* $\text{supp}(\hat{\phi}) \subset (-\sigma, \sigma)$ *and m prime,* $S_2(m; \phi) \ll \sigma\frac{\log m}{m}$*, implying that this term does not contribute to the main term in the 1-level density for any finite* σ*.*

Proof. We must analyze (for m prime)

$$S_2(m; \phi) = \frac{1}{m-2}\sum_{\substack{\chi(m) \\ \chi\neq\chi_0}}\sum_{p}\frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)[\chi^2(p) + \overline{\chi}^2(p)]p^{-1}. \tag{25}$$

The orthogonality relations immediately imply

$$S(m) \ := \ \sum_{\substack{\chi(m) \\ \chi\neq\chi_0}}[\chi^2(p) + \overline{\chi}^2(p)] = \begin{cases} 2(m-2) & \text{if } p \equiv \pm 1(m) \\ -2 & \text{if } p \not\equiv \pm 1(m). \end{cases} \tag{26}$$

The proof is straightforward as $\chi^2(p) = \chi(p^2)$ (and similarly for $\overline{\chi}$).

Let

$$\delta_m(p, \pm) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \bmod m \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

We argue as we did in our analysis of $S_1(m;\phi)$ in Lemma 2.4, and find

$$\begin{aligned}
S_2(m;\phi) &= \frac{1}{m-2}\sum_{\substack{\chi(m)\\ \chi\neq\chi_0}}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)[\chi^2(p)+\overline{\chi}^2(p)]p^{-1}\\
&= \frac{1}{m-2}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)\sum_{\substack{\chi(m)\\ \chi\neq\chi_0}}[\chi^2(p)+\overline{\chi}^2(p)]p^{-1}\\
&= \frac{1}{m-2}\sum_p^{m^{\sigma/2}} \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)p^{-1}[-2+(2m-2)\delta_m(p,\pm)]\\
&\ll \frac{1}{m-2}\sum_p^{m^{\sigma/2}} p^{-1} + \frac{2m-2}{m-2}\sum_{\substack{p=1\\ p\equiv\pm1(m)}}^{m^{\sigma/2}} p^{-1}\\
&\ll \frac{1}{m-2}\sum_{k=1}^{m^{\sigma/2}} k^{-1} + \sum_{\substack{k=m+1\\ k\equiv 1(m)}}^{m^{\sigma/2}} k^{-1} + \sum_{\substack{k=m-1\\ k\equiv -1(m)}}^{m^{\sigma/2}} k^{-1}\\
&\ll \frac{1}{m-2}\log(m^{\sigma/2}) + \frac{1}{m}\sum_{k=1}^{m^{\sigma/2}} k^{-1} + \frac{1}{m}\sum_{k=1}^{m^{\sigma/2}} k^{-1} + O\left(\frac{1}{m}\right)\\
&\ll \sigma\left(\frac{\log m}{m}+\frac{\log m}{m}+\frac{\log m}{m}+\frac{1}{m}\right).
\end{aligned}\tag{28}$$

Therefore $S_2(m;\phi) = O(\sigma\frac{\log m}{m})$, so for all fixed, finite σ there is no contribution. □

2.2 *Dirichlet Characters from Square-Free Conductors*

We now remove the restriction that m is prime and consider the more general case of square-free conductors. The purpose of this section is to highlight some of the issues that arise in the analysis of low-lying zeros in families of L-functions in a setting where the methods can be appreciated without being overwhelmed by technical details.

Specifically, we discuss the question of how to normalize these zeros (either locally or globally), as well as how to combine results from different cases. We find it is convenient to partition the space of characters by the number of prime factors, which we denote by r, of their conductors. We then generalize our bounds on the first and second sums, explicitly determining the r dependence. The proof is completed by standard results on sums of the divisor function. This procedure is used in the analysis of many other families. For example, in [ILS] the analysis of newforms is accomplished by using inclusion–exclusion to apply the Petersson formula to the various spaces of oldforms, removing their contributions and carefully combining the errors.

Our main result is Theorem 2.2. As the proof is similar to the proof of Theorem 2.1, we content ourselves below with highlighting the differences. The first choice is how to normalize the zeros of each Dirichlet L-function. We can split our family by the conductor, and note that the normalization of the zeros depends *only* on this quantity. Further, this number varies monotonically as we move from N to $2N$. While we could normalize by the average log-conductor, or even by $\log N$, there is no additional work to rescale each L-function's zeros by the logarithm of the conductor. The reason is that we will break the analysis below by the size of the conductor, and our first and second sums do not contribute. The situation is different for the contribution from the Gamma factor; however, by (14) there is no affect on the main terms. While the situation appears different if we looked at the 2-level density, as then we would have cross terms and would have to deal with sums of products of logarithms of conductors and Dirichlet characters, there is no difficulty here as the conductors are constant among characters with the same moduli, and monotonically increasing with the moduli. These properties allow us to again break the analysis into characters with the same moduli. The situation is very different for one-parameter families of elliptic curves. There, we have to be significantly more careful, as these cross terms become much harder to handle. For more on these issues, see [Mil1, Mil2].

Before proving Theorem 2.2, we first set some notation and isolate some useful results. Fix an $r \geq 1$ and distinct, odd primes $m_1, \dots, m_r$. Let

$$\begin{aligned} m &:= m_1 m_2 \cdots m_r \\ M_1 &:= (m_1-1)(m_2-1)\cdots(m_r-1) = \phi(m) \\ M_2 &:= (m_1-2)(m_2-2)\cdots(m_r-2). \end{aligned} \tag{29}$$

Note M_2 is the number of primitive characters mod m, each of conductor m. For each $\ell_i \in [1, m_i-2]$ we have the primitive character discussed in the previous section, χ_{ℓ_i}. A general primitive character mod m is given by a product of these characters:

$$\chi(u) = \chi_{\ell_1}(u)\chi_{\ell_2}(u)\cdots\chi_{\ell_r}(u). \tag{30}$$

Let $\mathscr{F}_m = \{\chi : \chi = \chi_{\ell_1}\chi_{\ell_2}\cdots\chi_{\ell_r}\}$. Then $|\mathscr{F}_m| = M_2$, and we are led to investigating the following sums:

$$\begin{aligned} S_1(m,r;\phi) &= \frac{1}{M_2}\sum_{\chi\in\mathscr{F}_m}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right)\frac{\chi(p)+\overline{\chi}(p)}{\sqrt{p}} \\ S_2(m,r;\phi) &= \frac{1}{M_2}\sum_{\chi\in\mathscr{F}_m}\sum_p \frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)\frac{\chi^2(p)+\overline{\chi}^2(p)}{p}; \end{aligned} \tag{31}$$

we have added an r in the notation above to highlight the fact that m has r distinct odd prime factors. We first bound these two sums in terms of r, and then sum over r to complete the proof of Theorem 2.2.

2.2.1 The First Sum $S_1(m, r; \phi)$ (m Square-Free)

Lemma 2.7 (Contribution from $S_1(m, r; \phi)$). *Notation as above (in particular, m has r factors),*

$$S_1(m, r; \phi) \ll \frac{1}{M_2} 2^r m^{\sigma/2}. \tag{32}$$

Proof. We must study $\sum_{\chi \in \mathscr{F}_m} \chi(p)$ (the sum with $\overline{\chi}$ is handled similarly). Earlier we showed

$$\sum_{\ell_i=1}^{m_i-2} \chi_{\ell_i}(p) = \begin{cases} m_i - 1 - 1 & \text{if } p \equiv 1 \bmod m_i \\ -1 & \text{otherwise.} \end{cases} \tag{33}$$

Define

$$\delta_{m_i}(p, 1) = \begin{cases} 1 & \text{if } p \equiv 1 \bmod m_i \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

Then

$$\begin{aligned} \sum_{\chi \in \mathscr{F}_m} \chi(p) &= \sum_{\ell_1=1}^{m_1-2} \cdots \sum_{\ell_r=1}^{m_r-2} \chi_{\ell_1}(p) \cdots \chi_{\ell_r}(p) \\ &= \prod_{i=1}^{r} \sum_{\ell_i=1}^{m_i-2} \chi_{\ell_i}(p) = \prod_{i=1}^{r} (-1 + (m_i - 1)\delta_{m_i}(p, 1)). \end{aligned} \tag{35}$$

Let us denote by $k(s)$ an s-tuple $(k_1, k_2, \ldots, k_s)$ with $k_1 < k_2 < \cdots < k_s$. This is just a subset of $\{1, 2, \ldots, r\}$. There are 2^r possible choices for $k(s)$. We use these to expand the above product. Define

$$\delta_{k(s)}(p, 1) = \prod_{i=1}^{s} \delta_{m_{k_i}}(p, 1). \tag{36}$$

If $s = 0$, we set $\delta_{k(0)}(p, 1) = 1$ for all p. Then

$$\prod_{i=1}^{r} (-1 + (m_i - 1)\delta_{m_i}(p, 1)) = \sum_{s=0}^{r} \sum_{k(s)} (-1)^{r-s} \delta_{k(s)}(p, 1) \prod_{i=1}^{s} (m_{k_i} - 1). \tag{37}$$

Let $h(p) = 2\frac{\log p}{\log(m/\pi)}\hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right) \ll ||\hat{\phi}||$. Then

$$\begin{aligned} S_1 &= \sum_p^{m^\sigma} \frac{1}{2} h(p) p^{-1/2} \frac{1}{M_2} \sum_{\chi \in \mathscr{F}} [\chi(p) + \overline{\chi}(p)] \\ &= \sum_p^{m^\sigma} h(p) p^{-1/2} \frac{1}{M_2} \sum_{s=0}^{r} \sum_{k(s)} (-1)^{r-s} \delta_{k(s)}(p,1) \prod_{i=1}^{s} (m_{k_i} - 1) \\ &\ll \sum_p^{m^\sigma} p^{-1/2} \frac{1}{M_2} \left(1 + \sum_{s=1}^{r} \sum_{k(s)} \delta_{k(s)}(p,1) \prod_{i=1}^{s} (m_{k_i} - 1) \right). \end{aligned} \tag{38}$$

Observing that $m/M_2 \le 3^r$ we see the $s = 0$ sum contributes

$$S_{1,0} = \frac{1}{M_2} \sum_p^{m^\sigma} p^{-1/2} \ll 3^r m^{\sigma/2-1}, \tag{39}$$

which is negligible for $\sigma < 2$, though it is also bounded by $m^{\sigma/2-1}/M_2$. Now we study

$$S_{1,k(s)} = \frac{1}{M_2} \prod_{i=1}^{s} (m_{k_i} - 1) \sum_p^{m^\sigma} p^{-1/2} \delta_{k(s)}(p,1). \tag{40}$$

The effect of the factor $\delta_{k(s)}(p,1)$ is to restrict the summation to primes $p \equiv 1(m_{k_i})$ for $k_i \in k(s)$. The sum will increase if instead of summing over primes satisfying the congruences we sum over all numbers n satisfying the congruences (with $n \ge 1 + \prod_{i=1}^{s} m_{k_i}$). As the sum is now over integers and not primes, we can use basic uniformity properties of integers to bound it. We are summing integers mod $\prod_{i=1}^{s} m_{k_i}$, so summing over integers satisfying these congruences is basically just $\prod_{i=1}^{s}(m_{k_i})^{-1} \sum_{n=1}^{m^\sigma} n^{-1/2} = \prod_{i=1}^{s}(m_{k_i})^{-1} m^{\sigma/2}$. We can do this as the sum of the reciprocals from the residue classes of $\prod_{i=1}^{s} m_{k_i}$ differ by at most their first term. Throwing out the first term of the class $1 + \prod_{i=1}^{s} m_{k_i}$ makes it have the smallest sum of the $\prod_{i=1}^{s} m_{k_i}$ classes, so adding all the classes and dividing by $\prod_{i=1}^{s} m_{k_i}$ increases the sum. Hence (recalling $m/M_2 \le 3^r$)

$$S_{1,k(s)} \ll \frac{1}{M_2} \prod_{i=1}^{s} (m_{k_i} - 1) \prod_{i=1}^{s} (m_{k_i})^{-1} m^{\sigma/2} \ll 3^r m^{\sigma/2-1}, \tag{41}$$

though it is also bounded by $m^{\sigma/2-1}/M_2$. Therefore, for all s the $S_{1,k(s)}$ contribute $3^r m^{\sigma/2-1}$. There are 2^r choices, yielding

$$S_1 \ll 6^r m^{\sigma/2-1}, \tag{42}$$

which is negligible as m goes to infinity *for fixed* r if $\sigma < 2$. If instead we do not use $m/M_2 \leq 3^r$, we obtain a bound of $O(2^r m^{\sigma/2}/M_2)$. □

The worst errors occur when m is the product of the first r primes. Let p_i denote the ith prime. The Prime Number Theorem implies for r large that

$$\log m = \sum_{p \leq p_r} \log p \ \sim \ p_r. \tag{43}$$

As $p_r \sim r \log r$, we find $\log m \sim r \log r$ or $r \sim \log m / \log\log m$. Thus

$$6^r \ \sim \ e^{r \log 6} \ \sim \ m^{\log 6/\log\log m}. \tag{44}$$

While this is $o(m^\epsilon)$ for any $\epsilon > 0$, this estimate is wasteful when m has few prime factors. For example, if $m = 10^{50}$, then $m^{\log 6/\log\log m} \sim m^{0.3775}$, which is sizable. We thus prefer to leave the estimate of $S_1(m, r; \phi)$ as a function of r, and then average over the number of square-free integers with exactly r distinct odd prime factors. Such a division will lead to significantly better results for the family of square-free conductors.

2.2.2 The Second Sum $S_2(m, r; \phi)$ (m Square-Free)

Lemma 2.8 (Contribution from $S_2(m, r; \phi)$). *Notation as above (in particular, m has r factors),*

$$S_2(m, r; \phi) \ll \frac{1}{M_2} 3^r m^{1/2}. \tag{45}$$

Proof. We must study $\sum_{\chi \in \mathscr{F}} \chi^2(p)$ (the sum with $\overline{\chi}$ is handled similarly). Earlier we showed

$$\sum_{\ell_i=1}^{m_i-2} \chi_{\ell_i}^2(p) = \begin{cases} m_i - 1 - 1 & \text{if } p \equiv \pm 1 \bmod m_i \\ -1 & \text{otherwise.} \end{cases} \tag{46}$$

Then

$$\begin{aligned} \sum_{\chi \in \mathscr{F}} \chi^2(p) &= \sum_{\ell_1=1}^{m_1-2} \cdots \sum_{\ell_r=1}^{m_r-2} \chi_{\ell_1}^2(p) \cdots \chi_{\ell_r}^2(p) \\ &= \prod_{i=1}^{r} \sum_{\ell_i=1}^{m_i-2} \chi_{\ell_i}^2(p) \\ &= \prod_{i=1}^{r} (-1 + (m_i - 1)\delta_{m_i}(p, 1) + (m_i - 1)\delta_{m_i}(p, -1)). \end{aligned} \tag{47}$$

Instead of having 2^r terms as in the first sum, now we have 3^r. Let $k(s)$ be as before, and let $j(s)$ be an s-tuple of ± 1's. As s ranges from 0 to r we get each of the 3^r possibilities, as for a fixed s there are $\binom{r}{s}$ choices for $k(s)$, each of these having 2^s choices for $j(s)$ (note $\sum_{s=0}^{r} 2^s\binom{r}{k} = (1+2)^r$). Let $h(p) = 2\frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right) \ll ||\hat{\phi}||$. Define

$$\delta_{k(s)}(p,j(s)) = \prod_{i=1}^{s} \delta_{m_{k_i}}(p,j_i). \tag{48}$$

Then

$$\sum_{\chi\in\mathscr{F}} \chi^2(p) = \sum_{s=0}^{r}\sum_{k(s)}\sum_{j(s)}(-1)^{r-s}\delta_{k(s)}(p,j(s))\prod_{i=1}^{s}(m_{k_i}-1). \tag{49}$$

Therefore

$$\begin{aligned}
S_2 &= \frac{1}{M_2}\sum_{p}\frac{\log p}{\log(m/\pi)}\hat{\phi}\left(2\frac{\log p}{\log(m/\pi)}\right)p^{-1}\sum_{\chi\in\mathscr{F}}[\chi^2(p)+\overline{\chi}^2(p)] \\
&= \frac{1}{M_2}\sum_{p} h(p)\sum_{s=0}^{r}\sum_{k(s)}\sum_{j(s)}p^{-1}(-1)^{r-s}\delta_{k(s)}(p,j(s))\prod_{i=1}^{s}(m_{k_i}-1) \\
&\ll \frac{1}{M_2}\sum_{p}\sum_{s=0}^{r}\sum_{k(s)}\sum_{j(s)}p^{-1}\delta_{k(s)}(p,j(s))\prod_{i=1}^{s}(m_{k_i}-1) \\
&= \sum_{s=0}^{r}\sum_{k(s)}\sum_{j(s)}S_{2,k(s),j(s)}.
\end{aligned} \tag{50}$$

The term where $s=0$ is handled easily (recall $m/M_2 \leq 3^r$):

$$S_{2,0,0} = \frac{1}{M_2}\sum_{p}^{m^\sigma}p^{-1} \ll 3^r\frac{\log m^\sigma}{m} \tag{51}$$

(we could also bound it by $\sigma\log(m)/M_2$).

We would like to handle the terms for $s \neq 0$ analogously as before. The congruences on p from $k(s)$ and $j(s)$ force us to sum only over certain primes mod $\prod_{i=1}^{s} m_{k_i}$, with each prime satisfying $p \geq m_{k_i} \pm 1$. We increase the sum by summing over all integers satisfying these congruences. As each congruence class mod $\prod_{i=1}^{s} m_{k_i}$ has basically the same sum, we can bound our sum over primes satisfying the congruences $k(s), j(s)$ by $\prod_{i=1}^{s}(m_{k_i})^{-1}\sum_{n=1}^{m^\sigma} n^{-1} = \prod_{i=1}^{s}(m_{k_i})^{-1}\log m^\sigma$.

There is one slight problem with this argument. Before each prime was congruent to 1 mod each prime m_{k_i}, hence the first prime occurred no earlier than at $1 + \prod_{k=1}^{s} m_{k_i}$. Now, however, some primes are congruent to $+1$ mod m_{k_i} while others are congruent to -1, and it is possible the first such prime occurs before $\prod_{k=1}^{s} m_{k_i}$.

For example, say the prime is congruent to $+1$ mod 11, and -1 mod $3, 5, 17$. We want the prime to be greater than $3 \cdot 5 \cdot 11 \cdot 17$, but $3 \cdot 5 \cdot 17 - 1$ is congruent to -1 mod $3, 5, 17$ and $+1$ mod 11. (Fortunately it equals 254, which is composite.)

So, for each pair $(k(s), j(s))$ we handle all but the possibly first prime as we did in the First Sum case. We now need an estimate on the possible error for low primes. Fortunately, there is at most one for each pair, and as our sum has a $1/p$, we can expect cancelation if it is large.

Fix now a pair (remember there are at most 3^r pairs). As we never specified the order of the primes m_i, without loss of generality (basically, for notational convenience) we may assume that our prime p is congruent to $+1$ mod $m_{k_1} \cdots m_{k_a}$, and -1 mod $m_{k_{a+1}} \cdots m_{k_s}$.

The contribution to the second sum from the possible low prime in this pair is

$$\frac{1}{M_2} \frac{1}{p} \prod_{i=1}^{s} (m_{k_i} - 1). \tag{52}$$

How small can p be? The $+1$ congruences imply that $p \equiv 1 (m_{k_1} \cdots m_{k_a})$, so p is at least $m_{k_1} \cdots m_{k_a} + 1$. Similarly the -1 congruences imply p is at least $m_{k_{a+1}} \cdots m_{k_s} - 1$. Since the product of these two lower bounds is greater than $\prod_{i=1}^{s} (m_{k_i} - 1)$, at least one must be greater than $\left(\prod_{i=1}^{s} (m_{k_i} - 1)\right)^{1/2}$. Therefore the contribution to the second sum from the possible low prime in this pair is bounded by (remember $m/M_2 \leq 3^r$)

$$\frac{1}{M_2} \left(\prod_{i=1}^{s} (m_{k_i} - 1) \right)^{1/2} \leq \frac{m^{1/2}}{M_2} \leq 3^r m^{-1/2}. \tag{53}$$

Combining this with the estimate for the primes larger than $\prod_{i=1}^{s} (m_{k_i} - 1)$ yields

$$S_{2,k(s),j(s)} \ll 3^r m^{-1/2} + \frac{3^r}{m} \log m^{\sigma}, \tag{54}$$

yielding (as there are 3^r pairs)

$$S_2 = \sum_{s=0}^{r} \sum_{k(s)} \sum_{j(s)} S_{2,k(s),j(s)} \ll 9^r m^{-1/2}; \tag{55}$$

if we don't use $m/M_2 \leq 3^r$ we find a bound of $3^r m^{1/2}/M_2$. □

2.2.3 Proof of Theorem 2.2

We now extend the results of the previous sections to consider the family $\mathscr{F}_{N;\text{sq-free}}$ of all primitive characters whose conductor is an odd square-free integer in $[N, 2N]$. Some of the bounds below can be improved, but as the improvements do not increase the range of convergence, they will only be sketched.

Proof (Proof of Theorem 2.2). First we calculate the number of primitive characters arising from odd square-free numbers $m \in [N, 2N]$. Let $m = m_1 m_2 \cdots m_r$. Then m contributes $(m_1 - 2) \cdots (m_r - 2)$ characters. On average we might expect the number of characters to be of order N, and as a positive percent of numbers are square-free, we expect there to be on the order of N^2 characters.

Instead we prove there are at least $N^2/\log^2 N$ primitive characters in the family; as we are winning by power savings and not logarithms, the $\log^2 N$ factor is harmless. There are at least $N/\log^2 N + 1$ primes in the interval. For each prime p (except possibly the first) we have $p - 2 \geq N$. Hence there are at least $N \cdot \frac{N}{\log^2 N} = N^2/\log^2 N$ primitive characters. Let $M = |\mathscr{F}_{N;\text{sq-free}}|$. Then

$$M \geq N^2 \log^{-2} N \quad \Rightarrow \quad \frac{1}{M} \leq \frac{\log^2 N}{N^2}. \tag{56}$$

We recall the results from the previous section. Fix an odd square-free number $m \in [N, 2N]$, and say m has $r = r(m)$ factors. Before we divided the First and Second Sums by $M_2 = (m_1 - 2) \cdots (m_r - 2)$, as this was the number of primitive characters in our family. Now we divide by M. Hence the contribution to the First and Second Sums from this m is

$$\begin{aligned} S_1(m, r; \phi) &\ll \frac{1}{M} 2^{r(m)} m^{\sigma/2} \\ S_2(m, r; \phi) &\ll \frac{1}{M} 3^{r(m)} m^{1/2}. \end{aligned} \tag{57}$$

Note that $2^{r(m)} = \tau(m)$, the number of divisors of m. While it is possible to prove

$$\sum_{n \leq x} \tau^\ell(n) \ll x(\log x)^{2^\ell - 1} \tag{58}$$

the crude bound

$$\tau(n) \leq c(\epsilon) n^\epsilon \tag{59}$$

yields the same region of convergence. Note $3^{r(m)} \leq \tau^2(m)$. Therefore by Lemma 2.7 the contributions to the first sum are majorized by

$$
\begin{aligned}
\sum_{\substack{m=N \\ m \text{ square-free}}}^{2N} S_1(m,r;\phi) &\ll \sum_{m=N}^{2N} \frac{1}{M} 2^{r(m)} m^{\sigma/2} \\
&\ll \frac{1}{M} N^{\sigma/2} \sum_{m=N}^{2N} \tau(m) \\
&\ll \frac{1}{M} N^{\sigma/2} c(\epsilon) N^{1+\epsilon} \\
&\ll \frac{\log^2 N}{N^2} N^{\sigma/2} c(\epsilon) N^{1+\epsilon} \\
&\ll c(\epsilon) N^{\frac{1}{2}\sigma+\epsilon-1} \log^2 N. \qquad (60)
\end{aligned}
$$

For $\sigma < 2$, choosing $\epsilon < 1 - \frac{1}{2}\sigma$ yields S_1 goes to zero as N tends to infinity. For the second sum, Lemma 2.8 bounds it by

$$
\begin{aligned}
\sum_{\substack{m=N \\ m \text{ square-free}}}^{2N} S_2(m,r;\phi) &\ll \sum_{m=N}^{2N} \frac{1}{M} 3^{r(m)} m^{1/2} \\
&\ll \frac{1}{M} N^{1/2} \sum_{m=N}^{2N} \tau^2(m) \\
&\ll c(\epsilon) \frac{\log^2 N}{N^2} N^{1/2} N^{1+2\epsilon} \\
&\ll c(\epsilon) N^{2\epsilon-\frac{1}{2}} \log^2 N, \qquad (61)
\end{aligned}
$$

which converges to zero as N tends to infinity for all σ and completes the proof. □

2.3 Dirichlet Characters from a Fixed Modulus

We thank the referee for the following theorem and proof, which extends Theorem 2.1 to the family of Dirichlet characters for any fixed modulus.

Theorem 2.9 (Dirichlet Characters from a Fixed Modulus). *Let $\mathscr{F}_m$ denote the family of primitive Dirichlet characters arising from a fixed m, and let $\hat{\phi}$ be an even Schwartz function with supp$(\hat{\phi}) \subset (-2, 2)$. Denote the conductor of χ by $c(\chi)$. Then*

$$
\frac{1}{\phi(m)} \sum_{\substack{\chi(m) \\ \chi \neq \chi_0}} \sum_{\gamma_\chi : L(\frac{1}{2}+i\gamma_\chi,\chi)=0} \phi\left(\gamma_\chi \frac{\log(c(\chi)/\pi)}{2\pi}\right) = \int_{-\infty}^{\infty} \phi(y)dy + O\left(\frac{1}{\log m}\right). \qquad (62)
$$

As $m \to \infty$, the above agrees only with the $m \to \infty$ limit of the 1-level density of $m \times m$ unitary matrices.

Proof. We argue similarly as in the proof of Theorem 2.1. From Eq. (3.8) of [IwKo] we have

$$\sum_{\chi(m)} \chi(p) = \sum_{d|(p-1,m)} \phi(d)\mu(m/d). \tag{63}$$

We can now bound the first prime sum, $S_1(m;\phi)$:

$$\begin{aligned} S_1(m;\phi) &= \frac{1}{\phi(m)} \sum_{d|m} \phi(d)\mu(m/d) \sum_{p\equiv 1(d)} \frac{\log p}{\log(m/\pi)} \hat{\phi}\left(\frac{\log p}{\log(m/\pi)}\right) \frac{2\mathrm{Re}(\chi(p))}{\sqrt{p}} \\ &\ll \frac{m^{\sigma/2}}{\phi(m)} \sum_{d|m} \frac{\phi(d)}{d} \leq \frac{\tau(m)}{\phi(m)} m^{\sigma/2}, \end{aligned} \tag{64}$$

which is $O(1/\log m)$, completing the proof. □

Remark 2.10. We could argue as in the proof of Theorem 2.9, and by applying trekhe Brun–Titchmarsh Theorem extend the support to $[-2, 2]$.

3 Convolutions of Families of *L*-Functions

The analysis of Dirichlet L-functions in Sect. 2 highlights the general framework for determining the behavior of the low-lying zeros in a family and identifying the corresponding symmetry group. In this section we describe how to find the symmetry group of a compound family in terms of its constituent pieces. In order to view these results in the proper context, we first briefly summarize the procedure used in most works to calculate 1-level densities, and refer the reader to [SaShTe] in this volume for a more detailed treatment.

These calculations break down into three steps. The first step is to understand and control conductors. In most families analyzed to date they are either constant, or monotonically increasing. Their importance stems from the fact that their logarithm controls the spacing of zeros near the central point, and constancy or monotonicity allows us to pass sums over the family past the test function to the Fourier coefficients. When these properties fail, the computations are significantly harder. A notable exception is in one-parameter families of elliptic curves over $\mathbb{Q}(T)$, where for $t \in [N, 2N]$ variations in the logarithms of the conductors, from $\log(N^d)$ to $\log(cN^d)$, greatly complicate the analysis and require careful sieving.

The second step is the classic explicit formula, which relates sums of our test function ϕ at the zeros of the L-functions to sums of its Fourier transform $\hat{\phi}$ at the primes (weighted by the coefficients of the L-function). This is very similar to the role the Eigenvalue Trace Lemma plays in random matrix theory. While we wish to understand the eigenvalues of a matrix, it is the matrix elements where we have information; the Eigenvalue Trace Lemma allows us to pass from knowledge of the matrix coefficients (which we have) to knowledge of the eigenvalues (which we desire). The explicit formulas in number theory play a similar role.

The explicit formula is useless, however, unless we have a way to execute the resulting sums. The final step is to use an averaging formula for weighted sums of L-function coefficients. Examples here include the orthogonality relations of Dirichlet characters, the Petersson formula for holomorphic cusp forms, and the Kuznetsov trace formula for Maass forms. Unfortunately, as our family becomes more complicated the averaging formulas become harder to use, and often yield smaller support. This can be seen in comparison of some recent work (such as [GolKon, MaTe, ShTe]).

The goal for the remainder of this section is to discuss how to identify the corresponding symmetry group for a family of L-functions, and to discuss the role the Fourier coefficients play in the rate of convergence of the 1-level density to the scaling limits of ensembles from the classical compact groups.

3.1 Identifying the Symmetry Group of a Family

Determining the corresponding symmetry group for a family of L-functions is one of the hardest questions in the subject. In many cases we cannot compute the 1-level density for large enough support to distinguish between the three orthogonal candidates (though we can uniquely determine which works by looking at the 2-level density). In many situations we are able to argue by analogy with a function field analogue, where the situation is clearer and the answer arises from the monodromy group. Another approach is to work with the Sato–Tate measure of the family as carried out in [ShTe].

A folklore conjecture stated that the symmetry was determined by the sign of the functional equations. For example, if all the signs were odd, then the family had to have SO(odd) symmetry. If the signs are all even, then there are two candidates: Symplectic and SO(even). Initially many thought that SO(even) symmetry happened when there was a corresponding family with odds signs that was being ignored (for example, splitting the family of weight k and level $N > 1$ cuspidal newforms by sign and ignoring the forms with odd sign), and that if there were no corresponding family with odd signs then the symmetry would be Symplectic. Dueñez and Miller [DuMil1] disproved this conjecture by analyzing a family suggested by Sarnak: $\{L(s, \phi \times \mathrm{sym}^2 f) : f \in H_k\}$, where ϕ is a fixed even Hecke–Maass cusp form and H_k is a Hecke eigenbasis for the space of holomorphic cusp forms of weight k for the full modular group. Their proof involved finding the symmetry group of a Rankin–Selberg convolution in terms of the symmetry groups of the constituents. They generalized their argument to many families in [DuMil2]. We quickly sketch the main ideas of that argument, and then conclude this section with an interpretation of convergence to the limiting densities in the spirit of the Central Limit Theorem.

We first need some standard notation and results.

- π: A cuspidal automorphic representation on GL_n.
- $Q_\pi > 0$: The analytic conductor of $L(s,\pi) = \sum \lambda_\pi(n)/n^s$.
- By GRH[5] the non-trivial zeros are $\frac{1}{2} + i\gamma_{\pi,j}$.
- $\{\alpha_{\pi,i}(p)\}_{i=1}^n$: The Satake parameters, and $\lambda_\pi(p^\nu) = \sum_{i=1}^n \alpha_{\pi,i}(p)^\nu$. Thus the p^ν-th coefficient of $L(s,\pi)$ is the ν-th moment of the Satake parameters.
- $L(s,\pi) = \sum_n \frac{\lambda_\pi(n)}{n^s} = \prod_p \prod_{i=1}^n (1 - \alpha_{\pi,i}(p)p^{-s})^{-1}$.

The explicit formula, applied to a given $L(s,\pi)$, yields

$$\sum_j g\left(\gamma_{\pi,j}\frac{\log Q_\pi}{2\pi}\right) = \hat{g}(0) - 2\sum_p \sum_{\nu=1}^{\infty} \hat{g}\left(\frac{\nu \log p}{\log Q_\pi}\right) \frac{\lambda_\pi(p^\nu)\log p}{p^{\nu/2}\log Q_\pi}. \tag{65}$$

For ease of exposition, we assume the conductors in our family are constant,[6] and thus $Q_\pi = Q$ say. Thus in calculating the 1-level density we can push the sum over our family $\mathscr{F}_N$ through the test function; here, $\mathscr{F}_N$ are all forms in our infinite family $\mathscr{F}$ with some restriction involving N on the conductor (frequent choices are the conductor equals N, lives in an interval $[N, 2N]$, or is at most N). The 1-level density is then found by taking the limit as $N \to \infty$. We rescale the zeros by $\log R$, where R is closely related to Q (it sometimes differs by a fixed, multiplicative constant; this extra flexibility simplifies some of the resulting expressions for various families).

We also assume sufficient decay in the $\lambda_\pi(p^\nu)$'s so that the sum over primes with $n \geq 3$ converges; this is known for many families. Determining the 1-level density, up to lower order terms which we will return to later, is equivalent to analyzing the $N \to \infty$ limits of

$$\begin{aligned} S_1(\mathscr{F}_N) &:= -2\sum_p \hat{g}\left(\frac{\log p}{\log R}\right)\frac{\log p}{\sqrt{p}\log R}\left[\frac{1}{|\mathscr{F}_N|}\sum_{\pi\in\mathscr{F}_N}\lambda_\pi(p)\right] \\ S_2(\mathscr{F}_N) &:= -2\sum_p \hat{g}\left(2\frac{\log p}{\log R}\right)\frac{\log p}{p\log R}\left[\frac{1}{|\mathscr{F}_N|}\sum_{\pi\in\mathscr{F}_N}\lambda_\pi(p^2)\right]. \end{aligned} \tag{66}$$

As

$$\lambda_\pi(p^\nu) = \alpha_{\pi,1}(p)^\nu + \cdots + \alpha_{\pi,n}(p)^\nu, \tag{67}$$

[5]The definition of the 1-level density as a sum of a test function at scaled zeros is well defined even if GRH fails; however, in that case the zeros are no longer on a line and we thus lose the ability to talk about spacings between zeros. Thus in many of the arguments in the subject GRH is only used to interpret the quantities studied, though there are exceptions (in [ILS] the authors use GRH for Dirichlet L-functions to expand Kloosterman sums).

[6]It is easy to handle the case where the conductors are monotone by rescaling the zeros by the average log-conductor; as remarked many times above the general case is more involved.

we see that only the first two moments of the Satake parameters enter the calculation. The sum over the remaining powers,

$$S_\nu(\mathscr{F}_N) \ := \ -2\sum_{\nu=3}^{\infty}\sum_p \hat{g}\left(\nu\frac{\log p}{\log R}\right)\frac{\log p}{p^{\nu/2}\log R}\left[\frac{1}{|\mathscr{F}_N|}\sum_{\pi\in\mathscr{F}_N}\lambda_\pi(p^\nu)\right], \tag{68}$$

is $O(1/\log R)$ under the Ramanujan Conjectures.[7]

To date, the only families where the first sum $S_1(\mathscr{F}_N)$ is not negligible are elliptic curve families with rank. The presence of non-zero terms here require trivial modifications to the classical random matrix ensembles, and effectively in the limit only result in additional independent zeros at the central point. Thus, if the family has rank r, the scaling limit is that of a block diagonal matrix, with an $r\times r$ identity matrix in the upper left, and then an $(N-r)\times(N-r)$ matrix in the lower right (with the other two rectangular blocks zero).

We introduce a symmetry constant for the family, $c_{\mathscr{F}}$, as follows:

$$c_{\mathscr{F}} \ := \ \lim_{N\to\infty}\frac{1}{|\mathscr{F}_N|}\sum_{\pi\in\mathscr{F}_N}\lambda_\pi(p^2), \tag{69}$$

which is the limit of the average second moment of the Satake parameters. The corresponding classical compact group is Unitary if $c_{\mathscr{F}}$ is 0, Symplectic if $c_{\mathscr{F}} = 1$, and Orthogonal if $c_{\mathscr{F}} = -1$. Equivalently, $c_{\mathscr{F}} = 0$ (respectively, 1 or -1) if the family $\mathscr{F}$ has Unitary (respectively, Symplectic or Orthogonal) symmetry.

3.2 *Identifying the Symmetry Group from Rankin–Selberg Convolutions*

In this section we assume we have two families of L-functions where we can determine the corresponding symmetry group. Under standard assumptions (which are proven in many cases), the Rankin–Selberg convolution exists and it makes sense to talk about the symmetry group of the family. We assume for simplicity below that π_2 is not the representation contragredient to π_1, and thus the L-function below will not have a pole, though with more book-keeping this case can readily be handled. The Satake parameters of the convolution $\pi_{1,p}\times\pi_{2,p}$ are

$$\{\alpha_{\pi_1\times\pi_2,k}(p)\}_{k=1}^{nm} = \{\alpha_{\pi_1,i}(p)\cdot\alpha_{\pi_2,j}(p)\}_{\substack{1\le i\le n\\ 1\le j\le m}}. \tag{70}$$

[7] The Satake parameters $|\alpha_{\pi,i}|$ are bounded by p^δ for some δ, and it is conjectured that we may take $\delta = 0$. While this conjecture is open in general, for many forms there is significant progress towards these bounds with some $\delta < 1/2$. See, for example, recent work of Kim and Sarnak [Kim, KimSa]. For our purposes, we only need to be able to take $\delta < 1/6$, as such an estimate and trivial bounding suffices to show that the sum over all primes and all $\nu \ge 3$ is $O(1/\log R)$.

The main result is that the symmetry of the new compound family is beautifully and simply related to the symmetry of the constituent pieces. See [DuMil2] for a statement of which families are nice (examples include Dirichlet L-functions and GL_2 families).

Theorem 3.1 (Dueñez–Miller [DuMil2]). *If $\mathscr{F}$ and $\mathscr{G}$ are nice families of L-functions, then $c_{\mathscr{F}\times\mathscr{G}} = c_{\mathscr{F}} \cdot c_{\mathscr{G}}$.*

Proof (Sketch of the proof). From (70), we find that the moments of the Satake parameters for $\pi_{1,p} \times \pi_{2,p}$ are

$$\sum_{k=1}^{nm} \alpha_{\pi_1\times\pi_2,k}(p)^\nu = \sum_{i=1}^{n} \alpha_{\pi_1,i}(p)^\nu \sum_{j=1}^{m} \alpha_{\pi_2,j}(p)^\nu. \tag{71}$$

Thus, if $\pi_1 \in \mathscr{G}_N$ and $\pi_2 \in \mathscr{G}_M$, we find

$$\begin{aligned} c_{\mathscr{F}\times\mathscr{G}} &= \lim_{N,M\to\infty} \frac{1}{|\mathscr{F}_N|\,|\mathscr{G}_M|} \sum_{\substack{\pi_1\in\mathscr{F}_N\\ \pi_2\in\mathscr{G}_M}} \lambda_{\pi_1\times\pi_2}(p^2) \\ &= \lim_{N,M\to\infty} \frac{1}{|\mathscr{F}_N|} \sum_{\pi_1\in\mathscr{F}_N} \lambda_{\pi_1}(p^2) \frac{1}{|\mathscr{G}_M|} \sum_{\pi_2\in\mathscr{G}_N} \lambda_{\pi_2}(p^2) = c_{\mathscr{F}} c_{\mathscr{G}}. \end{aligned} \tag{72}$$

The first sum is handled similarly, and the higher moments do not contribute by assumption on the family (the definition of a good family includes sufficient bounds towards the Ramanujan conjecture to handle the $\nu \geq 3$ terms).

3.3 Connections to the Central Limit Theorem

We end this section by interpreting our results in the spirit of the Central Limit Theorem, which we hope will shed some light on the universality of results.

Interestingly, random matrix theory does not seem to know about arithmetic. By this we mean that very different families of L-functions converge to one of the five flavors (unitary, symplectic, or one of the three orthogonals), independent of the arithmetic structure of the family. It doesn't matter if we have quadratic Dirichlet characters or the symmetric square of GL_2-forms; we see symplectic behavior. Similarly it doesn't matter if our family of elliptic curves have complex multiplication or not, or instead are holomorphic cusp forms of weight k or Maass forms; we see orthogonal behavior.[8]

[8]There are some situations where arithmetic enters. The standard example is that in estimating moments of L-functions one has a product $a_k g_k$, where a_k is an arithmetic factor coming from

One of the first places this universality was noticed was in the work of Rudnick and Sarnak [RudSa], who showed for suitable test functions that the n-level correlations of zeros arising from a fixed cuspidal automorphic representation agreed with the Gaussian Unitary Ensemble. The cause of their universality was that the answer was governed by the first and second moments of the Fourier coefficients, and explained why the behavior of zeros far from the central point was the same for all L-functions.

We have a similar explanation for the behavior of the zeros near the central point. Our universality is due to the fact that the main term of the limiting behavior depends only on the first two moments of the Satake parameters, which to date have very few possibilities. The effect of the higher moments are felt only in the $\nu \geq 3$ terms, which (under the Generalized Ramanujan Conjectures) contribute $O(1/\log R)$. While these contributions vanish in the limit, they can be felt in *how* the limiting density is approached.

Notice how similar this is to the Central Limit Theorem, which in its simplest form states that the normalized sum of independent random variables drawn from the same nice distribution (finite moments suffice) converges to being normally distributed. If the mean μ and the variance σ^2 of a random variable X are finite, we can always study instead the standardized random variable $Z = (X - \mu)/\sigma$, which has mean 0 and variance 1. Thus the first 'free' moment of our density is the third (or fourth if the distribution is symmetric). A standard proof is to look at the Fourier transform of the N-fold convolution, Taylor expand, and then show that the inverse Fourier transform converges to the Gaussian. The higher moments emerge only in the error terms, and while they have no contribution as $N \to \infty$ they do affect the rate in which the density of the convolution approaches the Gaussian.

Thus, for families of L-functions the higher moments of the Satake parameters help control the convergence to random matrix theory, and can depend on the arithmetic of family. This leads to the exciting possibility of isolating lower order terms in 1-level densities, and seeing the arithmetic of the family emerge.

Unfortunately, it is often very hard to isolate these lower order terms from other errors. For example, Dueñez and Miller [DuMil2] convolve two families of elliptic curves with ranks r_1 and r_2, and see a potential lower order term of size $r_1 r_2$ divided by the logarithm of the conductor. Thus, while this looks like a lower order term which is highly dependent on the arithmetic of the family, there are other error terms which can only be bounded by larger quantities (though we believe these bounds are far from optimal and that this product term should be larger in the limit). We discuss some of these issues in more detail in the concluding section.

the arithmetic of the form and g_k arises from random matrix theory. See, for example, [CFKRS, KeSn1, KeSn2].

4 Lower Order Terms and Rates of Convergence

In this section we discuss some work (see [Mil3, Mil6]) on lower order terms in families of elliptic curves, though similar results can be done for other families (especially families of Dirichlet L-functions [FioMil] or cusp forms [MilMo]). We first report on some families where these lower order terms have been successfully isolated (which is different than the example from convolving two families with rank from Sect. 3.3), and end with some current research about finer properties of the distribution of the Satake parameters in families of elliptic curves and lower order terms.

4.1 Arithmetic-Dependent Lower Order Terms in Elliptic Curve Families

The results below are from [Mil6], where many families of elliptic curves were studied. For families of elliptic curves, it is significantly easier to calculate and work with $\lambda_E(p)$ (which is an integer and computable via sums of Legendre symbols) then the Satake parameters $\alpha_{E,1}(p)$ and $\alpha_{E,2}(p)$. We thus first re-express the formula for the 1-level density to involve sums over the λ_E's, and then give several families with lower order terms depending on the arithmetic.

It is often convenient to study weighted moments (for example, in [ILS] much work is required to remove the harmonic weights, which facilitated applications of the Petersson formula). For a family $\mathscr{F}$ and a weight function w define

$$\begin{aligned} A_{r,\mathscr{F}}(p) &:= \frac{1}{W_R(\mathscr{F})} \sum_{\substack{f\in\mathscr{F}\\ f\in S(p)}} w_R(f)\lambda_f(p)^r \\ A'_{r,\mathscr{F}}(p) &:= \frac{1}{W_R(\mathscr{F})} \sum_{\substack{f\in\mathscr{F}\\ f\notin S(p)}} w_R(f)\lambda_f(p)^r \\ S(p) &:= \{f\in\mathscr{F} : p \nmid C_f\}, \end{aligned} \tag{73}$$

where C_f is the conductor of f (when doing the computations, there are sometimes differences at primes dividing the conductor, and it is worth isolating their contribution). The main difficulty in determining the 1-level density is evaluating

$$S(\mathscr{F}) := -2\sum_p \sum_{m=1}^{\infty} \frac{1}{W_R(\mathscr{F})} \sum_{f\in\mathscr{F}} w_R(f) \frac{\alpha_{f,1}(p)^m + \alpha_{f,2}(p)^m}{p^{m/2}} \frac{\log p}{\log R}\, \hat{\phi}\left(m\frac{\log p}{\log R}\right), \tag{74}$$

where we are assuming we have GL_2 forms.

The following alternative expansion for the explicit formula from [Mil6] is especially tractable for families of elliptic curves:

$$\begin{aligned}
S(\mathscr{F}) = &-2\sum_p \sum_{m=1}^{\infty} \frac{A'_{m,\mathscr{F}}(p)}{p^{m/2}} \frac{\log p}{\log R}\, \hat{\phi}\left(m\frac{\log p}{\log R}\right) \\
&-2\hat{\phi}(0)\sum_p \frac{2A_{0,\mathscr{F}}(p)\log p}{p(p+1)\log R} + 2\sum_p \frac{2A_{0,\mathscr{F}}(p)\log p}{p\log R}\, \hat{\phi}\left(2\frac{\log p}{\log R}\right) \\
&-2\sum_p \frac{A_{1,\mathscr{F}}(p)}{p^{1/2}} \frac{\log p}{\log R}\, \hat{\phi}\left(\frac{\log p}{\log R}\right) + 2\hat{\phi}(0)\frac{A_{1,\mathscr{F}}(p)(3p+1)}{p^{1/2}(p+1)^2}\frac{\log p}{\log R} \\
&-2\sum_p \frac{A_{2,\mathscr{F}}(p)\log p}{p\log R}\, \hat{\phi}\left(2\frac{\log p}{\log R}\right) + 2\hat{\phi}(0)\sum_p \frac{A_{2,\mathscr{F}}(p)(4p^2+3p+1)\log p}{p(p+1)^3\log R} \\
&-2\hat{\phi}(0)\sum_p \sum_{r=3}^{\infty} \frac{A_{r,\mathscr{F}}(p)p^{r/2}(p-1)\log p}{(p+1)^{r+1}\log R} + O\left(\frac{1}{\log^3 R}\right) \\
= &\, S_{A'}(\mathscr{F}) + S_0(\mathscr{F}) + S_1(\mathscr{F}) + S_2(\mathscr{F}) + S_A(\mathscr{F}) + O\left(\frac{1}{\log^3 R}\right).
\end{aligned} \tag{75}$$

Letting $\widetilde{A}_{\mathscr{F}}(p) := \frac{1}{W_R(\mathscr{F})}\sum_{f\in S(p)} w_R(f)\frac{\lambda_f(p)^3}{p+1-\lambda_f(p)\sqrt{p}}$, by the geometric series formula we may replace $S_A(\mathscr{F})$ with $S_{\widetilde{A}}(\mathscr{F})$, where

$$S_{\widetilde{A}}(\mathscr{F}) = -2\hat{\phi}(0)\sum_p \frac{\widetilde{A}_{\mathscr{F}}(p)p^{3/2}(p-1)\log p}{(p+1)^3\log R}. \tag{76}$$

We now state some results (see [Mil6] for the proofs). For comparison purposes we start with the family of cuspidal newforms, as this family is significantly easier to calculate and serves as a good baseline. In reading the formulas below, it is important to note that the contributions from the smaller primes are significantly more than those from the larger primes. For elliptic curves the primes 2 and 3 often behave differently; while they will have no affect on the main term, they will strongly influence the lower order terms.

In the subsections below, we assume the logarithms of the conductors are of size $\log R$, so that we are comparing zeros of similar size. In all families of elliptic curves we start with an elliptic curve over $\mathbb{Q}(T)$, and then form a one-parameter family by looking at the specializations from setting T equal to integers t.

4.1.1 $\mathscr{F}_{k,N}$ the Family of Even Weight k and Prime Level N Cuspidal Newforms, or Just the Forms with Even (or Odd) Functional Equation

Up to $O(\log^{-3} R)$, as $N \to \infty$ for test functions ϕ with $\text{supp}(\hat{\phi}) \subset (-4/3, 4/3)$ the (non-conductor) lower order term for either of these families is

$$C \cdot 2\hat{\phi}(0)/\log R, \tag{77}$$

with $C \approx -1.33258$. In other words, the difference between the Katz–Sarnak prediction and the 1-level density has a lower order term of order $1/\log R$, with the next correction $O(1/\log^3)$. Note the lower order corrections are independent of the distribution of the signs of the functional equations, and the weight k.

4.1.2 CM Example, with or Without Forced Torsion: Specializations of $y^2 = x^3 + B(6T+1)^\kappa$ Over $\mathbb{Q}(T)$, with $B \in \{1,2,3,6\}$ and $\kappa \in \{1,2\}$

This family of elliptic curves has complex multiplication. We consider the sub-family obtained by sieving and restricting T so that $(6T+1)$ is $(6/\kappa)$-power free. If $\kappa = 1$, then all values of B give the same result, while if $\kappa = 2$ then the four values of B have different lower order corrections. Note if $\kappa = 2$ and $B = 1$ then there is a forced torsion point of order three, $(0, 6T+1)$.

Up to errors of size $O(\log^{-3} R)$, the (non-conductor) lower order terms are again of size $C \cdot 2\hat{\phi}(0)/\log R$; we give numerical approximations for the C's for various choices of B and κ:

$$\begin{aligned}
B = 1, \kappa = 1: &\quad -2.124 \cdot 2\hat{\phi}(0)/\log R,\\
B = 1, \kappa = 2: &\quad -2.201 \cdot 2\hat{\phi}(0)/\log R,\\
B = 2, \kappa = 2: &\quad -2.347 \cdot 2\hat{\phi}(0)/\log R\\
B = 3, \kappa = 2: &\quad -1.921 \cdot 2\hat{\phi}(0)/\log R\\
B = 6, \kappa = 2: &\quad -2.042 \cdot 2\hat{\phi}(0)/\log R.
\end{aligned} \tag{78}$$

4.1.3 CM Example, with or Without Rank: Specializations of $y^2 = x^3 - B(36T+6)(36T+5)x$ Over $\mathbb{Q}(T)$, with $B \in \{1,2\}$

We consider another complex multiplication family. If $B = 1$, the family has rank 1 over $\mathbb{Q}(T)$, while if $B = 2$, the family has rank 0. We consider the sub-family obtained by sieving to $(36T+6)(36T+5)$ is cube-free. Again we find a lower order term of size $C \cdot 2\hat{\phi}(0)/\log R$, with next term of size $O(1/\log^3 R)$. The most important difference between these two families is the contribution from the $S_{\widetilde{\mathscr{A}}}(\mathscr{F})$

terms, where the $B = 1$ family is approximately $-0.11 \cdot 2\hat{\phi}(0)/\log R$, while the $B = 2$ family is approximately $0.63 \cdot 2\hat{\phi}(0)/\log R$. This large difference is due to biases of size $-r$ in the Fourier coefficients $a_t(p)$ in a one-parameter family of rank r over $\mathbb{Q}(T)$.

The main term of the average moments of the pth Fourier coefficients are given by the complex multiplication analogue of Sato–Tate in the limit, for each p there are lower order correction terms which depend on the rank.

4.1.4 Non-CM Example: Specializations of $y^2 = x^3 - 3x + 12T$ Over $\mathbb{Q}(T)$

Up to $O(\log^{-3} R)$, the (non-conductor) lower order correction for this family is $C \cdot 2\hat{\phi}(0)/\log R$, where $C \approx -2.703$. Note this answer is very different than the family of weight 2 cuspidal newforms of prime level N.

4.2 Second Moment Bias in One-Parameter Families of Elliptic Curves

In Sect. 4.1 we saw lower order terms to the 1-level density for families of elliptic curves which depended on the arithmetic of the family. In this section we report on work on progress on possible family-dependent lower order terms to the second moment of the Fourier coefficients in families of elliptic curve L-functions; see [MMRW] for a more complete investigation of these families, and Appendix for some initial results on other families. We then conclude in Sect. 4.3 by exploring the implications such a bias would have on low-lying zeros (in particular, in understanding the excess rank phenomenon).

We have observed an interesting property in the average second moments of the Fourier coefficients of elliptic curve L-functions over $\mathbb{Q}(T)$. Specifically, consider an elliptic curve $\mathcal{E} : y^2 = x^3 + A(T)x + B(T)$ over $\mathbb{Q}(T)$, where $A(T), B(T)$ are polynomials in $\mathbb{Z}[T]$ and the curve E_t (obtained by specializing T to t) has coefficient $a_t(p)$ (of size $2\sqrt{p}$) in the series expansion of its L-function. Define the average second moment $A_2(p)$ for the family by

$$A_2(p) := \frac{1}{p} \sum_{t \bmod p} a_t(p)^2 \tag{79}$$

(where for notational convenience we are suppressing the subscript $\mathcal{E}$ on A_2, as the family is fixed). Michel [Mic] proved that

$$A_2(p) = p^2 + O(p^{3/2}) \tag{80}$$

for families of elliptic curves with non-constant j-invariant $j(T)$, and cohomological arguments show that the lower-order terms[9] are of sizes $p^{3/2}, p, p^{1/2}$, and 1. In every case that we have proven or numerically analyzed, the following conjecture holds.

Conjecture 4.1 (Bias Conjecture). *For any family of elliptic curves $\mathscr{E}$ over $\mathbb{Q}(T)$, the largest lower order term in the second moment of $\mathscr{E}$ which does not average to* 0 *is on average negative. Explicitly, from Michel [Mic] we have*

$$A_2(p) = p^2 + \beta_{3/2}(p)p^{3/2} + \beta_1(p)p + \beta_{1/2}p^{1/2} + \beta_0(p) \tag{81}$$

where each $\beta_r(p)$ is of order 1; when we write the second moment thusly the first $\beta_r(p)$ term which does not average to zero will average to a negative value.

Below, we give several proven cases of the Bias Conjecture and some preliminary numerical evidence supporting the conjecture. We have made several additional observations about the terms in the second moments, though we do not know if these always hold.

- In families with constant j-invariant, the largest term is on average p^2 (rather than exactly p^2), and the Bias Conjecture appears to hold similarly.
- Every explicit second moment expression has a non-zero $p^{3/2}$ term or a non-zero p term (or both). The term of size $p^{3/2}$ always averages to 0, and the term of size p is always on average negative.
- In many cases the terms of size $p^{3/2}$ and/or p are governed by the values of an elliptic curve coefficient, that is, a sum of the form

$$\sum_{x \bmod p} \left(\frac{ax^3 + bx^2 + cx + d}{p}\right), \tag{82}$$

possibly squared, cubed, or multiplied by p, et cetera.

