ARITHMETIC QUANTUM CHAOS ON LOCALLY SYMMETRIC SPACES

LIOR SILBERMAN

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
MATHEMATICS

NOVEMBER, 2005

Abstract

We report progress on the equidistribution problem of automorphic forms on locally symmetric spaces. First, generalizing work of Zelditch-Wolpert we construct a representation theoretic analog of the micro-local lift, showing that (under a technical condition of non-degeneracy) every weak-* limit of the generalized Wigner measures associated to a sequence of Maass forms with divergent spectral parameters on a locally symmetric space $\Gamma\backslash G/K$ can be lifted to a measure on the homogeneous space $\Gamma\backslash G$ which is invariant by a maximal split torus A in G. Secondly, we consider the case where $G \simeq \mathrm{PGL}_d(\mathbb{R})$ and $\Gamma < G$ is a lattice associated to a division algebra over $\mathbb{Q}$ of prime degree d. When the measures are associated to Hecke-Maass eigenforms, we generalize the work of Bourgain-Lindenstrauss to show that every non-trivial $a \in A$ acts with positive entropy on each ergodic component of the lifted measure. Applying recent measure rigidity results of Einsiedler-Katok we find that the limit measure must be the Haar measure on $\Gamma\backslash G$. In particular we prove that a non-degenerate sequence of Hecke-Maass forms becomes equidistributed in $\Gamma\backslash G/K$ in the semiclassical limit.

These results arise from joint work with Akshay Venkatesh of the Courant Institute of Mathematical Sciences.

Acknowledgments

I am privileged to have been supervised by Peter Sarnak, who has shown me a lot of mathematics (and patience), and has directed me toward even more mathematics. Membership in his mathematical community is a great experience.

The community in the department has given me a lot, both mathematically and personally. I have lived and learned here, and among many names would like to mention those of Adrian Banner, Brent Doran, Harald Helfgott, Jordan Ellenberg and Andreea Nicoara. But this does not diminish the contribution of everyone else, who are always willing to give of their time to discuss mathematics. I would also like to thank the assistance of the staff of the department over the last five years, in particular Scott Kenney and Jill LeClair.

This thesis represents joint work with Akshay Venkatesh, a former fellow student at Princeton. Collaborating with him has been a pleasure, and I hope to continue our work in the future. This work started with our conversions at the Clay Mathematics Institute summer school in Toronto during June 2004, but has its foundations at an ersatz seminar convened by Prof. Sarnak that May to study the then-recent work of Elon Lindenstrauss on the arithmetic quantum chaos problem on hyperbolic surfaces. We have also benefited from conversations with Joseph Bernstein, Manfred Einsiedler, Erez Lapid, Elon Lindenstrauss, Dima Jakobson and Andre Reznikov.

I have immensely enjoyed the graduate student community at Princeton, a community that has made my stay here even more memorable. Special thanks to Stefan van der Elst, Thomas Horine, Peter Keevash, Christopher Mole and Juliet O'Brien, and to Eric Adelizzi, Morten Kloster and Manish Vachharajani. Also not forgotten are my fellow Officers on various Graduate College House Committees, and fellow members of the Graduate Student Government and other student organizations who give much to this community. Last, but certainly not least, I would like to thank Joanna Karczmarek. You have been wonderful.

To my parents, Giora and Silvia, you have brought me up to where I am today. I dedicate this thesis to you. I have missed you and my brother Noam, but your support and encouragement have always been with me.

To my parents

Contents

CHAPTER 1

Introduction

1.1. General starting point: the semi-classical limit on Riemannian manifolds

Let Y be a compact Riemannian manifold, with the associated Laplace operator Δ and Riemannian measure $d\rho$. An important problem of harmonic analysis (or mathematical physics) on Y is understanding the asymptotic behaviour of eigenfunctions of Δ in the large eigenvalue limit. The equidistribution problem asks whether for an eigenfunction ψ with a large eigenvalue λ, $|\psi(x)|$ is approximately constant on Y. This can be approached "pointwise" and "on average" (bounding $\|\psi\|_{L^\infty(Y)}$ and $\|\psi\|_{L^p(Y)}$ in terms of λ, respectively), or "weakly": asking whether as $|\lambda| \to \infty$, the probability measures defined by $d\bar{\mu}_\psi(y) = |\psi(y)|^2 \, d\rho(y)$ converge in the weak-* sense to the "uniform" measure $\frac{d\rho}{\mathrm{vol}(Y)}$. For example, Sogge [**29**] derives L^p bounds for $2 \leq p \leq \infty$, and in the special case of Hecke eigenfunctions on hyperbolic surfaces, Iwaniec and Sarnak [**15**] gave a non-trivial L^∞ bound. Here we will consider the weak-* equidistribution problem for a special class of manifolds and eigenfunctions.

A general approach to the weak-* equidistribution problem was found by Šnirel′man [**28**]. To an eigenfunction ψ he associates a distribution μ_ψ on the unit cotangent bundle S^*Y projecting to $\bar{\mu}_\psi$ on Y. Generalizing the "Wigner function" formalism of statistical physics (see, e.g. [**11**, pp. 58–59] or the original account [**32**]), this construction (the "microlocal lift") proceeds using the theory of pseudo-differential operators and has the property that, for any sequence $\{\psi_n\}_{n=1}^\infty \subset L^2(Y)$ with eigenvalues λ_n tending to infinity, any weak-* limit of the $\mu_n = \mu_{\psi_n}$ is a probability measure on the unit tangent bundle S^*Y, invariant under the geodesic flow. Since any weak-* limit of the μ_n projects to a weak-* limit of the $\bar{\mu}_n$, it suffices to understand these limits; Liouville's measure $d\lambda$ on S^*Y plays here the role of the Riemannian measure on Y.

This construction has a natural interpretation from the point of view of semi-classical physics. The geodesic flow on Y describes the motion of a free particle ("billiard ball"). S^*Y is (essentially) the *phase space* of this system, i.e. the state space of the motion. In this setting one calls a function $g \in C^\infty(S^*Y)$ an *observable*. The state space of the quantum-mechanical billiard is $L^2(Y)$, with the infinitesimal generator of time evolution $-\Delta$. "Observables" here are bounded self-adjoint operators $B\colon L^2(Y) \to L^2(Y)$. Decomposing a state $\psi \in L^2(Y)$ w.r.t. the spectral measure of B gives a probability measure on the spectrum of B (which is the set of possible "outcomes" of the measurement). The expectation value of the "measuring B while the system is in the state ψ" is then given by the matrix element $\langle B\psi, \psi\rangle$. In the particular case where B is a 0th-order pseudo-differential operator with symbol $g \in C^\infty(S^*Y)$, we think of B as a "quantization" of g, and any such a B will be denoted $\mathrm{Op}(g)$.

We can now describe Šnirel′man construction: it is given by $\mu_\psi(g) = \langle \mathrm{Op}(g)\psi, \psi\rangle$. This indeed lifts $\bar{\mu}_\psi$, since for $g \in C^\infty(Y)$ we can take $\mathrm{Op}(g)$ to be multiplication by g. If ψ is taken to be an eigenfunction then, asymptotically, this construction does not depend on the choice of "quantization scheme," that is to say, on the choice of the assignment $g \mapsto \mathrm{Op}(g)$. Indeed, if B_1, B_2 have the same symbol of order 0, and $-\Delta\psi = \lambda\psi$ (i.e. "ψ is an eigenstate of energy λ") then one has $\langle (B_1 - B_2)\psi, \psi\rangle = O(\lambda^{-1/2})$.

On a philosophical level we expect our quantum-mechanical description to approach the classical one at the limit of large energies. We will not formalize this idea (the "correspondence principle"), but depend on it for motivating our main question, whether ergodic properties of the classical system persist in the semi-classical limit of the "quantized" version:

PROBLEM 1.1.1. (Quantum Ergodicity) Let $\{\psi_n\}_{n=1}^\infty \subset L^2(Y)$ be an orthonormal basis consisting of eigenfunctions of the Laplacian.

(1) What measures occur as weak-* limits of the $\{\bar{\mu}_n\}$? In particular, when does $\bar{\mu}_n \xrightarrow[n\to\infty]{\text{wk-*}} d\rho$ hold?
(2) What measures occur as weak-* limits of the $\{\mu_n\}$? In particular, when does $\mu_n \xrightarrow[n\to\infty]{\text{wk-*}} d\lambda$ hold?

DEFINITION 1.1.2. Call a measure μ on S^*Y a (microlocal) *quantum limit* if it is a weak-* limit of a sequence of distributions μ_{ψ_n} associated, via the microlocal lift, to a sequence of eigenfunctions ψ_n with $|\lambda_n| \to \infty$.

In this language, the main problem is classifying the quantum limits of the classical system, perhaps showing that the Liouville measure is the unique quantum limit. As formalized by Zelditch [**35**] (for surfaces of constant negative curvature) and Colin de Verdière [**4**] (for general Y), the best general result known is still:

THEOREM 1.1.3. *(Šnirel′man-Zelditch-Colin de Verdière) Let Y be a compact manifold, $\{\psi_n\}_{n=1}^\infty \subset L^2(Y)$ an orthonormal basis of eigenfunctions of Δ, ordered by increasing eigenvalue. Then:*

(1) $\frac{1}{N}\sum_{n=1}^N \mu_n \xrightarrow[N\to\infty]{\text{wk-*}} d\lambda$ holds with no further assumptions.
(2) Under the additional assumption that the geodesic flow on S^*Y is ergodic, there exists a subsequence $\{n_k\}_{k=1}^\infty$ of density 1 s.t. $\mu_{n_k} \xrightarrow[k\to\infty]{\text{wk-*}} d\lambda$.

COROLLARY. *For this subsequence, $\bar{\mu}_{n_k} \xrightarrow[k\to\infty]{\text{wk-*}} d\rho$.*

It was proved by Hopf [**14**] that the geodesic flow on a manifold of contant negative curvature is ergodic. This was generalized to the case of non-constant negative sectional curvature by Anosov [**1**]. In that situation Rudnick and Sarnak [**25**] conjecture a simple situation:

CONJECTURE 1.1.4. *(Quantum unique ergodicity) Let Y be a compact manifold of strictly negative sectional curvature. Then:*

(1) *(QUE on Y) The $\bar{\mu}_n$ converge weak-* to the Riemannian measure on Y.*
(2) *(QUE on S^*Y) $d\lambda$ is the unique quantum limit on Y.*

We remark that [**25**] also give an example of a hyperbolic 3-manifold Y, a point $P \in Y$, and a sequence of eigenfunctions ψ_n with eigenvalues λ_n, such that $|\psi_n(P)| \gg \lambda_n^{1/4-\epsilon}$. The point P is a fixed point of many Hecke operators, and behaves in a similar fashion to the poles of a surface of revolution. This remarkable phenomenon does not seem to contradict Conjecture 1.1.4. The scarcity of such points and their higher-dimensional analogues will play an important role in the analysis of Chapters 4 and 5.

One difficulty associated with this problem is that of multiplicity of the spectrum. For a negatively curved manifold Y, it is believed that the multiplicities of the Laplacian Δ acting on $L^2(Y)$ are quite small, i.e. the λ-eigenspace has dimension $\ll_\epsilon \lambda^\epsilon$. This question seems extremely difficult even for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, and no better bound is known than the general $O(\lambda^{1/2}/\log(\lambda))$, valid for all negatively curved manifolds. The freedom associated with high degeneracy might allow the construction of "scarred" eigenfunctions which become concentrated on singular subsets of Y.

However, even lacking information on the multiplicities, it transpires that in many natural instances we have a *distinguished basis* for $L^2(Y)$. In that context, it is then natural to ask whether Conjecture 1.1.4 can be resolved with respect to this distinguished basis. Since it is believed that the Δ-multiplicities are small, this modification is, philosophically, not too far from the original question. However, it is in many natural cases far more tractable. The main example is that of *congruence* quotients of symmetric spaces, where the distinguished basis is that of Hecke eigenforms. This is discussed further below, after introducing the important work on surfaces of constant negative curvature.

1.2. Hyperbolic surfaces and automorphic forms

The quantum unique ergodicity question for hyperbolic surfaces has been intensely investigated over the last two decades. We recall some important results.

Zelditch's work [**34, 36**] on the case of compact surfaces Y of constant negative curvature provided a representation-theoretic alternative to the original construction of the microlocal lift via the theory of pseudo-differential operators. It is well-known that the universal cover of such a surface Y is the upper half-plane $\mathbb{H} \simeq \mathrm{PSL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$, so $Y = \Gamma\backslash\mathbb{H}$ for a uniform lattice $\Gamma < G = \mathrm{PSL}_2(\mathbb{R})$. Then the $\mathrm{SO}_2(\mathbb{R}) \simeq S^1$ bundle $X = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R}) \twoheadrightarrow Y$ is isomorphic to the unit cotangent bundle of Y. In this parametrization, the geodesic flow on S^*Y is given by the action of the maximal split torus $A = \left\{ \begin{pmatrix} e^{t/2} & \\ & e^{-t/2} \end{pmatrix} \right\}$ on X from the right. Zelditch's explicit microlocal lift starts with the observation that an eigenfunction ψ_n (considered as a K-invariant function on X) can be thought of as the spherical vector $\varphi_0^{(n)}$ in an irreducible G-subrepresentation of $L^2(X)$. He then constructs another ("generalized") vector in this subrepresentation, a distribution

$\delta^{(n)}$, and shows that the distribution given by $\mu_{\psi_n}(g) = \delta^{(n)}(g\overline{\varphi_0^{(n)}})$ for $g \in C_c^\infty(X)$ agrees (up to terms which decay as the λ_n grow) with the microlocal lift. He then observes that the distribution μ_{ψ_n} is exactly annihilated by a differential operator of the form $H + \frac{J}{r_n}$ where H is the infinitesimal generator of the geodesic flow, J a certain (fixed) second-order differential operator, and $\lambda_n = -\frac{1}{4} - r_n^2$. It is then clear that any weak-* limit taken as $|\lambda_n| \to \infty$ will be annihilated (in the sense of distributions) by the differential operator H, or in other words be invariant under the geodesic flow. Wolpert [**33**] made Zelditch's approach self-contained by showing that the limits are positive measures without using pseudo-differential calculus. One advantage of this approach is that it is based entirely on the right action of $\mathrm{PSL}_2(\mathbb{R})$ on X, and in particular respects structures on X that commute with this action.

When $\Gamma < \mathrm{PSL}_2(\mathbb{R})$ is a so-called *congruence* lattice, there are additional operators acting on functions on $X = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$: for each prime p (except for a finite set of "ramified primes" depending on Γ) there exists an operator $T_p \colon L^2(X) \to L^2(X)$ commuting with the right action of $\mathrm{PSL}_2(\mathbb{R})$. It arises from a $\mathrm{PSL}_2(\mathbb{R})$-equivariant foliation of X into $p+1$-regular graphs (the "Hecke Foliation"; almost all the leafs are trees), and T_p is the graph Laplacian operator on each leaf. In particular T_p also acts on functions on Y and commutes with Δ. These are the *Hecke operators*, and they all commute. The joint eigenfunctions of all the T_p and Δ are called Hecke-Maass forms. They encode considerable arithmetic information and are central objects of study in analytic number theory. They are the prototypical examples of the more general automorphic forms considered below, and will form our distinguished basis.

Much more results are known on the quantum chaos problem for Hecke-Maass forms. One example is the Iwaniec-Sarnak result mentioned above. Of interest to us, a quantum limit μ_∞ arising from micro-local lifts of these eigenfunctions is called an *arithmetic quantum limit*. The arithmetic quantum chaos problem (posed in general below generality) is the classification of such limits.

The study of arithmetic quantum limits started with the seminal result of Rudnick and Sarnak [**25**], that a weak-* limit $\bar{\mu}_\infty$ coming from $\bar{\mu}_{\psi_n}$ attached to Hecke-Maass eigenforms cannot be supported on a finite union of closed geodesics. One way to think of this result is as stating that arithmetic quantum limits cannot be too singular, due to the behaviour of Hecke eigenfunctions along the Hecke foliation: if a Hecke eigenfunction is too large on a piece of the geodesic, it must also be somewhat large at translates of this piece by the Hecke foliation. A clever choice of the prime p (depending on the closed geodesics under consideration) assured that the translates were all disjoint, and a contradiction was obtained to the fact that $\bar{\mu}_n(Y) = 1$.

Using many places at once, Bourgain and Lindenstrauss [**3**] obtained a significantly stronger result: they showed that the μ_∞-measure of an ϵ-neighbourhood of a piece of a geodesic must decay at least as fast as $\epsilon^{2/9}$. In the language of ergodic theory, they have shown that any $a \in A$ acts on every ergodic component of an arithmetic quantum limit μ_∞ with positive entropy.

Building on this result, Lindenstrauss [**19**] proved a theorem classifying A-invariant measures on X satisfying the positive entropy property as well as a "recurrence" property easily satisfied by arithmetic quantum limits: such measures must be proportional to the Haar measure dx. This (almost) answered the arithmetic QUE problem for congruence surfaces:

THEOREM 1.2.1. *(Lindenstrauss) Let $Y = \Gamma\backslash\mathbb{H}$ be a congruence quotient of the hyperbolic plane, and let μ_∞ be an arithmetic quantum limit on $X = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R}) \simeq S^*Y$. Then $\mu_\infty = c \cdot dx$ for some $c \in [0,1]$. If Γ is co-compact (ie arising from a quaternion algebra) then $c = 1$.*

The main theorem of this thesis is a generalization of this theorem to division algebras of degree greater than 2. As the basic strategy of the proof remains the same, we shall record it here:

(1) Start with a sequence of Hecke-Maass forms $\{\psi_n\}_{n=1}^\infty \subset L^2(Y)$ and their associated measures $\{\bar{\mu}_n\}$, converging to a limit measure $\bar{\mu}_\infty$.
(2) Passing to a subsequence, lift them to measures μ_n on the bundle $X \twoheadrightarrow Y$ converging to a limit μ_∞ which is invariant under a subgroup $A < \mathrm{PSL}_2(\mathbb{R})$. The lift is constructed in way which respects the Hecke-eigenform condition.
(3) Using the Hecke correspondence, show that an arithmetic limit μ_∞ cannot be too singular, in that it must have positive entropy w.r.t. the action of elements $a \in A$.
(4) Apply a measure-rigidity theorem to show that $\mu_\infty \propto \mu_{\text{Haar}}$.

We should remark that the special case of congruence surfaces can also be attacked from a different direction. A beautiful formula of Watson [**30**] relates the triple-product integral $\bar{\mu}_n(\psi_m)$ to a special value of an L-function attached to $\psi_n \times \bar{\psi}_n \times \psi_m$. Equidistribution of the $\bar{\mu}_n$ would follow from fast enough decay of this special value, which in turn would follow from an appropriate Generalized Riemann Hypothesis. The rate of decay obtained this way from the GRH is best possible (that was shown in [**22**]). Moreover, conditioned on the GRH the formula permits an evaluation of the normalization of μ_∞ in the non-compact case (the "escape-of-mass" problem) giving $\bar{\mu}_\infty(Y) = 1$ in that case as well. In fact, to show that $c = 1$ it suffices to give a sub-convex bound in the eigenvalue aspect for the Rankin-Selberg L-function $L(\frac{1}{2}, \psi_n \times \bar{\psi}_n)$.

1.3. Quantum unique ergodicity on locally symmetric spaces

Lindenstrauss's clear exposition [**18**] of the Zelditch-Wolpert microlocal lift actually considers the case of $Y = \Gamma\backslash(\mathbb{H} \times \cdots \times \mathbb{H})$ for an irreducible lattice Γ in $\mathrm{PSL}_2(\mathbb{R}) \times \cdots \times \mathrm{PSL}_2(\mathbb{R})$. The natural candidates for ψ_n there are not eigenfunctions of the Laplacian alone, but rather of all the "partial" Laplacians associated to each factor separately. Set now $G = \mathrm{PSL}_2(\mathbb{R})^h$, $K = \mathrm{SO}_2(\mathbb{R})^h$, $X = \Gamma\backslash G$, $Y = \Gamma\backslash G/K$, and take Δ_i to be the Laplacian operator associated with the ith factor (so that $\mathbb{C}[\Delta_1, \ldots, \Delta_h]$ is the ring of K-bi-invariant differential operators on G). Assume that $\Delta_i\psi_n + \lambda_{n,i}\psi_n = 0$, where $\lim_{n\to\infty} \lambda_{n,i} = \infty$ for each $1 \le i \le h$ separately. Generalizing the Zelditch-Wolpert construction, Lindenstrauss obtains distributions $\delta^{(n)}\overline{\varphi_0^{(n)}}$ on X, projecting to $\bar{\mu}_{\psi_n}$on Y, and so that every weak-* limit of these (a "quantum limit") is a finite positive measure invariant under the action of the full maximal split torus A^h.

It is important to note that the lift is to the bundle $X \twoheadrightarrow Y$ which is *not* the unit cotangent bundle of Y, in fact much smaller: it is $3h$-dimensional whereas S^*Y would be $4h-1$-dimensional. Moreover, the limits obtained are invariant by a much larger subgroup (of dimension h) rather than the 1-dimensional geodesic flow of S^*Y. The last fact is not entirely surprising, in that we have assumed that ψ_n are eigenfunctions of a family of h independent commuting differential operators. This phenomenon will repeat with our general representation-theoretic lift below.

Following the construction, Lindenstrauss proposes the following version of QUE, also due to Sarnak:

PROBLEM 1.3.1. (QUE on locally symmetric spaces) Let G be a connected semi-simple Lie group with finite center. Let K be a maximal compact subgroup of G, $\Gamma < G$ a lattice, $X = \Gamma\backslash G$, $Y = \Gamma\backslash G/K$. Let $\{\psi_n\}_{n=1}^{\infty} \subset L^2(Y)$ be a sequence of normalized eigenfunctions of the ring of G-invariant differential operators on G/K, with the eigenvalues w.r.t. the Casimir operator tending to ∞ in absolute value. Is it true that $\bar{\mu}_{\psi_n}$ converge weak-* to the normalized projection of the Haar measure to Y?

We remark that a central character should certainly play no role in this problem, and it is possible to consider instead the case where G is a reductive group, and $\psi_n \in L^2(Y, \omega_n)$ is a sequence of eigenfunctions which transform under unitary central characters $\omega_n \in \hat{Z}$ where Z is the center of G. The measures $\bar{\mu}_{\psi_n}$ are then probability measures on $Y_Z = Z\backslash Y$, since $|\psi_n(y)|^2$ is Z-invariant. We take this point of view from now on. We therefore also will use the notation $X_Z = Z\backslash X$.

Chapter 3 is devoted to showing the first result of this thesis (Theorem 1.3.2 below): the construction of the microlocal lift in this setting. We will impose a mild non-degeneracy condition on the sequence of eigenfunctions (see Section 3.3.2; the assumption essentially amounts to asking that all eigenvalues tend to infinity, at the same rate for operators of the same order.)

With K and G as in Problem 1.3.1, let A be as in the Iwasawa decomposition $G = NAK$, i.e. $A = \exp(\mathfrak{a})$ where $\mathfrak{a}$ is a maximal abelian subspace of $\mathfrak{p}$. (Full definitions are given in Section 2.2). For $G = \mathrm{GL}_n(\mathbb{R})$ and $K = \mathrm{O}_n(\mathbb{R})$, one may take A to be the subgroup of diagonal matrices with positive entries. Let $\pi\colon X_Z \to Y_Z$ be the projection. We denote by dx the G-invariant probability measures on X_Z, and by dy the projection of this measure to Y_Z.

The content of the Theorem that follows amounts, roughly, to a "G-equivariant microlocal lift" on Y. While our definitions have been specific to $\mathrm{GL}_n(\mathbb{R})$, the proof will not make any use of this fact. The theorem holds for any reductive group, with appropriate generalization of the non-degeneracy assumption.

THEOREM 1.3.2. *Let $\psi_n \subset L^2(Y, \omega_n)$ be a non-degenerate sequence of normalized eigenfunctions, whose eigenvalues approach ∞. Then, after replacing ψ_n by an appropriate subsequence, there exist functions $\tilde{\psi}_n \in L^2(X, \omega_n)$ and distributions μ_n on X_Z such that:*

(1) *The projection of μ_n to Y_Z coincides with $\bar{\mu}_n$, i.e. $\pi_* \mu_n = \bar{\mu}_n$.*
(2) *Let σ_n be the measure $|\tilde{\psi}_n(x)|^2 dx$ on X_Z. Then, for every $g \in C_c^\infty(X_Z)$, we have $\lim_{n\to\infty}(\sigma_n(g) - \mu_n(g)) = 0$.*
(3) *Every weak-* limit σ_∞ of the measures σ_n (necessarily a positive measure of mass ≤ 1) is A-invariant.*
(4) *(Equivariance). Let $E \subset \mathrm{End}_G(C^\infty(X_Z))$ be a $\mathbb{C}$-subalgebra of bounded endomorphisms of $C^\infty(X_Z)$, commuting with the G-action. Noting that each $e \in E$ induces an endomorphism of $C^\infty(Y)$, suppose that ψ_n is an eigenfunction for E (i.e. $E\psi_n \subset \mathbb{C}\psi_n$). Then we may choose $\tilde{\psi}_n$ so that $\tilde{\psi}_n$ is an eigenfunction for E with the same eigenvalues as ψ_n, i.e. for all $e \in E$ there exists $\lambda_e \in \mathbb{C}$ such that $e\psi_n = \lambda_e \psi_n, e\tilde{\psi}_n = \lambda_e \tilde{\psi}_n$.*

We first remark that the distributions μ_n (resp. the measures σ_n) generalize the constructions of Zelditch (resp. Wolpert). Although, in view of (2), they carry roughly equivalent information, it is convenient to work with both simultaneously: the distributions μ_n are canonically defined and easier to manipulate algebraically, whereas the measures σ_n are patently positive and are central to the arguments of Chapter 5.

PROOF. In Section 3.3.1 we define the distributions μ_n. (In the language of Definition 3.3.3, we take $\mu_n = \mu_{\psi_n}(\varphi_0, \delta)$).

Claim (1) is established in Lemma 3.3.6.

In Section 3.3.2 we introduce the non-degeneracy condition. Proposition 3.3.13 defines $\tilde{\psi}_n$ and establishes the claims (2) and (4). (Observe that this Proposition establishes (2) only for K-finite test functions g. Since the extension to general g is not necessary for any of our applications, we omit the proof.)

Finally, in section 3.4 we establish claim (3) (Corollary 3.4.9) by finding enough differential operators annihilating μ_n. □

REMARK 1.3.3.

(1) It is important to verify that non-degenerate sequences of eigenfunctions exist. We mostly consider here the case compact quotients X_Z, for which [**7, 6**] show that a positive proportion of the unramified spectrum lies in every open subcone of the Weyl chamber (for definitions see Theorem 3.2.7 and the discussion in Section 3.1). A similar statement for finite-volume *arithmetic* quotients Y should follow from the recent techniques of [**20**]. Earlier, [**23**, Thm. 5.3] has treated the case of $\mathrm{SL}_3(\mathbb{Z})\backslash\mathrm{SL}_3(\mathbb{R})/\mathrm{SO}_3(\mathbb{R})$.
(2) We shall use the phrase *non-degenerate quantum limit* to denote any weak-* limit of σ_n, where notations are as in Theorem 1.3.2. Note that if σ_∞ is such a limit, then claim (2) of the Theorem shows that there exists a subsequence (n_k) of the

integers such that $\sigma_\infty(g) = \lim_{n_k\to\infty} \mu_{n_k}(g)$ for all $g \in C_c^\infty(X_Z)$. Depending on the context, we shall therefore use the notation σ_∞ or μ_∞ for a non-degenerate quantum limit.

(3) It is not necessary to pass to a subsequence in Theorem 1.3.2. See Remark 3.3.12.

(4) It is likely that the A-invariance aspect of Theorem 1.3.2 could be established by standard microlocal methods; however, the equivariance property does not follow readily from these methods and is absolutely crucial for our application. For us the invariance arises from the action of the ring of invariant differential operators, which is a polynomial algebra in r generators where $r = \dim A$.

(5) The measures μ_n all are invariant by the compact group $M = Z_K(\mathfrak{a})$. In fact, Theorem 1.3.2 should strictly be interpreted as lifting measures to X_Z/M rather than X_Z.

(6) Theorem 1.3.2 admits a natural geometric interpretation. Informally, the bundle $X/M \to Y$ may be regarded as a bundle parameterizing maximal flats in Y, and the A-action on X/M corresponds to "translation along flats." We refer to [**27**, Sec. 5.3] for a further discussion of this point.

The existence of the microlocal lift already places a restriction on the possible weak-* limits of the measures $\{\bar{\mu}_n\}$ on Y_Z. For example, the A-invariance of μ_∞ shows that the support of any weak-* limit measure $\bar{\mu}_\infty$ must be a union of maximal flats.

More importantly, Theorem 1.3.2 allows us to pose a new version of the problem:

PROBLEM 1.3.4. (QUE on homogeneous spaces) In the setting of Problem 1.3.1, is the G-invariant measure on X_Z the unique non-degenerate quantum limit?

REMARK 1.3.5. When formulating Conjecture 1.1.4, Rudnick and Sarnak could rely on part (2) of Theorem 1.1.3 to guarantee that the conjectured unique limit is, in fact, a quantum limit. As the geodesic flow on the locally symmetric spaces we consider is also ergodic, this argument extends to the context of Problem 1.3.1 (at least when Y is compact). While our work on the arithmetic case outlined in the next section implies (in certain special cases) the analogous fact for Problem 1.3.4, it is likely that a direct proof is possible. This is especially so in the compact quotient case, when the main problem is of technical nature: showing that most of the spectrum is non-degenerate in our sense. This should follow from the results of [**7**].

1.4. Arithmetic QUE in the higher-rank case

The main result of this thesis is the resolution of Problem 1.3.4 for certain higher rank symmetric spaces, in the context of *arithmetic* quantum limits. We first recall their definition and significance.

As in the special case of congruence quotients of the hyperbolic plane, the situation of having (something close to) a distinguished basis occurs for $Y = \Gamma\backslash G/K$ and $\Gamma \subset G$ a congruence lattice. For almost all primes p there exists a commutative algebra $\mathcal{H}_p$ of operators acting on $L^2(X)$ arising from a discrete foliation. These operators commute with each other and with the G-invariant differential operators. This distinguished basis is obtained by simultaneously diagonalizing the action of the Hecke operators. Precise definitions of the foliation and the Hecke operators in the case under consideration are given in Section 2.3; in any case, we refer to quantum limits arising via the lift from subsequences of Hecke-Maass forms as *arithmetic quantum limits*. A special case of Problem 1.3.4 is then:

CONJECTURE 1.4.1. *(Arithmetic QUE) Let $\psi_n \in L^2(Y, \omega_n)$ be a (non-degenerate ?) sequence of Hecke-Maass eigenforms. Is dy the unique weak-* limit of the $\bar{\mu}_n$? Is the G-invariant measure on X_Z the unique (non-degenerate ?) arithmetic quantum limit?*

In Chapters 4 and 5 we study the properties of arithmetic quantum limits in the case where Γ arises from the multiplicative group of a division algebra of prime degree d over $\mathbb{Q}$. The case $d = 2$ is the theorem of Lindenstrauss discussed above.

For brevity, we state the result in the language of automorphic forms; in particular, $\mathbb{A}$ is the ring of adéles of $\mathbb{Q}$. Detailed discussion of the construction may be found in Chapter 2.

