



HAL
open science

THE MAXIMAL GROWTH OF TORIC PERIODS AND OSCILLATORY INTEGRALS FOR MAXIMAL FLAT SUBMANIFOLDS

Bart Michels

► **To cite this version:**

Bart Michels. THE MAXIMAL GROWTH OF TORIC PERIODS AND OSCILLATORY INTEGRALS FOR MAXIMAL FLAT SUBMANIFOLDS. 2022. hal-03695051

HAL Id: hal-03695051

<https://hal.science/hal-03695051>

Preprint submitted on 14 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THE MAXIMAL GROWTH OF TORIC PERIODS AND OSCILLATORY INTEGRALS FOR MAXIMAL FLAT SUBMANIFOLDS

BART MICHELS

ABSTRACT. We prove a new omega result for toric periods of Hecke-Maass forms on compact locally symmetric spaces associated to forms of \mathbf{PGL}_3 . This is motivated by conjectures on the maximal growth of L -functions as well as by questions about the size of automorphic periods. We also prove a mean square asymptotic result for maximal flat periods on more general locally symmetric spaces of non-compact type, which takes as main input bounds for real relative orbital integrals.

1. INTRODUCTION

In this article we prove that toric periods of Hecke-Maass forms on compact locally symmetric spaces associated to forms of \mathbf{PGL}_3 exhibit nontrivial growth. In particular, the result implies the following.

Theorem 1.1. *Let \mathbf{G} be an anisotropic \mathbb{Q} -form of \mathbf{PGL}_3 with $\mathbf{G}(\mathbb{R})$ noncompact. Let X be an associated locally symmetric space and $(f_j) \in L^2(X)$ an orthonormal basis of Hecke-Maass forms with Laplacian eigenvalues $\lambda_j \geq 0$. Let $\mathbf{H} \subset \mathbf{G}$ be a maximal torus with the same \mathbb{R} -rank r as \mathbf{G} , and denote by $\mathcal{P}_{\mathbf{H}}(f_j)$ the \mathbf{H} -period of f_j . Then*

$$\max_{\lambda_j \leq x} \lambda_j^{r/4} |\mathcal{P}_{\mathbf{H}}(f_j)| \gg (\log \log x)^{1/4+o(1)}$$

as $x \rightarrow +\infty$.

The more precise result is Theorem 1.2. We make two immediate remarks, on which we expand further in the remainder of this introduction. The first concerns the factor $\lambda_j^{r/4}$ in the left-hand side. Because the \mathbf{H} -period is not zero-dimensional, the quadratic mean of the periods $\mathcal{P}_{\mathbf{H}}(f_j)$ over eigenvalue intervals of size $\sqrt{\lambda}$ around $\lambda > 0$, decays by a theorem of Zelditch [56], with rate $\lambda^{-r/4}$. In other words, the periods have size $\asymp \lambda_j^{-r/4}$ “on average”, and the theorem says that the periods are unbounded when normalized accordingly. In fact, proving a finer spectral parameter version of the result of Zelditch is an essential part of the proof of Theorem 1.1, and we state it in Theorem 1.3. We add here that stating the bound in terms of the Laplacian eigenvalue is an oversimplification, and in fact, has the risk of making the growth rate meaningless because the average decay rate is likely to have a smaller exponent near singular spectral parameters; see also §1.3. We did so for the sake of exposition.

2010 *Mathematics Subject Classification.* 11F03, 11F72.

Université Sorbonne Paris Nord, LAGA, CNRS, UMR 7539, F-93430, Villetaneuse, France.
Email: michels@math.univ-paris13.fr.

The second remark concerns the double logarithmic growth rate in the right-hand side. At first sight, this may seem disappointing when compared with results of this type for (forms of) \mathbf{PGL}_2 [37, 39], or with results that give polynomial growth of periods (§1.3). But in fact we have a strong reason to believe that in this case the double logarithmic growth is best possible, up to the exponent $1/4$. And moreover, that this is related to the maximal growth of L -functions at the edge of their critical strip, and should be compared to the conjecture [21] that

$$\max_{1 \leq |t| \leq x} |\zeta(1 + it)| \sim e^\gamma \log \log x.$$

We expand on this at the end of §1.4.

1.1. The spectral growth of periods. Let \mathbf{G} be a semisimple algebraic \mathbb{Q} -group and \mathbf{H} a closed algebraic subgroup that we assume to be anisotropic, meaning that the adelic points $\mathbf{H}(\mathbb{A}_{\mathbb{Q}})$ have compact image $[\mathbf{H}]$ in the automorphic quotient $[\mathbf{G}] = \mathbf{G}(\mathbb{Q}) \backslash \mathbf{G}(\mathbb{A}_{\mathbb{Q}})$. For any automorphic form f on $[\mathbf{G}]$ one may consider the period $\mathcal{P}_{\mathbf{H}}(f)$ by integrating along a suitable translate of $[\mathbf{H}]$. We consider families of forms by allowing certain parameters to vary, and we also speak of aspects. Our focus is on the spectral aspect family of Hecke-Maass forms: Fix a maximal compact subgroup of $\mathbf{G}(\mathbb{R})$ and a level structure, and let \mathcal{F} be the family of spherical cusp forms on the locally symmetric space resulting from these choices. The aim of this article is to advance our understanding of the following question.

When $\lambda(f)$ denotes the Laplacian eigenvalue of $f \in \mathcal{F}$, how fast does

$$(1) \quad \max_{\substack{f \in \mathcal{F} \\ \lambda(f) \leq x}} |\mathcal{P}_{\mathbf{H}}(f)|$$

grow, if at all, as $x \rightarrow \infty$?

To state the question correctly one needs normalizing factors in front of the periods (as in (8) and Theorem 1.2 below). Further, $\lambda(f)$ is not always the correct measure of complexity for the \mathbf{H} -period (see §1.3). But we ignore these details for now. We will be interested in sub-polynomial growth rates, whose significance is best understood in the context of extreme values of L -functions. In fact, question (1) is in many ways similar to the following question, and this similarity will be our basis for revealing possible arithmetic information about periods.