Rosen and Silverman [RoSi] proved that the negative bias in the first moments is related to the rank of family by

$$\lim_{X\to\infty} \frac{1}{X} \sum_{p\le X} A_1(p)\frac{\log p}{p} = \operatorname{rank}\mathscr{E}(\mathbb{Q}(T)). \tag{83}$$

It is natural to ask whether the bias in the second moments is also related to the family rank. We are currently investigating this. More generally, we can ask if higher moments are also biased and if this bias is also related to the rank of the family.

[9]These bounds cannot be improved, as Miller [Mil3] found a family where there is a term of size $p^{3/2}$.

4.2.1 Evidence: Explicit Formulas

We have proven the conjecture for a variety of specific families and some restricted cases, and list a few of these below; these are a representative subset of families we have successfully studied, and we are currently investigating many more. The average bias refers to the average value of the coefficient of the largest lower order term not averaging to 0 (which in all of our cases is the p term).

Lemma 4.2. *Consider elliptic curve families of the form $y^2 = ax^3 + bx^2 + cx + d + eT$. These families have rank 0 over $\mathbb{Q}(T)$, and for primes $p > 3$ with $p \nmid a, e$ and $p \nmid b^2 - 3ac$,*

$$A_2(p) = p^2 - p\left(1 + \left(\frac{b^2 - 3ac}{p}\right) + \left(\frac{-3}{p}\right)\right). \tag{84}$$

These families obey the Bias Conjecture with an average bias of -1 in the p term.

Lemma 4.3. *Consider families of the form $y^2 = ax^3 + bx^2 + (cT + d)x$. These families have rank 0, and for primes $p > 3$ with $p \nmid a, b, c$,*

$$A_2(p) = p^2 - p\left(1 + \left(\frac{-1}{p}\right)\right). \tag{85}$$

These families obey the Bias Conjecture with an average bias of -1 in the p term.

Lemma 4.4. *Consider families of the form $y^2 = x^3 + T^n x$. These families have rank 0, and for primes $p > 3$,*

$$A_2(p) = \begin{cases} (p-1)\left(\sum_{x(p)}\left(\frac{x^3+x}{p}\right)\right)^2 & \text{if } n \equiv 0(2) \\ (p^2 - p)\left(1 + \left(\frac{-1}{p}\right)\right) & \text{if } n \equiv 1(2). \end{cases} \tag{86}$$

These families obey the Bias Conjecture with an average bias of $-4/3$ for $n \equiv 0(2)$ and -1 for $n \equiv 1(2)$ in the p term.

Lemma 4.5. *Consider families of the form $y^2 = x^3 + T^n$. These families have rank 0, and for primes $p > 3$,*

$$A_2(p) = \begin{cases} (p-1)\left(\sum_{x(p)}\left(\frac{x^3+1}{p}\right)\right)^2 & \text{if } n \equiv 0(3) \\ p^2 - p\left(1 + \left(\frac{-3}{p}\right)\right) & \text{if } n \equiv 1(3) \\ p^2 - p & \text{if } n \equiv 2(3). \end{cases} \tag{87}$$

These families obey the Bias Conjecture with an average bias of $-4/3$ for $n \equiv 0(3)$ and -1 for $n \equiv 1, 2(3)$ in the p term.

Lemma 4.6. *Consider families of the form $y^2 = x^3 + Tx^2 + (mt - 3m^2)x - m^3$ for m a non-zero integer. These families have rank* 0 *for m non-square and rank* 1 *for m a square, and for primes $p > 3$,*

$$A_2(p) = p^2 - p\left(2 + 2\left(\frac{-3}{p}\right)\right) - 1. \tag{88}$$

These families obey the Bias Conjecture with an average bias of -2.

Lemmas 4.2 and 4.3 prove the Bias Conjecture for a large number of families studied by Fermigier in [Fe2]. A more systematic study of Fermigier's families (which is in progress [MMRW]) will help determine whether the bias in second moments is correlated to the family rank. Lemmas 4.4 and 4.5 provide examples of complex-multiplication families where the Bias Conjecture holds. Lemma 4.6 proves the conjecture for a family with an unusual distribution of signs, providing stronger evidence for the conjecture.

4.2.2 Numerical Data

The following lemma is useful for analyzing Fermigier's rank 1 families [Fe2].

Lemma 4.7. *Consider families of the form $y^2 = ax^3 + cx^2 + (dT + e)x + g$. For $p \nmid d, g$,*

$$A_2(p) = p^2 + pc_1(p) - pc_0(p), \tag{89}$$

where $c_0(p)$ is the number of roots of the congruence $2ax^3 + cx^2 - g \equiv 0(p)$ and $c_1(p) = \sum_{x,y: axy^2+(ax^2+cx)y-g\equiv 0(p)} \left(\frac{xy}{p}\right)$.

We are not able to explicitly determine the $c_1(p)$ term in general, but the data in Table 1 suggests that on average this term is 0. We averaged these coefficients over the 6000th to the 7000th primes, and all averages are very small in absolute value. Thus, we believe that these families obey the Bias Conjecture with an average bias of $c_0(p)$, which in most cases is about 1. We collected additional data on rank 2 families, and found similar evidence from these families that the $p^{3/2}$ term coefficient is on average 0.

We also collected numerical data for several families that were too complicated to analyze explicitly. We used two averaging statistics,

$$\mathbb{E}_p\left(\frac{A_2(p) - p^2}{p^{3/2}}\right), \qquad \mathbb{E}_p\left(\frac{A_2(p) - p^2}{p}\right), \tag{90}$$

where the averages are taken over some range of primes. These statistics are meant to quantify the average bias in the cases where the largest lower term is of size $p^{3/2}$ and p, respectively. For these families, we calculated the second moment for

Table 1 Averages of $p^{3/2}$ term coefficients in rank 1 families

Family	Average($c_1(p)$)	Average($c_0(p)$)
$y^2 = 4x^3 - 7x^2 + 4tx + 4$	0.0068	0.974
$y^2 = 4x^3 + 5x^2 + (4t - 2)x + 1$	−0.0176	1.005
$y^2 = 4x^3 + 5x^2 + (4t + 2)x + 1$	−0.0174	1.005
$y^2 = 4x^3 + x^2 + (4t + 2)x + 1$	0.0399	0.993
$y^2 = 4x^3 + x^2 + 4tx + 4$	0.0068	0.985
$y^2 = 4x^3 + x^2 + (4t + 6)x + 9$	−0.0113	1.988
$y^2 = 4x^3 + 4x^2 + 4tx + 1$	0.0072	0.974
$y^2 = 4x^3 + 5x^2 + (4t + 4)x + 4$	0.0035	1.012
$y^2 = 4x^3 + 4x^2 + 4tx + 9$	0.0256	1.005
$y^2 = 4x^3 + 5x^2 + 4tx + 4$	0.0043	1.005
$y^2 = 4x^3 + 5x^2 + (4t + 6)x + 9$	−0.0143	1.037

the 100th–150th primes. In every case, the running $p^{3/2}$-normalized average was small in magnitude, further suggesting that the $p^{3/2}$ term coefficient is on average 0. In most families, the p-normalized statistic revealed a clear negative average bias, but two families showed a positive p-normalized average bias. The problem behind these statistics is the rate of decay of the $p^{3/2}$ term. In order for these statistics to reliably detect an average bias, the average coefficient of the $p^{3/2}$ term would need to exhibit enough cancelation that in the limit it would be smaller than the conjectured bias coming from the lower order terms. This is only a heuristic, but it suggests that we need to improve this method of analyzing general families. The positive average families were positive overall but had a negative average on the second half of the primes. However, here we feel as though we are trying to force out a negative average. For several families that support the conjecture, we tried averaging only over the second half of our sample to see if the bias was still negative in this reduced sample, and it was in each case.

In the last section we discuss connections of the negative bias with excess rank. It is important to note, however, that it is the smallest primes that contribute the most. Thus while there may be a negative bias overall, at the end of the day what might matter most is what occurs for the primes 2 and 3 (and other small primes).

4.3 Biases and Excess Rank

We end by very briefly discussing an application of the conjectured negative bias in the second moments to the observed excess rank in families. For more details, see [Mil3]. The purpose of this section is to show how the arithmetic in lower order terms can be used as a possible explanation for some interesting phenomena. The 1-level density, with an appropriate test function, is used to obtain upper bounds for the average rank; there were several papers using essentially the 1-level density for this purpose before Katz and Sarnak isolated the 1-level density as a statistic to study independent of rank estimation. We show that lower order terms arising from arithmetic contribute for finite conductors and require a very slight change in the upper bound of the average rank. Of course, this is not a proof of a connection between these factors and the average rank, as all we can show is that these affect the upper bound; however, it is worth noting the role they play in such calculations. For more on finite models and the behavior of elliptic curve zeros, see [DHKMS1, DHKMS2].

For a one-parameter family of elliptic curves $\mathscr{E}$ of rank r over $\mathbb{Q}(T)$, assuming the Birch and Swinnerton-Dyer conjecture by Silverman's specialization theorem eventually all curves E_t have rank at least r, and under natural standard conjectures a typical family will have equidistribution of signs of the functional equations. The minimalist conjecture on rank suggests that in the limit half should have rank r and half rank $r+1$, giving an average rank of $r+1/2$; however, in many families this is not observed. Instead, roughly 30 % have rank r and 20 % rank $r+2$, while about 48 % have rank $r+1$ and 2 % rank $r+3$. The question is whether or not the average rank stays on the order of $r+\frac{1}{2}+0.40$ (or anything larger than $r+1/2$), or if this is a result of small conductors and the limiting behavior not being seen. See [Fe1, Fe2, Wa] for numerical investigations and [BhSh1, BhSh2, Br, H-B, FoPo, Mic, Sil, Yo2] for theoretical bounds of the average rank.

Consider families where the average second moment of $a_t(p)^2$ is $p^2 - m_{\mathscr{E}}p + O(1)$ with $m_{\mathscr{E}} > 0$, and let $t \in [N, 2N]$ for simplicity. We have already handled the contribution from p^2 to the 1-level density, and the $-m_{\mathscr{E}}p$ term contributes

$$\begin{aligned} S_2 &\sim \frac{-2}{N}\sum_p \frac{\log p}{\log R}\hat{\phi}\left(2\frac{\log p}{\log R}\right)\frac{1}{p^2}\frac{N}{p}(-m_{\mathscr{E}}p) \\ &= \frac{2m_{\mathscr{E}}}{\log R}\sum_p \hat{\phi}\left(2\frac{\log p}{\log R}\right)\frac{\log p}{p^2}. \end{aligned} \tag{91}$$

Thus there is a contribution of size $\frac{1}{\log R}$. A good choice of test functions (see Appendix A of [ILS], or [FrMil] for optimal test functions for all classical compact groups and larger support) is the Fourier pair

$$\phi(x) = \frac{\sin^2(2\pi\frac{\sigma}{2}x)}{(2\pi x)^2}, \quad \hat{\phi}(u) = \begin{cases} \frac{\sigma-|u|}{4} & \text{if } |u| \le \sigma \\ 0 & \text{otherwise.} \end{cases} \tag{92}$$

Note $\phi(0) = \frac{\sigma^2}{4}$, $\hat{\phi}(0) = \frac{\sigma}{4} = \frac{\phi(0)}{\sigma}$, and evaluating the prime sum in (91) gives

$$S_2 \sim \left(\frac{0.986}{\sigma} - \frac{2.966}{\sigma^2 \log R}\right)\frac{m_{\mathscr{E}}}{\log R}\,\phi(0). \tag{93}$$

While we expect any σ to hold, in all theoretical work to date σ is greatly restricted. In [Mil3] the consequences of this are analyzed in detail for various values of σ. If $\sigma = 1$ and $m_{\mathscr{E}} = 1$, then the $1/\sigma$ term would contribute 1, the lower correction would contribute 0.03 for conductors of size 10^{12}, and thus the average rank is bounded by $1 + r + \frac{1}{2} + 0.03 = r + \frac{1}{2} + 1.03$. This is significantly higher than Fermigier's observed $r + \frac{1}{2} + 0.40$. If we were able to prove our 1-level density for $\sigma = 2$, then the $1/\sigma$ term would contribute $1/2$, and the lower order correction would contribute 0.02 for conductors of size 10^{12}. Thus the average rank would be bounded by $1/2 + r + 1/2 + 0.02 = r + 1/2 + 0.52$. While the main error contribution is from $1/\sigma$, there is still a noticeable effect from the lower order terms in $A_2(p)$. Moreover, we are now in the ballpark of Fermigier's bound; of course, we were already there without the potential correction term!

Acknowledgements This chapter is an extension of a talk given by the second named author at the Simons Symposium on Families of Automorphic Forms and the Trace Formula in Rio Grande, Puerto Rico, on January 28th, 2014. It is a pleasure to thank them and the organizers, as well as the anonymous referee who provided numerous suggestions which improved the paper (in particular the statement and proof of Theorem 2.9). Much of this work was conducted while the second named author was supported by NSF Grants DMS0600848, DMS0970067 and DMS1265673, and the others by NSF Grant DMS1347804 and Williams College.

Appendix: Biases in Second Moments in Additional Families

By Megumi Asada, Eva Fourakis, Steven J. Miller and Kevin Yang

This appendix describes work in progress on investigating biases in the second moments of other families. It is thus a companion to Sect. 4.2. Fuller details and proofs will be reported by the authors in [AFMY]; our purpose below is to quickly describe results on analogues of the Bias Conjecture.

Dirichlet Families

Let q be prime, and let $\mathscr{F}_q$ be the family of nontrivial Dirichlet characters of level q. In this family, the second moment is given by

$$M_2(\mathscr{F}_q; X) = \sum_{p<X} \sum_{\chi \in \mathscr{F}_q} \chi^2(p). \tag{94}$$

Denote the amalgamation of families by $\mathscr{F}_Y = \cup_{Y/2<q<Y} \mathscr{F}_q$, with the naturally defined second moment.

Computing $M_2(\mathscr{F}_q, X)$ is straightforward from the orthogonality relations, which as we've seen earlier yields a quantity related to the classical problem on the distribution of primes in residue classes. Approximating carefully $\pi(X)$ and $\pi(X, q, a)$ via the Prime Number Theorem, one can deduce the following.

Theorem 4.8. *The family $\mathscr{F}_q$ has positive bias, independent of q, in the second moments of the Fourier coefficients of the L-functions.*

Remark 4.9. Note that the behavior of Dirichlet L-functions is very different than that from families of elliptic curves.

Now, suppose $q \neq \ell$ is a prime such that $q \equiv 1(\ell)$. Let $\mathscr{F}_{q,\ell}$ be the family of nontrivial ℓ-torsion Dirichlet characters of level q, which is nonempty by the stipulated congruence condition. In this family, the second moment is given by

$$M_2(\mathscr{F}_{q,\ell}; X) = \sum_{p<X} \sum_{\chi \in \mathscr{F}_{q,\ell}} \chi^2(p). \tag{95}$$

Define $\mathscr{F}_Y := \cup_{Y/2<q<Y} \mathscr{F}_{q,\ell_q}$ for any choice of suitable ℓ_q for each q.

Theorem 4.10. *The family $\mathscr{F}_{q,\ell}$ has zero bias independent of q and ℓ. Thus, $\mathscr{F}_Y$ exhibits zero bias in the second moments of the Fourier coefficients of the L-functions.*

Families of Holomorphic Cusp Forms

Let $S_{k,q}(\chi_0)$ denote the space of cuspidal newforms of level q, weight k and trivial nebentypus, endowed with the structure of a Hilbert space via the Petersson inner product. Let $B_{k,q}(\chi_0)$ be any orthonormal basis of $S_{k,q}(\chi_0)$ and let $\mathscr{F}_X := \cup_{k<X:k\equiv 0(2)} \mathscr{B}_{k,q=1}(\chi_0)$. In this family, the second moment is given by the weighted Fourier coefficients[10]:

$$M_2(\mathscr{F}_X; \delta) = \sum_{p<X^\delta} \sum_{k<X:k\equiv 0(2)} \sum_{f \in B_{k,q}(\chi_0)} |\psi_f(p)|^2, \tag{96}$$

where $\psi_f(p) = \frac{(\Gamma(k-1))^{\frac{1}{2}}}{(4\pi p)^{\frac{k-1}{2}}} \lambda_f(p) \sqrt{\log p}$, with $\lambda_f(p)$ the Hecke eigenvalue of f for the Hecke operator T_p. Let $\mathscr{F}_{X;\varepsilon} = \cup_{q<X^\varepsilon} \mathscr{F}_X$ be the amalgamation of families with the second moment

$$M_2(\mathscr{F}_{X;\varepsilon}; \delta) = \sum_{p<X^\delta} \sum_{q<X^\varepsilon} \sum_{k<X:k\equiv 0(2)} \sum_{f \in B_{k,q}(\chi_0)} |\psi_f(p)|^2. \tag{97}$$

[10]Following [ILS] we can remove the weights, but their presence facilitates the application of the Petersson formula.

The Petersson Formula provides an explicit method of computing $M_2(\mathscr{F}_X;\delta)$ via Kloosterman sums and Bessel functions. Averaging over the level and weight to obtain asymptotic approximations as in [ILS], we prove the following theorem in [AFMY].

Theorem 4.11. *The family $\mathscr{F}_X$ has negative bias, independent of the level q of $\frac{1}{2}$, in the second moments of the Fourier coefficients of the L-functions. Thus, $\mathscr{F}_{X;\varepsilon}$ exhibits negative bias.*

Let us now let $H^*_{k,q}(\chi_0)$ denote a basis of newforms of Petersson norm 1 for prime level q and even weight k. We consider another weighted second moment, given by

$$M_2^{\text{weighted}}(\mathscr{F}_X;\delta) = \sum_{p<X^\delta} \sum_{k<X:k\equiv 0(2)} \sum_{f\in H^*_{k,q}(\chi_0)} \frac{\Gamma(k)}{(4\pi)^k}|\lambda_f(p)|^2. \tag{98}$$

Similarly, let $\mathscr{F}_{X;\varepsilon} = \cup_{q<X^\varepsilon}\mathscr{F}_X$ be the amalgamation of these families with the weighted second moment

$$M_2^{\text{weighted}}(\mathscr{F}_{X;\varepsilon};\delta) = \sum_{p<X^\delta} \sum_{q<X^\varepsilon} \sum_{k<X:k\equiv 0(2)} \sum_{f\in H^*_{k,q}(\chi_0)} \frac{\Gamma(k)}{(4\pi)^k}|\lambda_f(p)|^2. \tag{99}$$

We prove the following in [AFMY].

Theorem 4.12. *The family $\mathscr{F}_X$ has positive bias dependent on the level q. Moreover, the family $\mathscr{F}_{X,\varepsilon}$ exhibits positive bias as well.*

If we now consider the unweighted second moment given by

$$M_2(\mathscr{F}_X;\delta) = \sum_{p<X^\delta} \sum_{k<X:k\equiv 0(2)} \sum_{f\in H^*_{k,q}(\chi_0)} \lambda_f^2(p), \tag{100}$$

we prove the following in [AFMY] as well.

Theorem 4.13. *Assume $\delta < 1$ and $\varepsilon = 1$. The family $\mathscr{F}_X$ has positive bias dependent on q. Moreover, the family $\mathscr{F}_{X;\varepsilon}$ exhibits positive unweighted bias as well.*

References

[AAILMZ] L. Alpoge, N. Amersi, G. Iyer, O. Lazarev, S.J. Miller, L. Zhang, Maass waveforms and low-lying zeros, in *Analytic Number Theory: In Honor of Helmut Maier's 60th birthday* (Carl Pomerance, Michael Th. Rassias, editors), Springer-Verlag (2015)

[AFMY] M. Asada, E. Fourakis, A. Kwon, S.J. Miller, K. Yang, Biases in the second moments of Fourier coefficients in families of L-functions. Preprint (2015)

[AMil] L. Alpoge, S.J. Miller, Low lying zeros of Maass form L-functions. Int. Math. Res. Not. 24 pp. (2014). doi:10.1093/imrn/rnu012

[BFMT-B] O. Barrett, F.W.K. Firk, S.J. Miller, C. Turnage-Butterbaugh, From quantum systems to L-functions: pair correlation statistics and beyond, in *Open Problems in Mathematics*, ed. by J. Nash Jr., M.T. Rassias, Springer-Verlag (2016)

[BhSh1] M. Bhargava, A. Shankar, Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves (2010). arXiv:1006.1002

[BhSh2] M. Bhargava, A. Shankar, Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0 (2010). arXiv:1007.0052

[BS-D1] B. Birch, H.P.F. Swinnerton-Dyer, Notes on elliptic curves. I. J. Reine Angew. Math. **212**, 7–25 (1963)

[BS-D2] B. Birch, H.P.F. Swinnerton-Dyer, Notes on elliptic curves. II. J. Reine Angew. Math. **218**, 79–108 (1965)

[Br] A. Brumer, The average rank of elliptic curves I. Invent. Math. **109**, 445–472 (1992)

[CFKRS] J.B. Conrey, D. Farmer, P. Keating, M. Rubinstein, N. Snaith, Integral moments of L-functions. Proc. Lond. Math. Soc. (3) **91**(1), 33–104 (2005)

[CFZ1] J.B. Conrey, D.W. Farmer, M.R. Zirnbauer, Autocorrelation of ratios of L-functions. Commun. Number Theory Phys. **2**(3), 593–636 (2008)

[CFZ2] J.B. Conrey, D.W. Farmer, M.R. Zirnbauer, Howe pairs, supersymmetry, and ratios of random characteristic polynomials for the classical compact groups (2005). http://arxiv.org/abs/math-ph/0511024

[Con] J.B. Conrey, L-Functions and random matrices, in *Mathematics Unlimited — 2001 and Beyond* (Springer, Berlin, 2001), pp. 331–352

[ConSn] J.B. Conrey, N.C. Snaith, Applications of the L-functions ratios conjecture. Proc. Lond. Math. Soc. **93**(3), 594–646 (2007)

[ConSn2] J.B. Conrey, N.C. Snaith, Triple correlation of the Riemann zeros (2007). http://arxiv.org/abs/math/0610495

[Da] H. Davenport, *Multiplicative Number Theory*, 2nd edn. Graduate Texts in Mathematics, vol. 74 (Springer, New York, 1980). Revised by H. Montgomery

[DHKMS1] E. Dueñez, D.K. Huynh, J.C. Keating, S.J. Miller, N. Snaith, The lowest eigenvalue of Jacobi random matrix ensembles and Painlevé VI. J. Phys. A Math. Theor. **43**, 405204 (27pp.) (2010)

[DHKMS2] E. Dueñez, D.K. Huynh, J.C. Keating, S.J. Miller, N. Snaith, Models for zeros at the central point in families of elliptic curves (with Eduardo Dueñez, Duc Khiem Huynh, Jon Keating and Nina Snaith). J. Phys. A Math. Theor. **45**, 115207 (32pp.) (2012)

[DuMil1] E. Dueñez, S.J. Miller, The low lying zeros of a GL(4) and a GL(6) family of L-functions. Compos. Math. **142**(6), 1403–1425 (2006)

[DuMil2] E. Dueñez, S.J. Miller, The effect of convolving families of L-functions on the underlying group symmetries. Proc. Lond. Math. Soc. (2009). doi: 10.1112/plms/pdp018

[Ed] H.M. Edwards, *Riemann's Zeta Function* (Academic, New York, 1974)

[ER-GR] A. Entin, E. Roditty-Gershon, Z. Rudnick, Low-lying zeros of quadratic Dirichlet L-functions, hyper-elliptic curves and random matrix theory. Geom. Funct. Anal. **23**(4), 1230–1261 (2013)

[Fe1] S. Fermigier, Zéros des fonctions L de courbes elliptiques. Exp. Math. **1**(2), 167–173 (1992)

[Fe2] S. Fermigier, Étude expérimentale du rang de familles de courbes elliptiques sur $\mathbb{Q}$. Exp. Math. **5**(2), 119–130 (1996)

[FioMil] D. Fiorilli, S.J. Miller, Surpassing the ratios conjecture in the 1-level density of Dirichlet L-functions. Algebra and Number Theory **9**(1), 13–52 (2015)

[FirMil] F.W.K. Firk, S.J. Miller, Nuclei, primes and the random matrix connection. Symmetry **1**, 64–105 (2009). doi:10.3390/sym1010064

[FoIw] E. Fouvry, H. Iwaniec, Low-lying zeros of dihedral L-functions. Duke Math. J. **116** (2), 189–217 (2003)

[FoPo] E. Fouvrey, J. Pomykala, Rang des courbes elliptiques et sommes d'exponentelles. Monat. Math. **116**, 111–125 (1993)

[For] P. Forrester, *Log-Gases and Random Matrices*. London Mathematical Society Monograph, vol. 34 (Princeton University Press, Princeton, NJ, 2010)

[FrMil] J. Freeman, S.J. Miller, Determining Optimal Test Functions for Bounding the Average Rank in Families of L-Functions, in SCHOLAR – a Scientific Celebration Highlighting Open Lines of Arithmetic Research, Conference in Honour of M. Ram Murty's Mathematical Legacy on his 60th Birthday (A. C. Cojocaru, C. David and F. Pappaardi, editors), Contemporary Mathematics, AMS and CRM, **655** (2015)

[Gao] P. Gao, N-level density of the low-lying zeros of quadratic Dirichlet L-functions. Ph.D. thesis, University of Michigan (2005)

[GJMMNPP] J. Goes, S. Jackson, S.J. Miller, D. Montague, K. Ninsuwan, R. Peckner, T. Pham, A unitary test of the ratios conjecture. J. Number Theory **130**(10), 2238–2258 (2010)

[GoHuKe] S.M. Gonek, C.P. Hughes, J.P. Keating, A hybrid Euler-Hadamard product formula for the Riemann zeta function. Duke Math. J. **136**, 507–549 (2007)

[GolKon] D. Goldfeld, A. Kontorovich, On the GL(3) Kuznetsov formula with applications to symmetry types of families of L-functions. Automorphic representations and L-functions, 263–310, Tata Inst. Fundam. Res. Stud. Math., 22, Tata Inst. Fund. Res., Mumbai (2013)

[Gü] A. Güloğlu, Low-lying zeros of symmetric power L-functions. Int. Math. Res. Not. **2005**(9), 517–550 (2005)

[Ha] B. Hayes, The spectrum of Riemannium. Am. Sci. **91**(4), 296–300 (2003)

[H-B] R. Heath-Brown, The average analytic rank of elliptic curves. Duke Math. J. **122**(3), 591–623 (2004)

[HuMil] C. Hughes, S.J. Miller, Low-lying zeros of L-functions with orthogonal symmetry. Duke Math. J. **136**(1), 115–172 (2007)

[HuRud] C. Hughes, Z. Rudnick, Linear statistics of low-lying zeros of L-functions. Q. J. Math. Oxford **54**, 309–333 (2003)

[HuyKeSn] D.K. Huynh, J.P. Keating, N.C. Snaith, Lower order terms for the one-level density of elliptic curve L-functions. J. Number Theory **129**(12), 2883–2902 (2009)

[HuyMM] D.K. Huynh, S.J. Miller, R. Morrison, An elliptic curve family test of the ratios conjecture. J. Number Theory **131**, 1117–1147 (2011)

[ILS] H. Iwaniec, W. Luo, P. Sarnak, Low lying zeros of families of L-functions. Inst. Hautes Études Sci. Publ. Math. **91**, 55–131 (2000)

[Iw] H. Iwaniec, *Introduction to the Spectral Theory of Automorphic Forms*. Biblioteca de la Revista Matemática Iberoamericana (1995)

[IwKo] H. Iwaniec, E. Kowalski, *Analytic Number Theory*. AMS Colloquium Publications, vol. 53 (American Mathematical Society, Providence, RI, 2004)

[KaSa1] N. Katz, P. Sarnak, *Random Matrices, Frobenius Eigenvalues and Monodromy*. AMS Colloquium Publications, vol. 45 (American Mathematical Society, Providence, RI, 1999)

[KaSa2] N. Katz, P. Sarnak, Zeros of zeta functions and symmetries. Bull. Am. Math. Soc. **36**, 1–26 (1999)

[KeSn1] J.P. Keating, N.C. Snaith, Random matrix theory and $\zeta(1/2+it)$. Commun. Math. Phys. **214**(1), 57–89 (2000)

[KeSn2] J.P. Keating, N.C. Snaith, Random matrix theory and L-functions at $s = 1/2$. Commun. Math. Phys. **214**(1), 91–110 (2000)

[KeSn3] J.P. Keating, N.C. Snaith, Random matrices and L-functions. Random matrix theory. J. Phys. A **36**(12), 2859–2881 (2003)

[Kim] H. Kim, Functoriality for the exterior square of GL_2 and the symmetric fourth of GL_2. J. Am. Math. Soc. **16**(1), 139–183 (2003)

[KimSa] H. Kim, P. Sarnak, Appendix: Refined estimates towards the Ramanujan and Selberg conjectures. Appendix to [Kim]

[LevMil] J. Levinson, S.J. Miller, The n-level density of zeros of quadratic Dirichlet L-functions. Acta Arith. **161**, 145–182 (2013)

[MaTe] J. Matz, N. Templier, Sato-Tate equidistribution for families of Hecke-Maass forms on $\mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$. Preprint (2015), http://arxiv.org/abs/1505.07285

[Meh] M. Mehta, *Random Matrices*, 2nd edn. (Academic, Boston, 1991)

[Mic] P. Michel, Rang moyen de familles de courbes elliptiques et lois de Sato-Tate. Monat. Math. **120**, 127–136 (1995)

[Mil1] S.J. Miller, 1- and 2-level densities for families of elliptic curves: evidence for the underlying group symmetries. Ph.D. thesis, Princeton University (2002)

[Mil2] S.J. Miller, 1- and 2-level densities for families of elliptic curves: evidence for the underlying group symmetries. Compos. Math. **140**, 952–992 (2004)

[Mil3] S.J. Miller, Variation in the number of points on elliptic curves and applications to excess rank. C. R. Math. Rep. Acad. Sci. Canada **27**(4), 111–120 (2005)

[Mil5] S.J. Miller, A symplectic test of the *L*-functions ratios conjecture. Int. Math. Res. Not. **2008**, 36 pp. (2008). Article ID rnm146

[Mil6] S.J. Miller, Lower order terms in the 1-level density for families of holomorphic cuspidal newforms. Acta Arith. **137**, 51–98 (2009)

[Mil7] S.J. Miller, An orthogonal test of the *L*-functions ratios conjecture. Proc. Lond. Math. Soc. (2009). doi:10.1112/plms/pdp009

[MilMo] S.J. Miller, D. Montague, An orthogonal test of the *L*-functions ratios conjecture. II. Acta Arith. **146**, 53–90 (2011)

[MilPe] S.J. Miller, R. Peckner, Low-lying zeros of number field *L*-functions. J. Number Theory **132**, 2866–2891 (2012)

[MMRW] B. Mackall, S.J. Miller, C. Rapti, K. Winsor, Lower-order biases in elliptic curve fourier coefficients in families, in Frobenius Distributions: Lang-Trotter and Sato-Tate Conjectures (David Kohel and Igor Shparlinski, editors), Contemporary Mathematics **663**, AMS, Providence, RI (2016)

[Mon] H.L. Montgomery, The pair correlation of zeros of the zeta function, in *Analytic Number Theory, Proceedings of Symposia in Pure Mathematics*, vol. 24 (American Mathematical Society, Providence, RI, 1973), pp. 181–193

[MonVa] H.L. Montgomery, R.C. Vaughan, The large sieve. Mathematika **20**, 119–134 (1973)

[Od1] A. Odlyzko, On the distribution of spacings between zeros of the zeta function. Math. Comput. **48**(177), 273–308 (1987)

[Od2] A. Odlyzko, The 10^{22}-nd zero of the Riemann zeta function, *Proceedings of Conference on Dynamical, Spectral and Arithmetic Zeta-Functions*, ed. by M. van Frankenhuysen, M.L. Lapidus, Contemporary Mathematics Series (American Mathematical Society, Providence, RI, 2001). http://www.research.att.com/~amo/doc/zeta.html

[OS1] A.E. Özlük, C. Snyder, Small zeros of quadratic *L*-functions. Bull. Aust. Math. Soc. **47**(2), 307–319 (1993)

[OS2] A.E. Özlük, C. Snyder, On the distribution of the nontrivial zeros of quadratic *L*-functions close to the real axis. Acta Arith. **91**(3), 209–228 (1999)

[RiRo] G. Ricotta, E. Royer, Statistics for low-lying zeros of symmetric power *L*-functions in the level aspect. Preprint, Forum Math. **23**, 969–1028 (2011)

[Ro] E. Royer, Petits zéros de fonctions *L* de formes modulaires. Acta Arith. **99**(2), 147–172 (2001)

[RoSi] M. Rosen, J. Silverman, On the rank of an elliptic surface. Invent. Math. **133**, 43–67 (1998)

[Rub] M. Rubinstein, Low-lying zeros of *L*–functions and random matrix theory. Duke Math. J. **109**, 147–181 (2001)

[RudSa] Z. Rudnick, P. Sarnak, Zeros of principal *L*-functions and random matrix theory. Duke Math. J. **81**, 269–322 (1996)

[SaShTe] P. Sarnak, S.W. Shin, N. Templier, Families of *L*-functions and their symmetry, in *Families of Automorphic Forms and the Trace Formula*, ed. by W. Müller, S.W. Shin, N. Templier (Springer, New York, 2016)

[ShTe] S.W. Shin, N. Templier, *Sato-Tate Theorem for Families and low-lying zeros of automorphic L-functions.* With Appendix 1 by R. Kottwitz and Appendix 2 by R. Cluckers, J. Gordon and I. Halupczok. Invent. Math. **203**(1), 1–177 (2016)

[Sil] J. Silverman, The average rank of an algebraic family of elliptic curves. J. Reine Angew. Math. **504**, 227–236 (1998)

[Ti] E. Titchmarsh, D.R. Heath-Brown, *The Theory of the Riemann Zeta-Function* (Oxford University Press, Oxford, 1986)

[Wa] M. Watkins, Rank distribution in a family of cubic twists (2007). http://arxiv.org/pdf/math.NT/0412427

[Ya] A. Yang, Low-lying zeros of Dedekind zeta functions attached to cubic number fields. Preprint

[Yo1] M. Young, Lower-order terms of the 1-level density of families of elliptic curves. Int. Math. Res. Not. **2005**(10), 587–633 (2005)

[Yo2] M. Young, Low-lying zeros of families of elliptic curves. J. Am. Math. Soc. **19**(1), 205–250 (2006)

Asymptotics of Automorphic Spectra and the Trace Formula

Werner Müller

Abstract This paper is a survey article on the limiting behavior of the discrete spectrum of the right regular representation in $L^2(\Gamma\backslash G)$ for a lattice Γ in a semisimple Lie group G. We discuss various aspects of the Weyl law, the limit multiplicity problem, and the analytic torsion.

1991 *Mathematics Subject Classification.* Primary: 11F70, Secondary: 58J52, 11F75

1 Introduction

Let G be a connected linear semisimple Lie group of noncompact type with a fixed choice of a Haar measure. Let $\Pi(G)$ denote the set of all equivalence classes of irreducible unitary representations of G, equipped with the Fell topology [Di]. We fix a Haar measure on G. Let $\Gamma \subset G$ be a lattice in G, i.e., a discrete subgroup such that $\mathrm{vol}(\Gamma\backslash G) < \infty$. Let R_Γ be the right regular representation of G on $L^2(\Gamma\backslash G)$. Let $L^2_{\mathrm{disc}}(\Gamma\backslash G)$ be the span of all irreducible subrepresentations of R_Γ and denote by $R_{\Gamma,\mathrm{disc}}$ the restriction of R_Γ to $L^2_{\mathrm{disc}}(\Gamma\backslash G)$. Then $R_{\Gamma,\mathrm{disc}}$ decomposes discretely as

$$R_{\Gamma,\mathrm{disc}} \cong \hat{\bigoplus}_{\pi\in\Pi(G)} m_\Gamma(\pi)\pi, \tag{1}$$

where

$$m_\Gamma(\pi) = \dim \mathrm{Hom}_G(\pi, R_\Gamma) = \dim \mathrm{Hom}_G(\pi, R_{\Gamma,\mathrm{disc}})$$

is the multiplicity with which π occurs in R_Γ. The multiplicities are known to be finite under a weak reduction-theoretic assumption on (G,Γ) [OW], which

W. Müller (✉)
Universität Bonn, Mathematisches Institut, Endenicher Allee 60, D-53115 Bonn, Germany
e-mail: mueller@math.uni-bonn.de

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_12

is satisfied if G has no compact factors or if Γ is arithmetic. The study of the multiplicities $m_\Gamma(\pi)$ is one of the main concerns in the theory of automorphic forms. Apart from special cases like discrete series representations, one cannot hope in general to describe the multiplicity function on $\Pi(G)$ explicitly. A more feasible and interesting problem is the study of the asymptotic behavior of the multiplicities with respect to the growth of various parameters such as the level of congruence subgroups or the infinitesimal character of π. This is closely related to the study of families of automorphic forms (see [SST]).

The first problem in this context is the Weyl law. Let K be a maximal compact subgroup of G. Fix an irreducible representation σ of K. Let $\Pi(G;\sigma)$ be the subspace of all $\pi \in \Pi(G)$ such that $[\pi|_K : \sigma] > 0$. Especially, if σ is the trivial representation, then $\Pi(G;\sigma)$ is the spherical dual $\Pi^{\mathrm{sph}}(G)$. Given $\pi \in \Pi(G)$, denote by $\lambda_\pi = \pi(\Omega)$ the Casimir eigenvalue of π. For $\lambda \geq 0$ let the counting function be defined by

$$N_\Gamma^\sigma(\lambda) = \sum_{\substack{\pi \in \Pi(G;\sigma) \\ |\lambda_\pi| \leq \lambda}} m_\Gamma(\pi). \tag{2}$$

Then the problem is to determine the behavior of the counting function as $\lambda \to \infty$.

Another basic problem is the limit multiplicity problem, which is the study of the asymptotic behavior of the multiplicities if $\mathrm{vol}(\Gamma\backslash G) \to \infty$. For $G = \mathrm{GL}_n(\mathbb{R})$ this corresponds to the study of harmonic families of cuspidal automorphic representations of $\mathrm{GL}_n(\mathbb{A})$, $\mathbb{A}$ being the ring of adeles (see [SST]). More precisely, for a given lattice Γ define the discrete spectral measure μ_Γ on $\Pi(G)$, associated with Γ, by

$$\mu_\Gamma = \frac{1}{\mathrm{vol}(\Gamma\backslash G)} \sum_{\pi \in \Pi(G)} m_\Gamma(\pi)\delta_\pi, \tag{3}$$

where δ_π is the Dirac measure at π. Then the limit multiplicity problem is concerned with the study of the asymptotic behavior of μ_Γ as $\mathrm{vol}(\Gamma\backslash G) \to \infty$. For appropriate sequences of lattices (Γ_n) one expects that the measures μ_{Γ_n} converge to the Plancherel measure μ_{pl} on $\Pi(G)$.

There are closely related problems in topology and spectral theory. One of them concerns Betti numbers. Let K be a maximal compact subgroup of G and put $\widetilde{X} = G/K$. Let Γ be a uniform lattice in G and let (Γ_n) be a tower of normal subgroups of Γ. Put $X = \Gamma\backslash\widetilde{X}$ and $X_n = \Gamma_n\backslash\widetilde{X}$, $n \in \mathbb{N}$. Then $X_n \to X$ is a sequence of finite normal coverings of X. For any topological space Y let $b_k(Y)$ denote the k-th Betti number of Y. Then

$$\lim_{n\to\infty} \frac{b_k(X_n)}{\mathrm{vol}(X_n)} = b_k^{(2)}(X), \tag{4}$$

where $b_k^{(2)}(X)$ is the k-th L^2-Betti number of X. This was proved by Lück [Lu1] in the more general context of CW-complexes. In the case of locally symmetric spaces, it follows from the results about limit multiplicities. Again, it was extended by Abert et al. [AB1] to much more general sequences of uniform lattices.

A more sophisticated spectral invariant is the Ray-Singer analytic torsion $T_X(\rho)$ (see [RS]). It depends on a finite dimensional representation ρ of Γ and is defined in terms of the spectra of the Laplace operators $\Delta_p(\rho)$ on p-forms with coefficients in the flat bundle associated with ρ. Of particular interest are representations of Γ which arise as the restriction of a representation of G. For appropriate representations, called strongly acyclic, Bergeron and Venkatesh [BV] studied the asymptotic behavior of $\log T_{X_n}(\rho)$ as $n \to \infty$. One of their main results is

$$\lim_{n\to\infty} \frac{\log T_{X_n}(\rho)}{\mathrm{vol}(X_n)} = \log T_X^{(2)}(\rho), \tag{5}$$

where $T_X^{(2)}(\rho)$ is the L^2-torsion [Lo, MV]. Using the equality of analytic torsion and Reidemeister torsion [Ch, Mu1], (5) implies results about the growth of the torsion subgroup in the integer homology of arithmetic groups. Let $\mathbf{G}$ be a semisimple algebraic group over $\mathbb{Q}$, $G = \mathbf{G}(\mathbb{R})$ and $\Gamma \subset \mathbf{G}(\mathbb{Q})$ a co-compact, arithmetic subgroup. As shown in [BV], there are strongly acyclic representations ρ of G on a finite dimensional vector space V such that V contains a Γ-invariant lattice M. Let $\mathcal{M}$ be the local system of free $\mathbb{Z}$-modules over X, attached to M. Then the cohomology $H_*(X,\mathcal{M})$ of X with coefficients in $\mathcal{M}$ is a finite abelian group. Denote by $|H_*(X,\mathcal{M})|$ its order. Assume that $d = \dim(X)$ is odd. Then by [BV] one has

$$\lim_{n\to\infty} \sum_{p=1}^{d} (-1)^{p+\frac{d-1}{2}} \frac{\log |H_p(X_n,\mathcal{M})|}{[\Gamma\colon\Gamma_n]} = c_{M,G}\,\mathrm{vol}(X),$$

where $c_{M,G}$ is a constant that depends only on G and M. Moreover, if $\delta(G) := \operatorname{rank} G - \operatorname{rank} K = 1$, then $c_{M,G} > 0$. It is conjectured that the limit

$$\lim_{n\to\infty} \frac{\log |H_j(X_n,\mathcal{M})|}{[\Gamma\colon\Gamma_n]} \tag{6}$$

always exists and is equal to zero, unless $\delta(G) = 1$ and $j = (d-1)/2$. In the latter case it is equal to $c_{M,G}$ times $\mathrm{vol}(X)$. The conjecture is known to be true for $G = \mathrm{SL}_2(\mathbb{C})$.

An important problem is to extend these results to the non-compact case.

2 The Arthur Trace Formula

The trace formula is one of the main technical tools to study the kind of spectral problems mentioned in the introduction. For $\mathbb{R}$-rank one groups the Selberg trace formula is available [Wa1]. In the higher rank case the Selberg trace formula is replaced by the Arthur trace formula.

In this section we recall Arthur's trace formula, and in particular the refinement of the spectral expansion obtained in [FLM1].

2.1 Notation

We will mostly use the notation of [FLM1]. Let $\mathbf{G}$ be a reductive group defined over $\mathbb{Q}$ and let $\mathbb{A}$ be the ring of adeles of $\mathbb{Q}$. We fix a maximal compact subgroup $\mathbf{K} = \prod_v \mathbf{K}_v = \mathbf{K}_\infty \cdot \mathbf{K}_{\text{fin}}$ of $\mathbf{G}(\mathbb{A}) = \mathbf{G}(\mathbb{R}) \cdot \mathbf{G}(\mathbb{A}_{\text{fin}})$.

Let $\mathfrak{g}$ and $\mathfrak{k}$ denote the Lie algebras of $\mathbf{G}(\mathbb{R})$ and $\mathbf{K}_\infty$, respectively. Let θ be the Cartan involution of $\mathbf{G}(\mathbb{R})$ with respect to $\mathbf{K}_\infty$. It induces a Cartan decomposition $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$. We fix an invariant bi-linear form B on $\mathfrak{g}$ which is positive definite on $\mathfrak{p}$ and negative definite on $\mathfrak{k}$. This choice defines a Casimir operator Ω on $\mathbf{G}(\mathbb{R})$, and we denote the Casimir eigenvalue of any $\pi \in \Pi(\mathbf{G}(\mathbb{R}))$ by λ_π. Similarly, we obtain a Casimir operator $\Omega_{\mathbf{K}_\infty}$ on $\mathbf{K}_\infty$ and write λ_τ for the Casimir eigenvalue of a representation $\tau \in \Pi(\mathbf{K}_\infty)$ (cf. [BG, § 2.3]). The form B induces a Euclidean scalar product $(X, Y) = -B(X, \theta(Y))$ on $\mathfrak{g}$ and all its subspaces. For $\tau \in \Pi(\mathbf{K}_\infty)$ we define $\|\tau\|$ as in [CD, § 2.2].

We fix a maximal $\mathbb{Q}$-split torus $\mathbf{S}_0$ of $\mathbf{G}$ and let $\mathbf{M}_0$ be its centralizer, which is a minimal Levi subgroup defined over $\mathbb{Q}$. We assume that the maximal compact subgroup $\mathbf{K} \subset \mathbf{G}(\mathbb{A})$ is admissible with respect to $\mathbf{M}_0$ [Ar5, § 1]. Denote by A_0 the identity component of $\mathbf{S}_0(\mathbb{R})$, which is viewed as a subgroup of $\mathbf{S}_0(\mathbb{A})$. We write $\mathcal{L}$ for the (finite) set of Levi subgroups containing $\mathbf{M}_0$, i.e., the set of centralizers of subtori of $\mathbf{S}_0$. Let $W_0 = N_{\mathbf{G}(\mathbb{Q})}(\mathbf{S}_0)/\mathbf{M}_0$ be the Weyl group of $(\mathbf{G}, \mathbf{S}_0)$, where $N_{\mathbf{G}(\mathbb{Q})}(H)$ is the normalizer of H in $\mathbf{G}(\mathbb{Q})$. For any $s \in W_0$ we choose a representative $w_s \in \mathbf{G}(\mathbb{Q})$. Note that W_0 acts on $\mathcal{L}$ by $s\mathbf{M} = w_s \mathbf{M} w_s^{-1}$.

Let now $\mathbf{M} \in \mathcal{L}$. We write $\mathbf{S}_M$ for the split part of the identity component of the center of $\mathbf{M}$. Set $A_M = A_0 \cap \mathbf{S}_M(\mathbb{R})$ and $W(\mathbf{M}) = N_{\mathbf{G}(\mathbb{Q})}(\mathbf{M})/\mathbf{M}$, which can be identified with a subgroup of W_0. Denote by $\mathfrak{a}_M^*$ the $\mathbb{R}$-vector space spanned by the lattice $X^*(\mathbf{M})$ of $\mathbb{Q}$-rational characters of $\mathbf{M}$ and let $\mathfrak{a}_{M,\mathbb{C}}^* = \mathfrak{a}_M^* \otimes_{\mathbb{R}} \mathbb{C}$ be its complexification. We write $\mathfrak{a}_M$ for the dual space of $\mathfrak{a}_M^*$, which is spanned by the co-characters of $\mathbf{S}_M$. Let

$$H_M : \mathbf{M}(\mathbb{A}) \to \mathfrak{a}_M$$

be the homomorphism given by

$$e^{\langle \chi, H_M(m)\rangle} = |\chi(m)|_{\mathbb{A}} = \prod_v |\chi(m_v)|_v \tag{7}$$

for any $\chi \in X^*(\mathbf{M})$ and denote by $\mathbf{M}(\mathbb{A})^1 \subset \mathbf{M}(\mathbb{A})$ the kernel of H_M. Let $\mathcal{L}(\mathbf{M})$ be the set of Levi subgroups containing $\mathbf{M}$ and $\mathcal{P}(\mathbf{M})$ the set of parabolic subgroups of $\mathbf{G}$ with Levi part $\mathbf{M}$. We also write $\mathcal{F}(\mathbf{M}) = \mathcal{F}^G(\mathbf{M}) = \coprod_{\mathbf{L}\in\mathcal{L}(\mathbf{M})} \mathcal{P}(\mathbf{L})$ for the (finite) set of parabolic subgroups of $\mathbf{G}$ containing $\mathbf{M}$. Note that $W(\mathbf{M})$ acts on $\mathcal{P}(\mathbf{M})$ and $\mathcal{F}(\mathbf{M})$ by $s\mathbf{P} = w_s \mathbf{P} w_s^{-1}$. Denote by Σ_M the set of reduced roots of $\mathbf{S}_M$ on the Lie algebra of $\mathbf{G}$. For any $\alpha \in \Sigma_M$ we denote by $\alpha^\vee \in \mathfrak{a}_M$ the corresponding co-root. Let $L^2_{\mathrm{disc}}(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A}))$ be the discrete part of $L^2(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A}))$, i.e., the closure of the sum of all irreducible subrepresentations of the regular representation of $\mathbf{M}(\mathbb{A})$. We denote by $\Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))$ the countable set of equivalence classes of irreducible unitary representations of $\mathbf{M}(\mathbb{A})$ which occur in the decomposition of $L^2_{\mathrm{disc}}(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A}))$ into irreducible representations.

For any $\mathbf{L} \in \mathcal{L}(\mathbf{M})$ we identify $\mathfrak{a}_L^*$ with a subspace of $\mathfrak{a}_M^*$. We denote by $\mathfrak{a}_M^L$ the annihilator of $\mathfrak{a}_L^*$ in $\mathfrak{a}_M$. We set

$$\mathcal{L}_1(\mathbf{M}) = \{\mathbf{L} \in \mathcal{L}(\mathbf{M}) : \dim \mathfrak{a}_M^L = 1\}$$

and

$$\mathcal{F}_1(\mathbf{M}) = \bigcup_{\mathbf{L}\in\mathcal{L}_1(\mathbf{M})} \mathcal{P}(\mathbf{L}).$$

Note that the restriction of the scalar product $(\cdot,\cdot)$ on $\mathfrak{g}$ defined above gives $\mathfrak{a}_{M_0}$ the structure of a Euclidean space. In particular, this fixes Haar measures on the spaces $\mathfrak{a}_M^L$ and their duals $(\mathfrak{a}_M^L)^*$. We follow Arthur in the corresponding normalization of Haar measures on the groups $\mathbf{M}(\mathbb{A})$ [Ar1, § 1].

2.2 Intertwining Operators

The main ingredient of the spectral side of the Arthur trace formula are logarithmic derivatives of intertwining operators. We shall now describe the structure of the intertwining operators.

Let $\mathbf{P} \in \mathcal{P}(\mathbf{M})$. We write $\mathfrak{a}_P = \mathfrak{a}_M$. Let $\mathbf{U}_P$ be the unipotent radical of $\mathbf{P}$ and $\mathbf{M}_P$ the unique $\mathbf{L} \in \mathcal{L}(\mathbf{M})$ (in fact the unique $\mathbf{L} \in \mathcal{L}(\mathbf{M}_0)$) such that $\mathbf{P} \in \mathcal{P}(\mathbf{L})$. Denote by $\Sigma_P \subset \mathfrak{a}_P^*$ the set of reduced roots of $\mathbf{S}_M$ on the Lie algebra $\mathfrak{u}_P$ of $\mathbf{U}_P$. Let Δ_P be the subset of simple roots of $\mathbf{P}$, which is a basis for $(\mathfrak{a}_P^G)^*$. Write $\mathfrak{a}_{P,+}^*$ for the closure of the Weyl chamber of $\mathbf{P}$, i.e.

$$\mathfrak{a}_{P,+}^* = \{\lambda \in \mathfrak{a}_M^* : \langle \lambda, \alpha^\vee\rangle \geq 0 \text{ for all } \alpha \in \Sigma_P\} = \{\lambda \in \mathfrak{a}_M^* : \langle \lambda, \alpha^\vee\rangle \geq 0 \text{ for all } \alpha \in \Delta_P\}.$$

Denote by δ_P the modulus function of $\mathbf{P}(\mathbb{A})$. Let $\bar{\mathcal{A}}_2(\mathbf{P})$ be the Hilbert space completion of

$$\{\phi \in C^\infty(\mathbf{M}(\mathbb{Q})\mathbf{U}_P(\mathbb{A})\backslash\mathbf{G}(\mathbb{A})) : \delta_P^{-\frac{1}{2}}\phi(\cdot x) \in L^2_{\mathrm{disc}}(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A})),\ \forall x \in \mathbf{G}(\mathbb{A})\}$$

with respect to the inner product

$$(\phi_1, \phi_2) = \int_{A_M\mathbf{M}(\mathbb{Q})\mathbf{U}_P(\mathbb{A})\backslash\mathbf{G}(\mathbb{A})} \phi_1(g)\overline{\phi_2(g)}\, dg.$$

Let $\alpha \in \Sigma_M$. We say that two parabolic subgroups $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(\mathbf{M})$ are *adjacent* along α, and write $\mathbf{P}|^\alpha\mathbf{Q}$, if $\Sigma_P \cap -\Sigma_Q = \{\alpha\}$. Alternatively, $\mathbf{P}$ and $\mathbf{Q}$ are adjacent if the closure $\overline{\mathbf{PQ}}$ of $\mathbf{PQ}$ belongs to $\mathcal{F}_1(\mathbf{M})$. Any $\mathbf{R} \in \mathcal{F}_1(\mathbf{M})$ is of the form $\overline{\mathbf{PQ}}$ for a unique unordered pair $\{\mathbf{P}, \mathbf{Q}\}$ of parabolic subgroups in $\mathcal{P}(\mathbf{M})$, namely $\mathbf{P}$ and $\mathbf{Q}$ are maximal parabolic subgroups of $\mathbf{R}$, and $\mathbf{P}|^\alpha\mathbf{Q}$ with $\alpha^\vee \in \Sigma_P^\vee \cap \mathfrak{a}_M^R$. Switching the order of $\mathbf{P}$ and $\mathbf{Q}$ changes α to $-\alpha$.

For any $\mathbf{P} \in \mathcal{P}(\mathbf{M})$ let $H_P \colon \mathbf{G}(\mathbb{A}) \to \mathfrak{a}_P$ be the extension of the map H_M, which is defined by (7), to a left $\mathbf{U}_P(\mathbb{A})$-and right $\mathbf{K}$-invariant map. Denote by $\mathcal{A}^2(\mathbf{P})$ the dense subspace of $\bar{\mathcal{A}}^2(\mathbf{P})$ consisting of its $\mathbf{K}$- and $\mathcal{Z}(\mathfrak{g}_\mathbb{C})$-finite vectors, where $\mathcal{Z}(\mathfrak{g}_\mathbb{C})$ is the center of the universal enveloping algebra of $\mathfrak{g}_\mathbb{C} := \mathfrak{g} \otimes \mathbb{C}$. That is, $\mathcal{A}^2(\mathbf{P})$ is the space of automorphic forms ϕ on $\mathbf{U}_P(\mathbb{A})\mathbf{M}(F)\backslash\mathbf{G}(\mathbb{A})$ such that $\delta_P^{-\frac{1}{2}}\phi(\cdot k)$ is a square-integrable automorphic form on $A_M\mathbf{M}(F)\backslash\mathbf{M}(\mathbb{A})$ for all $k \in \mathbf{K}$. Let $\rho(\mathbf{P}, \lambda)$, $\lambda \in \mathfrak{a}^*_{M,\mathbb{C}}$, be the induced representation of $\mathbf{G}(\mathbb{A})$ on $\bar{\mathcal{A}}^2(\mathbf{P})$ given by

$$(\rho(\mathbf{P}, \lambda, y)\phi)(x) = \phi(xy)e^{\langle\lambda, H_P(xy) - H_P(x)\rangle}.$$

It is isomorphic to $\mathrm{Ind}_{\mathbf{P}(\mathbb{A})}^{\mathbf{G}(\mathbb{A})}\left(L^2_{\mathrm{disc}}(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A})) \otimes e^{\langle\lambda, H_M(\cdot)\rangle}\right)$.