Let $\mathbb{D}/\mathbb{Q}$ be a division algebra of prime degree d, and let $\mathbb{G} = \mathbb{D}^\times$ be the associated general linear group. Assume that $\mathbb{G}$ is split at ∞, ie that $G = \mathbb{G}(\mathbb{R}) \simeq \mathrm{GL}_d(\mathbb{R})$. Let K_f be an open compact subgroup of $\mathbb{G}(\mathbb{A}_\mathrm{f})$ such that $X = \mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})/K_\mathrm{f}$ contains a single G-orbit. Then there exists a discrete subgroup $\Gamma < \mathbb{G}(\mathbb{R})$ such that $X = \Gamma\backslash G$, and Section 2.3 develops a Hecke algebra $\mathcal{H}_{(R)}$ acting on functions on X via Hecke operators at almost all primes. There exists an abundance of open compact subgroups K_f satisfying the condition above. For example, quotients of G by congruence subgroups associated to Eichler orders are of this type (see Lemma 2.3.7 for details).

The subgroup Γ projects to a co-compact lattice in $G/Z \simeq \mathrm{PGL}_d(\mathbb{R})$ where Z is the center of G. As in the previous section we let $X_Z = Z\Gamma\backslash G$ denote the resulting compact homogeneous space of $\mathrm{PGL}_d(\mathbb{R})$, A denote the maximal split torus of diagonal matrices in G, and ω_n denote unitary characters of Z.

The second result of this thesis is:

THEOREM 1.4.2. *Let $\tilde{\psi}_n \in L^2(X, \omega_n)$ be a sequence of $\mathcal{H}_{(R)}$ eigenforms on X such that the associated probability measures σ_n on X_Z converge weak-* to an A-invariant probability measure σ_∞. Then every $a \in A \setminus Z$ acts on every A-ergodic component of σ_∞ with positive entropy.*

PROOF. This is essentially a rephrasing of Theorem 5.0.1, where the uniformity of the estimate means it carries over to weak-* limits. By that theorem we find an $\eta > 0$ such that for any fixed $C \subset M_a A_a$ as defined in Section 4.2 and small enough ϵ we have for all $x \in X_Z$ that $\sigma_\infty(xB(C, \epsilon)) \ll \epsilon^\eta$. For a proof that this bound implies that a acts with positive entropy see [**17**, Sec. 8]. While written for the case of quaternion algebras ($d = 2$), that discussion readily generalizes to our situation by modifying its “Step 2” to account for the action of a on the Lie algebra – compare our Section 4.2 and the definitions at the start of [**17**, Sec. 7]. $\square$

REMARK 1.4.3. The statement of Theorem 5.0.1 gives a direct bound on singular behaviour of the σ_n. Its proof follows the ideas of Rudnick-Sarnak and Bourgain-Lindenstrauss: translating the set $xB(C,\epsilon)$, which is an ϵ-neighbourhood of a piece C of a "generalized geodesic" (Levi subgroup) by the Hecke correspondence at many places we show that it must have small σ_n-measure.

1.5. The Main Theorem

Following the strategy proposed above, we now state and prove the main result of this thesis:

THEOREM 1.5.1. *Let $Y_Z \simeq \Gamma\backslash\mathrm{PGL}_d(\mathbb{R})/\mathrm{SO}_d(\mathbb{R})$ be a compact locally symmetric space, where the lattice Γ is associated to an Eichler order in a division algebra of the prime degree d over $\mathbb{Q}$, split over $\mathbb{R}$. Let $\{\psi_n\}_{n=1}^{\infty} \subset L^2(Y_Z)$ be a non-degenerate sequence of Maass forms which are also eigenforms of the Hecke algebra $\mathcal{H}_{(R)}$ of Section 2.3. Then the associated probability measures $\bar{\mu}_n$ converge weak-* to the normalized Haar measure on Y, as their lifts μ_n converge weak-* to the normalized Haar measure dx on $X_Z = \Gamma\backslash\mathrm{PGL}_d(\mathbb{R})$.*

In other words, then the normalized Haar measure is the unique non-degenerate arithmetic quantum limit in this case.

PROOF. In fact, the proof generalizes to the case where $\psi_n \in L^2(Y,\omega_n)$ for central characters ω_n. The case $d = 2$ is Lindenstrauss's Theorem quoted as Theorem 1.2.1 above, and we will thus assume $d \geq 3$. Passing to a subsequence, let $\psi_n \in L^2(Y,\omega_n)$ be a non-degenerate sequence of Hecke-Maass forms on Y such that $\bar{\mu}_n \to \bar{\mu}_\infty$ weakly. Passing to a subsequence, let $\tilde{\psi}_n$ and σ_n be as in Theorem 1.3.2 such that $\sigma_n \to \sigma_\infty$ weakly and σ_∞ lifts $\bar{\mu}_\infty$. Then σ_∞ is a non-degenerate arithmetic quantum limit on X_Z. By Theorem 1.4.2, σ_∞ is an A-invariant probability measure on X_Z such that every $a \in A \setminus Z$ acts on every A-ergodic component of σ_∞ with positive entropy. Then [**9**, Th. 4.1(iv)] shows that σ_∞ has a unique ergodic component, μ_{Haar}. □

REMARK 1.5.2.

(1) The assumption that Γ is associated to an Eichler order is of technical nature. The result certainly holds for Hecke eigenfunctions on an adelic double-coset space $\tilde{X} = \mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})/K_{\mathrm{f}}$ where $\mathbb{G}$ is the group of invertible elements of a $\mathbb{Q}$-division algebra which is $\mathbb{R}$-split. In general, however, such a space is a disjoint union of several components of the form $X = \Gamma\backslash G$ where Γ is a congruence subgroup, and we would like to consider eigenfunctions on the components themselves. It is not clear, whoever, whether we can form a sufficiently large explicit Hecke algebra acting on such a component. For this one is interested in the set of primes p such that each leaf of the p-Hecke foliation (defined in Section 2.3) is contained in a single component X of $\tilde{X}$.

(2) In all likelihood it is possible to obtain a version of Theorem 1.3.2 for degenerate sequences as well. The resulting quantum limits μ_∞ would be invariant under subtori $A_1 < A$ depending on the degeneracy of the limit parameter $\tilde{\nu}_\infty$.

A measure rigidity theorem generalizing [**19**], requiring only invariance under a one-parameter subgroup, positive entropy and recurrence would then allow us to drop the non-degeneracy assumption and resolve the AQUE problem for division algebras of prime degree $d \geq 3$.

(3) We expect the techniques developed for the proof of Theorem 1.5.1 will generalize at least to some other locally symmetric spaces, the case of $\mathbb{D}$ being the simplest; but there are considerable obstacles to obtaining a theorem for *any* arithmetic locally symmetric space at present. For a general $\mathbb{G}$, the results of Chapter 4 can be generalized to show that the intersections will be controlled by a proper $\mathbb{Q}$-subgroup. However, this subgroup can be quite large, making the analysis on the building much more difficult even when $\mathbb{G}(\mathbb{Q}_p) \simeq \mathrm{GL}_d(\mathbb{Q}_p)$. Moving from the building of GL_d to buildings of other types might present difficulties of its own.

(4) It is also possible to prove results for the case where $\mathbb{G}$ is split, i.e. isomorphic to GL_d over $\mathbb{Q}$. The proof is essentially the same except that the measure rigidity results of [**10**] are used instead. Since in that case the quotient is not compact this does not address the escape-of-mass question. Somewhat surprisingly, however, the normalization of the measure is already controlled by the degenerate Eisenstein series. Hence a sub-convexity result for the Rankin-Selberg L-function would control the escape as in the case of GL_2.

CHAPTER 2

Notation and Fundamentals

We define here standard notation and recall basic facts about division algebras over the rationals and the real, p-adic and adelic Lie groups associated to them.

2.1. Division Algebras

2.1.1. Central simple algebras and their general linear groups. Let K be an infinite field, $\mathbb{D}(K)/K$ a finite-dimensional central simple K-algebra, i.e. a K-algebra with no two-sided ideals and center K. Then for any field L/K, $\mathbb{D}(L) \overset{\text{def}}{=} \mathbb{D}(K) \otimes_K L$ is a central simple L-algebra of the same dimension. It is easy to see ([**31**, Prop. IX-1-2]) that such an algebra must be of the form $\mathbb{D}(K) \simeq M_n(H)$ where H is a central division algebra over K. In particular, $\dim_K \mathbb{D}(K) = n^2 \dim_K H$.

By the Cayley-Hamilton theorem, every element of $\mathbb{D}(L)$ is algebraic over L. In particular, if L/K is algebraically closed then L is the unique division algebra over L, and hence $\mathbb{D}(L) \simeq M_d(L)$ for some $d \geq 1$. We then have:

$$\dim_K \mathbb{D}(K) = \dim_L \mathbb{D}(L) = d^2.$$

In particular, the number d only depends on $\mathbb{D}(K)$ and is called the *degree* of $\mathbb{D}(K)$. It also follows that the dimension of every central simple K-algebra is a square. An field L/K for which $\mathbb{D}(L) \simeq M_d(L)$ is said to *split* $\mathbb{D}(K)$. Alternatively, we say that $\mathbb{D}(K)$ *splits over* L.

Fixing a linear basis $\{u_i\}_{i=1}^{d^2} \subset \mathbb{D}(K)$, we note that it is also a basis over L of $\mathbb{D}(L)$ for any L/K. We can then write any $x \in \mathbb{D}(L)$ uniquely in the form $\sum_{i=1}^{d^2} x_i u_i$. Working in this co-ordinate system $(x, y) \mapsto (x \cdot y)_i$ is then a bilinear map $K^{d^2} \times K^{d^2} \to K$ and hence there exist $a_{ijk} \in K$ such that

$$\left(\sum_{j=1}^{d^2} x_j u_j\right)\left(\sum_{k=1}^{d^2} y_k u_k\right) = \sum_{i=1}^{d^2}\left(\sum_{j,k=1}^{d^2} a_{ijk} x_i y_j\right) u_i.$$

NOTATION 2.1.1. We will use the notation $\{x_i\}_{i=1}^{d^2}$ to denote the co-ordinates of any $x \in \mathbb{D}(L)$ w.r.t. our basis, $\{x_i(g)\}_{i=1}^{d^2}$ for the co-ordinates of $g \in \mathbb{D}^\times(L)$.

We remark that an L-automorphism of $M_n(L)$ is given by a change of basis, i.e. by conjugation by an element of $\mathrm{GL}_n(L)$. It follows that if A is an L-algebra isomorphic to $M_n(L)$ then the pullback of the map $\det\colon M_n(L) \to L$ to A is well-defined independently of the choice of isomorphism.

FACT 2.1.2. [**31**, Prop. XI-2-6]

(1) *There exists a map* $\nu_{\mathbb{D}(K)} \colon \mathbb{D}(K) \to K$ *such that for any* L/K *where* $\mathbb{D}$ *splits we have* $\det \restriction_{\mathbb{D}(K)} = \nu_{\mathbb{D}(K)}$.
(2) *There exists a polynomial* $\nu_{\mathbb{D}} \in K[x_1, \ldots, x_{d^2}]$ *of degree* d *depending on the choice of basis such that* $\nu_{\mathbb{D}(L)}(x) = \nu_{\mathbb{D}}(x_1, \ldots, x_{d^2})$ *for any* L/K *(not necessarily split) and any* $x \in \mathbb{D}(L)$. *Here* $\nu_{\mathbb{D}(L)} \colon \mathbb{D}(L) \to L$ *is the map constructed in the first part.*
(3) *The maps* $\nu_{\mathbb{D}(K)}$ *and* $\nu_{\mathbb{D}}$ *are both known as the* reduced norm.

Passing to a split extension shows that for any L/K, $x \in \mathbb{D}(L)$ is invertible iff $\nu_{\mathbb{D}}(x) \neq 0$, in which case it is possible to compute the co-ordinates of x^{-1} by polynomial functions of its co-ordinates $\{x_i\}$ and $\nu_{\mathbb{D}}(x)^{-1}$. We now identify $\mathbb{G} \stackrel{\text{def}}{=} \mathbb{D}^\times$ with the set of solutions to $\nu_{\mathbb{D}}(x_1, \ldots, x_{d^2}) \cdot x_0 = 1$ in $d^2 + 1$-dimensional affine space. Since the multiplication operation in $\mathbb{D}$ is also polynomial in the co-ordinates (and $\nu_{\mathbb{D}}(xx') = \nu_{\mathbb{D}}(x)\nu_{\mathbb{D}}(x')$) this makes $\mathbb{G}$ into a linear algebraic group defined over K, with the maps $x_i \colon \mathbb{G} \to \mathbb{A}^1$ all algebraic and defined over K. We will thus supplement our previous notation by using $x_0(g) \in L^\times$ to denote the inverse of the reduced norm of $g \in \mathbb{D}^\times(L)$.

The center of $\mathbb{G}$ is precisely the invertible elements of the center of $\mathbb{D}$, i.e. the invertible elements of the ground field, and we set $\mathbb{G}^{\text{ad}} = \mathbb{G}/Z_{\mathbb{G}}$. We also set $\mathbb{G}^1 = \{g \mid \nu_{\mathbb{D}}(g) = 1\}$. This is a Zariski-closed subgroup.

If $\bar{K}$ is an algebraic closure of K then we have $\mathbb{D}(\bar{K}) \simeq M_d(\bar{K})$. It is then clear that $\mathbb{G}(\bar{K}) \simeq \mathrm{GL}_d(\bar{K})$, $\mathbb{G}^{\text{ad}}(\bar{K}) \simeq \mathrm{PGL}_d(\bar{K})$ and $\mathbb{G}^1(\bar{K}) \simeq \mathrm{SL}_d(\bar{K})$. In particular, the last isomorphism shows that $\mathbb{G}^1$ is simply connected as an algebraic group.

2.1.2. Algebras over local and global fields. Let D_p be a central simple algebra over the field $\mathbb{Q}_p$. An *order* $\mathcal{O}_p \subset D_p$ is a finitely-generated $\mathbb{Z}_p$-subalgebra which spans D_p. Equivalently, it is a compact open $\mathbb{Z}_p$-subalgebra.

Now let $\mathbb{D}$ be a central simple algebra over $\mathbb{Q}$. An *order* is a finitely-generated $\mathbb{Z}$-subalgebra $\mathcal{O} \subset \mathbb{D}(\mathbb{Q})$ which contains a basis for $\mathbb{D}(\mathbb{Q})$ over $\mathbb{Q}$. An order is *maximal* if it is not properly contained in another order. These exist (e.g. by Zorn's Lemma) and we choose a maximal order $\mathcal{O} \subset \mathbb{D}(\mathbb{Q})$. It is a torsion-free abelian group of rank d^2. We can thus fix a $\mathbb{Z}$-basis $\{u_i\}_{i=1}^{d^2} \subset \mathcal{O}$ once and for all. The structure coefficients a_{ijk} with respect to this basis then all lie in $\mathbb{Z}$.

For a place $v \in |\mathbb{Q}|$, the local field $\mathbb{Q}_v$ is an extension of $\mathbb{Q}$ and we set $D_v = \mathbb{D}(\mathbb{Q}_v) = \mathbb{D}(\mathbb{Q}) \otimes_{\mathbb{Q}} \mathbb{Q}_v$ and $G_v = \mathbb{G}(\mathbb{Q}_v) = D_v^\times$. In particular we denote $G = G_\infty$. We say that $\mathbb{D}$ *splits at* $v \in |\mathbb{Q}|$ if it splits over $\mathbb{Q}_v$.

FACT 2.1.3. *(for proofs, see*[**31**]*) For a finite prime* p *of* $\mathbb{Q}$ *let* $\mathcal{O}_p$ *denote the (topological) closure of* $\mathcal{O}$ *in* $\mathbb{D}(\mathbb{Q}_p)$.

(1) $\nu_{D_p}(D_p^\times) = \mathbb{Q}_p^\times$ *and hence* $\nu_{D_p}(\mathcal{O}_p^\times) = \mathbb{Z}_p^\times$.
(2) *relatively compact multiplicatively closed subset* $T \subset D_p$ *is contained in a maximal order. In particular* $K < G_p$ *is maximal compact iff* $K = R_p^\times$ *for a maximal order* $R_p \subset \mathbb{D}(\mathbb{Q}_p)$.

(3) *$\mathcal{O}_p$ is a maximal order in $\mathbb{D}(\mathbb{Q}_p)$, in particular a maximal compact subring; the maximal orders of $\mathbb{D}(\mathbb{Q}_p)$ are all conjugate; $M_d(\mathbb{Z}_p)$ is a maximal order of $M_d(\mathbb{Q}_p)$.*
(4) *For almost all p we have $\mathbb{D}(\mathbb{Q}_p) \simeq M_d(\mathbb{Q}_p)$. We then have $G_p \simeq \mathrm{GL}_d(\mathbb{Q}_p)$ can fix an isomorphism $\varphi_p \colon G_p \to \mathrm{GL}_d(\mathbb{Q}_p)$ such that $\varphi_p(\mathcal{O}_p^\times) = \mathrm{GL}_d(\mathbb{Z}_p)$.*
(5) *We have $\mathcal{O}_p = \oplus_{i=1}^{d^2} \mathbb{Z}_p u_i$. In particular $x_i(g) \in \mathbb{Z}_p$ for every $g \in \mathcal{O}_p^\times$.*

Following claim (3) we fix the maximal compact subgroups $K_p = \mathcal{O}_p^\times$ of G_p. The first part of claim (4) is that $\mathbb{D}$ splits at almost all places. We let R_0 denote the set of finite places were $\mathbb{D}$ does *not* split.

2.1.3. Division algebras of prime degree. We now make the assumption that $\mathbb{D}/\mathbb{Q}$ is a division algebra and that its degree d is a prime and at least 3. If $K/\mathbb{Q}$ is a field extension, then $\mathbb{D}(K)$ is a central simple K-algebra, hence a matrix algebra over a central division H/K. We then have

$$d^2 = \dim_K \mathbb{D}(K) = \dim_H \mathbb{D}(K) \cdot \dim_K H.$$

As d is prime and $\dim_H \mathbb{D}(K)$ and $\dim_K H$ are both squares, there are two possibilities. If $H = \mathbb{D}(K)$ (i.e. $\mathbb{D}(K)$ is also a division algebra), we say that $\mathbb{D}$ *ramifies* over K. Otherwise we have $H = K$, that is $\mathbb{D}$ splits over K.

If $\mathbb{D}$ ramifies at $v \in |\mathbb{Q}|$ then G_v^{ad} is compact. Since $\mathbb{R}$ itself and Hamilton's quaternions $\mathbb{H}$ are the unique central division algebras over $\mathbb{R} = \mathbb{Q}_\infty$, we see that $\mathbb{D}$ can ramify at ∞ only if $d = 2$, which does not hold by assumption. Denoting $G = G_\infty$, this amounts to saying that $G \simeq \mathrm{GL}_d(\mathbb{R})$.

2.2. The Real Group

We conform to the notation of [**16**].

We are considering the group $G \simeq GL_d(\mathbb{R})$, obtained as $\mathbb{D}(\mathbb{R})^\times$ where $\mathbb{D}/\mathbb{Q}$ is a division algebra of prime degree, split at ∞. We choose the Cartan involution $\Theta(g) = \{\}^t\, g^{-1}$ for G, and let $K = \mathrm{O}_d(\mathbb{R})$ be the Θ-fixed maximal compact subgroup, $Z \simeq \mathbb{R}^\times$ the center of G. Let $S = Z\backslash G/K$ be the symmetric space, with $x_K \in S$ the point with stabilizer KZ. We fix a G-invariant metric on S. To normalize it, we observe that the tangent space at the point $x_K \in S$ is identified with $\mathfrak{p}/\mathfrak{z}$ (see below), and we endow it with the Killing form:

Let $\mathfrak{g} = \mathrm{Lie}(G) \simeq M_d(\mathbb{R})$, and let $\theta(X) = -X^t$ denote the differential of Θ, giving the Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ with $\mathfrak{k} = \mathrm{Lie}(K)$ (the anti-symmetric matrices) and $\mathfrak{p}$ the symmetric matrices. The pairing $\langle X, Y \rangle = \mathrm{Tr}(XY^t) - \frac{1}{d} \mathrm{Tr}(X)\,\mathrm{Tr}(Y)$ is $\mathrm{Ad}\,G$-equivariant and positive semi-definite (positive definite on $\mathfrak{g}_{\mathrm{ss}} = [\mathfrak{g}, \mathfrak{g}]$, the subalgebra of matrices of trace 0). Its isotropic subspace is precisely the center $\mathrm{Lie}(Z) = Z_{\mathrm{Lie}(G)}$, where Z is the connected component of the center (in general we would take Z to be the split component of the torus $Z_{\mathbb{G}}(\mathbb{R})$). We fix a maximal abelian subalgebra $\mathfrak{a} \subset \mathfrak{p}$, the subalgebra of diagonal matrices isomorphic to $\mathbb{R}^d$.

We denote by $\mathfrak{a}_{\mathbb{C}}$ the complexification $\mathfrak{a} \otimes_{\mathbb{R}} \mathbb{C}$; we shall occasionally write $\mathfrak{a}_{\mathbb{R}}$ for $\mathfrak{a}$ for emphasis in some contexts. We denote by $\mathfrak{a}^*$ (resp. $\mathfrak{a}^*_{\mathbb{C}}$) the real dual (resp. the complex dual) of $\mathfrak{a}$; again, we shall occasionally write $\mathfrak{a}^*_{\mathbb{R}}$ for $\mathfrak{a}^*$. For $\nu \in \mathfrak{a}^*_{\mathbb{C}}$, we define $\mathrm{Re}(\nu), \mathrm{Im}(\nu) \in \mathfrak{a}^*_{\mathbb{R}}$ to be the real and imaginary parts of ν, respectively.

For $\alpha \in \mathfrak{a}^*$ set $\mathfrak{g}_\alpha = \{X \in \mathfrak{g} \mid \forall H \in \mathfrak{a} : \mathrm{ad}(H)X = \alpha(H)X\}$,

$$\Delta(\mathfrak{a}\colon\mathfrak{g}) = \{\alpha \in \mathfrak{a}^* \setminus \{0\} \mid \mathfrak{g}_\alpha \neq \{0\}\}$$

and call the latter the (restricted) *roots* of $\mathfrak{g}$ w.r.t. $\mathfrak{a}$. The subalgebra $\mathfrak{g}_0$ is θ invariant, and hence $\mathfrak{g}_0 = (\mathfrak{g}_0 \cap \mathfrak{p}) \oplus (\mathfrak{g}_0 \cap \mathfrak{k})$. By the maximality of $\mathfrak{a}$ in $\mathfrak{p}$, we must then have $\mathfrak{g}_0 = \mathfrak{a} \oplus \mathfrak{m}$ where $\mathfrak{m} = Z_{\mathfrak{k}}(\mathfrak{a})$ (here $\mathfrak{m} = \{0\}$). We have $\Delta(\mathfrak{a}\colon\mathfrak{g}) = \{\alpha_{ij}\}_{i \neq j \leq d}$ where $\alpha_{ij}(H) = H_{ii} - H_{jj}$. The root subspaces are $\mathfrak{g}_{ij} = \mathbb{R} \cdot E^{ij}$.

The Killing form also induces a natural pairing $\langle \cdot, \cdot \rangle$ on $\mathfrak{a}^*$ w.r.t. which $\Delta(\mathfrak{a}\colon\mathfrak{g}) \subset \mathfrak{a}^*$ is a root system. The associated Weyl group, generated by the root reflections s_α, will be denoted $W(\mathfrak{a}\colon\mathfrak{g})$. s_{ij} acts on $\mathbb{R}^d$ by exchanging the ith and jth co-ordinate, so that $W(\mathfrak{a}\colon\mathfrak{g}) \simeq S_d$. This group is also canonically isomorphic to the analytic Weyl groups $N_G(A)/Z_G(A)$ and $N_K(A)/Z_K(A)$, where a set of representatives is given by the permutation matrices. The fixed-point set of any s_α is a hyperplane in $\mathfrak{a}^*$, called a *wall*. The connected components of the complement of the union of the walls are cones, called the (open) *Weyl chambers*. A subset $\Pi \subset \Delta(\mathfrak{a}\colon\mathfrak{g})$ will be called a *system of simple roots* (by abuse of notation a "simple system") if every root can be uniquely expressed as an integral combination of elements of Π with either all coefficients non-negative or all coefficients non-positive. For a simple system Π, the open cone $C_\Pi = \{\nu \in \mathfrak{a}^* \mid \forall \alpha \in \Pi : \langle \nu, \alpha \rangle > 0\}$ is an open Weyl chamber, and the map $\Pi \mapsto C_\Pi$ is a 1-1 correspondence between simple systems and chambers. The Weyl group acts simply transitively on the chambers and simple systems. The closure of an open chamber will be called a closed chamber. The action of $W(\mathfrak{a}\colon\mathfrak{g})$ on $\mathfrak{a}^*$ extends in the complex-linear way to an action on $\mathfrak{a}^*_{\mathbb{C}}$ preserving $i\mathfrak{a}^* \subset \mathfrak{a}^*_{\mathbb{C}}$, and we call an element $\nu \in \mathfrak{a}^*_{\mathbb{C}}$ *regular* if it is fixed by no $w \in W(\mathfrak{a}\colon\mathfrak{g})$. We use $\rho = \frac{1}{2}\sum_{\alpha>0}(\dim \mathfrak{g}_\alpha)\alpha \in \mathfrak{a}^*$ to denote half the sum of the positive (restricted) roots.

Fixing the simple system $\Pi = \{e_{i,i+1}\}_{i=1}^{d-1}$ we get a notion of positivity. For $\mathfrak{n} = \oplus_{\alpha>0}\mathfrak{g}_\alpha$ (strictly upper-triangular matrices) and $\bar{\mathfrak{n}} = \Theta\mathfrak{n}$ we have $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{m} \oplus \bar{\mathfrak{n}}$ and (Iwasawa decomposition) $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$. By means of the Iwasawa decomposition, we may uniquely write every $X \in \mathfrak{g}$ in the form $X = X_{\mathfrak{n}} + X_{\mathfrak{a}} + X_{\mathfrak{k}}$. We sometimes also write $H_0(X)$ for $X_{\mathfrak{a}}$.

Let N, A be the subgroups of G corresponding to the subalgebras $\mathfrak{n}$, $\mathfrak{a} \subset \mathfrak{g}$ respectively (upper-triangular unipotent matrices and diagonal matrices with positive entries, respectively), and let $M = Z_K(\mathfrak{a})$ (diagonal matrices with entries in $\{\pm 1\}$). Then A is a maximal split torus in G, and $\mathfrak{m} = \mathrm{Lie}(M)$, though M is not necessarily connected. Moreover $P_0 = NAM$ is a minimal parabolic subgroup of G, with the map $N \times A \times M \to P_0$ being a diffeomorphism. The map $N \times A \times K \to G$ is a (surjective) diffeomorphism (Iwasawa decomposition), so for $g \in G$ there exists a unique $H_0(g) \in \mathfrak{a}$ such that $g = n\exp(H_0(g))k$ for some $n \in N$, $k \in K$. The map $H_0 : G \to \mathfrak{a}$ is continuous; restricted to A it is the inverse of the exponential map.

Let $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$ denote the complexification of $\mathfrak{g}$. It is a complex semi-simple Lie algebra. Let $\theta_{\mathbb{C}}$ denote the *complex-linear* extension of θ to $\mathfrak{g}_{\mathbb{C}}$. It is *not* a Cartan involution of $\mathfrak{g}_{\mathbb{C}}$. We fix a maximal abelian subalgebra $\mathfrak{b} \subset \mathfrak{m}$ and set $\mathfrak{h} = \mathfrak{a} \oplus \mathfrak{b}$. Then $\mathfrak{h}_{\mathbb{C}} = \mathfrak{h} \otimes \mathbb{C} \subset \mathfrak{g}_{\mathbb{C}}$ is a Cartan subalgebra, with the associated root system $\Delta(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$ satisfying $\Delta(\mathfrak{a} : \mathfrak{g}) = \{\alpha\restriction_{\mathfrak{a}}\}_{\alpha \in \Delta(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})} \setminus \{0\}$. Moreover, we can find a system of simple roots $\Pi_{\mathbb{C}} \subset \Delta(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$ and a system of simple roots $\Pi \subset \Delta(\mathfrak{a} : \mathfrak{g})$ such that the positive roots w.r.t. Π are precisely the nonzero restrictions of the positive roots w.r.t. $\Pi_{\mathbb{C}}$. We fix such a compatible pair of simple systems, and let $\rho_{\mathfrak{h}}$ denote half the sum of the roots in $\Delta(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$, positive w.r.t. $\Pi_{\mathbb{C}}$.

Let $F_0 \subset \Delta(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$ consist of the roots that restrict to 0 on $\mathfrak{a}$, $F_0^+ \subset F_0$ those positive w.r.t. $\Pi_{\mathbb{C}}$. Let $\mathfrak{n}_M = \oplus_{\alpha \in F_0^+} (\mathfrak{g}_{\mathbb{C}})_\alpha$, $\bar{\mathfrak{n}}_M = \oplus_{\alpha \in F_0^+} (\mathfrak{g}_{\mathbb{C}})_{-\alpha}$. Then $\mathfrak{m}_{\mathbb{C}} = \mathfrak{n}_M \oplus \mathfrak{b}_{\mathbb{C}} \oplus \bar{\mathfrak{n}}_M$ and $\mathfrak{g}_{\mathbb{C}} = \mathfrak{n}_{\mathbb{C}} \oplus \mathfrak{n}_M \oplus \mathfrak{h}_{\mathbb{C}} \oplus \bar{\mathfrak{n}}_M \oplus \bar{\mathfrak{n}}_{\mathbb{C}}$.

For $\nu \in \mathfrak{a}_{\mathbb{C}}^*$, set $\|\nu\|^2 = \langle \mathrm{Re}(\nu), \mathrm{Re}(\nu)\rangle + \langle \mathrm{Im}(\nu), \mathrm{Im}(\nu)\rangle$ (with the inner products taken in $\mathfrak{a}_{\mathbb{R}}^*$).

If $\mathfrak{l}_{\mathbb{C}}$ is a complex Lie algebra, then we denote by $U(\mathfrak{l}_{\mathbb{C}})$ its universal enveloping algebra, and by $\mathfrak{Z}(\mathfrak{l}_{\mathbb{C}})$ its center. In particular we set $\mathfrak{Z} = \mathfrak{Z}(\mathfrak{g}_{\mathbb{C}})$.

2.3. The Adelic group and its quotients, Components

Let $\mathbb{G}(\mathbb{A}_{\mathrm{f}})$ denote the subgroup of the Cartesian product $\prod_p G_p$ consisting of those sequences g such that $g_p \in K_p$ for almost all p. Declaring $K = \prod_p K_p$ (in the product topology) to be an open compact subgroup of $\mathbb{G}(\mathbb{A}_{\mathrm{f}})$ makes $\mathbb{G}(\mathbb{A}_{\mathrm{f}})$ into a totally disconnected locally compact topological group. Finally, we set $\mathbb{G}(\mathbb{A}) = G_\infty \cdot \mathbb{G}(\mathbb{A}_{\mathrm{f}})$. This is a locally compact group. This construction is called a *restricted direct product* and denoted:

$$\mathbb{G}(\mathbb{A}) = G_\infty \cdot \prod_p{}' (G_p, K_p).$$

For $g \in \mathbb{G}(\mathbb{A})$ (or $g \in \mathbb{G}(\mathbb{A}_{\mathrm{f}})$) we denote g_v (resp. g_p) its components at specific places.

REMARK 2.3.1. In the alternate one can define $\mathbb{G}(\mathbb{A})$ as the $\mathbb{A}$-points of the variety $\nu_{\mathbb{D}}(g) \cdot x_0(g) = 1$, with the topology as a subset of $\mathbb{A}^{d^2+1}$ (in both senses). This shows that the group obtained by the restricted direct product procedure above is independent of the choice of the maximal order $\mathcal{O}$.