When \mathcal{F} is a reasonable family of L -functions (not necessarily known to be automorphic) for \mathbf{GL}_n , how fast does

$$(2) \quad \max_{\substack{L \in \mathcal{F} \\ C(L) \leq x}} |L(\tfrac{1}{2})|$$

grow as $x \rightarrow \infty$?

Here $C(L)$ denotes the analytic conductor [27], and we have scaled all L -functions to have central value at $\frac{1}{2}$. A classical example is the family of unramified unitary Hecke characters, leading to vertical shifts of the Riemann zeta function: $L(|\cdot|_{\infty}^{it}, s) = \zeta(s + it)$ (the t -aspect for ζ), in which case $C(L) \asymp t$.

Many problems in analytic number theory are related to understanding the size of the central value $L(\frac{1}{2})$ in families. For every reasonable family we can formulate the Lindelöf hypothesis that $L(\frac{1}{2}) \ll_{\epsilon} C(L)^{\epsilon}$ and the subconvexity problem which asks for a $\delta > 0$ such that $L(\frac{1}{2}) \ll C(L)^{1/4 - \delta}$. In particular, the growth of $|L(\frac{1}{2})|$ is

sub-polynomial under the Lindelöf hypothesis. It is conjectured [18] that the true growth of (2) is

$$(3) \quad \max_{\substack{L \in \mathcal{F} \\ C(L) \leq x}} |L(\tfrac{1}{2})| = \exp\left((C_{\mathcal{F}} + o(1))\sqrt{\log x \cdot \log \log x}\right)$$

for some explicit constant $C_{\mathcal{F}}$ that depends on the family. Proofs conditional on GRH [41, 44] or using the resonance method of Soundararajan [49] typically yield lower bounds of quality

$$(4) \quad \max_{\substack{L \in \mathcal{F} \\ C(L) \leq x}} |L(\tfrac{1}{2})| \gg \exp\left(C\sqrt{\frac{\log x}{\log \log x}}\right)$$

for some $C > 0$. Recent progress as well as lower bounds with additional restrictions on the argument of $L(\frac{1}{2})$ can be found in [5], [6], [12].

1.2. Spectral resonance. The similarity between the problems about automorphic periods and (central) L -values is already visible in the resonance method in [49], where for each family of L -functions under consideration the spectral ingredient is a trace formula: The Fourier inversion formula in the case of ζ , and the Petersson trace formula in the case of L -functions of modular forms. In the seminal article [26], which predates this, the authors use the Selberg trace formula to prove the following result about a compact arithmetic hyperbolic surface X : Let $z \in X$ be a CM point. When (f_j) denotes the sequence of Hecke-Maass forms on X , ordered by increasing eigenvalue λ_j , then

$$(5) \quad \max_{\lambda_j \leq x} |f_j(z)| \gg \sqrt{\log \log x} \quad (x \rightarrow \infty).$$

Their method is known as the amplification method, which bears much similarity to the resonance method of Soundararajan. It may be summarized as follows: Construct a “resonator” $R(f) \geq 0$ with the property that a quotient of the form

$$(6) \quad \frac{\sum_j R(f_j) |f_j(z)|^2}{\sum_j R(f_j)}$$

is large. If this quotient is bigger than a real number $M > 0$, then at least one ‘period’ $|f_j(z)|$ must be bigger than M . When $R(f)$ is defined as a Hecke eigenvalue of f , a sum of the shape of the numerator naturally appears in the relative trace formula, and a sum of the shape of the denominator naturally appears in the trace formula.

The similarity between the two problems is even more explicit in [39], where the lower bound (5) is improved to

$$(7) \quad \max_{\lambda_j \leq x} |f_j(z)| \gg \exp\left(C\sqrt{\frac{\log x}{\log \log x}}\right) \quad (x \rightarrow \infty),$$

for some explicit $C > 0$, by employing a resonator inspired by the resonators used for L -functions in [49]. That this is similar to the growth rate (4) should not come as a surprise: The point evaluation $f_j(z)$ is (at least under a class number one hypothesis) precisely the period that appears in a formula of Waldspurger [53] that relates the period to the central value of a Rankin-Selberg L -function.

In [37] we proved an analogue of (7) with CM points replaced by closed geodesics: for certain compact arithmetic hyperbolic surfaces X and for a closed geodesic $\ell \subset X$,

$$(8) \quad \max_{\lambda_j \leq x} \lambda_j^{1/4} \left| \int_{\ell} f_j \right| \gg \exp \left(C \sqrt{\frac{\log x}{\log \log x}} \right) \quad (x \rightarrow \infty),$$

for some $C > 0$. The factor $\lambda_j^{1/4}$ is a normalizing factor (which in this case can be thought of as a gamma factor), which is there to make the geodesic periods of size 1 on average; see also Theorem 1.3 below.

1.3. Periods beyond \mathbf{GL}_2 . This article originates in an attempt to generalize the result (8) about geodesic periods to higher rank groups. Now, while question (1) is in some aspects similar to question (2), there are important differences. The first is in the problem statement itself: Where for L -functions the correct measure of complexity is the analytic conductor, the situation is much more delicate for periods. The Laplacian eigenvalue $\lambda(f)$ is a naive choice, and the correct replacement is most likely an integral involving an approximate spectral projector around f , which determines the size of mean square asymptotics for periods over $O(1)$ spectral windows (as in Theorem 1.3). For sufficiently generic spectral parameters, that integral should be of size $\lambda(f)^{(n-r-d)/2}$, with n the dimension of the symmetric space, r its rank, and d the dimension of the projection to the symmetric space of the \mathbf{H} -orbit underlying the period. But asymptotically evaluating this integral in terms of the spectral parameter of f , even for generic parameters, is a difficult problem.

The second difference is the following. Whereas in the case of L -functions one would expect the resonance method to always produce a nontrivial result of quality (4) (provided that we know enough about the family under consideration), this not the case for periods. In fact, the spectral resonance method is sometimes fruitless, and sometimes produces extreme values of periods with power growth. Moreover, a larger variety of techniques exist to prove lower bounds for periods. We give a brief historical account of these facts.