For $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(\mathbf{M})$ let

$$M_{Q|P}(\lambda) : \mathcal{A}^2(\mathbf{P}) \to \mathcal{A}^2(\mathbf{Q}), \quad \lambda \in \mathfrak{a}^*_{M,\mathbb{C}},$$

be the standard *intertwining operator* [Ar3, § 1], which is the meromorphic continuation in λ of the integral

$$[M_{Q|P}(\lambda)\phi](x) = \int_{\mathbf{U}_Q(\mathbb{A})\cap\mathbf{U}_P(\mathbb{A})\backslash\mathbf{U}_Q(\mathbb{A})} \phi(nx)e^{\langle\lambda, H_P(nx) - H_Q(x)\rangle}\, dn, \quad \phi \in \mathcal{A}^2(\mathbf{P}),\ x \in \mathbf{G}(\mathbb{A}).$$

These operators satisfy the following properties.

(1) $M_{P|P}(\lambda) \equiv \mathrm{Id}$ for all $\mathbf{P} \in \mathcal{P}(\mathbf{M})$ and $\lambda \in \mathfrak{a}^*_{M,\mathbb{C}}$.
(2) For any $\mathbf{P}, \mathbf{Q}, \mathbf{R} \in \mathcal{P}(\mathbf{M})$ we have $M_{R|P}(\lambda) = M_{R|Q}(\lambda) \circ M_{Q|P}(\lambda)$ for all $\lambda \in \mathfrak{a}^*_{M,\mathbb{C}}$. In particular, $M_{Q|P}(\lambda)^{-1} = M_{P|Q}(\lambda)$.
(3) $M_{Q|P}(\lambda)^* = M_{P|Q}(-\overline{\lambda})$ for any $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(\mathbf{M})$ and $\lambda \in \mathfrak{a}^*_{M,\mathbb{C}}$. In particular, $M_{Q|P}(\lambda)$ is unitary for $\lambda \in i\mathfrak{a}^*_M$.
(4) If $\mathbf{P}|^\alpha\mathbf{Q}$, then $M_{Q|P}(\lambda)$ depends only on $\langle\lambda, \alpha^\vee\rangle$.

Given $\pi \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))$, let $\mathcal{A}^2_\pi(\mathbf{P})$ be the space of all $\phi \in \mathcal{A}^2(\mathbf{P})$ for which the function $x \in \mathbf{M}(\mathbb{A}) \mapsto \delta_P^{-\frac{1}{2}}\phi(xg)$, $g \in \mathbf{G}(\mathbb{A})$, belongs to the π-isotypic subspace $L^2(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A}))$. For any $\mathbf{P} \in \mathcal{P}(\mathbf{M})$ we have a canonical isomorphism of $\mathbf{G}(\mathbb{A}_f) \times (\mathfrak{g}_{\mathbb{C}}, K_\infty)$-modules

$$j_P : \mathrm{Hom}(\pi, L^2(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A}))) \otimes \mathrm{Ind}_{\mathbf{P}(\mathbb{A})}^{\mathbf{G}(\mathbb{A})}(\pi) \to \mathcal{A}^2_\pi(\mathbf{P}).$$

If we fix a unitary structure on π and endow $\mathrm{Hom}(\pi, L^2(A_M\mathbf{M}(\mathbb{Q})\backslash\mathbf{M}(\mathbb{A})))$ with the inner product $(A, B) = B^*A$ (which is a scalar operator on the space of π), the isomorphism j_P becomes an isometry.

Suppose that $\mathbf{P}|^\alpha\mathbf{Q}$. The operator $M_{Q|P}(\pi, s) := M_{Q|P}(s\varpi)|_{\mathcal{A}^2_\pi(P)}$, where $\varpi \in \mathfrak{a}^*_M$ is such that $\langle\varpi, \alpha^\vee\rangle = 1$, admits a normalization by a global factor $n_\alpha(\pi, s)$ which is a meromorphic function in s. We may write

$$M_{Q|P}(\pi, s) \circ j_P = n_\alpha(\pi, s) \cdot j_Q \circ (\mathrm{Id} \otimes R_{Q|P}(\pi, s)) \tag{8}$$

where $R_{Q|P}(\pi, s) = \otimes_v R_{Q|P}(\pi_v, s)$ is the product of the locally defined normalized intertwining operators and $\pi = \otimes_v \pi_v$ [Ar3, § 6], (cf. [Mu6, (2.17)]). In many cases, the normalizing factors can be expressed in terms of automorphic L-functions [Sha1, Sha2]. For example, let $\mathbf{G} = \mathrm{GL}(n)$. Then the global normalizing factors n_α can be expressed in terms of Rankin-Selberg L-functions. The properties of these functions are collected and analyzed in [Mu4, Mu5, § 4,5]. Write $\mathbf{M} \simeq \prod_{i=1}^r \mathrm{GL}(n_i)$, where the root α is trivial on $\prod_{i\geq 3}\mathrm{GL}(n_i)$, and let $\pi \simeq \otimes\pi_i$ with representations $\pi_i \in \Pi_{\mathrm{disc}}(\mathrm{GL}(n_i, \mathbb{A}))$. Let $L(s, \pi_1 \times \tilde{\pi}_2)$ be the completed Rankin-Selberg L-function associated with π_1 and π_2. It satisfies the functional equation

$$L(s, \pi_1 \times \tilde{\pi}_2) = \epsilon(\frac{1}{2}, \pi_1 \times \tilde{\pi}_2)N(\pi_1 \times \tilde{\pi}_2)^{\frac{1}{2}-s}L(1-s, \tilde{\pi}_1 \times \pi_2) \tag{9}$$

where $|\epsilon(\frac{1}{2}, \pi_1 \times \tilde{\pi}_2)| = 1$ and $N(\pi_1 \times \tilde{\pi}_2) \in \mathbb{N}$ is the conductor. Then we have

$$n_\alpha(\pi, s) = \frac{L(s, \pi_1 \times \tilde{\pi}_2)}{\epsilon(\frac{1}{2}, \pi_1 \times \tilde{\pi}_2)N(\pi_1 \times \tilde{\pi}_2)^{\frac{1}{2}-s}L(s+1, \pi_1 \times \tilde{\pi}_2)}. \tag{10}$$

2.3 The Trace Formula

Arthur's trace formula gives two alternative expressions for a distribution J on $\mathbf{G}(\mathbb{A})^1$. Note that this distribution depends on the choice of $\mathbf{M}_0$ and $\mathbf{K}$. For $h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1)$, Arthur defines $J(h)$ as the value at the point $T = T_0$ specified in [Ar5, Lemma 1.1] of a polynomial $J^T(h)$ on $\mathfrak{a}_{M_0}$ of degree at most $d_0 = \dim \mathfrak{a}^G_{M_0}$. Here, the polynomial $J^T(h)$ depends in addition on the choice of a parabolic subgroup

$\mathbf{P}_0 \in \mathcal{P}(\mathbf{M}_0)$. Consider the equivalence relation on $\mathbf{G}(\mathbb{Q})$ defined by $\gamma \sim \gamma'$ whenever the semisimple parts of γ and γ' are $\mathbf{G}(\mathbb{Q})$-conjugate. Let $\mathcal{O}$ be the set of the resulting equivalence classes (which are in bijection with conjugacy classes of semisimple elements). The coarse geometric expansion [Ar1] is

$$J^T(h) = \sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}^T(h), \tag{11}$$

where the summands $J_{\mathfrak{o}}^T(h)$ are again polynomials in T of degree at most d_0. Write $J_{\mathfrak{o}}(h) = J_{\mathfrak{o}}^{T_0}(h)$, which depends only on $\mathbf{M}_0$ and $\mathbf{K}$. Then $J_{\mathfrak{o}}(h) = 0$ if the support of h is disjoint from all conjugacy classes of $\mathbf{G}(\mathbb{A})$ intersecting $\mathfrak{o}$ (cf. [Ar6, Theorem 8.1]). By [ibid., Lemma 9.1] (together with the descent formula of [Ar5, § 2]), for each compact set $\Omega \subset \mathbf{G}(\mathbb{A})^1$ there exists a finite subset $\mathcal{O}(\Omega) \subset \mathcal{O}$ such that for h supported in Ω only the terms with $\mathfrak{o} \in \mathcal{O}(\Omega)$ contribute to (11). In particular, the sum is always finite. The geometric side of the trace formula is then defined to be the distribution

$$J_{\mathrm{geo}}(h) = \sum_{\mathfrak{o}\in\mathcal{O}} J_{\mathfrak{o}}(h), \quad h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1). \tag{12}$$

When $\mathfrak{o}$ consists of the unipotent elements of $\mathbf{G}(\mathbb{Q})$, we write $J_{\mathrm{unip}}^T(h)$ for $J_{\mathfrak{o}}^T(h)$.

We now turn to the spectral side. Let $\mathbf{L} \supset \mathbf{M}$ be Levi subgroups in $\mathcal{L}$, $\mathbf{P} \in \mathcal{P}(\mathbf{M})$, and let $m = \dim \mathfrak{a}_L^G$ be the co-rank of $\mathbf{L}$ in $\mathbf{G}$. Denote by $\mathfrak{B}_{P,L}$ the set of m-tuples $\underline{\beta} = (\beta_1^\vee, \dots, \beta_m^\vee)$ of elements of $\Sigma_P^\vee$ whose projections to $\mathfrak{a}_L$ form a basis for $\mathfrak{a}_L^G$. For any $\underline{\beta} = (\beta_1^\vee, \dots, \beta_m^\vee) \in \mathfrak{B}_{P,L}$ let $\mathrm{vol}(\underline{\beta})$ be the co-volume in $\mathfrak{a}_L^G$ of the lattice spanned by $\underline{\beta}$ and let

$$\begin{aligned}\Xi_L(\underline{\beta}) &= \{(\mathbf{Q}_1, \dots, \mathbf{Q}_m) \in \mathcal{F}_1(M)^m : \ \beta_i^\vee \in \mathfrak{a}_M^{Q_i},\ i = 1, \dots, m\} \\ &= \{(\overline{\mathbf{P}_1\mathbf{P}_1'}, \dots, \overline{\mathbf{P}_m\mathbf{P}_m'}) : \ \mathbf{P}_i|^{\beta_i}\mathbf{P}_i',\ i = 1, \dots, m\}.\end{aligned}$$

For any smooth function f on $\mathfrak{a}_M^*$ and $\mu \in \mathfrak{a}_M^*$ denote by $D_\mu f$ the directional derivative of f along $\mu \in \mathfrak{a}_M^*$. For a pair $\mathbf{P}_1|^\alpha\mathbf{P}_2$ of adjacent parabolic subgroups in $\mathcal{P}(\mathbf{M})$ write

$$\delta_{P_1|P_2}(\lambda) = M_{P_2|P_1}(\lambda) D_\varpi M_{P_1|P_2}(\lambda) : \mathcal{A}^2(\mathbf{P}_2) \to \mathcal{A}^2(\mathbf{P}_2),$$

where $\varpi \in \mathfrak{a}_M^*$ is such that $\langle \varpi, \alpha^\vee \rangle = 1$. [1] Equivalently, writing $M_{P_1|P_2}(\lambda) = \Phi(\langle \lambda, \alpha^\vee \rangle)$ for a meromorphic function Φ of a single complex variable, we have

$$\delta_{P_1|P_2}(\lambda) = \Phi(\langle \lambda, \alpha^\vee \rangle)^{-1} \Phi'(\langle \lambda, \alpha^\vee \rangle).$$

[1]Note that this definition differs slightly from the definition of $\delta_{P_1|P_2}$ in [FL1].

For any m-tuple $\mathcal{X} = (\mathbf{Q}_1, \dots, \mathbf{Q}_m) \in \Xi_L(\underline{\beta})$ with $\mathbf{Q}_i = \overline{\mathbf{P}_i\mathbf{P}'_i}, \mathbf{P}_i|^{\beta_i}\mathbf{P}'_i$, denote by $\Delta_{\mathcal{X}}(\mathbf{P}, \lambda)$ the expression

$$\frac{\mathrm{vol}(\underline{\beta})}{m!} M_{P'_1|P}(\lambda)^{-1}\delta_{P_1|P'_1}(\lambda)M_{P'_1|P'_2}(\lambda)\cdots\delta_{P_{m-1}|P'_{m-1}}(\lambda)M_{P'_{m-1}|P'_m}(\lambda)\delta_{P_m|P'_m}(\lambda)M_{P'_m|P}(\lambda).$$

In [FLM1, pp. 179–180] we define a (purely combinatorial) map $\mathcal{X}_L : \mathfrak{B}_{P,L} \to \mathcal{F}_1(M)^m$ with the property that $\mathcal{X}_L(\underline{\beta}) \in \Xi_L(\underline{\beta})$ for all $\underline{\beta} \in \mathfrak{B}_{P,L}$. [2]

For any $s \in W(\mathbf{M})$ let $\mathbf{L}_s$ be the smallest Levi subgroup in $\mathcal{L}(\mathbf{M})$ containing w_s. We recall that $\mathfrak{a}_{L_s} = \{H \in \mathfrak{a}_M \mid sH = H\}$. Set

$$\iota_s = |\det(s-1)_{\mathfrak{a}_M^{L_s}}|^{-1}.$$

For $\mathbf{P} \in \mathcal{F}(\mathbf{M}_0)$ and $s \in W(\mathbf{M}_P)$ let $M(\mathbf{P}, s) : \mathcal{A}^2(\mathbf{P}) \to \mathcal{A}^2(\mathbf{P})$ be as in [Ar3, p. 1309]. $M(\mathbf{P}, s)$ is a unitary operator which commutes with the operators $\rho(\mathbf{P}, \lambda, h)$ for $\lambda \in \mathrm{i}\mathfrak{a}^*_{L_s}$. Now we can state the refined spectral expansion.

Theorem 2.1 ([FLM1]). *For any $h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1)$ the spectral side of Arthur's trace formula is given by*

$$J_{\mathrm{spec}}(h) := \sum_{[\mathbf{M}]} J_{\mathrm{spec},M}(h), \tag{13}$$

$[\mathbf{M}]$ ranging over the conjugacy classes of Levi subgroups of $\mathbf{G}$ (represented by members of $\mathcal{L}$), where

$$J_{\mathrm{spec},M}(h) = \frac{1}{|W(\mathbf{M})|}\sum_{s\in W(\mathbf{M})}\iota_s\sum_{\underline{\beta}\in\mathfrak{B}_{P,L_s}}\int_{\mathrm{i}(\mathfrak{a}_{L_s}^G)^*}\mathrm{tr}(\Delta_{\mathcal{X}_{L_s}(\underline{\beta})}(\mathbf{P},\lambda)M(\mathbf{P},s)\rho(\mathbf{P},\lambda,h))\,d\lambda \tag{14}$$

with $\mathbf{P} \in \mathcal{P}(\mathbf{M})$ arbitrary. The operators are of trace class and the integrals are absolutely convergent.

Note that the term corresponding to $\mathbf{M} = \mathbf{G}$ is $J_{\mathrm{spec},G}(h) = \mathrm{tr}\, R_{\mathrm{disc}}(h)$. Next assume that $\mathbf{M}$ is the Levi subgroup of a maximal parabolic subgroup $\mathbf{P}$. Furthermore, let $\mathbf{L} = \mathbf{M}$. Let $\bar{\mathbf{P}}$ be the opposite parabolic subgroup to $\mathbf{P}$. Then up to a constant, the contribution to the spectral side is given by

$$\sum_{\pi\in\Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A})^1)}\int_{i\mathfrak{a}^*}\mathrm{tr}\left(M_{\bar{P}|P}(\pi,\lambda)^{-1}\frac{d}{dz}M_{\bar{P}|P}(\pi,\lambda)M(\mathbf{P},s)\rho(\mathbf{P},\pi,\lambda,h)\right)\,d\lambda.$$

[2] The map $\mathcal{X}_L$ depends in fact on the additional choice of a vector $\underline{\mu} \in (\mathfrak{a}_M^*)^m$ which does not lie in an explicit finite set of hyperplanes. For our purposes, the precise definition of $\mathcal{X}_L$ is immaterial.

The trace formula is the statement that the spectral side equals the geometric side, i.e., the following equality holds:

$$J_{\text{spec}}(h) = J_{\text{geo}}(h), \quad h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1). \tag{15}$$

3 The Weyl Law

The Weyl law is concerned with the study of the asymptotic behavior of the counting function (2) as $\lambda \to \infty$. This is the first problem which needs to be solved in order to be able to pursue a deeper study of the cuspidal automorphic spectrum. For example, the study of statistical properties of the automorphic spectrum requires first of all to know that the spectrum is infinite and has the right asymptotic properties. This, in particular, concerns the study of families of automorphic forms (see [SST]).

The investigation of the asymptotic behavior of the counting function (2) is closely related to the study of the counting function of the eigenvalues of the Laplace operator on a compact Riemannian manifold. We briefly recall the Weyl law in this case. Let M be a smooth, compact Riemannian manifold of dimension n with smooth boundary ∂M (which may be empty). Let

$$\Delta = -\operatorname{div} \circ \operatorname{grad} = d^* d$$

be the Laplace-Beltrami operator associated with the metric g of M. We consider the Dirichlet eigenvalue problem

$$\Delta\phi = \lambda\phi, \quad \phi\big|_{\partial M} = 0. \tag{16}$$

As is well known, (16) has a discrete set of solutions

$$0 \le \lambda_0 \le \lambda_2 \le \cdots \to \infty$$

whose only accumulation point is at infinity and each eigenvalue occurs with finite multiplicity. The corresponding eigenfunctions ϕ_i can be chosen such that $\{\phi_i\}_{i\in\mathbb{N}_0}$ is an orthonormal basis of $L^2(M)$. For $\lambda \ge 0$ let

$$N(\lambda) = \#\{j : \lambda_j \le \lambda\}$$

be the counting function, where eigenvalues are counted with multiplicities. Let $\mathbf{\Gamma}(s)$ be the Gamma function. Then the Weyl law states

$$N(\lambda) = \frac{\operatorname{vol}(M)}{(4\pi)^{n/2}\mathbf{\Gamma}\left(\frac{n}{2}+1\right)}\lambda^{n/2} + o(\lambda^{n/2}), \quad \lambda \to \infty. \tag{17}$$

This was first proved by Weyl [We] for a bounded domain $\Omega \subset \mathbb{R}^3$. Written in a slightly different form it is known in physics as the Rayleigh-Jeans law. Garding [Ga] proved Weyl's law for a general elliptic operator on a domain in $\mathbb{R}^n$. For a closed Riemannian manifold (17) was proved by Minakshisundaram and Pleijel [MP]. Formula (17) does not say much about the finer structure of the distribution of the eigenvalues. A basic problem is the estimation of the remainder term

$$R(\lambda) := N(\lambda) - \frac{\operatorname{vol}(M)}{(4\pi)^{n/2}\Gamma\left(\frac{n}{2}+1\right)}\lambda^{n/2}. \tag{18}$$

For a closed Riemannian manifold, Avakumović [Av] established the Weyl law with the following optimal estimation of the remainder term

$$R(\lambda) = O(\lambda^{(n-1)/2}). \tag{19}$$

This result was extended to more general and higher order operators by Hörmander [Ho].

The connection with the estimation of the counting function (2) is established as follows. Let $\widetilde{X} = G/K$. It can be equipped with a G-invariant metric which is unique up to scaling. Let $X = \Gamma\backslash\widetilde{X}$. Assume that Γ is torsion free. Then X is a complete Riemannian manifold of finite volume. Let $\sigma \in \widehat{K}$ and let $\widetilde{E}_\sigma \to \widetilde{X}$ be the homogeneous vector bundle associated with σ, which is equipped with the invariant Hermitian metric induced by σ. Let $E_\sigma = \Gamma\backslash\widetilde{E}_\sigma$ be the corresponding locally homogeneous vector bundle over X. Let $C^\infty(X, E_\sigma)$ be the space of smooth sections of E_σ. There is a canonical isomorphism

$$C^\infty(X, E_\sigma) \cong (C^\infty(\Gamma\backslash G) \otimes V_\sigma)^K \tag{20}$$

(see [Mia, p. 4]). Let ∇^σ be the connection in E_σ induced by the canonical connection in $\widetilde{E}_\sigma$. Let $\Delta_\sigma = (\nabla^\sigma)^*\nabla^\sigma$ be the Bochner-Laplace operator, acting in $C^\infty(X, E_\sigma)$. It is an elliptic, second order, formally self-adjoint differential operator of Laplace type, i.e., its principal symbol is given by $\|\xi\|_x^2 \operatorname{Id}_{E_{\sigma,x}}$. Let $\Omega \in \mathcal{Z}(\mathfrak{g}_\mathbb{C})$ be the Casimir element and $R_\Gamma(\Omega)$ the Casimir operator acting in $C^\infty(\Gamma\backslash G)$. With respect to the isomorphism (20) the Bochner-Laplace operator is related to the Casimir operator $R_\Gamma(\Omega)$ by

$$\Delta_\sigma = -R_\Gamma(\Omega) + \lambda_\sigma \operatorname{Id}, \tag{21}$$

where λ_σ is the Casimir eigenvalue of σ. Assume that X is compact. Then Δ_σ has a pure discrete spectrum consisting of a sequence of eigenvalues $0 \le \lambda_1 \le \lambda_2 \le \cdots \to \infty$ of finite multiplicities. Let

$$N_\Gamma(\lambda; \sigma) = \#\{j : \lambda_j \le \lambda\}$$

be the counting function of the eigenvalues, where eigenvalues are counted with their multiplicity. Using (20) and (21), it follows that the counting function (2) has the same asymptotic behavior as $N_\Gamma(\lambda;\sigma)$. A generalization of (17) is the following Weyl law

$$N_\Gamma(\lambda;\sigma) = \frac{\dim(\sigma)\operatorname{vol}(\Gamma\backslash G/K)}{(4\pi)^{d/2}\mathbf{\Gamma}(d/2+1)}\lambda^{d/2} + o(\lambda^{d/2}), \quad \lambda \to \infty, \tag{22}$$

where $d = \dim(X)$. To prove (22) one can use the heat equation method [BGV, Gi]. It starts with the observation that the heat operator $e^{-t\Delta_\sigma}$ is an integral operator with a smooth kernel $K_\sigma(t,x,y)$. Since the underlying manifold is compact, it follows that the heat operator is a trace class operator and one has the following elementary "trace formula"

$$\sum_{j=1}^{\infty} e^{-t\lambda_j} = \operatorname{Tr}\left(e^{-t\Delta_\sigma}\right) = \int_X \operatorname{tr} K_\sigma(t,x,x)\,dx. \tag{23}$$

(see [BGV, Proposition 2.32]). The construction of an approximation of the heat kernel gives rise to an asymptotic expansion of the form

$$\int_X \operatorname{tr} K_\sigma(t,x,x)\,dx \sim t^{-d/2}\sum_{j=0}^{\infty} a_j t^j \tag{24}$$

as $t \to 0^+$. Moreover $a_0 = \dim(\sigma)\operatorname{vol}(X)/(4\pi)^{d/2}$ (see [BGV, Theorem 2.30], [Gi, Chap. 1, § 1.7]). Combined with (23), it follows that

$$\sum_{j=1}^{\infty} e^{-t\lambda_j} = \frac{\dim(\sigma)\operatorname{vol}(\Gamma\backslash G/K)}{(4\pi)^{d/2}} t^{-d/2} + O(t^{-d/2+1}) \tag{25}$$

as $t \to 0^+$. Applying Karamata's theorem [BGV, Theorem 2.42], we obtain the Weyl law (22). The heat equation method does not lead to any nontrivial estimation of the remainder term. The method of Avakumović [Av] and Hörmander [Ho] is based on the study of the wave equation (see [DG]). For a locally symmetric manifold this means to use the Selberg trace formula. So far estimations of the remainder term are only known if σ is the trivial representation, i.e., for the case of the Laplace operator on functions.

For a locally symmetric space $X = \Gamma\backslash\widetilde{X}$, $\widetilde{X} = G/K$, there is not only the Laplace operator, but the whole algebra of G-invariant differential operators $\mathcal{D}(\widetilde{X})$ on $\widetilde{X}$, which one needs to consider. The structure of $\mathcal{D}(\widetilde{X})$ can be described as follows. Let $G = NAK$ be the Iwasawa decomposition of G, W the Weyl group of (G,A), and $\mathfrak{a}$ be the Lie algebra of A. Let $S(\mathfrak{a}_\mathbb{C})$ be the symmetric algebra of the complexification $\mathfrak{a}_\mathbb{C} = \mathfrak{a} \otimes \mathbb{C}$ of $\mathfrak{a}$ and let $S(\mathfrak{a}_\mathbb{C})^W$ be the subspace of Weyl group invariants in

$S(\mathfrak{a}_{\mathbb{C}})$. Then by a theorem of Harish-Chandra [He, Chap. X, Theorem 6.15] there is a canonical isomorphism

$$\mu : \mathcal{D}(\widetilde{X}) \cong S(\mathfrak{a}_{\mathbb{C}})^W. \tag{26}$$

This shows that $\mathcal{D}(\widetilde{X})$ is commutative. The minimal number of generators equals the rank of $\widetilde{X}$ which is $\dim \mathfrak{a}$ [He, Chap. X, § 6.3]. Let $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$. Then by (26), λ determines an character

$$\chi_\lambda : \mathcal{D}(\widetilde{X}) \to \mathbb{C}$$

and $\chi_\lambda = \chi_{\lambda'}$ if and only if λ and λ' are in the same W-orbit. Since $S(\mathfrak{a}_{\mathbb{C}})$ is integral over $S(\mathfrak{a}_{\mathbb{C}})^W$ [He, Chap. X, Lemma 6.9], each character of $\mathcal{D}(\widetilde{X})$ is of the form χ_λ for some $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$. Thus the characters of $\mathcal{D}(\widetilde{X})$ are parametrized by $\mathfrak{a}_{\mathbb{C}}^*/W$.

Each $D \in \mathcal{D}(\widetilde{X})$ descends to a differential operator

$$D : C^\infty(\Gamma \backslash \widetilde{X}) \to C^\infty(\Gamma \backslash \widetilde{X}).$$

Assume that $\Gamma \backslash \widetilde{X}$ is compact. Let $\mathcal{E} \subset C^\infty(\Gamma \backslash \widetilde{X})$ be an eigenspace of the Laplace operator. Then $\mathcal{E}$ is a finite-dimensional vector space which is invariant under $D \in \mathcal{D}(\widetilde{X})$. For each $D \in \mathcal{D}(\widetilde{X})$, the formal adjoint D^* of D also belongs to $\mathcal{D}(\widetilde{X})$. Thus we get a representation

$$\rho : \mathcal{D}(\widetilde{X}) \to \operatorname{End}(\mathcal{E})$$

by commuting normal operators. Therefore, $\mathcal{E}$ decomposes into the direct sum of joint eigenspaces of $\mathcal{D}(\widetilde{X})$. Given $\lambda \in \mathfrak{a}_{\mathbb{C}}^*/W$, let

$$\mathcal{E}(\lambda) = \{\varphi \in C^\infty(\Gamma \backslash \widetilde{X}) : D\varphi = \chi_\lambda(D)\varphi,\ D \in \mathcal{D}(\widetilde{X})\}.$$

Let $m(\lambda) = \dim \mathcal{E}(\lambda)$. Then the spectrum $\Lambda(\Gamma)$ of $\Gamma \backslash \widetilde{X}$ is defined to be

$$\Lambda(\Gamma) = \{\lambda \in \mathfrak{a}_{\mathbb{C}}^*/W : m(\lambda) > 0\},$$

and we get an orthogonal direct sum decomposition

$$L^2(\Gamma \backslash \widetilde{X}) = \bigoplus_{\lambda \in \Lambda(\Gamma)} \mathcal{E}(\lambda). \tag{27}$$

If we pick a fundamental domain for W, we may regard $\Lambda(\Gamma)$ as a subset of $\mathfrak{a}_{\mathbb{C}}^*$. If $\operatorname{rank}(\widetilde{X}) > 1$, then $\Lambda(\Gamma)$ is multidimensional. In this setting, a generalization of the Weyl law has been established by Duistermaat et al. [DKV]. To describe the result, we need to introduce some notations. Let $\beta(i\lambda)$, $\lambda \in \mathfrak{a}^*$, be the Plancherel density. Let

$$\Lambda_{\text{temp}}(\Gamma) = \Lambda(\Gamma) \cap i\mathfrak{a}^*, \quad \Lambda_{\text{comp}}(\Gamma) = \Lambda(\Gamma) \setminus \Lambda_{\text{temp}}(\Gamma)$$

be the tempered and complementary spectrum, respectively. Given an open bounded subset Ω of $\mathfrak{a}^*$ and $t > 0$, let

$$t\Omega := \{t\mu \colon \mu \in \Omega\}. \tag{28}$$

One of the main results of [DKV] is the following asymptotic formula for the distribution of the tempered spectrum [DKV, Theorem 8.8]

$$\sum_{\lambda \in \Lambda_{\mathrm{temp}}(\Gamma) \cap (it\Omega)} m(\lambda) = \frac{\mathrm{vol}(\Gamma\backslash\widetilde{X})}{|W|} \int_{it\Omega} \beta(\lambda)\, d\lambda + O(t^{d-1}), \quad t \to \infty. \tag{29}$$

Note that the leading term is of order $O(t^d)$. The growth of the complementary spectrum is of lower order. Let $B_t(0) \subset \mathfrak{a}^*_{\mathbb{C}}$ be the ball of radius $t > 0$ around 0. There exists $C > 0$ such that for all $t \geq 1$

$$\sum_{\lambda \in \Lambda_{\mathrm{comp}}(\Gamma) \cap B_t(0)} m(\lambda) \leq C t^{d-2} \tag{30}$$

[DKV, Theorem 8.3]. The main tool to prove (29) and (30) is the Selberg trace formula.

The estimations (29) and (30) contain more information about the distribution of $\Lambda(\Gamma)$ then just the Weyl law. Indeed, the eigenvalue of Δ corresponding to $\lambda \in \Lambda_{\mathrm{temp}}(\Gamma)$ equals $\| \lambda \|^2 + \| \rho \|^2$. So if we choose Ω in (29) to be the unit ball, then (29) together with (30) reduces to Weyl's law for $\Gamma\backslash\widetilde{X}$.

We note that (29) and (30) can also be rephrased in terms of representation theory. Let R_Γ be the right regular representation of G in $L^2(\Gamma\backslash G)$ defined by

$$(R_\Gamma(g_1)f)(g_2) = f(g_2 g_1), \quad f \in L^2(\Gamma\backslash G),\ g_1, g_2 \in G.$$

Let $\Pi(G)$ denote the set equivalence classes of unitary irreducible representations of G. Since $\Gamma\backslash G$ is compact, R_Γ decomposes into the direct sum of irreducible unitary representations of G (see [GGP, § 2.3]). Given $\pi \in \Pi(G)$, let $m(\pi)$ be the multiplicity with which π occurs in R_Γ. Let $\mathcal{H}_\pi$ denote the Hilbert space in which π acts. Then

$$L^2(\Gamma\backslash G) \cong \bigoplus_{\pi \in \Pi(G)} m(\pi) \mathcal{H}_\pi.$$

Now observe that $L^2(\Gamma\backslash\widetilde{X}) = L^2(\Gamma\backslash G)^K$. Let $\mathcal{H}_\pi^K$ denote the subspace of K-fixed vectors in $\mathcal{H}_\pi$. Then

$$L^2(\Gamma\backslash\widetilde{X}) \cong \bigoplus_{\pi \in \Pi(G)} m(\pi) \mathcal{H}_\pi^K.$$

Note that $\dim \mathcal{H}_\pi^K \leq 1$. Let $\Pi^{\mathrm{sph}}(G) \subset \Pi(G)$ be the subset of all π with $\mathcal{H}_\pi^K \neq \{0\}$. This is the spherical dual. Given $\pi \in \Pi^{\mathrm{sph}}(G)$, let λ_π be the infinitesimal character of π. If $\pi \in \Pi^{\mathrm{sph}}(G)$, then $\lambda_\pi \in \mathfrak{a}_{\mathbb{C}}^*/W$. Moreover $\pi \in \Pi^{\mathrm{sph}}(G)$ is tempered, if π is unitarily induced from the minimal parabolic subgroup $P = NAM$. In this case we have $\lambda_\pi \in i\mathfrak{a}^*/W$. So (29) can be rewritten as

$$\sum_{\substack{\pi \in \Pi^{\mathrm{sph}}(G) \\ \lambda_\pi \in it\Omega}} m(\pi) = \frac{\mathrm{vol}(\Gamma\backslash G)}{|W|} \int_{it\Omega} \beta(\lambda)\, d\lambda + O(t^{n-1}), \quad t \to \infty. \tag{31}$$

If Γ is not co-compact, then Δ_σ has a nonempty continuous spectrum which consists of a half-line $[c, \infty)$ for some $c \geq 0$. This makes it much more difficult to study the discrete spectrum of this operator, because almost all eigenvalues, if they exist, will be embedded into the continuous spectrum. It is well known from mathematical physics that embedded eigenvalues are unstable under perturbations. One of the basic tools to study the cuspidal automorphic spectrum is the trace formula.

3.1 Rank One

In the non-compact case, a general Weyl law was first derived by Selberg for a hyperbolic surface $X = \Gamma\backslash\mathbb{H}$ of finite area, where $\mathbb{H} = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ is the upper half-plane. We briefly recall the method which is based on the trace formula. It illustrates the basic idea which is also used in the higher rank case.

Let $\Delta = d^*d$ be the Laplace operator with respect to the hyperbolic metric. Then Δ, regarded as operator in $L^2(X)$ with domain $C^\infty(X)$, is essentially self-adjoint. The spectrum of Δ is the union of a pure point spectrum and the absolutely continuous spectrum. The pure point spectrum consists of a sequence of eigenvalues

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$$

of finite multiplicities. If X is non-compact then, in general, we only know that λ_0 exists. We slightly change the definition of the counting function by

$$N_\Gamma(\lambda) := \#\{j\colon \sqrt{\lambda_j} \leq \lambda\}.$$

The new terms in the trace formula, which are due to the non-compactness of $\Gamma\backslash\mathbb{H}$ arise from the parabolic conjugacy classes in Γ and the Eisenstein series. Let us recall the definition of Eisenstein series. Let $a_1, \ldots, a_m \in \mathbb{R} \cup \{\infty\}$ be representatives of the Γ-conjugacy classes of parabolic fixed points of Γ. The a_i's are called *cusps*. For each a_i let Γ_{a_i} be the stabilizer of a_i in Γ. Choose $\sigma_i \in \mathrm{SL}(2,\mathbb{R})$ such that

$$\sigma_i(\infty) = a_i, \quad \sigma_i^{-1}\Gamma_{a_i}\sigma_i = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{Z} \right\}.$$

Then the Eisenstein series $E_i(z,s)$ associated with the cusp a_i is defined as

$$E_i(z,s) = \sum_{\gamma \in \Gamma_{a_i} \backslash \Gamma} \mathrm{Im}(\sigma_i^{-1}\gamma z)^s, \quad \mathrm{Re}(s) > 1. \tag{32}$$

The series converges absolutely and uniformly on compact subsets of the half-plane $\mathrm{Re}(s) > 1$ and it satisfies the following properties.

(1) $E_i(\gamma z, s) = E_i(z,s)$ for all $\gamma \in \Gamma$.
(2) As a function of s, $E_i(z,s)$ admits a meromorphic continuation to $\mathbb{C}$ which is regular on the line $\mathrm{Re}(s) = 1/2$.
(3) $E_i(z,s)$ is a smooth function of z and satisfies $\Delta_z E_i(z,s) = s(1-s)E_i(z,s)$.

The contribution of the Eisenstein series to the Selberg trace formula is given by their zeroth Fourier coefficients of the Fourier expansion in the cusps. The zeroth Fourier coefficient of the Eisenstein series $E_k(z,s)$ at the cusp a_l is given by

$$\int_0^1 E_k(\sigma_l(x+iy),s)\,dx = \delta_{kl}y^s + C_{kl}(s)y^{1-s},$$

where δ_{kl} is Kronecker's delta function and $C_{kl}(s)$ is a meromorphic function of $s \in \mathbb{C}$. Put

$$C(s) := (C_{kl}(s))_{k,l=1}^m .$$

This is the so-called *scattering matrix*. Let $g \in C_c^\infty(\mathbb{R})$ and let $h = \hat{g}$ be the Fourier transform of g. Let $\phi(s) := \det C(s)$. Denote by $\{\gamma\}$ the hyperbolic Γ-conjugacy classes. For every hyperbolic element γ, denote by γ_0 the primitive hyperbolic element such that $\gamma = \gamma_0^k$ for some $k \in \mathbb{N}$. Every nontrivial hyperbolic conjugacy class $\{\gamma\}$ corresponds to a unique closed geodesic c_γ. Let $l(\gamma)$ denote its length. Write the eigenvalues as

$$\lambda_j = \frac{1}{4} + r_j^2, \quad r_j \in [0,\infty) \cup i(0,1/2].$$

Then the trace formula is the following identity

$$\sum_j h(r_j) - \frac{1}{4\pi}\int_{-\infty}^{\infty} h(r)\frac{\phi'}{\phi}(1/2+ir)\,dr + \frac{1}{4}\phi(1/2)h(0)$$
$$= \frac{\operatorname{Area}(\Gamma\backslash\mathbb{H})}{4\pi}\int_{\mathbb{R}} h(r)r\tanh(\pi r)\,dr + \sum_{\{\gamma\}}\frac{l(\gamma_0)}{2\sinh\left(\frac{l(\gamma)}{2}\right)}g(l(\gamma))$$
$$-\frac{m}{2\pi}\int_{-\infty}^{\infty} h(r)\frac{\mathbf{\Gamma}'}{\mathbf{\Gamma}}(1+ir)dr + \frac{m}{4}h(0) - m\ln 2\, g(0) \qquad (33)$$

(see [Se1, (9.31)]). The left-hand side is the spectral side, which contains all terms associated with the spectrum and the right-hand side is the geometric side. The trace formula holds for every discrete subgroup $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ with co-finite area. In analogy to the counting function of the eigenvalues we introduce the winding number

$$M_\Gamma(\lambda) = -\frac{1}{4\pi}\int_{-\lambda}^{\lambda}\frac{\phi'}{\phi}(1/2+ir)\,dr, \qquad (34)$$

which measures the continuous spectrum. Using the cut-off Laplacian of Lax-Phillips [CV] one can deduce the following elementary bounds

$$N_\Gamma(\lambda) \ll \lambda^2, \quad M_\Gamma(\lambda) \ll \lambda^2, \quad \lambda \geq 1. \qquad (35)$$

These bounds imply that the trace formula (33) holds for a larger class of functions. In particular, it can be applied to the heat kernel k_t. Its spherical Fourier transform equals $h_t(r) = e^{-t(1/4+r^2)}$, $t > 0$. If we insert h_t into the trace formula, we get the following asymptotic expansion as $t \to 0$.

$$\sum_j e^{-t\lambda_j} - \frac{1}{4\pi}\int_{\mathbb{R}} e^{-t(1/4+r^2)}\frac{\phi'}{\phi}(1/2+ir)\,dr$$
$$= \frac{\operatorname{Area}(\Gamma\backslash\mathbb{H})}{4\pi t} + \frac{a\log t}{\sqrt{t}} + \frac{b}{\sqrt{t}} + O(1) \qquad (36)$$

for certain constants $a, b \in \mathbb{R}$. Using [Se1, (8.8), (8.9)] it follows that the winding number $M_\Gamma(\lambda)$ is monotonically increasing for $\lambda \gg 0$. Therefore we can apply a Tauberian theorem to (36) and we get the following Weyl law, established by Selberg [Se1]. As $\lambda \to \infty$ we have

$$N_\Gamma(\lambda) + M_\Gamma(\lambda) \sim \frac{\operatorname{Area}(\Gamma\backslash\mathbb{H})}{4\pi}\lambda^2. \qquad (37)$$

In general, we cannot estimate separately the counting function and the winding number. For congruence subgroups, however, the entries of the scattering matrix can be expressed in terms of well-known analytic functions. For $\Gamma(N)$ the determinant of the scattering matrix $\phi(s)$ has been computed by Huxley [Hu]. It has the form

$$\phi(s) = (-1)^l A^{1-2s} \left(\frac{\boldsymbol{\Gamma}(1-s)}{\boldsymbol{\Gamma}(s)} \right)^k \prod_\chi \frac{L(2-2s, \bar{\chi})}{L(2s, \chi)}, \tag{38}$$

where $k, l \in \mathbb{Z}$, $A > 0$, the product runs over Dirichlet characters χ to some modulus dividing N and $L(s, \chi)$ is the Dirichlet L-function with character χ. Especially for $\Gamma(1)$ we have

$$\phi(s) = \sqrt{\pi} \frac{\boldsymbol{\Gamma}(s-1/2)\zeta(2s-1)}{\boldsymbol{\Gamma}(s)\zeta(2s)}, \tag{39}$$

where $\zeta(s)$ denotes the Riemann zeta function.

Using Stirling's approximation formula to estimate the logarithmic derivative of the Gamma function and standard estimations for the logarithmic derivative of Dirichlet L-functions on the line $\mathrm{Re}(s) = 1$ [Pr, Chap. V, Theorem 7.1], we get

$$\frac{\phi'}{\phi}(1/2 + ir) = O(\log(4 + |r|)), \quad |r| \to \infty. \tag{40}$$

This implies that

$$M_{\Gamma(N)}(\lambda) \ll \lambda \log \lambda. \tag{41}$$

Together with (37) we obtain Weyl's law for the point spectrum of the Laplacian on $X(N) = \Gamma(N)\backslash\mathbb{H}$:

$$N_{\Gamma(N)}(\lambda) \sim \frac{\mathrm{Area}(X(N))}{4\pi}\lambda^2, \quad \lambda \to \infty, \tag{42}$$

which is due to Selberg [Se1, p. 668]. A similar formula holds for other congruence groups such as $\Gamma_0(N)$. In particular, (42) implies that for congruence groups there exist infinitely many linearly independent Maass cusp forms.

By a more sophisticated use of the Selberg trace formula one can estimate the remainder term (see [Mu7]). For congruence subgroups one gets

Theorem 3.1. *For every $N \in \mathbb{N}$ we have*

$$N_{\Gamma(N)}(\lambda) = \frac{\mathrm{Area}(X(N))}{4\pi}\lambda^2 + O(\lambda \log \lambda) \tag{43}$$

as $\lambda \to \infty$.

A finite area hyperbolic surface for which the Weyl law holds is called by Sarnak *essentially cuspidal*. Now it is strongly believed that essential cuspidality is limited to special arithmetic surfaces. This is based on work by Phillips and Sarnak who studied the behavior of the discrete spectrum when Γ is deformed in the corresponding Teichmüller space. We refer to [Sa1] for a detailed discussion of their method. This led Phillips and Sarnak to the following conjectures.

Conjecture 1. *(1) The generic Γ in a given Teichmüller space of finite area hyperbolic surfaces is not essentially cuspidal.*

(2) Except for the Teichmüller space of the once punctured torus, the generic Γ has only a finite number of discrete eigenvalues.

Reznikov [Rez] has extended the method described above to deal with arithmetic quotients of rank one globally symmetric spaces. He has shown that for congruence quotients the determinant of the scattering matrix can be expressed as a ratio of automorphic L-functions. Using the properties of the L-functions, it follows that the determinant of the scattering matrix is a meromorphic function of order one. As above, this implies the following theorem.

Theorem 3.2 ([Rez]). *Any congruence subgroup of the unit group of a rational quadratic form in the group of motions of the hyperbolic space is essentially cuspidal.*

A similar result holds for congruence quotients of the complex hyperbolic space.

3.2 Higher Rank

We turn now to the general case. We assume that $G = \mathbf{G}(\mathbb{R})$, where $\mathbf{G}$ is a connected semisimple algebraic group over $\mathbb{Q}$. Let $X = \Gamma\backslash\widetilde{X} = \Gamma\backslash G/K$ and $E_\sigma \to X$ be as above. Let $\Delta_\sigma\colon C^\infty(X, E_\sigma) \to C^\infty(X, E_\sigma)$ be the Bochner-Laplace operator. As operator in $L^2(X, E_\sigma)$ it is essentially self-adjoint. Let $L^2_{\text{disc}}(X, E_\sigma)$ be the subspace of $L^2(X, E_\sigma)$ which is the closure of the span of all L^2-eigensections of Δ_σ. Recall that a cusp form for Γ is a smooth K-finite function $\phi\colon \Gamma\backslash G \to \mathbb{C}$ which is a joint eigenfunction of the center of the universal enveloping algebra $\mathcal{Z}(\mathfrak{g}_\mathbb{C})$ and which satisfies

$$\int_{\Gamma\cap N_P\backslash N_P} \phi(nx)\, dn = 0$$

for all unipotent radicals N_P of proper parabolic subgroups P of G, which are of the form $P = \mathbf{P}(\mathbb{R})$ for a rational parabolic subgroup $\mathbf{P}$ of $\mathbf{G}$. Put

$$L^2_{\text{cus}}(X, E_\sigma) := (L^2_{\text{cus}}(\Gamma\backslash G) \otimes V_\sigma)^K.$$

Then $L^2_{\text{cus}}(X, E_\sigma)$ is contained in $L^2_{\text{disc}}(X, E_\sigma)$. The orthogonal complement $L^2_{\text{res}}(X, E_\sigma)$ of $L^2_{\text{cus}}(X, E_\sigma)$ in $L^2_{\text{disc}}(X, E_\sigma)$ is called the *residual subspace*. By Langland's theory of Eisenstein series it follows that $L^2_{\text{res}}(X, E_\sigma)$ is spanned by iterated residues of cuspidal Eisenstein series (see [La2]). By definition we have an orthogonal decomposition

$$L^2_{\text{disc}}(X, E_\sigma) = L^2_{\text{cus}}(X, E_\sigma) \oplus L^2_{\text{res}}(X, E_\sigma).$$

Let $N^{\text{disc}}_\Gamma(\lambda; \sigma)$, $N^{\text{cus}}_\Gamma(\lambda; \sigma)$, and $N^{\text{res}}_\Gamma(\lambda; \sigma)$ be the counting function of the eigenvalues with eigensections belonging to the corresponding subspace. The following results about the growth of the counting functions hold for any lattice Γ in a real semisimple Lie group. Let $d = \dim X$. Donnelly [Do] has proved the following bound for the cuspidal spectrum

$$\limsup_{\lambda \to \infty} \frac{N^{\text{cus}}_\Gamma(\lambda, \sigma)}{\lambda^{d/2}} \le \frac{\dim(\sigma)\operatorname{vol}(X)}{(4\pi)^{d/2}\mathbf{\Gamma}\left(\frac{d}{2}+1\right)}. \tag{44}$$

For the full discrete spectrum, we have at least an upper bound for the growth of the counting function. The main result of [Mu2] states that

$$N^{\text{disc}}_\Gamma(\lambda, \sigma) \ll (1 + \lambda^{2d}). \tag{45}$$

This result implies that invariant integral operators are of trace class on the discrete subspace which is the starting point for the trace formula. The proof of (45) relies on the description of the residual subspace in terms of iterated residues of Eisenstein series.

Let $N^{\text{cus}}_\Gamma(\lambda)$ be the counting function with respect to the trivial representation of K, i.e., the counting function of the cuspidal spectrum of the Laplacian on functions. Then Sarnak [Sa2] conjectured that if $\operatorname{rank}(G/K) > 1$, Weyl's law holds for $N^{\text{cus}}_\Gamma(\lambda)$, which means that equality holds in (44). Furthermore, one expects that the growth of the residual spectrum is of lower order than the cuspidal spectrum.

In the meantime Sarnak's conjecture has been verified in quite a number of cases. A. Reznikov proved it for congruence groups in a group G of real rank one, Miller [Mi] proved it for $\mathbf{G} = \mathrm{SL}(3)$ and $\Gamma = \mathrm{SL}(3, \mathbb{Z})$, the author [Mu5] established it for $\mathbf{G} = \mathrm{SL}(n)$ and a congruence subgroup Γ. The most general result is due to Lindenstrauss and Venkatesh [LV] who proved the following theorem.

Theorem 3.3. *Let* $\mathbf{G}$ *be a split adjoint semi-simple group over* $\mathbb{Q}$ *and let* $\Gamma \subset \mathbf{G}(\mathbb{Q})$ *be a congruence subgroup. Let* $d = \dim S$. *Then*

$$N^{\text{cus}}_\Gamma(\lambda) \sim \frac{\operatorname{vol}(\Gamma\backslash\widetilde{X})}{(4\pi)^{d/2}\mathbf{\Gamma}\left(\frac{d}{2}+1\right)}\lambda^{d/2}, \quad \lambda \to \infty. \tag{46}$$

The method used by Lindenstrauss and Venkatesh is based on the construction of convolution operators with pure cuspidal image. It avoids the delicate estimates of the contributions of the Eisenstein series to the trace formula. This proves existence of many cusp forms for these groups.

For an arbitrary K-type, we have the following theorem proved in [Mu3].

Theorem 3.4. *Let $n \geq 2$ and $\widetilde{X} = \mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$. Let $d = \dim \widetilde{X} = n(n+1)/2-1$. For every principal congruence subgroup Γ of $\mathrm{SL}(n,\mathbb{Z})$ and every irreducible unitary representation σ of $\mathrm{SO}(n)$ such that $\sigma|_{Z_\gamma} = \mathrm{Id}$, we have*

$$N_\Gamma^{\mathrm{cus}}(\lambda,\sigma) \sim \frac{\dim(\sigma)\operatorname{vol}(\Gamma\backslash\widetilde{X})}{(4\pi)^{d/2}\mathbf{\Gamma}(d/2+1)}\lambda^{d/2} \tag{47}$$

as $\lambda \to \infty$.

The residual spectrum for $\mathrm{SL}(n)$ has been described by Moeglin and Waldspurger [MW]. Combined with (44) it follows that for $\mathbf{G} = \mathrm{SL}(n)$ we have

$$N_{\Gamma(N)}^{\mathrm{res}}(\lambda,\sigma) \ll \lambda^{d/2-1}, \tag{48}$$

where $d = \dim \mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$ and $\Gamma(N) \subset \mathrm{SL}(n,\mathbb{Z})$ is the principal congruence subgroup of level N.

The proof of Theorem 3.4 uses the Arthur trace formula combined with the heat equation method similar to the proof of (42). The application of the Arthur trace formula requires the adelic reformulation of the problem.

We briefly describe the method. For all details we refer to [Mu5]. For simplicity we consider only the trivial K_∞-type, i.e., we consider the counting function $N_\Gamma^{\mathrm{cus}}(\lambda)$. By (48) we can replace the counting function $N_\Gamma^{\mathrm{cus}}(\lambda)$ by $N_\Gamma^{\mathrm{disc}}(\lambda)$. Let $\mathbf{G} = \mathrm{GL}(n)$ be regarded as an algebraic group over $\mathbb{Q}$. Denote by A_G the split component of the center of $\mathbf{G}$ and let $A_G(\mathbb{R})^0$ be the component of 1 in $A_G(\mathbb{R})$. Let $\Pi_{\mathrm{disc}}(\mathbf{G}(\mathbb{A}),\xi_0)$ be the set of all irreducible subrepresentations of the regular representation of $G(\mathbb{A})$ in $L^2(\mathbf{G}(\mathbb{Q})A_G(\mathbb{R})^0\backslash\mathbf{G}(\mathbb{A}))$. Given a representation $\pi \in \Pi_{\mathrm{disc}}(\mathbf{G}(\mathbb{A}),\xi_0)$, let $m(\pi)$ denote the multiplicity with which π occurs in $L^2(\mathbf{G}(\mathbb{Q})A_G(\mathbb{R})^0\backslash\mathbf{G}(\mathbb{A}))$. For any irreducible representation $\pi = \pi_\infty \otimes \pi_f$ of $\mathbf{G}(\mathbb{A})$, let $\mathcal{H}_{\pi_\infty}$ and $\mathcal{H}_{\pi_f}$ denote the Hilbert space of the representation π_∞ and π_f, respectively. Let K_f be an open compact subgroup of $\mathbf{G}(\mathbb{A}_f)$. Denote by $\mathcal{H}_{\pi_f}^{K_f}$ the subspace of K_f-invariant vectors in $\mathcal{H}_{\pi_f}$ and by $\mathcal{H}_{\pi_\infty}^{K_\infty}$ the subspace of K_∞-invariant vectors in $\mathcal{H}_{\pi_\infty}$. Given $\pi \in \Pi(\mathbf{G}(\mathbb{A}),\xi_0)$, denote by λ_{π_∞} the Casimir eigenvalue of the restriction of π_∞ to $\mathbf{G}(\mathbb{R})^1$. Assume that $-1 \neq K_f$. Then (47) for the trivial K_∞-type follows by Karamata's theorem [BGV, Theorem 2.42] from the existence of an asymptotic expansion of the form

$$\sum_{\pi\in\Pi_{\mathrm{disc}}(\mathbf{G}(\mathbb{A}),\xi_0)} m(\pi)e^{t\lambda_{\pi_\infty}}\dim\big(\mathcal{H}_{\pi_f}^{K_f}\big)\dim\big(\mathcal{H}_{\pi_\infty}^{K_\infty}\big) \sim \frac{\operatorname{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1/K_f)}{(4\pi)^{d/2}}t^{-d/2} \tag{49}$$

as $t \to +0$.