LEMMA 2.3.2. *Let K_{f} be an open compact subgroup of $\mathbb{G}(\mathbb{A}_{\mathrm{f}})$. Then there exists a finite set R_1 of finite places and an open compact subgroup $K_{R_1} < \prod_{p \in R_1} G_p$ such that $K_{\mathrm{f}} = K_{R_1} \times \prod_{p \notin R_1} K_p$.*

PROOF. A set of basic neighbourhoods of the identity in $\mathbb{G}(\mathbb{A}_{\mathrm{f}})$ is given by the sets of the form $\prod_p U_p$ where $U_p \subset G_p$ are open and $U_p = K_p$ for almost all p. Since K_{f} is open we conclude that there exists a finite set R_1 of finite places such that $\prod_{p \notin R_1} K_p$ is contained in K_{f}. For any $p \notin R_1$ the projection map $\mathbb{G}(\mathbb{A}_{\mathrm{f}}) \to G_p$ is continuous and the image of K_{f} under this map is a compact subgroup containing the maximal compact subgroup K_p. It follows that K_{f} is an open compact subgroup of $\prod_{p \in R_1} G_p \times \prod_{p \notin R_1} K_p$ which contains the second factor. It is hence obviously of the form $K_{R_1} \times \prod_{p \notin R_1} K_p$ for a subgroup $K_{R_1} < \prod_{p \in R_1} G_p$. This subgroup equals the image of K_{f} under a quotient map and in particular is compact and open. $\square$

Letting $R = R_0 \cup R_1$ we set $K_R = K_{R_1} \times \prod_{p \in R_0 \setminus R_1} K_p$ so that $K_\mathrm{f} = K_R \times \prod_{p \notin R} K_p$. In that case we have for every $p \notin R$ an isomorphism $\varphi_p \colon G_p \to \mathrm{GL}_d(\mathbb{Q}_p)$ such that $\varphi(K_p) = \mathrm{GL}_d(\mathbb{Z}_p)$.

FACT 2.3.3. *Let $\mathbb{G}$ be a linear algebraic group. We identify $\mathbb{G}(\mathbb{Q})$ with its image in the Cartesian product $\prod_{v \in |\mathbb{Q}|} G_v$ under the diagonal embedding, and let $K_\mathrm{f} < \mathbb{G}(\mathbb{A}_\mathrm{f})$ be an open compact subgroup.*

(1) *$\mathbb{G}(\mathbb{Q})$ lies in $\mathbb{G}(\mathbb{A})$. It is discrete in the adelic topology of $\mathbb{G}(\mathbb{A})$.*
(2) *(finiteness of class number) The space $\tilde{X} = \mathbb{G}(\mathbb{Q}) \backslash \mathbb{G}(\mathbb{A}) / K_\mathrm{f}$ has finitely many connected components.*

In certain cases we can say more about the space described in claim (2):

FACT 2.3.4. *Returning to our previous notation, let $\mathbb{G}$ be the group of invertible elements of a division algebra defined over $\mathbb{Q}$, $\mathbb{G}^1$ the group of elements of norm 1. Then:*

(1) *The quotient $\mathbb{G}(\mathbb{Q}) Z_\mathbb{G}(\mathbb{A}) \backslash \mathbb{G}(\mathbb{A})$ is compact.*
(2) *$\mathbb{G}^1(\mathbb{Q})$ is dense in $\mathbb{G}^1(\mathbb{A}_\mathrm{f})$. Equivalently (see below), $\mathbb{G}^1(\mathbb{Q}) \backslash \mathbb{G}^1(\mathbb{A}) / K_\mathrm{f}^1$ is connected for any open compact $K_\mathrm{f} < \mathbb{G}^1(\mathbb{A}_\mathrm{f})$.*

DEFINITION 2.3.5. A *component* of $\tilde{X}$ is a G-orbit $X = \tilde{x} G$ for some $\tilde{x} \in \tilde{X}$.

Since G is not connected the 'components' we have just defined do not coincide with the topological connected components of $\tilde{X}$. These are closely related concepts though:

LEMMA 2.3.6. *$\tilde{X}$ is the disjoint union of its components, of which there is a finite number. Furthermore, let X be a component of $\tilde{X}$. Then:*

(1) *X is a union of connected components of $\tilde{X}$.*
(2) *There exists a discrete subgroup $\Gamma < G$ such that $X \simeq \Gamma \backslash G$.*
(3) *$X_Z = Z \backslash X$ is compact. Hence $\Gamma Z / Z$ is a co-compact lattice in $G^{ad} = G/Z$.*

PROOF. As K_f is open in $\mathbb{G}(\mathbb{A}_\mathrm{f})$, the quotient $\mathbb{G}(\mathbb{A}_\mathrm{f}) / K_\mathrm{f}$ is discrete. The space of components

$$\tilde{X}/G = \mathbb{G}(\mathbb{Q}) \backslash \mathbb{G}(\mathbb{A}) / G \cdot K_\mathrm{f} = \mathbb{G}(\mathbb{Q}) \backslash \mathbb{G}(\mathbb{A}_\mathrm{f}) / K_\mathrm{f}$$

is a quotient of this and hence discrete as well. By the second claim of the previous Lemma it has finitely many connected components, i.e. it is finite. Now every component X is closed and open in $\tilde{X}$ (being the inverse image under the quotient map of a point of $\tilde{X}/G$), that is a union of connected components. Since the components are closed and open, $\tilde{X}$ is their disjoint union.

Now let X be a component of $\tilde{X}$, and let $g_\mathrm{f} \in \mathbb{G}(\mathbb{A}_\mathrm{f})$ be a representative for the class of X in $\mathbb{G}(\mathbb{Q}) \backslash \mathbb{G}(\mathbb{A}_\mathrm{f}) / K_\mathrm{f}$. We can then set:

$$\Gamma = \left\{ \gamma_\infty \mid \gamma \in \mathbb{G}(\mathbb{Q}) : (\gamma_p)_{p < \infty} \in g_\mathrm{f} K_\mathrm{f} g_\mathrm{f}^{-1} \right\}.$$

We first verify that this is a discrete subgroup of G. For this let $U \subset G$ be a relatively compact open neighbourhood of the identity. Then $\tilde{U} = U \times g_\mathrm{f} K_\mathrm{f} g_\mathrm{f}^{-1}$ is a relatively compact open neighbourhood of the identity in $\mathbb{G}(\mathbb{A})$, and it follows that $\Gamma \cap U \simeq \mathbb{G}(\mathbb{Q}) \cap \tilde{U}$ is finite.

To see that $\Gamma\backslash G \simeq X$ we start with the surjective map $\varphi\colon G \to X$ given by $\varphi(g_\infty) = \mathbb{G}(\mathbb{Q})g_\infty g_{\mathrm{f}} K_{\mathrm{f}}$. By definition we have $\varphi(g_\infty) = \varphi(g'_\infty)$ iff there exist $\gamma \in \mathbb{G}(\mathbb{Q})$, $k_{\mathrm{f}} \in K_{\mathrm{f}}$ such that $\gamma g_\infty g_{\mathrm{f}} k_{\mathrm{f}} = g_\infty g_{\mathrm{f}}$. At the finite places this reads $\gamma = g_{\mathrm{f}} k_{\mathrm{f}} g_{\mathrm{f}}^{-1}$, at the infinite place $\gamma g_\infty = g'_\infty$ and we conclude that $\varphi(g_\infty) = \varphi(g'_\infty)$ iff there exists $\gamma \in \Gamma$ such that $\gamma g_\infty = g'_\infty$. Since Γ is closed in G this means our map φ induces a bijective continuous map $\Gamma\backslash G \to X$. It is open since the map $G \to \mathbb{G}(\mathbb{A})/K_{\mathrm{f}}$ given by $g_\infty \mapsto g_\infty g_{\mathrm{f}} K_{\mathrm{f}}$ is open (the space $\mathbb{G}(\mathbb{A}_{\mathrm{f}})/K_{\mathrm{f}}$ is discrete).

Finally we note that X_Z is a closed subset of $\tilde{X}_Z = Z\backslash\tilde{X}$. It thus suffices to verify the compactness of the latter space. A direct computation shows: $Z_{\mathbb{G}}(\mathbb{Q})\backslash Z_{\mathbb{G}}(\mathbb{A}) = Z^+ \times \prod_p \mathbb{Z}_p^\times$ where Z^+ is the connected component of Z. It follows that $\tilde{X}_Z/\prod_p \mathbb{Z}_p^\times = \mathbb{G}(\mathbb{Q})Z_{\mathbb{G}}(\mathbb{A})\backslash\mathbb{G}(\mathbb{A})/K_{\mathrm{f}}$. In fact, for $p \notin R_1$ we have $\mathbb{Z}_p^\times \subset \mathcal{O}_p^\times = K_p$ while the fact that $K_{R_1} < G_{R_1}$ is open implies that is contains a subgroup of finite index of $\prod_{p\in R_1} \mathbb{Z}_p^\times$. We conclude that $\tilde{X}_Z$ is a finite cover of the compact space $\mathbb{G}(\mathbb{Q})Z_{\mathbb{G}}(\mathbb{A})\backslash\mathbb{G}(\mathbb{A})/K_{\mathrm{f}}$. □

LEMMA 2.3.7. *Let $K_{\mathrm{f}} < \mathbb{G}(\mathbb{A}_{\mathrm{f}})$ be an open compact subgroup.*

(1) *Assume $\nu_{\mathbb{D}}(K_{\mathrm{f}}) = \prod_p \mathbb{Z}_p^\times$. Then $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A}_{\mathrm{f}})/K_{\mathrm{f}}$ reduces to a single point.*
(2) *For any maximal order $\mathcal{O}$, the maximal compact subgroup $K_{\mathrm{f}} = \prod_p \mathcal{O}_p^\times$ satisfies the above condition. The same holds for the intersection of two such subgroups, in which case we say that K_{f} is associated to an "Eichler order".*

PROOF. For the first part, let $g_{\mathrm{f}} \in \mathbb{G}(\mathbb{A}_{\mathrm{f}})$. We then have $\nu_{\mathbb{D}}(g_{\mathrm{f}}) = \prod_p p^{k_p} u_p \in \mathbb{A}_{\mathrm{f}}^\times$ for some $k_p \in \mathbb{Z}$ (almost all of which are zero) and $u_p \in \mathbb{Z}_p^\times$.

By a Eichler's Theorem (see [**31**, Prop. XI-3-3] and note that our $\mathbb{D}$ is $\mathbb{R}$-split) there exists $\gamma \in \mathbb{G}(\mathbb{Q})$ such that $r = \nu_{\mathbb{D}}(\gamma) = \prod_p p^{-k_p}$. Moreover, for any p the number rp^{k_p} is p-integral, so that $rp^{k_p}u_p \in \mathbb{Z}_p^\times$. By assumption there exists $k_{\mathrm{f}} \in K_{\mathrm{f}}$ such that $\nu_{\mathbb{D}}(k_{\mathrm{f}}) = \prod_p \left(rp^{k_p}u_p\right)^{-1}$. It follows that the element $h_{\mathrm{f}} = \gamma g_{\mathrm{f}} k_{\mathrm{f}}$ of $\mathbb{G}(\mathbb{A}_{\mathrm{f}})$ has $\nu_{D_p}((h_{\mathrm{f}})_p)) = 1$ at every prime p, i.e. $h_{\mathrm{f}} \in \mathbb{G}^1(\mathbb{A}_{\mathrm{f}})$. Now since $\mathbb{G}^1(\mathbb{A}_{\mathrm{f}})$ is a topological subgroup of $\mathbb{G}^1(\mathbb{A}_{\mathrm{f}})$, $K_{\mathrm{f}}^1 = K_{\mathrm{f}} \cap \mathbb{G}^1(\mathbb{A}_{\mathrm{f}})$ is an open subgroup there. By part (2) of Fact 2.3.4 we can thus find $\gamma' \in \mathbb{G}^1(\mathbb{Q})$ and $k'_{\mathrm{f}} \in K_{\mathrm{f}}^1$ such that $\gamma^1 k_{\mathrm{f}}^1 = h_{\mathrm{f}}$. We thus have:

$$1 \equiv \gamma' k'_{\mathrm{f}} \equiv \gamma g_{\mathrm{f}} k_{\mathrm{f}} \equiv g_{\mathrm{f}}$$

as desired, where equivalence is read in the double coset space $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A}_{\mathrm{f}})/K_{\mathrm{f}}$.

The second part follows immediately from part (1) of Fact 2.1.3. If $K_{\mathrm{f}}, K'_{\mathrm{f}}$ are both associated to maximal orders $\mathcal{O}, \mathcal{O}'$ it suffices to show that $\nu_{D_p}(K_p \cap K'_p) = \mathbb{Z}_p^\times$ at every place separately. At every place p where $\mathbb{D}$ ramifies D_p has a unique maximal order R_p. We then have $K_p = K'_p = R_p^\times$ and are in the same situation as before. At a place where p splits we use the isomorphism with $\mathrm{GL}_d(\mathbb{Q}_p)$ and the computation (up to conjugation) in Lemma 5.1.16 of the joint stabilizer of two vertices in the associated building to obtain an explicit form (up to conjugation) for $K_p \cap K'_p$. One sees that the intersection must contain a conjugate of the subgroup $\left\{\mathrm{diag}(u, 1, \ldots, 1) \mid u \in \mathbb{Z}_p^\times\right\}$ of $\mathrm{GL}_d(\mathbb{Q}_p)$ and hence an element with reduced norm u for any $u \in \mathbb{Z}_p^\times$. In this case $K_{\mathrm{f}} \cap K'_{\mathrm{f}}$ is the open compact subgroup associated to the order $\mathcal{O} \cap \mathcal{O}'$, which in the case $d = 2$ is the class of orders constructed in [**8**, pp. 48–55]. □

We henceforth assume that K_{f} satisfies the condition of the Lemma, so that we can identify $\tilde{X}$ and its single component X. The quotient $Y\backslash G/K$ is then a locally symmetric space of non-positive curvature (this appellation is sometimes reserved for $Y_Z = Z\backslash Y$). We also remark that (unlike G) the symmetric space G/K is always connected. Hence Y and Y_Z are connected manifolds even when X isn't.

We normalize the Haar measures dx on X_Z, dk on K and dy on Y_Z to have total mass 1 (here dy is the pushforward of dx under the the projection from X_Z to Y_Z given by averaging w.r.t. dk).

For any unitary character $\omega \in \hat{Z}$ consider the space of functions $\psi\colon X \to \mathbb{C}$ such that $\psi(xz) = \omega(z)\psi(x)$ for all $z \in Z, x \in X$. Since ω is unitary the map $x \mapsto |\psi(x)|^2$ is Z-invariant, and hence a function on the compact space X_Z. We let $L^2(X,\omega)$ denote the space of the functions ψ as above such that $\int_{X_Z} |\psi(x)|^2\, dx < \infty$. We also let $L^2(Y,\omega)$ be the subspace consisting of K-invariant functions. We note that $L^2(X,\omega)$ is a unitary representation of G under right translation and that as a G-representation $L^2(X)$ is the direct integral of the $L^2(X,\omega)$.

2.4. The p-adic groups and Hecke operators

For a prime $p \notin R_0$ we have $G_p \simeq \mathrm{GL}_d(\mathbb{Q}_p)$, $K_p \simeq \mathrm{GL}_d(\mathbb{Z}_p)$. Let $\mathcal{H}_p$ denote the convolution algebra of the bi-K_p-invariant functions of compact support on G_p. It is commutative, and generated by the elements $K_p a_p K_p$ with $a_p \in A_p$, the subgroup of diagonal matrices of $\mathrm{GL}_d(\mathbb{Q}_p)$ (see Fact 5.1.15).

By assumption we can think of functions on X as functions on $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})$ which are invariant by K_{f} on the right. For almost all primes p (except for $p \in R_1$), K_p is a direct factor of K_{f} and hence if $p \notin R = R_0 \cup R_1$, $\mathcal{H}_p$ acts on the space of functions on X by convolution on the right. Moreover, the actions of $\mathcal{H}_p$ and $\mathcal{H}_{p'}$ for $p \neq p'$ commute, and they both commute with the right regular action of $G = G_\infty$.

We call an operator T_h on functions of X associated to an element $h \in \mathcal{H}_p$ a *Hecke operator*, and think of it as arising from a discrete foliation of the manifold X where the leaf of $\mathbb{G}(\mathbb{Q})g_\infty g_{\mathrm{f}} K_{\mathrm{f}}$ is given by $\{\mathbb{G}(\mathbb{Q})g_\infty g_{\mathrm{f}} x_p K_{\mathrm{f}}\}_{x_p \in G_p}$: each leaf is of the form $H_p\backslash G_p/K_p$ for some (generically trivial) subgroup $H_p < G_p$ and the action of T_h on $f\colon X \to \mathbb{C}$ is given by restricting f to each leaf and convolving on the right with h. This action is analyzed in detail in Chapter 5, where the control of subgroups H_p causing the Hecke correspondence to *return* is achieved using the results of Chapter 4.

Together the Hecke operators at all places $p \notin R$ generate the (commutative) *Hecke algebra* $\mathcal{H}_{(R)}$. This is the algebra we have in mind when we apply Theorem 1.3.2.

DEFINITION 2.4.1. We call $\psi \in L^2(X,\omega)$ a *Hecke eigenfunction* if it is a joint eigenfunction of the Hecke algebra $\mathcal{H}_{(R)}$.

CHAPTER 3

The Micro-Local Lift

3.1. Introduction and motivation

Let $\psi \in L^2(Y,\omega)$ be normalized as well as an eigenfunction of $\mathfrak{Z}$. The aim of the present section is to construct a distribution μ_ψ on X_Z that lifts the measure $\bar{\mu}_\psi$ on Y_Z, and establish some basic properties of μ_ψ. We will of course take $\psi = \psi_n$, and the corresponding distribution will be the distribution μ_n discussed in the proof of Theorem 1.3.2. The functions $\tilde{\psi}_n$ will then be chosen so that the measures $|\tilde{\psi}_n(x)|^2 dx$ approximate μ_n; finally, both $|\tilde{\psi}_n(x)|^2$ and μ_n will become A-invariant as $n \to \infty$.

We begin by fixing notation and providing some motivation for the relatively formal definitions that follow.

Setting $\psi(x) = \psi(xK)$ for any $x \in X$, we can think of ψ as a function in $L^2(X,\omega)$. By the uniqueness of spherical functions [**13**, Th. 4.3 & 4.5], ψ generates an irreducible spherical G-subrepresentation of $L^2(X)$. As discussed in Section 3.2.1 below, we can then find $\nu \in \mathfrak{a}^*_{\mathbb{C}}$ such that this representation is isomorphic to a principal series representation π_ν (in particular, π_ν is unitarizable). We will assume for the rest of this section that $Re(\nu) = 0$, i.e. π_ν is *tempered*, and that ν is regular. This will eventually be the only case of interest to us in view of the non-degeneracy assumption (Definition 3.3.8). In this case the induced representation (V_K, I_ν) (living on a space of functions on K defined below, including the constant function φ_0) is irreducible and isomorphic to π_ν.

It follows that there is a unique G-homomorphism $R_\psi : (V, I_\nu) \to L^2(X)$ such that $R_\psi(\varphi_0) = \psi$. The normalization $\|\psi\|_{L^2(X,\omega)} = 1$ now implies $\|R_\psi(f)\|_{L^2(X)} = \|f\|_{L^2(K)}$ for any $f \in V_K$, i.e. that R_ψ is an isometry.

We now give the rough idea of the construction that follows in the language of Wolpert and Lindenstrauss; the language we shall use later is slightly different, so the discussion here also provides a translation. The strategy of proof is similar to theirs; in a sense, the main difficulty is finding the "correct" definitions in higher rank. For instance, the proofs of Wolpert and Lindenstrauss use heavily the fact that K-types for $\mathrm{PSL}_2(\mathbb{R})$ have multiplicity one, and the explicit action of the Lie algebra by raising and lowering operators. We shall need a more intrinsic approach to handle the general case.

The measure $\bar{\mu}_\psi$ on Y_Z is defined by $g \mapsto \int_{Y_Z} g(y)|\psi(y)|^2 dy$. More generally, suppose that $\psi' \in L^2(X)$ belongs to the G-subrepresentation generated by ψ, i.e. $\psi' \in R_\psi(V)$. Then $\psi(x)\overline{\psi'(x)}$ is Z-invariant and we can consider the (signed) measure on X_Z given by:

$$\sigma\colon g \mapsto \int_{X_Z} \psi(x)\overline{\psi'(x)} g(x) dx. \tag{3.1.1}$$

If $g(x)$ is K-invariant, then so is the product $\psi(x)g(x)$, and it follows that the right-hand side of (3.1.1) depends only on the projection of ψ' onto $R_\psi(V)^K$. The space $R_\psi(V)^K$ is one-dimensional, spanned by ψ, and it follows that if $\psi' - \psi \perp \psi$ then the measure σ on X projects to the measure $\bar{\mu}_\psi$ on Y.

The distribution μ_ψ we shall construct will be in the spirit of (3.1.1), but with ψ' a "generalized vector" in $R_\psi(V)$. Suppose, in fact, that $\psi'_1, \psi'_2, \ldots \psi'_n, \ldots$ are an infinite sequence of elements of $R_\psi(V)$ that transform under different K-types, and suppose further that $g \in C_c^\infty(X_Z)_K$. Then, by considering K-types, the integral $\int_X \psi(x)\overline{\psi'_j(x)}g(x)dx$ vanishes for all sufficiently large j. It follows that, if one sets ψ' to be the *formal sum* $\sum_{j=1}^\infty \psi'_j$, one can make sense of (3.1.1) by interpreting it as:

$$\sigma(g) = \sum_{j=1}^{\infty} \int_X \psi(x)\overline{\psi'_j(x)}g(x)dx$$

In other words, if $g \in C_c^\infty(X_Z)_K$, we may make sense of (3.1.1) while allowing ψ' to belong to the space $\widehat{V_K}$ of "infinite formal sums of K-types." Our definition of μ_ψ will, indeed, be of the form (3.1.1) but with ψ' an "infinite formal sum" of this kind.

For a certain choice of ψ' (denoted Φ_∞ in [**18**]), we will wish to show that (3.1.1) is "approximately a positive measure" and "approximately A-invariant," where both statements become true in the large eigenvalue limit in an appropriate sense. For the "approximate positivity," we shall integrate (3.1.1) by parts to show that there exists another unit vector $\psi'' \in R_\psi(V)$ such that $\int_X \psi(x)\overline{\psi'(x)}g(x)dx \approx \int_X |\psi''(x)|^2 g(x)dx$, where the right-hand side is evidently a positive measure. For the "approximate A-invariance," we will construct differential operators that annihilate $\psi(x)\overline{\psi'(x)}$; this reduces to a purely algebraic question of constructing elements in $U(\mathfrak{g})$ that annihilate a vector in a certain tensor product representation.

The space $\widehat{V_K}$ is very closely linked to the dual V'_K of the K-finite vectors: the conjugate linear isomorphism $T : V \to V'$ (3.2.1) extends to a conjugate-linear isomorphism $T : \widehat{V_K} \to V'_K$. For formal reasons, it is simpler to work with V'_K than $\widehat{V_K}$; this is the viewpoint we shall take in Definition 3.3.1. To motivate this viewpoint, let us rewrite (3.1.1) in a different fashion. Let $v' \in V$ be chosen so that $\psi' = R_\psi(v')$, and let P be the orthogonal projection of $L^2(X)$ onto $R_\psi(V)$. We may rewrite (3.1.1) – using the notations of Definition 3.2.3 – as follows:

$$\begin{aligned} \sigma(g) &= \langle \psi(x)g(x), \psi'(x)\rangle_{L^2(X)} = \langle P(\psi(x)g(x)), \psi'(x)\rangle_{L^2(X)} \\ &= \langle R_\psi^{-1} \circ P(\psi(x)g(x)), v'\rangle_V = T(v') \circ R_\psi^{-1} \circ P(\psi(x)g(x)) \end{aligned} \tag{3.1.2}$$

Now, if $g \in C_c^\infty(X_Z)_K$, then the quantity $R_\psi \circ P(\psi(x)g(x))$ is K-finite, i.e. belongs to V_K. It follows that, if $g \in C_c^\infty(X_Z)_K$, the last expression of (3.1.2) makes formal sense if we replace $T(v')$ by any functional $\Phi \in V'_K$.

3.2. More on Real Lie Groups

3.2.1. Spherical Representations and the model (V_K, I_ν). We recall some facts from the representation theory of compact and reductive groups. At the end of this section we analyze a model (the "compact picture") for the spherical dual of G.

THEOREM 3.2.1. [**16**, Th. 1.12] *Let K be a compact topological group and let $\hat{K}_{\text{fin}}$ be the set of equivalence classes of irreducible finite-dimensional unitary representations of K.*

(1) *(Peter-Weyl) Every $\rho \in \hat{K}_{\text{fin}}$ occurs discretely in $L^2(K)$ with multiplicity equal to its dimension $d(\rho)$. Moreover, $L^2(K)$ is isomorphic to the Hilbert direct sum of its isotypical components $\{L^2(K)_\rho\}_{\rho\in\hat{K}_{\text{fin}}}$.*
(2) *Let $\pi : K \to \mathrm{GL}(W)$ be a representation of K on the Frêchet space W. Then $\oplus_{\rho\in\hat{K}} W_\rho$ is dense in W, where W_ρ is the ρ-isotypical subspace.*
(3) *Every irreducible representation of on a Frêchet space is finite-dimensional and hence unitarizable. In particular, $\hat{K}_{\text{fin}}$ is the unitary dual of K.*
(4) *For K as in Section 2.2, $\hat{K}$ is countable.*

Note that while [**16**, Th. 1.12(c-e)] are only claimed for unitary representations on Hilbert spaces, their proofs only rely on the action of the convolution algebra $C(K)$ on representations of K, and hence carry over with little modification to the more general context needed here. The last conclusion follows from the separability of $L^2(K)$, which in turn follows from the separability of K.

NOTATION 3.2.2. Let $\pi : K \to \mathrm{GL}(W)$ be as above. The algebraic direct sum

$$W_K \stackrel{\text{def}}{=} \oplus_{\rho\in\hat{K}} W_\rho$$

consists precisely of these $w \in W$ which generate a finite-dimensional K-subrepresentation. We refer to W_K as the space of K-finite vectors. We will use W^K to denote these vectors of W fixed by K.

DEFINITION 3.2.3. Set $V = L^2(M\backslash K)$, and set $V_K \subset V$ to be the space of K-finite vectors. Let $C^\infty(M\backslash K)$ be the smooth subspace, $C^\infty(M\backslash K)'$ the space of distributions on $M\backslash K$. Let V_K' (resp. V') be the dual to V_K (resp. V). Then we have natural inclusions $V_K \subset C^\infty(M\backslash K) \subset V$ and $V_K' \supset C^\infty(M\backslash K)' \supset V'$; further, we have (Riesz representation) a conjugate-linear isomorphism

$$V \stackrel{T}{\hookrightarrow} V' \tag{3.2.1}$$

where the map $T\colon V \to V'$ is defined via the rule $T(f)(g) = \langle g, f\rangle_V = \int_{M\backslash K} g\overline{f}\,dk$.

Fix an increasing exhaustive sequence of finite dimensional K-stable subspaces of V_K, i.e. a sequence $V_1 \subset V_2 \subset \cdots \subset V_N \subset V_{N+1} \subset \ldots$ of subspaces such that $\cup_{i=1}^\infty V_i = V_K$ and each V_i is a K-subrepresentation.

For $\Phi \in V_K'$ and $1 \le N \in \mathbb{Z}$, define the N-truncation of Φ as the unique element $\Phi_N \in V_N$ such that $T(\Phi_N) - \Phi$ annihilates V_N.

Finally let $\varphi_0 \in V_K$ be the function that is identically 1.

DEFINITION 3.2.4. Let μ be a regular Borel measure on a space X. Call a sequence of non-negative functions $\{f_j\} \in L^1(\mu)$ a *δ-sequence* at $x \in X$ if, for every j, $\int f_j d\mu = 1$, and moreover if, for every $g \in C(X)$, $\lim_{j\to\infty} \int f_j \cdot g d\mu = g(x)$.

LEMMA 3.2.5. *There exists a sequence $\{f_j\}_{j=1}^{\infty} \subset V_K$ such that $|f_j|^2$ is a δ-sequence on $M\backslash K$.*

PROOF. Let $\{h_j\}_{j=1}^{\infty} \subset C(M\backslash K)$ be a δ-sequence. By the Peter-Weyl theorem V_K is dense in $C(M\backslash K)$, so that for every j we can choose $f'_j \in V_K$ such that the difference $\left\| \sqrt{h_j(k)} - f'_j(k) \right\|_{\infty} \leq \frac{1}{2^j}$. Then one may take $f_j = \frac{f'_j}{\|f'_j\|_2}$. □

Secondly, we recall the construction of the spherical principal series representations of a reductive Lie group. An irreducible representation of G is *spherical* if it contains a K-fixed vector. Such a vector is necessarily unique up to scaling.

To any $\nu \in \mathfrak{a}^*_{\mathbb{C}}$ we associate the character $\chi_\nu(p) = \exp(\nu(H_0(p)))$ of P_0 and the unitarily induced representation with $(\mathfrak{g}, K)$-module

$$\operatorname{Ind}_{P_0}^G \chi_\nu = \left\{ f \in C^\infty(G)_K \mid \forall p \in P, g \in G : f(pg) = e^{\langle \nu+\rho, H_0(p)\rangle} f(g) \right\}. \tag{3.2.2}$$

By the Iwasawa decomposition, every $f \in \operatorname{Ind}_{P_0}^G \chi_\nu$ is determined by its restriction to K; this restriction defines an element of the space V_K. Conversely, every $f \in V_K$ extends uniquely to a member of $\operatorname{Ind}_{P_0}^G \chi_\nu$.

DEFINITION 3.2.6. For $\nu \in \mathfrak{a}^*_{\mathbb{C}}$, we denote by (I_ν, V_K) the representation of $\mathfrak{g}$ on V_K fixed by the discussion above; we shall also use I_ν to denote the corresponding action of $\mathfrak{g}$ on $C^\infty(M\backslash K)$ and of G on V. We shall denote by I'_ν the dual action of $\mathfrak{g}$ on either V'_K or $C^\infty(M\backslash K)'$.

Note also that $\varphi_0 \in V_K$ (see Definition 3.2.3) is a spherical vector for the representation (I_ν, V_K).

THEOREM 3.2.7. *(The unitary spherical dual; references are drawn from* [**16**]*)*

(1) *For any $\nu \in \mathfrak{a}^*_{\mathbb{C}}$, $\operatorname{Ind}_{P_0}^G \chi_\nu$ has a unique spherical irreducible subquotient, to be denoted π_ν. [Th. 8.37] Any spherical irreducible unitary representation of G is isomorphic to π_ν for some ν. [Th. 8.38] We have $\pi_{\nu_1} \simeq \pi_{\nu_2}$ iff there exists $w \in W(\mathfrak{a}\colon\mathfrak{g})$ such that $\nu_2 = w\nu_1$.*
(2) *[§7.1-3] If $\operatorname{Re}(\nu) = 0$ then $\operatorname{Ind}_{P_0}^G \chi_\nu$ is unitarizable, with the invariant Hermitian form given by $\langle f, g\rangle = \int_{M\backslash K} f(k)\overline{g(k)}dk$. This representation has a unique spherical summand (necessarily isomorphic to π_ν), and we let $j_\nu \colon V_K \to \pi_\nu$ denote the orthogonal projection map. [Th. 7.2] If ν is regular then $\operatorname{Ind}_{P_0}^G \chi_\nu$ is irreducible.*
(3) *[§16.5(7) & Th. 16.6] If π_ν is unitarizable then $\operatorname{Re}(\nu)$ belongs to the convex hull of $\{w\rho\}_{\{w\in W(\mathfrak{a}\colon\mathfrak{g})\}} \subset \mathfrak{a}^*$, a compact set. Moreover, there exists $w \in W(\mathfrak{a}\colon\mathfrak{g})$ such that $w^2 = 1$ and $w\nu = -\bar{\nu}$. In particular if $\operatorname{Re}(\nu) \neq 0$, then $w \neq 1$, and since Im(ν) is w-fixed it is not regular.*

Note that the norm on π_ν is only unique up to scaling. If $\operatorname{Re}(\nu) = 0$ and Im(ν) is regular (the main case under consideration), we choose $\|\varphi_0\|_{\pi_\nu} = 1$.