The first example of a different technique is given in [47], which is about discrete periods on certain hyperbolic 3-manifolds. The authors use what is known as a distinction method. It employs a vanishing property that says that only a sparse subsequence of Hecke-Maass forms, lying in the image of a theta lift, have nonzero \mathbf{H} -period. They are “distinguished” by \mathbf{H} . This is then contrasted with a mean square asymptotic, which gives the full average of the periods (and which does not see the arithmetic underlying the distinction). If the periods are supported on a sparse subsequence and the average does not know about this, it follows that that the periods must attain large values on the distinguished subsequence. In fact, the sequence is polynomially sparse, and the authors obtain periods with power growth. Note that while the arithmetic properties used are crucial, the Hecke operators do not play a direct role.

The article [40] characterizes the hyperbolic 3-manifolds to which the proof extends, as being those of Maclachlan-Reid type. Moreover, it gives the first example of the second phenomenon mentioned above: power growth obtained from the resonance method. Further power growth results that use the distinction method include the following settings: discrete periods on hyperbolic n -manifolds with $n \geq 5$

[13], and discrete unitary periods for \mathbf{GL}_n (which includes the case of hyperbolic 2 and 3-manifolds) [31].

A vast generalization of power growth results in the other direction, using the resonance method, is the article [10]. In certain situations, one can obtain exceptional sequences of eigenfunctions with periods of power growth using both methods, and show that they are related. We refer to [10, 40] for a discussion of the relation between the two methods. We do note the following: All \mathbb{Q} -groups in these examples are such that $\mathbf{G}(\mathbb{R})$ is not split.

1.4. Sub-polynomial growth. This article is concerned with applications of the resonance method to periods of Hecke-Maass forms in the spectral aspect, in situations that have so far remained gray zones. By “gray zone”, we mean the following. A first condition is that no explicit connection to central L -values is known or conjectured. Indeed, when conjecture (3) does not apply, without heuristics that come from applications of random matrix theory to L -functions, it becomes an interesting question whether there are sufficiently strong arithmetic reasons for the existence of unusually large periods, and if so, what the maximal growth should be. A second condition is that no power growth is expected, or at least that established distinction techniques or resonance techniques do not produce power growth. In the gray zone it is not clear what period growth, if present at all, should be attributed to: heuristics motivated by random behavior of Laplacian eigenfunctions in negative curvature (Berry’s random wave conjecture [2]), random behavior of more exotic underlying arithmetic objects, or neither of these?

In fact, our hope when exploring the gray zone is to reveal arithmetic information: If in some new situation we obtain lower bounds of quality (4), this would be a very strong hint that there is an as of yet unknown relation with central L -values. If we obtain periods with power growth, this might indicate the existence of a functorial lift (although it could be attributed to other factors as well; see [10]). If we obtain growth rates smaller than (4) or nothing at all, this still leaves the possibility of a relation with L -values further to the right of $\frac{1}{2}$, or at worst in the half-plane of convergence.

An example of such an unexplored situation is that of toric periods on locally symmetric spaces of non-compact type associated to a semisimple group \mathbf{G} . Unless \mathbf{G} is isogenous to a product of forms of \mathbf{PGL}_2 , these fall in the gray zone. When the locally symmetric space X is viewed as a disjoint union of classical locally symmetric spaces, then the period along a maximal \mathbb{R} -split torus corresponds to the integral along a flat submanifold of X , which has the property that the intersection with each component is either maximal flat or empty; see §6.4.

To state the main theorems we introduce some minimal notation for spectral parameters. When G is a connected semisimple Lie group with finite center, make a choice of Iwasawa decomposition $G = NAK$. Let $\mathfrak{a} = \text{Lie}(A)$ and define the generic set $(\mathfrak{a}^*)^{\text{gen}} \subset \mathfrak{a}^*$ as the set of elements that are regular and that do not lie in a proper subspace spanned by roots. A locally symmetric space X is assumed to be compatible with the choice of K . When \mathbf{G} is a semisimple group over \mathbb{Q} , we may define the above notation with respect to $\mathbf{G}(\mathbb{R})^0$, and we again assume an associated adelic locally symmetric space X to be compatible with K . The spectral parameters of Hecke-Maass forms on X can then be viewed as elements of $(\mathfrak{a}_{\mathbb{C}})^*$. For any additional notation and terminology used in the theorems below, we refer to §2.1, §2.2 and §6.

We can now state the following theorem.

Theorem 1.2. *Let \mathbf{G} be an anisotropic \mathbb{Q} -form of \mathbf{PGL}_3 with $\mathbf{G}(\mathbb{R})$ noncompact. Let X be an associated adelic locally symmetric space and $(f_j) \in L^2(X)$ an orthonormal basis of Hecke-Maass forms with spectral parameters $\nu_j \in (\mathfrak{a}_{\mathbb{C}})^*$. Let $\mathbf{H} \subset \mathbf{G}$ be a maximal torus with the same \mathbb{R} -rank r as \mathbf{G} , and denote by $\mathcal{P}_{\mathbf{H}}(f_j)$ the \mathbf{H} -period of f_j . Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\text{gen}}$ be compact. There exist $C > 0$ and $\delta > 0$ such that uniformly for $\nu \in D_{\mathfrak{a}^*}$ and $t \in \mathbb{R}$ we have*

$$\max_{\|\nu_j - t\nu\| \leq C} (1+t)^r |\mathcal{P}_{\mathbf{H}}(f_j)|^2 \gg (\log \log(2+t))^{\delta+o(1)}.$$

Moreover, when E is the splitting field of \mathbf{H} , we may take $\delta = 6/[E : \mathbb{Q}]$.

In §6.2 we give the list of groups to which Theorem 1.2 can be applied. There we also show that the associated Lie group $\mathbf{G}(\mathbb{R})$ is either $\mathbf{PGL}_3(\mathbb{R})$ or the quasi-split projective unitary group $\mathbf{PU}(2, 1)$, and in these cases the \mathbb{R} -rank equals 2 or 1 respectively.