To establish (49) we apply the Arthur trace formula as follows. We choose a certain family of test functions $\tilde{\phi}_t^1 \in C_c^\infty(\mathbf{G}(\mathbb{A})^1)$, depending on $t > 0$, which at the infinite place are given by the heat kernel $h_t \in C^\infty(\mathbf{G}(\mathbb{R})^1)$ of the Laplacian $\widetilde{\Delta}$ on $\widetilde{X}$, multiplied by a certain cutoff function φ_t, and which at the finite places is given by the normalized characteristic function of an open compact subgroup K_f of $\mathbf{G}(\mathbb{A}_f)$. Then by the non-invariant trace formula [Ar1] we have the equality

$$J_{\text{spec}}(\tilde{\phi}_t^1) = J_{\text{geo}}(\tilde{\phi}_t^1), \quad t > 0.$$

Then we study asymptotic behavior of the spectral and the geometric side as $t \to 0$. To deal with the geometric side, we use the fine $\mathfrak{o}$-expansion [Ar6]

$$J_{\text{geo}}(f) = \sum_{\mathbf{M}\in\mathcal{L}} \sum_{\gamma\in(\mathbf{M}(\mathbb{Q}_S))_{M,S}} a^M(S,\gamma) J_M(\gamma, f), \tag{50}$$

which expresses the distribution $J_{\text{geo}}(f)$ in terms of weighted orbital integrals $J_M(\gamma, f)$. Here $\mathbf{M}$ runs over the set of Levi subgroups $\mathcal{L}$ containing the Levi component $\mathbf{M}_0$ of the standard minimal parabolic subgroup $\mathbf{P}_0$, S is a finite set of places of $\mathbb{Q}$, and $(\mathbf{M}(\mathbb{Q}_S))_{M,S}$ is a certain set of equivalence classes in $\mathbf{M}(\mathbb{Q}_S)$. This reduces our problem to the investigation of weighted orbital integrals. The key result is that

$$\lim_{t\to 0} t^{d/2} J_M(\tilde{\phi}_t^1, \gamma) = 0,$$

unless $\mathbf{M} = \mathbf{G}$ and $\gamma = 1$. This follows from the description of the local weighted orbital integrals by [Ar4, Corollary 6.2]. The contributions to (50) of the terms where $\mathbf{M} = \mathbf{G}$ and $\gamma = 1$ are easy to determine. Using the behavior of the heat kernel $h_t(1)$ as $t \to 0$, it follows that

$$J_{\text{geo}}(\tilde{\phi}_t^1) \sim \frac{\text{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1/K_f)}{(4\pi)^{d/2}} t^{-d/2} \tag{51}$$

as $t \to 0$. To deal with the spectral side we use Theorem 2.1. This theorem allows us to replace $\tilde{\phi}_t^1$ by a similar function $\phi_t^1 \in \mathcal{C}^1(G(\mathbb{A})^1)$ which is given as the product of the heat kernel h_t at infinity and the normalized characteristic function of K_f. The term in $J_{\text{spec}}(\phi_t^1)$ corresponding to $\mathbf{M} = \mathbf{G}$ is $J_{\text{spec},G}(\phi_t^1) = \operatorname{tr} R_{\text{disc}}(\phi_t^1)$, which is equal to the left-hand side of (49). If $\mathbf{M}$ is a proper Levi subgroup of $\mathbf{G}$, then $J_{\text{spec},M}(\phi_t^1)$ is given by (14), which is a finite sum of integrals. The main ingredient of the integrals are logarithmic derivatives of intertwining operators and the estimation of these integrals is reduced to the estimation of the logarithmic derivatives. Using (8) this problem is reduced to the estimation of the logarithmic derivatives of the normalizing factors and the local intertwining operators. In the case of $\mathbf{G} = \mathrm{GL}(n)$, the normalizing factors are expressed in terms of Ranking-Selberg L-functions (10). Using the analytic properties of

Rankin-Selberg L-functions, it follows that there exist $C > 0$ and $T > 1$ such that for $\pi = \pi_1 \otimes \pi_2$, $\pi_i \in \Pi_{\mathrm{disc}}(\mathrm{GL}(n_i, \mathbb{A}))$, we have

$$\int_T^{T+1} \left| \frac{n'_\alpha(\pi, i\lambda)}{n_\alpha(\pi, i\lambda)} \right| d\lambda \leq C \log(T + \nu(\pi_1 \times \tilde{\pi}_2)), \tag{52}$$

where $\nu(\pi_1 \times \tilde{\pi}_2) = N(\pi_1 \times \tilde{\pi}_2)(2 + c(\pi_1 \times \tilde{\pi}_2)$, $N(\pi_1 \times \tilde{\pi}_2)$ is the conductor occurring in the functional equation (9) and $c(\pi_1 \times \tilde{\pi}_2)$ is the analytic conductor defined in [Mu5, (4.21)]. For the proof of (52) see [Mu5, Proposition 5.1]. In the case of $\mathrm{SL}(2, \mathbb{R})$ we have the pointwise estimate (40). If we integrate it, we get the analogue of (52) which would suffice to derive the Weyl law for the principal congruence subgroups of $\mathrm{SL}(2, \mathbb{Z})$.

Finally we have to deal with normalized intertwining operators

$$R_{Q|P}(\pi, s) = \otimes_v R_{Q|P}(\pi_v, s).$$

Since the open compact subgroup K_{fin} of $\mathbf{G}(\mathbb{A}_{\mathrm{fin}})$ is fixed, there are only finitely many places v for which we have to consider $R_{Q|P}(\pi_v, s)$. The main ingredient for the estimation of the logarithmic derivative of $R_{Q|P}(\pi_v, s)$, which is uniform in π_v, is a weak version of the Ramanujan conjecture (see [MS, Proposition 0.2]).

Combining these estimations, it follows that for every proper Levi subgroup $\mathbf{M}$ of $\mathbf{G}$ we have

$$J_{\mathrm{spec},M}(\phi_t^1) = O(t^{-(d-1)/2}) \tag{53}$$

as $t \to +0$. This proves (49).

The next problem is to estimate the remainder term in the Weyl law. For $\mathbf{G} = \mathrm{SL}(n)$ this problem has been studied by E. Lapid and the author in [LM]. Actually, we consider not only the cuspidal spectrum of the Laplacian, but also the cuspidal spectrum of the whole algebra of $\mathrm{SL}(n, \mathbb{R})$-invariant differential operators $\mathcal{D}(\widetilde{X})$ on $\widetilde{X} = \mathrm{SL}(n, \mathbb{R}) / \mathrm{SO}(n)$.

As $\mathcal{D}(\widetilde{X})$ preserves the space of cusp forms, we can proceed as in the compact case and decompose $L^2_{\mathrm{cus}}(\Gamma \backslash \widetilde{X})$ into joint eigenspaces of $\mathcal{D}(\widetilde{X})$. Recall that by (26) the characters of $\mathcal{D}(\widetilde{X})$ are parametrized by $\mathfrak{a}^*_{\mathbb{C}}/W$. Given $\lambda \in \mathfrak{a}^*_{\mathbb{C}}/W$, denote by χ_λ the corresponding character of $\mathcal{D}(\widetilde{X})$ and let

$$\mathcal{E}_{\mathrm{cus}}(\lambda) = \left\{ \varphi \in L^2_{\mathrm{cus}}(\Gamma \backslash \widetilde{X}) : D\varphi = \chi_\lambda(D)\varphi \right\}$$

be the associated joint eigenspace. Each eigenspace is finite-dimensional. Let $m(\lambda) = \dim \mathcal{E}_{\mathrm{cus}}(\lambda)$. Define the cuspidal spectrum $\Lambda_{\mathrm{cus}}(\Gamma)$ to be

$$\Lambda_{\mathrm{cus}}(\Gamma) = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}}/W : m(\lambda) > 0\}.$$

Then as in (27) we have an orthogonal direct sum decomposition

$$L^2_{\mathrm{cus}}(\Gamma\backslash\widetilde{X}) = \bigoplus_{\lambda\in\Lambda_{\mathrm{cus}}(\Gamma)} \mathcal{E}_{\mathrm{cus}}(\lambda).$$

Let $\beta(\lambda)$ be the Plancherel measure on $i\mathfrak{a}^*$. Then in [LM] we established the following extension of main results of [DKV] to congruence quotients of $\widetilde{X} = \mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$.

Theorem 3.5. *Let $d = \dim\widetilde{X}$. Let $\Omega \subset \mathfrak{a}^*$ be a bounded domain with piecewise smooth boundary. Then for $N \geq 3$ we have*

$$\sum_{\substack{\lambda\in\Lambda_{\mathrm{cus}}(\Gamma(N))\\ \lambda\in it\Omega}} m(\lambda) = \frac{\mathrm{vol}(\Gamma(N)\backslash\widetilde{X})}{|W|}\int_{it\Omega}\beta(\lambda)\,d\lambda + O\left(t^{d-1}(\log t)^{\max(n,3)}\right), \tag{54}$$

as $t\to\infty$, and

$$\sum_{\substack{\lambda\in\Lambda_{\mathrm{cus}}(\Gamma(N))\\ \lambda\in B_t(0)\backslash i\mathfrak{a}^*}} m(\lambda) = O\left(t^{d-2}\right), \quad t\to\infty. \tag{55}$$

If we apply (54) and (55) to the unit ball in $\mathfrak{a}^*$, we get the following corollary.

Corollary 3.6. *Let $\widetilde{X} = \mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$ and $d = \dim\widetilde{X}$. Let $\Gamma(N)$ be the principal congruence subgroup of $\mathrm{SL}(n,\mathbb{Z})$ of level N. Then for $N\geq 3$ we have*

$$N^{\mathrm{cus}}_{\Gamma(N)}(\lambda) = \frac{\mathrm{vol}(\Gamma(N)\backslash\widetilde{X})}{(4\pi)^{d/2}\mathbf{\Gamma}\left(\frac{d}{2}+1\right)}\lambda^{d/2} + O\left(\lambda^{(d-1)/2}(\log\lambda)^{\max(n,3)}\right), \quad \lambda\to\infty.$$

The condition $N \geq 3$ in Theorem 3.5 is imposed for technical reasons. It guarantees that the principal congruence subgroup $\Gamma(N)$ is neat in the sense of Borel, and in particular, has no torsion. This simplifies the analysis by eliminating the contributions of the non-unipotent conjugacy classes in the trace formula. In fact, in the recent paper [MT], Matz and Templier have eliminated the assumption $N\geq 3$ at the expense of the remainder term which is only $O(t^{d-1/2})$ (see [MT, (1.1)]). Moreover, [MT, Remark 1.9] contains a discussion of a possible improvement of the estimation of the remainder term.

Note that $\Lambda_{\mathrm{cus}}(\Gamma(N))\cap i\mathfrak{a}^*$ is the cuspidal tempered spherical spectrum. The Ramanujan conjecture [Sa3] for $\mathrm{GL}(n)$ at the Archimedean place states that

$$\Lambda_{\mathrm{cus}}(\Gamma(N))\subset i\mathfrak{a}^*$$

so that (55) is empty, if the Ramanujan conjecture is true. However, the Ramanujan conjecture is far from being proved. Moreover, it is known to be false for other groups $\mathbf{G}$ and (55) is what one can expect in general.

The method to prove Theorem 3.5 is an extension of the method of [DKV]. The Selberg trace formula, which is one of the basic tools in [DKV], is replaced by the non-invariant Arthur trace formula. Again, one of the main issues in the proof is the estimation of the logarithmic derivatives of the intertwining operators occurring on the spectral side of the trace formula.

3.3 Upper and Lower Bounds

In some cases it suffices to have upper or lower bounds for the counting function. For example, Donnelly's result (44) implies that there exists a constant $C > 0$ such that

$$N_\Gamma^{\mathrm{cus}}(\lambda;\sigma) \le C(1+\lambda^{d/2}), \quad \lambda \ge 0. \tag{56}$$

For the full discrete spectrum we have the bound (45). However, the exponent is not the optimal one. For some applications it is necessary to have such a bound which is uniform in Γ. For the cuspidal spectrum this problem has been studied by Deitmar and Hoffmann [DH]. To state the result, we have to introduce some notations. Let $\Gamma_n(N)$ be the principal congruence subgroup of $\mathrm{GL}(n,\mathbb{Z})$ of level N. Let $\mathbf{G}$ be a connected reductive linear algebraic group over $\mathbb{Q}$. Let $\eta\colon \mathbf{G} \to \mathrm{GL}(n)$ be a faithful $\mathbb{Q}$-rational representation. A family $\mathcal{T}$ of subgroups of $\mathbf{G}(\mathbb{Q})$ is called a *family of bounded depth* in $\mathbf{G}(\mathbb{Q})$ if there exists $D \in \mathbb{N}$ which satisfies the following property: For every $\Gamma \in \mathcal{T}$ there exists $N \in \mathbb{N}$ such that $\Gamma_n(N) \cap \eta(\mathbf{G}(\mathbb{Q}))$ is a subgroup of $\eta(\Gamma)$ of index at most D. Then the result of Deitmar and Hoffmann [DH, Corollary 18] is the following theorem.

Theorem 3.7. *Let $\Gamma_0 \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. Let $\mathcal{T}$ be a family subgroups of Γ_0 which is of bounded depth in $\mathbf{G}(\mathbb{Q})$. There exists $C > 0$ such that for all $\Gamma \in \mathcal{T}$ and all $\lambda \ge 0$ we have*

$$N_\Gamma^{\mathrm{cus}}(\lambda;\sigma) \le C[\Gamma_0\colon\Gamma](1+\lambda)^{d/2}. \tag{57}$$

Conjecture 2. *The estimation (57) holds for $N_\Gamma^{\mathrm{disc}}(\lambda;\sigma)$.*

Given the description of the residual spectrum for $\mathrm{GL}(n)$ by [MW], it seems possible to establish this conjecture for $\mathrm{GL}(n)$.

As for lower bounds there is the weak Weyl law established in [LM]. For $\sigma \in \widehat{K}$ let

$$c_\sigma(\Gamma) = \frac{\dim(\sigma)\,\mathrm{vol}(\Gamma\backslash\widetilde{X})}{(4\pi)^{d/2}\mathbf{\Gamma}(d/2+1)}$$

be the constant in Weyl's law, where $d = \dim(\widetilde{X})$. Let $\mathbf{G}$ be a semisimple algebraic group defined over $\mathbb{Q}$ and let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be a congruence subgroup defined by an open compact subgroup $K_{\text{fin}} = \prod_p K_p$ of $\mathbf{G}(\mathbb{A}_{\text{fin}})$. Let S be a finite set of primes. We will say that Γ is deep enough with respect to S, if for every prime $p \in S$, K_p is a subgroup of some minimal parahoric subgroup of $\mathbf{G}(\mathbb{Q}_p)$. Then the main result of [LM] is the following theorem.

Theorem 3.8. *Let* $\mathbf{G}$ *be an almost simple connected and simply connected semisimple algebraic group defined over* $\mathbb{Q}$ *such that* $\mathbf{G}(\mathbb{R})$ *is noncompact. Let S be a finite set of primes containing at least two primes. Then for every congruence subgroup* $\Gamma \subset \mathbf{G}(\mathbb{Q})$ *there exists a nonnegative constant* $c_S(\Gamma) \leq 1$ *such that for every* $\sigma \in \widehat{K}$ *with* $\sigma|_{Z_\Gamma} = \mathrm{Id}$ *we have*

$$c_\sigma(\Gamma) c_S(\Gamma) \leq \liminf_{\lambda \to \infty} \frac{N_\Gamma^{\text{cus}}(\lambda, \sigma)}{\lambda^{d/2}}.$$

Moreover $c_S(\Gamma) > 0$ *if* Γ *is deep enough with respect to S.*

3.4 Self-Dual Automorphic Representations

So far, we considered only the family of all cusp forms of $\mathrm{GL}(n, \mathbb{A})$. A nontrivial subfamily is formed by the family of self-dual automorphic representations. They arise as functorial lifts of automorphic representations of classical groups. Functoriality from quasisplit classical groups to general linear groups has been established by Cogdell et al. [CKP] for generic automorphic representations and then by Arthur [Ar8] for all representations. In his thesis, V. Kala has studied the counting function of self-dual cuspidal automorphic representations of $\mathrm{GL}(n, \mathbb{A})$. For $N \in \mathbb{N}$ with prime decomposition $N = \prod_p p^{r(p)}$ let

$$K_p(N) := \left\{ k \in \mathrm{GL}(n, \mathbb{Z}_p) \colon k \equiv 1 \mod p^{r(p)} \mathbb{Z}_p \right\}$$

Let $K(N)$ be the principal congruence subgroup defined by

$$K(N) := O(n) \times \prod_p K_p(N).$$

Let

$$N_{\text{sd}}^{K(N)}(\lambda) := \sum_{\substack{\lambda(\Pi) \leq \lambda \\ \Pi \cong \widetilde{\Pi}}} \dim \Pi^{K(N)},$$

where the sum ranges over all self-dual cuspidal automorphic representations Π of $\mathrm{GL}(n, \mathbb{A})$ with Casimir eigenvalues $\leq \lambda$. Then the main result of [Ka] is the following theorem.

Theorem 3.9. *Let $n = 2m + \varepsilon$ with $\varepsilon = 0, 1$. Put $d = m^2 + m$. For all $N \in \mathbb{N}$ there exist constants $C_1, C_2 > 0$ such that for $\lambda \gg 0$ one has*

$$C_1 \lambda^{d/2} \le N_{\mathrm{sd}}^{K(N)}(\lambda) \le C_2 \lambda^{d/2}.$$

By Corollary 3.6, the counting function of all cuspidal representations, counted similarly, is asymptotic to $C\lambda^{d/2}$, where $d = (n^2 + n - 2)/2$. Hence for $n > 2$, the density of self-dual cusp forms is zero.

The main idea of the proof of Theorem 3.9 is to consider the descent π of each self-dual cuspidal automorphic representation Π of $\mathrm{GL}(n, \mathbb{A})$ to one of the quasisplit classical groups $\mathbf{G}(\mathbb{A})$ and to use results towards the Weyl law on $\mathbf{G}(\mathbb{A})$. The number $d = m^2 + m$ is related to the dimension of the corresponding symmetric space $\mathbf{G}(\mathbb{R})/K_\infty$ (see [Ka, p. 17]). The key problem of the proof is to relate the Casimir eigenvalue and the existence of $K(N)$-fixed vectors for Π and π.

In a special case Kala's method leads to an exact asymptotic formula. Let $n = 2m$ and $d = m^2 + m$. Let $K = O(n) \times \prod_p K_p$ with $K_p = \mathrm{GL}(n, \mathbb{Z}_p)$. Then there exists $C > 0$ such that

$$N_{\mathrm{sd}}^{K}(\lambda) = C\lambda^{d/2} + o(\lambda^{d/2}) \tag{58}$$

(see [Ka, Corollary 6.2.2]). One may conjecture that this is true in general.

3.5 Weyl's Law for Hecke Operators

An important extension of the Weyl law is the study of the asymptotic distribution of infinitesimal characters of cuspidal automorphic representations weighted by the eigenvalues of Hecke operators acting on cusp forms of $\mathrm{GL}(n)$. For details we refer to the recent papers by Matz [Ma1], Matz and Templier [MT] and the survey article of Matz in these proceedings.

4 The Limit Multiplicity Problem

The limit multiplicity problem is another basic problem which is concerned with the asymptotic behavior of automorphic spectra.

In this section we summarize some of the known results about the limit multiplicity problem. Let G be a semisimple Lie group, $\Gamma \subset G$ a lattice in G, and μ_Γ the measure (3) on $\Pi(G)$. To begin with we recall some facts concerning the Plancherel measure μ_{pl} on $\Pi(G)$. First of all, the support of μ_{pl} is the tempered dual $\Pi(G)_{\mathrm{temp}}$, consisting of the equivalence classes of the irreducible unitary tempered representations. Up to a closed subset of Plancherel measure zero, the topological space $\Pi(G)_{\text{temp}}$ is homeomorphic to a countable union of Euclidean spaces of bounded dimensions. Under this homeomorphism the Plancherel density is given by a continuous function. We call the relatively quasi-compact subsets of $\Pi(G)$ *bounded*. We note that $\mu_\Gamma(A) < \infty$ for bounded sets $A \subset \Pi(G)$ under the reduction-theoretic assumptions on (G, Γ) mentioned above (see [BG]). A bounded subset A of $\Pi(G)_{\text{temp}}$ is called a Jordan measurable subset, if $\mu_{\text{pl}}(\partial A) = 0$, where $\partial A = \bar{A} - \text{int}(A)$ is the boundary of A in $\Pi(G)_{\text{temp}}$. Furthermore, a Riemann integrable function on $\Pi(G)_{\text{temp}}$ is a bounded, compactly supported function which is continuous almost everywhere with respect to the Plancherel measure.

Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of Borel measures on $\Pi(G)$. We say that the sequence $(\mu_n)_{n\in\mathbb{N}}$ has the *limit multiplicity property* (property (LM)), if the following two conditions are satisfied.

(1) For every Jordan measurable set $A \subset \Pi(G)_{\text{temp}}$ we have

$$\mu_n(A) \to \mu_{\text{pl}}(A), \quad \text{as } n \to \infty.$$

(2) For every bounded subset $A \subset \Pi(G) \setminus \Pi(G)_{\text{temp}}$ we have

$$\mu_n(A) \to 0, \quad \text{as } n \to \infty.$$

We note that condition (1) can be restated as

(1a) For every Riemann integrable function f on $\Pi(G)_{\text{temp}}$ one has

$$\lim_{n\to\infty} \mu_n(f) = \mu_{\text{pl}}(f).$$

Now let $(\Gamma_n)_{n\in\mathbb{N}}$ be a sequence of lattices in G. The sequence $(\Gamma_n)_{n\in\mathbb{N}}$ is said to have the limit multiplicity property (LM), if the sequence of measures $(\mu_{\Gamma_n})_{n\in\mathbb{N}}$ has property (LM).

The limit multiplicity problem can be formulated as follows: under which conditions does the sequence of measures μ_{Γ_n} satisfy property (LM)?

The limit multiplicity problem has been studied to a great extent in the case of uniform lattices. In this case, R_Γ decomposes discretely. It started with the work of DeGeorge and Wallach [DW1, DW2], who considered towers of normal subgroups, i.e., descending sequences of normal subgroups of finite index of a given uniform lattice with trivial intersection. For such sequences they dealt with the case of discrete series representations and the tempered spectrum, if the split rank of G is 1. Subsequently, Delorme [De] solved the limit multiplicity problem affirmatively for normal towers of cocompact lattices. Recently, there has been great progress in proving limit multiplicity for much more general sequences of uniform lattices by Abert et al. [AB1, AB2]. In particular, families of non-commensurable lattices were considered for the first time. The basic idea is the notion of Benjamini–Schramm convergence (BS-convergence), which originally was introduced for sequences of finite graphs of bounded degree and has been adopted by Abert et al. to sequences of Riemannian manifolds. For a Riemannian manifold M and $R > 0$ let

$$M_{<R} = \{x \in M\colon \mathrm{injrad}_M(x) < R\}.$$

Let (Γ_n) be a sequence of lattices in G. Then the orbifolds $M_n = \Gamma_n \backslash X$ are said to *BS-converge* to X, if for every $R > 0$ one has

$$\lim_{n\to+\infty} \frac{\mathrm{vol}((M_n)_{<R})}{\mathrm{vol}(M_n)} = 0. \tag{59}$$

To find examples of sequences (Γ_n) which satisfy this condition, consider a cocompact arithmetic lattice $\Gamma_0 \subset G$. By [AB1, Theorem 5.2] there exist constants $c, \mu > 0$ such that for any congruence subgroup $\Gamma \subset \Gamma_0$ and any $R > 1$ one has

$$\mathrm{vol}((\Gamma\backslash X)_{<R}) \le e^{cR}\,\mathrm{vol}(\Gamma\backslash X)^{1-\mu}. \tag{60}$$

Thus any sequence (Γ_n) of congruences subgroups of Γ_0 such that $\mathrm{vol}(\Gamma_n\backslash G) \to \infty$ as $n \to \infty$ satisfies (59).

A family of lattices in G is called to be uniformly discrete, if there exists a neighborhood of the identity in G that intersects trivially all of their conjugates. For torsion-free lattices Γ_n this is equivalent to the condition that there is a uniform lower bound of the injectivity radii of the manifolds $\Gamma_n\backslash X$. In particular, any family of normal subgroups (Γ_n) of a fixed uniform lattice Γ is uniformly discrete. Now the following theorem is one of the main results of [AB1, Theorem 1.2].

Theorem 4.1 ([AB1]). *Let (Γ_n) be a uniformly discrete sequence of lattices in G such that the orbifolds $\Gamma_n\backslash X$ BS-converge to X. Then the sequence (Γ_n) has the (LM) property.*

It follows from the discussions above that any sequence of congruence subgroups (Γ_n) of a given cocompact arithmetic lattice Γ_0 of G satisfies the assumptions of the theorem.

A special case of the limit multiplicity property is the case of a singleton $A = \{\pi\}$. Let $\Pi(G)_d \subset \Pi(G)$ be the discrete series and $d(\pi)$ the formal degree of $\pi \in \Pi(G)_d$. If (Γ_n) is a sequence of lattices in G which satisfies the (LM) property, then it follows that

$$\lim_{n\to\infty} \frac{m_{\Gamma_n}(\pi)}{\mathrm{vol}(\Gamma_n\backslash G)} = \begin{cases} d(\pi), & \pi \in \Pi(G)_d,\\ 0, & \text{else.}\end{cases} \tag{61}$$

It was first proved by DeGeorge and Wallach [DW1] that (61) holds for any tower of normal subgroups of a given uniform lattice of G.

An important problem is to extend these results to the non-cocompact case. Then the spectrum contains a continuous part and much less is known. The limit multiplicity problem has been solved for normal towers of arithmetic lattices and discrete series L-packets of representations (with regular parameters) by Rohlfs and Speh [RoS]. Then Savin [Sav] solved the limit multiplicity problem for the discrete series and normal towers of congruence subgroups.

In [FLM2] we dealt with the general case. Let F be a number field and denote by $\mathcal{O}_F$ its ring of integers. For the non-compact lattice $\mathrm{SL}(n, \mathcal{O}_F) \subset \mathrm{SL}(n, F \otimes \mathbb{R})$ we have the following result.

Theorem 4.2. *Let F be a number field. Then the collection of principal congruence subgroups* (Γ_N) *of* $\mathrm{SL}(n, \mathcal{O}_F)$ *has the limit multiplicity property.*

In [FL2], T. Finis and E. Lapid extended this result to the collection of all congruence subgroups of $\mathrm{SL}(n, \mathcal{O}_F)$, not containing non-trivial central elements. In [FLM2], we also discussed the case of a general reductive group.

4.1 The Density Principle and the Trace Formula

A standard approach to the limit multiplicity problem is to use integration against test functions on G and the trace formula. Let K be a maximal compact subgroup of G. Denote by $C^\infty_{c,\mathrm{fin}}(G)$ the space of smooth, compactly supported bi-K-finite functions on G. Given $f \in C^\infty_{c,\mathrm{fin}}(G)$, define $\hat{f}(\pi)$ for $\pi \in \Pi(G)$ by $\hat{f}(\pi) := \operatorname{tr} \pi(f)$. The function $\pi \in \Pi(G) \mapsto \hat{f}(\pi)$ on $\Pi(G)$ is the "Fourier transform" of f. Let μ be a Borel measure on $\Pi(G)$. Then $\mu(\hat{f})$ is defined (of course, it might be divergent). In particular, we have the two Borel measures μ_{pl} and μ_Γ defined on $\Pi(G)$. For these measures we have $\mu_{pl}(\hat{f}) = f(1)$ and

$$\mu_\Gamma(\hat{f}) = \frac{1}{\mathrm{vol}(\Gamma\backslash G)} \operatorname{tr} R_{\Gamma,\mathrm{disc}}(f). \tag{62}$$

By [Mu2], $R_{\Gamma,\mathrm{disc}}(f)$ is a trace class operator. Thus the right-hand side is well defined. Furthermore, by the Plancherel theorem we have $\mu_{\mathrm{pl}}(\hat{f}) = f(1)$. The density principle of Sauvageot [Sau], which is a refinement of the work of Delorme, can be stated as follows.

Theorem 4.3. *Let* $(\mu_n)_{n\in\mathbb{N}}$ *be a sequence of Borel measures on* $\Pi(G)$ *and assume that for all* $f \in C^\infty_{c,\mathrm{fin}}(G)$ *we have*

$$\mu_n(\hat{f}) \to \mu_{pl}(\hat{f}) = f(1), \quad \text{as } n \to \infty. \tag{63}$$

Then $(\mu_n)_{n\in\mathbb{N}}$ *satisfies (LM).*

Now let $(\Gamma_n)_{n\in\mathbb{N}}$ be a sequence of lattices in G. Then by Theorem 4.3 it follows that $(\Gamma_n)_{n\in\mathbb{N}}$ satisfies (LM), if

$$\mu_{\Gamma_n}(\hat{f}) \to f(1), \quad n \to \infty, \tag{64}$$

for all $f \in C^\infty_{c,\mathrm{fin}}(G)$. A standard approach to verify (64) is to use the trace formula. In the case of co-compact lattices this is rather simple. Let Γ be a cocompact lattice in G. In this the Selberg trace formula we obtain

$$\mathrm{vol}(\Gamma\backslash G)\mu_\Gamma(\hat{f}) = \mathrm{tr}\, R_\Gamma(f) = \sum_{\{\gamma\}\in C(\Gamma)} \mathrm{vol}(\Gamma_\gamma\backslash G_\gamma) \int_{G_\gamma\backslash G} f(x^{-1}\gamma x)\, dx,$$

where $C(\Gamma)$ denotes the Γ-conjugacy classes of Γ, and G_γ (resp. Γ_γ) denotes the centralizer of γ in G (resp. Γ). Let $\Gamma_1 \subset \Gamma$ be a finite index subgroup. For $\gamma \in \Gamma$ let

$$c_{\Gamma_1}(\gamma) = |\{\delta \in \Gamma_1\backslash\Gamma \colon \delta\gamma\delta^{-1} \in \Gamma_1\}|. \tag{65}$$

In [Co], Corwin shows that the elements on the right-hand side of the trace formula for Γ_1 can be grouped together in a way to give

$$\mu_{\Gamma_1}(\hat{f}) = \frac{1}{\mathrm{vol}(\Gamma\backslash G)} \sum_{\{\gamma\}\in C(\Gamma)} \mathrm{vol}(\Gamma_\gamma\backslash G_\gamma) \frac{c_{\Gamma_1}(\gamma)}{[\Gamma\colon\Gamma_1]} \int_{G_\gamma\backslash G} f(x^{-1}\gamma x)\, dx. \tag{66}$$

For a central element γ we obviously have $c_{\Gamma_1}(\gamma) = [\Gamma\colon\Gamma_1]$. Assume that the center of Γ is trivial. Let $(\Gamma_n)_{n\in\mathbb{N}}$ be a sequence of finite index subgroups of Γ. Then we have

$$\mu_{\Gamma_n}(\hat{f}) = f(1) + \frac{1}{\mathrm{vol}(\Gamma\backslash G)} \sum_{\{\gamma\}\in C(\Gamma)\backslash\{1\}} \mathrm{vol}(\Gamma_\gamma\backslash G_\gamma) \frac{c_{\Gamma_n}(\gamma)}{[\Gamma\colon\Gamma_n]} \int_{G_\gamma\backslash G} f(x^{-1}\gamma x)\, dx. \tag{67}$$

By dominated convergence, it follows that in order to establish (63) for the sequence $(\Gamma_n)_{n\in\mathbb{N}}$, it suffices to show that for every $\gamma \in \Gamma$, $\gamma \neq 1$, we have

$$\frac{c_{\Gamma_n}(\gamma)}{[\Gamma\colon\Gamma_n]} \to 0, \quad \text{as } n \to \infty. \tag{68}$$

Now note that if Γ_1 is a normal subgroup of Γ, then $c_{\Gamma_1}(\gamma)/[\Gamma\colon\Gamma_1]$ is the characteristic function of Γ_1. Thus for normal towers of finite index subgroups of Γ the condition (68) holds trivially. This implies Delorme's result.

If Γ is not co-compact, the Selberg trace formula is only available in the rank one case. We have to switch to the adelic framework so that we can use the Arthur trace formula.

Thus let now $\mathbf{G}$ be an arbitrary reductive group defined over $\mathbb{Q}$. Let $\mathbb{A} = \mathbb{R}\times\mathbb{A}_{\mathrm{fin}}$ be the locally compact adele ring of $\mathbb{Q}$. For every place v of $\mathbb{Q}$ (i.e., $v = \infty$ or $v = p$ a prime) let $|\cdot|_v$ be the normalized absolute value of $\mathbb{Q}$. As usual, $\mathbf{G}(\mathbb{R})^1$ denotes the intersection of the kernels of the homomorphisms $|\chi|\colon \mathbf{G}(\mathbb{R}) \to \mathbb{R}^+$, where χ runs over the $\mathbb{Q}$-rational characters of $\mathbf{G}$. Similarly we define the normal subgroup $\mathbf{G}(\mathbb{A})^1$ of $\mathbf{G}(\mathbb{A})$. Every $\pi \in \Pi(\mathbf{G}(\mathbb{A})^1)$ can be written as $\pi = \pi_\infty \otimes \pi_{\mathrm{fin}}$, where $\pi_\infty \in \Pi(\mathbf{G}(\mathbb{R})^1)$ and $\pi_{\mathrm{fin}} \in \Pi(\mathbf{G}(\mathbb{A}_{\mathrm{fin}}))$. Fix a Haar measure on $\mathbf{G}(\mathbb{A})$. For any open compact subgroup K_f of $\mathbf{G}(\mathbb{A}_{\text{fin}})$, let $\mu_K = \mu_K^G$ be the measure on $\Pi(\mathbf{G}(\mathbb{R})^1)$ defined by

$$\begin{aligned}\mu_K &= \frac{1}{\operatorname{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1/K)} \sum_{\pi\in\Pi(G(\mathbb{R})^1)} \operatorname{Hom}_{\mathbf{G}(\mathbb{R})^1}(\pi, L^2(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1/K))\delta_\pi \\ &= \frac{\operatorname{vol}(K)}{\operatorname{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1)} \sum_{\pi\in\Pi(\mathbf{G}(\mathbb{A})^1)} \dim \operatorname{Hom}_{\mathbf{G}(\mathbb{A})^1}(\pi, L^2(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1)) \dim(\pi_{\text{fin}})^K \delta_{\pi_\infty}.\end{aligned} \tag{69}$$

We say that a sequence $(K_n)_{n\in\mathbb{N}}$ of open compact subgroups of $\mathbf{G}(\mathbb{A}_{\text{fin}})$ has the limit multiplicity property, if $\mu_{K_n} \to \mu_{\text{pl}}$, $n \to \infty$, in the sense that

(1) For every Jordan measurable subset $A \subset \Pi(\mathbf{G}(\mathbb{R})^1))_{\text{temp}}$ we have $\mu_{K_n}(A) \to \mu_{\text{pl}}(A)$ as $n \to \infty$, and
(2) For every bounded subset $A \subset \Pi(\mathbf{G}(\mathbb{R})^1) \setminus \Pi(\mathbf{G}(\mathbb{R})^1))_{\text{temp}}$, we have $\mu_{K_n}(A) \to 0$ as $n \to \infty$.

Again we can rephrase the first condition by saying that for any Riemann integrable function f on $\Pi(\mathbf{G}(\mathbb{R})^1)_{\text{temp}}$ we have

$$\mu_{K_n}(f) \to \mu_{\text{pl}}(f), \quad \text{as} \quad n \to \infty. \tag{70}$$

Note that when $\mathbf{G}$ satisfies the strong approximation property (which is the case if $\mathbf{G}$ is semisimple, simply connected, and without any $\mathbb{Q}$-simple factor $\mathbf{H}$ for which $\mathbf{H}(\mathbb{R})$ is compact) and K is an open compact subgroup of $\mathbf{G}(\mathbb{A}_{\text{fin}})$, then we have

$$\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})/K \cong \Gamma_K\backslash\mathbf{G}(\mathbb{R}),$$

where $\Gamma_K = \mathbf{G}(\mathbb{Q}) \cap K$ is a lattice in the connected semisimple Lie group $\mathbf{G}(\mathbb{R})$.

Now for $f \in C_{c,\text{fin}}^\infty(\mathbf{G}(\mathbb{R})^1)$ we have

$$\mu_K(\hat{f}) = \frac{1}{\operatorname{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(A)^1)} \operatorname{tr} R_{\text{disc}}(f \otimes \mathbf{1}_K) \tag{71}$$

and

$$\mu_{\text{pl}}(\hat{f}) = f(1). \tag{72}$$

Sauvageot's density principle [Sau] can now be reformulated as follows.

Theorem 4.4. *Let $(K_n)_{n\in\mathbb{N}}$ be a sequence of open compact subgroups of $\mathbf{G}(\mathbb{A}_{\text{fin}})$. Suppose that for every $f \in C_{c,\text{fin}}^\infty(\mathbf{G}(\mathbb{R})^1)$ we have*

$$\mu_{K_n}(\hat{f}) \to f(1), \quad n \to \infty. \tag{73}$$

Then $(K_n)_{n\in\mathbb{N}}$ has the limit multiplicity property.

To try to verify (73), it is natural to use Arthur's (non-invariant) trace formula, which is an equality

$$J_{\mathrm{spec}}(h) = J_{\mathrm{geo}}(h), \quad h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1),$$

of two distribution on $\mathbf{G}(\mathbb{A})^1$ [Ar1, Ar2, Ar3]. The distribution J_{spec} is expressed in terms of spectral data and J_{geo} in terms of geometric data. The main terms on the geometric side are the elliptic orbital integrals. In particular, the contribution $\mathrm{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1)h(1)$ of the identity element occurs on the geometric side. The main term on the spectral side is $\mathrm{tr}\, R_{\mathrm{disc}}(h)$. By (71) it follows that (73) can be broken down into the following two statements. For every $f \in C^\infty_{c,\mathrm{fin}}(G(\mathbb{R})^1)$ we have

$$J_{\mathrm{spec}}(f \otimes \mathbf{1}_{K_n}) - \mathrm{tr}\, R_{\mathrm{disc}}(f \otimes \mathbf{1}_{K_n}) \to 0, \quad n \to \infty, \tag{74}$$

and

$$J_{\mathrm{geo}}(f \otimes \mathbf{1}_{K_n}) \to \mathrm{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1)f(1), \quad n \to \infty. \tag{75}$$

We call (74) the spectral—and (75) the geometric limit property.

4.2 Bounds on Co-rank One Intertwining Operators

In this section we formulate two conditions on the behavior of the intertwining operators $M_{Q|P}$ which imply the spectral limit property for a given $\mathbf{G}$. They also imply Weyl's law for the group $\mathbf{G}$. We call these properties (TWN) (tempered winding number) and (BD) (bounded degree). The first property is global and second local. The first property is connected with analytic problems in the theory of automorphic L-functions.

We will use the notation $A \ll B$ to mean that there exists a constant c (independent of the parameters under consideration) such that $A \le cB$. If c depends on some parameters (say F) and not on others, then we will write $A \ll_F B$.

Fix a faithful $\mathbb{Q}$-rational representation $\rho : \mathbf{G} \to \mathrm{GL}(V)$ and a $\mathbb{Z}$-lattice Λ in the representation space V such that the stabilizer of $\hat{\Lambda} = \hat{\mathbb{Z}} \otimes \Lambda \subset \mathbb{A}_{\mathrm{fin}} \otimes V$ in $\mathbf{G}(\mathbb{A}_{\mathrm{fin}})$ is the group $\mathbf{K}_{\mathrm{fin}}$. (Since the maximal compact subgroups of $\mathrm{GL}(\mathbb{A}_{\mathrm{fin}} \otimes V)$ are precisely the stabilizers of lattices, it is easy to see that such a lattice exists.) For any $N \in \mathbb{N}$ let

$$\mathbf{K}(N) = \{g \in \mathbf{G}(\mathbb{A}_{\mathrm{fin}}) \,:\, \rho(g)v \equiv v \pmod{N\hat{\Lambda}}, \quad v \in \hat{\Lambda}\} \tag{76}$$

be the principal congruence subgroup of level N, an open normal subgroup of $\mathbf{K}_{\mathrm{fin}}$. The groups $\mathbf{K}(N)$ form a neighborhood basis of the identity element in $\mathbf{G}(\mathbb{A}_{\mathrm{fin}})$. For an open subgroup K of $\mathbf{K}_{\mathrm{fin}}$ let the *level* of K be the smallest integer N such that $\mathbf{K}(N) \subset K$. Analogously, define $\mathrm{level}(K_v)$ for open subgroups $K_v \subset \mathbf{K}_v$.

As in [Mu6], for any $\pi \in \Pi(\mathbf{M}(\mathbb{R}))$ we define $\Lambda_\pi = \sqrt{\lambda_\pi^2 + \lambda_\tau^2}$, where τ is a lowest $\mathbf{K}_\infty$-type of $\mathrm{Ind}_{\mathbf{P}(\mathbb{R})}^{\mathbf{G}(\mathbb{R})}(\pi)$ and λ_π and λ_τ are the Casimir eigenvalues of π and τ, respectively. Note that this is well defined, because λ_τ is independent of τ. Roughly speaking, Λ_π measures the size of π. For $\mathbf{M} \in \mathcal{L}$, $\alpha \in \Sigma_M$ and $\pi \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))$ let $n_\alpha(\pi, s)$ be the global normalizing factor defined by (8).

Definition 4.5. We say that the group $\mathbf{G}$ satisfies the property (TWN) (tempered winding number) if for any $\mathbf{M} \in \mathcal{L}$, $\mathbf{M} \neq \mathbf{G}$, and any finite subset $\mathcal{F} \subset \Pi(\mathbf{K}_{M,\infty})$ there exists an integer $k > 1$ such that for any $\alpha \in \Sigma_M$ and any $\epsilon > 0$ we have

$$\int_{\mathrm{i}\mathbb{R}} \left| \frac{n'_\alpha(\pi, s)}{n_\alpha(\pi, s)} \right| (1 + |s|)^{-k} \, ds \ll_{\mathcal{F},\epsilon} (1 + \Lambda_{\pi_\infty})^k \, \mathrm{level}(K_M)^\epsilon \tag{77}$$

for all open compact subgroups K_M of $\mathbf{K}_{M,\mathrm{fin}}$ and all $\pi = \pi_\infty \otimes \pi_{\mathrm{fin}} \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))$ such that π_∞ contains a $\mathbf{K}_{M,\infty}$-type in the set $\mathcal{F}$ and $\pi_{\mathrm{fin}}^{K_M} \neq 0$.

Since the normalizing factors $n_\alpha(\pi, s)$ arise from co-rank one situations, the property (TWN) is hereditary for Levi subgroups.

Remark 4.6. If we fix an open compact subgroup K_M, then the corresponding bound

$$\int_{\mathrm{i}\mathbb{R}} \left| \frac{n'_\alpha(\pi, s)}{n_\alpha(\pi, s)} \right| (1 + |s|)^{-k} \, ds \ll_{K_M} (1 + \Lambda_{\pi_\infty})^k$$

is the content of [Mu6, Theorem 5.3]. So, the point of (TWN) lies in the dependence of the bound on K_M.

Remark 4.7. In fact, we expect that

$$\int_T^{T+1} \left| \frac{n'_\alpha(\pi, \mathrm{i}t)}{n_\alpha(\pi, \mathrm{i}t)} \right| \, dt \ll 1 + \log(1 + T) + \log(1 + \Lambda_{\pi_\infty}) + \log \mathrm{level}(K_M) \tag{78}$$

for all $T \in \mathbb{R}$ and $\pi \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))^{K_M}$. This would give the following strengthening of (TWN):

$$\int_{\mathrm{i}\mathbb{R}} \left| \frac{n'_\alpha(\pi, s)}{n_\alpha(\pi, s)} \right| (1 + |s|)^{-2} \, ds \ll 1 + \log(1 + \Lambda_{\pi_\infty}) + \log \mathrm{level}(K_M)$$

for any $\pi \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))^{K_M}$.

Remark 4.8. If G' is simply connected, then by [Lub, Lemma 1.6] (cf. also [FLM2, Proposition 1]) we can replace $\mathrm{level}(K_M)$ by $\mathrm{vol}(K_M)^{-1}$ in the definition of (TWN) (as well as in (78)).

For $\mathrm{GL}(n)$ the normalizing factors are expressed in terms of Rankin-Selberg L functions (see (10)). The known properties of Rankin-Selberg L-functions lead to the estimation (52), which implies the desired estimation. By [FLM2, Lemma 5.4], the case of $\mathrm{SL}(n)$ can be reduced to $\mathrm{GL}(n)$. In this way we get (see [FLM2]).

Theorem 4.9. *The estimate* (78) *holds for* $\mathbf{G} = \mathrm{GL}(n)$ *or* $\mathrm{SL}(n)$ *with an implied constant depending only on n. In particular, the groups* $\mathrm{GL}(n)$ *and* $\mathrm{SL}(n)$ *satisfy the property (TWN).*

Remark 4.10. For general groups $\mathbf{G}$ the normalizing factors are given, at least up to local factors, by quotients of automorphic L-functions associated with the irreducible constituents of the adjoint action of the L-group ${}^L M$ of $\mathbf{M}$ on the unipotent radical of the corresponding parabolic subgroup of ${}^L G$ [La1]. To argue as above, we would need to know that these L-functions have finitely many poles and satisfy a functional equation with the associated conductor bounded by an arbitrary power of $\mathrm{level}(K_M)$ for automorphic representations $\pi \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))^{K_M}$. Unfortunately, finiteness of poles and the expected functional equation are not known in general. It is possible that for classical groups these properties are within reach.

Now we come to the second condition, which is a condition on the local intertwining operators. Recall that for a finite prime p, the matrix coefficients of the local normalized intertwining operators $R_{Q|P}(\pi_p, s)^{K_p}$ are rational functions of p^s. Moreover, their denominators can be controlled in terms of π_p, and the degrees of these denominators are bounded in terms of $\mathbf{G}$ only. For any Levi subgroup $\mathbf{M} \in \mathcal{L}$ let $\mathbf{G_M}$ be the closed subgroup of $\mathbf{G}$ generated by the unipotent radicals $\mathbf{U}_P$, where $\mathbf{P} \in \mathcal{P}(\mathbf{M})$. It is a connected semisimple normal subgroup of $\mathbf{G}$.

Definition 4.11. We say that $\mathbf{G}$ satisfies (BD) (bounded degree) if there exists a constant c (depending only on $\mathbf{G}$ and ρ), such that for any $\mathbf{M} \in \mathcal{L}$, $\mathbf{M} \neq \mathbf{G}$, and adjacent parabolic groups $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(\mathbf{M})$, any prime p, any open subgroup $K_p \subset \mathbf{K}_p$ and any smooth irreducible representation π_p of $\mathbf{M}(\mathbb{Q}_p)$, the degrees of the numerators of the linear operators $R_{Q|P}(\pi_p, s)^{K_p}$ are bounded by $c \log_p \mathrm{level}^{G_M}(K_P)$ if $\mathbf{K}_p$ is hyperspecial, and by $c(1 + \log_p \mathrm{level}^{\mathbf{G_M}}(K_p))$, otherwise.

Property (BD) has been studied in [FLM3]. By [FLM3, Theorem 1, Proposition 6] we have the following theorem.

Theorem 4.12. *The groups* $\mathrm{GL}(n)$ *and* $\mathrm{SL}(n)$ *satisfy (BD).*

The property (BD) has the following consequence.

Proposition 4.13. *Suppose that* $\mathbf{G}$ *satisfies (BD). Let* $\mathbf{M} \in \mathcal{L}$ *and let* $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(\mathbf{M})$ *be adjacent parabolic subgroups. Then for all* $\pi \in \Pi_{\mathrm{disc}}(\mathbf{M}(\mathbb{A}))$, *for all open subgroups* $K \subset \mathbf{K}_{\mathrm{fin}}$ *and all* $\tau \in \Pi(\mathbf{K}_\infty)$ *we have*

$$\int_{i\mathbb{R}} \left\| R_{Q|P}(\pi, s)^{-1} \frac{d}{ds} R_{Q|P}(\pi, s)\Big|_{I_P^G(\pi)^{\tau,K}} \right\| (1 + |s|^2)^{-1} \, ds$$
$$\ll 1 + \log(\|\tau\| + \mathrm{level}(K; \mathbf{G}_{\mathbf{M}}^{+})). \quad (79)$$

The proof of the proposition follows from a generalization of Bernstein's inequality [BE]. Suppose that $\mathbf{G}$ satisfies (TWN) and (BD). Combining (77) and (79) we get an appropriate estimate for the corresponding integral involving the logarithmic derivative of the intertwining operators.

4.3 Application to the Limit Multiplicity Problem

The limit multiplicity property is a consequence of properties (TWN) and (BD). The proof proceeds by induction over the Levi subgroups of $\mathbf{G}$. The property that is suitable for the induction procedure is not the spectral limit property, but a property that we call *polynomial boundedness* (PB). This is a weaker version of the statement of Conjecture 2.

We write $\mathcal{D}$ for the set of all conjugacy classes of pairs (M,δ) consisting of a Levi subgroup M of $\mathbf{G}(\mathbb{R})^1$ and a discrete series representation δ of M^1, where $M = A_M \times M^1$ and A_M is the largest central subgroup of M isomorphic to a power of $\mathbb{R}^{>0}$. For any $\underline{\delta} \in \mathcal{D}$ let $\Pi(\mathbf{G}(\mathbb{R})^1)_{\underline{\delta}}$ be the set of all irreducible unitary representations which arise by the Langlands quotient construction from the irreducible constituents of $I_M^L(\delta)$ for Levi subgroups $L \supset M$. Here, I_M^L denotes (unitary) induction from an arbitrary parabolic subgroup of L with Levi subgroup M to L.

Definition 4.14. Let $\mathfrak{M}$ be a set of Borel measures on $\Pi(\mathbf{G}(\mathbb{R})^1)$. We call $\mathfrak{M}$ *polynomially bounded* (PB), if for all $\underline{\delta} \in \mathcal{D}$ there exist $N_{\underline{\delta}} > 0$ such that

$$\mu\left(\{\pi \in \Pi(\mathbf{G}(\mathbb{R})^1)_{\underline{\delta}} \colon |\lambda_\pi| \le R\}\right) \ll_{\underline{\delta}} (1+R)^{N_{\underline{\delta}}}$$

for all $\mu \in \mathfrak{M}$ and $R > 0$.

Now consider the measures μ_K defined by (69). Let $\mathbf{M} \in \mathcal{L}$ and denote by $\mathbf{K}_M(N)$ the congruence subgroups of $\mathbf{M}(\mathbb{A}_{\text{fin}})$, defined by (76). Denote by $\mu^{\mathbf{M}}_{\mathbf{K}_M(N)}$ the measure defined by (71) with $\mathbf{M}$ in place of $\mathbf{G}$. Then the key result is the following lemma.

Lemma 4.15. *Suppose that* $\mathbf{G}$ *satisfies (TWN) and (BD). Then for each* $\mathbf{M} \in \mathcal{L}$*, the collection of measures* $\{\mu^{\mathbf{M}}_{\mathbf{K}_M(N)}\}$*,* $N \in \mathbb{N}$*, is polynomially bounded.*

This has the consequence that if $\mathbf{G}$ satisfies (TWN) and (BD), then for every $\mathbf{M} \neq \mathbf{G}$ and $f \in C^\infty_{c,\text{fin}}(\mathbf{G}(\mathbb{R})^1)$ we have

$$J_{\text{spec},M}(f \otimes \mathbf{1}_{\mathbf{K}(N)}) \to 0$$

as $N \to \infty$. Thus by Theorem 2.1 it follows that if $\mathbf{G}$ satisfies (TWN) and (BD), then for every $f \in C^\infty_{c,\text{fin}}(\mathbf{G}(\mathbb{R})^1)$ we have

$$J_{\text{spec}}(f \otimes \mathbf{1}_{\mathbf{K}(N)}) - \operatorname{tr} R_{\text{disc}}(f \otimes \mathbf{1}_{\mathbf{K}(N)}) \to 0$$

for $n \to \infty$. Thus the spectral limit property is satisfied in this case. By Theorems 4.9 and 4.12, the groups $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$ satisfy (TWN) and (BD) and therefore, the spectral limit property holds for $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$.

To deal with the geometric limit property we use the coarse geometric expansion

$$J^T(h) = \sum_{\mathfrak{o}\in\mathcal{O}} J^T_{\mathfrak{o}}(h), \quad h \in C^\infty_c(\mathbf{G}(\mathbb{A})^1), \tag{80}$$

(see (11) for the notation). Write $J_{\mathfrak{o}}(f) = J_{\mathfrak{o}}^{T_0}(f)$, which depends only on $\mathbf{M}_0$ and $\mathbf{K}$. Let J_{unip}^T be the contribution of the unipotent elements of $G(\mathbb{Q})$ to the trace formula (11), which is a polynomial in $T \in \mathfrak{a}_{M_0}$ of degree at most $d_0 = \dim \mathfrak{a}_{M_0}^G$ [Ar7]. It can be split into the contributions of the finitely many $G(\bar{\mathbb{Q}})$-conjugacy classes of unipotent elements of $\mathbf{G}(\mathbb{Q})$. It is well known [[Ar7, Corollary 4.4]] that the contribution of the unit element is simply the constant polynomial $\mathrm{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1)h(1)$. Write

$$J_{\mathrm{unip}-\{1\}}^T(h) = J_{\mathrm{unip}}^T(h) - \mathrm{vol}(\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})^1)h(1), \quad h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1).$$

Define the distributions J_{unip} and $J_{\mathrm{unip}-\{1\}}$ as $J_{\mathrm{unip}}^{T_0}$ and $J_{\mathrm{unip}-\{1\}}^{T_0}$, respectively. Since the groups $\mathbf{K}(N)$ form a neighborhood basis of the identity element in $\mathbf{G}(s\mathbb{A}_{\mathrm{fin}})$, it is easy to see that for a given $h \in C_c^\infty(\mathbf{G}(\mathbb{A})^1)$, for all but finitely many N one has

$$J(h \otimes \mathbf{1}_{\mathbf{K}(N)}) = J_{\mathrm{unip}}(h \otimes \mathbf{1}_{\mathbf{K}(N)}). \tag{81}$$

For any compact subset $\Omega \subset G(\mathbb{R})^1$ we write $C_\Omega^\infty(\mathbf{G}(\mathbb{R})^1)$ for the Fréchet space of all smooth functions on $G(\mathbb{R})^1$ supported in Ω equipped with the seminorms $\sup_{x\in\Omega}|(Xh)(x)|$, where X ranges over the left-invariant differential operators on $\mathbf{G}(\mathbb{R})$. The key result is the following proposition.

Proposition 4.16. *For any compact subset $\Omega \subset \mathbf{G}(\mathbb{R})^1$ there exists a seminorm $||\cdot||$ on $C_\Omega^\infty(\mathbf{G}(\mathbb{R})^1)$ such that*

$$|J_{\mathrm{unip}-\{1\}}(h \otimes \mathbf{1}_{\mathbf{K}(N)})| \le \frac{(1+\log(N))}{N}\|h\|$$

for all $h \in C_\Omega^\infty(\mathbf{G}(\mathbb{R})^1)$ and all $N \in \mathbb{N}$.

The proof of Proposition 4.16 consists of a slight extension of Arthur's arguments in [Ar7]. Combining (81) and Proposition 4.16 the geometric limit property follows. This completes the proof of Theorem 4.2 for $F = \mathbb{Q}$. The case of a general F is proved similarly. For details see [FLM2].

5 Analytic Torsion and Torsion in the Cohomology of Arithmetic Groups

The theorem of DeGeorge and Wallach on limit multiplicities for discrete series [DW1] implies the statement (4) on the approximation of L^2-Betti numbers by normalized Betti numbers of finite covers [AB2]. For towers of normal subgroups of finite index, Lück [Lu1] proved this in the more general context of finite CW complexes. This is part of his study of the approximation of L^2-invariants by their classical counterparts [Lu2]. A more sophisticated spectral invariant is the analytic torsion introduced by Ray and Singer [RS]. The study of the corresponding approximation problem has interesting applications to the torsion in the cohomology of arithmetic groups.