For future reference we compute the action of $\mathfrak{g}$ on V_K via I_ν. First, note that the action of K on $V = L^2(M\backslash K)$ is given by right translation, and the action of $\mathfrak{k} \subset \mathfrak{g}$ on V_K is then given by right differentiation.

Secondly, recall that if $U \subset \mathbb{R}^n$ is open, a *differential operator* $\mathbf{D}$ on U is an expression of the form $\sum_{i=1}^K f_i \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n}$, where the f_i are smooth and $\alpha_j \geq 0$. If M is a smooth n-manifold, we say a map $\mathbf{D} : C^\infty(M) \to C^\infty(M)$ is a differential operator if it is defined by a differential operator in each coordinate chart.

LEMMA 3.2.8. *Let $f \in V_K$ and let $X \in \mathfrak{g}$. Then there exists a differential operator $\mathbf{D}_X$ on $M\backslash K$ (depending linearly on X and independent of ν) such that for every $k \in K$,*

$$(I_\nu(X)f)(k) = \langle \nu + \rho, H_0(\mathrm{Ad}(k)X)\rangle\, f(k) + (\mathbf{D}_X f)(k).$$

PROOF. Let $t \in \mathbb{R}$ be small, and consider $f(k\exp(tX)) = f(\exp(t\,\mathrm{Ad}(k)X) \cdot k)$. We write the Iwasawa decomposition of $\mathrm{Ad}(k)X \in \mathfrak{g}$ as $\mathrm{Ad}(k)X = X_\mathfrak{n}(k) + X_\mathfrak{a}(k) + X_\mathfrak{k}(k)$ where $X_\mathfrak{a}(k) = H_0(\mathrm{Ad}(k)X)$. By the Baker-Campbell-Hausdorff formula, $\exp(t\,\mathrm{Ad}(k)X)$ has the form $\exp(tX_\mathfrak{n}(k)) \cdot \exp(tX_\mathfrak{a}(k)) \cdot \exp(tX_\mathfrak{k}(k)) + O(t^2)$, so that:

$$(I_\nu(X)f)(k) = \frac{d}{dt} f\left(\exp(tX_\mathfrak{n}(k)) \cdot \exp(tX_\mathfrak{a}(k)) \cdot k\right) \restriction_{t=0} + \frac{d}{dt} f\left(\exp(tX_\mathfrak{k}(k))k\right) \restriction_{t=0} .$$

To conclude, observe that $f \mapsto \frac{d}{dt} f\left(\exp(tX_\mathfrak{k}(k))k\right) \restriction_{t=0}$ defines a differential operator $\mathbf{D}_X$ on $M\backslash K$. □

Lemma 3.2.8 will be used in the following way: as $\|\nu\| \to \infty$, the operator $I_\nu(\frac{X}{\|\nu\|})$ acts on V_K in a very simple fashion, *modulo* certain error terms of order $\|\nu\|^{-1}$. The simplicity of this "rescaled" action as $\|\nu\| \to \infty$ will be of importance in our analysis.

3.2.2. Some Functional Analysis. We collect here some simple functional analysis facts that we shall have need of.

Let $C_c^\infty(X_Z)$ denote the space of smooth functions of compact support on X. It is endowed with the usual "direct-limit" topology: fix a sequence of K-invariant compact sets $C_1 \subset C_2 \subset \ldots$ such that their interiors exhaust X. Then the $C_c^\infty(C_i)$ exhaust $C_c^\infty(X_Z)$. $C_c^\infty(C_i)$ is endowed as usual with a family of seminorms, viz. for any $\mathcal{D} \in U(\mathfrak{g}_\mathbb{C})$ we define $\|f\|_{C_i,\mathcal{D}} = \sup_{x \in C_i} |\mathcal{D}f|$. These seminorms induce a topology on each $C_c^\infty(C_i)$. We give $C_c^\infty(X_Z)$ the topology of the union of $C_c^\infty(C_i)$, i.e. a map from $C_c^\infty(X_Z)$ is continuous if and only if its restriction to each $C_c^\infty(C_i)$ is continuous.

In other words: a sequence of functions converges in $C_c^\infty(X_Z)$ if their supports are all contained in a fixed compact set, and all their derivatives converge uniformly on that compact set.

$C_c^\infty(X_Z)$is a locally convex complete space in this topology. In particular, its subspace $C_c^\infty(X_Z)_K$ of K-finite vectors is dense. We denote by $C_c^\infty(X_Z)'$ (resp. $C_c^\infty(X_Z)'_K$) the topological dual to $C_c^\infty(X_Z)$ (resp. the algebraic dual to $C_c^\infty(X_Z)_K$). Both spaces will be endowed with the weak-* topology. We shall refer to an element of $C_c^\infty(X_Z)'$ as a *distribution* on X.

Let $C_0(X)$ be the Banach space of continuous functions on X decaying at infinity, endowed with the supremum norm. Let $C_0(X)'$ be the continuous dual of $C_0(X)$; the

Riesz representation theorem identifies it with the space of finite (signed) Borel measures on X. We endow $C_0(X)'$ with the weak-* topology.

It is easy to see that $C_c^\infty(X_Z)_K$ is dense in $C_0(X)$. In particular any (algebraic) linear functional on $C_c^\infty(X_Z)_K$ which is bounded w.r.t. the $\sup$-norm extends to a finite signed measure on X, with total variation equal to the norm of the functional. Moreover, if this functional is non-negative on the non-negative members of $C_c^\infty(X_Z)_K$ then the associated measure is a positive measure.

3.3. Representation-Theoretic Lift

3.3.1. Lifting a single (non-degenerate) eigenfunction.

DEFINITION 3.3.1. Let $\Phi \in V_K'$ be an (algebraic) functional, and $f \in V_K$. Let $\mu_\psi(f, \Phi)$ be the functional on $C_c^\infty(X_Z)_K$ defined by the rule:

$$\mu_\psi(f, \Phi)(g) = \Phi \circ R_\psi^{-1} \circ P(R_\psi(f) \cdot g) \tag{3.3.1}$$

where $g \in C_c^\infty(X_Z)_K$, $P : L^2(X, \omega) \to R_\psi(V)$ is the orthogonal projection, and $R_\psi(f) \cdot g$ denotes pointwise multiplication of functions on X.

REMARK 3.3.2. In fact, if $\Phi \in C^\infty(M\backslash K)'$ (see equation (3.2.1)) then $\mu_\psi(f, \Phi)$ extends to an element of $C_c^\infty(X_Z)'$, i.e. defines a distribution on X: μ_ψ is the composite

$$C_c^\infty(X_Z) \overset{g \mapsto R_\psi(f) g}{\longrightarrow} C_c^\infty(X_Z) \overset{R_\psi^{-1} P}{\longrightarrow} C^\infty(M\backslash K) \overset{\Phi}{\to} \mathbb{C},$$

and it is easy to verify that each of these maps is continuous. This is never used in our arguments: we use this observation only to refer to certain μ_ψ as “distributions”.

DEFINITION 3.3.3. Let $\delta \in V_K'$ be the distribution $\delta(f) = f(1)$, and call $\mu_\psi \overset{\text{def}}{=} \mu_\psi(\varphi_0, \delta)$ the (non-degenerate) *microlocal lift* of $\bar{\mu}_\psi$.

The rest of the section will exhibit basic formal properties of this definition. We will establish most of the formal properties of μ_ψ by restricting Φ to be of the form $T(f_2)$, where the conjugate-linear mapping T is as defined in (3.2.1). This situation will occur sufficiently often that, for typographical ease, it will be worth making the following definition:

DEFINITION 3.3.4. Let $f_1, f_2 \in V_K$. We then set $\mu_\psi^T(f_1, f_2) = \mu_\psi(f_1, T(f_2))$.

LEMMA 3.3.5. *Suppose* $f_1, f_2 \in V_K$. *Then*

$$\mu_\psi^T(f_1, f_2)(g) = \int_{X_Z} R_\psi(f_1)(x)\overline{R_\psi(f_2)(x)} g(x) dx. \tag{3.3.2}$$

and μ_ψ^T *defines a signed measure on* X_Z *of total variation at most* $||f_1||_{L^2(K)}||f_2||_{L^2(K)}$. *If* $f_1 = f_2$, *then* $\mu_\psi^T(f_1, f_1)$ *is a positive measure of mass* $||f_1||_{L^2(K)}^2$.

PROOF. (3.3.2) is a consequence of the definition of μ. The Cauchy-Schwarz inequality implies that $|\mu_\psi^T(f_1, f_2)(g)| \le ||f_1||_{L^2(K)}||f_2||_{L^2(K)}||g||_{L^\infty(X)}$, whence the second conclusion. The last assertion is immediate. □

In fact, it may be helpful to think of μ_ψ as being given by a distributional extension of the formula (3.3.2); see the discussion of Section 3.1.

LEMMA 3.3.6. *The distribution $\mu_\psi(\varphi_0, \delta)$ on X projects to the measure $|\psi|^2 dy$ on Y.*

PROOF. In view of the previous Lemma, it will suffice to show that the distribution $\mu_\psi(\varphi_0, \delta) - \mu_\psi^T(\varphi_0, \varphi_0)$ on X projects to 0 on Y_Z. This amounts to showing that $\mu_\psi(\varphi_0, \delta - T(\varphi_0))$ annihilates any K-invariant function $g \in C_c^\infty(X_Z)^K$. Taking into account that the functional $\delta - T(\varphi_0)$ on V_K annihilates any K-invariant vector, the claim follows from the definition of μ_ψ. □

LEMMA 3.3.7. *The map $\mu_\psi : V_K \otimes V_K' \to C_c^\infty(X_Z)_K'$ is equivariant for the natural $\mathfrak{g}$-actions on both sides.*

PROOF. This follows directly from the definition of μ_ψ. □

Concretely speaking, this says that for $f \in V_K, \Phi \in V_K', g \in C_c^\infty(X_Z)_K, X \in \mathfrak{g}$ we have

$$\mu_\psi(Xf_1, \Phi)(g) + \mu_\psi(f_1, X\Phi)(g) + \mu_\psi(f_1, \Phi)(Xg) = 0 \tag{3.3.3}$$

where X acts on V_K via I_ν and on V_K' via I_ν'. In particular, if $f_1, f_2 \in V_K$ we have

$$\mu_\psi^T(Xf_1, f_2)(g) + \mu_\psi^T(f_1, Xf_2)(g) + \mu_\psi^T(f_1, f_2)(Xg) = 0 \tag{3.3.4}$$

3.3.2. Sequences of eigenfunctions and quantum limits. In what follows we shall consider $\psi_n \in L^2(Y, \omega_n)$, a sequence of eigenfunctions with parameters $\{\nu_n\}_{n=1}^\infty$ diverging to ∞ (i.e. with $\|\nu_n\| \to \infty$). Set $\tilde{\nu}_n = \frac{\nu_n - d\omega_n}{\|\nu_n\|}$ (i.e. remove the central character part of the parameter). For $f_1, f_2 \in V_K$ and $\Phi \in V_K'$, we abbreviate $\mu_{\psi_n}^T(f_1, f_2)$ (resp. $\mu_{\psi_n}(f, \Phi)$) to $\mu_n^T(f_1, f_2)$ (resp. $\mu_n(f, \Phi)$), and we abbreviate the microlocal lift μ_{ψ_n} ($:= \mu_n(\varphi_0, \delta)$) to μ_n.

DEFINITION 3.3.8. ($G^{\mathrm{ad}} = G/Z$ simple) We say a sequence ψ_n is *non-degenerate* if every limit point of the sequence $\tilde{\nu}_n$ is regular.

We say that it is *conveniently arranged* if it is non-degenerate, $\mathrm{Re}(\nu_n) = 0$ for all n, the limit in $\lim_{n\to\infty} \tilde{\nu}_n$ exists, the ν_n are all regular, and for all $f_1, f_2 \in V_K$ the measures $\mu_n^T(f_1, f_2)$ converge in $C_0(X_Z)'$ as $n \to \infty$. In this situation we denote $\lim_{n\to\infty} \tilde{\nu}_n$ by $\tilde{\nu}_\infty$.

The existence of non-degenerate sequences of eigenfunctions was discussed in Remark 1.3.3. This follows from strong versions of Weyl's Law on Y. By Theorem 3.2.7, the non-degeneracy of a sequence ψ_n as in the Definition implies $\mathrm{Re}(\nu_n) = 0$ for all large enough n. For fixed $f_1, f_2 \in V_K$ the total variation of the measures $\mu_n^T(f_1, f_2)$ is bounded independently of n (Lemma 3.3.5); in view of the (weak-*) compactness of the unit ball in $C_0(X)'$ it follows that this sequence of measures has a convergent subsequence. Combining this remark with the fact that V_K has a countable basis, a diagonal argument shows that every non-degenerate sequence of eigenfunctions has a conveniently arranged subsequence.

Now suppose $\{\psi_n\}$ is a conveniently arranged sequence and fix $f_1 \in V_K, \Phi \in V_K', g \in C_c^\infty(X_Z)_K$. Let Φ_N be the N-truncation of Φ (see Definition 3.2.3). In view of (3.3.1), if we choose $N := N(f_1, g)$ sufficiently large, then $\mu_n(f_1, \Phi)(g) = \mu_n^T(f_1, \Phi_N)(g)$. It follows that the limit $\lim_{n\to\infty} \mu_n(f_1, \Phi)(g)$ exists.

We may consequently define $\mu_\infty\colon V_K \times V_K' \to C_c^\infty(X_Z)_K'$ and $\mu_\infty^T\colon V_K \times V_K \to C_c^\infty(X_Z)_K'$ by the rules:

$$\mu_\infty(f,\Phi)(g) = \lim_{n\to\infty} \mu_n(f_1,\Phi)(g), \quad (g \in C_c^\infty(X_Z)_K)$$
$$\mu_\infty^T(f_1,f_2) = \mu_\infty(f_1,T(f_2)) \tag{3.3.5}$$

LEMMA 3.3.9. *For fixed $f_1 \in V_K$, the map $\Phi \to \mu_\infty(f_1,\Phi)$ is continuous as a map $V_K' \to C_c^\infty(X_Z)_K'$, both spaces being endowed with the weak topology.*

PROOF. This is an easy consequence of the definitions. □

It is natural to ask whether $\mu_\infty(f_1,\Phi)$ extend to an element of $C_c^\infty(X_Z)'$, at least when $\Phi \in C^\infty(M\backslash K)'$. Indeed a uniform bound on the distributions $\mu_n(f_1,\Phi)$ follows from making the argument of Remark 3.3.2 quantitative. This is not needed for our choice of (f_1,Φ), however, when we can address this directly.

Henceforth $\{\psi_n\}_{n=1}^\infty$ will be a conveniently arranged sequence. We will show that $\mu_\infty(\varphi_0,\delta)$ is positive and bounded w.r.t. the L^∞ norm on $C_c^\infty(X_Z)_K$. It hence extends to a finite positive measure.

The key to the positivity of the limits is the following lemma (cf. [**33**, Prop. 3.3], [**18**, Th. 3.1]).

LEMMA 3.3.10. *(Integration by parts) Let $\{\psi_n\}$ be conveniently arranged. Then, for any $f,\ f_1,\ f_2 \in V_K$ we have:*

$$\mu_\infty^T(f_1, f\cdot f_2) = \mu_\infty^T(\bar{f}\cdot f_1, f_2). \tag{3.3.6}$$

Here e.g. $f\cdot f_2$ denotes pointwise multiplication of functions on $M\backslash K$.

PROOF. We start by exhibiting explicit functions f for which (3.3.6) is valid.

Extend every $\nu \in \mathfrak{a}_\mathbb{C}^*$ to $\mathfrak{g}_\mathbb{C}^*$ via the Iwasawa decomposition $\mathfrak{g} = \mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$. For any $X \in \mathfrak{g}_{\text{ss}}$, let $p_X(k) = \frac{1}{i}\langle\tilde{\nu}_\infty, \text{Ad}(k)X\rangle$. For fixed X, $k \mapsto p_X(k)$ defines a K-finite element of $L^2(M\backslash K)$.

By (3.3.4), for every X, f_1, f_2, g, and n, we have

$$\mu_n^T(Xf_1,f_2)(g) + \mu_n^T(f_1,X\,f_2)(g) + \mu_n^T(f_1,f_2)(Xg) = 0. \tag{3.3.7}$$

Divide by $\|\nu_n\|$ and apply Lemma 3.2.8 (as well as $\langle d\omega_n, \mathfrak{g}_{\text{ss}}\rangle = 0$) to see:

$$\mu_n^T(ip_n\cdot f_1,f_2)(g) + \mu_n^T(f_1,ip_n\cdot f_2)(g) = -\frac{\mu_n^T(\mathbf{D}_Xf_1,f_2)(g) + \mu_n^T(f_1,\mathbf{D}_Xf_2)(g) + \mu_n^T(f_1,f_2)(Xg)}{\|\nu_n\|}, \tag{3.3.8}$$

where $p_n(k) = \frac{1}{i}\left\langle \tilde{\nu}_n + \frac{\rho}{\|\nu_n\|}, \text{Ad}(k)X\right\rangle$.

As $n \to \infty$, the right-hand side of (3.3.8) tends to zero by Lemma 3.3.5. On the other hand, p_nf_i (considered as continuous functions on K) converge uniformly to p_Xf_i. Another application of Lemma 3.3.5 shows that the left-hand side of (3.3.8) converges to $i\mu_\infty^T(p_Xf_1,f_2) - i\mu_\infty^T(f_1,p_X\cdot f_2)$. Since $p_X = \overline{p_X}$ this shows that (3.3.6) holds with $f = p_X$.

Now let $\mathcal{F} \subset C(M\backslash K)$ be the $\mathbb{C}$-subalgebra generated by the p_X and the constant function 1. Clearly (3.3.6) holds for all $f \in \mathcal{F}$. This subalgebra is K-stable since

$p_X(kk_1) = p_{\mathrm{Ad}(k_1)X}(k)$ and hence $\mathcal{F} \cap V_\rho \subset \mathcal{F}$ for all $\rho \in \hat{K}$. Showing $\mathcal{F}$ is dense in $L^2(M\backslash K)$ suffices to conclude that $\mathcal{F} = V_K$.

We will prove the stronger assertion that $\mathcal{F}$ is dense in $C(M\backslash K)$ using the Stone-Weierstrass theorem. Note that $1 \in \mathcal{F}$, and $\mathcal{F}$ is closed under complex conjugation since $p_X = \overline{p_X}$. It therefore suffices to show that $\mathcal{F}$ separates the points of $M\backslash K$. To this end, let $k_1, k_2 \in K$ be such that $p_X(k_1) = p_X(k_2)$ for all $X \in \mathfrak{g}$. Then $\langle \tilde{\nu}_\infty, \mathrm{Ad}(k_1)X\rangle = \langle \tilde{\nu}_\infty, \mathrm{Ad}(k_2)X\rangle$ for all $X \in \mathfrak{g}$, i.e. $\langle \mathrm{Ad}(k_1)^{-1}\tilde{\nu}_\infty - Ad(k_2)^{-1}\tilde{\nu}_\infty, X\rangle = 0$ for all $X \in \mathfrak{g}$. This implies that $\mathrm{Ad}(k_1^{-1})\tilde{\nu}_\infty = \mathrm{Ad}(k_2)^{-1}\tilde{\nu}_\infty$; by the non-degeneracy assumption, $Z_K(\tilde{\nu}_\infty) = Z_K(A) = M$, so $Mk_1 = Mk_2$, i.e. k_1 and k_2 represent the same point of $M\backslash K$. $\square$

Lemma 3.3.10 shows easily that $\mu_\infty(\varphi_0, \delta)$ extends to a positive measure. Indeed, choosing f_j as in Lemma 3.2.5, we see that

$$\mu_\infty(\varphi_0, \delta) = \lim_{j\to\infty} \mu_\infty^T(\varphi_0, |f_j|^2) = \lim_{j\to\infty} \mu_\infty^T(f_j, f_j). \tag{3.3.9}$$

Here we have invoked Lemma 3.3.9 for the first equality. It is clear that $\mu_\infty^T(f_j, f_j)$ defines a positive measure on X; thus $\mu_\infty(\varphi_0, \delta)$, initially defined as an (algebraic) functional on $C_c^\infty(X_Z)_K$, extends to a positive measure on X. To obtain the slightly stronger conclusion implicit in (2) of Theorem 1.3.2, we will analyze this argument more closely.

COROLLARY 3.3.11. *Notations as in Lemma 3.3.10, there exist a constant $C_{f_1,f_2,f}$ and a seminorm $||\cdot||$ on $C_c^\infty(X_Z)$ such that*

$$\left|\mu_n^T(f_1, f\cdot f_2)(g) - \mu_n^T(\overline{f}\cdot f_1, f_2)(g)\right| \leq C_{f_1,f_2,f}\|g\| \left[\|\tilde{\nu}_\infty - \tilde{\nu}_n\| + \|\nu_n\|^{-1}\right] \tag{3.3.10}$$

PROOF. We keep track of the error term in the proof of of Lemma 3.3.10.

Fix a basis $\{X_i\}$ for $\mathfrak{g} = \mathfrak{g}_{\mathrm{ss}} \oplus Z_{\mathfrak{g}}$, and define a seminorm on $C_c^\infty(X_Z)$ by $\|g\| = \|g\|_{L^\infty(X_Z)} + \sum_i \|X_i g\|_{L^\infty(X_Z)}$. With this seminorm, (3.3.10) holds for $f_1, f_2 \in V_K$ and $f = p_X$. This follows from (3.3.8), utilizing Lemma 3.3.5 and the fact that $\|p_X - p_n\|_{L^\infty(M\backslash K)} \ll \|\tilde{\nu}_\infty - \tilde{\nu}_n\|$.

Next suppose $f_1, f_2, f, f' \in V_K$ and $\alpha, \alpha' \in \mathbb{C}$. Then, if (3.3.10) is valid for (f_1, f_2, f) and (f_1, f_2, f'), it is also valid for $(f_1, f_2, \alpha f + \alpha' f')$. Further, if (3.3.10) is valid for $(f_1, f' \cdot f_2, f)$ and for $(\overline{f} f_1, f_2, f')$, then it is also valid for $(f_1, f_2, f \cdot f')$.

Consider now the set of $f \in V_K$ for which (3.3.10) holds for all $f_1, f_2 \in V_K$. The remarks above show that this is a subalgebra of V_K that contains each p_X. The Corollary then follows from the equality $\mathcal{F} = L^2(M\backslash K)_K$ established in the Lemma. $\square$

REMARK 3.3.12. It is possible to obtain a bound of the form $C_{f_1,f_2,f,\tilde{\nu}_n}\|g\|\|\nu_n\|^{-1}$, with the constant uniformly bounded if the $\tilde{\nu}_n$ are uniformly bounded away from the walls. This result can be used to avoid passing to a subsequence in Theorem 1.3.2 or the following Proposition; this is unnecessary for our applications, however.

PROPOSITION 3.3.13. *(Positivity and equivariance: (2) and (4) of Theorem 1.3.2).*

Let $\{\psi_n\}$ be non-degenerate. After replacing $\{\psi_n\}$ by an appropriate subsequence, there exist functions $\tilde{\psi}_n$ on X with the following properties:

(1) *Define the measure σ_n via the rule $\sigma_n(g) = \int_X g(x)|\tilde{\psi}_n(x)|^2 dx$. Then, for each $g \in C_c^\infty(X_Z)_K$ we have $\lim_{n\to\infty}(\sigma_n(g) - \mu_n(g)) = 0$.*

(2) *Let* $E \subset \mathrm{End}_G(C^\infty(X_Z))$ *be a* $\mathbb{C}$*-subalgebra of endomorphisms of* $C^\infty(X_Z)$*, commuting with the* G*-action. Note that each* $e \in E$ *induces an endomorphism of* $C^\infty(Y)$*. Assume in addition that* ψ_n *is an eigenfunction for* E*. Then we may choose* $\tilde{\psi}_n$ *so that each* $\tilde{\psi}_n$ *is an eigenfunction for* E *with the same eigenvalues as* ψ_n*.*

PROOF. Without loss of generality we may assume that $\{\psi_n\}$ are conveniently arranged.

Let $\{f_j\}_{j=1}^\infty \subset V_K$ be the sequence of functions provided by Lemma 3.2.5, so that $T(|f_j|^2)$ approximates δ. The main idea is, as in (3.3.9), to approximate $\mu_n = \mu_n(\varphi_0, \delta)$ using $\mu_n^T(f_j, f_j)$.

For any $g \in C_c^\infty(X_Z)_K$ we have:

$$\begin{aligned} \left|\mu_n(g) - \mu_n^T(f_j, f_j)(g)\right| \le & \left|\mu_n(\varphi_0, \delta)(g) - \mu_n(\varphi_0, |f_j|^2)(g)\right| \\ & + \left|\mu_n(\varphi_0, |f_j|^2)(g) - \mu_n(f_j, f_j)(g)\right|. \end{aligned} \tag{3.3.11}$$

Corollary (3.3.11) provides a seminorm $\|\cdot\|$ on $C_c^\infty(X_Z)$ and a constant C_j such that

$$\left|\mu_n(\varphi_0, |f_j|^2)(g) - \mu_n(f_j, f_j)(g)\right| \le C_j ||g|| \cdot \left[\|\tilde{\nu}_n - \tilde{\nu}_\infty\| + \|\nu_n\|^{-1}\right].$$

Choose a sequence of integers $\{j_n\}_{n=1}^\infty$ such that $j_n \to \infty$ and:

$$C_{j_n} \cdot \left[\|\tilde{\nu}_n - \tilde{\nu}_\infty\| + \|\nu_n\|^{-1}\right] \xrightarrow[n\to\infty]{} 0$$

We now estimate the other term on the right-hand side of (3.3.11). Choosing $N = N(g)$ large enough so that $\mu_n(\varphi_0, \delta)(g) = \mu_n(\varphi_0, \delta_N)(g)$, we have

$$\left|\mu_n(\varphi_0, \delta)(g) - \mu_n(\varphi_0, |f_j|^2)(g)\right| \le \left\| |f_j|_N^2 - \delta_N \right\|_{L^2(M\backslash K)} \|g\|_\infty .$$

As $j \to \infty$ (in particular, if $j = j_n$), $|f_j|_N^2 \to \delta_N$ in V_N, so this term tends to zero. It follows that

$$\lim_{n\to\infty} \left|\mu_n(g) - \mu_n^T(f_{j_n}, f_{j_n})(g)\right| = 0. \tag{3.3.12}$$

Setting $\tilde{\psi}_n = R_{\psi_n}(f_{j_n})$, we deduce that

$$\lim_{n\to\infty} \left(\mu_n(g) - \int_{X_Z} |\tilde{\psi}_n|^2 g(x) dx\right) = 0 \tag{3.3.13}$$

holds for every $g \in C_c^\infty(X_Z)_K$. In particular, we obtain (1) of the Proposition.

To obtain the equivariance property note that the representation I_{ν_n} is irreducible as a $(\mathfrak{g}, K)$-module. By [**16**, Corollary 8.11], there exists $u_n \in U(\mathfrak{g})$ such that $I_{\nu_n}(u_n)\varphi_0 = f_{j_n}$. Thus $\tilde{\psi}_n = u_n \psi_n$. Now every $e \in E$ commutes with the right G-action; in particular, $e u_n = u_n e$. It follows that $\tilde{\psi}_n$ transforms under the same character of E as ψ_n. □

3.4. Cartan invariance of quantum limits

In this section we show that a non-degenerate quantum limit μ_∞ is invariant under the action of $A < G$. This invariance follows from differential equations satisfied by the intermediate distributions μ_n. The construction of these differential equations is a purely algebraic problem: construct elements in the $U(\mathfrak{g}_{\mathbb{C}})$-annihilator of $\varphi_0 \otimes \delta \in V_K \otimes V'_K$, where the $U(\mathfrak{g}_{\mathbb{C}})$-action is by $I_\nu \otimes I'_\nu$.

Ultimately, these differential equations are derived from the fact that each $z \in \mathfrak{Z} = \mathfrak{Z}(\mathfrak{g}_{\mathbb{C}})$ acts by a scalar on the representation (V_K, I_{ν_n}). To motivate the method and provide an example, we first work out the simplest case, that of $\mathrm{PSL}_2(\mathbb{R})$, in detail. In this case the resulting operator is due to Zelditch.

This section is written without reference to the central character – assume it to be trivial. Allowing the central character to vary would amount to writing $\mathfrak{Z}(\mathfrak{g}) = \mathfrak{Z}(\mathfrak{g}_{\mathrm{ss}}) \otimes \mathfrak{Z}(Z_{\mathfrak{g}})$ and only working with the first part.

3.4.1. Example of $G = \mathrm{PSL}_2(\mathbb{R})$. Set $G = \mathrm{PSL}_2(\mathbb{R})$, $\Gamma \leq G$ a lattice, and A the subgroup of diagonal matrices. Let H (explicitly given below) be the infinitesimal generator of A, thought of first as a differential operator acting on $X = \Gamma\backslash G$ via the differential of the regular representation. If $\{\psi_n\}$ is a conveniently arranged sequence of eigenfunctions on $\Gamma\backslash G/K$, and μ_n the corresponding distributions (Definition 3.3.3), we will exhibit a second-order differential operator J such that for all $g \in C_c^\infty(X_Z)_K$,

$$\mu_n((H - \frac{J}{r_n})g) = 0, \tag{3.4.1}$$

where $r_n \sim |\lambda_n|^{1/2}$. Since the $\mu_n(Jg)$ are bounded (they converge to $\mu_\infty(Jg)$), we will conclude that $\mu_\infty(Hg) = 0$, in other words that μ_∞ is A-invariant. This operator in equation (3.4.1) is given in [**36**]. Its discovery was motivated by the proof (via Egorov's theorem) of the invariance of the usual microlocal lift under the geodesic flow. We show here how it arises naturally in the representation-theoretic approach.

By Lemma 3.3.7, it will suffice to find an operator annihilating the element $\varphi_0 \otimes \delta \in V_K \otimes V'_K$, where $U(\mathfrak{g}_{\mathbb{C}})$ acts via $I_\nu \otimes I'_\nu$.

Let $H = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}$, $X_+ = \begin{pmatrix} 0 & 1 \\ & 0 \end{pmatrix}$, $X_- = \begin{pmatrix} 0 & \\ 1 & 0 \end{pmatrix}$ be the standard generators of $\mathfrak{SL}_2$, with the commutation relations $[H, X_\pm] = \pm 2X_\pm$, $[X_+, X_-] = H$. The roots w.r.t. the maximal split torus $\mathfrak{a} = \mathbb{R} \cdot H$ are given by $\pm\alpha(H) = \pm 2$. We also set $W = X_+ - X_-$, so that $\mathbb{R}{\cdot}W = \mathfrak{k}$. Letting $+\alpha$ be the positive root, $\mathfrak{n} = \mathbb{R}{\cdot}X_+$, we have $\rho(H) = \frac{1}{2}\alpha(H) = 1$. Set $\exp \mathfrak{a} = A$ as in the introduction.