We now come back to the question raised earlier: What must the nontrivial growth in Theorem 1.2 be attributed to? We believe the reason for this growth is arithmetic, for the following reason. There is an exceptional theta-correspondence between \mathbf{PGL}_3 and the exceptional group \mathbf{G}_2 , which is related to the maximal toric periods in the theorem. In fact, the period $\mathcal{P}_{\mathbf{H}}(f_j)$ should be, up to other arithmetic factors that are equally mysterious, related to a product of L -values

$$L(\pi_{f_j}, 1)L(\tilde{\pi}_{f_j}, 1),$$

where π_{f_j} denotes the representation generated by f_j and $\tilde{\pi}_{f_j}$ denotes its contragredient. This period relation, while not proven, explains a great deal about the lower bound in Theorem 1.2. First, it leads us to believe that the lower bound can, perhaps, not simply be attributed to random behaviour of Laplacian eigenfunctions on Riemannian manifolds of nonpositive curvature. But in fact, much more can be said. The period formula explains why the growth rate we obtain should be polynomial in $\log \log C(L(\pi_{f_j}, s))$, the double log of the analytic conductor! Indeed, the double logarithm reminds us of the following result of Levinson [33]: There exists a constant $C > 0$ such that for arbitrarily large $t \in \mathbb{R}$, one has

$$|\zeta(1+it)| \geq e^\gamma \log \log t - C.$$

This is certainly not the best known result, but the main term in the right-hand side is conjectured to be optimal [21]. We refer to [1] for the state of the art on the extreme values of $\zeta(1+it)$, and results with lower order terms. The results and conjectures for $\zeta(1+it)$ are exemplary for the more general situation. In fact, there is the following conditional statement [3]: When π is a unitary cuspidal tempered automorphic representation for \mathbf{GL}_n whose Godement-Jacquet L -function $L(\pi, s)$ satisfies the generalized Riemann hypothesis, then

$$(\log \log C(\pi))^{-n} \ll |L(\pi, 1)| \ll (\log \log C(\pi))^n,$$

where the implicit constants depend on n only and we denote $C(\pi)$ for the analytic conductor of $L(\pi, s)$. Moreover, if π is not tempered, one still has similar bounds but with bigger exponents.

Coming back to the interpretation of Theorem 1.2, we conclude the following. First, ignoring for the sake of the argument the other factors in the period formula

mentioned above, it is to be expected that the maximal toric periods exhibit oscillations that are polynomial in $\log \log \lambda_j$. Indeed, this is (at least conjecturally) the largest permitted oscillation of the L -value $L(\pi_{f_j}, 1)L(\tilde{\pi}_{f_j}, 1)$, and based on what we know about $\zeta(1 + it)$ we would expect this maximal oscillation to be almost realized. Second, if forced to make a conjecture about the maximal growth of the periods in Theorem 1.2, it would be that the exponent of $\log \log(2 + t)$ in the right-hand side can be replaced by $12 + o(1)$. (Taking into account the fact the periods appear with a square in the left-hand side.)

About the exponent in the right-hand side, we remark the following: The splitting field E of \mathbf{H} is Galois, and the Galois group embeds naturally as a subgroup of $\mathrm{GL}_2(\mathbb{Z})$ [9, §1.7]. It is well known that finite subgroups of $\mathrm{GL}_2(\mathbb{Z})$ have cardinality at most 12; this is most easily seen by observing that the eigenvalues of a matrix of finite order must be roots of unity. Thus $[E : \mathbb{Q}] \leq 12$, meaning that $\delta = \frac{1}{2}$ is admissible. In the best case, the Galois group is of cardinality 3 and we obtain the exponent 2; still far from the (naive) conjecture that it can be replaced by 12.

The main arithmetic ingredient that goes into the proof of Theorem 1.2 is the optimization problem of finding the best possible resonator sequence. This is what explains the restriction to forms of \mathbf{PGL}_3 . The optimization problem takes as input asymptotics for local p -adic integrals arising in the geometric main term in a relative trace formula, and requires us to construct a suitable Hecke operator. For (forms of) \mathbf{PGL}_n with $n \geq 3$, only for $n = 3$ have we found a suitable winning construction that, when plugged into the optimization problem, yields nonconstant growth. We strongly believe that for $n \geq 4$ no such construction exists, and that the resonance method cannot produce any growth of toric periods for $n \geq 4$. If indeed true, we believe that, roughly speaking, this should be attributed to the heuristic that tori inside \mathbf{PGL}_3 are still relatively large, while they are too small inside \mathbf{PGL}_n , $n \geq 4$. We refer to §8.3 and Remark 8.13 for these negative statements.

1.5. A mean square asymptotic. The main analytic ingredient in the proof of Theorem 1.2 is a (amplified) mean square asymptotic for maximal flat periods. We require averages over spectral windows of bounded size, and even a non-amplified version requires a considerable amount of work. For this ingredient there is no reason to restrict to the Lie groups associated to groups in Theorem 1.2, and we prove a more general result, formulated classically in terms of periods along maximal flat submanifolds.

Theorem 1.3. *Let G be a noncompact connected semisimple Lie group with finite center and rank r , and let X be an associated compact Riemannian locally symmetric space. Assume that at least one of the following holds:*

- G has rank 1;
- $G = \mathrm{SL}_p(\mathbb{R})$ or $\mathrm{SU}(k, p - k)$ with p prime and $0 < k < p$, and X arises from a \mathbb{Q} -form of \mathbf{SL}_p .

Let $(f_j) \in L^2(X)$ be an orthonormal basis of Maass forms. Let $\mathcal{F} \subset X$ be a compact maximal flat submanifold and denote by $\mathcal{P}_{\mathcal{F}}(f_j)$ the period of f_j along \mathcal{F} . Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\mathrm{gen}}$ be compact. There exists $C > 0$ such that uniformly for $\nu \in D_{\mathfrak{a}^*}$ and $t \in \mathbb{R}$ we have

$$\sum_{\|\nu_j - t\nu\| \leq C} |\mathcal{P}_{\mathcal{F}}(f_j)|^2 \asymp \beta(t\nu) \cdot (1 + t)^{-r},$$

where β denotes the Plancherel density.