5.1 Analytic Torsion and L^2-Torsion

Let X be a compact Riemannian manifold of dimension n and let $\rho\colon \pi_1(X) \to \mathrm{GL}(V)$ a finite dimensional representation of its fundamental group. Let $E_\rho \to X$ be the flat vector bundle associated with ρ. Choose a Hermitian fiber metric in E_ρ. Let $\Delta_p(\rho)$ be the Laplace operator on E_ρ-valued p-forms with respect to the metrics on X and in E_ρ. It is an elliptic differential operator, which is formally self-adjoint and non-negative. Since X is compact, $\Delta_p(\rho)$ has a pure discrete spectrum consisting of sequence of eigenvalues $0 \le \lambda_0 \le \lambda_1 \le \cdots \to \infty$ of finite multiplicity. Let

$$\zeta_p(s;\rho) := \sum_{\lambda_j > 0} \lambda_j^{-s} \tag{82}$$

be the zeta function of $\Delta_p(\rho)$. The series converges absolutely and uniformly on compact subsets of the half-plane $\mathrm{Re}(s) > n/2$ and admits a meromorphic extension to $s \in \mathbb{C}$, which is holomorphic at $s = 0$ [Shu]. Then the Ray-Singer analytic torsion $T_X(\rho) \in \mathbb{R}^+$ is defined by

$$T_X(\rho) := \exp\left(\frac{1}{2} \sum_{p=1}^{n} (-1)^p p \frac{d}{ds} \zeta_p(s;\rho)\big|_{s=0} \right). \tag{83}$$

It depends on the metrics on X and E_ρ. However, if $\dim(X)$ is odd and ρ acyclic, which means that $H^*(X, E_\rho) = 0$, then $T_X(\rho)$ is independent of the metrics [Mu3]. The analytic torsion has a topological counterpart. This is the Reidemeister torsion $T_X^{\mathrm{top}}(\rho)$ (usually it is denoted by $\tau_X(\rho)$), which is defined in terms of a smooth triangulation of X [RS, Mu1]. It is known that for unimodular representations ρ (meaning that $|\det \rho(\gamma)| = 1$ for all $\gamma \in \pi_1(X)$) one has the equality

$$T_X(\rho) = T_X^{\mathrm{top}}(\rho) \tag{84}$$

[Ch, Mu1]. In the general case of a non-unimodular representation the equality does not hold, but the defect can be described [BZ].

Let $X_i \to X$, $i \in \mathbb{N}$, be sequence of finite coverings of X. Let $\mathrm{inj}(X_j)$ denote the injectivity radius of X_j and assume that $\mathrm{inj}(X_j) \to \infty$ as $j \to \infty$. Then the question is: Does

$$\frac{\log T_{X_j}(\rho)}{\mathrm{vol}(X_j)} \tag{85}$$

converge as $j \to \infty$ and if so, what is the limit? For a tower of normal coverings and the trivial representation ρ_0 a conjecture of Lück [Lu2, Conjecture 7.4] states that the sequence (85) converges and the limit is the L^2-torsion, first introduced by Lott [Lo] and Mathei [MV]. The L^2-torsion is defined as follows. Recall that the zeta function $\zeta_p(s)$ can be expressed in terms of the heat operator

$$\zeta_p(s) = \frac{1}{\Gamma(s)} \int_0^\infty (\mathrm{Tr}\left(e^{-t\Delta_p}\right) - b_p) t^{s-1}\, dt,$$

where b_p is the p-th Betti number and $\mathrm{Re}(s) > n/2$. Let $e^{-t\widetilde{\Delta}_p}$ be the heat operator of the Laplace operator $\widetilde{\Delta}_p$ on p-forms on the universal covering $\widetilde{X}$ of X. Let $\widetilde{K}_p(t,x,y)$ be the kernel of $e^{-t\widetilde{\Delta}_p}$. Note that $\widetilde{K}_p(t,x,y)$ is a homomorphism of $\Lambda^p T_y^*(X)$ to $\Lambda^p T_x^*(X)$. Let $F \subset \widetilde{X}$ be a fundamental domain for the action of $\Gamma := \pi_1(X)$ on $\widetilde{X}$. Then the Γ-trace of $e^{-t\widetilde{\Delta}_p(\rho)}$ is defined as

$$\mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}_p}\right) := \int_F \mathrm{tr}\, \widetilde{K}_p(t,x,x)\, dx. \tag{86}$$

The L^2-Betti number $b_p^{(2)}$ is defined as

$$b_p^{(2)} := \lim_{t\to\infty} \mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}_p}\right).$$

In order to be able to define the Mellin transform of the Γ-trace one needs to know the asymptotic behavior of $\mathrm{Tr}_\Gamma(e^{-t\widetilde{\Delta}_p})$ as $t \to 0$ and $t \to \infty$. Using a parametrix for the heat kernel which is pulled back from a parametrix on X, one can show that for $t \to 0$, $\mathrm{Tr}_\Gamma(e^{-t\widetilde{\Delta}_p})$ has an asymptotic expansion similar to the compact case [Lo]. For the large time behavior we need to introduce the Novikov-Shubin invariants

$$\tilde{\alpha}_p = \sup\left\{\beta_p \in [0,\infty)\colon \mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}_p}\right) - b_p^{(2)} = O(t^{-\beta_p/2}) \text{ as } t \to \infty\right\} \tag{87}$$

Assume that $\tilde{\alpha}_p > 0$ for all $p = 1, \dots, n$. Then the L^2- torsion $T_X^{(2)} \in \mathbb{R}^+$ can be defined by

$$\begin{aligned}\log T_X^{(2)} = \frac{1}{2}\sum_{p=1}^n (-1)^p p\ \Bigg[\frac{d}{ds}\left(\frac{1}{\Gamma(s)}\int_0^1 \mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}'_p}\right) t^{s-1}\, dt\right)\Bigg|_{s=0} \\ + \int_1^\infty t^{-1}\, \mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}'_p}\right)\, dt\Bigg],\end{aligned} \tag{88}$$

where $\widetilde{\Delta}'_p$ denotes the restriction of $\widetilde{\Delta}_p$ to the orthogonal complement of $\ker \widetilde{\Delta}_p$ and the first integral is defined near $s = 0$ by analytic continuation. This definition can be generalized to all finite dimensional representations ρ of Γ, if the corresponding Novikov-Shubin invariants are all positive. Then the L^2-torsion $T_X^{(2)}(\rho)$ is defined as in (88). If there exists $c > 0$ such that the spectrum of $\Delta_p(\rho)$ is bounded from below by c, then the integral

$$\int_0^\infty \mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}_p(\rho)}\right) t^{s-1}\, dt$$

converges for $\mathrm{Re}(s) > n/2$ and admits a meromorphic continuation to $\mathbb{C}$ which is holomorphic at $s = 0$. Thus, if there is a positive lower bound of the spectrum of all $\Delta_p(\rho)$, $p = 1, \ldots, n$, then $T_X^{(2)}(\rho)$ can be defined in the usual way by

$$\log T_X^{(2)}(\rho) = \frac{1}{2}\sum_{p=1}^{n}(-1)^p p \frac{d}{ds}\left(\frac{1}{\Gamma(s)}\int_0^\infty \mathrm{Tr}_\Gamma\left(e^{-t\widetilde{\Delta}_p(\rho)}\right) t^{s-1}\, dt\right)\bigg|_{s=0}.$$

Let $\Gamma = \pi_1(X, x_0)$ and let $(\Gamma_i)_{i\in\mathbb{N}_0}$ be a tower of normal subgroups of finite index of $\Gamma = \Gamma_0$. Let $X_i = \Gamma_i\backslash\widetilde{X}$, $i \in \mathbb{N}_0$, be the corresponding covering of X. Let T_X and $T_X^{(2)}$ denote the analytic torsion and L^2-torsion with respect to the trivial representation. Lück [Lu2, Conjecture 7.4] has made the following conjecture.

Conjecture 3. *For every closed Riemannian manifold X the L^2-torsion $T_X^{(2)}$ exists and for a sequence of coverings $(X_i \to X)_{i\in\mathbb{N}}$ as above one has*

$$\lim_{i\to\infty}\frac{\log T_{X_i}}{[\Gamma\colon\Gamma_i]} = \log T_X^{(2)}.$$

One is tempted to make this conjecture for any finite dimensional representation ρ.

5.2 Compact Locally Symmetric Spaces

Now we turn to the locally symmetric case. Let $X = \Gamma\backslash\widetilde{X}$, where $\widetilde{X} = G/K$ is a Riemannian symmetric space of non-positive curvature and $\Gamma \subset G$ is a discrete, torsion free, cocompact subgroup. Let τ be an irreducible finite dimensional complex representation of $\mathbf{G}$. Let $E_\tau \to X$ be the flat vector bundle associated with the representation $\tau|_\Gamma$ of Γ. By [MM], E_τ can be equipped with a canonical Hermitian fiber metric, called admissible, which is unique up to scaling. Let $\Delta_p(\tau)$ be the Laplace operator on p-forms with values in E_τ, with respect to the choice of any admissible fiber metric in E_τ. Let $T_X(\tau)$ be the corresponding analytic torsion. Let $\widetilde{\Delta}_p(\tau)$ be the Laplace operator on $\widetilde{E}_\tau$-valued p-forms on $\widetilde{X}$. Let $\widetilde{E}_\tau \to \widetilde{X}$ be the homogeneous vector bundle defined by $\tau|_K$. By [MM] there is a canonical isomorphism

$$E_\tau \cong \Gamma\backslash\widetilde{E}_\tau$$

and the metric on E_τ is induced by the homogeneous metric on $\widetilde{E}_\tau$. Thus

$$C^\infty(\widetilde{X}, \widetilde{E}_\tau) \cong (C^\infty(G) \otimes V_\tau)^K. \tag{89}$$

Let R be the right regular representation of G in $C^\infty(G)$ and let $R(\Omega)$ be the operator in $(C^\infty(G) \otimes V_\tau)^K$ induced by the Casimir element. Then with respect to the isomorphism (89) we have

$$\widetilde{\Delta}_p(\tau) = -R(\Omega) + \lambda_\tau \,\mathrm{Id}$$

(see [MM]). This implies that the heat operator $e^{-t\widetilde{\Delta}_p(\tau)}$ is a convolution operator given by a kernel

$$H_t^{p,\tau} : G \to \mathrm{End}(\Lambda^p \mathfrak{p}^* \otimes V_\tau).$$

Let $h_t^{p,\tau} \in C^\infty(G)$ be defined by $h_t^{p,\tau}(g) = \operatorname{tr} H_t^{p,\tau}(g)$, $g \in G$. Then it follows from (86) that

$$\mathrm{Tr}_\Gamma \left(e^{-t\widetilde{\Delta}_p(\tau)} \right) = \mathrm{vol}(X) h_t^{p,\tau}(1). \tag{90}$$

Now one can use the Plancherel theorem to compute $h_t^{p,\tau}(1)$ and determine its asymptotic behavior as $t \to 0$ and $t \to \infty$. For the trivial representation this was carried out in [Ol] and for strongly acyclic τ in [BV]. So let $\widetilde{\Delta}_p(\tau)'$ be the restriction of $\widetilde{\Delta}_p(\tau)$ to the orthogonal complement of the kernel of $\widetilde{\Delta}_p(\tau)$. Now let

$$\tilde{\alpha}_p(X,\tau) := \sup \left\{ \beta_p \in [0,\infty) \colon \mathrm{Tr}_\Gamma \left(e^{-t\widetilde{\Delta}_p(\tau)'} \right) = O(t^{-\beta_p/2}) \ \text{ as } \ t \to \infty \right\}, \tag{91}$$

$p = 0, \dots, n$, be the twisted Novikov-Shubin invariants. Assume that $\tilde{\alpha}_p(X,\tau) > 0$, $p = 0, \dots, n$. Then the L^2-torsion $T_X^{(2)}(\tau)$ is defined. By [Ol, Theorem 1.1] this is the case for the trivial representation. Furthermore, if τ is strongly acyclic, then $\tilde{\alpha}_p(X,\tau) = \infty$ for all p. Using the definition of the L^2-torsion, it follows that

$$\log T_X^{(2)}(\tau) = \mathrm{vol}(X) t_{\widetilde{X}}^{(2)}(\tau), \tag{92}$$

where $t_{\widetilde{X}}^{(2)}(\tau)$ is a constant that depends only on $\widetilde{X}$ and τ.

Now let (Γ_j) be sequence of torsion free cocompact lattices in G. Let $X_j = \Gamma_j \backslash \widetilde{X}$ and assume that $\mathrm{inj}(X_j) \to \infty$ if $j \to \infty$. A representation $\tau \colon G \to \mathrm{GL}(V)$ is called *strongly acyclic*, if there is $c > 0$ such that the spectrum of $\Delta_{X_j,p}(\tau)$ is contained in $[c, \infty)$ for all $j \in \mathbb{N}$ and $p = 0, \dots, n$.

Now let $\mathbf{G}$ be a connected semisimple algebraic $\mathbb{Q}$-group. Let $G = \mathbf{G}(\mathbb{R})$. Then it is proved in [BV] that strongly acyclic representations exist. For such representations Bergeron and Venkatesh [BV, Theorem 4.5] established the following theorem.

Theorem 5.1. *Let $\tau\colon G \to \mathrm{GL}(V)$ be strongly acyclic. Then*

$$\lim_{j\to\infty} \frac{\log(T_{X_j}(\tau))}{\mathrm{vol}(X_j)} = t_X^{(2)}(\tau), \tag{93}$$

where $X_j = \Gamma_j\backslash\widetilde{X}$ and $\mathrm{inj}(X_j) \to \infty$ as $j \to \infty$.

The number $t_X^{(2)}(\tau)$ can be computed using the Plancherel theorem. Let $\delta(G) = \mathrm{rank}(G) - \mathrm{rank}(K)$ be the fundamental rank or "deficiency" of G. By [BV, Proposition 5.2] one has

Proposition 5.2. *If $\delta(G) \neq 1$, then $t_X^{(2)}(\tau) = 0$. For $\delta(G) = 1$ one has*

$$(-1)^{\frac{\dim \widetilde{X}-1}{2}} t_X^{(2)}(\tau) > 0.$$

We note that the simple Lie groups G with $\delta(G) = 1$ are $\mathrm{SL}_3(\mathbb{R})$ and $\mathrm{SO}(p,q)$ with pq odd, especially $G = \mathrm{SO}^0(2m+1,1)$ is a group with fundamental rank 1.

Next we briefly recall the main steps of the proof of Theorem 5.1. To indicate the dependence of the heat operator and other quantities on the covering X_j, we use the subscript X_j. The uniform spectral gap at 0 implies that there exist constants $C, c > 0$ such that for all $p = 0,\dots,n$, $j \in \mathbb{N}$ and $t \geq 1$ one has

$$\mathrm{Tr}\left(e^{-t\Delta_{X_j,p}(\tau)}\right) \leq Ce^{-tc}\,\mathrm{vol}(X_j) \tag{94}$$

(see [BV]). This is the key result that makes the method to work. Let

$$K_{X_j}(t,\tau) := \frac{1}{2}\sum_{p=1}^{n}(-1)^p p\,\mathrm{Tr}\left(e^{-t\Delta_{X_j,p}(\tau)}\right). \tag{95}$$

Using (94) it follows that the analytic torsion can be defined by

$$\log T_{X_j}(\tau) = \frac{d}{ds}\left(\frac{1}{\Gamma(s)}\int_0^\infty K_{X_j}(t,\tau)t^{s-1}\,dt\right)\bigg|_{s=0}. \tag{96}$$

Let $T > 0$. Then we can split the integral and rewrite the right-hand side as

$$\log T_{X_j}(\tau) = \frac{d}{ds}\left(\frac{1}{\Gamma(s)}\int_0^T K_{X_j}(t,\tau)t^{s-1}\,dt\right)\bigg|_{s=0} + \int_T^\infty K_{X_j}(t,\tau)t^{-1}\,dt.$$

By (94) there exist $C, c > 0$ such that

$$\frac{1}{\mathrm{vol}(X_j)}\left|\int_T^\infty K_{X_j}(t,\tau)t^{-1}\,dt\right| \leq Ce^{-cT} \tag{97}$$

for all $j \in \mathbb{N}_0$ and $T > 1$. To deal with the first term one can use the Selberg trace formula. Put

$$k_t^\tau := \frac{1}{2}\sum_{p=1}^{n}(-1)^p p h_t^{p,\tau}.$$

Then the Selberg trace formula gives

$$K_{X_j}(t,\tau) = \operatorname{vol}(X_j)k_t^\tau(1) + H_{X_j}(k_t^\tau),$$

where $H_{X_j}(k_t^\tau)$ is the contribution of the hyperbolic conjugacy classes. Using (90) and the definition of k_t^τ, it follows that

$$\frac{d}{ds}\left(\frac{1}{\Gamma(s)}\int_0^T k_t^\tau(1)t^{s-1}\,dt\right)\Big|_{s=0} = t_{\widetilde{X}}^{(2)}(\tau) + O\left(e^{-cT}\right)$$

as $T \to \infty$. Regrouping the terms of the hyperbolic contribution $H_{X_j}(k_t^\tau)$ as in (67) it follows that the corresponding integral divided by $\operatorname{vol}(X_j)$ converges to 0 as $j \to \infty$. This proves the theorem.

One expects Theorem 5.1 to be true in general. However, if there is no spectral gap at zero, one cannot argue as above. The key problem is to control the small eigenvalues as $j \to \infty$. Sufficient conditions on the behavior of the small eigenvalues are discussed in [Lu2] and in the 3-dimensional case also in [BSV].

In view of the potential applications to the cohomology of arithmetic groups, discussed in the next section, it is very desirable to extend Theorem 5.1 to the non-compact case. The first problem one faces is that the corresponding Laplace operators have a nonempty continuous spectrum and therefore, the heat operators are not trace class and the analytic torsion cannot be defined as above. This problem has been studied by Raimbault [Ra1] for hyperbolic 3-manifolds and in [MP2] for hyperbolic manifolds of any dimension.

So let $G = \mathrm{SO}^0(n,1)$, $K = \mathrm{SO}(n)$ and $\widetilde{X} = G/K$. Equipped with a suitably normalized G-invariant metric, $\widetilde{X}$ becomes isometric to the n-dimensional hyperbolic space $\mathbb{H}^n$. Let $\Gamma \subset G$ be a torsion free lattice. Then $X = \Gamma\backslash\widetilde{X}$ is an oriented n-dimensional hyperbolic manifold of finite volume. As above, let $\tau\colon G \to \mathrm{GL}(V)$ be a finite dimensional complex representation of G. The first step is to define a regularized trace of the heat operators $e^{-t\Delta_p(\tau)}$. To this end one uses an appropriate height function to truncate X at sufficient high level $Y > Y_0$ to get a compact manifold $X(Y) \subset X$ with boundary $\partial X(Y)$, which consists of a disjoint union of $n-1$-dimensional tori. Let $K^{p,\tau}(t,x,y)$ be the kernel of the heat operator $e^{-t\Delta_p(\tau)}$. Using the spectral resolution of $\Delta_p(\tau)$, it follows that there exist $\alpha(t) \in \mathbb{R}$ such that $\int_{X(Y)} \operatorname{tr} K^{p,\tau}(t,x,x)\,dx - \alpha(t)\log Y$ has a limit as $Y \to \infty$. Then we define the regularized trace as

$$\operatorname{Tr}_{\mathrm{reg}}\left(e^{-t\Delta_p(\tau)}\right) := \lim_{Y\to\infty}\left(\int_{X(Y)} \operatorname{tr} K^{p,\tau}(t,x,x)\,dx - \alpha(t)\log Y\right). \tag{98}$$

We note that the regularized trace is not uniquely defined. It depends on the choice of truncation parameters on the manifold X. However, if $X_0 = \Gamma_0\backslash\mathbb{H}^n$ is given and if truncation parameters on X_0 are fixed, then every finite covering X of X_0 is canonically equipped with truncation parameters, namely one simply pulls back the height function on X_0 to a height function on X via the covering map.

Let θ be the Cartan involution of G with respect to $K = \mathrm{SO}(n)$. Let $\tau_\theta = \tau \circ \theta$. If $\tau \ncong \tau_\theta$, it can be shown that $\mathrm{Tr}_{\mathrm{reg}}\left(e^{-t\Delta_p(\tau)}\right)$ is exponentially decreasing as $t \to \infty$ and admits an asymptotic expansion as $t \to 0$. Therefore, the regularized zeta function $\zeta_{\mathrm{reg},p}(s;\tau)$ of $\Delta_p(\tau)$ can be defined as in the compact case by

$$\zeta_{\mathrm{reg},p}(s;\tau) := \frac{1}{\Gamma(s)} \int_0^\infty \mathrm{Tr}_{\mathrm{reg}}\left(e^{-t\Delta_p(\tau)}\right) t^{s-1}\, dt. \tag{99}$$

The integral converges absolutely and uniformly on compact subsets of the half-plane $\mathrm{Re}(s) > n/2$ and admits a meromorphic extension to the whole complex plane, which is holomorphic at $s = 0$. So in analogy with the compact case, the regularized analytic torsion $T_X(\tau) \in \mathbb{R}^+$ can be defined by the same formula (83).

In even dimension the analytic torsion is rather trivial. Therefore, we assume that $n = 2m + 1$. Furthermore, for technical reasons we assume that every lattice $\Gamma \subset G$ satisfies the following condition: For every Γ-cuspidal parabolic subgroup P of G one has

$$\Gamma \cap P = \Gamma \cap N_P, \tag{100}$$

where N_P denotes the unipotent radical of P. Let Γ_0 be a fixed lattice in G and let $X_0 = \Gamma_0\backslash\widetilde{X}$. Let $\Gamma_j, j \in \mathbb{N}$, be a sequence of finite index torsion free subgroups of Γ_0. This sequence is called to be *cusp uniform*, if the tori which arise as cross sections of the cusps of the manifolds $X_J := \Gamma_j\backslash\widetilde{X}$ satisfy some uniformity condition (see [MP2, Definition 8.2]).

The following theorem and its corollaries are established in [MP2]. One of the main results of [MP2] is the following theorem which may be regarded as an analog of Theorem 5.1 for oriented finite volume hyperbolic manifolds.

Theorem 5.3. *Let Γ_0 be a lattice in G and let Γ_i, $i \in \mathbb{N}$, be a sequence of finite-index normal subgroups which is cusp uniform and such that each Γ_i, $i \geq 1$, is torsion-free and satisfies* (100). *If* $\lim_{i\to\infty}[\Gamma_0 : \Gamma_i] = \infty$ *and if each* $\gamma_0 \in \Gamma_0 - \{1\}$ *only belongs to finitely many* Γ_i*, then for each* τ *with* $\tau \neq \tau_\theta$ *one has*

$$\lim_{i\to\infty} \frac{\log T_{X_i}(\tau)}{[\Gamma : \Gamma_i]} = t^{(2)}_{\mathbb{H}^n}(\tau)\,\mathrm{vol}(X_0). \tag{101}$$

In particular, if under the same assumptions Γ_i *is a tower of normal subgroups, i.e.* $\Gamma_{i+1} \subset \Gamma_i$ *for each* i *and* $\cap_i \Gamma_i = \{1\}$, *then* (101) *holds.*

For hyperbolic 3-manifolds, Theorem 5.3 was proved by Raimbault [Ra1] under additional assumptions on the intertwining operators. We emphasize that the above theorem holds without any additional assumptions.

Now we specialize to arithmetic groups. First consider $\Gamma_0 := \mathrm{SO}^0(n,1)(\mathbb{Z})$. Then Γ_0 is a lattice in $\mathrm{SO}^0(n,1)$. For $q \in \mathbb{N}$ let $\Gamma(q)$ be the principal congruence subgroup of Γ_0 of level q. Using a result of Deitmar and Hoffmann [DH], it follows that the family of principal congruence subgroups $\Gamma(q)$ is cusp uniform [MP2, Lemma 10.1]. Thus Theorem 5.3 implies the following corollary (see [MP2, Corollary 1.3]).

Corollary 5.4. *For any finite dimensional irreducible representation τ of $\mathrm{SO}^0(n,1)$ with $\tau \not\cong \tau_\theta$ the principal congruence subgroups $\Gamma(q)$, $q \geq 3$, of $\Gamma_0 := \mathrm{SO}^0(n,1)(\mathbb{Z})$ satisfy*

$$\lim_{q\to\infty} \frac{\log T_{X_q}(\tau)}{[\Gamma\colon\Gamma(q)]} = t^{(2)}_{\mathbb{H}^n}(\tau)\,\mathrm{vol}(X_0),$$

where $X_q := \Gamma(q)\backslash\mathbb{H}^n$ and $X_0 := \Gamma_0\backslash\mathbb{H}^n$.

We recall that by Proposition 5.2 we have $(-1)^{\frac{n-1}{2}} t^{(2)}_{\mathbb{H}^n}(\tau) > 0$.

Next we consider the 3-dimensional case. We note that every lattice $\Gamma \subset \mathrm{SO}^0(3,1)$ can be lifted to a lattice $\Gamma' \subset \mathrm{Spin}(3,1)$. Moreover, recall that there is a natural isomorphism $\mathrm{Spin}(3,1) \cong \mathrm{SL}_2(\mathbb{C})$. If ρ is the standard representation of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{C}^2$, then the finite dimensional irreducible representations of $\mathrm{SL}_2(\mathbb{C})$ are given by $\mathrm{Sym}^p\,\rho \otimes \mathrm{Sym}^q\,\bar{\rho}$, $p,q \in \mathbb{N}$, where Sym^k denotes the k-th symmetric power and $\bar{\rho}$ denotes the complex conjugate representation to ρ. One has $(\mathrm{Sym}^p\,\rho \otimes \mathrm{Sym}^q\,\bar{\rho})_\theta = \mathrm{Sym}^q\,\rho \otimes \mathrm{Sym}^p\,\bar{\rho}$. For $D \in \mathbb{N}$ square free let $\mathcal{O}_D$ be the ring of integers of the imaginary quadratic field $\mathbb{Q}(\sqrt{-D})$ and let $\Gamma(D) := \mathrm{SL}_2(\mathcal{O}_D)$. Then $\Gamma(D)$ is a lattice in $\mathrm{SL}_2(\mathbb{C})$. If $\mathfrak{a}$ is a non-zero ideal in $\mathcal{O}_D$, let $\Gamma(\mathfrak{a})$ be the associated principal congruence subgroup of level $\mathfrak{a}$. Then Theorem 5.1 implies the following corollary (see [MP2, Corollary 1.4]).

Corollary 5.5. *Let $D \in \mathbb{N}$ be square free. Let $\mathfrak{a}_i$ be a sequence of non-zero ideals in $\mathcal{O}_D$ such that each $N(\mathfrak{a}_i)$ is sufficiently large and such that $\lim_{i\to\infty} N(\mathfrak{a}_i) = \infty$. Put $X_D := \Gamma(D)\backslash\mathbb{H}^3$ and $X_i := \Gamma(\mathfrak{a}_i)\backslash\mathbb{H}^3$. Let $\tau = \mathrm{Sym}^p\,\rho \otimes \mathrm{Sym}^q\,\bar{\rho}$ with $p \neq q$. Then one has*

$$\lim_{i\to\infty} \frac{\log T_{X_i}(\tau)}{[\Gamma(D)\colon\Gamma(\mathfrak{a}_i)]} = t^{(2)}_{\mathbb{H}^3}(\tau)\,\mathrm{vol}(X_D).$$

5.3 Applications to the Cohomology of Arithmetic Groups: The Cocompact Case

Theorem 5.1 has interesting consequences for the cohomology of arithmetic groups. Let $\Gamma \subset G$ be a discrete, torsion free, cocompact subgroup. Let $\tau\colon G \to \mathrm{GL}(V)$ be a finite dimensional real representation and let $E \to X$ be the associated vector bundle. Choose a fiber metric h in E. Assume that there exist a Γ-invariant lattice $M \subset V$. Let $\mathcal{M}$ be the associated local system of free $\mathbb{Z}$-modules over X. Then we have $E = \mathcal{M} \otimes \mathbb{R}$. Let $H^*(X, \mathcal{M})$ be the cohomology of X with coefficients in $\mathcal{M}$. Each $H^q(X, \mathcal{M})$ is a finitely generated $\mathbb{Z}$-module. Let $H^q(X, \mathcal{M})_{\mathrm{tors}}$ be the torsion subgroup and

$$H^q(X; \mathcal{M})_{\mathrm{free}} = H^q(X, \mathcal{M})/H^q(X, \mathcal{M})_{\mathrm{tors}}.$$

We identify $H^q(X, \mathcal{M})_{\mathrm{free}}$ with a subgroup of $H^q(X, E)$. Let $\langle \cdot, \cdot \rangle_q$ be the inner product in $H^q(X, E)$ induced by the L^2-metric on $\mathcal{H}^q(X, E)$. Let $e_1, \dots, e_{r_q}$ be a basis of $H^q(X, \mathcal{M})_{\mathrm{free}}$ and let G_q be the Gram matrix with entries $\langle e_k, e_l \rangle$. Put

$$R_q(\tau, h) = \sqrt{|\det G_q|}, \quad q = 0, \dots, n.$$

Define the "regulator" $R(\tau, h)$ by

$$R(\tau, h) = \prod_{q=0}^{n} R_q(\tau, h)^{(-1)^q}. \tag{102}$$

Recall that the Reidemeister torsion $T_X^{\mathrm{top}}(\tau, h)$ depends on the metric h through the choice of an orthonormal basis in the cohomology $H^*(X, E)$, where the inner product in $H^*(X, E_\tau)$ is defined as above. The key result relating Reidemeister torsion and cohomology is the following proposition.

Proposition 5.6. *Let τ be a unimodular representation of Γ on a finite-dimensional $\mathbb{R}$-vector space V. Let $M \subset V$ be a Γ-invariant lattice and let $\mathcal{M}$ be the associated local system of finitely generated free $\mathbb{Z}$-modules on X. Let h be a fiber metric in the flat vector bundle $E = \mathcal{M} \otimes \mathbb{R}$. Then we have*

$$T_X^{\mathrm{top}}(\tau, h) = R(\tau, h) \cdot \prod_{q=0}^{n} \left| H^q(X, \mathcal{M})_{\mathrm{tors}} \right|^{(-1)^{q+1}}. \tag{103}$$

Especially, if $\tau|_\Gamma$ is acyclic, i.e., if $H^*(X, E) = 0$, then $T_X^{\mathrm{top}}(\tau, h)$ is independent of h and we denote it by $T_X^{\mathrm{top}}(\tau)$. Moreover, $R(\tau, h) = 1$. Then $H^*(X, \mathcal{M})$ is a torsion group and one has

$$T_X^{\mathrm{top}}(\tau) = \prod_{q=0}^{n} \left| H^q(X, \mathcal{M}) \right|^{(-1)^{q+1}}.$$

Representations τ of G which admit a Γ-invariant lattice arise in the following arithmetic situation. Let $\mathbf{G}$ be a semisimple algebraic group defined over $\mathbb{Q}$ and let $G = \mathbf{G}(\mathbb{R})$. Let $\Gamma \subset \mathbf{G}(\mathbb{Q})$ be an arithmetic subgroup. Let V_0 be a $\mathbb{Q}$-vector space and let $\rho\colon \mathbf{G} \to \mathrm{GL}(V_0)$ be a rational representation. Then there exists a lattice $M \subset V_0$ which is invariant under Γ and $V_0 = M \otimes_{\mathbb{Z}} \mathbb{Q}$. Let $V = V_0 \otimes_{\mathbb{Q}} \mathbb{R}$ and let $\tau\colon G \to \mathrm{GL}(V)$ be the representation induced by ρ. Then $M \subset V$ is a Γ-invariant lattice.

Assume that $\Gamma \subset \mathbf{G}(\mathbb{Q})$ is cocompact in G (equivalently assume that $\mathbf{G}$ is anisotropic). Then it is proved in [BV] that strongly acyclic arithmetic Γ-modules M exist. Assume that $\delta(G) = 1$. Let M be a strongly acyclic arithmetic Γ-module. Then by (84), Theorem 5.1 and Proposition 5.2 it follows that there exists a constant $C > 0$, which depends on G and M, such that

$$\lim_{j\to\infty} \sum_{k=0}^{n} (-1)^{k+\frac{\dim(\widetilde{X})-1}{2}} \frac{\log |H_k(X_j, \mathcal{M})|}{[\Gamma\colon \Gamma_j]} = C \operatorname{vol}(\Gamma\backslash\widetilde{X}) \tag{104}$$

(see [BV, (1.4.2)]). This implies the following theorem of Bergeron and Venkatesh [BV, Theorem 1.4].

Theorem 5.7. *Suppose that* $\delta(\widetilde{X}) = 1$. *Then strongly acyclic arithmetic* Γ*-modules exist. For any such module* M*,*

$$\liminf_{j} \sum_{k\equiv a \,(\mathrm{mod}\ 2)} \frac{\log |H_k(X_j, \mathcal{M})|}{[\Gamma\colon \Gamma_j]} \geq C \operatorname{vol}(\Gamma\backslash\widetilde{X}),$$

where $a = (\dim(\widetilde{X}) - 1)/2$ *and* $C > 0$ *depends only on* G *and* M*.*

In Theorem 5.7, one cannot in general isolate the degree which produces torsion. A conjecture of Bergeron and Venkatesh [BV, Conjecture 1.3] claims the following.

Conjecture 4. *The limit*

$$\lim_{j\to\infty} \frac{\log |H_k(X_j, \mathcal{M})_{\mathrm{tors}}|}{[\Gamma\colon \Gamma_j]}$$

exists for each k *and is zero unless* $\delta(G) = 1$ *and* $k = \frac{\dim(\widetilde{X})-1}{2}$. *In that case, it is always positive and equal to a positive constant* $C_{G,M}$*, which can be explicitly described, times* $\operatorname{vol}(\Gamma\backslash\widetilde{X})$.

An example, for which this conjecture can be verified is $G = \mathrm{SL}(2, \mathbb{C})$.

If the representation τ of G is not acyclic, various difficulties occur. First of all, the spectrum of the Laplace operators has no positive lower bound which causes the problem with the small eigenvalues discussed above in the context of analytic torsion. Secondly the regulator $R(\tau, h)$ is in general nontrivial. It turns out to be rather difficult to control the growth of the regulator. Of particular interest is the case of the trivial representation, i.e., the integer homology $H_k(X_j, \mathbb{Z})$. The 3-dimensional case has been studied in [BSV]. In this paper the authors discuss conditions which imply that the results of [BV] on strongly acyclic local systems can be extended to the case of the trivial local system. There are conditions on the cohomology and the spectrum of the Laplace operator on 1-Forms. The conditions on the spectrum are as follows. Let $(\Gamma_i)_{i\in\mathbb{N}}$ be a sequence of cocompact congruence subgroups of a fixed arithmetic subgroup $\Gamma \subset \mathrm{SL}(2,\mathbb{C})$. Let $X_i = \Gamma_i\backslash\mathbb{H}^3$ and put $V_i := \mathrm{vol}(X_i)$. Let $\lambda_j^{(i)}$ $j \in \mathbb{N}$, be the eigenvalues of the Laplace operator on 1-forms of X_i. Assume:

(1) For every $\varepsilon > 0$ there exists $c > 0$ such that

$$\limsup_{i\to\infty} \frac{1}{V_i} \sum_{0<\lambda_j^{(i)}\le c} |\log \lambda_j^{(i)}| \le \varepsilon.$$

(2) $b_1(X_i, \mathbb{Q}) = o(\frac{V_i}{\log V_i})$.

Let T_{X_i} be the analytic torsion with respect to the trivial local system. As shown in [BSV], conditions (1) and (2) imply that

$$\frac{\log T_{X_i}}{V_i} \longrightarrow t^{(2)}_{\mathbb{H}^3} = -\frac{1}{6\pi}, \quad i \to \infty.$$

Unfortunately, it seems to be difficult to verify (1) and (2). The other problem is to estimate the growth of the regulator (see [BSV]). We note that condition (1) is equivalent to the following condition (1′).

(1′) Let $d\mu_1$ be the spectral measure of $\widetilde{\Delta}_1$. For every $c > 0$ one has

$$\frac{1}{V_i} \sum_{0<\lambda_j^{(i)}\le c} \log \lambda_j^{(i)} \longrightarrow \int_0^c \log \lambda \, d\mu_1(\lambda), \quad i \to \infty.$$

There is a certain similarity with the limit multiplicity problem.

Finally we note that there is related work by Calegari and Venkatesh [CaV] who use analytic torsion to compare torsion in the cohomology of different arithmetic subgroups of $\mathrm{SL}(2,\mathbb{C})$ and establish a numerical form of a Jacquet-Langlands correspondence in the torsion case.

5.4 The Finite Volume Case

Many important arithmetic groups are not cocompact. So it is desirable to extend the results of the previous section to the finite volume case. In order to achieve this one has to deal with the following problems.

(1) Define an appropriate regularized version $T_X^{\mathrm{reg}}(\rho)$ of the analytic torsion for a finite volume locally symmetric space $X = \Gamma\backslash\widetilde{X}$ and establish the analog of (93). So let $\Gamma_j \subset \Gamma$ be a sequence of subgroups of finite index and $X_j := \Gamma_j\backslash\widetilde{X}$, $j \in \mathbb{N}$. Assume that $\mathrm{vol}(V_j) \to \infty$. Under appropriate additional assumptions on the sequence $(\Gamma_j)_{j\in\mathbb{N}}$ one has to show that

$$\lim_{j\to\infty} \frac{\log T_{X_j}^{\mathrm{reg}}(\rho)}{\mathrm{vol}(X_j)} = t_{\widetilde{X}}^{(2)}(\rho).$$

(2) Show that $T_X^{\mathrm{reg}}(\rho)$ has a topological counterpart $T_X^{\mathrm{top}}(\rho)$, possibly the Reidemeister torsion of an intersection complex.
(3) If E_ρ is arithmetic, i.e., if there is a local system of finite rank free $\mathbb{Z}$-modules $\mathcal{M}$ over X such that $E_\rho = \mathcal{M} \otimes \mathbb{R}$, establish an analog of (103).
(4) Estimate the growth of the regulator.

For hyperbolic manifolds (1) has been proved in [Ra1] in the 3-dimensional case and in [MP1] and [MP2] in general. It would be very interesting to extend these results to the higher rank case. $\mathrm{SL}(3,\mathbb{R})$ seems to be doable.

Raimbault [Ra2] has studied (2) in the 3-dimensional case and established a kind of asymptotic equality of analytic and Reidemeister torsion, which is sufficient for the present purpose. Of course, the goal is to prove an exact equality. For hyperbolic manifolds there is some recent progress [AR]. Unfortunately, this paper does not cover the relevant flat bundles. The method requires that the flat bundle can be extended to the boundary at infinity. This is not the case for the flat bundles which arise from representations of G by restriction to Γ. Pfaff [Pf] has established a gluing formula for the regularized analytic torsion of a hyperbolic manifold, which reduces the problem to the case of a cusp.

(4) has been studied by Raimbault [Ra2] for 3-dimensional hyperbolic manifolds. It turns out to be very difficult. The real cohomology never vanishes. There is always the part of the cohomology coming from the boundary. This is the Eisenstein cohomology introduced by Harder [Ha]. These cohomology classes are represented by Eisenstein classes, which are rational cohomology classes. The problem is to estimate the denominators of the Eisenstein classes which seems to be a hard problem.

References

[AB1] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, I. Samet, On the growth of Betti numbers of locally symmetric spaces. C. R. Math. Acad. Sci. Paris **349**(15–16), 831–835 (2011)

[AB2] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, I. Samet, *On the Growth of L^2-Invariants for Sequences of Lattices in Lie Groups* (2012). arXiv:1210.2961

[AR] P. Albin, F. Rochon, D. Sher, *Analytic Torsion and R-Torsion of Witt Representations on Manifolds with Cusps* (2014). arXiv:1504.02418

[Ar1] J. Arthur, A trace formula for reductive groups. I. Terms associated to classes in $G(\mathbb{Q})$. Duke Math. J. **45**(4), 911–952 (1978)
[Ar2] J. Arthur, On a family of distributions obtained from Eisenstein series. I. Application of the Paley-Wiener theorem. Am. J. Math., 104(6), 1243–1288 (1982)
[Ar3] J. Arthur, On a family of distributions obtained from Eisenstein series. II. Explicit formulas. Am. J. Math., 104(6), 1289–1336 (1982)
[Ar4] J. Arthur, The local behavior of weighted orbital integrals. Duke Math. J. **56** (1988), 223–293
[Ar5] J. Arthur, The trace formula in invariant form. Ann. of Math. (2), **114**(1), 1–74 (1981)
[Ar6] J. Arthur, On a family of distributions obtained from orbits. Can. J. Math. **38**(1), 179–214 (1986)
[Ar7] J. Arthur, A measure on the unipotent variety. Can. J. Math. **37**(6), 1237–1274 (1985)
[Ar8] J. Arthur, *The Endoscopic Classification of Representations Orthogonal and Symplectic Groups*. American Mathematical Society Colloquium Publications, vol. 61. (American Mathematical Society, Providence, RI, 2013)
[Av] V.G. Avakumović, Über die Eigenfunktionen auf geschlossenen Riemannschen Mannigfaltigkeiten. Math. Z. **65**, 327–344 (1956)
[BGV] N. Berline, E. Getzler, M. Vergne, Heat kernels and Dirac operators, in *Grundlehren Text Editions.* (Springer, Berlin, 2004)
[BV] N. Bergeron, A. Venkatesh, The asymptotic growth of torsion homology for arithmetic groups. J. Inst. Math. Jussieu **12**(2), 391–447
[BSV] N. Bergeron, M.H. Sengün, A. Venkatesh, *Torsion Homology and Cycle Complexity of Arithmetic Manifolds* (2014). arXiv: 1401.6989v1
[BE] P. Borwein, T. Erdélyi, Sharp extensions of Bernstein's inequality to rational spaces. Mathematika **43**(2), 413–423 (1996)
[BZ] J.-M. Bismut, W. Zhang, An extension of a theorem by Cheeger and Ḿuller. With an appendix by Franǫois Laudenbach. Astérisque **205**, 235 pp. (1992); C.J. Bushnell, G. Henniart, An upper bound on conductors for pairs. J. Number Theory **65**(2),183–196 (1997)
[BG] A. Borel, H. Garland, Laplacian and the discrete spectrum of an arithmetic group. Am. J. Math. **105**(2), 309–335 (1983).
[CaV] F. Calegari, A. Venkatesh, *A Torsion Jacquet–Langlands Correspondence* (2012). arXiv:1212.3847
[Ch] J. Cheeger, Analytic torsion and the heat equation. Ann. Math. (2) **109**(2), 259–322 (1979)
[CD] L. Clozel, P. Delorme, Le théorème de Paley-Wiener invariant pour les groupes de Lie réductifs. Invent. Math. **77**(3), 427–453 (1984)
[CKP] J. Cogdell, H. Kim, I. Piatetski-Shapiro, F. Shahidi, Functoriality for the classical groups. Publ. Math. Inst. Hautes Études Sci **99**, 163–233 (2004)
[CV] Y. Colin de Verdière, Pseudo-laplaciens. II. Ann. Inst. Fourier **33**(2), 87–113 (1983)
[Co] L. Corwin, The Plancherel measure in nilpotent Lie groups as a limit of point measures, Math. Zeitschrift **155**, 151–162 (1977)
[DW1] D.L. DeGeorge, N.R. Wallach, Limit formulas for multiplicities in $L^2(\Gamma\backslash G)$. Ann. Math. (2) **107**, 133–150 (1978)
[DW2] D.L. DeGeorge, N.R. Wallach, Limit formulas for multiplicities in $L^2(\Gamma\backslash G)$. II. The tempered spectrum. Ann. Math. (2) **109**(3), 477–495 (1979)
[DH] A. Deitmar, W. Hoffmann, Spectral estimates for towers of noncompact quotients. Can. J. Math. **51**(2), 266–293 (1999)
[De] P. Delorme, Formules limites et formules asymptotiques pour les multiplicités dans $L^2(G/\Gamma)$. Duke Math. J. **53**, 691–731 (1986)
[Di] J. Dixmier, Les C^*-algèbres et leurs représentations. Deuxième édition. Cahiers Scientifiques, Fasc. XXIX. Gauthier-Villars Éditeur, Paris (1969)
[Do] H. Donnelly, On the cuspidal spectrum for finite volume symmetric spaces. J. Differ. Geom. **17**(2), 239–253 (1982)
[DG] J.J. Duistermaat, V.W. Guillemin, The spectrum of positive elliptic operators and periodic bicharacteristics. Invent. Math. **29**(1), 39–79 (1975)

[DKV] J.J. Duistermaat, J.A.C. Kolk, V.S. Varadarajan, Spectra of compact locally symmetric manifolds of negative curvature. Invent. Math. **52**, 27–93 (1979)

[FL1] T. Finis, E. Lapid, On the spectral side of Arthur's trace formula—combinatorial setup. Ann. Math. (2) **174**(1), 197–223 (2011)

[FL2] T. Finis, E. Lapid, *An Approximation Principle for Congruence Subgroups II: Application to the Limit Multiplicity Problem* (2015). arXiv:1504.04795

[FLM1] T. Finis, E. Lapid, W. Müller, On the spectral side of Arthur's trace formula— absolute convergence. Ann. Math. (2) **174**, 173–195 (2011)

[FLM2] T. Finis, E. Lapid, W. Müller, Limit multiplicities for principal congruence subgroups of GL(n) and SL(n). J. Inst. Math. Jussieu **14**(3), 589–638 (2015)

[FLM3] T. Finis, E. Lapid, W. Müller, On the degrees of matrix coefficients of intertwining operators. Pac. J. Math. **260** (2), 433–456 (2012)

[Ga] L. Garding, Dirichlet's problem for linear elliptic partial differential equations. Math. Scand. **1**, 55–72 (1953)

[GGP] I.M. Gel'fand, M.I. Graev, I.I. Pyatetskii-Shapiro, Representation theory and automorphic functions. Reprint of the 1969 edition, in *Generalized Functions*, vol. 6 (Academic, Boston, MA, 1990)

[Gi] P.B. Gilkey, Invariance theory, the heat equation, and the Atiyah-Singer index theorem, in *Studies in Advanced Mathematics*, 2nd edn. (CRC Press, Boca Raton, FL, 1995)

[Ha] G. Harder, Eisenstein cohomology of arithmetic groups. The case GL2. Invent. Math. **89**(1), 37–118 (1987)

[He] S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962)

[Ho] L. Hörmander, The spectral function of an elliptic operator. Acta Math. **121**, 193–218 (1968)

[Hu] M.N. Huxley, Scattering matrices for congruence subgroups, in *Modular forms (Durham, 1983). Ellis Horwood Series in Mathematics and its Applications: Statistics, Operational Research*, 141–156 (Horwood, Chichester, 1984)

[Ka] V. Kala, *Density of Self-Dual Automorphic Representations of* $\mathrm{GL}_N(\mathbb{A}_{\mathbb{Q}})$ (2014). arXiv:1406.0385

[La1] R.P. Langlands, *Euler Products* (Yale University Press, New Haven, CT, 1971); A. James, K. Whittemore, *Lecture in Mathematics* given at Yale University, 1967, Yale Mathematical Monographs, 1

[La2] R.P. Langlands, *On the Functional Equations Satisfied by Eisenstein Series*. Lecture Notes in Mathematics, vol. 544 (Springer, Berlin/New York, 1976)

[LM] E. Lapid, W. Müller, Spectral asymptotics for arithmetic quotients of SL(n,R)/SO(n), Duke Math. J. **149**(1), 117–155 (2009)

[LV] E. Lindenstrauss, A. Venkatesh, Existence and Weyl's law for spherical cusps forms, Geom. Funct. Anal. **17**(1), 220–251 (2007)

[Lo] J. Lott, Heat kernels on covering spaces and topological invariants. J. Differ. Geom. **35**(2), 471–510 (1992)

[Lub] A. Lubotzky, Subgroup growth and congruence subgroups. Invent. Math. **119**, 267–295 (1995)

[Lu1] W. Lück, Approximating L^2-invariants by their finite-dimensional analogues, Geom. Funct. Anal. **4**(4), 455–481 (1994)

[Lu2] W. Lück, *Survey on Approximating L^2-Invariants by Their Classical Counterparts: Betti Numbers, Torsion Invariants and Homological Growth* (2015). arXiv:1501.07446

[MV] V. Mathai, L^2-analytic torsion. J. Funct. Anal. **107**(2), 369–386 (1992)

[MM] Y. Matsuhsima, S. Murakami, On vector bundle valued harmonic forms and automorphic forms on symmetric Riemannian manifolds. Ann. Math. (2) **78**, 365–416 (1963)

[Ma1] J. Matz, *Weyl's Law for Hecke Operators on GL(n) Over Imaginary Quadratic Number Fields* (2013). arXiv:1310.6525

[MT] J. Matz, N. Templier, *Sato-Tate Equidistribution for Families of Hecke-Maass Forms on SL(n,R)/SO(n)* (2015). arXiv:1505.07285

[Mia] R.J. Miatello, The Minakshisundaram-Pleijel coefficients for the vector-valued heat kernel on compact locally symmetric spaces of negative curvature. Trans. Am. Math. Soc. **260**(1), 1–33 (1980)

[Mi] S.D. Miller, On the existence and temperedness of cusp forms for $SL_3(\mathbb{Z})$. J. Reine Angew. Math. **533**, 127–169 (2001)

[MP] S. Minakshisundaram, A. Pleijel, Some properties of the eigenfunctions of the Laplace-operator on Riemannian manifolds. Can. J. Math. **1**, 242–256 (1949)

[MW] C. Mœglin, J.-L. Waldspurger, Le spectre résiduel de GL(n), Ann. Sci. l'École Norm. Sup. (4) **22**(4), 605–674 (1989)

[Mu1] W. Müller, Analytic torsion and R-torsion of Riemannian manifolds. Adv. Math. **28**(3), 233–305 (1978)

[Mu2] W. Müller, The trace class conjecture in the theory of automorphic forms. Ann. Math. (2), **130**(3), 473–529 (1989)

[Mu3] W. Müller, Analytic torsion and R-torsion for unimodular representations. J. Am. Math. Soc. **6**(3), 721–753 (1993)

[Mu4] W. Müller, B. Speh, Absolute convergence of the spectral side of the Arthur trace formula for GL_n. Geom. Funct. Anal. **14**, 58–93 (2004). With an appendix by E.M. Lapid

[Mu5] W. Müller, Weyl's law for the cuspidal spectrum of SL_n. Ann. Math. (2) **165**(1), 275–333 (2007)

[Mu6] W. Müller, On the spectral side of the Arthur trace formula. Geom. Funct. Anal. **12**, 669–722 (2002)

[Mu7] W. Müller, Weyl's law in the theory of automorphic forms, in *Groups and Analysis*. London Mathematical Society Lecture Note Series, vol. 354 (Cambridge University Press, Cambridge, 2008), pp. 133–163

[MP1] W. Müller, J. Pfaff, Analytic torsion of complete hyperbolic manifolds of finite volume. J. Funct. Anal. **263**(9), 2615–2675 (2012)

[MP2] W. Müller, J. Pfaff, The analytic torsion and its asymptotic behavior for sequences of hyperbolic manifolds of finite volume, J. Funct. Anal. **267**(8), 2731–2786 (2014)

[MS] W. Müller, B. Speh, Absolute convergence of the spectral side of the Arthur trace formula for GL_n. With an appendix by E.M. Lapid. Geom. Funct. Anal. **14**(1), 58–93 (2004)

[Ol] M. Olbrich, L^2-invariants of locally symmetric spaces. Doc. Math. **7**, 219–237 (2002)

[OW] S. Osborne, G. Warner, The theory of Eisenstein systems, in *Pure and Applied Mathematics*, vol. 99 (Academic, [Harcourt Brace Jovanovich, Publishers], New York/London, 1981)

[Pf] J. Pfaff, *A Gluing Formula for the Analytic Torsion on Hyperbolic Manifolds With Cusps* (2013). arXiv:1312.6384

[Pr] K. Prachar, *Primzahlverteilung* (Springer, Berlin/Ǵottingen/Heidelberg, 1957)

[Ra1] J. Raimbault, *Asymptotics of Analytic Torsion for Hyperbolic Three-Manifolds* (2012). arXiv:1212.3161

[Ra2] J. Raimbault, *Analytic, Reidemeister and Homological Torsion for Congruence Three-Manifolds* (2013). arXiv:1307.2845

[RS] D.B. Ray, I.M. Singer, *R*-torsion and the Laplacian on Riemannian manifolds. Adv. Math. **7**, 145–210 (1971)

[Rez] A. Reznikov, Eisenstein matrix and existence of cusp forms in rank one symmetric spaces. Geom. Funct. Anal. **3**(1), 79–105 (1993)

[RoS] J. Rohlfs, B. Speh, On limit multiplicities of representations with cohomology in the cuspidal spectrum. Duke Math. J. **55**, 199–211 (1987)

[Sa1] P. Sarnak, Spectra of hyperbolic surfaces. Bull. Am. Math. Soc. (N.S.) **40**(4), 441–478 (2003)

[Sa2] P. Sarnak, On cusp forms, in *The Selberg Trace Formula and Related Topics (Brunswick, Maine, 1984)*. Contemporary Mathematics, vol. 53 (American Mathematical Society, Providence, RI, 1986), pp. 393–407

[Sa3] P. Sarnak, Notes on the generalized Ramanujan conjectures, in *Harmonic Analysis, the Trace Formula, and Shimura Varieties. Clay Mathematics Proceedings*, vol. 4 (American Mathematical Society, Providence, RI, 2005), pp. 659–685

[SST] P. Sarnak, S.-W. Shin, N. Templier, *Families of L-Functions and Their Symmetry* (2014). arXiv:1401.5507

[Sau] F. Sauvageot, Principe de densité pour les groupes réductifs. Compositio Math. **108**, 151–184 (1997)

[Sav] G. Savin, Limit multiplicities of cusp forms. Invent. Math., **95**, 149–159 (1989)

[Se1] A. Selberg, *Harmonic Analysis*, *Collected Papers*, vol. I (Springer, Berlin, 1989), pp. 624–674

[Sha1] F. Shahidi, On certain L-functions. Am. J. Math. **103**(2), 297–355 (1981)

[Sha2] F. Shahidi, On the Ramanujan conjecture and finiteness of poles for certain L-functions. Ann. Math. (2) **127**(3), 547–584 (1988)

[Shi] S.-W. Shin, Automorphic Plancherel density theorem. Israel J. Math. **192**(1), 83–120 (2012)

[Shu] M.A. Shubin, Pseudodifferential operators and spectral theory, 2nd edn. (Springer, Berlin, 2001)

[Wa1] G. Warner, Selberg's trace formula for nonuniform lattices: the R-rank one case, in *Studies in Algebra and Number Theory*. Advances in Mathematics Supplementary Studies, vol. 6 (Academic, New York/London, 1979)

[We] H. Weyl, Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). Math. Annalen **71**, 441–479 (1912)

Families of L-Functions and Their Symmetry

Peter Sarnak, Sug Woo Shin, and Nicolas Templier

Abstract A few years ago the first-named author proposed a working definition of a family of automorphic L-functions. Then the work by the second and third-named authors on the Sato–Tate equidistribution for families made it possible to give a conjectural answer for the universality class introduced by Katz and the first-named author for the distribution of the zeros near $s = 1/2$. In this article we develop these ideas fully after introducing some structural invariants associated to the arithmetic statistics of a family.