The Casimir element $C \in \mathfrak{Z}(\mathfrak{SL}_2\mathbb{C})$ is given by $4C = H^2 + 2X_+X_- + 2X_-X_+$. For the parameter $\nu \in i\mathfrak{a}^*$ given by $\nu(H) = 2ir$ ($r \in \mathbb{R}$), C acts on π_ν with the eigenvalue $\lambda = -\frac{1}{4} - r^2$. The Weyl element acts by mapping $\nu \mapsto -\nu$. On $S = G/K$ with the metric normalized to have constant curvature -1, C reduces to the hyperbolic Laplacian. In particular, every eigenfunction $\psi \in L^2(\Gamma\backslash G/K)$ with eigenvalue $\lambda < -\frac{1}{4}$ generates a unitary principal series subrepresentation. Definition 3.3.3 associates to ψ a distribution $\mu_\psi(\varphi_0, \delta)$ on $\Gamma\backslash G$.

As in Definition 3.2.6, we have an action I_ν of G on V and of $\mathfrak{g}$ on V_K. Note that for $g \in NA$, $f \in V_K$, $(I_\nu(g)f)(1) = f(g) = e^{\langle \nu+\rho, H_0(g)\rangle} f(1)$. Since $\delta(f) = f(1)$ and the pairing between V_K and V'_K is G-invariant, it follows that for $X \in \mathfrak{a} \oplus \mathfrak{n}$, $I'_\nu(X)\delta = -\langle \nu + \rho, H_0(X)\rangle \delta$.

Suppressing I_ν from now on this means that $X\cdot(f\otimes\delta) = (Xf)\otimes\delta - \langle \nu+\rho, H_0(X)\rangle\, f\otimes\delta$. Extend $\nu+\rho$ trivially on $\mathfrak{n}$ to obtain a functional on $\mathfrak{a}\oplus\mathfrak{n}$. Then

$$(X + (\nu+\rho)(X))\cdot(f\otimes\delta) = (Xf)\otimes\delta. \tag{3.4.2}$$

Now since $\mathfrak{a}$ normalizes $\mathfrak{n}$ and $\nu+\rho$ is trivial on $\mathfrak{n}$, the map $X\mapsto X+(\nu+\rho)(X)$ is a Lie algebra homomorphism $\mathfrak{a}\oplus\mathfrak{n}\to\mathfrak{a}\oplus\mathfrak{n}$, and hence extends to an algebra homomorphism $\tau_{\nu+\rho}\colon U(\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})\to U(\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})$. (3.4.2) shows that, for $u\in U(\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})$,

$$\tau_{\nu+\rho}(u)\cdot(f\otimes\delta) = (uf)\otimes\delta \tag{3.4.3}$$

In view of (3.4.3) any operator $u\in U(\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})$ annihilating φ_0 gives rise to an operator annihilating $\varphi_0\otimes\delta$.

The natural starting point is the eigenvalue equation $(4C+1+4r^2)\varphi_0 = 0$. Of course, C is not an element of $U(\mathfrak{n}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C})$. Fortunately, it "nearly" is: there exists an $C'\in U(\mathfrak{n}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C})$ such that $C-C'$ annihilates φ_0.

In detail, we use the commutation relations and the fact that $X_- = X_+ - W$ to write $4C = H^2 - 2H + 4X_+^2 - 4X_+W$. Since ϕ_0 is spherical, it follows that $W\phi_0 = 0$. Thus

$$\left(H^2 - 2H + 4X_+^2 + 1 + 4r^2\right)\varphi_0 = 0 \tag{3.4.4}$$

Since $(\nu+\rho)(H) = 2ir+1$, we conclude from (3.4.3) that:

$$\left((H+2ir+1)^2 - 2(H+2ir+1) + 4X_+^2 + 1 + 4r^2\right)\cdot\varphi_0\otimes\delta = 0.$$

Collecting terms in powers of r we see that this may be written as:

$$\left((2H)(2ir) + (H^2+4X_+^2)\right)\varphi_0\otimes\delta = 0$$

Setting $J = \frac{H^2+4X_+^2}{4i}$ and dividing by $4ir$ we see that the operator $H+\frac{J}{r}$ annihilates $\varphi_0\otimes\delta$, and so also the distribution μ_n. One then deduces the A-invariance of μ_∞ as discussed in the start of this section.

Notice that the terms involving r^2 in (3.4.4) canceled. This is a general feature which will be of importance.

3.4.2. The general proof. We now generalize these steps in order. Notations being as in Sections 2.2,3.2 and in Definition 3.3.3, we first compute the action of $U(\mathfrak{m}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})$ on δ (Lemma 3.4.1) and then on $\varphi_0\otimes\delta$ (Corollary 3.4.2). Secondly we find an appropriate form for the elements of $\mathfrak{Z}(\mathfrak{g}_\mathbb{C})$ (Corollary 3.4.5), which gives us the exact differential equation (3.4.6). We then show that the elements we constructed annihilating μ_ψ are (up to scaling) of an appropriate form $H+\frac{J}{\|\nu\|^*}$ (Lemma 3.4.6), and "take the limit as $\nu\to\infty$" (Corollary 3.4.7) to see that μ_∞ is invariant under a sub-torus of A.

A final step (not so apparent in the $\mathrm{PSL}_2(\mathbb{R})$ case) is to verify that we have constructed *enough* differential operators to obtain invariance under the full split torus (Lemma 3.4.8). In fact, even in the rank-1 case one needs to verify that the "H" part is non-zero.

Given $\lambda\in\mathfrak{a}_\mathbb{C}^*$, we extend it to a linear map $\mathfrak{m}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C}\to\mathbb{C}$. Since $\mathfrak{m}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C}$ is an ideal of this Lie algebra, λ is a Lie algebra homomorphism; thus it extends to an algebra homomorphism $\lambda\colon U(\mathfrak{m}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})\to\mathbb{C}$. We denote by τ_λ the translation automorphism of $U(\mathfrak{m}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C})$ given by $X\mapsto X+\lambda(X)$ on $\mathfrak{m}_\mathbb{C}\oplus\mathfrak{a}_\mathbb{C}\oplus\mathfrak{n}_\mathbb{C}$. Similarly, given $\chi\in\mathfrak{h}_\mathbb{C}^*$, we define $\tau_\chi : U(\mathfrak{h}_\mathbb{C}) \to U(\mathfrak{h}_\mathbb{C})$. We shall write $U(\mathfrak{g}_\mathbb{C})^{\leq d}$ for the elements of $U(\mathfrak{g}_\mathbb{C})$ of degree $\leq d$, and similarly for other enveloping algebras and $\mathfrak{Z} = \mathfrak{Z}(\mathfrak{g}_\mathbb{C})$ (e.g. $\mathfrak{Z}^{\leq d} = \mathfrak{Z} \cap U(\mathfrak{g}_\mathbb{C})^{\leq d}$).

Let $\nu \in \mathfrak{a}^*_\mathbb{C}$. Let $\chi_\nu : \mathfrak{Z} \to \mathbb{C}$ be the infinitesimal character corresponding to I_ν (that is, the scalar by which $\mathfrak{Z}$ acts in (I_ν, V_K).) Recall that $\rho_\mathfrak{h}$ denotes the half-sum of positive roots for $(\mathfrak{h}_\mathbb{C} : \mathfrak{g}_\mathbb{C})$, ρ the half-sum for $(\mathfrak{a} : \mathfrak{g})$.

LEMMA 3.4.1. *For* $X \in \mathfrak{m} \oplus \mathfrak{a} \oplus \mathfrak{n}$, $I_\nu(X)\delta = -\langle \nu + \rho, X\rangle \delta$.

PROOF. This follows from the definitions. □

COROLLARY 3.4.2. *For any* $u \in U(\mathfrak{m}_\mathbb{C} \oplus \mathfrak{a}_\mathbb{C} \oplus \mathfrak{n}_\mathbb{C})$ *and* $f \in V_K$,

$$I_\nu \otimes I'_\nu(\tau_{\nu+\rho}(u)) \cdot (f \otimes \delta) = (I_\nu(u)f) \otimes \delta.$$

PROOF. This follows from the previous Lemma. □

REMARK 3.4.3. Denote by $\mathcal{D}_G(G/K)$ the ring of G-invariant differential operators on $S = G/K$. There is an evident homomorphism $U(\mathfrak{g}_\mathbb{C})^K \to \mathcal{D}_G(G/K)$ with kernel $U(\mathfrak{g}_\mathbb{C})^K \cap \mathfrak{k}U(\mathfrak{g}_\mathbb{C})$, see [**12**, 2.6] or [**5**, 9.2]. We recall that "projection to $U(\mathfrak{a}_\mathbb{C})$" under the Poincaré-Birkhoff-Witt isomorphism $U(\mathfrak{g}_\mathbb{C}) = U(\mathfrak{n}_\mathbb{C}) \otimes U(\mathfrak{a}_\mathbb{C}) \otimes U(\mathfrak{k}_\mathbb{C})$ descends (after composing with an appropriate translation on $U(\mathfrak{a}_\mathbb{C})$) to an isomorphism of $\mathcal{D}_G(G/K)$ with $U(\mathfrak{a}_\mathbb{C})^{W(\mathfrak{g}:\mathfrak{a})}$. We shall need some very slightly refined information about this decomposition. There is an evident map $\mathfrak{Z} \to \mathcal{D}_G(G/K)$ and it will suffice for our purpose to understand the decomposition on the image of $Z(\mathfrak{g}_\mathbb{C})$. (Although we do not need this, the map from $\mathfrak{Z} \to \mathcal{D}_G(G/K)$ is in most cases nearly surjective; in all cases the quotient field of the image coincides with the quotient field of $\mathcal{D}_G(G/K)$, c.f. [**13**, 3.16].)

DEFINITION. Let $\mathrm{pr} : U(\mathfrak{g}_\mathbb{C}) \to U(\mathfrak{h}_\mathbb{C})$ be the projection corresponding to the decomposition $U(\mathfrak{g}_\mathbb{C}) = U(\mathfrak{h}_\mathbb{C}) \oplus [(\mathfrak{n}_\mathbb{C} \oplus \mathfrak{n}_M)U(\mathfrak{g}_\mathbb{C}) + U(\mathfrak{g}_\mathbb{C})(\bar{\mathfrak{n}}_\mathbb{C} \oplus \bar{\mathfrak{n}}_M)]$ (arising from the decomposition $\mathfrak{g}_\mathbb{C} = \mathfrak{n}_\mathbb{C} \oplus \mathfrak{n}_M \oplus \mathfrak{h}_\mathbb{C} \oplus \bar{\mathfrak{n}}_\mathbb{C} \oplus \bar{\mathfrak{n}}_M$ by the Poincaré-Birkhoff-Witt Theorem).

LEMMA 3.4.4. *For* $z \in \mathfrak{Z}^{\leq d}$, *we have*

$$z - \mathrm{pr}(z) \in U(\mathfrak{n}_\mathbb{C})U(\mathfrak{a}_\mathbb{C})^{\leq d-2}U(\mathfrak{k}_\mathbb{C}).$$

PROOF. It suffices to show that $z - \mathrm{pr}(z) \in U(\mathfrak{n}_\mathbb{C})U(\mathfrak{g}_\mathbb{C})^{\leq d-2}U(\mathfrak{k}_\mathbb{C})$, since $\mathfrak{g}_\mathbb{C} = \mathfrak{n}_\mathbb{C} \oplus \mathfrak{a}_\mathbb{C} \oplus \mathfrak{k}_\mathbb{C}$.

Let $\mathcal{B}(\mathfrak{n}_\mathbb{C})$, $\mathcal{B}(\bar{\mathfrak{n}}_\mathbb{C})$, $\mathcal{B}(\mathfrak{n}_M)$ and $\mathcal{B}(\bar{\mathfrak{n}}_M)$ be bases for $\mathfrak{n}_\mathbb{C}$, $\bar{\mathfrak{n}}_\mathbb{C}$, $\mathfrak{n}_M$ and $\bar{\mathfrak{n}}_M$, respectively, consisting of $\mathfrak{h}_\mathbb{C}$-eigenvectors. Let $\mathcal{B}(\mathfrak{a}_\mathbb{C})$ and $\mathcal{B}(\mathfrak{b}_\mathbb{C})$ be bases for $\mathfrak{a}_\mathbb{C}$ and $\mathfrak{b}_\mathbb{C}$, respectively.

By Poincaré-Birkhoff-Witt, one may uniquely express z as a linear combination of terms of the form:

$$\mathcal{D} = X_1 \dots X_n Y_1 \dots Y_m A_1 \dots A_t B_1 \dots B_r \overline{X}_1 \dots \overline{X}_k \overline{Y}_1 \dots \overline{Y}_l$$

where $X_* \in \mathcal{B}(\mathfrak{n}_\mathbb{C})$, $Y_* \in \mathcal{B}(\mathfrak{n}_M)$, $A_* \in \mathcal{B}(\mathfrak{a}_\mathbb{C})$, $B_* \in \mathcal{B}(\mathfrak{b}_\mathbb{C})$,$\overline{X}_* \in \mathcal{B}(\bar{\mathfrak{n}}_\mathbb{C})$ and $\overline{Y}_* \in \mathcal{B}(\bar{\mathfrak{n}}_M)$. Then $z - \mathrm{pr}(z)$ consists of the sum of all terms $\mathcal{D}$ for which $n + m + k + l \neq 0$. We show that each such term satisfies $\mathcal{D} \in U(\mathfrak{n}_\mathbb{C})U(\mathfrak{g}_\mathbb{C})^{\leq d-2}U(\mathfrak{k}_\mathbb{C})$.

In view of the fact that $z - \mathrm{pr}(z)$ commutes with $\mathfrak{a}_\mathbb{C}$, one has $n = 0$ iff $k = 0$. Further, if $n = k = 0$, then the fact that $z - \mathrm{pr}(z)$ commutes with $\mathfrak{b}_\mathbb{C}$ implies $m = 0$ iff $l = 0$. Also one has $n + m + t + r + k + l \leq d$.

We now proceed in a case-by-case basis, using either the inclusion $\mathfrak{n}_M \oplus \mathfrak{b}_{\mathbb{C}} \oplus \bar{\mathfrak{n}}_M = \mathfrak{m}_{\mathbb{C}} \subset \mathfrak{k}_{\mathbb{C}}$, or the observation that for $\overline{X} \in \bar{\mathfrak{n}}_{\mathbb{C}}$ we have $\theta_{\mathbb{C}}\overline{X} \in \mathfrak{n}_{\mathbb{C}}$, while $\overline{X} + \theta_{\mathbb{C}}\overline{X} \in \mathfrak{k}_{\mathbb{C}}$ (it is $\theta_{\mathbb{C}}$-stable!).

(1) $k = l = 0$ is impossible, for this would force $n = m = 0$.
(2) $k \geq 1$ and $l \geq 1$. Then $n \geq 1$ so that $X_1 \dots X_n \in U(\mathfrak{n}_{\mathbb{C}})$, $\overline{Y}_1 \dots \overline{Y}_l \in U(\mathfrak{k}_{\mathbb{C}})$, and $m + t + r + k \leq d - 2$.
(3) $k = 0$ and $l \geq 1$. Then $n = 0$ and $m \geq 1$, so $t \leq d - 2$. Since $[\mathfrak{a}, \mathfrak{m}] = 0$ we may commute the A-terms past the Y-terms, so that $\mathcal{D}$ is the product of the A-terms (at most $d-2$ of them) and $Y_1 \dots Y_m B_1 \dots B_r \overline{Y}_1 \dots \overline{Y}_l \in U(\mathfrak{k}_{\mathbb{C}})$.
(4) $k \geq 1$ and $l = 0$. Then $n \geq 1$. Set $s = Y_1 \dots Y_m A_1 \dots A_t B_1 \dots B_r \overline{X}_1 \dots \overline{X}_{k-1}$ so that $\mathcal{D} = X_1 \dots X_n \cdot s \cdot \overline{X}_k$. Since $m + t + r + (k-1) \leq d - 1 - n \leq d - 2$, we have $s \in U(\mathfrak{g}_{\mathbb{C}})^{\leq d-2}$. Then (recall $\theta_{\mathbb{C}}$ is the complex-linear extension of the Cartan involution θ to $\mathfrak{g}_{\mathbb{C}}$),

$$\begin{aligned} \mathcal{D} = X_1 \dots X_n s \overline{X}_k &= X_1 \dots X_n \cdot s \cdot (\overline{X}_k - \theta_{\mathbb{C}}(\overline{X}_k)) \\ &+ X_1 \dots X_n \theta_{\mathbb{C}}(\overline{X}_k) s \\ &+ X_1 \dots X_n (s\theta_{\mathbb{C}}(\overline{X}_k) - \theta_{\mathbb{C}}(\overline{X}_k) s). \end{aligned} \tag{3.4.5}$$

From the observation above, the first two terms on the right clearly belong to $U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{g}_{\mathbb{C}})^{\leq d-2}U(\mathfrak{k}_{\mathbb{C}})$. Moreover, $\left[s, \theta_{\mathbb{C}}(\overline{X}_k)\right] \in U(\mathfrak{g}_{\mathbb{C}})^{\leq d-2}$ (the general fact that $[p, q] \in U(\mathfrak{g}_{\mathbb{C}})^{d_p + d_q - 1}$ whenever $p \in U(\mathfrak{g}_{\mathbb{C}})^{\leq d_p}$, $q \in U(\mathfrak{g}_{\mathbb{C}})^{\leq d_q}$ follows by induction on the degrees from the formula $[ab, c] = a[b, c] + [a, c]b$). Thus the third term of (3.4.5) belongs to $U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{g}_{\mathbb{C}})^{\leq d-2}U(\mathfrak{k}_{\mathbb{C}})$ also.

□

COROLLARY 3.4.5. *Let $z \in \mathfrak{Z}^{\leq d}$. Then there exists $b = b(z) \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})^{\leq d-2}$ such that $z - \mathrm{pr}(z) + b(z) \in U(\mathfrak{g}_{\mathbb{C}}) \cdot \mathfrak{k}_{\mathbb{C}}$.*

Since $I_\nu(\mathfrak{k}_{\mathbb{C}})$ annihilates φ_0 and $z \cdot \varphi_0 = \chi_\nu(z)\varphi_0$, we have $I_\nu(\chi_\nu(z) - \mathrm{pr}(z) + b(z)) \cdot \varphi_0 = 0$. In view of Corollary 3.4.2inf} we obtain:

$$I_\nu \otimes I'_\nu(\tau_{\nu+\rho}\,\mathrm{pr}(z) - \tau_{\nu+\rho} b(z) - \chi_\nu(z))(\varphi_0 \otimes \delta) = 0 \tag{3.4.6}$$

In what follows, we shall freely identify the algebra $U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$ with the Weyl-invariant polynomial functions on $\mathfrak{h}^*_{\mathbb{C}}$.

Given $\mathcal{P} \in U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$, we denote by $\mathcal{P}' \colon \mathfrak{h}^*_{\mathbb{C}} \to \mathfrak{h}_{\mathbb{C}}$ its differential. In other words, we identify $\mathcal{P}$ with a polynomial function on $\mathfrak{h}^*_{\mathbb{C}}$, and $\mathcal{P}'$ denotes the derivative of this function; it takes values in the cotangent space of $\mathfrak{h}^*_{\mathbb{C}}$, which is canonically identified at every point with $\mathfrak{h}_{\mathbb{C}}$.

We shall use the notation $U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq r}$ to denote polynomials of degree $\leq r$ on $\mathfrak{a}^*_{\mathbb{C}}$, valued in the vector space $U(\mathfrak{g}_{\mathbb{C}})$. Note that given $J \in U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq r}$ and $\nu \in \mathfrak{a}^*_{\mathbb{C}}$ we can speak of the "value of J at ν." We denote it by $J(\nu)$ and it belongs to $U(\mathfrak{g}_{\mathbb{C}})$.

LEMMA 3.4.6. *Let $\mathcal{P} \in U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$ have degree $\leq d$. Set $H = \frac{\mathcal{P}'(\nu)}{\|\nu\|^{d-1}} \in \mathfrak{h}_{\mathbb{C}}$. Then there exists $J \in U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq d-2}$ such that*

$$I_\nu \otimes I'_\nu \left(H + \frac{J(\nu)}{\|\nu\|^{d-1}} \right) \cdot \varphi_0 \otimes \delta = 0.$$

(As defined in Section 2.2, $\|\nu\|$ denotes the norm of $\nu \in \mathfrak{a}_{\mathbb{C}}^*$ w.r.t. the Killing form.)

PROOF. The map $\gamma_{HC} \colon \mathfrak{Z} \to U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$ given by $\gamma_{HC}(z) = \tau_{\rho_{\mathfrak{h}}} \operatorname{pr}(z)$ is an isomorphism of algebras, the Harish-Chandra homomorphism. With the above identification, the infinitesimal character of (V_K, I_ν) corresponds to "evaluation at $\nu + \rho - \rho_{\mathfrak{h}}$," i.e. for $\mathcal{P} \in U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$:

$$\chi_\nu(\gamma_{HC}^{-1}(\mathcal{P})) = \mathcal{P}(\nu + \rho_{\mathfrak{a}} - \rho_{\mathfrak{h}}) \tag{3.4.7}$$

(See [**16**, Prop 8.22]; w.r.t. the maximal torus $\mathfrak{b}_{\mathbb{C}} \subset \mathfrak{m}_{\mathbb{C}}$, the infinitesimal character of the trivial representation of $\mathfrak{m}_{\mathbb{C}}$ is (the Weyl-group orbit of) $\rho - \rho_{\mathfrak{h}}$).

Given $\mathcal{P} \in U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$ of degree d, we set $z = \gamma_{HC}^{-1}(\mathcal{P})$ in (3.4.6), writing $b(\mathcal{P})$ for the element $b(z)$. Note that $z \in Z(\mathfrak{g}_{\mathbb{C}})^{\leq d}$, as the Harish-Chandra homomorphism "preserves degree" (see [**5**, 7.4.5(c)]), and hence $b(\mathcal{P}) \in U(\mathfrak{n}_{\mathbb{C}})U(\mathfrak{a}_{\mathbb{C}})^{\leq d-2}$.

Combining (3.4.6) and (3.4.7): $\varphi_0 \otimes \delta$ is then annihilated by the operator

$$\left(\tau_{\nu+\rho-\rho_{\mathfrak{h}}}\mathcal{P} - \mathcal{P}(\nu + \rho_{\mathfrak{a}} - \rho_{\mathfrak{h}}) - \tau_{\nu+\rho} b(\mathcal{P})\right) \varphi_0 \otimes \delta = 0 \tag{3.4.8}$$

Let $\underline{x} = (x_1, \ldots, x_n), \underline{y} = (y_1, \ldots, y_n)$. If a polynomial $p \in \mathbb{C}[\underline{x}]$ has degree d, $p(\underline{x} + \underline{y}) - p(\underline{y}) = p'(\underline{y})(\underline{x}) + q(\underline{x}, \underline{y})$ where $q \in (\mathbb{C}[\underline{x}])\,[\underline{y}]$ has degree at most $d-2$ in $\underline{y}$, and the derivative $p'(\underline{y})$ is understood to act as a linear functional on $\underline{x}$.

Applying this to $p = \mathcal{P}, \underline{y} = \nu + \rho - \rho_{\mathfrak{h}}$ we see that there exists $J_1 \in U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq d-2}$ with $\deg(J) \leq d-2$ and

$$\tau_{\nu+\rho-\rho_{\mathfrak{h}}}\mathcal{P} - \mathcal{P}(\nu + \rho - \rho_{\mathfrak{h}}) = \mathcal{P}'(\nu + \rho - \rho_{\mathfrak{h}}) + J_1(\nu) \tag{3.4.9}$$

Now $b(\mathcal{P}) \in U(\mathfrak{n}_{\mathbb{C}}) \cdot U(\mathfrak{a}_{\mathbb{C}})^{\leq d-2}$, so the map $\nu \mapsto \tau_{\nu+\rho} b(\mathcal{P})$ can be regarded as an element $J_2 \in U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq d-2}$. Similarly $\nu \mapsto \mathcal{P}'(\nu + \rho - \rho_{\mathfrak{h}}) - \mathcal{P}'(\nu)$ defines an element $J_3 \in U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq d-2}$.

Combining these remarks with (3.4.8) and (3.4.9), we see that

$$(\mathcal{P}'(\nu) + J_1(\nu) + J_2(\nu) + J_3(\nu))\varphi_0 \otimes \delta = 0$$

Set $J \equiv J_1 + J_2 + J_3$ and divide by $\|\nu\|^{d-1}$ to conclude. □

COROLLARY 3.4.7. *Let $\mathcal{P} \in U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}}:\mathfrak{g}_{\mathbb{C}})}$. Notations being as in Definition 3.3.8 and Lemma 3.3.9, suppose $\{\psi_n\}$ is conveniently arranged. Then $\mu_\infty(\varphi_0, \delta)$ is $\mathcal{P}'(\tilde{\nu}_\infty)$-invariant.*

PROOF. It suffices to verify this for $\mathcal{P}$ homogeneous, say of degree d. Combining Lemma 3.4.6 and Lemma 3.3.7, and using the homogeneity of $\mathcal{P}$, we see that there exists $J \in U(\mathfrak{g}_{\mathbb{C}})[\mathfrak{a}_{\mathbb{C}}]^{\leq d-2}$ so that

$$\left(\mathcal{P}'(\tilde{\nu}_n) + \frac{J(\nu_n)}{||\nu_n||^{d-1}} \right) \cdot \mu_n(\varphi_0 \otimes \delta) = 0$$

Here $(\mathcal{P}' + \dots)$ acts on $\mu_n(\varphi_0 \otimes \delta)$ according to the natural action of $U(\mathfrak{g}_{\mathbb{C}})$ on $C_c^\infty(X_Z)'_K$. Now fix $g \in C_c^\infty(X_Z)_K$. Let $u \to u^t$ be the unique $\mathbb{C}$-linear anti-involution of $U(\mathfrak{g}_{\mathbb{C}})$ such that $X^t = -X$ for $X \in \mathfrak{g}_{\mathbb{C}} \subset U(\mathfrak{g}_{\mathbb{C}})$. Then we have for each d

$$\mu_n(\varphi_0 \otimes \delta)\left(\left(\mathcal{P}'(\tilde{\nu}_n) - \frac{J^t(\nu_n)}{\|\nu_n\|^{d-1}}\right) g\right) = 0. \tag{3.4.10}$$

Note that, as n varies, the quantity $\left(\mathcal{P}'(\tilde{\nu}_n) - \frac{J^t(\nu_n)}{\|\nu_n\|^{d-1}}\right) g$ remains in a fixed finite dimensional subspace of $C^\infty(X)_K$. Further, it converges in that subspace to $\mathcal{P}'(\tilde{\nu}_\infty)g$.

With these remarks in mind, we can pass to the limit $n \to \infty$ in (3.4.10) to obtain $\mu_\infty(\varphi_0 \otimes \delta)(\mathcal{P}'(\nu_\infty)g) = 0$, i.e. $\mathcal{P}'(\nu_\infty)$ annihilates μ_∞ as required. □

It remains to show that the subspace

$$S = \left\{\mathcal{P}'(\tilde{\nu}_\infty) \mid \mathcal{P} \in U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})}\right\} \subset \mathfrak{h}_{\mathbb{C}} \tag{3.4.11}$$

contains $\mathfrak{a}_{\mathbb{C}}$. By the Corollary this will show that $\mathfrak{a}$ annihilates any limit measure, or that this measure is A-invariant.

LEMMA 3.4.8. *Let $W_0 \subset W(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$ be the stabilizer of $\tilde{\nu}_\infty \in \mathfrak{a}_{\mathbb{C}}^*$, and define S as in (3.4.11). Then $S = \mathfrak{h}_{\mathbb{C}}^{W_0}$. In particular, if $\tilde{\nu}_\infty$ is regular, then S contains $\mathfrak{a}_{\mathbb{C}}$.*

PROOF. This can be seen either from the fact that S is the image of the map on cotangent spaces induced by the quotient map $\mathfrak{h}_{\mathbb{C}}^* \to \mathfrak{h}_{\mathbb{C}}^*/W_0$, or more explicitly: first construct many elements in $U(\mathfrak{h}_{\mathbb{C}})^{W(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})}$ by averaging over $W(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$, and then directly compute derivatives to obtain the claimed equality.

W_0 is generated by the reflections in $W(\mathfrak{h}_{\mathbb{C}} : \mathfrak{g}_{\mathbb{C}})$ fixing $\tilde{\nu}_\infty$. In the case where $\tilde{\nu}_\infty$ is regular as an element in $i\mathfrak{a}_{\mathbb{R}}^*$, the corresponding roots must be trivial on all of $\mathfrak{a}_{\mathbb{C}}^*$. In particular, any element of W_0 fixes all of $\mathfrak{a}_{\mathbb{C}}$. □

COROLLARY 3.4.9. *Let notations be as in Proposition 3.3.13. Then any weak-* limit σ_∞ of the measures σ_n is A-invariant.*

PROOF. After passing to an appropriate subsequence, we may assume that $\{\psi_n\}$ is conveniently arranged. Proposition 3.3.13, (1), shows that $\sigma_\infty(g) = \mu_\infty(\varphi_0, \delta)(g)$ whenever $g \in C_c^\infty(X_Z)_K$. Corollary 3.4.7 and Lemma 3.4.8, together with the fact that $C_c^\infty(X_Z)_K$ is dense in $C_0(X)$, show that σ_∞ is A-invariant. □

CHAPTER 4

The Method of Hecke Translates I: Tubular Neighbourhood, Translates, and Diophantine Geometry

4.1. Overview

In the previous chapter we have seen that a Maass form cannot be too concentrated – its associated measure can be approximately lifted to a measure which is approximately A-invariant (in the non-degenerate case). This chapter and the next one are devoted to showing a similar conclusion for Hecke eigenforms. We hence fix a unitary Hecke character $\omega \in \hat{Z}$ and a normalized Hecke eigenfunction $\psi \in L^2(X, \omega)$, and let μ denote the measure $\mu_\psi(f) = \int_{X_Z} f\,|\psi|^2\,dx$ on X_Z. We will of course take ψ to be one of the functions $\tilde{\psi}_n$ of Theorem 1.3.2. As before, π will denote any of the quotient maps $G \twoheadrightarrow X \twoheadrightarrow X_Z$.

In Section 4.2 we will describe certain relatively compact open sets $B_a(C,\epsilon) \subset G$ to be called "tubular neighbourhoods", depending on a parameter ϵ, their "width" , which will tend to zero. We will also use the term for sets of the form $xB_a(C,\epsilon) \subset X_Z$ where $x \in X_Z$.

Using the isomorphism $X_Z = \Gamma Z\backslash G \simeq \mathbb{G}(\mathbb{Q})Z\backslash G \times \mathbb{G}(\mathbb{A}_\mathrm{f})/K_\mathrm{f}$, any pair $g_\infty \in G$ and $g_\mathrm{f} \in \mathbb{G}(\mathbb{A}_\mathrm{f})$ gives rise to a relatively compact open subset

$$g_\infty B_a(C,\epsilon) g_\mathrm{f} \overset{\text{def}}{=} \{Z\mathbb{G}(\mathbb{Q})g_\infty b g_\mathrm{f} K_\mathrm{f} \mid b \in B_a(C,\epsilon)\} \subset X_Z,$$

which we will call a *Hecke translate* of the tubular neighbourhood $xB_a(C,\epsilon)$ of $x = Z\Gamma g_\infty$. Writing $g_\mathrm{f} = \gamma k_\mathrm{f}$ for some $\gamma \in \Gamma$, $k_\mathrm{f} \in K_\mathrm{f}$, we see that $g_\infty B_a(C,\epsilon) = x' B_a(C,\epsilon)$ where $x' = Z\Gamma(\gamma_\infty^{-1} g_\infty)$. In other words, a Hecke translate of a tubular neighbourhood is again a tubular neighbourhood of the same type.