The restriction to certain locally symmetric spaces in Theorem 1.3 comes from our bounds for orbital integrals in the relative trace formula. These are defined in terms of a parameter $g \in G$, and behave differently when g centralizes a proper portion of the group underlying the period. That is, when g lies in the standard Levi subgroup of a semistandard parabolic other than G or the minimal one (see (2.4)). We have not been able to deal with the integrals for such g , which is why we restrict to settings where they do not appear in the relative trace formula.

Theorem 1.3 tells us that we may view the factor $(1+t)^r$ in the left-hand side in Theorem 1.2 as a normalizing factor that makes the squared periods of size 1 on average. It is consistent with the mean square asymptotic of Zelditch [56] over eigenvalue intervals, but there is no obvious implication between the two results. The relation to Zelditch's result is analogous to the relation of the spectral parameter Weyl law [14] to the classical Weyl law for compact Riemannian manifolds.

The stationary phase analysis in the proof of Theorem 1.3 draws inspiration from the proof of bounds for spherical functions in [15] and generalizes work of Marshall [34] for PGL_2 . Where for PGL_2 one relies on classical facts about the geometry of geodesics in the Poincaré upper half-plane model, the analogues of those facts were not available in the generality needed here and are established in the companion article [38].

The quantification over ν and t in Theorem 1.2 and Theorem 1.3 may be visualized as follows. Fix a closed cone in the interior of $(\mathfrak{a}^*)^{\mathrm{gen}}$. Then $t\nu$ tends to ∞ inside the cone, and the maximum is taken over a ball of bounded radius around $t\nu$. Figure 1 gives a picture when $\mathfrak{g} = \mathfrak{sl}_3(\mathbb{R})$, and the simple roots are denoted by α and β . When $\mathfrak{g} = \mathfrak{sl}_3(\mathbb{R})$, the condition on tempered spectral parameters to not lie in a proper subspace spanned by roots, is equivalent to not being self-dual, and it also appears in [4].

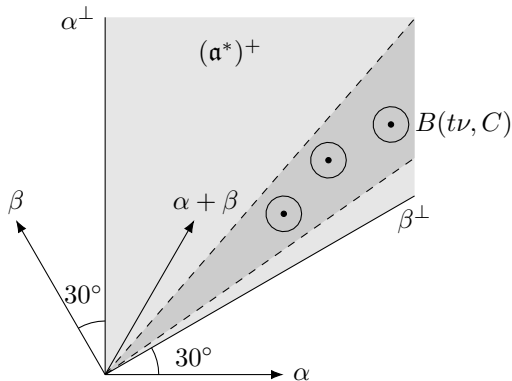


FIGURE 1. A cone in $(\mathfrak{a}^*)^{\mathrm{gen}}$ when $G = \mathrm{SL}_3(\mathbb{R})$, with balls $B(t\nu, C)$ as in Theorem 1.3.

2. PRELIMINARIES

2.1. Lie groups and Lie algebras. Let G be a (possibly disconnected) semisimple Lie group which is reductive in the sense of Harish-Chandra [22]; see also [29, Chapter VII]. That is, G has finite center and acts by inner automorphisms on its

Lie algebra \mathfrak{g} . The group G will almost always be connected semisimple with finite center and we will then simply say G is semisimple; only in §6 and §7 it will be possibly disconnected. Let $K \subset G$ be a maximal compact subgroup and θ a Cartan involution of G whose fixed point set is K . It induces an involution θ of \mathfrak{g} , whose $+1$ - and -1 -eigenspaces we denote by \mathfrak{k} and \mathfrak{p} . We denote the exponential of $X \in \mathfrak{g}$ by $\exp(X)$. Denote by $\langle \cdot, \cdot \rangle$ the Killing form on \mathfrak{g} ; it is positive definite on \mathfrak{k} and negative definite on \mathfrak{p} . Define $\langle \cdot, \cdot \rangle_\theta = \langle \cdot, -\theta(\cdot) \rangle$, a positive definite symmetric bilinear form. All statements on \mathfrak{g} involving norms, orthogonality and adjoints will be with respect to $\langle \cdot, \cdot \rangle_\theta$. Let $\mathfrak{a} \subset \mathfrak{p}$ be a maximal abelian subalgebra and $A = \exp(\mathfrak{a})$. The choices of \mathfrak{a} are all conjugate under K . Define $P = \exp(\mathfrak{p})$. Multiplication $P \times K \rightarrow G$ is a diffeomorphism, known as the Cartan decomposition. In particular, K meets all components of G .

2.2. Symmetric spaces. References for the following facts about symmetric spaces are [16, 24].

Assume here that G is (connected) semisimple. The quotient $S = G/K$ carries a left- G -invariant Riemannian metric induced by the Killing form on \mathfrak{p} . It is a symmetric space of non-compact type, and every such space arises in this way.

The maximal flat submanifolds of S are of the form gAK with $g \in G$. Such g is uniquely determined by the submanifold up to multiplication on the right by $N_G(A)$. When $\dim(A) = 1$, the maximal flats are precisely the geodesics. The rank of G is defined to be $\dim(A)$.

We equip G with any Riemannian metric and the associated distance function $d(\cdot, \cdot)$. We will care only about distances on compact sets and up to constant factors, so we do not need to impose any invariance properties of $d(\cdot, \cdot)$.

2.3. Iwasawa decomposition. Let Σ be the set of restricted roots of \mathfrak{a} in \mathfrak{g} . By convention, $0 \notin \Sigma$. We denote by \mathfrak{g}_α the root space of a root $\alpha \in \Sigma$ and by $H_\alpha \in \mathfrak{a}$ the element corresponding to α under the isomorphism $\mathfrak{a} \cong \mathfrak{a}^*$ given by $\langle \cdot, \cdot \rangle$. Fix a set of positive roots Σ^+ with basis Π . Let $\mathfrak{n} \oplus \mathfrak{a} \oplus \mathfrak{k}$ and $N \times A \times K$ be the corresponding Iwasawa decompositions of \mathfrak{g} and G . Define $M = Z_K(A)$ and $M' = N_K(A)$ and denote by \mathfrak{m} the Lie algebra of M .