2010 *Mathematics Subject Classification.* 11F70, 14G10

1 Definition of Families and Conjectures

The zoo of automorphic cusp forms π on $G = \mathrm{GL}_n$ over $\mathbb{Q}$ correspond bijectively to their standard completed L-functions $\Lambda(s, \pi)$ and they constitute a countable set containing species of different types. For example, there are self-dual forms, ones corresponding to finite Galois representations, to Hasse–Weil zeta functions of varieties defined over $\mathbb{Q}$, to Maass forms, etc. From a number of points of view (including the nontrivial problem of isolating special forms) one is led to

P. Sarnak
Institute for Advanced Study, Princeton, NJ 08540, USA

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
e-mail: sarnak@math.princeton.edu

S.W. Shin
Department of Mathematics, UC Berkeley, Berkeley, CA 94720, USA

Korea Institute for Advanced Study, 85 Hoegiro, Dongdaemun-gu, Seoul 130-722, Republic of Korea
e-mail: sug.woo.shin@berkeley.edu

N. Templier (✉)
Department of Mathematics, Cornell University, Ithaca, NY 14853, USA
e-mail: templier@math.cornell.edu

W. Müller et al. (eds.), *Families of Automorphic Forms and the Trace Formula*, Simons Symposia, DOI 10.1007/978-3-319-41424-9_13

study such $\Lambda(s,\pi)$'s in families in which the π's have similar characteristics. Some applications demand the understanding of the behavior of the L-functions as π varies over a family. Other applications involve questions about an individual L-function. In practice a *family* is investigated as it arises.

For example, the density theorems of Bombieri [Bomb] and Vinogradov [Vin65] are concerned with showing that in a suitable sense most Dirichlet L-functions have few violations of the Riemann hypothesis, and as such it is a powerful substitute for the latter. Other examples are the GL_2 subconvexity results which are proved by deforming the given form in a family (see [IS00] and [MV10] for accounts). In the analogous function field setting the notion of a family of zeta functions is well defined, coming from the notion of a family of varieties defined over a base. Here too the power of deforming in a family in order to understand individual members is amply demonstrated in the work of Deligne [Del80]. In the number field setting there is no formal definition of a family $\mathfrak{F}$ of L-functions.

Our aim is to give a working definition [Sarn08] for the formation of a family which will correspond to parametrized subsets of $\mathsf{A}(G)$, the set of isobaric automorphic representations on $G(\mathbb{A})$. As far as we can tell these include almost all families that have been studied. For the most part our families can be investigated using the trace formula, monodromy groups in arithmetic geometry and the geometry of numbers. We introduce below the following invariants of a family: Sato–Tate measure, indicators, homogeneity type, rank and average root number. Thereby we put forth some structural properties of the arithmetic statistics of families and wish it to contribute towards a general framework. These invariants lead to a determination of the distribution of the zeros near $s=\frac{1}{2}$ of members of the family. For the high zeros of a given $\Lambda(s,\pi)$, it was shown in [RS96] that the local scaled spacing statistics follows the universal GUE laws (Gaussian Unitary Ensemble). We find that the low-lying zeros (i.e., near $s=\frac{1}{2}$) of a family $\mathfrak{F}$ follow one of the three universality classes computed in [KS-b] as the scaling limits of monodromy groups.

For the purpose of defining a family we will assume freely various standard conjectures when convenient. While many of these are well out of reach, important special cases are known and in passing to families they become approachable. We begin by reviewing some notation and invariants associated with individual π's.

Any $\pi\in\mathsf{A}(G)$ decomposes as an isobaric sum $\pi=\pi_1\boxplus\pi_2\boxplus\cdots\boxplus\pi_r$ with π_j an automorphic cusp form on GL_{n_j}, $n_1+n_2+\cdots+n_r=n$ [JPSS83]. Correspondingly $\Lambda(s,\pi)=\Lambda(s,\pi_1)\Lambda(s,\pi_2)\cdots\Lambda(s,\pi_r)$ and this reduces the study to that of cusp forms, which will be our main focus. Here and elsewhere the central character of π is normalized to be unitary and the functional equation relates $\Lambda(s,\pi)$ to $\Lambda(1-s,\tilde{\pi})$, where $\tilde{\pi}$ is the representation contragradient to π. Furthermore we assume that the central character of π is trivial when restricted to $\mathbb{R}_{>0}$ (equivalently is of finite order). Denote by $\mathsf{A}_{\mathrm{cusp}}(G)$ the subset of cuspidal automorphic representations on G. By our normalization this is a countable set. For $\pi\in\mathsf{A}_{\mathrm{cusp}}(G)$ its conductor $N(\pi)$ is a positive integer defined as the product over appropriate powers of various primes v at which π_v is ramified (here $\pi\simeq\otimes_v\pi_v$). It is the integer appearing in the functional equation for $\Lambda(s,\pi)$ (see [GJ72]). The analytic conductor $C(\pi)$ as defined in [IS00] is the product of $N(\pi)$ with a factor coming from π_∞. The analytic conductor measures the "complexity" of π (and also the local density of zeros of $\Lambda(s,\pi)$ near $s=\frac{1}{2}$) much like the height of rational points in diophantine analysis. As in that setting the set $S(x)=\{\pi,\ C(\pi)<x\}$ is finite (see [Brum06]). It would be interesting to derive a *Weyl–Schanuel* type theorem for this "universal" family, giving the asymptotic behavior of $S(x)$ as x goes to infinity.[1] We will use $C(\pi)$ to order the elements of a family $\mathfrak{F}\subset \mathsf{A}_{\mathrm{cusp}}(G)$. The root number $\varepsilon(\pi)=\varepsilon(\frac{1}{2},\pi)$ is a complex number of unit modulus that occurs as the sign of the functional equation relating $\Lambda(s,\pi)$ to $\Lambda(1-s,\tilde{\pi})$ ([GJ72]). We say that π is self-dual if $\pi=\tilde{\pi}$ and in this case $\varepsilon(\pi)=\pm 1$. For a self-dual π, $\Lambda(s,\pi\times\pi)=\Lambda(s,\pi,\mathrm{sym}^2)\Lambda(\pi,s,\wedge^2)$ and π is said to be *orthogonal* or *symplectic* according to the first or the second factor above carrying the pole at $s=1$ (in the orthogonal case π is a standard functorial transfer of a form on a symplectic group or an even orthogonal group and similarly for the symplectic case from an odd orthogonal group). The symplectic case can only occur if n is even, and if π is orthogonal then $\varepsilon(\pi)=1$ ([Lapid] and [Art13, Theorem 1.5.3.(b)]).

The question of the distribution of π_v as v varies over the primes is the generalized *Sato–Tate* problem and its formulation is problematic. Each π_v is a point in the unitary dual of $G(\mathbb{Q}_v)$ and according to the generalized Ramanujan conjectures it lies in the tempered dual $\widehat{G(\mathbb{Q}_v)}^{\mathrm{temp}}$ if π is cuspidal (see [Sarn05]). Moreover for v large π_v is unramified and hence can be identified with a diagonal unitary matrix $(\alpha_{\pi_v}(1),\ldots,\alpha_{\pi_v}(n))$ that is a point in an n-dimensional torus quotient T_c/W, where T_c is the product of n unit circles and W is the permutation group on n letters (we divide by W since the matrix is only determined up to $\mathrm{GL}_n(\mathbb{C})$ conjugacy). The generalized Sato–Tate conjecture asserts that these π_v's become equidistributed with respect to a measure $\mu_{\mathrm{ST}}(\pi)$ on T_c (or more precisely T_c/W) as $v\to\infty$. If π corresponds to a finite irreducible Galois representation ρ, whose image is denoted $B\subset \mathrm{GL}_n(\mathbb{C})$, then $\mu_{\mathrm{ST}}(\pi)$ exists by the Chebotarev density theorem and is equal to the push forward μ_B of Haar measure on B to the tempered conjugacy classes $G_c^{\#}\simeq T_c/W$ of $G_c\simeq U(n)$, a maximal compact subgroup which is isomorphic to a compact unitary group. Langlands [Lan04] suggests that for any π there is a (possibly non-connected) reductive algebraic subgroup B of $\mathrm{GL}_n(\mathbb{C})$ such that $\mu_{\mathrm{ST}}(\pi)=\mu_B$ where the latter denotes the pushforward of the Haar measure on $B\cap G_c$. In [Ser12] Serre gives a precise formulation in terms of Lie group data and a constructive approach when π comes from geometry. In any case it follows from the analytic properties of $\Lambda(s,\pi)$ and $\Lambda(s,\pi\times\tilde{\pi})$ that

$$
\begin{aligned}
\int_{T_c}\chi(t)\mu_{\mathrm{ST}}(t)&=\int_{B_c}(\alpha_1(\theta)+\cdots+\alpha_n(\theta))\mu_B(\theta)=0\\
\int_{T_c}|\chi(t)|^2\,\mu_{\mathrm{ST}}(t)&=\int_{B_c}|\alpha_1(\theta)+\cdots+\alpha_n(\theta)|^2\,\mu_B(\theta)=1
\end{aligned}
\tag{1}
$$

[1] Brumley–Milicevic [BM] have recently done so for $\mathrm{GL}(2)/\mathbb{Q}$.

where $\chi(t) = \mathrm{tr}(t)$. Hence B is irreducible in $\mathrm{GL}_n(\mathbb{C})$. In general it may happen that $\mu_{B_1} = \mu_{B_2}$ for B_1 not conjugate to B_2 in $\mathrm{GL}_n(\mathbb{C})$ (see [AYY]), so that B may not be determined up to conjugacy. For our purposes it is μ_B that is important, so let $I(T) := I^{\#}(\mathrm{GL}_n(\mathbb{C}))$ denote the countable set of probability measures that come from irreducible subgroups B as above. Langlands's assertion is that $\mu_{\mathrm{ST}}(\pi)$ is in $I(T)$ and we will then loosely speak of π being of type B if $\mu_{\mathrm{ST}}(\pi) = \mu_B$, even if B is not unique.

We turn to our formulation of a *parametric family* $\mathfrak{F}$ of automorphic representations on G. $\mathfrak{F} = (W, F)$ consists of a parameter space W and a map $F : W \to \mathsf{A}(G)$, and is based on two very general conjectural means of constructing automorphic forms: spectral and geometric.

Harmonic Families Let H be a connected reductive algebraic group defined over $\mathbb{Q}$ and $\mathsf{A}(H)$ the set of discrete automorphic representations on $H(\mathbb{A})$. A harmonic (spectral) set $\mathfrak{H}$ of forms on H is a subset of $\mathsf{A}(H)$ consisting of forms π which are unramified outside of a finite set of places, or for which $\pi_v \in B_v$ for v in a finite set of places and B_v is a nice subset of positive Plancherel measure in the unitary dual $\widehat{H(\mathbb{Q}_v)}$, or a hybrid of these conditions. The important thing is that these sets $\mathfrak{H}$ can be isolated using the trace formula on $H(\mathbb{Q})\backslash H(\mathbb{A})$. Let $r : {}^L H \to {}^L G$ be a representation of the corresponding Langlands dual group, then functoriality gives a map $r_* : \mathfrak{H} \to \mathsf{A}(G)$ and defines a parametric family $\mathfrak{F} = (\mathfrak{H}, r_*)$ of automorphic representations on G.

Geometric Families These parametric families come from zeta functions which are formed from counting solutions to algebraic equations over finite fields, namely Dedekind zeta functions and Hasse–Weil zeta functions. Let $\mathcal{W}$ be an open dense subscheme of $\mathbb{A}^m_{\mathbb{Q}} = \mathrm{Spec}\,\mathbb{Q}[W_1, \ldots, W_m]$ (or $\mathbb{Z}[W_1, \ldots, W_m]$ if we work over $\mathbb{Z}$) with $W_1, \ldots, W_m$ transcendental parameters. Let X be a smooth and proper scheme over $\mathcal{W}$ with integral fibers. So specializing the base to $w = (w_1, \ldots, w_m) \in \mathcal{W}(\mathbb{Q})$ yields a smooth proper variety X_w over $\mathbb{Q}$.

As part of the data defining the corresponding parametric family we may restrict the w's locally over $\mathbb{R}$ to lie in a real projective cone C to ensure that the *discriminant* $D(w)$ (see Remark (i) below) corresponding to the family has controlled size in terms of the height of w as a point of $\mathbb{P}^m(\mathbb{Q})$. Put $W = C \cap \mathcal{W}(\mathbb{Q})$. For the w's in W we get in this way a Hasse–Weil L-function (if X_w is zero dimensional, a Dedekind zeta function) on the étale cohomology group in a fixed degree d

$$L(s, H^d_{et}(X_w \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \overline{\mathbb{Q}}_l)) \tag{2}$$

by specializing to w. (See Appendix 1 below for the definition. It involves a choice of a field isomorphism $\iota : \overline{\mathbb{Q}}_l \simeq \mathbb{C}$, though the expectation is that (2) is independent of the choice.) Note that the dimension n of the d-th cohomology of the closed fibers of X is constant over W. Assuming the modularity conjecture (Conjecture 4 in Appendix 1) we get a map $F : W \to \mathsf{A}(G) = \mathsf{A}(\mathrm{GL}_n)$ such that $F(w)$ is the $|\det|^{d/2}$-

twist of the automorphic representation corresponding to (2) (so that $L(s, F(w)) = L(s + \frac{d}{2}, H^d_{et}(X_w \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \overline{\mathbb{Q}}_l)))$. This gives us a parametric family $\mathfrak{F} = (W, F)$ of automorphic forms.

Remarks. (i) Our aim is a statistical study of members of the family. For parametric families $\mathfrak{F} = (W, F)$ this means ordering the members according to the sets

$$\{w \in W : \ C(F(w)) < x\},$$

and this can be achieved with the caveat that one first replaces $C(F(w))$ by a dominating gauge function $D(w) = \mathrm{Disc}(X_w)$ which approximates $C(F(w))$. There are many cases for which F is essentially one-to-one and then $F(W)$ is a parametrized subset of $\mathsf{A}(G)$. We call such a subset a *parametrized family*, where we can drop the parameter space W since the study of $F(W)$ when ordered by conductor does not depend on the parametrization.

(ii) Various operations can be performed on parametric families such as union; $\mathfrak{F} \cup \mathfrak{F}'$ which is the family with parameter space $W \sqcup W'$ and the corresponding map F or F'. For the product $\mathfrak{F} \times \mathfrak{F}'$ we take as parameters $W \times W'$ and the map $F(w) \times F(w')$, where the last is the Rankin product giving a form on $GL_{nn'}$ if $\mathfrak{F}$ is on GL_n and $\mathfrak{F}'$ on $\mathrm{GL}_{n'}$. (The product $\pi \times \pi'$ corresponds to the functorial map $\rho \otimes \rho'$ where ρ and ρ' are the standard representations of GL_n and $\mathrm{GL}_{n'}$.) In this product setting we allow one of the factors to be a singleton in forming the product family. One is tempted to form other Boolean operations such as intersections on parametrized families and this can be done (yielding new families) in many cases. However in general global diophantine equations on the parameters W intervene and these can lead to subsets of $\mathsf{A}(G)$ which are not families in our sense and which don't obey any of the predictions below (see Sect. 3).

(iii) There are various subsets of $\mathsf{A}(G)$ which aren't realized in terms of our general constructions which form natural families and which probably obey the conjectures below. These are defined through Galois and class groups and other arithmetic invariants. For example, the set of π's which correspond to finite Galois representations, and among these the set of π's for which the image of the corresponding Galois representation is a given group B (up to conjugation). Another is the set of Hecke zeta functions of class groups of number fields of a given degree, cf. Sect. 3.5 below. While abelian H's above can be studied to the same extent as our general families using class field theory, we don't know how to study these families in any generality and hence we do not include them as part of the general definition. Note, however, that one can often produce large parametric subfamilies of these arithmetic "families."

(iv) The twist by $|\det|^{d/2}$ in the definition of a geometric family (W, F) is introduced to ensure that the (non-archimedean) local components of $\pi = F(w)$ are unitary, cf. the remark below Conjecture 1 and the last paragraph of Appendix 1.

With the definition of a parametric family in place we put forth the basic conjectures about them. These may look far-fetched at first, but unlike the study of individual forms, they can be studied and there is ample evidence (by way of proof) for the conjectures. We will give various examples in Sect. 2.

For $\pi \in \mathsf{A}(G)$ we write the finite part of its standard L-function as

$$L(s,\pi) = \prod_{v<\infty} L(s,\pi_v) = \sum_{n=1}^{\infty} \frac{\lambda_\pi(n)}{n^s}. \tag{3}$$

In studying a (harmonic or geometric) parametric family $\mathfrak{F} = (W, F)$ the first thing one needs to count asymptotically is

$$|\mathfrak{F}(x)| = \sum_{w\colon\, C(F(w))<x} 1. \tag{4}$$

Since with our normalization there are finitely many automorphic representations $\pi \in \mathsf{A}(G)$ of conductor less than x, this count is indeed finite as soon as F is finite-to-one. This means that there are obvious cases that should be excluded, for example if the L-map r were to factor through ${}^{L}H \to W_{\mathbb{Q}}$ for harmonic families or if X were isotrivial for geometric families. If F is not finite-to-one, we impose suitable constraints on the parameter space such as restriction to a projective cone in the geometric setting which renders the finiteness (see Sect. 2.1).

Also implicit in our definition is the requirement that a family has infinite cardinality. This infiniteness is not strictly necessary at first since, for example, Conjecture 1 below reduces to the Sato–Tate conjecture for an individual representation but then as we move on to finer arithmetic invariants and to the universality conjecture this becomes critical. Thus we assume from now on that the parameter space W is infinite. Then $|\mathfrak{F}(x)| \to \infty$ as $x \to \infty$ and we expect an asymptotic for $|\mathfrak{F}(x)|$ that is a power of x, possibly with logarithms attached.

The following more general vertical limits should exist as $x \to \infty$ with a modest uniformity in $n \geqslant 1$:

$$\sum_{w\colon\, C(F(w))<x} \lambda_{F(w)}(n) = t_{\mathfrak{F}}(n) \cdot |\mathfrak{F}(x)| + O(n^A\, |\mathfrak{F}(x)|^{\delta}) \tag{5}$$

for some $A < \infty$ and $\delta < 1$. As mentioned before it is understood that in practice the ordering by conductor is often replaced by a closely related ordering involving an approximation in terms of the parameters in the family. Also in some explicit cases one might look at shells $\{w : \ x < C(F(w)) < x + H\}$ rather than balls, as smaller sets give finer individual information.

The structure of the limits in (5) can be described in terms of p-adic densities. Each $\pi \in \mathsf{A}(G)$ determines a point $(\pi_\infty, \pi_2, \pi_3, \pi_5, \ldots)$ in $\prod_v \widehat{\mathrm{GL}_n(\mathbb{Q}_v)}$, with its product topology (and π_v is unramified for v large enough).

Conjecture 1 (Sato–Tate Conjecture for $\mathfrak{F}$). *There is $p_0 = p_0(\mathfrak{F}) > 0$ such that if we order the w's in W by $C(F(w))$ then $F(w)$ is equidistributed in $Y := \prod_{p \geqslant p_0} \widehat{\mathrm{GL}_n(\mathbb{Q}_p)}$ with respect to a measure $\mu(\mathfrak{F})$ satisfying:*

(i) it is a probability measure and is supported on the tempered spectrum, hence the same holds for $\mu_p(\mathfrak{F})$ the projection of $\mu(\mathfrak{F})$ on $\widehat{\mathrm{GL}_n(\mathbb{Q}_p)}$,

(ii) it has a decomposition as a convex sum $\mu(\mathfrak{F}) = \nu_1 + \nu_2 + \cdots + \nu_r$ of positive measures such that each ν_j is a product measure on Y,

(iii) the average of the $\mu_p(\mathfrak{F})$ over p exists and defines the Sato–Tate *measure $\mu_{\mathrm{ST}}(\mathfrak{F})$ on T, that is*

$$\lim_{x\to\infty} \frac{1}{x} \sum_{p_0 \leqslant p < x} \log p \cdot \mu_p(\mathfrak{F})_{|T} =: \mu_{\mathrm{ST}}(\mathfrak{F}) \tag{6}$$

(for many families there is no need to average over p as $\lim_{p\to\infty} \mu_p(\mathfrak{F})_{|T} = \mu_{\mathrm{ST}}(\mathfrak{F})$),

(iv) $\mu_{\mathrm{ST}}(\mathfrak{F})$ is a probability measure and lies in the convex hull of $I(T)$.

The intuition for (iv) is clear enough, $\mu_{\mathrm{ST}}(\mathfrak{F})$ is a mixture of the measures $\mu_{\mathrm{ST}}(\pi)$ for the "generic" π in $F(W)$. The decomposition asserts that only finitely many B-types occur generically in $\mathfrak{F}$.

Remark. An analogous conjecture can be stated for the harmonic set $\mathfrak{H}$ itself without any reference to the L-morphism r. In this case $\widehat{\mathrm{GL}_n(\mathbb{Q}_p)}$ should be replaced by the unitary dual $\widehat{H(\mathbb{Q}_p)}$, and $\mathfrak{H}$ would be ordered by an invariant analogous to the conductor. This analogue of Conjecture 1 is treated in [ST16], for example. See Sect. 2.5 below for more details.

Remark. A priori $F(w)$ may not define a point in Y but one can simply interpret the equidistribution in the conjecture as asserting in particular that the number of w such that $F(w)$ does not lie in Y is statistically negligible. In other words, we need not assume that the local components π_p (for $p \geqslant p_0$) are unitary for each $\pi = F(w)$ to make sense of the conjecture, though we do expect them to be always unitary. For harmonic families, the unitarity of π_p is standard (assuming the Langlands functoriality map for r_* is compatible with the transfer of A-parameters via r) and comes down to the fact that the local A-parameters for $\mathrm{GL}_n(\mathbb{Q}_p)$ correspond to unitary representations. For geometric families, the unitarity is known in the case of good reduction but generally conditional on the weight-monodromy conjecture, cf. Remark (iv) above and Appendix 1.

For our purpose only some cruder invariants of $\mu_{\mathrm{ST}}(\mathfrak{F})$ are critical. These are the following indicators:

$$\begin{aligned} i_1(\mathfrak{F}) &= \int_T |\chi(t)|^2 \; \mu_{\mathrm{ST}}(\mathfrak{F})(t) \\ i_2(\mathfrak{F}) &= \int_T \chi(t)^2 \, \mu_{\mathrm{ST}}(\mathfrak{F})(t) \\ i_3(\mathfrak{F}) &= \int_T \chi(t^2) \, \mu_{\mathrm{ST}}(\mathfrak{F})(t) \end{aligned} \tag{7}$$

where $\chi(t) = \mathrm{tr}(t)$. We note the following equality:

$$i_3(\mathfrak{F}) = \lim_{x\to\infty} \frac{1}{x} \sum_{p<x} t_{\mathfrak{F}}(p^2) \log p. \tag{8}$$

Assuming (5) and the Riemann hypothesis for the relevant L-functions one can show the following.

(i) $i_1(\mathfrak{F}) \geqslant 1$ and $i_1(\mathfrak{F}) = 1$ iff almost all $F(w)$'s are cuspidal. In this case we say that $\mathfrak{F}$ is *essentially cuspidal* and for the most part we assume that this is the case. So for our statistical distribution questions the family is in $\mathsf{A}_{\mathrm{cusp}}(G)$.
(ii) $0 \leqslant i_2(\mathfrak{F}) \leqslant 1$ and $i_2(\mathfrak{F}) = 1$ iff almost all $F(w)$'s are self-dual and $i_2(\mathfrak{F}) = 0$ iff almost all $F(w)$'s are not self-dual. In the former case we say that $\mathfrak{F}$ is *essentially self-dual* and in the latter case $\mathfrak{F}$ is *non self-dual*. Note that $i_2(\mathfrak{F}) = 0 \Rightarrow i_3(\mathfrak{F}) = 0$.
(iii) $-1 \leqslant i_3(\mathfrak{F}) \leqslant 1$ and $i_3(\mathfrak{F}) = 1$ iff almost all $F(w)$'s are orthogonal and $i_3(\mathfrak{F}) = -1$ iff almost all $F(w)$'s are symplectic (called *essentially orthogonal* and *essentially symplectic*, respectively).

The above analysis allows one to compute for any $\mathfrak{F}$ satisfying (5) the Sato–Tate measures corresponding to the equidistribution of the $F(w)$'s for each of the three types. This gives positive measures $\mu_{\mathrm{U}}(\mathfrak{F})$, $\mu_{\mathrm{O}}(\mathfrak{F})$, and $\mu_{\mathrm{Sp}}(\mathfrak{F})$ on T such that

$$\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_{\mathrm{U}}(\mathfrak{F}) + \mu_{\mathrm{O}}(\mathfrak{F}) + \mu_{\mathrm{Sp}}(\mathfrak{F}). \tag{9}$$

The proportions of type of $F(w)$ in $\mathfrak{F}$ are determined from our indicators:

$$\begin{aligned} \mu_{\mathrm{U}}(\mathfrak{F})(T) + \mu_{\mathrm{O}}(\mathfrak{F})(T) + \mu_{\mathrm{Sp}}(\mathfrak{F})(T) &= 1 = i_1(\mathfrak{F}) \\ \mu_{\mathrm{O}}(\mathfrak{F})(T) + \mu_{\mathrm{Sp}}(\mathfrak{F})(T) &= i_2(\mathfrak{F}) \\ \mu_{\mathrm{O}}(\mathfrak{F})(T) - \mu_{\mathrm{Sp}}(\mathfrak{F})(T) &= i_3(\mathfrak{F}). \end{aligned} \tag{10}$$

As a complement it is helpful to note the following

$$\int_T \chi(t)\mu_{\mathrm{ST}}(\mathfrak{F})(t) = 0,$$

which follows from the fact that $\mathfrak{F}$ is essentially cuspidal and hence the absence of pole at $s = 1$ for almost all $F(w)$'s. Equivalently the limit

$$\lim_{x\to\infty} \frac{1}{x} \sum_{p<x} t_{\mathfrak{F}}(p) \log p$$

exists and always is equal to zero. This is to be compared with (8) above and (11) below.

The interpretation of these indicators in terms of B-types is clear. If $\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_B$ for some B, then by classical representation theory of compact groups, $i_1(\mathfrak{F}) = 1$ asserts that B is irreducible in $\mathrm{GL}_n(\mathbb{C})$, $i_2(\mathfrak{F}) = 1$ asserts that B is self-dual (as a subgroup of $\mathrm{GL}_n(\mathbb{C})$) and $i_3(\mathfrak{F})$ is the Frobenius–Schur indicator of B in $\mathrm{GL}_n(\mathbb{C})$. If assertion (iv) of Conjecture 1 holds, that is $\mu_{\mathrm{ST}}(\mathfrak{F})$ is a convex combination of μ_B's, then even though this decomposition need not be unique,[2] collecting the B-types according to their indices i_2, i_3 will reproduce the unique decomposition of $\mu_{\mathrm{ST}}(\mathfrak{F})$ given in (10).

The assertion (ii) of Conjecture 1 suggests that there is a stronger decomposition $\mathfrak{F} = \mathfrak{F}_1 \cup \cdots \cup \mathfrak{F}_r$, although this is not formally part of the conjecture. Here each subfamily $\mathfrak{F}_i$ of $\mathfrak{F}$ has asymptotic density $p_i \in [0, 1]$ and $\nu_i = p_i \mu(\mathfrak{F}_i)$. A family $\mathfrak{F}_i$ such that $\mu(\mathfrak{F}_i)$ is a direct product of measures on $\widehat{\mathrm{GL}_n(\mathbb{Q}_p)}$ is irreducible in some sense. For example, it is plausible that it implies that its horizontal average $\mu_{\mathrm{ST}}(\mathfrak{F}_i)$ be of the form μ_B for some irreducible B as above and thus $\mathfrak{F}_i$ is essentially homogeneous.

Indeed in many of the examples discussed in Sect. 2 such a B will be shown to exist (see notably Sects. 2.5 and 2.11). Then we can say that we have attached a Sato–Tate group $H(\mathfrak{F}) = B$ to the (irreducible) family $\mathfrak{F}$. We abstain from attempting a general conjecture about $H(\mathfrak{F})$ for at least two reasons, first because $H(\mathfrak{F})$ is not uniquely determined by $\mu_{\mathrm{ST}}(\mathfrak{F})$ so that a consistent definition seems hopeless, and second because for certain thin families the existence of $H(\mathfrak{F})$ is at the same level of difficulty as the existence of the Langlands group H_π for an individual π (see Sect. 2.8).

To put forth our prediction for the distribution of the zeros near $s = \frac{1}{2}$ of members of a family $\mathfrak{F}$ we need two further invariants attached to the family. The first is the rank, $r(\mathfrak{F})$, which is typically zero. The only case where we expect it might not be zero is for geometric families for which $s = \frac{1}{2}$ is a special value of $\Lambda(s, \pi)$ connected with a version of the generalization of the Birch and Swinnerton-Dyer conjecture. In the case of elliptic curves, if there are parametric, global rational solutions to the equations defining X (namely solutions in $\mathbb{Q}(W_1, \ldots, W_m)$) they will specialize to solutions of X_w for $w = (w_1, \ldots, w_m) \in W$. In general one considers not only rational points but rational algebraic cycles as in the conjecture by Tate, Lichtenbaum, Deligne, Bloch–Kato, Beilinson, and others.

[2] Jun Yu [Yu13] has given examples of this non-uniqueness.

The rank of the family is concerned with the rate of convergence of $\mu_p(\mathfrak{F})$ to $\mu_{\mathrm{ST}}(\mathfrak{F})$, and is defined to be

$$r(\mathfrak{F}) := \lim_{x\to\infty} \frac{1}{x} \sum_{p<x} -t_{\mathfrak{F}}(p)\sqrt{p}\log p. \tag{11}$$

For these geometric families one can show that $t_{\mathfrak{F}}(p) \ll p^{-1/2}$, so that (11) measures the next to leading term.

This formula (11) in the context of families and rank of elliptic surfaces has been proposed by Nagao [Nagao]. For X a family of elliptic curves forming an elliptic surface the equality of $r(\mathfrak{F})$ and the rank of $X/\mathbb{Q}(W)$ follows from the Tate conjecture for the surface, see [RS98]. The universal distributions for zeros near $s = \frac{1}{2}$ are concerned with fluctuations over the family after removing these persistent zeros at $s = \frac{1}{2}$. In what follows we assume that these have been removed or more simply that $r(\mathfrak{F}) = 0$ (according to definition (11)).[3]

The final invariant of $\mathfrak{F}$ that we need concerns the symplectic π's in $\mathfrak{F}$. For these the epsilon factor or root number $\varepsilon(\pi)$ can be $+1$ or -1 and it is not dictated by the Sato–Tate measure of $\mathfrak{F}$. According to (10) we can decompose the family into essential subfamilies $\mathfrak{F}_{\mathrm{U}}$, $\mathfrak{F}_{\mathrm{O}}$, $\mathfrak{F}_{\mathrm{Sp}}$ and we would like to decompose $\mathfrak{F}_{\mathrm{Sp}}$ further as $\mathfrak{F}_{\mathrm{Sp},+}$ and $\mathfrak{F}_{\mathrm{Sp},-}$ according to $\varepsilon = 1$ or -1. Since $\varepsilon(\pi)$ is given in terms of a product of local ε-factors at the ramified places of π, one can compute this decomposition analytically in many cases. However to do so in general involves computing averages over our parametric family of the Möbius function μ. Namely cancellations in sums

$$\sum_{w} \mu(M(w)) \tag{12}$$

where w varies over a large set in $\mathbb{Z}^m$ and $M \in \mathbb{Z}[W_1, \dots, W_m]$. These are predicted by natural generalization of Chowla's conjectures and are known in special cases [Helf04].

Assuming these allows one to refine the decomposition (10) as

$$\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_{\mathrm{U}}(\mathfrak{F}) + \mu_{\mathrm{O}}(\mathfrak{F}) + \mu_{\mathrm{Sp},+}(\mathfrak{F}) + \mu_{\mathrm{Sp},-}(\mathfrak{F}), \tag{13}$$

as well as the corresponding decomposition into essentially homogeneous subfamilies. In particular this reduces the study of the distribution of the low-lying zeros (as well as other statistical questions for $\mathfrak{F}$) to the case of $\mathfrak{F}$ being one of these four homogeneous families.

We now move to the main statistics of families that we will study, namely low-lying zeros. There are other statistics of interest notably moments of L-values, which

[3] For a homogeneous symplectic family of positive rank the third and fourth rows of Conjecture 2 below should read $\epsilon = (-1)^{r(\mathfrak{F})}$ and $\epsilon = -(-1)^{r(\mathfrak{F})}$, respectively.

are known since the work of Keating–Snaith [KS00] to relate to the symmetry type. See [CFKRS] and [Mich07] for a broad review of results and applications (there has been much progress since the appearance of these reviews). Our definition of families captures most of the examples studied to date (see Sect. 2), although not all of them (see Sect. 3). Our Conjecture 1 is a precise formulation of all the local statistics expected for families. In fact our notion of families provides a natural setting for the axiomatic recipes in [CFKRS], specifically Conjecture 1 as well as Conjecture 2 below are consistent with the family averaging assumptions made in [CFKRS, p. 82].

Write the zeros of $\Lambda(s,\pi)$ as $\frac{1}{2}+i\gamma_j^{(\pi)}$ (with multiplicities). For the purpose of studying the zeros near $s=\frac{1}{2}$ we scale the $\gamma_j^{(\pi)}$'s setting

$$\tilde{\gamma}_j^{(\pi)} := \gamma_j^{(\pi)} \frac{\log C(\pi)}{2\pi}. \tag{14}$$

This normalization is universal (i.e., there are no parameters in this process, the conductor $C(\pi)$ measures the local density). The four universality classes of distributions determined in [KS-b] are

(1) $\mathrm{U}(\infty)$: the scaling limit of the distribution near 1 of eigenvalues of matrices in $U(N)$, $N\to\infty$,
(2) $\mathrm{Sp}(\infty)$: the scaling limit of the distribution near 1 of eigenvalues of matrices in $\mathrm{USp}(2N)$, $N\to\infty$,
(3) $\mathrm{SO}_{\mathrm{even}}(\infty)$: the scaling limit of the distribution near 1 of eigenvalues of matrices in $\mathrm{SO}(2N)$, $N\to\infty$,
(4) $\mathrm{SO}_{\mathrm{odd}}(\infty)$: the scaling limit of the distribution near 1 of the eigenvalues of matrices in $\mathrm{SO}(2N+1)$, $N\to\infty$.

In the theoretical (rather than numerical) study of the $\tilde{\gamma}_j^{(\pi)}$'s as π varies over $\mathfrak{F}$ one computes the fluctuation r-level densities $W^{(r)}$, $r\geqslant 1$ (see [KS-b, KS99] and also the examples in Sect. 2), and these determine all other statistics.

We can finally state the

Conjecture 2 (Universality Conjecture). *Let $\mathfrak{F}$ be a rank* 0 *essentially homogeneous family. Then the low-lying zeros of the members of $\mathfrak{F}$ follow the laws in the following table:*

Homogeneity type of $\mathfrak{F}$	Symmetry type of $\mathfrak{F}$	Fluctuation r-level density
Non self-dual	$\mathrm{U}(\infty)$	$W_0^{(r)}$, $r\geqslant 1$
Orthogonal	$\mathrm{Sp}(\infty)$	$W_-^{(r)}$, $r\geqslant 1$
Symplectic $\varepsilon=1$	$\mathrm{SO}_{\mathrm{even}}(\infty)$	$W_+^{(r)}$, $r\geqslant 1$
Symplectic $\varepsilon=-1$	$\mathrm{SO}_{\mathrm{odd}}(\infty)$	$W_-^{(r)}$, $r\geqslant 1$

The r-variable densities $W^{(r)}$ are those from [KS-b]. Note that for the type Symplectic $\varepsilon = -1$, we omit the zero at $s = \frac{1}{2}$, which is there because of the sign of the functional equation when forming the densities of each member. The fact that $W_{-}^{(r)}$ is entered on lines 2 and 4 of this table is surprising but can be related to a similar coincidence at the level of the Weyl integration formula which is already observed in [Weyl].

In the formulation of Conjecture 2 above we have restricted ourselves to homogeneous families. This is for simplicity since one could easily consider families of forms which have mixed types, for example it often happens that essentially symplectic families have a root number that takes both the values 1 and -1 with positive proportion (see Sect. 2 for more examples). The low-lying zeros of such mixed families will be distributed according to the densities above, with weights determined by the decomposition (10).

The Sato–Tate conjecture for families (Conjecture 1) is in fact a theorem under mild assumptions as we shall explain with examples in the next section (see notably Sect. 2.11 for general geometric families and Sect. 2.5 for general harmonic families). The conjecture is independent of the analytic continuation of the corresponding L-functions and it only captures a portion of the arithmetic of the families.

This is in contrast to the universality conjecture (Conjecture 2) which is far reaching. It involves arithmetic cancellations which if true lie much deeper. Also its formulation relies on the zeros $\gamma_j^{(\pi)}$ and thus assumes the analytic continuation of $\Lambda(s, \pi)$ inside the critical strip, which is often a conditional statement. It seems an interesting question to find a substitute towards an unconditional formulation of the universality conjecture in all cases since the Symmetry Type is an intrinsic invariant of a family that should be independent of functoriality or modularity conjectures. One important source of additional invariants of families are p-adic ones (Selmer groups, p-adic L-functions, etc.) which also can be closely tied with the Symmetry Type, see notably Heath-Brown [Hea94], Bhargava–Shankar [BS13] as well as the recent [BKLPR] and the references therein.

Besides theoretical results yielding Conjecture 2 for restricted supports of test functions, an important piece of evidence comes from numerical experiments. There are robust algorithms [Rub05] to numerically compute the zeros and there is ample and excellent agreement for families of L-functions of low degrees.

Another important part of the picture is the function field analogue, where we work with the function field $\mathbb{F}_q(X)$ of a curve X and an ℓ-adic sheaf F of dimension d. See [ST16, p. 5] and [Katz01] for a discussion. For example, if $\mathcal{F}$ is irreducible self-dual orthogonal, then there is a natural pairing on $H^1(X, \mathcal{F})$ which is symplectic invariant by the action of Frobenius. This is consistent with Conjecture 2 and even stronger since it provides a spectral interpretation which is lacking over number fields.

As a corollary to the universality above we conclude that if n is odd, and $\mathfrak{F}$ a pure self-dual family (i.e., all members are self-dual) then its symmetry type is $\mathrm{Sp}(\infty)$ without any further assumptions (in this case $r(\mathfrak{F}) = 0$ since $s = \frac{1}{2}$ is not critical in the context of Deligne's special value conjectures [Del79]; see Appendix 2). Similarly a harmonic family $\mathfrak{F}$ arising from automorphic forms on split E_8, F_4 or G_2 will have symmetry type $\mathrm{Sp}(\infty)$ since all irreducible representations of their dual groups are self-dual and orthogonal [Ste68].

2 Examples

In this section we collect various examples of families, some old some new, which explicate the notions above and which prove in part the various claims and conjectures. It is this wealth of examples that we have tried to unify.

2.1 n=1

For $G = \mathrm{GL}(1)$, the set $\mathsf{A}(G)$ consists of all the primitive (nontrivial) Dirichlet characters χ so that parametrized families can be described explicitly. The most basic such family is

$$\mathfrak{F}^{(2)} = \{\chi : \ \chi^2 = 1\}. \tag{15}$$

In terms of our formation it arises either as all the self-dual forms on GL_1 or as the geometric family coming from the curve $Z^2 = W$ over $\mathbb{Z}[W]$, i.e., the Dedekind zeta function of quadratic extensions of $\mathbb{Q}$ after removing the constant factor of $\zeta(s)$. The last gives a parametric family which after a standard square-free sieving argument renders $\mathfrak{F}^{(2)}$ as a parametrized family. According to Conjecture 2 the Symmetry Type of $\mathfrak{F}^{(2)}$ should be $\mathrm{Sp}(\infty)$. There is ample evidence for this both numerical and theoretical (see Rubinstein's thesis [Rub01]). In this case where $\mathrm{GL}_1(\mathbb{C})$ is abelian and 1-dimensional, $I(T)$ corresponds bijectively to the finite subgroups of $T_c = \{z : \ |z| = 1\}$ together with T_c itself. The Sato–Tate measure for $\mathfrak{F}^{(2)}$ exists and is equal to μ_B where $B = \{1, -1\} \subset T$. In fact $\mu(\mathfrak{F}^{(2)}) = \prod_v \mu_B$ (that is μ_B at each place v), $r(\mathfrak{F}) = 0$ and $i_1(\mathfrak{F}^{(2)}) = i_2(\mathfrak{F}^{(2)}) = i_3(\mathfrak{F}^{(2)}) = 1$.

The precise statement about the low-lying zeros of $L(s, \chi)$ is as follows. For χ primitive of period q its conductor $N(\chi)$ is q and since $\chi_\infty = 1$ or sgn, the analytic conductor $C(\chi) = q$ as well. To form the r-level density sums write the zeros of $\Lambda(s, \chi)$, $\chi \in \mathfrak{F}^{(2)}$ as

$$\frac{1}{2} + i\gamma_j^{(\chi)}, \quad \text{with } j = \pm 1, \pm 2, \cdots$$

where $\gamma_j^{(\chi)} \geqslant 0$ if $j \geqslant 1$ and $\gamma_{-j}^{(\chi)} = -\gamma_j^{(\chi)}$.

For $\Phi \in \mathcal{S}(\mathbb{R}^r)$ even in each variable, form the r-level (scaled) densities for the low-lying zeros of $\Lambda(s, \chi)$:

$$D(\chi, \Phi) := \sum_{j_1, j_2, \cdots, j_r}^{*} \Phi\left(\frac{\gamma_{j_1}^{(\chi)} \log C(\chi)}{2\pi}, \ldots, \frac{\gamma_{j_1}^{(\chi)} \log C(\chi)}{2\pi}\right), \tag{16}$$

where $*$ denotes the sum is over $j_k = \pm 1, \pm 2, \ldots$ and $j_{k_1} \neq j_{k_2}$ if $k_1 \neq k_2$. The full $\mathrm{Sp}(\infty)$ conjecture for $\mathfrak{F}^{(2)}$ is equivalent to

$$\frac{1}{\mathfrak{F}^{(2)}(x)} \sum_{\chi \in \mathfrak{F}^{(2)}(x)} D(\chi, \Phi) \to \int_{\mathbb{R}^r} \Phi(u) W_{-}^{(r)}(u)\, du, \quad \text{as } x \to \infty \tag{17}$$

for any $r \geqslant 1$ and $\Phi \in \mathcal{S}(\mathbb{R}^r)$, where

$$W_{-}^{(r)}(x_1, \cdots, x_r) = \det(K_{-}(x_i, x_j))_{\substack{i=1,\ldots,r,\\ j=1,\cdots,r}}$$

$$K_{-}(x, y) := \frac{\sin \pi(x-y)}{\pi(x-y)} - \frac{\sin \pi(x+y)}{\pi(x+y)}$$

and

$$\mathfrak{F}^{(2)}(x) = \left\{\chi \in \mathfrak{F}^{(2)} :\ C(\chi) < x\right\}. \tag{18}$$

The first to consider the 1-level density for this family were Özlük and Snyder [OS93], who proved (17) for $r = 1$ and support of the Fourier transform $\widehat{\Phi}$ of Φ contained in $(-\frac{2}{3}, \frac{2}{3})$. Rubinstein [Rub01] established (17) for any $r \geqslant 1$ as long as the support $\widehat{\Phi} \subset \left\{\xi :\ \sum_{j=1}^{r} \left|\xi_j\right| < 1\right\}$. Later Gao [Gao] proved that the limit on the l.h.s. of (17) exists for support $\widehat{\Phi} \subset \left\{\xi :\ \sum_{j=1}^{r} \left|\xi_j\right| < 2\right\}$ but attempts to prove that his answer agrees with the r.h.s. in (17) failed until recently. What remained was an apparently difficult series of combinatorial identities. These are recently proven in [ERR13] thus establishing (17) in this bigger range. An interesting feature of their proof is that it uses the function field analogues to verify the identities and in this sense it is similar to the recent proof of the Fundamental Lemma ([Ngo] and the references therein). The point is that replacing $\mathbb{Q}$ by $\mathbb{F}_q(t)$ and computing the analogue r-level densities for the family of quadratic extensions of $\mathbb{F}_q(t)$ leads to the same answers and ranges as the case of $\mathbb{Q}$. But now averaging over q and keeping track of uniformity to switch orders leads to the setting in which [KS-b] prove the full $\mathrm{Sp}(\infty)$ conjecture and hence the combinatorial identities must hold in the case of $\mathbb{Q}$! An alternative combinatorial proof of the identities should also be possible along the line of [CS].

2.2 Number Fields and Artin L-Functions

The zero dimensional cases of the geometric families are already very rich. Let $K = \mathbb{Q}(W_1, \ldots, W_m)$ with $W_1, \ldots, W_m$ indeterminates and let $f \in K[x]$ be irreducible with splitting field L and Galois group B. According to Hilbert's irreducibility theorem the set of $w = (w_1, \ldots, w_m)$ in $\mathbb{Q}^m$ for which $f(x, w)$ is irreducible over $\mathbb{Q}$ and the Galois group of its splitting field $L_w/\mathbb{Q}$ is equal to B is the complement of a thin set ([Ser97, p. 123]). We call such w's f-generic and these are almost all of the points when counting the w's by height ([Ser97, § 13.1]). Let $\rho : B \to \mathrm{GL}_n(\mathbb{C})$ be an irreducible n-dimensional representation and let $H = \rho(B)$. To each generic w we have the corresponding irreducible Galois representation $\rho_w : \mathrm{Gal}(L_w/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$. This gives a family of n-dimensional Artin L-functions $L(s, \rho_w)$ and (conjecturally) automorphic cuspforms π_w on $G = \mathrm{GL}_n(\mathbb{A})$. That is, we have a parametrized family $\mathfrak{F} = (W, F)$ where $F(w) = \pi_w$ for w generic. By the Chebotarev density theorem for each such w, the Sato–Tate measure μ_{π_w} exists and is equal to μ_H. So we expect that $\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_H$ as well. This is indeed so if we order the π_w's by the height of w. For p large the asymptotics in (5) with $n = p^e$ holds. This follows by considering the w's mod p and then studying the variety $f(x, w_1, \ldots, w_m) = 0$ over $\mathbb{F}_p$ and using the theory of Artin's congruence zeta and L-functions for curves over finite fields in the case of the variable w_1, and [Weil, LW54] in general. This leads to the existence of the vertical limits $\mu_p(\mathfrak{F})$ and also that these converge to μ_H as $p \to \infty$. That is $\mu_{ST}(\mathfrak{F})$ exists and is equal to μ_H in this ordering. A more appropriate ordering of the w's is by the size of $D(w)$ where $D = D(W_1, \ldots, W_m)$ is the discriminant of f. The analogue of (5) can be carried out for this ordering as well, at least if w keeps away from directions in which $D(w)$ vanishes. The conductor of π_w is essentially the content of $D(w)$ and (5) can be carried out if the degree of D is small compared to the number of variables W. In all cases we find that $\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_H$. Once we have μ_H the key indicators $i_2(\mathfrak{F})$ and $i_3(\mathfrak{F})$ (here $i_1(\mathfrak{F}) = 1$) are then determined by the corresponding Schur indicators of H. Conjecture 2 can be established for $\mathfrak{F}$ for test functions of limited support (as discussed in Sect. 2.1) if $D(w)$ is of low degree.

Some very interesting *parametrized* families arise in connection with Dedekind zeta functions of number fields of fixed degree k. For $k = 2$ this is the family $\mathfrak{F}^{(2)}$ in Sect. 2.1. For $k = 3$ consider the parameters W_1, W_2, W_3, W_4 and the corresponding binary cubic forms (it is convenient to work projectively here) $f(W) = W_1x^3 + W_2x^2y + W_3xy^2 + W_4y^3$. The Galois group of f over $\mathbb{Q}(W)$ is S_3. Let $V(\mathbb{Q})$ be the $\mathbb{Q}$-vector space of such forms with $w = (w_1, w_2, w_3, w_4) \in \mathbb{Q}^4$. Let $V_{\mathrm{gen}}(\mathbb{Q})$ denote the points $w \in V(\mathbb{Q})$ for which the splitting field L_w of f_w is an S_3 extension of $\mathbb{Q}$. This together with a fixed irreducible representation ρ of S_3 yields a parametric family $\mathfrak{F}$ as above. The group $\mathrm{GL}_2(\mathbb{Q})$ acts on $V(\mathbb{Q})$ by linear change of variables and it preserves the fields L_w. The quotient $\mathrm{GL}_2(\mathbb{Q})\backslash V_{\mathrm{gen}}(\mathbb{Q})$ parameterizes exactly the S_3 splitting fields of degree 3 polynomials over $\mathbb{Q}$ (see [WY92]). In order to count these when ordered by conductor it is best to work over $\mathbb{Z}$ rather than $\mathbb{Q}$ as was done in [DH71] who parametrized and counted the cubic extensions of $\mathbb{Q}$ when ordered by discriminant. With $\mathrm{GL}_2(\mathbb{Z})$ acting on $V(\mathbb{Z})$ and $V(\mathbb{R})$ one determines a fundamental domain Ω and then orders points in $\Omega(\mathbb{Z})$ by the discriminant $D(w_1, w_2, w_3, w_4)$ which has degree 4. Furthermore one can sieve to fundamental discriminants and to points in $\Omega_{\mathrm{gen}}(\mathbb{Z})$. The most delicate point technically is dealing with w's in $\Omega(\mathbb{Z})$ with $D(w) \leqslant X$ and w near the directions where $D(w) = 0$. To each f in this parametrized reduced set correspond three conjugate cubic fields K_f', K_f'', K_f''' gotten by adjoining to $\mathbb{Q}$ one of the roots of f and $\mathrm{disc}(K_f^{(j)}) = D(f)$. In this way one obtains a parametrization of the cubic extensions of $\mathbb{Q}$ with Galois group S_3. Now $\zeta_{K_f^{(j)}}(s)/\zeta(s) = L(s, \rho_f)$ where ρ_f is the corresponding 2-dimensional irreducible representation of S_3. Thus this family π_{ρ_f} of GL_2-cuspforms (which are known to exist in this case since ρ_f is dihedral) is the parametrized family $\mathfrak{F}_3$ of Dedekind zeta functions of cubic extensions. We have $\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_H$ where H is the dihedral group D_3 in $\mathrm{GL}_2(\mathbb{C})$. It is orthogonal and hence $i_j(\mathfrak{F}) = 1$ for $j = 1, 2, 3$. In particular $\mathfrak{F}_3$ has an $\mathrm{Sp}(\infty)$ symmetry. This example is due to Yang [Yang].

For $k = 4, 5$ the parametrization over $\mathbb{Q}$ of degree k extensions with S_k Galois groups in terms of $G(\mathbb{Q})$ orbits of points in certain G-prehomogeneous vector spaces V is given in [WY92]. The theory over $\mathbb{Z}$ as needed to determine the density of such quartic and quintic fields is due to Bhargava ([Bha05, Bha10]). In all of these cases (including $k \geqslant 6$ if they could be suitably parametrized) $\mu_{\mathrm{ST}}(\mathfrak{F}_k) = \mu_{H_k}$, where H_k is the $k-1$ dimensional representation of S_k realized as the symmetries of the $k-1$ simplex. Since this representation is orthogonal we have $i_j(\mathfrak{F}_k) = 1$, for $j = 1, 2, 3$ and all of these parametrized families have an $\mathrm{Sp}(\infty)$ symmetry. A detailed treatment of families of Artin representations is the subject of [SST15].

2.3 Families of Elliptic Curves

We next consider geometric families $E \to \mathcal{W}$ of curves of genus one. The 1-parameter families are geometrically the same as elliptic surfaces fibered over the affine line. The singular fibers are classified by Kodaira and Néron and can be determined with Tate's algorithm. A 1-parameter family is given by polynomials in $\mathbb{Z}[w]$ which are the coefficients of the equation of a plane algebraic curve. A well-studied example is that of quadratic twists of a given elliptic curve which can be written in Weierstrass form as $wy^2 = x^3 + ax + b$. It can be viewed as a twist of a fixed elliptic curve with the quadratic family from Sect. 2.1 (for quadratic twists of any fixed automorphic form see Sect. 2.8 below). There is a natural 2-parameter family $\mathfrak{F}^{(\mathrm{ell})}$ given by $y^2 = x^3 + w_1 x + w_2$, where every elliptic curve over $\mathbb{Q}$ appears as a fiber with $a, b \in \mathbb{Z}$. The discriminant function is $D(w) = 4w_1^3 + 27w_2^2$. By modularity we obtain in each situation a parametric family $\mathfrak{F}$ of automorphic cusp forms on $\mathrm{PGL}(2)$.

Conjecture 1 can be verified for each of these families $\mathfrak{F}$ of elliptic curves and the Sato–Tate measure $\mu_{ST}(\mathfrak{F})$ exists with indicators $i_1(\mathfrak{F}) = i_2(\mathfrak{F}) = 1$ and $i_3(\mathfrak{F}) = -1$. Hence these families are homogeneous symplectic and correspondingly have symmetry type $O(\infty)$. For $\mathfrak{F}^{(\mathrm{ell})}$ this follows from a theorem of Birch [Birch] while in general see Sect. 2.11 below.

There is a caveat that we order the elliptic curves by height rather than conductor. Ordering by height for $\mathfrak{F}^{(\mathrm{ell})}$ means that we restrict to a box,

$$\max(4|w_1|^3, 27|w_2|^2) < x, \quad \text{with } x \to \infty.$$

It is desirable to be able to order by conductor $C(w) < x$ with $x \to \infty$ which yields interesting questions related to the square-free sieve for the discriminant polynomial $D(w)$. For $\mathfrak{F}^{(\mathrm{ell})}$ it follows from [FNT92, DK00] that the number of non-isogeneous semistable elliptic curves of conductor $C(w) < x$ is at least $x^{\frac{5}{6}}$ and at most $x^{1+\varepsilon}$. The average conductor is also important and it leads one to consider the ratio $\frac{\log C(w)}{\log |D(w)|}$ which is less than 1 and according to a conjecture of Szpiro should be greater than $\frac{1}{6} - \varepsilon$ with a finite number of exceptions. For $\mathfrak{F}^{(\mathrm{ell})}$ the ratio can be shown to be one on average using the square-free sieve which is known for polynomials in 2-variables of degree $\leqslant 6$ by Greaves [Greaves] (for 1-parameter families it is known for degree $\leqslant 3$ by [Hoo76]).