In the next chapter we will analyze the behaviour of ψ on a large number of such translates to bound the measures $\mu_\psi(xB_a(C,\epsilon))$ and show they must decay as a power of ϵ ("positive entropy for the action of a"). We would like to choose a disjoint set of translates, and hence we need to understand the intersection pattern of such a set. The main result of this chapter (Theorem 4.4.4 of Section 4.4) is the first step in that direction, showing that under certain conditions on $S \subset \mathbb{G}(\mathbb{A}_\mathrm{f})$ there exists a proper $\mathbb{Q}$-subalgebra of our division algebra which controls the intersection pattern of the Hecke translates $\{xB_a(C,\epsilon)g_\mathrm{f}\}_{g_\mathrm{f} \in S}$. Since we assume $\mathbb{D}$ to be of prime degree d, this proper subalgebra will then be commutative, i.e. a number field of dimension d contained in $\mathbb{D}(\mathbb{Q})$.

The key idea is the analysis in section 4.3 of the polynomial nature of the condition "the subalgebra of $\mathbb{D}(\mathbb{Q})$ generated by the subset $U \subset \mathbb{D}(\mathbb{Q})$ is proper".

For a very concrete version the ideas of this chapter (in the case $d = 2$) the see Lemmata 3.1, 3.2 and 3.3 of [**3**].

4.2. Algebra in tubular neighbourhoods

Let $a \in A \setminus Z$. We then set $\mathfrak{n}_a \subset \mathfrak{n}$ be the Lie subalgebra spanned by the root spaces $\mathfrak{n}_\alpha$ associated to roots α such that $\alpha(a) > 0$. This is the unipotent radical of a parabolic subalgebra associated to that set of roots. The Levi factor is $M_a A_a = Z_G(a)$, where $A_a = Z(M_a A_a) = \cap_{\alpha(a)=0} \operatorname{Ker}(\alpha)$ is its center. We then set $\bar{\mathfrak{n}}_a = \overline{\mathfrak{n}_a}$, the unipotent radical of the opposite parabolic subgroup.

For any $\epsilon > 0$ let $\mathfrak{n}_a^\epsilon = \{X \in \mathfrak{n}_a \mid \|X\| < \epsilon\}$ (note that $\bar{\mathfrak{n}}_a^\epsilon = \theta \mathfrak{n}_a^\epsilon$), and set $N_a(\epsilon) = \exp \mathfrak{n}_a^\epsilon$, $\bar{N}_a(\epsilon) = \Theta N_a(\epsilon) = \exp \bar{\mathfrak{n}}_a^\epsilon$. For any relatively compact symmetric neighbourhood of the identity $C \subset A_a M_a$ and every $\epsilon > 0$, we set $B_a(C, \epsilon) = \bar{N}_a(\epsilon) \cdot C \cdot N_a(\epsilon) \subset G$. We call $\tilde{C} \subset G$ a *neighbourhood* of C if $\tilde{C}$ is neighbourhood of every $c \in \bar{C}$ (topological closure).

DEFINITION 4.2.1. We call a set of the form $B_a(C, \epsilon)$ a *tubular neighbourhood* of the *piece* $C \subset A_a M_a$.

Thinking of ϵ as small (C will be fixed), the elements of $B_a(C, \epsilon)$ are very close to lying on $A_a M_a$. Since $A_a M_a$ is a subgroup, one expects that a set of the form $B_a(C, \epsilon) B_a(C', \epsilon')$ also consists of elements close to a neighbourhood of the identity of $A_a M_a$.

LEMMA 4.2.2. *(c.f.* [**3**, Lemma 3.2]*) For any C, C', any relatively compact symmetric neighbourhoods $C \subset \tilde{C}$ resp. $CC' \subset \tilde{C}$, and any ϵ, ϵ' small enough w.r.t. the choice of $C, C', \tilde{C}$ we have:*

$$B_a(C, \epsilon)^{-1} \subset B(\tilde{C}, O_C(\epsilon))$$

resp.

$$B_a(C, \epsilon) B_a(C', \epsilon') \subset B_a(\tilde{C}, O_{C,C'}(\epsilon + \epsilon')).$$

PROOF. This is a direct computation. We only prove the first assertion.

Since the adjoint action is differentiable and $B(C, 1)$ is a relatively compact subset of G, there exists a constant r_C such that $\|\operatorname{Ad}(m) X\| \leq r_C \|X\|$for any $m \in B(C, 1)$ and $X \in \mathfrak{g}$. Secondly, for any (vector-space) direct sum decomposition $\mathfrak{g} = \oplus_i V_i$ the map $(X_i)_i \mapsto \prod_i \exp X_i$ is a local diffeomorphism. Thus there exists $\delta, r' > 0$ such that if $X_1, X_2, Y \in \mathfrak{g}$ are all of norms$\leq \delta$ then there exists $X' \in \mathfrak{g}$ and $(Y_1, Y_2, Y_3) \in \bar{\mathfrak{n}}_a \oplus \mathfrak{m}_a \oplus \mathfrak{n}_a \simeq \mathfrak{g}$ such that $\|X'\| \leq r' (\|X_1\| + \|X_2\|)$, $\|Y_i\| \leq r' \|Y\|$ and $\exp X' = \exp X_1 \exp X_2$ and $\exp Y = \exp Y_1 \exp Y_2 \exp Y_3$ (the existence of X' follows from the smoothness of the multiplication operation in the co-ordinate system $\exp : \mathfrak{g} \to G$, the existence of Y_i from the equivalence of that co-ordinate system with the one induced from the direct sum decomposition). We also observe that if $C_1 \subset G$ is any relatively compact subset, and $C_2 \supset C_1$ is a neighbourhood then for ϵ small enough we have $C_1 \exp \{X \in \mathfrak{g} \mid \|X\| \leq \epsilon\} \subset C_2$.

Assume that $r_C \epsilon \leq 1$, $r_C \epsilon \leq \delta$, $r' r_C \epsilon \leq \delta$. As the sets $\mathfrak{n}_a^\epsilon$, $\bar{\mathfrak{n}}_a^\epsilon$ are symmetric, so are $N_a(\epsilon)$ and $\bar{N}_a(\epsilon)$, so for the first assertion it suffices to show that if $b = \bar{n} m n \in B_a(C, \epsilon)$ then $n m \bar{n} \in B_a(\tilde{C}, O_C(\epsilon))$. For this write $n = \exp X$, $\bar{n} = \exp Y$ with $\|X\|, \|Y\| \leq \epsilon$ and write $n' = m^{-1} n m$ so that:

$$n m \bar{n} = m n' \bar{n} = m n' \bar{n} n'^{-1} n'.$$

We next note that $n' = \exp(\mathrm{Ad}(m^{-1})X) \in B(C,1)$ since $\|\mathrm{Ad}(m)X\| \le r_C \|X\| \le 1$. Since $n'\bar{n}n'^{-1} = \exp(\mathrm{Ad}(n')Y)$ we conclude $\|\mathrm{Ad}(n')Y\| \le r_C \|Y\| \le r_C\epsilon$. We can thus write $\exp(\mathrm{Ad}(n')Y) = \exp(Y_1)\exp(Y_2)\exp(Y_3)$ with $Y_1 \in \bar{\mathfrak{n}}_a$, $Y_2 \in \mathfrak{m}_a$ and $Y_3 \in \mathfrak{n}_a$ and $\|Y_i\| \le r' r_C \epsilon$. It follows that:

$$nm\bar{n} = \exp(\mathrm{Ad}(m)Y_1)\,(m\exp(Y_2))\,(\exp(Y_3)\exp(\mathrm{Ad}(m)X))\,.$$

Finally, choosing ϵ small enough will insure that $m\exp(Y_2) \in \tilde{C}$, while $\|\mathrm{Ad}(m)Y_2\| \le r' r_C^2 \epsilon$ and $\exp(Y_3)\exp(\mathrm{Ad}(m)X) = \exp X'$ for some $X' \in \mathfrak{n}_a$ such that $\|X'\| \le r'(1+r')r_C\epsilon$. □

The next Lemma formalizes the notion that the $xB_a(C,\epsilon)$ are "tubular neighbourhoods" of the pieces xC, uniformly over compacta in G:

LEMMA 4.2.3. *Let $\Omega_\infty \subset G$ be compact. Then there exists a constant $c > 0$ (depending on Ω_∞, C) such that for every $x_\infty \in \Omega_\infty$ and small enough $\epsilon > 0$, if $g \in x_\infty B_a(C,\epsilon)x_\infty^{-1}$ then there exists $g_\infty \in x_\infty C x_\infty^{-1} \subset x_\infty A_a M_a x_\infty^{-1}$ with $|x_i(g_\infty) - x_i(g)| \le c\epsilon$ for every i.*

PROOF. We know that $g = x_\infty n\tilde{g}\bar{n}x_\infty^{-1}$ with $\tilde{g} \in C$, $n \in N_a(\epsilon)$ and $\bar{n} \in \bar{N}_a(\epsilon)$. Now set $g_\infty = x_\infty \tilde{g} x_\infty^{-1}$ and consider the maps $t_i \colon \mathfrak{n}_a \times \bar{\mathfrak{n}}_a \times G \to \mathbb{R}$ given by

$$(X, \bar{X}, x_\infty) \mapsto x_i\left(\exp\left(\mathrm{Ad}(x_\infty)X\right) g_\infty \exp\left(\mathrm{Ad}(x_\infty)\bar{X}\right)\right).$$

Being the composition of smooth maps this is one as well; in particular it is continuously differentiable on a relatively compact open neighbourhood of $\mathfrak{n}_a^1 \times \bar{\mathfrak{n}}_a^1 \times \Omega_\infty$. We can hence find $c > 0$ such that for $X \in \mathfrak{n}_a^1$, $\bar{X} \in \bar{\mathfrak{n}}_a^1$ $\left|t_i(X,\bar{X},x_\infty) - t_i(0,0,x_\infty)\right| \le c\max\left\{\|X\|, \left\|\bar{X}\right\|\right\}$. Since we have $g_\infty = t(0,0,x_\infty)$ and $g = t(X,\bar{X},x_\infty)$ for some $X \in \mathfrak{n}_a^\epsilon$, $\bar{X} \in \bar{\mathfrak{n}}_a^\epsilon$ we have the desired result. □

4.3. Diophantine Geometry of Division Algebras

We will show that if $I \subset \mathbb{D}(\mathbb{Q})$ are close to a proper $\mathbb{R}$-subalgebra $D'_\mathbb{R} \subset \mathbb{D}(\mathbb{R})$ and have small denominators, they generate a proper $\mathbb{Q}$-subalgebra of $\mathbb{D}$ of dimension at most $\dim_\mathbb{R} D'_\mathbb{R}$. Along the way we will make extensive use the co-ordinates x_i on $\mathbb{D}$ introduced as Notation 2.1.1.

LEMMA 4.3.1. *Let*

$$V_{r,s}(K) = \left\{\underline{x} = \left(x^{(1)},\ldots,x^{(s)}\right) \in \mathbb{D}(K)^s \,\middle|\, \dim_K \mathrm{Sp}_K \underline{x} \le r\right\},$$

where $\mathrm{Sp}_K\underline{x}$ is the K-subspace of $\mathbb{D}(K) = \mathbb{D}(\mathbb{Q}) \otimes K$ spanned by $\underline{x}$. Then $V_{r,s}(K) = \cap_{p\in F}\{p = 0\}$ where F is a finite family of homogeneous polynomials in the co-ordinates of the $x^{(i)}$, with coefficients in $\mathbb{Q}$ and degrees bounded as a function of d.

PROOF. We need to verify that the statement "the rank of the $s \times d^2$ matrix M is at most r" is equivalent to the joint vanishing of some polynomials in the entries of M. Indeed, M having rank$\le r$ is equivalent to the vanishing of the determinant of every $(r+1)\times(r+1)$ minor of M, and the coefficients of the polynomials are in fact ± 1. □

LEMMA 4.3.2. *For each $1 \leq k \leq d^2$ and any field K extending $\mathbb{Q}$, let*

$$V_r(K) = \left\{ \underline{x} \in \mathbb{D}(K)^{d^2} \middle| \dim_K K[\underline{x}] \leq k \right\},$$

where $K[\underline{x}]$ is the K-subalgebra of $\mathbb{D}(K) = \mathbb{D}(\mathbb{Q}) \otimes K$ generated by $\underline{x}$. Then V_r is an algebraic variety defined over $\mathbb{Q}$, when defined in terms of the co-ordinates of the elements of $\underline{x}$ in the basis. Moreover, it is defined by a finite set of homogeneous polynomials of a-priori bounded degrees.

PROOF. For $\underline{x} \in \mathbb{D}(K)^{d^2}$ let $W_t(\underline{x})$ denote the subspace of $\mathbb{D}(K)$ spanned by all products of length at most t in the elements $x^{(1)}, \ldots, x^{(d^2)}$. This is a non-decreasing sequence of subspace of the d^2-dimensional K vector space $\mathbb{D}(K)$. Hence there must exist a $1 \leq t \leq d^2$ such that $W_{t+1} = W_t$. This means that the subspace W_t, spanned by products of the $x^{(i)}$, is closed under left and right multiplication by them. In other words, $W_t = K[\underline{x}]$, the subalgebra generated by the $\underline{x}$. Since the W_t cannot increase further we conclude that $W_{d^2} = K[\underline{x}]$ in all cases. Now, the set of products of at most d^2 of the $\underline{x}$ is an element of $(\mathbb{D}(K))^{\sum_{l=0}^{d^2} d^{2l}}$ which depends polynomially on $\underline{x}$ (fix an ordering). Since the previous Lemma showed that $V_{r,\sum_{l=0}^{d^2} d^{2l}}$ is defined by homogeneous polynomial equations, we are done. Note that the "structure constants" a_{ijk} enter into the coefficients of these polynomials. □

DEFINITION 4.3.3. Let $x \in \mathbb{Q}$. We define its *denominator* $\mathrm{d}(x)$ and *height* $h(x)$ by:

$$\begin{aligned} \mathrm{d}(x) &= \prod_{p : v_p(x) < 0} p^{-v_p(x)}, \\ h(x) &= \prod_{p < \infty} p^{|v_p(x)|}. \end{aligned}$$

For a sequence $\underline{x} \in \mathbb{Q}^r$ we set $\mathrm{d}(\underline{x}) = \gcd \{\mathrm{d}(x_i)\}$.

LEMMA 4.3.4. *(Properties of denominators and heights) Let $x, x' \in \mathbb{Q}$.*

(1) $\mathrm{d}(xx') \leq \mathrm{d}(x)\mathrm{d}(x')$ *and* $\mathrm{d}(x + x') \leq \mathrm{d}(x)\mathrm{d}(x')$.
(2) $h(xx') \leq h(x)h(x')$. *If $x \in \mathbb{Q}^\times$ then* $h(x^{-1}) = h(x)$.
(3) *Let $P \in \mathbb{Q}[\underline{x}]$ be a multivariate polynomial in r variables. Then there exist $C_P > 0$ and an integer t_P such that for all $\underline{x} \in \mathbb{Q}^r$:*

$$\mathrm{d}(P(\underline{x})) \leq C_P \mathrm{d}(\underline{x})^{t_P}.$$

PROOF. Direct calculation and induction. □

DEFINITION 4.3.5. For $x_p \in \mathbb{Q}_p$ we set $\mathrm{d}_p(x_p) = 1$ if $x_p \in \mathbb{Z}_p$, $\mathrm{d}_p(x_p) = p^{-v_p(x_p)}$ otherwise and $h_p(x_p) = p^{|v_p(x_p)|}$. If $x \in \mathbb{A}_\mathrm{f}$ we set $\mathrm{d}(x) = \prod_p \mathrm{d}_p(x_p)$, $h(x) = \prod_p h_p(x_p)$.

LEMMA 4.3.6. *$d \colon \mathbb{A}_\mathrm{f} \to \mathbb{Z}$ is uniformly continuous in the adelic topology; Our definitions are compatible with the standard embeddings $\mathbb{Q}_p \hookrightarrow \mathbb{A}_\mathrm{f}$ and $\mathbb{Q} \hookrightarrow \mathbb{A}_\mathrm{f}$.*

PROOF. Let $x \in \mathbb{A}_\mathrm{f}$, $x' \in \prod_p \mathbb{Z}_p$. Then $\mathrm{d}(x + x') = \mathrm{d}(x)$: at places where x is integral, so is $x + x'$. At places where x is not, we have $v_p(x + x') = v_p(x)$ by the ultrametric behaviour of v_p. The second assertion is obvious. □

We now extend the notion of denominator to elements of $\mathbb{G}$:

DEFINITION 4.3.7. For $\gamma \in \mathbb{D}(\mathbb{Q})^\times$ set $\mathrm{d}(\gamma) = \gcd\left\{\mathrm{d}(x_i(\gamma))\right\}_{i=1}^{d^2}$. As above we extend this notion to maps $\mathrm{d}_p\colon G_p \to \mathbb{Z}$ and $\mathrm{d}\colon \mathbb{G}(\mathbb{A}_\mathrm{f}) \to \mathbb{Z}$. For $g \in \mathbb{G}(\mathbb{A}_\mathrm{f})$ we will also write $h_p(g_p) = h_p(\nu_\mathbb{D}(g_p))$ and $h(g) = \prod_p h_p(g_p) = h(\nu_\mathbb{D}(g))$.

COROLLARY 4.3.8. *$d(gg') \le d(g)d(g')$ and $h(g^{-1}) = h(g)$ for any $g \in \mathbb{G}(\mathbb{A}_\mathrm{f})$. The map $d\colon \mathbb{G}(\mathbb{A}_\mathrm{f}) \to \mathbb{Z}$ is continuous.*

PROOF. We have $x_i(gg') = \sum_{j,k} a_{ijk} x_j(g) x_k(g')$ with all the $a_{ijk} \in \mathbb{Z}$. It is also clear that the map is locally constant. □

The continuity of the denominator map implies, in particular, its boundedness on the compact subgroup K_R, and we will assume that $d(k) \le c_1$ for any $k \in K_R$. If $p \notin R$ and $k_p \in K_p$ and by Fact 2.1.3(4) we have $x_i(k_p) \in \mathbb{Z}_p$ for all $1 \le i \le d^2$ and hence $\mathrm{d}_p(k_p) = 1$. In fact, if $g_p \in \mathbb{D}(\mathbb{Q}_p)$ then $x_i(g_p k_p)$ and $x_i(k_p g_p)$ are all linear combinations with integral coefficients of the $x_i(g_p)$. This means $\mathrm{d}_p(k_p g_p), \mathrm{d}_p(g_p k_p) \le \mathrm{d}_p(g_p)$. Multiplying by k_p^{-1} (also an element of K_p) we get:

LEMMA 4.3.9. *Let $p \notin R$, $k_p \in K_p$ and $g_p \in G_p$. Then $\mathrm{d}_p(g_p k_p) = \mathrm{d}_p(k_p g_p) = \mathrm{d}_p(g_p)$. If $g \in \mathbb{G}(\mathbb{A}_\mathrm{f})$, $k \in K_\mathrm{f}$ and $k_p = 1$ for $p \in R$ then $\mathrm{d}(kg) = \mathrm{d}(gk) = \mathrm{d}(g)$.*

4.4. Intersections of Hecke Translates

Let $a \in A \setminus Z$, $C \subset A_a M_a$. Let $\Omega_\infty \subset G$ be compact such that $\pi(\Omega_\infty) = X_Z$, and choose a neighbourhood $\tilde{C} \subset A_a M_a$ so that for small enough ϵ, $B_a(C,\epsilon) B_a(C,\epsilon)^{-1}$ is contained in $\tilde{B} = B_a(\tilde{C}, O(\epsilon))$ as in Lemma 4.2.2. We will also shorten $B = B_a(C,\epsilon)$.

Fixing some $x \in X_Z$ we are ready to analyze intersections of Hecke translates of xB. Writing $x = \Gamma Z x_\infty$ for some $x_\infty \in \Omega_\infty$, let $g, g' \in \mathbb{G}(\mathbb{A}_\mathrm{f})$ be such that $x_\infty B g \cap x_\infty B g'$ is nonempty. This entails the existence of $\gamma \in \mathbb{G}(\mathbb{Q})$, $z_\infty \in Z$, $k \in K_\mathrm{f}$ and $b, b' \in B$ such that

$$\gamma z_\infty x_\infty b g = x_\infty b' g' k \tag{4.4.1}$$

holds as an equality of adéles, where γ is embedded diagonally in $\mathbb{G}(\mathbb{A})$. We will say that such γ *cause an intersection*. At the infinite place this reads $\gamma z_\infty x_\infty b = x_\infty b'$ or:

$$\gamma z_\infty = x_\infty b' b x_\infty^{-1} \in x_\infty B B^{-1} x_\infty^{-1} \subset x_\infty \tilde{B} x_\infty^{-1}, \tag{4.4.2}$$

where the last inclusion follows from the choice of $\tilde{C}$. Now lemma 4.2.3 shows that γz_∞ is $O(\epsilon)$-close to $x_\infty A_a M_a x_\infty^{-1}$ (independently of g or g'!). Recalling that $A_a M_a$ is contained in a proper subalgebra of $\mathbb{D}(\mathbb{R})$, we see that these γ are very close to satisfying the polynomial equations we have just constructed, except for the annoyance of the factor z_∞: γ is $O(\epsilon/z_\infty)$-close to an element of the subalgebra, and $|z_\infty|$ might be very small. Lemma 4.4.3 addresses this problem.

Noting that the sets $\tilde{B}$ only decrease with ϵ, we also see that (assuming $\epsilon < 1$) if γ causes an intersection there exists $z_\infty \in \mathbb{R}^\times$ such that γz_∞ belongs to the relatively compact set:

$$\Omega_\infty B_a(\tilde{C}, O(1)) \Omega_\infty^{-1} \subset G_\infty. \tag{4.4.3}$$

Given $S \subset \mathbb{G}(\mathbb{A}_{\mathrm{f}})$, we denote $I(a, x_\infty, \tilde{B}, S)$ the set of $\gamma \in \mathbb{G}(\mathbb{Q})$ that cause an intersection for translates of $x_\infty B$ by *some* $g, g' \in S$, where γz_∞ lies in $x_\infty \tilde{B} x_\infty^{-1}$.

DEFINITION 4.4.1. Let $T > 0$. We will say that $S \subset \mathbb{G}(\mathbb{A}_{\mathrm{f}})$ *is T-bounded* if every $g \in S$ satisfies:

(1) $\mathrm{d}(g), \mathrm{d}(g^{-1}) \leq \epsilon^{-T}$;
(2) $h(\nu_{\mathbb{D}}(g)) \leq \epsilon^{-T}$ (where $\nu_{\mathbb{D}} \colon \mathbb{G}(\mathbb{A}_{\mathrm{f}}) \to \mathbb{A}_{\mathrm{f}}$ is the reduced norm);
(3) $(g_p)_{p \in R} \in K_R$.

LEMMA 4.4.2. *Let $S \subset \mathbb{G}(\mathbb{A}_{\mathrm{f}})$ be T-bounded, and let $x_\infty \in \Omega_\infty$, $\gamma \in I(a, x_\infty, \tilde{B}, S)$. Then $\mathrm{d}(\gamma), |\nu_{\mathbb{D}}(\gamma)|_\infty \ll \epsilon^{-2T}$ where the implied constant depends only on the choice of co-ordinates and on the compact subgroup K_{f}.*

PROOF. By definition we have $g \neq g' \in S$ and $k \in K_{\mathrm{f}}$ such that $\gamma = g' k g^{-1}$ holds as an equality of finite adéles. We now use this place-by-place. Projecting to $G_R = \prod_{p \in R} G_p$, we have $g_R, g'_R \in K_R$ and hence $\gamma \in K_R$ under the diagonal embedding. It follows that $\prod_{p \in R} \mathrm{d}_p(\gamma)$ is uniformly bounded by the number c we fixed above. If $p \notin R$ we use Corollary 4.3.8 and Lemma 4.3.9 to get:

$$\mathrm{d}_p(\gamma_p) \leq \mathrm{d}_p(g'_p k_p) \mathrm{d}_p(g_p^{-1}) = \mathrm{d}_p(g'_p) \mathrm{d}_p(g_p^{-1}).$$

Multiplying over all primes we conclude:

$$\mathrm{d}(\gamma) \ll e^{-2T}.$$

Next, by the product formula we have $|\nu_{\mathbb{D}}(\gamma)|_\infty = \prod_p |\nu_{\mathbb{D}}(\gamma_p)|_p^{-1}$. The absolute value of the reduced norm is a positive multiplicative quasi-character. In particular it is the constant 1 on any compact subgroup such as K_{f}. Also, we have defined the height so that $\left|\nu_{\mathbb{D}}(g_p^{-1})\right|_p \leq h_p(g_p)$. It follows again that:

$$|\nu_{\mathbb{D}}(\gamma)|_\infty \leq \prod_p h_p(g'_p) h_p(g_p) \leq h(g) h(g') \leq \epsilon^{-2T}.$$

□

LEMMA 4.4.3. *There exists a constant $c > 0$ (depending on $\Omega_\infty, \tilde{C}$) such that for every $x_\infty \in \Omega_\infty$, small enough $\epsilon > 0$, $S \subset \mathbb{G}(\mathbb{A}_{\mathrm{f}})$ and $\gamma \in I(a, x_\infty, \tilde{B}, S)$ there exist $z_\infty \in \mathbb{R}^\times_{>0}$ with $|z_\infty| \gg |\nu(\gamma)|_\infty^{-1/d}$ and $g_\infty \in x_\infty \tilde{C} x_\infty^{-1} \subset x_\infty A_a M_a x_\infty^{-1}$ such that $|x_i(g_\infty) - z_\infty x_i(\gamma)| \leq c\epsilon$ for every $1 \leq i \leq d^2$.*

PROOF. Let $\gamma \in I(a, x_\infty, \tilde{B}, S)$, $\mathrm{diag}(z_\infty) \in Z$ such that $\gamma \,\mathrm{diag}(z_\infty) \in x_\infty \tilde{B} x_\infty^{-1}$ (note that $x_i(\gamma \,\mathrm{diag}(z_\infty)) = z_\infty x_i(\gamma)$ by definition). Assuming, as we may, that $\epsilon < 1$, we have observed above that γz_∞ must belong to a specific compact subset of G. The continuity of the map $\det : G \to \mathbb{R}^\times$ shows that $\det(\gamma \,\mathrm{diag}(z_\infty))$ is uniformly bounded below. As $\nu(\gamma) = \det(\gamma)$ when on the right we embed γ in $\mathbb{D}(\mathbb{R})$, and $\det(\mathrm{diag}(z_\infty)) = z_\infty^d$ we are done.

□

THEOREM 4.4.4. *(c.f.* [**3**, Lemma 3.3]*) There exists* $T > 0$ *such that, given a* T*-bounded* S *and a point* $x_\infty \in \Omega_\infty$ *we can find a number field* $F \subset \mathbb{D}(\mathbb{Q})$ *of dimension* d *such that* $I(a, x_\infty, \tilde{B}, S) \subset F^\times \subset \mathbb{G}(\mathbb{Q})$*, whenever* ϵ *is small enough. Moreover, there exist* $\left\{\gamma'^{(j)}\right\}_{j=1}^{d} \subset \mathcal{O}$*, each of height at most* $O(\epsilon^{-T'})$ *for some known* T'*, such that* $F = \mathbb{Q}(\gamma'^{(1)}, \ldots, \gamma'^{(d)})$.

PROOF. Choose $\left\{\gamma^{(j)}\right\}_{j=1}^{d^2} \subset I(a, x_\infty, \tilde{B}, S)$ which have the same $\mathbb{Q}$-span as all of $I(a, x_\infty, \tilde{B}, S)$, and let $\mathbb{D}'$ denote the $\mathbb{Q}$-subalgebra of $\mathbb{D}$ generated by the$\left\{\gamma^{(j)}\right\}$, necessarily a division algebra. It suffices to show that this is a proper subalgebra of $\mathbb{D}$, since in that case $\dim_{\mathbb{Q}} \mathbb{D}'(\mathbb{Q})$ will be a proper divisor of $\dim_{\mathbb{Q}} \mathbb{D}(\mathbb{Q}) = d^2$. With d assumed prime, the possibilities are $\dim_{\mathbb{Q}} \mathbb{D}'(\mathbb{Q}) = 1$ where $\mathbb{D}'(\mathbb{Q}) = \mathbb{Q}$ and $\dim_{\mathbb{Q}} \mathbb{D}'(\mathbb{Q}) = d$, where $\mathbb{D}'(\mathbb{Q})$ must be monogenic and hence commutative, i.e. a number field.

By Lemma 4.4.2 we have $\mathrm{d}(\gamma^{(j)}) \ll \epsilon^{2T}$ for all j. Applying Lemma 4.4.3 as well we can find for each j

$$g_\infty^{(j)} \in x_\infty \tilde{C} x_\infty^{-1} \subset x_\infty A_a M_a x_\infty^{-1} \subset \Omega_\infty A_a M_a \Omega_\infty^{-1}$$

and $z_\infty^{(j)} \in Z_{\mathbb{G}}(\mathbb{R})$ such that $\left|z_\infty^{(j)}\right| \gg \epsilon^{2T/d}$ and such that for each $1 \le i \le d^2$ we have:

$$\left|x_i(g_\infty^{(j)}) - z_\infty^{(j)} x_i(\gamma_\infty^{(j)})\right| \ll \epsilon.$$

Now let $r = \dim_{\mathbb{R}} A_a M_a$, and let $\{f_m\}_{m=1}^{M}$ be a set of polynomials with integer coefficients defining V_r, each homogeneous of degree h_m in the $\underline{x}^{(j)}$. That $\left\{g_\infty^{(j)}\right\}_{j=1}^{d^2} \in V_r(\mathbb{R})$ (they generate an $\mathbb{R}$-subalgebra of at most that dimension!) can be written as $f_m\left(\left\{g_\infty^{(j)}\right\}_j\right) = 0$. Since we can uniformly bound the gradient of f_m in a $c\epsilon$-neighbourhood of the relatively compact set $\left(\Omega_\infty B_a(\tilde{C}, O(1))\Omega_\infty^{-1}\right)^{d^2}$ for ϵ small enough, we have:

$$\left|f_m\left(\left\{z_\infty^{(j)}\gamma_\infty^{(j)}\right\}\right)\right| \ll \epsilon,$$

and hence:

$$\left|f_m\left(\left\{\gamma_\infty^{(j)}\right\}\right)\right| \ll \epsilon^{1-2h_m T/d}.$$

Now since the coefficients of f_m are integers, the denominator of $f_m\left(\left\{\gamma_\infty^{(j)}\right\}_j\right)$ is at most the h_m-th power of the maximal denominator of an $x_i(\gamma_\infty^{(j)})$, i.e. at most $\epsilon^{-2h_m T}$. The crucial observation is then that if $f_m\left(\left\{\gamma_\infty^{(j)}\right\}_j\right) \neq 0$, it is then at least $\epsilon^{2h_m T}$. Choosing T so that for all m

$$T < \frac{2d}{d+1} h_m,$$

we can make sure that for ϵ small enough we will have $f_m\left(\left\{\gamma_\infty^{(j)}\right\}\right) = 0$ for all m, i.e. that $\dim \mathbb{D}' \le r$.

Finally, since $[F : \mathbb{Q}] = d$, we can assume w.l.g. that $\left\{\gamma^{(j)}\right\}_{j=1}^{d}$ generate F, and the same will hold if we replace $\gamma^{(j)}$ with $\gamma'^{(j)} = d(\gamma^{(j)})\gamma^{(j)} \in \mathcal{O}$. Now the discriminant of γ' is at most that of the characteristic polynomial of the automorphism $x \mapsto \gamma' x$ of $\mathbb{D}(\mathbb{Q})$. Since the discriminant of a polynomial is a polynomial expression in its coefficients, and since in our case these coefficients are polynomials in the $x_i(\gamma')$, which are bounded by the height of γ'. From this one can recover an exponent T' as in the statement of the Theorem. $\square$

REMARK 4.4.5.