Denote the Lie algebra Iwasawa projections by $E_{\mathfrak{n}}$, $E_{\mathfrak{a}}$ and $E_{\mathfrak{k}}$. We have the orthogonal restricted root space decomposition

$$(9) \quad \mathfrak{g} = \mathfrak{a} \oplus \mathfrak{m} \oplus \bigoplus_{\alpha \in \Sigma} \mathfrak{g}_\alpha.$$

Denote the Iwasawa projections from G onto N , A and K by n , a and κ . Define the height $H(g) = \log(a(g)) \in \mathfrak{a}$, the logarithm being the Lie logarithm on A .

2.4. Centralizers. Denote by \mathcal{L} the set of centralizers in G of subgroups of A . They are the standard Levi subgroups of semistandard parabolic subgroups of G . We will denote such a centralizer typically by L . When $L \in \mathcal{L}$ with Lie algebra \mathfrak{l} , define $\mathfrak{a}_L = \mathfrak{z}(\mathfrak{l}) \cap \mathfrak{a}$ and \mathfrak{a}^L its orthogonal complement in \mathfrak{a} . The set \mathcal{L} contains A and G , and $\mathfrak{a}^A = \mathfrak{a}_G = 0$ and $\mathfrak{a}^A = \mathfrak{a}^G = \mathfrak{a}$.

Define the positive Weyl chamber $\mathfrak{a}^+ = \{H \in \mathfrak{a} : \forall \alpha \in \Sigma^+ : \alpha(H) > 0\}$ and the regular set

$$\mathfrak{a}^{\text{reg}} = \mathfrak{a} - \bigcup_{L \neq M} \mathfrak{a}_L = \mathfrak{a} - \bigcup_{\alpha \in \Sigma} \ker(\alpha).$$

We have that $H \in \mathfrak{a}^{\text{reg}}$ if and only if its centralizer equals the centralizer of \mathfrak{a} . Define also the generic set

$$\mathfrak{a}^{\text{gen}} = \mathfrak{a}^{\text{reg}} - \bigcup_{L \in \mathcal{L} - \{G\}} \mathfrak{a}^L.$$

Combined superscripts correspond to intersections: $\mathfrak{a}^{\text{gen},+} = \mathfrak{a}^{\text{gen}} \cap \mathfrak{a}^+$.

We also define $(\mathfrak{a}^*)^{\text{reg}}$, $(\mathfrak{a}^*)^{\text{gen}}$ and $(\mathfrak{a}^*)^+$ to be the corresponding subsets under the isomorphism $\mathfrak{a} \cong \mathfrak{a}^*$ defined by the inner product $\langle \cdot, \cdot \rangle$. When $H \in \mathfrak{a}$ corresponds to $\lambda \in \mathfrak{a}^*$ under the isomorphism given by the Killing form, then $H \in \mathfrak{a}^{\text{reg}}$ if and only if λ is not orthogonal to any roots, and $H \in \mathfrak{a}^{\text{gen}}$ if and only if λ is in addition not contained in a proper subspace spanned by roots.

We will frequently use the facts that $Z_G(A) = MA$ and $N_G(A) = M'A$, as well as the following lemma.

Lemma 2.1. *Let $g \in G$ and $H \in \mathfrak{a}$. If $\text{Ad}_g(H) \in \mathfrak{m} \oplus \mathfrak{a}$, then $g \in M'Z_G(H)$.*

Proof. See for example [38, Lemma 2.2]. □

2.5. Derivatives. When G is any Lie group with Lie algebra \mathfrak{g} and b is an element of the universal enveloping algebra $U(\mathfrak{g})$, we denote by L_b the corresponding left invariant differential operator on $C^\infty(G)$. When $X \in \mathfrak{g} \subset U(\mathfrak{g})$, by definition

$$(L_X f)(g) = \left. \frac{d}{dt} \right|_{t=0} f(ge^{tX}).$$

When $f : M \rightarrow N$ is a differentiable map between differentiable manifolds, denote its differential at $m \in M$ by $(Df)_m$. Using left translation we identify all tangent spaces $T_g G$ with \mathfrak{g} . When $g \in G$, denote by L_g and R_g the left and right multiplication by g on G . With our convention on tangent spaces, we then have for all $g, h \in G$ that

$$\begin{aligned} (DL_g)_h &= \text{id}, \\ (DR_g)_h &= \text{Ad}_{g^{-1}}. \end{aligned} \tag{10}$$

When $X, Y \in \mathfrak{g}$ we have

$$L_X \text{Ad}_g(Y) = \text{Ad}_g([X, Y]), \tag{11}$$

$$L_X \text{Ad}_{g^{-1}}(Y) = -[X, \text{Ad}_{g^{-1}}(Y)]. \tag{12}$$

Lemma 2.2. *Let G be semisimple as in §2.1. The differential of the Iwasawa projection $H : G \rightarrow \mathfrak{a}$ at $g \in G$ is as follows:*

$$(DH)_g = E_{\mathfrak{a}} \circ \text{Ad}_{\kappa(g)}.$$

Proof. See for example [38, Lemma 2.3]. □

2.6. Measures and convolution. We fix any Haar measure dg on G and let dk be the Haar measure on K for which $\text{Vol}(K) = 1$. We fix the Haar measure on A that coincides with the measure induced from the left-invariant Riemannian metric on the symmetric space G/K on the submanifold $A \subset G/K$. If H is a subgroup between A and MA , this then defines a unique Haar measure on H .

The universal enveloping algebra $U(\mathfrak{g})$ acts on $C^\infty(G)$ as in §2.5. The convolution algebra $C_c^\infty(K \backslash G/K)$ acts on $C^\infty(G/K)$ by right translation, and it is commutative [25, Theorem IV.3.1].

2.7. Maass forms. Let G be semisimple and $\Gamma \subset G$ a co-compact and torsion-free lattice. The quotient $X = \Gamma \backslash G/K$ is a compact locally symmetric space, and $C^\infty(\Gamma \backslash G)$ carries an action of $U(\mathfrak{g})$. The algebra $Z(U(\mathfrak{g}))$ is commutative, and we may find an orthonormal basis $(f_j)_{j \geq 0}$ of $L^2(X)$ consisting of simultaneous eigenfunctions for $Z(U(\mathfrak{g}))$, the Maass forms.