The next interesting invariant is the rank $r(\mathfrak{F})$ defined in (11). For $\mathfrak{F}^{(\mathrm{ell})}$ it follows from [Birch] that $t_{\mathfrak{F}^{(\mathrm{ell})}}(p) \ll p^{-1}$ and thus $r(\mathfrak{F}^{(\mathrm{ell})}) = 0$. For a 1-parameter family it is shown in Miller [Miller] using the Tate conjecture proven in Rosen-Silverman [RS98] that $r(\mathfrak{F})$ coincides with the rank of the elliptic surface over $\mathbb{Q}(w)$. There are examples of 1-parameter families where $r(\mathfrak{F})$ is greater than 18 and indeed such families have been used via specialization to produce rational elliptic curves of high rank [Elk07].

Mazur showed that there are finitely many possibilities for the torsion subgroup of elliptic curves over $\mathbb{Q}$. Harron-Snowden [HS13] recently established various bounds towards counting elliptic curves with prescribed torsion subgroup. In the process they actually show that for each prescribed torsion subgroup, elliptic curves are parametrized by a corresponding moduli space which is close to being an open subscheme of the affine line $\mathbb{A}^1$. Thus these are parametric families according to our definition (e.g., see, [HS13, § 3] where each family is explicitly given by polynomial equations with one free parameter).

The root number is the subtlest of the invariants. In the family $(7 + 7w^4)y^2 = x^3 - x$ found by Cassels–Schinzel [CS82], the root number $\epsilon(\frac{1}{2}, E_w) = -1$ for all $w \in \mathbb{Z}$,[4] whereas the rank $r(\mathfrak{F}) = 0$. Another example [Wash87] is the 1-parameter family $y^2 = x^3 + wx^2 - (w + 3)x + 1$ which has root number $\epsilon(\frac{1}{2}, E_w) = -1$ for all $w \in \mathbb{Z}$ and for which $r(\mathfrak{F}) = 1$. Thus the rank $r(\mathfrak{F})$ and the root numbers of member of $\mathfrak{F}$ can behave independently from one another and this explains why in Sect. 1 we treat them as distinct invariants.

The average root number is governed by the polynomial $M \in \mathbb{Z}[w_1, \dots, w_m]$ whose zero set is the locus of the fibers E_w with nodal (multiplicative) singularity.

[4]In fact $\epsilon(\frac{1}{2}, E_w) = -1$ also if we let $w \in \mathbb{Q}$ which should be viewed a 2-parameter family by writing $w = \frac{w_1}{w_2}$ and ordering by height $\max(|w_1|, |w_2|) < x$.

Note that M is a polynomial factor of the discriminant D. It is shown by Helfgott [Helf04] how the average root number in these cases is reduced to sums of the type (12) and thus if M is non-constant, that is if the family has at least one nodal geometric fiber, then the average root number should be zero. In the two examples from the preceding paragraph M is constant and indeed one can find in [Rizzo, Helf04] further examples of families of elliptic curves with M constant, where the average root number can assume any value in a dense subset of $[-1, 1]$. The sum (12) can be estimated unconditionally for polynomials of low degree, for example [Helf05]

$$\frac{1}{x^2} \sum_{|w_1|,|w_2|<x} \mu(w_1^3 + 2w_2^3) = o(1), \quad \text{as } x \to \infty. \tag{19}$$

An example where the root number is shown to average to zero unconditionally is the 2-parameter family $y^2 = x(x + w_1)(x + w_2)$ which contains every elliptic curve over $\mathbb{Q}$ with full rational 2-torsion $(\mathbb{Z}/2\mathbb{Z})^2$ as a fiber over $(w_1, w_2) \in \mathbb{Z}^2$. The case of $\mathfrak{F}^{(\mathrm{ell})}$ is more difficult. The method of proof of (19) is closely related to the work of Friedlander–Iwaniec and Heath-Brown on primes represented by polynomials in 2-variables.

The upshot is that Conjecture 2 is verified for families of quadratic twists in [Rub01], for $\mathfrak{F}^{(\mathrm{ell})}$ in [Young, BZ08] and under the above assumptions for 1-parameter families in [Miller]. This yields upper-bounds for the average analytic[5] rank as a corollary, see, for example, the articles in the proceedings [LMS07].

2.4 Dwork Families

In this section we investigate a certain parametric family of Dwork hypersurfaces, which were prominent examples in Dwork's detailed study of hypersurfaces in the 1960s. (See introduction of [Katz09] for a commentary on the literature.) Let $U = \operatorname{Spec} \mathbb{Z}[\frac{1}{n+1}, w]$, a subscheme of the affine line over $\mathbb{Z}[\frac{1}{n+1}]$. Consider the subscheme X of $\mathbb{P}^n_U$ cut out by the equation

$$\sum_{i=0}^{n} x_i^{n+1} = (n+1)w \prod_{i=0}^{n} x_i,$$

where $(x_0 : \cdots : x_n)$ and w are the coordinates for $\mathbb{P}^n$ and U, respectively. The family $X \to U$ is a family of elliptic curves for $n = 2$ and that of K3 surfaces for $n = 3$. In general the fibers of $X \to U$ have dimension $n - 1$, so the cohomology in degree

[5]The average rank of Selmer groups, which yields upper-bounds for the average Mordell-Weil rank, can be bounded by other methods, see [Hea94, FIMR] for the 1-parameter families of quadratic twists and [BS13, BS14] for $\mathfrak{F}^{(\mathrm{ell})}$.

$n-1$ is the most interesting. We could work with the whole of H^{n-1} cohomology but it is convenient to deal with a piece of cohomology by exploiting a group action on X. Let μ_{n+1} be the set of $(n+1)$-st roots of unity. (One may view μ_{n+1} as a group scheme over U.) Let H be the quotient group $(\mu_{n+1})^{n+1}/\Delta(\mu_{n+1})$, where Δ is the diagonal embedding. Then H acts on X by letting $(\alpha_0 : \cdots : \alpha_n)$ act by $(x_0 : \cdots : x_n) \mapsto (\alpha_0 x_0 : \cdots : \alpha_n x_n)$ on X. Let H_0 denote the subgroup of H which is a quotient of $\{(\alpha_0 : \cdots : \alpha_n) : \prod_{i=0}^n \alpha_i = 1\}$ by $\Delta(\mu_{n+1})$.

Consider the setup and notation for geometric families in Sect. 1. Take C to be the set of $w \in \mathbb{Z}$ such that $w \nmid (n+1)$, viewed as a set of closed points of U. Denote by X_w the fiber of X over $w \in U$. Use the discriminant function $D(w) = w^{n+1} - 1$ on C. Define the map $F : C \to \mathsf{A}(GL_n)$ such that $F(w)$ is the $|\det|^{\frac{n-1}{2}}$-twist of the automorphic representation corresponding to the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-representation

$$H^{n-1}_{\mathrm{et}}(X_w \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \mathbb{Q}_l)^{H_0} \tag{20}$$

via Conjecture 4 (or Conjecture 5). Note that $F(w) \in \mathsf{A}(GL_n)$ since (20) has dimension n over $\mathbb{Q}_l$ as can be shown by computing its dimension for $w = 0$ ([LNM82, Lemma 1.1], cf. [HSBT, Lemma 1.1]). Since X_w has good reduction modulo p whenever $p \nmid D(w)$, cf. [Katz09, § 3], the Galois representation (20) is unramified at such p, hence $F(w)$ should be unramified outside the prime divisors of $D(w)$.

Suppose that n is *even*.

The monodromy of the Dwork family $\mathfrak{F}$ is shown by Dwork to be the full symplectic group (if one is only interested in the symplectic pairing it can also be constructed by Poincaré duality, cf. [HSBT, Lemma 1.10, Corollary 1.11]). The two main conjectures from Sect. 1 yield the following: first, $\mu_{\mathrm{ST}}(\mathfrak{F})$ arises from the push-forward of a Haar measure on a maximal compact subgroup of $\mathrm{Sp}(n,\mathbb{C})$ in $\mathrm{GL}(n,\mathbb{C})$. This is proved as explained in Sect. 2.11 below using the Deligne–Katz equidistribution theorem. In other words the family has a Sato–Tate group $H(\mathfrak{F}) = \mathrm{Sp}(n,\mathbb{C})$. Second, Conjecture 2 says that the Symmetry Type of $\mathfrak{F}$ should be a superposition of $\mathrm{SO}_{\mathrm{even}}(\infty)$ and $\mathrm{SO}_{\mathrm{odd}}(\infty)$. The superposition depends on the distribution of $\varepsilon = 1$ and $\varepsilon = -1$ which we expect will be 50 %.

Finally when n is *odd*, (20) is even dimensional and equipped with a perfect symmetric pairing and the exact monodromy is also computed by Dwork. Thus in this case $\mu_{\mathrm{ST}}(\mathfrak{F})$ arises from an even orthogonal group and Conjecture 2 says that the Symmetry Type of $\mathfrak{F}$ should be $\mathrm{Sp}(\infty)$. It would be desirable to test all these low-lying zeros predictions for this family numerically.

2.5 Harmonic Families and Plancherel Equidistribution

Consider a spectral set $\mathfrak{H} \subset \mathsf{A}(H)$ of automorphic representations of a connected reductive group H over $\mathbb{Q}$ and an L-map $r : {}^L H \to {}^L\mathrm{GL}_n$. These data give rise to a harmonic family $\mathfrak{F}$. We discuss the Sato–Tate equidistribution for $\mathfrak{F}$ as formulated in Conjecture 1. In fact we need not assume the functoriality conjecture for r to make sense of the conjecture. Namely for each $\sigma \in \mathfrak{H}$ unramified outside of the finite set of places S, we can attach [Borel] the partial L-function $L^S(s, \sigma, r)$, which should be the partial L-function for $r_*\sigma$ if we assumed that $r_*\pi$ was an automorphic representation of GL_n. The prime p_0 is chosen large enough so that $p \geqslant p_0 \Rightarrow p \notin S$ and thus the unramified representation $\pi_p = r_*\sigma_p$ is known.

The asymptotic count of (4) is a Weyl's law or limit multiplicity problem. This has a long history with a vast literature. For limit multiplicities for towers of subgroups it starts with the classical article of DeGeorge–Wallach [DeGW78]. In the case that $\sigma \in \mathfrak{H}$ have discrete series σ_∞ at infinity the asymptotic count is well understood and it is natural to first focus on this case for studying harmonic families. See the end of this subsection for a discussion of the Maass forms case.

The next step is the quantitative equidistribution (5) which is much more difficult to obtain. The PGL(2) case is treated in [ILS01], see Sect. 2.7 below. A generalization to higher rank groups was recently achieved by the second and third-named authors [ST16].

To fix notation, the spectral set $\mathfrak{H}$ will be the set of automorphic representations σ of $H(\mathbb{A})$ which are cohomological at infinity with regular weight. (This means that the infinite component of σ has nonzero Lie algebra cohomology against an irreducible algebraic representation with regular highest weight.) Such σ is always cuspidal by a theorem of Wallach. If we consider the weight aspect it will be convenient to fix a level at finite places. Also the weights will be restricted to a cone inside the positive Weyl chamber. (This condition is parallel to the cone condition for geometric families and is important for similar reasons such as the uniform control of the analytic conductor.) If we consider the level aspect, then we fix a regular weight at infinity and consider a sequence of principal congruence subgroups of level $N \to \infty$.

The main theorem of [ST16] is a quantitative Plancherel equidistribution theorem for the local factors σ_p of representations $\sigma \in \mathfrak{H}$. Fix a test function φ which is a Weyl invariant polynomial on the dual maximal torus of H. For each prime p large enough one can evaluate φ against the unramified representations σ_p of $H(\mathbb{Q}_p)$ and we have

$$\sum_{\sigma \in \mathfrak{H}(x)} \varphi(\sigma_p) = |\mathfrak{H}(x)| \int \varphi(\sigma) \mu_p^{\mathrm{pl}}(d\sigma) + O(|\mathfrak{H}(x)|^\delta p^A) \tag{21}$$

where μ_p^{pl} is the unramified Plancherel measure on $\widehat{H(\mathbb{Q}_p)}$ and $\delta < 1$. The main term comes from the contribution of the identity on the geometric side of Arthur's cohomological trace formula [Art89]. The remainder term comes from bounding the other orbital integrals. The multiplicative constant in $O(\cdot)$ is uniform in p and x. This uniformity is a major difficulty in the proof because the number of conjugacy classes $\mathcal{O}$ to be considered on the geometric side is unbounded. In particular we have a weak control on the regularity of $\mathcal{O}$, it can, for example, ramify at several arbitrary large primes. We refer to [ST16, § 1.7] for a summary of the harmonic analysis techniques that we use to resolve this difficulty.

We deduce from (21) that each $\mu_p(\mathfrak{F})$ comes from the restriction of the Plancherel measure on $\widehat{H(\mathbb{Q}_p)}$. Precisely $\mu_p(\mathfrak{F})$ is the pushforward of μ_p^{pl} under the functorial lift attached to $r : {}^L H \to \mathrm{GL}(n,\mathbb{C})$. This is the assertion (i) of Conjecture 1. The main term $t_{\mathfrak{F}}(n)$ in the asymptotic (5) is expressed in terms of these p-adic densities. We also get assertion (ii) and the global measure $\mu(\mathfrak{F})$ by inserting a more general test function φ that is supported at finitely many places.

Maass forms are automorphic forms invariant under a maximal compact subgroup at infinity. They correspond to automorphic representations whose archimedean factors are spherical which is a condition that fits well in our formation of harmonic families. We expect the results to be similar to the case discussed above. The classical case of Maass forms on GL(2) can be treated using the Selberg trace formula. In higher rank the asymptotic Weyl's law is established in general by Lindenstrauss–Venkatesh [LV07]. Weyl's law with remainder term and the quantitative equidistribution (21) are more difficult despite the harmonic analysis on the spherical unitary dual being well understood [Helgason, DKV83].[6] These difficulties revolve around the presence of Eisenstein series: notably there is not yet a satisfactory description of the residual spectrum for general groups. The absolute convergence of the Arthur trace formula recently established by Finis–Lapid–Müller [FLM11] is an important step forward. J. Matz and the third-named author [MT] have recently established the case of Maass forms on GL(n).

2.6 Invariants of Harmonic Families

We form the Sato–Tate measure $\mu_{\mathrm{ST}}(\mathfrak{F}) = \lim_{p\to\infty} \mu_p(\mathfrak{F})_{|T}$ in assertion (iii) of Conjecture 1. Using the formula of Macdonald for the unramified Plancherel measure one can show this limit exists.[7] The measure $\mu_{\mathrm{ST}}(\mathfrak{F})$ coincides with the Sato–Tate measure attached to the image of ${}^L H$ viewed as a subgroup of $\mathrm{GL}(n,\mathbb{C})$. This can be taken as the Sato–Tate group $H(\mathfrak{F})$ of the family, thus for harmonic families the existence of such a group is proven.

Next we examine the three indicators $i_1(\mathfrak{F})$, $i_2(\mathfrak{F})$, and $i_3(\mathfrak{F})$ in (7). From now on we make the assumption that the representation $r : {}^L H \to \mathrm{GL}(n,\mathbb{C})$ is irreducible which can be seen to be equivalent to $i_1(\mathfrak{F}) = 1$. Thus the family $\mathfrak{F}$ is essentially cuspidal. This implies under the GRH that the functorial lift $r_*\sigma$ is cuspidal for most $\sigma \in \mathfrak{H}$ which needs to be established by a separate unconditional argument. The strategy is to the relate the non-cuspidality of $r_*\sigma$ to the vanishing of certain periods of σ (which is a well-studied and difficult problem, see the works of Jacquet,

[6]The difficulty is with the contribution of the continuous spectrum and in fact allowing noncongruence groups Weyl's law may fail [PS85].

[7]This holds literally if H is a split group. For a general H the Plancherel measure at a prime p depends on the splitting behavior (it is "Frobenian"). The *average* of $\mu_p(\mathfrak{F})_{|T}$ over the primes $p < x$ as in (6) converges and assertion (iii) follows from Chebotarev equidistribution theorem.

Jiang, Soudry, and many others), that is that σ is distinguished and then to show that this doesn't happen generically for almost all members σ of $\mathfrak{F}$.

The indicator $i_2(\mathfrak{F})$ is either 1 or 0, depending on whether r is self-dual or not. The indicator $i_3(\mathfrak{F})$ is denoted $s(r)$ in [ST16]. It is the Frobenius–Schur indicator of r which is either -1, 1 or 0, depending on whether r is symplectic, orthogonal, or not self-dual, respectively. Thus the family $\mathfrak{F}$ is essentially homogeneous if r is irreducible and the homogeneity type is determined.

The rank $r(\mathfrak{F})$ is zero for harmonic families. This follows from the defining Eq. (11) and the Macdonald formula for the Plancherel measure which implies in every case the estimate $t_{\mathfrak{F}}(p) = O(p^{-1})$, see [ST16, § 2]. This vanishing of the rank reflects the fact that the central L-value (or the L-derivative if the root number is -1) is expected to vanish only for arithmetical reason which should happen only for a few exceptional members of the family $\mathfrak{F}$.

The root number is the most subtle of the invariants attached to the family $\mathfrak{F}$. It is relevant for essentially symplectic families and corresponds to a decomposition

$$\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_{\mathrm{Sp},+}(\mathfrak{F}) + \mu_{\mathrm{Sp},-}(\mathfrak{F}).$$

For families in the level aspect the root number is related to the Möbius function. See [ILS01] and the discussion below for the case of PGL(2). In the weight aspect the root number could be dealt with along the lines of [ST16] although we have omitted the details there.

As we have noted repeatedly Conjecture 2 lies deeper. Its formulation assumes the analytic continuation of the completed L-functions $\Lambda(s, \sigma, r)$ inside the critical strip in order to define the zeros. This is known in many cases notably via Rankin–Selberg integrals and the Langlands–Shahidi method. The functoriality conjecture of Langlands asserts that the L-functions should be attached to an isobaric representation $r_*\sigma \in \mathsf{A}(GL_n)$.

In this regard let us observe that under the Ramanujan conjecture for GL_n (resp. with a bound $\theta < \frac{1}{2}$ towards Ramanujan), each of the local factors $L_v(s, \sigma, r)$ has no pole for $\Re\mathrm{e}(s) > 0$ (resp. $\Re\mathrm{e}(s) > \theta$). Hence any zero ρ with $\Re\mathrm{e}(\rho) > 0$ (resp. $\Re\mathrm{e}(\rho) > \theta$) of the partial L-function $L^S(\rho, \sigma, r) = 0$ cannot be cancelled by a potential pole of a local factor $L_v(s, \sigma, r)$ at $s = \rho$. The set of non-trivial zeros of $L^S(s, \sigma, r)$ (i.e., within the critical strip) will coincide with the set of zeros of $\Lambda(s, \sigma, r)$. Thus Conjecture 2 only depends on the analytic continuation of the *partial L-functions*. The formulation is robust because it is independent of the ramified factors $L_v(s, \sigma, r)$ (the analysis of which is the most delicate aspect in all known constructions of L-functions and the expected properties aren't fully established in many cases).

Once the above invariants $\mu_{\mathrm{ST}}(\mathfrak{F})$, $i_2(\mathfrak{F})$, $i_3(\mathfrak{F})$, $r(\mathfrak{F})$ and eventually $\mu_{\mathrm{Sp},\pm}(\mathfrak{F})$ are found one can verify Conjecture 2 for a test function with restricted support. The size of the support depends directly on the quality of the estimate (21). The details are found in [ST16, § 12] while the Criterion 1.2 in [ST16] is the insight which has motivated our present formulation of Conjecture 2.

We note that there is ample flexibility in choosing the spectral set $\mathfrak{H} \subset \mathsf{A}(H)$. For example one can add harmonic analysis constraints at finitely many places. As soon as $\mathfrak{H}$ is "large enough," the invariants of the family are independent of the choice and thus the Symmetry Type remains the same. The analogue for geometric families is to add congruences constraints on the parameters which is also very natural.

2.7 Classical Modular Forms

As mentioned above the case of $H = \mathrm{PGL}(2)$ is treated in [ILS01]. One might wonder what an arbitrary parametrized spectral subset of $\mathsf{A}(H)$ should look like since our definition allows flexibility in choosing the local harmonic constraints.[8] The problematic case of forms of weight $k = 1$ is discussed in Sect. 3. In this subsection we focus on the results of [ILS01] which correspond to the spectral set of holomorphic cuspforms $S_k(N)$ of weight $k \geqslant 2$ and square-free level N where either $k, N \to \infty$ with a possible additional average in dyadic intervals.

Suppose for simplicity that r is the embedding $\mathrm{SL}(2,\mathbb{C}) \to \mathrm{GL}(2,\mathbb{C})$ and denote by $\mathfrak{S}$ the corresponding family of standard Hecke L-functions. The conductor is k^2N and thus $|\mathfrak{S}(x)|$ which is the number of forms $f \in \mathfrak{S}$ with $C(f) < x$ is asymptotic to x up to a multiplicative constant.

Conjecture 1 holds for $\mathfrak{S}$ as consequence of [ILS01] and the Plancherel equidistribution results [ST16] described in the previous subsections. The measure $\mu_{ST}(\mathfrak{S})$ is obtained from the conjugacy classes of $\mathrm{SU}(2)$ and hence coincides with the classical Sato–Tate measure. If we let T^1 be the one-dimensional torus of $\mathrm{SL}(2,\mathbb{C})$ and parametrize T^1/W by $\left(\begin{smallmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{smallmatrix}\right)$ with $0 \leqslant \theta \leqslant \pi$, then

$$\mu_{ST}(\mathfrak{S}) = \frac{2}{\pi} \sin^2\theta \, d\theta.$$

The indicators are given by $i_1(\mathfrak{S}) = i_2(\mathfrak{S}) = 1$ and $i_3(\mathfrak{S}) = -1$. (More generally the Frobenius–Schur indicator of the k-th symmetric power representation $\mathrm{SL}(2,\mathbb{C}) \to \mathrm{GL}(k+1,\mathbb{C})$ is equal to $(-1)^k$.) Thus the family $\mathfrak{S}$ is essentially symplectic and this is in accordance with the $SO(\infty)$ Symmetry Type.

To go further we decompose the family $\mathfrak{S} = \mathfrak{S}_+ \cup \mathfrak{S}_-$ according to the root number being $+1$ or -1, respectively. The proportion of each piece is 50 %. The root number is $\varepsilon(f) = i^k\mu(N)\lambda_f(N)N^{\frac{1}{2}}$, so this statement is equivalent to cancellations in sums of the type $\sum\limits_{f\in\mathfrak{S}(x)} \lambda_f(N)N^{\frac{1}{2}}$ which is an example of the Möbius type sums discussed in (12). This sum can be analyzed directly via the Petersson trace formula as in [ILS01] or alternatively using representation theory and the results in [ST16]. Above a prime $p \mid N$, the p-component of f is tamely ramified with trivial central character and thus is either the Steinberg representation or a twist of the Steinberg representation by the unramified quadratic character; each representation carries

[8]In the context where H is the unit group of a division algebra, P. Nelson has recently proposed [Nelson] conditions for certain test functions to isolate such "nice" spectral sets.

50 % of the mass of $\mu_p(\mathfrak{S})$ which comes from restriction of the Plancherel measure on $\widehat{\mathrm{PGL}_2(\mathbb{Q}_p)}$.

For $\Phi \in \mathcal{S}(\mathbb{R})$ and $f \in \mathfrak{S}$ we denote by $D(f, \Phi)$ the one-level distribution of the low-lying zeros of $\Lambda(s,f)$ (removing one zero at $s = \frac{1}{2}$ if $f \in \mathfrak{S}_-$). Then Conjecture 2 reads

$$\frac{1}{|\mathfrak{S}_\pm(x)|} \sum_{f \in \mathfrak{S}_\pm(x)} D(f, \Phi) \to \int_{-\infty}^{\infty} \Phi(u) W_\pm^{(1)}(u)\, du, \quad \text{as } x \to \infty. \tag{22}$$

In other words the Symmetry Type of $\mathfrak{S}_+$ (resp. $\mathfrak{S}_-$) is $\mathrm{SO}_{\mathrm{even}}(\infty)$ (resp. $\mathrm{SO}_{\mathrm{odd}}(\infty)$).

Unconditionally the asymptotic (22) holds if the support of $\widehat{\Phi}$ is restricted to $(-1, 1)$. Under the GRH for Dirichlet L-functions one can extend the support to $(-2, 2)$. This extension is significant because then the one-level density distinguishes between the $\mathrm{Sp}(\infty)$, $\mathrm{SO}_{\mathrm{even}}(\infty)$ and $\mathrm{SO}_{\mathrm{odd}}(\infty)$ Symmetry Types since the distributions $W_+^{(1)}$ and $W_1^{(1)}$ agree in $u \in (-1, 1)$ but split at $u = \pm 1$.

There are many interesting applications of GL(1) and GL(2) families, notably the non-vanishing of L-values, distribution of prime numbers, quantum chaos, subconvexity, equidistribution of arithmetic cycles, and more. Here we have shown how to generalize the Symmetry Type with restricted support to higher rank families. We view the low-lying zeros statistics as a first step towards these other arithmetic features and applications.

2.8 GL(1) Twists

We fix π a cuspidal automorphic representation of $\mathrm{GL}(n)$ over $\mathbb{Q}$. If χ is a Dirichlet character, we can consider the twist $\pi \otimes \chi$ which is again a cuspidal automorphic representation of $\mathrm{GL}(n)$. In Sect. 2.1 we have discussed GL(1) families, for example the family $\mathfrak{F}^{(2)}$ of quadratic characters. One can construct a parametric family

$$\mathfrak{F} = \left\{\pi \otimes \chi,\ \chi \in \mathfrak{F}^{(2)}\right\}.$$

As we have discussed in the remarks following the definition of families we allow one of the factor, to be a singleton $\{\pi\}$ when considering the Rankin–Selberg product of families.

The quantitative equidistribution (5) is easily verified as well as the first two assertions of Conjecture 1. The assertion (iii), however, is as difficult as the individual Sato–Tate conjecture for π itself. We identify the n-dimensional torus T with the diagonal of $\mathrm{GL}(n, \mathbb{C})$ and thus with the product of n copies of $\mathbb{C}^\times$. Assume the Sato–Tate conjecture holds for π with a certain limit measure $\mu_{ST}(\pi)$ on T and recall the Sato–Tate measure $\mu_{ST}(\mathfrak{F}^{(2)}) = \mu_B$ for $\mathfrak{F}^{(2)}$ where $B = \{1, -1\} \subset \mathbb{C}^\times$. We have a natural multiplication homomorphism $m : \mathbb{C}^\times \times T \to T$ given by pointwise multiplication of each coordinate. The assertion (iii) of Conjecture 1 holds and the Sato–Tate measure of the family $\mathfrak{F}$ is the direct image

$$\mu_{ST}(\mathfrak{F}) = m_*(\mu_B \times \mu_{ST}(\pi)). \tag{23}$$

Equivalently $\mu_{ST}(\mathfrak{F})$ is half the sum of $\mu_{ST}(\pi)$ and the image of $\mu_{ST}(\pi)$ under $t \mapsto -t$. Note that since the family $\mathfrak{F}$ is thin the average over the primes $p < x$ in (6) is critical (see also the footnote 7 on page 551 for another example).

Often $\mu_{ST}(\mathfrak{F}) = \mu_{ST}(\pi)$, for example in the case that π is a holomorphic modular form on GL(2) of weight at least two for which the individual Sato–Tate is known. On the other hand, the two measures may differ. The simplest example is when π is a cubic Dirichlet character on GL(1) in which case $\mu_{ST}(\pi)$ is the Haar measure on the group $\left\{1, e^{\frac{2i\pi}{3}}, e^{\frac{4i\pi}{3}}\right\}$ while $\mu_{ST}(\mathfrak{F})$ is the Haar measure on $\left\{1, e^{\frac{i\pi}{3}}, e^{\frac{2i\pi}{3}}, -1, e^{\frac{4i\pi}{3}}, e^{\frac{5i\pi}{3}}\right\}$.

In view of (23) and $\operatorname{tr}(-t) = -\operatorname{tr}(t)$ the indicators can be computed as

$$i_2(\mathfrak{F}) = \int_T \operatorname{tr}(t)^2 \, \mu_{ST}(\pi)(t)$$

$$i_3(\mathfrak{F}) = \int_T \operatorname{tr}(t^2) \, \mu_{ST}(\pi)(t)$$

and $i_1(\mathfrak{F}) = 1$ since π is cuspidal. Thus we expect that $\mathfrak{F}$ is essentially homogeneous and its homogeneous type is dictated by π. In fact we only need to know which of $L(s, \pi, \operatorname{sym}^2)$ or $L(s, \pi, \wedge^2)$ has a pole at $s = 1$, which is very little information about the Sato–Tate group H_π. So even if the Sato–Tate measure of π remains mysterious we can verify the universality Conjecture 2 for $\mathfrak{F}$ unconditionally, see [Rub01]. For example, if π is self-dual orthogonal, then $\mathfrak{F}$ is essentially orthogonal and the Symmetry Type is $\operatorname{Sp}(\infty)$.

We can consider other GL(1) twists as, for example, the family $\mathfrak{F}' := \{\pi \otimes \chi\}$ as χ ranges through all Dirichlet characters of conductor $q \leqslant Q$ with $Q \to \infty$. Then the same analysis applies where we should replace B by the full unit circle $S^1 \subset \mathbb{C}^\times$. Thus we expect the Sato–Tate measure

$$\mu_{ST}(\mathfrak{F}') = m_*(\mu_{S^1} \times \mu_{ST}(\pi)).$$

The indicators are easier to compute in this case since we have $i_1(\mathfrak{F}') = 1$ and $i_2(\mathfrak{F}') = i_3(\mathfrak{F}') = 0$. Thus the family $\mathfrak{F}'$ is non self-dual and the Symmetry Type is $U(\infty)$ independently of any property of π. One simply uses that π is cuspidal and thus $L(s, \pi \times \tilde{\pi})$ has a simple pole at $s = 1$ which controls all the restricted n-level densities universally. This is entirely analogous to the universality of high zeros found in [RS96]. This surprising universality and the behavior of the families $\mathfrak{F}$ and $\mathfrak{F}'$ fit nicely into our main conjectures.

We can analyze the previous example using the Sato–Tate group $H_\pi \subset \mathrm{GL}(n, \mathbb{C})$, assuming it exists. Then we would associate with the family $\mathfrak{F}'$ the group $H(\mathfrak{F}')$ generated by H_π and $\mathbb{C}^\times$. In the same way that $\mu_{ST}(\pi)$ corresponds to H_π, we have that $\mu_{ST}(\mathfrak{F}')$ corresponds to $H(\mathfrak{F}')$.

Conversely we don't know what $H(\mathfrak{F}')$ is unless we are willing to assume the existence of H_π. In fact this example shows that if the family $\mathfrak{F}$ is thin like this one, knowing $H(\mathfrak{F})$ is tantamount to knowing H_π and so one may as well face having to define H_π conjecturally, for every π, if we want $H(\mathfrak{F})$ in general.

One expects that H_π would be either a torus or semisimple. On the other hand, $H(\mathfrak{F}')$ obviously isn't and this immediately explains the vanishing of the indicators $i_2(\mathfrak{F}') = i_3(\mathfrak{F}') = 0$. In general a family whose Sato–Tate group has infinite center has to have $U(\infty)$ Symmetry Type.

2.9 Rankin–Selberg Products

In [DM06] Dueñez–Miller investigate an interesting example of a parametric family of L-functions obtained by a $\mathrm{GL}(2) \times \mathrm{GL}(3)$ Rankin–Selberg product. Let π be a fixed even unramified Hecke–Maass form on $\mathrm{PGL}(2)$. Consider the spectral set of holomorphic cusp forms $f \in S_k(1)$ with $k \to \infty$. We can form the family

$$\mathfrak{F} := \left\{\pi \times \mathrm{sym}^2(f),\ f \in S_k(1)\right\}$$

which consists of L-functions of degree 6. By the work of Kim and Shahidi functoriality is known in this case so $\mathfrak{F}$ is a family of automorphic representations on $\mathrm{GL}(6)$. By construction all these forms are self-dual symplectic and the root number $\varepsilon(\frac{1}{2}, \pi \times \mathrm{sym}^2(f))$ can be verified to be 1 for all f.

If we assume the Sato–Tate conjecture for π, then we can verify Conjecture 1 for $\mathfrak{F}$. The measure $\mu_{ST}(\mathfrak{F})$ on the 6-dimensional torus is associated with the subgroup $SU(2) \times PSU(2)$ of $U(6)$, where the embedding is given by $(\theta_1, \theta_2) \mapsto \theta_1 \otimes \mathrm{sym}^2 \theta_2$. Since

$$\mathrm{tr}(\theta_1 \otimes \mathrm{sym}^2 \theta_2) = \mathrm{tr}(\theta_1)\,\mathrm{tr}(\mathrm{sym}^2 \theta_2),$$

the indicators can be easily computed to be $i_1(\mathfrak{F}) = i_2(\mathfrak{F}) = 1$ and $i_3(\mathfrak{F}) = -1$. Thus the family is essentially symplectic as we expect. In fact as usual we don't need to assume the full Sato–Tate conjecture for π to compute these indicators, only the knowledge of the simple pole of $\Lambda(s, \pi \times \tilde{\pi})$ at $s = 1$ suffices.

In [DM06] the 1-level and 2-level densities for a small restricted support are obtained unconditionally. This determines the Symmetry Type as $SO_{\mathrm{even}}(\infty)$ in Conjecture 2. This family $\mathfrak{F}$ has the feature that each L-function has even functional equation without having to decompose a bigger family according to the root number, a feature which is present for any family with a $\mathrm{Sp}(\infty)$ Symmetry Type. Thus we can conclude following [DM06] that the Symmetry Type is not just a theory of signs of functional equations, which is also apparent in our Conjecture 2. More generally as studied in a subsequent paper [DM09] the Symmetry Type has a certain predicted behavior under Rankin–Selberg product of families. This can also be explained by Conjecture 2 since if $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are two essentially cuspidal homogeneous families we expect that $\mathfrak{F}_1 \times \mathfrak{F}_2$ be homogeneous and in view of the properties of the Frobenius–Schur indicator that $i_3(\mathfrak{F}_1 \times \mathfrak{F}_2) = i_3(\mathfrak{F}_1)i_3(\mathfrak{F}_2)$.

Another family that is constructed with Rankin–Selberg type integral consists of adjoint L-functions. For a family attached to the spectral set of Maass forms $\mathfrak{H}$ on $\mathrm{SL}(3,\mathbb{Z})$ this is studied recently by Goldfeld–Kontorovich [GK13] using their version of the Kuznetsov trace formula. They consider the harmonic family $\mathfrak{F} = (\mathfrak{H}, \mathrm{Ad}_*)$ where Ad_* corresponds to the adjoint representation. The main result of [GK13] is that the family has Symmetry type $\mathrm{Sp}(\infty)$ when the density sums (17) with $r = 1$ are weighted by special values at 1 of L-functions of members of the family. (These weights are not expected to affect the Symmetry type.) This is consistent with our Conjecture 2 since $\mathfrak{F}$ is a homogeneous family which is essentially orthogonal. Indeed if π is a cuspidal automorphic representation on $\mathrm{SL}(3,\mathbb{Z})$ then $L(s,\pi,\mathrm{Ad})$ is self-dual and orthogonal (it is always cuspidal because we are in full level, thus π is not a base change).

Actually this example generalizes nicely: let H be any split connected quasi-simple group over $\mathbb{Q}$. Form the adjoint representation which is an L-map from ${}^LH$ to GL_n where $n = \dim H$. Consider a generic spectral set $\mathfrak{H}$ as above and the family $(\mathfrak{H}, \mathrm{Ad}_*)$. The adjoint representation is irreducible and it preserves the Killing form on $\mathrm{Lie}({}^LH)$ which is bilinear symmetric and non-degenerate. Thus we expect almost all L-functions to be cuspidal and self-dual orthogonal, thus the family to be essentially orthogonal. Therefore according to Conjecture 2 we expect that any universal family of adjoint L-functions have Symmetry Type $\mathrm{Sp}(\infty)$. For $H = \mathrm{PGL}(2)$ the adjoint representation is the same as the symmetric square and this is a result in [ILS01].

The case of $H = \mathrm{PGSp}(4)$ is recently studied by Kowalski–Saha–Tsimerman [KST12]. Namely they consider the spectral set $S_k^*(\mathrm{Sp}(4,\mathbb{Z})) \subset \mathsf{A}(H)$ of Siegel cusp forms of weight $k \to \infty$. Let r be the degree four spin representation of ${}^LH = \mathrm{Spin}(5,\mathbb{C})$. We can form the family of L-functions

$$\mathfrak{F} := \left\{L(s,F,r),\ F \in S_k^*(\mathrm{Sp}(4,\mathbb{Z}))\right\}$$

which by functoriality for classical groups are known to correspond to automorphic representations of GL(4).

The main result of [KST12] is a (weighted) equidistribution result which is essentially related to Conjecture 1 for $\mathfrak{F}$. The measure $\mu_p(\mathfrak{F})$ is a (relative) Plancherel measure whose limit $\mu_{ST}(\mathfrak{F})$ exists as $p \to \infty$ and coincides with the Sato–Tate measure associated with the subgroup $r({}^LH) \subset \mathrm{GL}(4,\mathbb{C})$.

One finds that $i_1(\mathfrak{F}) = 1$, thus the family is essentially cuspidal. The members $F \in S_k^*(\mathrm{Sp}(4,\mathbb{Z}))$ such that $L(s,F,r)$ is not cuspidal are precisely the Saito–Kurokawa lifts from $\mathrm{SL}(2,\mathbb{Z})$. These form a (spectral) subset which is asymptotically negligible which confirms that almost all members of the family are cuspidal.

Next we have $i_2(\mathfrak{F}) = 1$ and $i_3(\mathfrak{F}) = -1$ and thus the family is essentially symplectic. In view of the isomorphism $\mathrm{Spin}(5, \mathbb{C}) \simeq \mathrm{Sp}(4, \mathbb{C})$, the representation r is self-dual symplectic which is consistent. The root number is $(-1)^k$ thus we expect according to Conjecture 2 an $\mathrm{SO}_{\mathrm{even}}(\infty)$ or $\mathrm{SO}_{\mathrm{odd}}(\infty)$ Symmetry Type, depending on the parity of the weight k.

The analysis of the low-lying zeros with a test function of restricted support is carried out in [KST12] but the results are altered by the presence of a weighting factor for each F. Since this weight is itself a *central* value of L-function by a formula conjectured by Böcherer and Furusawa–Martin, it carries much fluctuation which apparently yields a symmetry which is not consistent with our conjectures. If these weights are removed, we expect that this feature will disappear. Here this means that the weights which appear naturally from the application of the Petersson trace formula would need to be removed in order to interpret the symmetry type, see [Kow13] for further discussions.

2.10 Universal Families

For the universal family of all cuspidal automorphic forms on $\mathrm{GL}_n(\mathbb{A})$ we expect that the Sato–Tate Conjecture 1 still holds. The measure $\mu_p(\mathfrak{F})$ is closely related to the Plancherel measure. Precisely for each integer $k \geqslant 0$, let $\mu^{\mathrm{pl}}[p^k]$ be the restriction of the Plancherel measure to the subset of representations in $\widehat{\mathrm{GL}_n(\mathbb{Q}_p)}$ of conductor p^k. Then $\mu_p(\mathfrak{F})$ will be an explicit linear combination of the measures $\mu^{\mathrm{pl}}[p^k]$.

Example. This can be verified for $n = 1$, the universal family $\mathfrak{F}$ of all Dirichlet characters, see also [Kow13]. The total mass of $\mu^{\mathrm{pl}}[p^k]$ is $\varphi(p^k)$, the Euler function. A direct calculation shows that

$$\mu_p(\mathfrak{F}) = a \sum_{k=0}^{\infty} \frac{1}{p^{2k}} \mu^{\mathrm{pl}}[p^k]$$

where $a = \frac{p^3}{(p-1)(p+1)^2}$.

Note, however, that for the "family" of forms of level $n!$ or product of consecutive primes $2 \cdot 3 \cdot 5 \cdot 7 \ldots$, the Sato–Tate conjecture in the form (5) fails (as observed by Junehyuk Jung). The universality of the low-lying zeros in Conjecture 2 is still expected to hold here, but for deeper reasons. The case of families of Dirichlet characters can be verified directly, the case of GL(2) is done in [ILS01] and the general case is done in [ST16].

2.11 Deligne–Katz Equidistribution and Geometric Families

In this subsection we consider geometric families. Our goal is to explain how to approach Conjecture 1 using monodromy groups. There are many technical issues that we ignore and we confine ourselves to an outline.

We begin with a general geometric family as in the definition in Sect. 1. Thus $\mathcal{W}$ is an open dense subscheme of $\mathbb{A}^m_{\mathbb{Z}}$, and $f : X \to \mathcal{W}$ is smooth and proper with integral fibers. To concentrate on examples of geometric nature, we assume the fibers to be geometrically connected. For any $w \in W := \mathcal{W}(\mathbb{Z}) \cap C$ we denote the fiber by X_w. This gives rise to a parametric family $\mathfrak{F}$ of Hasse-Weil L-functions.

The local L-factor can be described using Grothendieck's l-adic monodromy theorem. (We need a result in p-adic Hodge theory when $p = l$ but it is harmless to assume $p > l$ for our purpose.) Let ρ_w be the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-representation acting on the space $H^d_{et}(X_w \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \mathbb{Q}_l)$. For any prime p we consider the Weil–Deligne representation

$$r_{w,p} := \iota \mathrm{WD}_v(\rho|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}),$$

see Appendix 1 for details. Also let $\pi_{w,p} := \mathrm{rec}^{-1}(\rho_{w,p}) \otimes |\det|^{d/2}$ viewed as an element of $\widehat{G(\mathbb{Q}_p)}$ where $G = \mathrm{GL}_n$. (As remarked in the previous section, the fact that $\pi_{w,p}$ is unitary is conditional on the weight-monodromy conjecture if X_w has bad reduction.) The local L-factor at p is given by

$$L(s, \pi_{w,p}) = L(s, r_{w,p}) = \det(1 - \mathrm{Frob}_p p^{-s} | V^{I_p} \cap \ker N)^{-1}$$

where V is the underlying space of $r_{w,p}$ and N is the corresponding nilpotent operator.

As a preliminary step we examine the ramification of the representations $\pi_{w,p}$. If $\pi_{w,p}$ is ramified, then p is a prime of bad reduction for X_w and also $D(w) \equiv 0 \pmod{p}$, where D is the discriminant function of the family. Conjecture 1 is rather precise because the assertions (i) and (ii) include the statistics of the ramified representations. The depth of the representations $\pi_{w,p} \in \widehat{G(\mathbb{Q}_p)}$ is bounded by a constant [ST14, § 3] independent of w, p because its field of rationality is $\mathbb{Q}$.

For each unramified $\pi_{w,p}$ we obtain an element $t_{w,p} \in T/W$. A crucial observation is that it depends only on w modulo p. Thus the measure $\mu_p(\mathfrak{F})|_T$ (and more generally $\mu_p(\mathfrak{F})$) is atomic, in fact supported on a finite subset of T/W. It is given explicitly by the following sum of Dirac measures:

$$\mu_p(\mathfrak{F})|_T = \frac{1}{|\mathcal{W}(\mathbb{F}_p)|} \sum_{\substack{w \in \mathcal{W}(\mathbb{F}_p),\\ D(w) \not\equiv 0}} \delta_{t_{w,p}}, \tag{24}$$

where the sum has been restricted to those w such that $\pi_{w,p}$ is unramified by demanding that $D(w) \not\equiv 0 \pmod{p}$. It implies by the Lang-Weil bound [LW54],

$$\mu_p(\mathfrak{F})(T) = 1 - O\left(\frac{1}{p}\right). \tag{25}$$

In view of (25) the ramified representations play no role in the assertions (iii) and (iv) of Conjecture 1 and hence also in the construction of $\mu_{\mathrm{ST}}(\mathfrak{F})$ which is our main interest. Thus from now on we shall focus on (24) and those representations $\pi_{w,p}$ which are unramified.

The analysis involves sets of integer points $w = (w_1, \dots, w_m)$ in sectors W in $\mathbb{Z}^m$ in regions defined by a homogeneous polynomial which approximates the conductor, for example a height condition that w lies in a large box (that is, each w_i lies in an interval). The sectors defining C are chosen to make these sets finite by avoiding the projective zero locus of the discriminant D. The assertion (24) is deduced from the convergence:

$$\frac{1}{|\mathfrak{F}(x)|} \sum_{\substack{w \in \mathcal{W}(\mathbb{Z}) \cap C, \\ |w_i|^{d_i} < x, \forall i}} \delta_{t_{w,p}} \rightharpoonup \frac{1}{|\mathcal{W}(\mathbb{F}_p)|} \sum_{\substack{w \in \mathcal{W}(\mathbb{F}_p), \\ D(w) \not\equiv 0}} \delta_{t_{w,p}},$$

which follows from the fact that $t_{w,p}$ depends only on w modulo p.

The above reasoning is the key arithmetic input. And indeed this argument occurs often in number theory such as in the circle method. This localization away from the zero locus of D makes the problem easier and in general it forces us to count the parametrized elements π_w in the family with some natural multiplicity.[9]

To establish assertions (iii) and (iv) of Conjecture 1 it remains to study the measures $\mu_p(\mathfrak{F})|_T$ and thus we are reduced to a problem over finite fields. The reduction is possible because we have chosen $\mathcal{W}$ to be affine in the definition of geometric families. In fact we see from the argument that we could relax this assumption somewhat, but not entirely see Sect. 3.1 below.

It is convenient to formulate the problem over finite fields by introducing the sheaf $\mathcal{G} := R^d f_* \overline{\mathbb{Q}_l}$, which is the "$H^d$ along the fibers X_w". It is a lisse ℓ-adic sheaf over $\mathcal{W}$ of rank n. The Grothendieck base change theorem implies that there is an action of the arithmetic fundamental group $\pi_1(\mathcal{W})$ on a finite dimensional $\overline{\mathbb{Q}_l}$-vector space which can be identified with the cohomology group of the fibers [KS-b]. Specifically there is a linear action by automorphism which yields the monodromy representation $\pi_1(\mathcal{W}) \to \mathrm{GL}(n, \overline{\mathbb{Q}_l})$, which is well defined up to conjugation. The geometric fundamental group $\pi_1^{\mathrm{geom}}(\mathcal{W})$ is a normal subgroup of $\pi_1(\mathcal{W})$, and we denote by G_{geom} the Zariski closure of its image. By a theorem of Deligne G_{geom} is semisimple. The Zariski closure of the image of the arithmetic fundamental group $\pi_1(\mathcal{W})$ is denoted G_{arith}.[10] Thus G_{geom} is a normal subgroup of G_{arith} and from now

[9]We note that we make analogous simplifying assumption in the case of harmonic families, see Sect. 2.5, where we have allowed some mild weights such as $\dim(\pi_v)^{U_v}$ which doesn't change the final answer but makes the problem easier to analyze with the trace formula.

[10]Here we are assuming as in [Katz13] geometric connectedness.

on we impose the hypothesis that $G_{\text{arith}} \subset \mathbb{G}_m G_{\text{geom}}$, see the recent article of Katz [Katz13] for details, cf. Hypothesis (H) there. This essentially amounts to a purity assumption on the sheaf $\mathcal{G}$, which gives a uniform control on the size of Frobenius eigenvalues.

For each prime p and $w \in \mathcal{W}(\mathbb{F}_p)$, the image of Frobenius under the monodromy representation lies in G_{arith}. Thanks to the hypothesis above, we can rescale it by a scalar and obtain an element $\text{Frob}_{w,p} \in G_{\text{geom}}$ well defined up to the choice of an l-adic unit and up to conjugation. Moreover by purity all the eigenvalues of $\iota\text{Frob}_{w,p}$ lie on the unit circle and therefore $\iota\text{Frob}_{w,p}$ may be viewed up to conjugation as an element of B_c, the maximal compact subgroup of G_{geom}, again we refer to [Katz13] for details.

We form the probability measure

$$\mu_p(\mathcal{G}) := \frac{1}{|\mathcal{W}(\mathbb{F}_p)|} \sum_{w \in \mathcal{W}(\mathbb{F}_p)} \delta_{\text{Frob}_{w,p}}$$

on $B_c^{\#}$. The key point of these constructions is that the pushforward of $\mu_p(\mathcal{G})$ under $B_c^{\#} \to T_c/W$ coincides up to $O\left(\frac{1}{p}\right)$ with the measure $\mu_p(\mathfrak{F})|_T$ defined in (24) above.

It remains to let the prime $p \to \infty$. The equidistribution of the measures $\mu_p(\mathcal{G})$, with respect to the Haar measure of B, is Katz's variant of Deligne's equidistribution theorem, see [Katz13] and [KS-b, § 9]. It is important here that it can be proven that the monodromy depends only on the topology of the family $X \to \mathcal{W}$. In other words the geometric fundamental group is independent of p for p large, see [Katz13, Theorem 2.1].

Specifically we apply Theorem 5.1 of [Katz13] (with all n_i equal to 1) to the sheaf $\mathcal{G}$, which is ι-pure by [KS-b, 9.1.15], to obtain that

$$\mu_p(\mathcal{G}) \rightharpoonup \mu_{ST}(\mathfrak{F}), \quad \text{as } p \to \infty, \tag{26}$$

where $\mu_{ST}(\mathfrak{F})$ is the pushforward of the Haar measure under $B_c^{\#} \to T_c/W$. Note that the base scheme S for us is of the form $\text{Spec}\mathbb{Z}[1/N]$ and therefore the Hypothesis (AFG) in [Katz13] involves removing finitely many primes p. This finishes the outline of the proof of the assertions (iii) and (iv) of Conjecture 1 for $\mathfrak{F}$.

For example, for the family of all elliptic curves which we have discussed in Sect. 2.3, the equidistribution theorem is an early result of Birch [Birch]. The example of 1-parameter families of hyperelliptic curves of genus g is treated in [KS-b], where we have $G_{\text{geom}} = \text{Sp}(2g)$ and $G_{\text{arith}} = \text{GSp}(2g)$. Another interesting example is the universal family of smooth projective hypersurfaces of given dimension and degree, which is also in [KS-b]. Finally the above equidistribution applies to the Dwork families discussed in Sect. 2.4.

Conjecture 2 can be established for $\mathfrak{F}$ for test functions of limited support and conditionally on the modularity conjecture for the X_w. Both for harmonic families (see §2.6) and for geometric families we have attached a group $H(\mathfrak{F})$ such that the associated Sato–Tate measure $\mu_{ST}(\mathfrak{F})$ is computed in terms of $H(\mathfrak{F})$. As we observed earlier the measure $\mu_{ST}(\mathfrak{F})$ need not determine the group $H(\mathfrak{F})$ uniquely, however there is a natural choice which comes from the method of proof of Conjecture 1, namely $H(\mathfrak{F}) := r({}^L H)$ for harmonic families and $H(\mathfrak{F}) := G_{\text{geom}}$ for geometric families.

Serre has recently put forward a Sato–Tate conjecture for schemes which is related to the above discussion. Let $X \to \mathcal{W}$ be a scheme of finite type. If $\mathcal{W}$ is a point, then this is the usual Sato–Tate conjecture for the Hasse–Weil L-functions attached to X. If $\mathcal{W}$ satisfies some suitable conditions, it is a direct consequence of (26) as explained in [Katz13] because it asks for the convergence for $x \to \infty$ of the average for $p^r < x$ of the measures $\mu_{p^r}(\mathcal{G})$. There are differences of this to our Sato–Tate conjecture for families: one being that the Sato–Tate conjecture for scheme is expected to be true for any base $\mathcal{W}$ (and is proven in [Katz13] under mild assumptions if $\mathcal{W}$ is not a point), whereas it is easy to construct counterexamples to our Conjectures 1 and 2 for families if the base $\mathcal{W}$ were arbitrary (see Sect. 3.1).

2.12 Prospects

Under certain assumptions we have verified for the above families the concepts introduced in Sect. 1. It is desirable to lift these assumptions as much as possible since this would strengthen our knowledge and make certain results unconditional. We summarize here the nature of these issues and give some plausible outlook of how some could be addressed in future work. We shall focus solely on the Sato–Tate equidistribution for families as formulated in Conjecture 1.

For general harmonic families, the Sato–Tate equidistribution for families implies working with general test functions, which raises important questions on the global harmonic analysis of the trace formula. One such question is formulated in [FLM15] in the context of limit multiplicities and concerns a uniform estimate on the winding number of normalizing scalars of intertwinning operators. Another challenge concerns the description of the residual spectrum which is known for $\mathrm{GL}(n)$ and used crucially in establishing quantitative error terms in the Weyl's law [LM09, MT]. These and related problems now seem within reach in the context of classical groups from the work of Arthur and others.

Local harmonic analysis and representation theory of p-adic groups and real Lie groups also play a major role in Conjecture 1. One would like to capture a portion of the spectrum that is as fine as possible. Over the reals this means discrete series versus stable packets and short spectral windows for Maass forms. For p-adic groups this means working with congruence subgroups beyond principal towers, see, e.g., [FL15], and possibly working with a single supercuspidal representation, a question discussed in [KST16] which will appear in this proceedings volume. Another property concerns uniform control on the matrix coefficients of intertwinning operators, which is studied in [MS04] over the reals and in [FLM12] over the p-adics. Finally the analytic conductor of representations, which is used in the present formulation of Conjecture 1, is difficult to define in complete generality. For this it is essential to clarify the relation between depth and conductor, see [Kala] for work in this direction, and it would be important to improve our understanding of the local Langlands correspondence in the tame case.

For geometric families it is a difficult problem in each specific example to identify the monodromy group. Also it is difficult to make the parametrization F one-to-one; this is related to the implementation of the square-free sieve, which a major step in the work of Bhargava on counting number fields with bounded discriminant. Analogously to the question of depth versus conductor mentioned above for automorphic representations, there is a question of the relation between height and conductor for Hasse-Weil L-functions.

3 Non-Examples

In this section we give some "families" of automorphic forms that do not fit into our prescription in Sect. 1. While some of these are natural and Conjectures 1 and 2 probably apply to them, they lack parametrizations and hence any known means of study and hence remain very speculative.

3.1 Limitations

We begin by pointing to limitations in forming families. The base space $\mathcal{W}$ of parameters in our definition of a geometric family is allowed to be $\mathbb{P}^m/\mathbb{Q}$, $\mathbb{A}^m/\mathbb{Z}$ or products of such. Unlike the algebro-geometric setting of families over finite fields, we cannot allow a general base $\mathcal{W}$ which is defined by equations over $\mathbb{Z}$ (or $\mathbb{Q}$). According to the solution of Hilbert's 10th problem [Mat93] one cannot say much about such sets $\mathcal{W}(\mathbb{Z})$, for example deciding if they are finite or not, and in general these sets may be unwieldy (see the example below). In particular the averages (5), or for that matter any other statistics associated with the family, need not exist. What would suffice for $\mathcal{W}(\mathbb{Z})$ in order for us to analyze the family to the extent that is described in Sects. 1 and 2 is that $\mathcal{W}$ be "strongly Hardy–Littlewood" in the sense of [BR95].