(1) The discriminant of F as above is also $O(\epsilon^{-T''})$, since it is at most the product of the discriminants of the $\gamma'^{(j)}$.

(2) We have used the observation that every γ causing an intersection must satisfy $\gamma_\infty \in \Omega_\infty B_a(\tilde{C}, O(1))\Omega_\infty^{-1}$, and the latter is a compact set independent of ϵ.

CHAPTER 5

The Method of Hecke Translates II: Geometry and Harmonic Analysis on the Building

As discussed in the introduction, this method was initiated by [**25, 15**]. Our analysis stems from the very concrete Lemmata 3.3 and 3.4 of [**3**].

The following is the second main ingredient of the main theorem:

THEOREM 5.0.1. *Let $a \in A \setminus Z$. Then there exists $\eta > 0$ such that for every Hecke Eigenfunction $\psi \in L^2(X, \omega)$ and every relatively compact neighbourhood of the identity $C < M_a$,*

$$\mu_\psi(xB_a(C, \epsilon)) \ll_a \epsilon^\eta$$

holds for ϵ small enough and any $x \in X_Z$. In particular, the implied constant is independent of f.

In the previous chapter we saw that given a T-bounded $S \subset \mathbb{G}(\mathbb{A}_\mathrm{f})$, intersections between sets of the type $x_\infty Bg$ and $x_\infty Bg'$ for $g, g' \in S$ could occur only if $\gamma g K_\mathrm{f} = g' K_\mathrm{f}$ for some $\gamma \in F^\times$, where $F \hookrightarrow \mathbb{D}(\mathbb{Q})$ is a number field of degree d and bounded discriminant. The proof hinges on choosing an S that will on the one hand be small enough to (almost) fail to admit such intersections, while on the other be large enough to have $\mu_\psi(xB)$ bounded by a small multiple of $\sum_{g \in S} \mu_\psi(xBg)$. The resulting disjointness of the translates implies that the latter sum is bounded by 1, hence that $\mu_\psi(xB)$ is small. The proof will be broken up in several stages, alternating geometrical considerations and harmonic analysis estimates.

We first analyze the intersection pattern at a single place, in other words the action of the torus $F^\times$ on the quotient G_p/K_p. We will do this by embedding the discrete set G_p/K_p in a geometric structure, the *building* of G_p. In Section 5.1 we give a summary of the properties of the building and use its geometry to construct a subset $S_p \subset G_p$ of translates for which we understand the (local) intersection pattern. In Section 5.2 we then bound $\sum_{g_p \in S_p} |\psi(x_\infty b g_p)|^2$ from below by a not-too-small multiple of $|\psi(x_\infty b)|^2$ for any $x_\infty \in \Omega_\infty$ and $b \in B$. We use there bounds toward the Ramanujan conjecture due to Luo-Rudnick-Sarnak.

The next step is to combine the information from many places. Section 5.3 shows that by taking the union of the S_p over many places we still have no intersections, allowing us to complete the proof of Theorem 5.0.1 in Section 5.4 .

5.1. The buildings of GL_n and PGL_n.

Let p be a finite rational prime, v the p-adic valuation on $\mathbb{Q}$, with completion $\mathbb{Q}_p$, valuation ring $\mathbb{Z}_p$ and maximal ideal $\mathfrak{p} = p\mathbb{Z}_p \lhd Z_p$. The residue field $\mathbb{Z}_p/\mathfrak{p} \simeq \mathbb{F}_p$ is then finite, and we will denote its cardinality q. We choose a uniformizer $\varpi \in \mathfrak{p} \setminus \mathfrak{p}^2$ (e.g. $\varpi = p$) and normalize the absolute value on $\mathbb{Q}_p$ so that $|\varpi| = q^{-1}$.

REMARK 5.1.1. The distinction between p and q appears silly. It amounts to distinguishing between a finite cardinal and the associated integer, thought of as an object of arithmetic. However, the discussion below would remain unchanged if $\mathbb{Q}_p$ was replaced by a field complete w.r.t. a discrete valuation, $\mathbb{Z}_p$ by its valuation ring $\mathcal{O}$, and $\mathfrak{p}$ by the maximal ideal of $\mathcal{O}$. In that case read q for the cardinality of the residue field, p for its characteristic. The buildings (defined below) are locally finite exactly when q is a finite cardinal.

NOTATION 5.1.2. For this Section (5.1) and the next only, we drop the subscript 'p' we use elsewhere, writing $G = \mathrm{GL}_n(\mathbb{Q}_p)$, $K = \mathrm{GL}_n(\mathbb{Z}_p)$. Thus A will denote the subgroup of invertible diagonal matrices and $Z = Z_G \simeq \mathbb{Q}_p^\times$ the center, i.e. the subgroup of non-zero scalar matrices.

DEFINITION 5.1.3. Let $\mathcal{B}^0 = \mathrm{PGL}_n(\mathbb{Q}_p)/\mathrm{PGL}_n(\mathbb{Z}_p)$, $\tilde{\mathcal{B}}^0 = \mathcal{B}^0 \times \mathbb{Z}$. We let G act on $\mathcal{B}^0$ via the quotient $\mathrm{PGL}_n(\mathbb{Q}_p)$, on $\tilde{\mathcal{B}}^0$ by:

$$g \cdot (x, n) = (gx, n + v(\det(g)))\,.$$

REMARK 5.1.4. One can identify G/K with the space of $\mathbb{Z}_p$-lattices in $\mathbb{Q}_p^n$, $\mathcal{B}^0$ with the space of homothety classes of lattices.

Noting that $\det(k) \in \mathcal{O}^\times$ for any $k \in K$, for $\tilde{x} = gK \in G/K$ the integer $\nu(\det g)$ is independent of the choice of representative $g \in \tilde{x}$. We will denote it $c(\tilde{x})$. Then the map $\varphi\colon G/K \to \tilde{\mathcal{B}}^0$ given by $\varphi(\tilde{x}) = (\tilde{x}Z, c(\tilde{x}))$ is a G-equivariant embedding. However, it is not surjective: for $\tilde{x} \in G/K$ and $z \in \mathbb{Q}_p^\times$ thought of as an element of Z we have $c(z\tilde{x}) = nv(z) + c(\tilde{x})$. Hence, to each $x \in \mathcal{B}^0$ we can associate a residue class $a_x \in \mathbb{Z}/n\mathbb{Z}$ so that the image of φ is precisely $\{(x,t) \mid x \in \mathcal{B}^0, t \in a_x\}$.

DEFINITION 5.1.5. Call a sequence $\{x_i\}_{i=0}^d \subset \mathcal{B}^0$ an *oriented d-simplex* if there exist representative lattices $\Lambda_i \in x_i$ (also let $\Lambda_{d+1} = \mathfrak{p}\Lambda_0$) such that $\Lambda_i \supset \Lambda_{i+1}$ for all $0 \leq i \leq d$. Denote by $\mathcal{B}^d$ the set of d-simplexes (forgetting the orientation for the moment). n-dimensional simplexes are called *chambers*.

FACT 5.1.6. $\mathcal{B} = \left\{\mathcal{B}^d\right\}_{d=0}^n$ *is a chamber complex:*

(1) *It is a simplicial complex, i.e. the intersection of two simplexes is again a simplex.*
(2) *Every simplex is contained in a chamber.*

Moreover, the G action on $\mathcal{B}^0$ is simplicial.

DEFINITION 5.1.7. The complex $\mathcal{B}$ is called the *simplicial building* of $\mathrm{PGL}_n(\mathbb{Q}_p)$. Endowing $\mathbb{Z}$ with its standard 1-dimensional simplicial complex structure, the simplicial complex $\tilde{B} = \mathcal{B} \times \mathbb{Z}$ (whose set of vertices is precisely $\tilde{\mathcal{B}}^0$) is called the (poly-)*simplicial building* of $\mathrm{GL}_n(\mathbb{Q}_p)$. The elements of $\mathcal{B}^0$ and $\tilde{B}^0$ will be called the *vertices* of the respective buildings. In particular, we will thing of the cosets of G/K as vertices of $\tilde{\mathcal{B}}$.

Now let $x_0 \in \mathcal{B}^0$ be the vertex stabilized by ZK (i.e. the identity coset), and let $\mathcal{A}_0^0 \subset \mathcal{B}^0$ be the orbit Ax_0. Identifying $A \simeq \left(\mathbb{Q}_p^\times\right)^n$, we have $\mathrm{Stab}_A(x_0) = Z \cdot \left(\mathbb{Z}_p^\times\right)^n$ (with the center acting diagonally). We then identify $\mathcal{A}_0^0$ with $A/\operatorname{Stab}_A(x_0) \simeq \mathbb{Z}^n/\mathbb{Z}(1,\ldots,1)$. The vertex corresponding to $\underline{r} + \mathbb{Z}(1,\ldots,1)$ is the homothety class of the lattice generated by $\{\varpi^{r_i}\underline{e}_i\}_{i=1}^n$ where $\{\underline{e}_i\}_{i=1}^n$ is the standard basis of $\mathbb{Q}_p^n$.

FACT 5.1.8. *The subcomplex $\mathcal{A}_0 \subset \mathcal{B}$ consisting of those simplexes supported on $\mathcal{A}_0^0$ is a chamber subcomplex; for every two simplexes $\Delta_1, \Delta_2 \in \mathcal{B}$ there exists $g \in G$ such that $g\Delta_1, g\Delta_2 \in \mathcal{A}_0$.*

We have $\{g \in G \mid g\mathcal{A}_0 = \mathcal{A}_0\} = N_G(A)$, and $N_G(A) = W \ltimes A$ where $W < K$ is the subgroup of permutation matrices (henceforth called the Weyl group *of G w.r.t. A).*

DEFINITION 5.1.9. A subcomplex $\mathcal{A}$ of the form $g\mathcal{A}_0$ for some $g \in G$ is called an *apartment* of $\mathcal{B}$. We have just seen that any two simplexes are contained in an apartment. In particular, any two points x, y of the geometric realization $|\mathcal{B}|$ lie in the geometric realization $|\mathcal{A}|$ of an apartment $\mathcal{A}$.

In order to give a canonical metric on $\mathcal{B}$, we start with the positive semidefinite bilinear form

$$\langle \underline{u}, \underline{v} \rangle = \sum_{i=1}^n u_i v_i - \frac{1}{n}\left(\sum_{i=1}^n u_i\right)\left(\sum_{i=1}^n v_i\right)$$

on $\mathbb{R}^n$. Its isotropic subspace is one-dimensional and spanned by the vector $(1,\ldots,1)$. In particular, this pairing descends to a positive-definite bilinear form on $\mathbb{R}^n/\mathbb{R}(1,\ldots,1)$. This defines a norm and hence a metric on this space.

FACT 5.1.10. *The identification of the vertices of the standard apartment with the quotient $\mathbb{Z}^n/\mathbb{Z}(1,\ldots,1)$ extends uniquely to a piecewise-linear isomorphism of the geometric realization$|\mathcal{A}_0|$ and $\mathbb{R}^n/\mathbb{R}(1,\ldots,1)$.*

Pulling back the norm we have defined gives a metric on $|\mathcal{A}_0|$. We have remarked before that the elements of G that preserve $\mathcal{A}_0$ are generated by $N_G(A) = WA$. Since W acts by permuting the co-ordinates, and A by affine translations, the metric we have defined is $N_G(A)$-invariant.

FACT 5.1.11. *This metric extends uniquely to a G-invariant metric $d(\cdot,\cdot)$ on $|\mathcal{B}|$, in particular on the set of vertices $\mathcal{B}^0$. $|\mathcal{B}|$ is a simply connected complete CAT(0) metric space. The geometric realization of an apartment is a (flat) geodesic subspace.*

The standard metric on $\mathbb{Z}$ extends to a metric on its realization $\mathbb{R}$ as a simplicial complex. We extend our metric d (with the same notation) to $\left|\tilde{\mathcal{B}}\right| \simeq |\mathcal{B}| \times \mathbb{R}$ by taking the Euclidean product ($d^2\left((x,s),(y,t)\right) = d^2(x,y) + |s-t|^2$).

DEFINITION 5.1.12. We call the metric space $(|\mathcal{B}|, d)$ together with the isometric G-action the *building* of $\mathrm{PGL}_n(\mathbb{Q}_p)$, the metric space $\left(\left|\tilde{\mathcal{B}}\right|, d\right)$ the *building* of G. Both spaces are simply connected complete CAT(0) spaces.

FACT 5.1.13. *Let (X, d) be a CAT(0) metric space. Then:*

(1) *(X, d) is uniquely geodesic. We will use $[x, y]$ to denote the unique geodesic segment between x, y.*

(2) *Let $Y \subset X$ be convex ($x, y \in Y \Rightarrow [x,y] \subset Y$) and closed. Then for any $x \in X$ the function $y \mapsto d(x,y)$ on Y is strictly convex and has a unique minimum, say $\pi_Y(x)$.*
(3) *The map $x \mapsto \pi_Y(x)$ is a retraction of X to Y. It does not increase distances.*

COROLLARY 5.1.14. *Let $x, y \in |\mathcal{B}|$ satisfy $\mathrm{pr}_{|\mathcal{A}_0|}(x) = \mathrm{pr}_{|\mathcal{A}_0|}(y) = x_0$. Then we have $\{a \in A \mid ax = y\} \subset A \cap KZ$.*

PROOF. Since the (isometric!) action of a on $|\mathcal{B}|$ preserves $|\mathcal{A}_0|$, and since the orthogonal projection operator pr is defined via the metric, we have $\mathrm{pr}_{|\mathcal{A}_0|}(ax) = a\,\mathrm{pr}_{|\mathcal{A}_0|}(x)$ for all $x \in |\mathcal{B}|$, $a \in A$. In particular, if $ax = y$ we have $ax_0 = x_0$, i.e. $a \in \mathrm{Stab}_G(x_0) = KZ$. □

Finally, we introduce a co-ordinate system of sorts on G, in terms of the set $A^+ = \{\mathrm{diag}(\varpi^{r_1}, \cdots, \varpi^{r_n}) \mid 0 = r_1 \leq r_2 \leq \cdots \leq r_n\} \subset A$. It is easy to see that $A^+ x_0 \subset \mathcal{A}_0^0$ is a set of representatives for the orbits of W. From this one gets:

FACT 5.1.15. *(Cartan decomposition) Let $x, y \in \mathcal{B}^0$ be vertices. Then there exists a unique $a \in A^+$ such for some $g \in G$ we have $gx = x_0$, $gy = ax_0$. In particular ("KAK decomposition"), for any $y \in \mathcal{B}^0$ there exists a unique $a \in A^+$ and some $k \in \mathrm{Stab}_G(x_0) = K$ such that $ky = ax_0$.*

We call a the *relative position* of y w.r.t. x (in that order!). Clearly this is a G-equivariant notion. It is also clear that the distance $d(x,y)$ only depends on the relative position of x and y.

LEMMA 5.1.16. *Let $\underline{r} \in A^+$ and let $N(x_0, \underline{r})$ be the set of vertices of relative position $\underline{r}$ to x_0. Then*

$$\#N(x_0, \underline{r}) \sim q^{2\delta(\underline{r})},$$

asymptotically as $q \to \infty$ where $\underline{r}$ is fixed. Here $\delta(\underline{r}) = -\sum_{i=1}^{n} \left(\frac{n+1}{2} - i\right) r_i$.

PROOF. The KAK decomposition shows $N(x_0, \underline{r}) = \{kax_0 \mid k \in K\}$ where $a = \mathrm{diag}(\varpi^{\underline{r}})$. Now it suffices to compute the index of $\mathrm{Stab}_K(ax_0)$ in K. A direct computation shows that

$$\mathrm{Stab}_K(ax_0) = \{k \in K \mid \forall i,j : v(k_{ij}) \geq r_i - r_j\}.$$

Since $v(k_{ij}) \geq 0$ for all i, j, the condition only has meaning for $i > j$. We also let $N = r_n$ and set $K_N = \{k \in K \mid k \equiv I_n(p^N)\}$, the kernel of the quotient map $Q\colon \mathrm{GL}_n(\mathbb{Z}_p) \to \mathrm{GL}_n(\mathbb{Z}/p^N\mathbb{Z})$. By the choice of N we see that $p^{r_i - r_j} \mid p^N$ for all $i > j$, and hence that $K_N \subset K_{\underline{r}}$. Setting $\bar{K}_{\underline{r}} = Q(K_{\underline{r}})$, $\bar{G} = \mathrm{GL}_n(\mathbb{Z}/p^N\mathbb{Z})$ and letting $\bar{B} < \bar{G}$ denote the subgroup of upper-triangular matrices, we then have:

$$\#N(x_0, \underline{r}) = [K\colon K_{\underline{r}}] = \left[\bar{G}\colon \bar{K}_{\underline{r}}\right] = \frac{\#\bar{G}}{\#\bar{B}\left[\bar{K}_{\underline{r}}\colon \bar{B}\right]}.$$

Next, let $\bar{U} < \bar{G}$ be the subgroup of lower-triangular unipotent matrices, let $\bar{W} < \bar{G}$ denote the subgroup of permutation matrices, and set $\bar{U}_{\underline{r}} = \bar{U} \cap \bar{K}_{\underline{r}}$. We will show that

$\#\bar{U}_{\underline{r}} \leq \left[\bar{K}_{\underline{r}} \colon \bar{B}\right] \leq \#W \cdot \#\bar{U}_{\underline{r}}$, and hence that:

$$\#N(x_0, \underline{r}) \sim \frac{\#\bar{G}}{\#B \cdot \#\bar{U}_{\underline{r}}}.$$

It now suffices to compute the orders of the finite groups $\bar{G}$, B, $\bar{U}_{\underline{r}}$. Since $\bar{U}_{\underline{r}}$ is the set of $\bar{g} \in \bar{G}$ such that $\bar{g}_{ij} = \delta_{ij}$ for $i \leq j$ and $\bar{g}_{ij} \in p^{r_i - r_j}\mathbb{Z}/p^N\mathbb{Z}$ for $i > j$, we have

$$\#\bar{U}_{\underline{r}} \sim \prod_{i>j} p^{N-(r_i - r_j)} = \left(p^N\right)^{n(n-1)/2} p^{-2\delta(\underline{r})}.$$

It is also clear that for N fixed and $p \to \infty$ (recall that N only depends on $\underline{r}$), we have $\#\bar{G} \sim \left(p^N\right)^{n^2}$ and

$$\#\bar{B} = \varphi(p^N)^n \cdot (p^N)^{n(n-1)/2} \sim \left(p^N\right)^{n(n+1)/2}.$$

It remains to estimate $\left[\bar{K}_{\underline{r}} \colon \bar{B}\right]$. Since both $\bar{U}_{\underline{r}}, \bar{B}$ are subgroups of $\bar{K}_{\underline{r}}$, we have $\bar{U}_{\underline{r}}\bar{B} \subset \bar{K}_{\underline{r}}$. Standard linear algebra (Gaussian elimination) shows that any element of $\bar{G}$ has at most a unique representation in the form $\bar{u}\bar{b}$ where $\bar{u} \in \bar{U}$, $\bar{b} \in \bar{B}$. This shows that $\#\bar{U}_{\underline{r}} \leq \left[\bar{K}_{\underline{r}} \colon \bar{B}\right]$. On the other side, we will use Gaussian elimination show $\bar{W}\bar{U}_{\underline{r}}\bar{B} = \bar{K}_{\underline{r}}$: Start with $k \in K_{\underline{r}}$. If $r_2 > r_1$, we must have $k_{11} \in \mathbb{Z}_p^\times$ (every other entry in the first column is in $p\mathbb{Z}_p^\times$). If $r_1 = r_2 \cdots = r_s$ then the top-left $s \times s$ minor of k must be invertible for the same reason (all entries below it are divisible by p) and we thus can permute the first s rows to ensure the pivots of the image of this minor in $\mathrm{GL}_n(\mathbb{Z}/p^N\mathbb{Z})$ are the diagonal elements. Moreover, the permuted matrix is still in $K_{\underline{r}}$. Multiplying on the right by an element of B (the set of upper-triangular matrices in K) we can now assume that the first s rows of k are the first s standard unit vectors. Next, we assume $r_{s+1} = \cdots = r_{s+t}$. Permuting these rows (an operation which essentially commutes with what we have done so far), we can assume that the pivots for the $t \times t$ minor on the diagonal at position $s+1, \ldots, s+t$ has its pivots on the diagonal, and continue the elimination by induction. □

LEMMA 5.1.17. *Let $x \in N(x_0, \underline{e}_n)$ (we think of $\underline{e}_n$ as a representative of a coset modulu $\mathbb{Z}(1, \ldots, 1)$). Then either $\#\left(N(x) \cap \mathcal{A}_0^0\right) \geq 2$ or $\mathrm{pr}_{|\mathcal{A}_0|}(x) = x_0$.*

PROOF. We may assume $x \notin \mathcal{A}_0$, $z_0 = \mathrm{pr}_{|\mathcal{A}_0|}(x) \neq x_0$. Being strictly convex, the function $y \mapsto d(x, y)$ is strictly decreasing on the geodesic segment $[x_0, z_0]$. Let $x_0 \in \Delta \in \mathcal{A}_0$ be a chamber such that $[x_0, z] = |\Delta| \cap [x_0, z_0]$ has positive length. Let $\mathcal{A}$ be an apartment containing the simplexes Δ and $\{x_0, x\}$. Then $x, x_0 \in |\mathcal{A}|$ and $\mathrm{pr}_{|\Delta_2|}(x) \neq x_0$, since $z \in |\Delta_2|$ is closer to x.

Without loss of generality we can identify the vertices of $\mathcal{A}$ with $\mathbb{Z}^n/\mathbb{Z}$ as before, with $x_0 = \underline{0}$, Δ_2 being the standard simplex with vertices $x_i = \sum_{j \geq n-i} \underline{e}_j$, and $x = \underline{e}_k$ for some k (there are all the neighbours with the correct relative position). The assumptions on the existence of z amount to saying that $x - x_0$ has a positive projection on $x_i - x_0$ for some $1 \leq i \leq n$ (otherwise $x - x_0$ would have non-positive projection on the direction $z - x_0$, contradicting $d(x,z) < d(x,x_0)$). But the inner product is:

$$\left\langle \underline{e}_k, \sum_{j\geq n-i} \underline{e}_j \right\rangle = \epsilon - \frac{i}{n}$$

where $\epsilon = 0$ if $k < n-i$, $\epsilon = 1$ otherwise. To make this positive, we must have $\epsilon = 1$, i.e. $k \geq n-i$. But then x is in fact a neighbour of $x_i \in \Delta_2^0 \subset \mathcal{A}_0^0$.

Conversely, if x is a neighbour of two vertices of $\mathcal{A}_0$ then (since all edges in $\mathcal{B}^1$ have the same length) the projection of x to $|\mathcal{A}_0|$ is closer to x than any of them, and in particular cannot be a vertex. □

LEMMA 5.1.18. *Let* $x \in N(x_0, \underline{e}_n)$ *be such that* $\mathrm{pr}_{|\mathcal{A}_0|}(x) = x_0$ *and let* $x' \in N(x, -\underline{e}_n)$. *Then* $\mathrm{pr}_{|\mathcal{A}_0|}(x') = x_0$.

PROOF. Assume $x' \neq x_0$, $z_0 = \mathrm{pr}_{|\mathcal{A}_0|}(x) \neq x_0$. As before let $x_0 \in \Delta$ be a simplex containing an initial segment $[x,z]$ of $[x_0,z_0]$ and let $\mathcal{A}$ be an apartment containing Δ and $\{x,x'\}$. As before we can choose co-ordinates such that $\Delta = \{x_i\}_{i=0}^{n-1}$ and $x = \underline{e}_1$ (the last by the proof of the previous lemma). We then have $x' = \underline{e}_1 - \underline{e}_k$ for some k, and the assumption $x' \neq x_0$ amount to assuming $k \neq 1$. We now show that x_0 is the point of $|\Delta|$ nearest to x' by computing the inner products $\langle x' - x_0, x_i - x_0\rangle$ and showing they are all non-positive. Indeed, with ϵ depending on i,k as in the previous lemma, we have:

$$\langle \underline{e}_1 - \underline{e}_k, x_i \rangle = -\epsilon \leq 0.$$

□

We now set $S_1 = \left\{x_1 \in N(x_0,a) \mid \mathrm{pr}_{|\mathcal{A}_0|}(x_1) = x_0\right\}$. Continuing outward we set $N_{x_1} = N(x_1, -\underline{e}_n) \setminus \{x_0\}$ for $x_1 \in S_1$.

LEMMA 5.1.19. *The union* $S_2 = \cup_{x_1 \in S_1} N_{x_1}$ *is disjoint.* S_1 *and* S_2 *are disjoint.*

PROOF. The second assertion is immediate (the elements of S_1 and S_2clearly have different relative positions to x_0). For the first, let $x_1 \in S_1$, $x_2 \in N_{x_1}$, and let $x' \in S_1$ be distinct from x_1. Let $\mathcal{A}$ be an apartment containing the simplexes $\{x_0,x_1\}$ and $\{x_1', x_2\}$. We can choose the co-ordinates in such a way that $x_1 = \underline{e}_1$, $x_1' = \underline{e}_2$, and $x_2 = \underline{e}_1 - \underline{e}_k$ for some $k \neq 1$. Assuming $k \neq 2$, the distance squared between x_1' and x_2 is:

$$\langle \underline{e}_1 - \underline{e}_2 - \underline{e}_k, \underline{e}_1 - \underline{e}_2 - \underline{e}_k \rangle = 9 - \frac{1}{n},$$

while the squared distance between adjoining vertices is easily seen to be $1 - \frac{1}{n}$. In the case $k = 2$ the squared distance is $5 - \frac{1}{n}$. □

COROLLARY 5.1.20. *Let* $S = S_1 \cup S_2$*, let* $x, y \in S$*, and let* $a \in A$ *satisfy* $ax = y$*. Then* $a \in KZ$.

LEMMA 5.1.21. $\#S_1 \sim q^{n-1}$, $\#S_2 \sim q^{2(n-1)}$.

PROOF. By Lemma 5.1.16, we have $\#N(x,\underline{e}_n), \#N(x,-\underline{e}_n) \sim q^{n-1}$ for any vertex x. Since $\#N_{x_1} = \#N(x_1,-\underline{e}_n) - 1$ for any $x_1 \in S_1$ and since these are all disjoint, we see that $\#S_2 \sim q^{n-1}\#S_1$ and that it suffices to show $\#N(x_0,\underline{e}_n) - \#S_1 \ll q^{n-2}$.

Since $N(x_0, \underline{e}_n) \setminus S_1$ consists of those elements of $N(x_0, \underline{e}_n)$ that are also neighbours of another vertex of $\mathcal{A}_0$, its cardinality is at most the size of the link of x_0 in $\mathcal{A}_0$ times a bound for the number of 2-simplexes in $\mathcal{B}^2$ containing a fixed 1-simplex. Fixing a neighbour $x \in \mathrm{Lk}^0(x_0)$ is equivalent to giving a lattice $p\Lambda_0 < \Lambda < \Lambda_0$. In this language we need to enumerate lattices Λ' of index p in Λ_0 containing Λ. Reducing mod $p\Lambda_0$ this is equivalent to enumerating the subspaces of $\mathbb{F}_q^n$ of codimension 1 containing a fixed non-trivial subspace. Dualizing, we need to bound the number of 1-dimensional subspaces of $\mathbb{F}_q^n$ contained in a fixed proper subspace, a number which is easily verified to be $\sim q^{d-1}$ when $d \leq n-1$ is the dimension of the subspace.

Since the structure of the apartment $\mathcal{A}_0$ is independent of q, the size of the link is uniformly bounded (in fact by $(n-1) \cdot n!$) and we are done. □

DEFINITION 5.1.22. Let $\tilde{S}_1 = \{(x,1) \mid x \in S_1\}$, $\tilde{S}_2 = \{(x',0) \mid x' \in S_2\}$, $\tilde{S} = \tilde{S}_1 \cup \tilde{S}_2$.

PROPOSITION 5.1.23. *(transversals) $\tilde{S} \subset \tilde{\mathcal{B}}^0$ is contained in the G-orbit of $(x_0, 0)$, i.e. in the image of G/K. Moreover it does not intersect the A-orbit of $(x_0, 0)$ and if $a \in A$ carries $x \in \tilde{S}$ to $y \in \tilde{S}$ then $a \in K$.*

PROOF. The first assertion is clear by construction. The second assertion follows from noting that every $x \in S$ satisfies $\mathrm{pr}_{|\mathcal{A}|}(x) = x_0$ while every $x' \in Ax_0$ is fixed by this projection, and that $x_0 \notin S$. For the last claim assume $a(x, \epsilon) = (y, \delta)$ with $\epsilon, \delta \in \{0, 1\}$. From $ax = y$ we conclude $a = kz$ for some $k \in K$, $z \in Z$. From $v(\det a) + \epsilon = \delta$ we conclude that $|v(\det(z))| \leq 1$, and since it is a multiple of n we must have $v(\det(z)) = 0$. This actually implies $z \in K$ and we are done. □

5.2. Hecke eigenfunctions – the local contribution

We keep here the notation of the previous section. However, we assume $p \notin R$ and identify $G_p \simeq \mathrm{GL}_d(\mathbb{Q}_p)$ with $\mathbb{D}^\times(\mathbb{Q}_p)$. Let $\psi \in L^2(X, \omega)$ be our Hecke eigenfunction. To any $x_\infty \in G_\infty$ we have associated its *p-Hecke orbit* $\{\mathbb{G}(\mathbb{Q}) x_\infty g_p K_{\mathrm{f}}\}_{g_p \in G_p} \subset X$. This is isomorphic to a quotient G_p/K_p, and by assumption the restriction f of ψ to this orbit is a Hecke eigenfunction on G_p/K_p. The following Proposition can best be described by saying f cannot be too concentrated on apartments: if $f(1)$ is large, then f must also be large on the transversal $\tilde{S}$ which lies away from the apartment. It demonstration relies on a bound toward the Generalized Ramanujan Conjecture, the proof of which is reproduced below.

FACT 5.2.1. *There exists $\delta > 0$ such that the Hecke eigenvalue λ considered below satisfies:*

$$|\lambda| \ll (\#S_1)^{\frac{1}{2}} \cdot q^{\frac{1}{2} - \delta}. \tag{5.2.1}$$

PROPOSITION 5.2.2. *("part of the tree" method) Let f be obtained as above. Then*

$$\sum_{\tilde{x} \in \tilde{S}} |f(\tilde{x})|^2 \gg \frac{1}{q^{1-2\delta}} |f(1)|^2 .$$

PROOF. Let $L_1 = \sum_{\tilde{x}\in\tilde{S}_1} f(\tilde{x})$, $L_2 = \sum_{\tilde{x}\in\tilde{S}_2} f(\tilde{x})$. There exist $\lambda = \lambda_f(\underline{e}_n, 1)$ depending only on f and the relative position $(\underline{e}_n, 1)$ such that for each $x \in S_1$ we have:

$$\sum_{x'\in N_x} f(x', 0) + f(x_0, 0) = \lambda \cdot f(x, 1).$$

Summing over S_1 and using the disjointness of the union defining S_2 we have:

$$L_2 + \#S_1 \cdot f(1) = \lambda \cdot L_1.$$

Therefore, at least one of the following holds:

$$\begin{aligned} |L_1| &\gg \frac{\#S_1}{|\lambda|}\,|f(1)| \\ |L_2| &\gg \#S_1\,|f(1)|\,. \end{aligned}$$

Squaring, and using Cauchy-Schwartz, we get one of:

$$\begin{aligned} \sum_{\tilde{x}\in\tilde{S}_1} |f(\tilde{x})|^2 &\gg \frac{\#S_1}{|\lambda|^2}\,|f(1)|^2 \\ \sum_{\tilde{x}\in\tilde{S}_2} |f(\tilde{x})|^2 &\gg \frac{(\#S_1)^2}{\#S_2}\,|f(1)|^2\,. \end{aligned}$$

Using Lemma 5.1.21 and the bound (5.2.1) complete the proof. □

Digression: proof of the estimate 5.2.1. Our eigenfunction $\psi \in L^2(X,\omega)$ generates an subrepresentation $\tilde{\pi} \subset L^2(\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A}),\omega)$ of $\mathbb{G}(\mathbb{A})$. Since ψ is invariant under right translations by the maximal compact subgroup K_p , there must exist an irreducible $\pi \subset \tilde{\pi}$ containing a non-zero K_p-invariant vector. It is easy to see that any K_p-invariant vector in $\tilde{\pi}$ must have the same eigenvalue λ w.r.t. the Hecke operator under consideration as ψ, and we may thus switch to the case where ψ is a K_p-spherical vector in an irreducible subrepresentation $\pi \subset L^2(\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A}),\omega)$.