For $\nu \in (\mathfrak{a}^*)_{\mathbb{C}}$, the complexified dual of \mathfrak{a} , let $\varphi_\nu : G \rightarrow \mathbb{C}$ be the spherical function of parameter ν [25, Chapter IV]. It is given explicitly by Harish-Chandra's integral representation as

$$(13) \quad \varphi_\nu(g) = \int_K \exp((\rho + i\nu)(H(kg))) dk,$$

where $H : G \rightarrow \mathfrak{a}$ is the Iwasawa projection and $\rho \in \mathfrak{a}^*$ the half-sum of positive roots (§2.1) [25, Theorem IV.4.3]. We say f_j has spectral parameter $\nu_j \in (\mathfrak{a}^*)_{\mathbb{C}}$ if it has the same $Z(U(\mathfrak{g}))$ -eigenvalues as φ_{ν_j} ; such ν_j is uniquely determined up to the action of the Weyl group.

The f_j are eigenfunctions for the algebra $C_c^\infty(K \backslash G/K)$. For any compactly supported smooth kernel $k \in C_c^\infty(K \backslash G/K)$, define the Harish-Chandra transform

$$(14) \quad \begin{aligned} \widehat{k} : (\mathfrak{a}^*)_{\mathbb{C}} &\rightarrow \mathbb{C} \\ \nu &\mapsto \int_G k(g) \varphi_{-\nu}(g) dg. \end{aligned}$$

Then $k * f_j = \widehat{k}(\nu_j) f_j$ (the convolution being on G). We recover k from \widehat{k} by

$$(15) \quad k(g) = \frac{1}{|W|} \int_{\mathfrak{a}^*} \varphi_\nu(g) \widehat{k}(\nu) \beta(\nu) d\nu$$

where $\beta(\nu)$ is the Plancherel density [25, §IV.7], provided that the Haar measure on G in (14) is appropriately normalized (see [25, Exercise IV.C.4]).

2.8. Periods. When $\mathcal{F} \subset X$ is a compact maximal flat submanifold, we may lift it to a maximal flat submanifold of the symmetric space G/K . The lift is the image of a subset of G of the form gA . Such g is uniquely determined by the choice of the lift, up to multiplication on the right by $N_G(A)$. Because \mathcal{F} is compact, Γ intersects gAg^{-1} in a lattice. We define the period of f_j along \mathcal{F} to be the integral

$$(16) \quad \mathcal{P}_{\mathcal{F}}(f_j) = \int_{(\Gamma \cap gAg^{-1}) \backslash gA} f_j.$$

Remark 2.3. The definition (16) does not depend on the choice of g or even on the choice of Iwasawa A -component A . But do note that the inclusion $gA \subset G$ induces a closed embedding of $N_\Gamma(gAg^{-1}) \backslash gA$ in $\Gamma \backslash G/K$ with image \mathcal{F} , and not (always) of the quotient of gA by $\Gamma \cap gAg^{-1}$. That is, one may also consider the integral $\int_{\mathcal{F}} f_j$ with respect to the induced metric on the submanifold $\mathcal{F} \subset G/K$. By our choice of Haar measure on A , the latter integral equals $\mathcal{P}_{\mathcal{F}}(f_j)$ divided by the finite number $|N_\Gamma(gAg^{-1})/(\Gamma \cap gAg^{-1})|$.

3. MEAN SQUARE ASYMPTOTICS FOR MAXIMAL FLAT PERIODS

This section contains the skeleton of the proof of Theorem 1.3. We set up a relative pre-trace formula in §3.1 and prove the theorem in §3.3, using the analytic results from the later sections §4 and §5. For most of this section, G can be any

semisimple Lie group and X any associated compact locally symmetric space. Only in §3.3 will we use that G and X as in Theorem 1.3.

3.1. Classical setup. Let $k \in C_c^\infty(K \backslash G / K)$ be a compactly supported smooth kernel. The automorphic kernel

$$k^{\text{aut}}(x, y) = \sum_{\gamma \in \Gamma} k(x^{-1}\gamma y)$$

acts on $L^2(\Gamma \backslash G)$, and its image lies in the space of right K -invariant functions. If this operator is positive on K -invariants (which will always be the case for us), then it has spectral expansion

$$(17) \quad \sum_{\gamma \in \Gamma} k(x^{-1}\gamma y) = \sum_j \widehat{k}(\nu_j) f_j(x) \overline{f_j(y)},$$

and both sums are uniformly convergent on $X \times X$. When Γ is torsion free and orientation-preserving on the Riemannian manifold G/K , this is a theorem of Mercer [36, p. 445] (see also [54, Satz VI.4.2.]). For the general case, see for example [51].

As in §2.8, let $g \in G$ be such that gA projects to \mathcal{F} in the quotient $\Gamma \backslash G / K$. Define $H = gAg^{-1}$ and $\Gamma_H = \Gamma \cap H$.

Lemma 3.1 (Partitions of unity). *There exists a nonnegative $b \in C_c^\infty(A)$ satisfying $\sum_{\gamma \in \Gamma_H} b(g^{-1}\gamma ga) = 1$ for all $a \in A$.*

Proof. We may find $b_0 \in C_c^\infty(\mathbb{R})$ with the property that $\sum_{n \in \mathbb{Z}} b_0(x+n) = 1$ for all $x \in \mathbb{R}$. Because $\log(\Gamma_H)$ is a lattice of full rank in $\log(H) = \text{Lie}(H)$, out of b_0 we may construct $b \in C_c^\infty(A)$ as in the statement. \square

The group $N_G(H)/H$ is compact, which implies that the discrete subgroup $N_\Gamma(H)/\Gamma_H \subset N_G(H)/H$ is finite.