The same difficulty arises if we try to perform simple Boolean operations on our families. If $\mathfrak{F}_1 = (W_1, F_1)$ and $\mathfrak{F}_2 = (W_2, F_2)$ are two parametric families in $\mathsf{A}(G)$, then a natural parametric definition of their intersection is $\mathfrak{F}_{12} = (W_{12}, F_{12})$ where $W_{12} = \{(w_1, w_2) : F_1(w_1) = F_2(w_2)\} \subset W_1 \times W_2$ and $F_{12}((w_1, w_2)) = F_1(w_1)$ ($= F_2(w_2)$) for $(w_1, w_2) \in W_{12}$. Note that if F_1 and F_2 are embeddings (so that $F_1(w_1)$ an $F_2(w_2)$ are parametrized sets in $\mathsf{A}(G)$) then $\mathfrak{F}_{12}$ parametrizes $\mathfrak{F}_1(W_1) \cap \mathfrak{F}_2(W_2)$. The problem is that $W_{12} \subset W_1 \times W_2$ encodes a general diophantine set and again we are dealing with unwieldy sets for which the various statistical averages over the family need not exist.

A concrete example of the above where we allow various operations on a parametric family is the following: Let $R \in \mathbb{N}$ be a recursive set [Mat93]. There is a polynomial $P = P_R \in \mathbb{Z}[W_1, \ldots, W_{10}]$ such that $P(\mathbb{Z}^{10}) \cap \mathbb{N} = R$ (see [Mat93]). Consider the parametric family $\mathfrak{F}$ in $\mathsf{A}(\mathrm{GL}_1)$ given by

$$\mathfrak{F} : \ X^2 = p(W_1, \ldots, W_{10})$$

so that

$$F((w_1, \ldots, w_{10})) = \chi_{D(w_1, \ldots, w_{10})}$$

where $D(w_1, \ldots, w_{10})$ is the square-free part of $p(w_1, \ldots, w_{10})$ and χ the Dirichlet character corresponding to the quadratic field $\mathbb{Q}(\sqrt{p(w)})$. Then $\mathfrak{F} = (W, F)$ is a parametric family in our sense and the discussion in Sects. 1 and 2 applies to it. However if we consider the image $T = F(\mathbb{Z}^{10})$ in $\mathsf{A}(\mathrm{GL}_1)$ and impose the condition that the field corresponding $F(w)$ is real (that is we intersect T with $\mathbb{N}$) then we arrive at the subset R of $\mathbb{N}$, realized as a subsect of $\mathfrak{F}^{(2)}$. The set of recursive subsets of $\mathbb{N}$ is very general and certainly any statement such as (5) will not hold for such a general R (when ordered by height).

3.2 Fields of Rationality

In this section we introduce a construction of families via field of rationality. Let π be an automorphic representation of $\mathrm{GL}_n(\mathbb{A})$. The field of rationality $\mathbb{Q}(\pi)$ for π is by definition the fixed field in $\mathbb{C}$ under

$$\{\sigma \in \mathrm{Aut}(\mathbb{C}) : \pi^\sigma \simeq \pi\}$$

where $\pi^\sigma := \pi \otimes_{\mathbb{C},\sigma} \mathbb{C}$. A well-known conjecture states that $[\mathbb{Q}(\pi) : \mathbb{Q}] < \infty$ if and only if π is algebraic in the sense of Clozel [Clozel]. (These notions and the conjecture extend to arbitrary connected reductive groups, cf. [BG11].)

Let $\mathfrak{F} = (\mathfrak{H}, F)$ be a harmonic family as in Sect. 1. For a number field K (as a subfield of $\mathbb{C}$) define $\mathfrak{F}_{\subseteq K}$ to be the subset consisting of $\pi \in \mathfrak{F}$ such that $\mathbb{Q}(\pi) \subseteq K$. Similarly for an integer $A \geqslant 1$ define $\mathfrak{F}_{\leqslant A} := \{\pi \in \mathfrak{F} : [\mathbb{Q}(\pi) : \mathbb{Q}] \leqslant A\}$. Observe that each of $\mathfrak{F}_{\subseteq K}$ and $\mathfrak{F}_{\leqslant A}$ is supposed to contain only algebraic members by the conjecture just mentioned. If $\mathfrak{F}$ is ramified at only finitely many primes, then $\mathfrak{F}_{\subseteq K}$ and $\mathfrak{F}_{\leqslant A}$ are conjectured to be finite sets, cf. [ST14, Conj 5.10], and verified to be finite when G is a general linear group or a quasi-split classical group. (See Theorem 1.6 and Corollary 6.8 of [ST14].)

Example. In the setup for harmonic families take $H = G = \mathrm{GL}_2$. Let $\mathfrak{H}$ be the family of all cuspidal automorphic representations π of $\mathrm{GL}_2(\mathbb{A})$ such that π_∞ is the discrete series of lowest weight (so that π correspond to classical modular forms of weight 2). Suppose that F comes from the identity L-morphism r. Then $\mathfrak{F}_{\subseteq \mathbb{Q}} = \mathfrak{F}_{\leqslant 1}$ is nothing but the family of all normalized cuspforms of weight 2 whose Fourier coefficients are rational numbers.

The family $\mathfrak{F}_{\subseteq \mathbb{Q}}$ in the example is identified with the family of all elliptic curves over $\mathbb{Q}$, cf. Appendix 1 below. The family corresponds to the moduli stack of elliptic curves over $\mathbb{Q}$ or a moduli scheme if a suitable level structure is added. So this example almost fits in the framework of geometric families considered earlier, to which the two main conjectures apply. This leads us to the question as to when the families $\mathfrak{F}_{\subseteq K}$ and $\mathfrak{F}_{\leqslant A}$ can be realized as geometric families. Moreover we may ask.

Question 3.3. *Suppose that the family $\mathfrak{F}_{\subseteq K}$ (resp. $\mathfrak{F}_{\leqslant A}$) has infinite cardinality. Are Conjectures 1 and 2 true for the family $\mathfrak{F}_{\subseteq K}$ (resp. $\mathfrak{F}_{\leqslant A}$)?*

To shed light on the question, let us pursue the connection with geometric families further when the family $\mathfrak{F}_{\subseteq \mathbb{Q}}$ is constructed as in the above example except that the weight is a general integer $k \geqslant 2$, following [PR15]. (Also see [Kha10, § 7.2].) A conjecture of Paranjape and Ramakrishnan states that each $\pi \in \mathfrak{F}_{\subseteq \mathbb{Q}}$ should be associated with a two-dimensional $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-subrepresentation of $H^{k-1}(X_\pi \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \overline{\mathbb{Q}}_l)$ for some Calabi-Yau variety X_π over $\mathbb{Q}$ of dimension $k-1$ (such that the two-dimensional piece should be cut out by the part with Hodge numbers $(k-1,0)$ and $(0,k-1)$). If true, this suggests that $\mathfrak{F}_{\subseteq \mathbb{Q}}$ might be a family of 2-dimensional motives appearing in the family of H^{k-1}-cohomology arising from a family of $(k-1)$-dimensional Calabi-Yau varieties. When $k=2$ this reduces to the discussion of the family of elliptic curves over $\mathbb{Q}$ above. In case $k=3$, where all π are of CM type and X_π are K3 surfaces, see [ES13] for a recent result due to Elkies and Schütt. A partial result towards the general case is worked out in [PR15]. However it is known that there are only finitely many π which are of CM type, correspond to a weight 3 cuspform, and have $\mathbb{Q}$ as field of rationality, and similarly for all odd $k \geqslant 3$ under the GRH, cf. [ES13, § 3] for more details. So the assumption of the above question is not superfluous. In fact the authors do not know a criterion for $\mathfrak{F}_{\subseteq K}$ to be infinite.

More generally these conjectures about other rationality for algebraic representations all point to geometric families again. So philosophically perhaps many families obtained by specifying the field of definition are already included in our geometric families. (However it may be too bold to predict that all such families obtained by constraining the field of rationality can be constructed via geometry. For instance, the case of $\mathrm{GL}(n)$ for $n \geqslant 3$ is unclear.) On the other hand, we note that a result on the degree of the field of rationality by two of us [ST14] can be interpreted as the following statement: a harmonic family cannot be defined by a geometric construction, at least when the components at infinity are discrete series, because then the degree of the field of rationality would be bounded.

There are other examples such as the family of all Maass forms of eigenvalue $\frac{1}{4}$, say with integer coefficients. A letter [Sarn02], extended in [Brum03] shows that these forms are the same as certain Galois representations with a given H-type (see below). So this family too can be thought of in two ways.

3.3 Local Conditions with Measure Zero

In the construction of harmonic families we allowed ourselves to restrict a local component π_v to a nice subset $B_v \subset \widehat{H(\mathbb{Q}_v)}$ only for B_v of *positive* Plancherel volume. It is of interest to study some cases where B_v has measure zero. In doing so our main tool for studying the family, namely the trace formula, cannot be used effectively to isolate members of the family.

An important special case is to take π_∞ in a specified finite subset. For a fixed irreducible algebraic representation ξ of H over $\mathbb{C}$, take B_∞ to be the set of $\pi_\infty \in \widehat{H(\mathbb{R})}$ such that π_∞ is cohomological for ξ, namely $\pi_\infty \otimes \xi$ has nonzero Lie algebra cohomology in some degree. Then B_∞ is a finite set and often has Plancherel measure zero, for instance when $H = GL_n$ for $n \geqslant 3$ and ξ is arbitrary. Then $\pi \in \mathsf{A}(H)$ is such that $\pi_\infty \in B_\infty$ captures the information about the cohomology of the corresponding locally symmetric space for H with coefficients in a local system arising from ξ. One could refine the above choice of B_∞ by taking B_∞ to be a singleton $\pi_\infty \in \widehat{H(\mathbb{R})}$ which is cohomological for ξ but not a discrete series. As a further generalization of the special case above, one can take B_∞ to be a finite set consisting of $\pi_\infty \in \widehat{H(\mathbb{R})}$ which are C-algebraic in the sense of [BG11, Definition 2.3.3]. Roughly speaking, it means that the infinitesimal character of π_∞ is integral after twisting by the half sum of all positive roots of H. For example, we get the family of all weight 1 cuspforms and the family of all Maass cuspforms with Laplace eigenvalue 1/4 when $H = \mathrm{GL}(2)$ and B_∞ is a suitably chosen singleton.

In all of these cases it is already a difficult problem to enumerate $\mathfrak{H}$ as analytic conductor grows, in other words to study the asymptotic growth of (4). The answer to the last is well known when B_∞ consists of discrete series (and thus has a positive volume) by work of de George-Wallach [DeGW78].[11] In the case at hand concerning these families for which B_∞ is as above, there have been some conjectures and results concerning the sizes of these sets, see [SX91, CE09, Mar12]. Here we take a step further to pose the question of whether our main conjectures (Conjectures 1 and 2) are true for such families. The same question can be asked when we prescribe constraints at finite places by subsets of Plancherel measure zero.

3.4 Universal H-Types

As discussed above any of our pure families $\mathfrak{F}$ has a H-type associated with it, namely an H such that $\mu_{\mathrm{ST}}(\mathfrak{F}) = \mu_H$. Conversely one might try form universal families with a given H-type. Given H, the set of π's in $\mathsf{A}_{\mathrm{cusp}}(G)$ with $H_\pi = H$ would be such a family, or we could impose this condition on π's in any one of our families. There are some basic difficulties with such a construction. The first is

[11] A general *uniform* such limit multiplicity theorem has been derived recently in [ABBGNRS].

that we don't know how to define H_π in general. To begin with we can get around this problem by restricting to π's which are algebraic. The second problem is more serious and this is, in any generality we have no means of understanding such an $\mathfrak{F}_H$ and even the simplest requisite (4) is mysterious. Nevertheless it would seem safe to expect that the H-type of $\mathfrak{F}_H$ is H, and that Conjectures 1 and 2 would hold for any rich enough such $\mathfrak{F}_H$ (for example, it should at least be an infinite set). A numerical study of such "families," even for GL_2-forms, would be revealing. The difficulty with a theoretical study of such π's is closely related (but easier since we only ask asymptotic questions) to the analytic problem of recognizing π's in $\mathsf{A}_{\mathrm{cusp}}(G)$ with a given H_π that is raised by Langlands in his "Beyond Endoscopy" paper [Lan04].

While we can't attack these H-type families, we can in all cases (at least where the Noether conjecture is known) produce geometric parametric subfamilies of any of these types. In many cases these subfamilies are probably close to being a positive proportion of the H-types. In fact one of the standard approaches to the inverse Galois problem for special finite H's is to make a H-extension of $\mathbb{Q}(T_1, T_2, \ldots, T_m)$ and then to specialize the t's and use Hilbert irreducibility by counting (see [Ser97]). This very construction is a geometric parametric family according to our definition and of course it gives a large subfamily of such a H-type in our context.

There are some H's for which $\mathfrak{F}_H$ can be studied, primarily using class field theory. For $G = \mathrm{GL}(1)$ and H a finite cyclic subgroup of $\mathbb{C}^\times$, $\mathfrak{F}_H$ consists of all Dirichlet characters of order $|H|$ (for $|H| = 2$ this is the family $\mathfrak{F}^{(2)}$ from §2.1). Conjecture 1 is established without much trouble and $\mu_{\mathrm{ST}}(\mathfrak{F}_H) = \mu_H$ and for $|H| \geqslant 3$, $i_2(\mathfrak{F}_H) = 0$ and the symmetry type is $U(\infty)$. Conjecture 2 has been established for test functions of restricted support and numerically for $|H| = 3$ [GZ11, DFK04].

For $G = \mathrm{GL}(2)$ an interesting family related to H-types, with H not fixed but varying itself over a class of groups, was constructed by Hecke. Namely π's which are holomorphic cusp forms of weight 1 for which H_π is (finite) dihedral. One can study a refined version of Conjectures 1 and 2 for this family by collecting these forms into smaller packets which correspond to Hecke characters of the class group of $\mathbb{Q}(\sqrt{D})$, where $D \to -\infty$. This was done in [FI03] who show that the symmetry type is $\mathrm{Sp}(\infty)$. From our point of view this is "clear" since $H_{\mathfrak{F}}$ is a dihedral subgroup of $\mathrm{GL}_2(\mathbb{C})$ and in particular has Frobenius–Schur indicator equal to 1. Other than using class field theory and specifically 1-dimensional characters, we know of few examples where universal families of H-types can be studied.

3.5 Closing Comments

There are obvious variations on these constructions. We can combine number field (geometric) families and harmonic families. For example, let $\{K_i\}_{i \in I}$ be a family of number fields over $\mathbb{Q}$ of fixed degree d such that $\mathrm{disc}(K_i) \to \infty$. A further option is to require that in addition that K_i's have isomorphic Galois groups, that they satisfy a constraint on primes of ramification, or some other reasonable properties. Let H be a connected reductive group over $\mathbb{Q}$, with an L-group representation $r : {}^L H \to GL(m, \mathbb{C})$. The latter gives rise to an L-group representation $R : {}^L(\mathrm{res}_{K_i/\mathbb{Q}} H) \to$

$GL(md, \mathbb{C})$ by applying r on each copy of the dual group of H. The functorial lift corresponding to R is the functorial lift with respect to r over K_i followed by the automorphic induction from GL_m over K_i to GL_{md} over $\mathbb{Q}$. The resulting family $\mathfrak{F}$ is a family of automorphic L-functions of degree md. If functoriality for r (over each K_i) is known, then we may think of $\mathfrak{F}$ as a family of automorphic representations of $\mathrm{GL}(md, \mathbb{A})$ whose standard L-functions are as above. Sometimes it happens that every $L(s, \pi, R)$ factorizes as a product of L-functions and has a certain factor in common. In that case we may as well remove the common factor altogether. This construction yields examples which are not covered by families of the first chapter.

Finally note that for any of our parametric families one can impose further restrictions in exhausting $\mathfrak{F}$ or placing arithmetic conditions on the conductors. For example, one can collect the π's in $\mathfrak{F}$ in shells of given conductor (going to infinity) if these sets are large, or one can restrict to π's in $\mathfrak{F}$ with conductor a prime number. We view these as simple variations of our formation of families, albeit often technically more problematic. We have emphasized families which are cuspidal and pure, however mixed types arise naturally enough. A good example is that of Dedekind zeta functions of quartic field extensions of $\mathbb{Q}$. For these a positive proportion has Galois closure S_4 (as in Sect. 2.2) but there is also a positive proportion with Galois group D_4 whose invariants are quite different (see [Bha05]).

Appendix 1. Hasse–Weil L-Functions

Here we recall the definition of the Hasse–Weil L-function (2) and the modularity conjecture. The modularity conjecture (Conjecture 4 below) states that the L-functions arising from algebraic varieties over $\mathbb{Q}$ should be automorphic L-functions. In fact we will explain how L-functions are attached to l-adic Galois representations, in particular the étale cohomology space appearing in (2). To do so we recall the local Langlands correspondence for general linear groups in order to be precise about the matching of L-functions at ramified places. We also reformulate the modularity conjecture as a bijective correspondence between certain l-adic Galois representations and automorphic representations preserving L-functions, incorporating observations by Clozel and Fontaine-Mazur. The reader is referred to [Tay04] for an excellent survey of many topics discussed in this Appendix.

Let p be a prime and K a finite extension of $\mathbb{Q}_p$ with residue field cardinality q_K. Write W_K for the Weil group of K. For an algebraically closed field Ω of characteristic 0, denote by $\mathrm{Rep}_n(W_K)_\Omega$ (resp. $\mathrm{Rep}(\mathrm{GL}_n(K))_\Omega$) the set of isomorphism classes of n-dimensional Frobenius-semisimple Weil-Deligne representations of W_K (resp. irreducible smooth representations of $\mathrm{GL}_n(K)$) on k-vector spaces. For simplicity an element of $\mathrm{Rep}_n(W_K)$ will be called an (n-dimensional) WD-representation of W_K. Recall that such a representation is represented by (V, ρ, N) where V is an n-dimensional space over Ω, $\rho : W_K \to \mathrm{GL}_\Omega(V)$ is a representation such that $\rho(I_K)$ is finite and $\rho(w)$ is semisimple for every $w \in W_K$, and $N \in \mathrm{End}_\Omega(V)$ is a nilpotent operator such that $wNw^{-1} = |w|N$ where $|\cdot| : W_K \to \mathbb{R}^\times_{>0}$ is the transport of the modulus character on $K^\times$ via class field theory. The local Langlands reciprocity map is a bijection

$$\mathrm{rec}_K : \mathrm{Rep}(\mathrm{GL}_n(K))_{\mathbb{C}} \to \mathrm{Rep}_n(W_K)_{\mathbb{C}}$$

uniquely characterized by a list of properties, cf. [HT01]. In particular $L(s, \pi) = L(s, \mathrm{rec}(\pi))$, $\varepsilon(s, \pi, \psi) = \varepsilon(\mathrm{rec}(\pi), \psi)$ for any nontrivial additive character $\psi : F \to \mathbb{C}^\times$ (and a fixed Haar measure on F), and we also have an equality of conductors $f(\pi) = f(\mathrm{rec}_K(\pi))$. Here the local L and ε factors as well as conductors are independently defined on the left and right-hand sides. Here we will only recall the definition of the conductor and L-factor for WD-representations, which is due to Grothendieck, leaving the rest of definitions and further references to [Tate] and [Tay04]. For $(V, \rho, N) \in \mathrm{Rep}_n(W_K)_\Omega$ the conductor is given by

$$f(V) := \dim(V/V^{I_K} \cap \ker N) + \int_0^\infty \dim V/V^{I_K^u} du,$$

where I_K^u is the upper numbering filtration on the inertia group I_K. Now let Frob_K denote the geometric Frobenius in W_K/I_K. The local L-factor is defined to be

$$L(s, V) := \det(1 - \mathrm{Frob}_K q_K^{-s} | V^{I_K} \cap \ker N)^{-1}$$

so that we have the equality $L(s, \pi) = \det(1 - \mathrm{Frob}_K q_K^{-s} | \mathrm{rec}(\pi)^{I_K} \cap \ker N)^{-1}$ for $\pi \in \mathrm{Rep}(\mathrm{GL}_n(K))_{\mathbb{C}}$.

Now fix a field isomorphism $\iota : \overline{\mathbb{Q}}_l \simeq \mathbb{C}$ and let $\rho : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$ be a continuous semisimple Galois representation which is unramified at almost all primes and potentially semistable (equivalently de Rham) at places of F above l. Such a ρ is to be called *algebraic*. At each finite place v of F, there is a functor WD_v from continuous representations of $\mathrm{Gal}(\overline{F}_v/F_v) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$ (assumed potentially semistable if $v|l$) to WD-representations of W_{F_v}. The construction of WD_v relies on Grothendieck's monodromy theorem when $v \nmid l$ and Fontaine's work in l-adic Hodge theory if $v|l$.

The (global) conductor for ρ is $\prod_v \mathfrak{p}_v^{f_v}$ where $\mathfrak{p}_v$ is the prime ideal of $\mathcal{O}_F$ corresponding to v, and $f_v = f(\rho|_{\mathrm{Gal}(\overline{F}_v/F_v)})$. With ρ is associated a product function in a complex variable s, which is a priori formal infinite product:

$$L(s, \rho) := \prod_{v:\mathrm{finite}} L_v(s, \rho), \qquad L_v(s, \rho) := L(s, \iota \mathrm{WD}_v(\rho|_{\mathrm{Gal}(\overline{F}_v/F_v)})).$$

When ρ arises as a subquotient in the l-adic cohomology of an algebraic variety over F, one can apply Deligne's purity theorem to show that $L(s, \rho)$ converges absolutely for $\mathrm{Re}(s) \gg 1$ (with often explicit lower bound). Further there is a recipe for the archimedean factor $L_\infty(s, \rho)$ in terms of Hodge-Tate weights of ρ at places above l. (See the definition of $\Gamma(R, s)$ in [Tay04, § 2], taking R to be the induced representation of ρ from $\mathrm{Gal}(\overline{F}/F)$ to $\mathrm{Gal}(\overline{F}/\mathbb{Q})$.) This leads to a completed L-function

$$\Lambda(s,\rho) := L(s,\rho)L_\infty(s,\rho).$$

In the main body of the paper we were interested in the L-functions for Galois representations arising from varieties. Let X be a smooth projective variety over $\mathbb{Q}$, so X has good reduction modulo p for all but finitely many primes p. Then a reciprocity law for X on a concrete level would be a description of the number of points of X in $\mathbb{F}_p$ (and its finite extensions) in terms of automorphic data at p (i.e., local invariants at p of several automorphic representations of general linear groups) as p runs over the set of primes with good reduction, cf. [Lan76]. This may be thought of as a non-abelian reciprocity law generalizing the Artin reciprocity law in class field theory as well as an observation about elliptic modular curves by Eichler-Shimura. Now we say that $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$ *comes from geometry* if

- ρ is unramified away from finitely many primes,
- there exists a finite collection of smooth projective varieties X_i and integers $d_i, m_i \in \mathbb{Z}$ (indexed by $i \in I$) such that ρ appears as a subquotient of

$$\bigoplus_{i\in I} H^{d_i}_{\mathrm{et}}(X \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \overline{\mathbb{Q}}_l)(m_i).$$

As usual (m_i) denotes the Tate twist. One can speak of the obvious analogue with $\mathbb{Q}$ replaced by any finite extension F over $\mathbb{Q}$. In the language of L-functions the following conjecture presents a precise form of the reciprocity law as above.

Conjecture 4. *Let $\iota : \overline{\mathbb{Q}}_l \simeq \mathbb{C}$ be an isomorphism. If $\rho : \mathrm{Gal}(\overline{F}/F) \to \mathrm{GL}_n(\overline{\mathbb{Q}}_l)$ comes from geometry, then $L(s,\rho)$ is automorphic, namely there exists an isobaric automorphic representation Π of $\mathrm{GL}_n(\mathbb{A}_F)$ such that $L_v(s,\Pi) = L_v(s,\rho)$ at every finite place v and $v = \infty$ (so that $L(s,\Pi) = L(s,\rho)$ and $\Lambda(s,\Pi) = \Lambda(s,\rho)$).*

The Hasse-Weil conjecture predicts that $L(s,\rho)$ should have nice analytic properties such as analytic continuation, functional equation, and boundedness in vertical strips. If we believe in the Hasse-Weil conjecture, the converse theorem (discovered by Weil and then developed notably by Piatetskii-Shapiro and Cogdell) gives us a good reason to also believe that Conjecture 4 is true.

The conjecture begs two natural questions, namely a useful characterization of ρ coming geometry and a description of Π that arise from such ρ. The conjectural answers have been provided by Fontaine-Mazur and Clozel, respectively. Indeed a conjecture by Fontaine-Mazur asserts that a continuous semisimple l-adic representation ρ comes from geometry if and only if it is algebraic. Following Clozel a cuspidal automorphic representation Π of $\mathrm{GL}_n(\mathbb{A}_F)$ is said to be *L-algebraic* if, roughly speaking, the L-parameters for Π at infinite places consist of algebraic characters in a suitable sense (see [BG11] for the definition; this differs from [Clozel] in that no adjustment by the $\frac{n-1}{2}$-th power is made, cf. comments below

Conjecture 5). An isobaric sum of cuspidal representations $\boxplus_{i=1}^{r}\Pi_i$ is algebraic if every Π_i is algebraic. Then we can reformulate Conjecture 4 as one about the existence of the global Langlands correspondence preserving L-functions:

Conjecture 5. *Fix ι as above. Then there exists a bijection $\Pi \leftrightarrow \rho$ between the set of L-algebraic isobaric automorphic representations of* $\mathrm{GL}_n(\mathbb{A}_F)$ *and the set of algebraic n-dimensional semisimple l-adic representations of* $\mathrm{Gal}(\overline{F}/F)$ *(up to isomorphism) such that the local L-factors are the same, so that* $L(s,\Pi) = L(s,\rho)$ *and* $\Lambda(s,\Pi) = \Lambda(s,\rho)$.

Remark. The strong multiplicity one theorem and the Chebotarev density theorem imply that if there is a correspondence $\Pi \leftrightarrow \rho$ as above then it should be a bijective correspondence and unique (but it does depend on the choice of ι). It is expected that the set of cuspidal Π maps onto the set of irreducible ρ. A stronger property, often referred to as the local-global compatibility, is believed to be true at finite places v: it says that $\mathrm{rec}_{F_v}(\Pi_v) = \iota\mathrm{WD}(\rho|_{\mathrm{Gal}(\overline{F}_v/F_v)})$. (This is stronger only at ramified places.) In particular it should be true that ρ and Π have the same conductor (at finite places). Since we are concerned with unitary duals, we have adopted the unitary normalization for the Langlands correspondence and algebraicity. For arithmetic considerations it is customary to twist Π by the $\frac{1-n}{2}$-th power of the modulus character in the conjecture. If so, one should replace "L-algebraic" by "C-algebraic," cf. [BG11].

It is worth noting that Conjecture 4 suffices for our purpose in discussing geometric families. An important part of the Langlands program has been to confirm Conjecture 4 when ρ is the l-adic cohomology of a Shimura variety (in any degree), which in turn led to many instances of the map $\Pi \mapsto \rho$ in Conjecture 5. Another remarkable result toward the conjectures is the modularity of elliptic curves over $\mathbb{Q}$ due to Wiles and Breuil-Conrad-Diamond-Taylor, who identified $L(s,\rho)$ with the L-function of a weight 2 modular form when ρ is the étale H^1 of an elliptic curve over $\mathbb{Q}$. Recent developments include modularity lifting and potential modularity theorems. As we have no capacity to make a long list of all known cases of either Conjecture 4 or 5, we mention survey articles [Tay04] and [Harr10] for the reader to begin reading about progress until 2009.

We close the discussion with a comment on the unitarity of local components and the issue of correct twist, cf. Remark (iv) below the definition of geometric families in Sect. 1. Consider the automorphic representation Π corresponding via the above conjectures to $\rho = H^d_{\mathrm{et}}(X \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \overline{\mathbb{Q}}_l)$ for a smooth proper variety X over $\mathbb{Q}$ (which is not necessarily geometrically connected). Set $\Pi' := \Pi \otimes |\det|^{d/2}$. If X has good reduction modulo a prime p, then the geometric Frobenius acts on the H^d-cohomology with absolute values $p^{d/2}$ under any choice of ι. (This is Deligne's theorem on the Weil Conjectures if $p \neq l$. The argument extends to $p = l$ by work of Katz-Messing.) Hence the twist the Satake parameters of Π'_p have absolute value 1, so Π'_p is unitary. In general when X has bad reduction modulo p, the unitarity of Π'_p can be deduced from the weight-monodromy conjecture in mixed characteristic (as stated in [Saito]). Despite recent progress, cf. [Sch12], the latter conjecture is still open. What we said of ρ should remain true when ρ is a subquotient of $H^d_{\mathrm{et}}(X \times_{\mathbb{Q}} \overline{\mathbb{Q}}, \overline{\mathbb{Q}}_l)$.

Appendix 2. Non-Criticality of the Central Value for Orthogonal Representations

Deligne ([Del79]) made a conjecture on special values of motivic L-functions. For a given L-function there is a set of the so-called *critical* values of s to which his conjecture applies. For our purpose we take on faith a motivic version of Conjecture 5 (cf. [Lan12, § 6] and Remark 3.5 above) on the existence of a bijection between absolutely irreducible pure motives M of rank n over $\mathbb{Q}$ and cuspidal C-algebraic automorphic representations π of $\mathrm{GL}_n(\mathbb{A})$ such that

$$L(s + \frac{n-1}{2}, M) = L(s, \pi).$$

Thereby Deligne's conjecture translates to a conjecture on automorphic L-functions. We copy the definition of s being critical from the motivic side to the automorphic side in the obvious way. We are particularly interested in the question of whether the central value $s = 1/2$ is critical for a cuspidal automorphic L-function which is unitarily normalized (for this a twist by a suitable power of the modulus character may be needed). The goal of appendix is to show

Proposition 3.6. *Suppose that a cuspidal automorphic representation* π *of* $\mathrm{GL}_n(\mathbb{A})$ *is*

(1) orthogonal (i.e., π *is self-dual and* $L(s, \pi, \mathrm{Sym}^2)$ *has a pole) and*
(2) regular and C-algebraic.

Then $s = 1/2$ *is not critical for* $L(s, \pi)$*.*

The statement, in particular the definition of criticality, is unconditional in that no unproven assertions need to be assumed. However the proof is conditional on Conjecture 5 as well as various conjectures around motives that are supposed to be true (see Sect. 1 of [Del79] for the latter). We freely assume them below.

Proof. There should be a pure irreducible rank n motive M over $\mathbb{Q}$ corresponding to π. We follow the conventional normalization so that the weight of M is $w = n - 1$. (Note that the second assumption on π implies that M has Hodge numbers 0 or 1. In the Hodge realization the dimension of $M^{p,q}$ is at most one, and zero if $p+q \neq n-1$.) Since π is self-dual, M is self-dual up to twist. More precisely there is a perfect pairing

$$M \otimes M \to \mathbb{Q}(1-n)$$

where $\mathbb{Q}(1-n)$ is the $(1-n)$-th power of the Tate motive.

The center of symmetry for $L(s, M)$, the L-function associated with M, is at $s = (1 + w)/2 = n/2$. The necessary condition (which may not be sufficient) for it to be critical is that $n/2 \in \mathbb{Z}$, namely that n is even (so w is odd). Hence we may and will assume that n is even. Now consider the l-adic realization

$$M_l \otimes M_l \to \mathbb{Q}_l(1-n),$$

where M_l is now an irreducible l-adic representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. By a result of Bellaiche-Chenevier's ([BC11]) the sign of M_l is equal to $(-1)^{n-1} = -1$, meaning that the above pairing on M_l is symplectic. (To apply their result we need both assumptions (1) and (2) on π.) Translating back to the automorphic side we deduce that π is also symplectic. We have shown that if $s = 1/2$ is critical then π is symplectic, completing the proof. $\square$

Example. When $n = 1$ and π corresponds to a Dirichlet character χ, it is well known that the central value $s = 1/2$ for $L(s,\chi)$ is not critical. In this case π is clearly orthogonal and the proposition applies.

Example. Consider the case of $n = 2$ where π corresponds to weight k cuspforms ($k \geqslant 1$). Since we are concerned with self-dual representations, we normalize the correspondence such that π is self-dual. Then π is regular algebraic if and only if k is even. (To deal with odd weight forms, one could twist π by a half-power of the modulus character, but then π would be self-dual only up to a twist.) In case k is even, we associate with π a pure motive M of rank 2 and weight 1 such that $\dim M^{1-k/2,k/2} = \dim M^{1-k/2,k/2} = 1$. It is equipped with a symplectic pairing $M \times M \to \mathbb{Q}(-1)$.

Acknowledgements Thanks to Manjul Bhargava, Daniel Bump, Brian Conrey, Dorian Goldfeld, Henryk Iwaniec, Philippe Michel, Emmanuel Kowalski, Erez Lapid, Gopal Prasad, Zeev Rudnick, Jean-Pierre Serre, Arul Shankar, Anders Södergren, Kannan Soundararajan, Akshay Venkatesh, Jun Yu, and especially Nicholas Katz, for discussions and insights on various aspects of the paper. S.W.S. is grateful to Princeton University and the Institute for Advanced Study for their hospitality during several short visits. He acknowledges partial supports from NSF grant DMS-1162250 and a Sloan Fellowship. N.T. acknowledges partial support from NSF grant DMS-1200684.

References

[ABBGNRS] M. Abert, N. Bergeron, I. Biringer, T. Gelander, N. Nikolov, J. Raimbault, I. Samet, On the growth of Betti numbers of locally symmetric spaces. C. R. Math. Acad. Sci. Paris **349** (15–16), 831–835 (2011)

[AYY] J. An, J.-K. Yu, J. Yu, On the dimension data of a subgroup. J. Differ. Geometry (to appear)

[Art89] J. Arthur, The L^2-Lefschetz numbers of Hecke operators. Invent. Math. **97**(2), 257–290 (1989)

[Art13] J. Arthur, *The Endoscopic Classification of Representations*. American Mathematical Society Colloquium Publications, vol. 61 (American Mathematical Society, Providence, RI, 2013). Orthogonal and symplectic groups

[BZ08] S. Baier, L. Zhao, On the low-lying zeros of Hasse-Weil L-functions for elliptic curves. Adv. Math. **219**(3), 952–985 (2008)

[BC11] J. Bellaïche, G. Chenevier, The sign of Galois representations attached to automorphic forms for unitary groups. Compos. Math. **147**(5), 1337–1352 (2011)

[BKLPR] M. Bhargava, D. Kane, H. Lenstra, B. Poonen, E. Rains, *Modeling the Distribution of Ranks, Selmer Groups, and Shafarevich-Tate Groups of Elliptic Curves* (2013). Preprint. Available at arXiv:1304.3971

[BS13] M. Bhargava, A. Shankar, *The Average Size of the* 5*-Selmer Group of Elliptic Curves is* 6*, and the Average Rank is Less Than* 1 (2013). Preprint. arXiv:1312.7859

[BS14] M. Bhargava, C. Skinner, *A Positive Proportion of Elliptic Curves Over $\mathbb{Q}$ Have Rank One* (2014). Preprint. arXiv:1401.0233

[Bha05] M. Bhargava, The density of discriminants of quartic rings and fields. Ann. Math. (2) **162** (2), 1031–1063 (2005)

[Bha10] M. Bhargava, The density of discriminants of quintic rings and fields. Ann. Math. (2) **172**(3), 1559–1591 (2010)

[Birch] B.J. Birch, How the number of points of an elliptic curve over a fixed prime field varies. J. London Math. Soc. **43**, 57–60 (1968)

[Bomb] E. Bombieri, On the large sieve. Mathematika **12**, 201–225 (1965)

[Borel] A. Borel, Automorphic L-functions, in *Automorphic Forms, Representations and L-Functions (Proceedings of the Symposia Pure Mathematics, Oregon State University, Corvallis, OR, 1977), Part 2* (1979), pp. 27–61

[BR95] M. Borovoi, Z. Rudnick, Hardy-Littlewood varieties and semisimple groups. Invent. Math. **119** (1), 37–66 (1995)

[BM] F. Brumley, D. Milicevic, Counting cusp forms by analytic conductor (in preparation)

[Brum03] F. Brumley, Maass cusp forms with quadratic integer coefficients. Int. Math. Res. Not. **18**, 983–997 (2003)

[Brum06] F. Brumley, Effective multiplicity one on GL_N and narrow zero-free regions for Rankin-Selberg L-functions. Am. J. Math. **128**(6), 1455–1474 (2006)

[BG11] K. Buzzard, T. Gee, The conjectural connections between automorphic representations and Galois representations, in *Proceedings of the LMS Durham Symposium 2011* (to appear)

[CE09] F. Calegari, M. Emerton, Bounds for multiplicities of unitary representations of cohomological type in spaces of cusp forms. Ann. Math. (2) **170**(3), 1437–1446 (2009)

[CS82] J.W.S. Cassels, A. Schinzel, Selmer's conjecture and families of elliptic curves. Bull. London Math. Soc. **14** (4), 345–348 (1982)

[Clozel] L. Clozel, Motifs et formes automorphes: applications du principe de fonctorialité, in *Auto-Morphic Forms, Shimura Varieties, and L-Functions*, vol. I, Ann Arbor, MI, 1988 (1990), pp. 77–159

[CFKRS] J.B. Conrey, D.W. Farmer, J.P. Keating, M.O. Rubinstein, N.C. Snaith, Integral moments of L-functions. Proc. London Math. Soc. (3) **91**(1), 33–104 (2005)

[LMS07] J.B. Conrey, D.W. Farmer, F. Mezzadri, N.C. Snaith (eds.), *Ranks of Elliptic Curves and Random Matrix Theory*. London Mathematical Society Lecture Note Series, vol. 341 (Cambridge University Press, Cambridge, 2007)

[CS] J.B. Conrey, N.C. Snaith, In support of n-correlation, J Commun. Math. Phys. **330**(2), 639–653 (2014)

[DH71] H. Davenport, H. Heilbronn, On the density of discriminants of cubic fields. II. Proc. R. Soc. London Ser. A **322**(1551), 405–420 (1971)

[DFK04] C. David, J. Fearnley, H. Kisilevsky, On the vanishing of twisted L-functions of elliptic curves. Exp. Math. **13**(2), 185–198 (2004)

[DeGW78] D.L. de George, N.R. Wallach, Limit formulas for multiplicities in $\mathrm{L}^2(\Gamma\backslash G)$. Ann. Math. (2) **107**(1), 133–150 (1978)

[Del79] P. Deligne, Valeurs de fonctions L et périodes d'intégrales, in *Automorphic Forms, Representations and L-Functions (Proceedings of the Symposia Pure Mathematics, Oregon State Univ., Corvallis, Ore., 1977), Part 2* (1979), pp. 313–346. With an appendix by N. Koblitz and A. Ogus

[Del80] P. Deligne, La conjecture de Weil. II. Inst. Hautes Études Sci. Publ. Math. **52**, 137–252 (1980)

[LNM82] P. Deligne, J.S. Milne, A. Ogus, K.-Y. Shih, *Hodge Cycles, Motives, and Shimura Varieties*. Lecture Notes in Mathematics, vol. 900 (Springer, Berlin, 1982)

[DM06] E. Dueñez, S.J. Miller, The low-lying zeros of a GL(4) and a GL(6) family of L- functions. Compos. Math. **142**(6), 1403–1425 (2006)

[DM09] E. Dueñez, S.J. Miller, The effect of convolving families of L-functions on the underlying group symmetries. Proc. Lond. Math. Soc. (3) **99**(3), 787–820 (2009)

[DKV83] J.J. Duistermaat, J.A.C. Kolk, V.S. Varadarajan, Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups. Compos. Math. **49**(3), 309–398 (1983)

[DK00] W. Duke, E. Kowalski, A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations. Invent. Math. **139**(1), 1–39 (2000). With an appendix by Dinakar Ramakrishnan

[Elk07] N. Elkies, *Three Lectures on Elliptic Surfaces and Curves of High Rank*, Oberwolfach Report (2007). arXiv:0709.2908

[ES13] N. Elkies, M. Schütt, Modular forms and K3 surfaces. Adv. Math. **240**, 106–131 (2013)

[ERR13] A. Entin, E. Roditty-Gersho, Z. Rudnick, Low-lying zeros of quadratic Dirichlet L-functions, hyper-elliptic curves and Random Matrix Theory. Geom. Funct. Anal. **23**(4), 307–319 (2013)

[FL15] T. Finis, E. Lapid, *An Approximation Principle for Congruence Subgroups II: Application to the Limit Multiplicity Problem*. Available at arXiv:1504.04795

[FLM11] T. Finis, E. Lapid, W. Müller, On the spectral side of Arthur's trace formula—absolute convergence. Ann. Math. (2) **174** (1), 173–195 (2011)

[FLM12] T. Finis, E. Lapid, W. Müller, On the degrees of matrix coefficients of intertwining operators. Pac. J. Math. **260**(2), 433–456 (2012)

[FLM15] T. Finis, E. Lapid, W. Müller, Limit multiplicities for principal congruence subgroups of GL(n) *and* SL(n). J. Inst. Math. Jussieu **14**(3), 589–638 (2015)

[FI03] E. Fouvry, H. Iwaniec, Low-lying zeros of dihedral L-functions. Duke Math. J. **116**(2), 189–217 (2003)

[FNT92] E. Fouvry, M. Nair, G. Tenenbaum, L'ensemble exceptionnel dans la conjecture de Szpiro. Bull. Soc. Math. France **120**(4), 485–506 (1992)

[FIMR] J.B. Friedlander, H. Iwaniec, B. Mazur, K. Rubin, The spin of prime ideals. Invent. Math. **193**(3), 697–749 (2013)

[Gao] P. Gao, *n-level density of the low-lying zeros of quadratic Dirichlet L-functions*, Thesis (Ph.D.), University of Michigan, ProQuest LLC, Ann Arbor, MI, 2005

[GZ11] P. Gao, L. Zhao, One level density of low-lying zeros of families of L-functions. Compos. Math. **147**(1), 1–18 (2011)

[GJ72] R. Godement, H. Jacquet, *Zeta Functions of Simple Algebras*. Lecture Notes in Mathematics, vol. 260 (Springer, Berlin, 1972)

[GK13] D. Goldfeld, A. Kontorovich, *On the GL(3) Kuznetsov Formula with Applications to Symmetry Types of Families of L-Functions*. Automorphic representations and *L*-functions. Tata Inst. Fundam. Res. Stud. Math. **22**, 263–310 (2013); Tata Inst. Fund. Res. Mumbai

[Greaves] G. Greaves, Power-free values of binary forms. Quart. J. Math. Oxford Ser. (2) **43**(169), 45–65 (1992)

[Harr10] M. Harris, Arithmetic applications of the Langlands program. Jpn. J. Math. **5**(1), 1–71 (2010)

[HSBT] M. Harris, N. Shepherd-Barron, R. Taylor, A family of Calabi-Yau varieties and potential automorphy. Ann. Math. (2) **171**(2), 779–813 (2010)

[HT01] M. Harris, R. Taylor, *The Geometry and Cohomology of Some Simple Shimura Varieties.* Annals of Mathematics Studies, vol. 151 (Princeton University Press, Princeton, NJ, 2001). With an appendix by Vladimir G. Berkovich

[HS13] R. Harron, A. Snowden, Counting elliptic curves with prescribed torsion. J. Reine Angew. Math. (to appear)

[Hea94] D.R. Heath-Brown, The size of Selmer groups for the congruent number problem. II. Invent. Math. **118** (2), 331–370 (1994). With an appendix by P. Monsky

[Helf04] H. Helfgott, *On the Behaviour of Root Numbers in Families of Elliptic Curves* (2004). arXiv:math/0408141

[Helf05] H. Helfgott, *The Parity Problem of Irreducible Cubic Forms* (2005). Preprint. arXiv:math/0501177

[Helgason] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*. Graduate Studies in Mathematics, vol. 34 (American Mathematical Society, Providence, RI)

[Hoo76] C. Hooley, *Applications of Sieve Methods to the Theory of Numbers*. Cambridge Tracts in Mathematics, vol. 70 (Cambridge University Press, Cambridge, 1976)

[IS00] H. Iwaniec, P. Sarnak, Perspectives on the analytic theory of L-functions. Geom. Funct. Anal. Special Volume, Part II, 705–741 (2000)

[ILS01] H. Iwaniec, W. Luo, P. Sarnak, Low-lying zeros of families of L-functions. Inst. Hautes Études Sci. Publ. Math. **91**, 55–131 (2001)

[JPSS83] H. Jacquet, I.I. Piatetskii-Shapiro, J.A. Shalika, Rankin-Selberg convolutions. Am. J. Math. **105** (2), 367–464 (1983)

[Kala] V. Kala, *Density of Self-Dual Automorphic Representations of* GL_N $(\mathbb{A}_{\mathbb{Q}})$, PhD thesis, Purdue, 2014. arXiv:1406.0385

[Katz01] N.M. Katz, Frobenius-Schur indicator and the ubiquity of Brock-Granville quadratic excess. Finite Fields Appl. **7**(1), 45–69 (2001). Dedicated to Professor Chao Ko on the occasion of his 90th birthday

[Katz09] N.M. Katz, Another look at the Dwork family, in *Algebra, Arithmetic, and Geometry: In Honor of Yu. I. Manin*, vol. II (2009), pp. 89–126

[Katz13] N.M. Katz, *Sato–Tate in the higher dimensional case: elaboration of 9.5.4 in Serre's* $N_X(p)$ *book*. J Enseign. Math. **59**(3–4), 359–377 (2013). Available at math.princeton.edu/~nmk/

[KS-b] N.M. Katz, P. Sarnak, *Random Matrices, Frobenius Eigenvalues, and Monodromy*. American Mathematical Society Colloquium Publications, vol. 45 (American Mathematical Society, Providence, RI, 1999)

[KS99] N.M. Katz, P. Sarnak, Zeroes of zeta functions and symmetry. Bull. Am. Math. Soc. (N.S.) **36**(1), 1–26 (1999)

[KS00] J.P. Keating, N.C. Snaith, Random matrix theory and L-functions at $s = 1/2$. Commun. Math. Phys. **214**(1), 91–110 (2000)

[Kha10] C. Khare, Serre's conjecture and its consequences. Jpn. J. Math. **5**(1), 103–125 (2010)

[KST16] J.L. Kim, S.W. Shin, N. Templier, Asymptotics and local constancy of characters of p-adic groups, in *Proceedings of Simons Symposium on Automorphic Forms and the Trace Formula*

[Kow13] E. Kowalski, Families of cusp forms, Actes de la Conférence. Théorie des Nombres et Applications. Publ. Math. Besanç con Algébre Théorie Nr. (2013), pp. 5–40

[KST12] E. Kowalski, A. Saha, J. Tsimerman, Local spectral equidistribution for Siegel modular forms and applications. Compos. Math. **148**(2), 335–384 (2012)

[LW54] S. Lang, A. Weil, Number of points of varieties in finite fields. Am. J. Math. **76**, 819–827 (1954)

[Lan76] R. Langlands, Some contemporary problems with origins in the Jugendtraum, in *Mathematical Developments Arising from Hilbert Problems (Proceedings of the Symposia Pure Mathematics, Vol. XXVIII, Northern Illinois University, De Kalb, IL, 1974)* (1976), pp. 401–418

[Lan04] R. Langlands, Beyond endoscopy, in *Contributions to Automorphic Forms, Geometry, and Number Theory* (Johns Hopkins University Press, Baltimore, 2004), pp. 611–697

[Lan12] R. Langlands, A prologue to "Functoriality and reciprocity" Part I. Pac. J. Math. **260**(2), 582–663 (2012)

[Lapid] E. Lapid, On the root number of representations of orthogonal type. Compos. Math. **140**(2), 274–286 (2004)

[LM09] E. Lapid, W. Müller, Spectral asymptotics for arithmetic quotients of $\mathrm{SL}(n, \mathbb{R})/\mathrm{SO}(n)$. Duke Math. J. **149**(1), 117–155 (2009)

[LV07] E. Lindenstrauss, A. Venkatesh, Existence and Weyl's law for spherical cusp forms. Geom. Funct. Anal. **17**(1), 220–251 (2007)

[Mar12] S. Marshall, Bounds for the multiplicities of cohomological automorphic forms on GL_2. Ann. Math. (2) **175** (3), 1629–1651 (2012)

[Mat93] Y.V. Matiyasevich, *Hilbert's Tenth Problem*. Foundations of Computing Series (MIT Press, Cambridge, MA, 1993). Translated from the 1993 Russian original by the author, With a foreword by Martin Davis

[MT] J. Matz, N. Templier, *Sato–Tate Equidistribution for Families of Hecke–Maass Forms on SL(n,R)/SO(n)* (2015). Available at arXiv:1505.07285
[Mich07] Ph. Michel, Analytic number theory and families of automorphic L-functions, in *Automorphic Forms and Applications*. Park City Mathematics Series, vol. 12 (American Mathematical Society, Providence, 2007), pp. 181–295
[MV10] Ph. Michel, A. Venkatesh, The subconvexity problem for GL_2. Publ. Math. Inst. Hautes Études Sci. **111**, 171–271 (2010)
[Miller] S.J. Miller, One- and two-level densities for rational families of elliptic curves: evidence for the underlying group symmetries. Compos. Math. **140**(4), 952–992 (2004)
[MS04] W. Müller, B. Speh, Absolute convergence of the spectral side of the Arthur trace formula for GL_n. Geom. Funct. Anal. **14**(1), 58–93 (2004). With an appendix by E.M. Lapid
[Nagao] K.-I. Nagao, $Q(T)$-rank of elliptic curves and certain limit coming from the local points. Manuscripta Math. **92** (1), 13–32 (1997). With an appendix by Nobuhiko Ishida, Tsuneo Ishikawa and the author
[Nelson] P. Nelson, *Quadratic Statistics on Quaternion Algebras*, I. arXiv:1601.02526
[Ngo] B.C. Ngô, Le lemme fondamental pour les algébres de Lie. Publ. Math. Inst. Hautes Études Sci. **111**, 1–169 (2010)
[OS93] A.E. Özlük, C. Snyder, Small zeros of quadratic L-functions. Bull. Aust. Math. Soc. **47**(2), 307–319 (1993)
[PS85] K. Paranjape, D. Ramakrishnan, Modular forms and Calabi-Yau varieties. Arithmetic and Geometry. London Mathematical Society Lecture Note Series, vol. 420 (Cambridge University Press, Cambridge, 2015), pp. 351–372
[PR15] D. Ramakrishnan, K. Paranjape, *Modular Forms and Calabi-Yau Varieties* (2008). Preprint
[Rizzo] O.G. Rizzo, Average root numbers for a nonconstant family of elliptic curves. Compositio Math. **136**(1), 1–23 (2003)
[RS98] M. Rosen, J.H. Silverman, On the rank of an elliptic surface. Invent. Math. **133**(1), 43–67 (1998)
[Rub01] M. Rubinstein, Low-lying zeros of L-functions and random matrix theory. Duke Math. J. **109**(1), 147–181 (2001)
[Rub05] M. Rubinstein, Computational methods and experiments in analytic number theory, in *Recent Perspectives in Random Matrix Theory and Number Theory*. London Mathematical Society Lecture Note Series, vol. 322 (Cambridge University Press, Cambridge, 2005), pp. 425–506
[RS96] Z. Rudnick, P. Sarnak, Zeros of principal L-functions and random matrix theory. Duke Math. J. **81**(2), 269–322 (1996). A celebration of John F. Nash, Jr.
[Saito] T. Saito, Weight spectral sequences and independence of l. J. Inst. Math. Jussieu **2**(4), 583–634 (2003)
[Sarn08] P. Sarnak, *On the Definition of Families* (2008). Available at http://publications.ias.edu/sarnak
[Sarn02] P. Sarnak, Maass cusp forms with integer coefficients, in *A Panorama of Number Theory or the View from Baker's Garden (Zürich, 1999)* (Cambridge University Press, Cambridge, 2002), pp. 121–127
[Sarn05] P. Sarnak, Notes on the generalized Ramanujan conjectures, in *Harmonic Analysis, the Trace Formula, and Shimura Varieties*. Clay Mathematics Proceedings, vol. 4 (American Mathematical Society, Providence, 2005), pp. 659–685
[SX91] P. Sarnak, X. Xue, Bounds for multiplicities of automorphic representations. Duke Math. J. **64**(1), 207–227 (1991)
[Sch12] P. Scholze, Perfectoid spaces. Publ. Math. Inst. Hautes Ëtudes Sci. **116**, 245–313 (2012)
[Ser97] J.-P. Serre, *Lectures on the Mordell-Weil Theorem*. Third, Aspects of Mathematics (Friedr. Vieweg & Sohn, Braunschweig, 1997). Translated from the French and edited by Martin Brown from notes by Michel Waldschmidt, With a foreword by Brown and Serre
[Ser12] J.-P. Serre, *Lectures on* $N_X(p)$. Research Notes in Mathematics, vol. 11 (Chapman & Hall/CRC Press, Boca Raton, FL, 2012)

[SST15] A. Shankar, A. Södergren, N. Templier, *Low-Lying Zeros of Certain Families of Artin L-Functions* (2015). Available at arXiv:1507.07031

[ST16] S.W. Shin, N. Templier, Sato–Tate Theorem for Families and low-lying zeros of automorphic L-functions. With Appendix 1 by R. Kottwitz and Appendix 2 by R. Cluckers, J. Gordon and I. Halupczok. Invent. Math. **203**(1), 1–177 (2016)

[ST14] S.W. Shin, N. Templier, On fields of rationality for automorphic representations. Compos. Math. **150**(12), 2003–2053 (2014)

[Ste68] R. Steinberg, *Lectures on Chevalley Groups* (Yale University, New Haven, CT, 1968). Notes prepared by John Faulkner and Robert Wilson

[Tate] J. Tate, Number theoretic background, in *Proceedings of the Symposia Pure Mathematics* (Oregon State University, Corvallis, OR, Part 2, 1979), pp. 3–26

[Tay04] R. Taylor, Galois representations. Ann. Fac. Sci. Toulouse **13**, 73–119 (2004)

[Vin65] A.I. Vinogradov, The density hypothesis for Dirichet L-series. Izv. Akad. Nauk SSSR Ser. Mat. **29**, 903–934 (1965)

[Wash87] L.C. Washington, Class numbers of the simplest cubic fields. Math. Comp. **48**(177), 371–384 (1987)

[Weil] A. Weil, On the Riemann hypothesis in function fields. Proc. Natl. Acad. Sci. U. S. A. **27**, 345–347 (1941)

[Weyl] H. Weyl, *The Classical Groups. Their Invariants and Representations* (Princeton University Press, Princeton, NJ, 1939)

[WY92] D.J. Wright, A. Yukie, Prehomogeneous vector spaces and field extensions. Invent. Math. **110**(2), 283–314 (1992)

[Yang] A. Yang, *Distribution problems associated to zeta functions and invariant theory*, PhD thesis, Princeton University, 2009

[Young] M.P. Young, Low-lying zeros of families of elliptic curves. J. Am. Math. Soc. **19**(1) (2006), 205–250

[Yu13] J. Yu, *On the Dimension Datum Problem and the Linear Dependence Problem* (2013). Available at arXiv:1303.0811