The component π_p of this representation at the place p is then a spherical representation of $G_p \simeq \mathrm{GL}_d(\mathbb{Q}_p)$ and hence isomorphic to the spherical constituent of the representation of $\mathrm{GL}_d(\mathbb{Q}_p)$ induced from the character $\mathrm{diag}(a_1,\ldots,a_d) \mapsto \prod_j |a_j|_p^{\mu_j+it/d}$ of A_p (where $t \in \mathbb{R}$ and $\sum_j \mu_j = 0$).

The eigenvalue λ is (up to normalization) the eigenvalue of the convolution operator $\pi_p(1_{N(x_0,\underline{e}_n)})$ acting on the spherical vector of π_p, where $1_{N(x_0,\underline{e}_n)}$ is the characteristic function of the subset $\cup_{x\in N(x_0,\underline{e}_n)} xK_p$ of G_p. This can be computed explicitly in terms of the parameters μ_j:

THEOREM 5.2.3. *(the Satake Isomorphism; see* [**26**]*) To the convolution operator associated to the characteristic function of KaK with $a \in A$ it is possible to associate a permutation-invariant polynomial $P(x_1,\ldots,x_n)$ such that its eigenvalue acting on the spherical function of π_p is given by $P(q^{\mu_1+it/d},\ldots,q^{\mu_d+it/d})$. If $a = \mathrm{diag}(q^{r_1},\ldots,q^{r_d})$ with $0 \le r_1 \le \cdots \le r_d$, a monomial of maximal degree in P is $\prod_j x_j^{r_j}$.*

In our case we get a symmetric polynomial P of degree 1 in d variables such that

$$\lambda = P(q^{\mu_1}, \dots, q^{\mu_n})q^{it}$$

for all choices of $\{\mu_j\}, t$. Since there is a unique such polynomial up to rescaling, we have:

$$\lambda_f = cq^{it}(\sum_j q^{\mu_j}).$$

We evaluate the constant c by considering the special case $t = 0$, $\mu_j = \frac{d+1}{2} - j$, where π_p is the trivial representation. In that case f (the restriction of ψ to an orbit G_p/K_p) is constant and we have the explicit evaluation $\lambda_f = \#N(x_0, \underline{e}_d) \sim q^{d-1}$. We conclude that as $q \to \infty$

$$c = \frac{\#N(x_0, \underline{e}_d)}{\sum_j q^{\frac{d+1}{2}-j}} \sim q^{\frac{d-1}{2}}.$$

As $\#S_1 \sim q^{d-1}$ this means:

$$|\lambda_f| \sim (\#S_1)^{1/2} \left|\sum_j q^{\mu_j}\right|.$$

We now obtain a bound on the parameters μ_j in three steps.

First, we construct an automorphic representation of $\mathrm{GL}_d(\mathbb{A})$ which also has π_p as its local component at p:

THEOREM 5.2.4. *(Arthur-Clozel; see* [**2**]*) Let π be an automorphic representation on $\mathbb{D}^\times(\mathbb{A})$. Then there exists an automorphic representation Π on $\mathrm{GL}_d(\mathbb{A})$ in the discrete spectrum such that for every finite place v where $\mathbb{D}$ splits we have $\pi_v \simeq \Pi_v$.*

Secondly, we argue that in the case where d is prime Π is in fact a *cuspidal* representation: the classification of the residual spectrum due to Mœglin-Waldspurger [**24**] implies that for d prime the discrete non-cuspidal spectrum of $\mathrm{GL}_d(\mathbb{A})$ consists of 1-dimensional representations. Π is not a character since π_p isn't.

Thirdly, the cuspidality implies a bound on the spectral parameters of $\Pi_p \simeq \pi_p$:

THEOREM 5.2.5. *(Luo-Rudnick-Sarnak; see* [**21**]*) Let Π be a cuspidal automorphic representation of $\mathrm{GL}_d(\mathbb{A})$. At every place v where Π_v is unramified, let it be the unitary spherical constituent of the representation induced from the character $\mathrm{diag}(a_1, \dots, a_d) \mapsto \prod_j |a_j|_v^{\mu_j + it/d}$ of A_v (where $t \in \mathbb{R}$ and $\sum_j \mu_j = 0$). Then*

$$|\Re \mu_j| \le \frac{1}{2} - \frac{1}{d^2+1}.$$

This gives $|\lambda| \ll (\#S_1)^{\frac{1}{2}} \cdot q^{\frac{1}{2}-\delta}$ where $\delta = \frac{1}{d^2+1}$. We also note that our estimate of c above shows that the implied constant is independent of q.

5.3. Split Tori

From here on we return to the usual notations: $G = \mathbb{G}(\mathbb{R}) \simeq \mathrm{GL}_d(\mathbb{R})$ etc. In order to apply the Diophantine results of the previous chapter, we need to fix $\tilde{C} \subset A_a M_a$ and

$\Omega_\infty \subset G$ as in the beginning of Section 4.4. We retain the notations $B = B_a(C, \epsilon)$ and $\tilde{B}$ as defined there.

We first estimate the denominator of an element of $\mathbb{G}(\mathbb{A}_f)$ in terms of the geometry of the building. Recall that for each $p \notin R$ we fixed an algebra isomorphism $\varphi_p \colon \mathbb{D}(\mathbb{Q}_p) \to M_d(\mathbb{Q}_p)$ such that $\varphi_p(\mathcal{O}_p) = M_d(\mathbb{Z}_p)$, and let $\mathbb{T}(\mathbb{Q}_p) \subset \mathbb{G}(\mathbb{Q}_p)$ be the inverse image under φ of the torus of diagonal matrices $A_p < \mathrm{GL}_d(\mathbb{Q}_p)$. Pulling back the Cartan decomposition (Fact 5.1.15) we see that for every $g_p \in G_p$ there exists a unique $a_p \in A_p^+$ and some $k_p, k_{p'} \in \mathcal{O}_p^\times = K_p$ such that $g_p = k_p \varphi_p^{-1}(a_p) k_p'$. If a_p has co-ordinates $\underline{r} = (r_1 \leq \cdots \leq r_n)$ we write[1] $r_p(g_p) = \max\{-r_1, r_n\}$. Necessarily a non-negative number, we call it the *radius* of g_p. It is immediate that $r_p(g_p^{-1}) = r(g_p)$.

LEMMA 5.3.1. $\mathrm{d}_p(g_p) \leq p^{r_p(g_p)}$, *hence also* $\mathrm{d}_p(g_p^{-1}) \leq p^{r_p(g_p)}$.

PROOF. Let a_p be defined as above, and let $z \in Z_{\mathrm{GL}_d(\mathbb{Q}_p)}$ be the scalar matrix p^{-r_1} so that $z a_p \in M_d(\mathbb{Z}_p)$. Since φ_p is an algebra homomorphism, we conclude that $p^{-r_1} g_p \in \mathcal{O}_p$, and hence that $x_i(g_p) \in p^{r_1}\mathbb{Z}_p$ for all i. □

Now let T be as in Theorem 4.4.4. For each prime $p \notin R$, let $\bar{S}_p$ denote all elements of G_p of radius at most 1 such that $h_p(g_p) \leq p$ (we set $\bar{S}_p = \emptyset$ if $p \in R$) also identify every $h \in \bar{S}_p$ with the element $g \in \mathbb{G}(\mathbb{A}_f)$ such that $g_p = h$ and $g_{p'} = 1$ for all primes $p' \neq p$. Finally, given $\epsilon > 0$ let:

$$\bar{S}_\epsilon = \cup_{p^2 \leq \epsilon^{-T}} \bar{S}_p \subset \mathbb{G}(\mathbb{A}_f).$$

This family of sets is T-bounded by construction. It follows that for ϵ small enough, we can associate to each $x \in \Omega$ a commuting subset $\left\{\gamma^{(j)}\right\}_{j=1}^d \subset \mathcal{O}$, of discriminants bounded by $O(\epsilon^{-T'})$, such that every $\gamma \in \mathbb{G}(\mathbb{Q})$ causing an intersection for xB w.r.t. Hecke translation by $\bar{S}_\epsilon$ lies in the subalgebra $F = \mathbb{Q}\left(\gamma^{(1)}, \ldots, \gamma^{(d)}\right) \subset \mathbb{D}(\mathbb{Q})$, which is isomorphic to a number field also to be denoted F. Given this data we let $E = \mathbb{Z}[\gamma^{(1)}, \ldots, \gamma^{(d)}]$denote the subring of $\mathcal{O}$ generated by the $\gamma^{(j)}$, and let $D = O(\epsilon^{-dT'})$ denote the product of their discriminants, a multiple of the discriminant of$E$. Since $\mathcal{O}$ is a $\mathbb{Z}$-algebra of finite type all its elements are integral over Z. In particular we have $E \subseteq \mathcal{O}_F$ and hence the discriminant of F divides D. Reflecting this we set

$$R_D = R \cup \{p \mid p|D\}, \qquad P_\epsilon = \left\{p \leq \epsilon^{-T/2}\right\} \setminus R_D.$$

Let $\mathbb{T}_F \subset \mathbb{G}$ be the (maximal) $\mathbb{Q}$-torus such that $\mathbb{T}_F(\mathbb{Q}) = F^\times$. We will be interested in the $\mathbb{Q}_p$-points of this torus, a subtorus of G_p. Clearly $\mathbb{T}_F(\mathbb{Q}_p) = (F \otimes_\mathbb{Q} \mathbb{Q}_p)^\times \subset \mathbb{D}(\mathbb{Q}_p)^\times$. As is well-known, $F \otimes_\mathbb{Q} \mathbb{Q}_p \simeq \oplus_{v|p} F_v$ where the direct sum is over the places of F lying over p. We thus have:

$$(F \otimes_\mathbb{Q} \mathbb{Q}_p)^\times \simeq \prod_{v|p} F_v^\times.$$

We now assume $p \notin R_D$. Then every $v \in |F|$ lying above p is unramified, and hence p is still a uniformizer of F_v so that $F_v^\times = \mathbb{Q}_p^\times \mathcal{O}_{F_v}^\times$. In fact, $F_v^\times = \mathbb{Q}_p^\times F_v^1$, where $F_v^1 =$

[1] This definition is independent of the choice of isomorphism φ_p, but we shall not need this fact.

$\{x \in F_v \mid N_{\mathbb{Q}_v}^{F_v}(x) = 1\} \subset \mathcal{O}_{F_v}$. We thus have:

$$\mathbb{T}_F(\mathbb{Q}_p) \simeq \left(\prod_{v|p} \mathbb{Q}_p^\times\right) \times \left(\prod_{v|p} F_v^1\right).$$

Setting $\mathbb{T}_F^{\mathrm{sp}}(\mathbb{Q}_p) = \prod_{v|p} \mathbb{Q}_p^\times$, $\mathbb{T}_F^{\mathrm{an}}(\mathbb{Q}_p) = \prod_{v|p} F_v^1$ we note that these are the $\mathbb{Q}_p$-points, respectively, of the maximal split and anisotropic $\mathbb{Q}_p$-subtori of $\mathbb{T}_F$. In other words, we have just written our torus as an almost-direct product of its split and anisotropic parts.

LEMMA 5.3.2. *(torus orbit contained in an apartment) Assume $p \notin R_D$. Then there exist $k_p \in K_p$ for which $k_p^{-1}\mathbb{T}_F^{sp}(\mathbb{Q}_p)k_p \subset \mathbb{T}(\mathbb{Q}_p)$. In addition, $\mathbb{T}_F^{an}(\mathbb{Q}_p) \subset K_p$, so that $\mathbb{T}_F(\mathbb{Q}_p) \subset k_p\mathbb{T}(\mathbb{Q}_p)K_p$.*

PROOF. We first note that since $\mathcal{O}$ is a free $\mathbb{Z}$-module of finite rank, every element of $\mathcal{O}$ is integral over $\mathbb{Z}$. In particular, the elements of the ring $E \subset F$ (defined above) are algebraic integers of F. Since the $\gamma^{(j)}$ generate F, we see that E is an order of F, and its discriminant divides D. Since p does not divide D this implies that E is dense in $\mathcal{O}_{F_v}$ for any place $v \in |F|$ lying above p. For the remainder of this proof v will denote such a place, and sums or products will be over the set of places of F lying above p.

The proof that F is dense in $\oplus_v F_v$ can be extended to show that $E \otimes \mathbb{Z}_p$ is dense in $\oplus_v \mathcal{O}_{F_v}$ and hence that

$$\oplus_v \mathcal{O}_{F_v} = E \otimes \mathbb{Z}_p \subset \mathcal{O}_p.$$

Restricting our attention to the invertible elements we conclude that

$$\mathbb{T}_F^{\mathrm{an}}(\mathbb{Q}_p) \subset \prod_v \mathcal{O}_{F_v}^\times \subset K_p.$$

In order to diagonalize the split part we let $x_v \in F \otimes \mathbb{Q}_p$ denote the idempotent given by the identity element of F_v under the isomorphism of $F \otimes \mathbb{Q}_p$ with $\oplus_v F_v$. Since

$$\mathbb{T}_F^{\mathrm{sp}}(\mathbb{Q}_p) = \left\{\sum_v a_v x_v \mid a_v \in \mathbb{Q}_p^\times\right\},$$

it suffices to simultaneously diagonalize the $\{x_v\}$. Since $x_v \in \mathcal{O}_{F_v}$, the previous discussion shows that $x_v \in \mathcal{O}_p$. Applying the isomorphism φ_p it now suffices to show that a family $\{x_v\}$ of commuting idempotents in $\varphi(\mathcal{O}_p) = M_d(\mathbb{Z}_p)$ can be diagonalized by an element of $\mathrm{GL}_d(\mathbb{Z}_p)$. Equivalently we need to find a minimal generating set of the standard lattice $\Lambda = \oplus_i \mathbb{Z}_p \underline{e}_i \subset \mathbb{Q}_p^d$ which consists of joint eigenvectors of the x_v.

For each choice of eigenvalues $\varepsilon_v \in \{0, 1\}$, we set $P(\underline{\varepsilon}) = \prod_v (-1)^{\varepsilon_v}(x_v - \varepsilon_v) \in M_d(\mathbb{Z}_p)$. Since the x_v commute, this is a collection of commuting idempotents as well. Furthermore, it is easy to check that $\sum_{\underline{\varepsilon}} P(\underline{\varepsilon}) = \prod_v (x_v + (1 - x_v)) = 1$. This way we obtain a direct sum decomposition $\Lambda = \oplus_{\underline{\varepsilon}} \Lambda_{\underline{\varepsilon}}$ where $\Lambda_{\underline{\varepsilon}} = P(\underline{\varepsilon})\Lambda$. It is clear that each $\Lambda_{\underline{\varepsilon}}$ is a torsion-free $\mathbb{Z}_p$-module consisting of those elements $t \in \Lambda$ for which $x_v t = \varepsilon_v t$ for every v. In particular, we can choose a $\mathbb{Z}_p$-basis for each of them. Combining these bases we obtain the desired basis for Λ, and hence an element of $\mathrm{GL}_d(\mathbb{Z}_p)$ conjugating the $\{x_v\}$ to diagonal $0 - 1$ matrices. □

For each prime $p \in P_\epsilon$ now fix k_p as in the lemma and let $\tilde{S}$ be the transversal constructed in Definition 5.1.22. The first claim of Proposition 5.1.23 assures us we can choose a set of representatives $g_s \in \mathrm{GL}_d(\mathbb{Q}_p)$ such that $\{g_s(x_0, 0)\} = \tilde{S}$.

COROLLARY 5.3.3. *Let $S_p = \left\{k_p\varphi_p^{-1}(g_s)k_p^{-1}\right\}_{s\in\tilde{S}}$. Then:*

(1) *$S_p \subset \bar{S}_p$. In other words, every $g_p \in S_p$ is of radius at most 1 and height at most p.*
(2) *We have $S_p \subset \bar{S}_p \setminus \mathbb{T}_F(\mathbb{Q}_p)K_p$.*
(3) *For $x \neq y \in S_p$ and $g \in \mathbb{T}_F(\mathbb{Q}_p)$ $gx = y$ implies $g \in k_p^{-1}K_pk_p = K_p$.*
(4) *For every $x_\infty \in G$ we have*

$$\sum_{g_p\in S_p} \left|\psi(Z\Gamma x_\infty g_p)\right|^2 \gg \frac{1}{p^{1-\delta}} \left|\psi(Z\Gamma x_\infty)\right|^2 .$$

PROOF. (1) The elements of $\tilde{S}$ are of relative positions $\underline{e}_n$ and $-\underline{e}_1 + \underline{e}_n$ to the identity coset, respectively. Their pull-backs by φ_p^{-1} are thus of radius 1 and reduced norm of absolute value either p or 1. Since multiplication by an element of K_p on the left or right does not change the radius or the height of an element of G_p, the same holds for the elements of S_p.

(2) and (3) follow directly from the corresponding claims of Proposition 5.1.23 via the Lemma. Part (4) follows from 5.2.2. □

Again constructing our set of translates place-by-place, we set:

$$S_\epsilon = \bigcup_{p^2\leq\epsilon^{-T}} S_p.$$

LEMMA 5.3.4. *(intersections only occur place-by-place) Let γ cause an intersection for Hecke translates of $x_\infty B$ by S_ϵ. Then there exists a prime p such that $\gamma_{p'} \in K_{p'}$ for all $p' \neq p$.*

PROOF. Recall the basic observation from the previous chapter: if $g, g' \in S_\epsilon$ are distinct and γ causes an intersection between $x_\infty Bg$ and $x_\infty Bg'$ then (the finite part of equation (4.4.1)):

$$\gamma \in g'Kg^{-1}.$$

Let $g \in S_p$, $g' \in S_{p'}$. If $p'' \neq p, p'$ then for $g_{p''}, g'_{p''} \in K_{p''}$ so $\gamma_{p''} \in K_{p''}$. If $p = p'$ we are done. In the case $p' \neq p$ the p'-component of g, is an element $g_{p'} \in K_{p'}$. We then read off $\gamma_{p'} \in g'_{p'}K_{p'}$. Since $\gamma \in \mathbb{T}_F(\mathbb{Q}_p)$ this means $g'_{p'} \in \mathbb{T}_F(\mathbb{Q}_p)K_p$, which contradicts the construction of $S_{p'}$ as interpreted in the first part of Corollary 5.3.3. □

The main geometric property of our construction is now clear:

PROPOSITION 5.3.5. *There exists finite subset $I \subset \Gamma$ such that for ϵ small enough, the set of $\gamma \in \mathbb{G}(\mathbb{Q})$ that cause intersections for $\{x_\infty Bg\}_{g\in S_\epsilon}$ ($x_\infty \in \Omega_\infty$ fixed) is contained in I. In particular, any point of the union $x_\infty B \bigcup \cup_{g\in S_\epsilon} x_\infty Bg \subset X_Z$ is contained in at most $|I|$ of the translates forming the union.*

PROOF. Recalling the observation leading to equation (4.4.3), we set

$$Q = \left\{ g \in G_\infty^1 \mid |\det(g)| = 1 \wedge \exists z_\infty \in \mathbb{R}^\times : g z_\infty \in \Omega_\infty B_a(\tilde{C}, O(1))\Omega_\infty^{-1} \right\}.$$

For $g \in Q$ and z_∞ as in the definition, we have $|\det(z_\infty)| = |\det(g z_\infty)|$ belonging to a compact subset of $\mathbb{R}^\times$. It follows that Q is relatively compact, and we will see that $I = \Gamma \cap Q$ works as claimed.

Let $\gamma \in F^\times = \mathbb{T}_F(\mathbb{Q})$ cause an intersection between $x_\infty B g$ and $x_\infty B g'$. By the Lemma we have $g, g' \in S_p$ for some p, so that $\gamma_{p'} \in K_{p'}$ for all $p' \neq p$. At the place p itself the second claim of Corollary 5.3.3 now gives $\gamma_p \in K_p$, so that $\gamma \in K_{\mathrm{f}}$, i.e. $\gamma \in K_{\mathrm{f}} \cap \mathbb{G}(\mathbb{Q}) = \Gamma$.

Next, $\gamma \in K_{\mathrm{f}}$ implies $|\nu(\gamma)|_\infty = \prod_p |\nu(\gamma_p)|^{-1} = 1$ so that $|\det(\gamma)| = 1$ and $\gamma \in Q$.

Finally, let $y \in x_\infty B g$ with $g \in S_p$. Then, if $y \in x_\infty B g'$ for some other $g' \in S_\epsilon$ we must have $g' \in S_p$ and some $\gamma \in I$ causing that intersection. As usual we write this in the form:

$$\gamma_p g_p K_p = g'_p K_p.$$

In particular, the coset $g'_p K_p \in G_p / K_p$ can be recovered from γ and g. Now since S_p was chosen to be a system of representatives for a set of such cosets, it follows that g' is uniquely determined by γ, so that there can be at most $|I|$ such g'. □

5.4. The proof of theorem 5.0.1

In summary, we fixed an open compact subgroup $K_{\mathrm{f}} < \mathbb{G}(\mathbb{A}_{\mathrm{f}})$, an element $a \in A \setminus Z$, a compact fundamental domain $\Omega_\infty \subset G$, and relatively compact neighbourhood $C, \tilde{C} \subset M_a A_a$.

Then, for $\epsilon > 0$ small enough, we have found in order a number field F with discriminant bound D controlling the intersections, a set of primes P_ϵ avoiding ramification, and finally a set of Hecke translates S_ϵ satisfying both geometric and spectral properties. We now show that for any central character ω unramified at R, Hecke eigenfunction ψ and any $x \in X_Z$, $\mu_\psi(xB)$ decays polynomially with ϵ.

PROOF. Choose some $x_\infty \in \Omega_\infty$ projecting to $x \in X_Z$, and let $1_{x_\infty B g}$ denote the characteristic function of the translate $x_\infty B g \subset X_Z$, Proposition 5.3.5 can be interpreted to read:

$$\sum_{g \in S_\epsilon} 1_{x_\infty B g}(y) \leq |I|.$$

Multiplying by $|\psi(y)|^2$ and integrating over X_Z we conclude:

$$\sum_{g \in S_\epsilon} \mu_\psi(x_\infty B g) \leq |I|.$$

Recall the construction of S_ϵ as $\cup_{p \in P_\epsilon} S_p$. Changing the order of summation and integration, we obtain:

$$\sum_{g \in S_p} \mu_\psi(x_\infty B g) = \int_B \sum_{g \in S_p} |\psi(x_\infty b g)|^2 \, dm(b)$$

Where dm is the Haar measure on X_Z. We now apply part (4) of Corollary 5.3.3 and conclude

$$\sum_{g\in S_p} \mu_\psi(x_\infty Bg) \gg \frac{1}{p^{1-\delta}} \int_B |\psi(x_\infty b)|^2 \, dm(b) = \frac{1}{p^{1-\delta}} \mu_\psi(xB).$$

Summing over $p \in P_\epsilon$ we get:

$$\sum_{g\in S_\epsilon} \mu_\psi(x_\infty Bg) \gg \left(\sum_{p\in P_\epsilon} \frac{1}{p^{1-\delta}} \right) \mu_\psi(xB).$$

Since

$$\sum_{p\in R_D} \frac{1}{p^{1-\delta}} \leq \sum_{p\in R} \frac{1}{p^{1-\delta}} + \log D \ll \log \epsilon^{-1},$$

while

$$\sum_{p^2 \leq \epsilon^{-T}} \frac{1}{p^{1-\delta}} \gg \epsilon^{-T\delta/3},$$

the latter expression also bounds the asymptotics of $\sum_{p\in P_\epsilon} p^{-1+\delta}$. We thus have:

$$\mu_\psi(xB_a(C,\epsilon)) \ll \epsilon^{T\delta/3}.$$

We remark that the implicit constant indeed only depends on a, on properties of $\mathbb{D}$ and K_{f} such that the set R of ramified places and the structure constants a_{ijk}, and finally on the choices of $\Omega_\infty, \tilde{C}$. The exponent $\eta = T\delta/3$, furthermore, only depends on the degree d of the division algebra (since T depends on that and on the dimension $r < d^2$ of the subalgebra spanned by $M_a A_a$). $\square$

Bibliography

1. D. V. Anosov, *Geodesic flows on closed Riemannian manifolds of negative curvature*, Trudy Mat. Inst. Steklov. **90** (1967), 209. MR MR0224110 (36 #7157)
2. James Arthur and Laurent Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Annals of Mathematics Studies, vol. 120, Princeton University Press, Princeton, NJ, 1989. MR **MR1007299 (90m:**22041)
3. Jean Bourgain and Elon Lindenstrauss, *Entropy of quantum limits*, Comm. Math. Phys. **233** (2003), no. 1, 153–171.
4. Yves Colin de Verdière, *Ergodicité et fonctions propres du laplacien*, Comm. Math. Phys. **102** (1985), no. 3, 497–502. MR **87d:**58145
5. Jacques Dixmier, *Enveloping algebras*, Graduate Studies in Mathematics, vol. 11, American Mathematical Society, Providence, RI, 1996, Revised reprint of the 1977 translation. MR **97c:**17010
6. J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan, *Erratum: "Spectra of compact locally symmetric manifolds of negative curvature"*, Invent. Math. **54** (1979), no. 1, 101. MR **82a:**58050b
7. ______, *Spectra of compact locally symmetric manifolds of negative curvature*, Invent. Math. **52** (1979), no. 1, 27–93. MR **82a:**58050a
8. Martin Eichler, *Lectures on modular correspondences*, Lectures on Mathematics, vol. 9, Tata Institue of Fundamental Research, Bombay, 1957.
9. Manfred Einsiedler and Anatole Katok, *Invariant measures on G/Γ for split simple Lie groups G*, Comm. Pure Appl. Math. **56** (2003), no. 8, 1184–1221, Dedicated to the memory of Jürgen K. Moser. MR **2004e:**37042
10. Manfred Einsiedler, Anatole Katok, and Elon Lindenstrauss, *Invariant measures and the set of exceptions to littlewood's conjecture*, http://www.math.princeton.edu/ẽlonl/Publications/EKLslnr.pdf.
11. Richard P. Feynman, *Statistical mechanics: a set of lectures*, Frontiers in Physics, W. A. Benjamin, Inc., Reading, Massachusetts, 1972.
12. Ramesh Gangolli and V. S. Varadarajan, *Harmonic analysis of spherical functions on real reductive groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete [Results in Mathematics and Related Areas], vol. 101, Springer-Verlag, Berlin, 1988. MR **89m:**22015
13. Sigurdur Helgason, *Groups and geometric analysis*, Mathematical Surveys and Monographs, vol. 83, American Mathematical Society, Providence, RI, 2000, Integral geometry, invariant differential operators, and spherical functions, Corrected reprint of the 1984 original. MR **2001h:**22001
14. Eberhard Hopf, *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939), 261–304. MR 1,243a
15. H. Iwaniec and P. Sarnak, *L^∞ norms of eigenfunctions of arithmetic surfaces*, Ann. of Math. (2) **141** (1995), no. 2, 301–320. MR **96d:**11060
16. Anthony W. Knapp, *Representation theory of semisimple groups*, Princeton Mathematical Series, vol. 36, Princeton University Press, Princeton, NJ, 1986, An overview based on examples. MR **87j:**22022
17. Elon Lindenstrauss, *Adelic dynamics and arithmetic quantum unique ergodicity*, to be published in the Proceedings of the 2005 Current Developments in Mathematics Conference.
18. ______, *On quantum unique ergodicity for $\Gamma\backslash\mathbb{H}\times\mathbb{H}$*, Internat. Math. Res. Notices (2001), no. 17, 913–933. MR **2002k:**11076
19. ______, *Invariant measures and arithmetic quantum unique ergodicity*, preprint (2003), (54 pages).

20. Elon Lindenstrauss and Akshay Venkatesh, *Existence and weyl's law for spherical cusp forms*, math.NT/0503724.
21. Wenzhi Luo, Zeév Rudnick, and Peter Sarnak, *On the generalized Ramanujan conjecture for* $\mathrm{GL}(n)$, Automorphic forms, automorphic representations, and arithmetic (Fort Worth, TX, 1996), Proc. Sympos. Pure Math., vol. 66, Amer. Math. Soc., Providence, RI, 1999, pp. 301–310. MR **2000e:**11072
22. Wenzhi Luo and Peter Sarnak, *Quantum variance for Hecke eigenforms*, Ann. Sci. École Norm. Sup. (4) **37** (2004), no. 5, 769–799. MR MR2103474
23. Stephen D. Miller, *On the existence and temperedness of cusp forms for* $\mathrm{SL}_3(\mathbb{Z})$, J. Reine Angew. Math. **533** (2001), 127–169. MR **2002b:**11070
24. C. Mœglin and 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]. MR **MR1361168 (97d:**11083)
25. Zeév Rudnick and Peter Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. **161** (1994), no. 1, 195–213. MR **95m:**11052
26. Ichirô Satake, *Theory of spherical functions on reductive algebraic groups over* $\mathfrak{p}$*-adic fields*, Inst. Hautes Études Sci. Publ. Math. (1963), no. 18, 5–69. MR MR0195863 (33 #4059)
27. Lior Silberman and Akshay Venkatesh, *On quantum unique ergodicity for locally symmetric spaces I*, preprint available at http://arxiv.org/abs/math/407413.
28. A. I. Šnirel′man, *Ergodic properties of eigenfunctions*, Uspekhi Mat. Nauk **29** (1974), no. 6(180), 181–182. MR 53 #6648
29. Christopher D. Sogge, *Concerning the* L^p *norm of spectral clusters for second-order elliptic operators on compact manifolds*, J. Funct. Anal. **77** (1988), no. 1, 123–138. MR **89d:**35131
30. Thomas Watson, *Rankin triple products and quantum chaos*, Ph.D. thesis, Princeton University, 2003.
31. André Weil, *Basic number theory*, third ed., Springer-Verlag, New York, 1974, Die Grundlehren der Mathematischen Wissenschaften, Band 144. MR MR0427267 (55 #302)
32. Eugene Wigner, *On the quantum correction for thermodynamic equilibrium*, Phys. Rev. **40** (1932), no. 5, 749–759.
33. Scott A. Wolpert, *Semiclassical limits for the hyperbolic plane*, Duke Math. J. **108** (2001), no. 3, 449–509. MR **2003b:**11051
34. Steven Zelditch, *Pseudodifferential analysis on hyperbolic surfaces*, J. Funct. Anal. **68** (1986), no. 1, 72–105. MR **87j:**58092
35. ______, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. **55** (1987), no. 4, 919–941. MR **89d:**58129
36. ______, *The averaging method and ergodic theory for pseudo-differential operators on compact hyperbolic surfaces*, J. Funct. Anal. **82** (1989), no. 1, 38–68. MR **91e:**58194