Lemma 3.2. *There exists a nonnegative $b \in C_c^\infty(A)$ with the property that for all $k \in C_c^\infty(K \backslash G / K)$ we have*

$$\begin{aligned} \sum_j \widehat{k}(\nu_j) |\mathcal{P}_{\mathcal{F}}(f_j)|^2 &= \text{Vol}(\Gamma_H \backslash H) \cdot |N_\Gamma(H)/\Gamma_H| \int_A k(a) da \\ &+ \sum_{\gamma \in \Gamma - N_\Gamma(H)} \int_{A \times A} b(a_1) b(a_2) k(a_1^{-1} g^{-1} \gamma g a_2) da_1 da_2. \end{aligned}$$

Proof. This follows from fairly standard manipulations. Take b to be a cutoff function given by Lemma 3.1. Integrating (17) over $(\Gamma_H \backslash H)^2$ and using (16) gives

$$\begin{aligned} \sum_j \widehat{k}(\nu_j) |\mathcal{P}_{\mathcal{F}}(f_j)|^2 &= \int_{(\Gamma_H \backslash H)^2} \sum_{\gamma \in \Gamma} k(x^{-1}\gamma y) dx dy \\ &= \int_{((g^{-1}\Gamma_H g) \backslash A)^2} \sum_{\gamma \in \Gamma} k(a_1^{-1} g^{-1} \gamma g a_2) da_1 da_2. \end{aligned}$$

We split the γ -sum into sums over the disjoint subsets $N_\Gamma(H)$ and $\Gamma - N_\Gamma(H)$. Because these sets are stable on both sides under Γ_H we may distribute the integral over the two terms. Concretely, the integral is the sum of the “diagonal term”

$$(18) \quad \int_{((g^{-1}\Gamma_H g) \backslash A)^2} \sum_{\gamma \in N_\Gamma(H)} k(a_1^{-1} g^{-1} \gamma g a_2) da_1 da_2$$

and an “off-diagonal term” where the sum goes over $\Gamma - N_\Gamma(H)$. For (18), unfolding in a_2 gives

$$= \int_{((g^{-1}\Gamma_H g \backslash A) \times A)} \sum_{\gamma \in N_\Gamma(H)/\Gamma_H} k(a_1^{-1}g^{-1}\gamma g a_2) da_1 da_2.$$

The γ -sum is now finite. We may change the order of summation and make for fixed γ and a_1 the change of variables in the a_2 -integral that makes the argument of k equal to a_2 , keeping in mind that $g^{-1}N_G(H)g = N_G(A) \subset N_K(A)A$ and that k is bi- K -invariant. This then gives the main term in the statement.

For the off-diagonal term, inserting the partitions of unity from Lemma 3.1, making a change of variables in the γ -sum and unfolding gives that the integral equals

$$\begin{aligned} &= \int_{(g^{-1}\Gamma_H g \backslash A)^2} \sum_{\gamma_1, \gamma_2 \in \Gamma_H} b(g^{-1}\gamma_1 g a_1) b(g^{-1}\gamma_2 g a_2) \\ &\quad \cdot \sum_{\gamma \in \Gamma - N_\Gamma(H)} k(a_1^{-1}g^{-1}\gamma_1^{-1}\gamma\gamma_2 g a_2) da_1 da_2 \\ &= \int_{A \times A} b(a_1) b(a_2) \sum_{\gamma \in \Gamma - N_\Gamma(H)} k(a_1^{-1}g^{-1}\gamma g a_2) da_1 da_2. \end{aligned}$$

Note that this $A \times A \times \Gamma$ -integral is indeed absolutely convergent, even compactly supported thanks to the compact support of b and k . Applying Fubini gives us the sum in the statement. \square

3.2. Test functions. To prove Theorem 1.3 we make a choice of $k \in C_c^\infty(K \backslash G / K)$ to filter out $O(1)$ sums on the spectral side of the pre-trace formula (17). For all spectral parameters ν_j we either have $\nu \in \mathfrak{a}^*$ or $\text{Re}(\nu)$ is singular and $\|\text{Im}(\nu)\| \leq \|\rho\|$ (see [25, §IV.8, Theorem 8.2] for the bound on $\|\text{Im}(\nu)\|$ and [28, Theorem 16.6] for the statement about singularity).

Lemma 3.3. *Let $R > 0$. We may find for every $\nu \in \mathfrak{a}^*$ a $k_\nu \in C_c^\infty(K \backslash G / K)$ with the following properties:*

- (1) $\widehat{k}_\nu(\mu) \geq 0$ for $\mu \in \mathfrak{a}^*$;
- (2) $\widehat{k}_\nu(\mu) \geq 1$ for $\mu \in \mathfrak{a}^*$ with $\|\mu - \nu\| \leq R$;
- (3) $\widehat{k}_\nu(\mu) \ll_N (1 + \|\nu - \mu\|)^{-N}$ uniformly in ν and $\mu \in (\mathfrak{a}_\mathbb{C})^*$ with $\|\text{Im}(\mu)\| \leq R$;
- (4) k_ν has support bounded independently of ν ;
- (5) If G is simple modulo its center, $k_\nu(g) \ll \beta(\nu)(1 + \|\nu\|d(g, K))^{-1/2}$ uniformly in ν and g .

Proof. Such k_ν can be constructed using the Paley-Wiener theorem of Gangolli [19, Theorem 3.5]; see for example [35, §2.2] for a construction, and [10, §4.1] for a proof of the last condition. \square

We will apply Lemma 3.3 with any fixed $R > \|\rho\|$ in order to get decay on the spectrum of $L^2(X)$.

Remark 3.4. The last condition in Lemma 3.3 is only used in the proof of Theorem 1.2. Specifically, in Proposition 7.7.

3.3. Diagonal and off-diagonal estimates. Assume now that G and the locally symmetric space are as in Theorem 1.3. The theorem will be proven by estimating the different terms that appear in the right-hand side in Lemma 3.2. Using the inversion formula (15) it will suffice to bound the analogous terms with the test function k replaced by the spherical function. The results in question are proven in §4 and §5.

Lemma 3.5. *If $\gamma \in \Gamma$ is such that $g_\infty^{-1}\gamma g_\infty \notin M'A$, then $g_\infty^{-1}\gamma g_\infty \notin \bigcup_{L \in \mathcal{L} - \{G\}} M'L$.*

Proof. When G has rank 1 there is nothing to prove because $\mathcal{L} = \{MA, G\}$. Assume now that $G = \mathrm{SL}_p(\mathbb{R})$ or $\mathrm{SU}(k, p-k)$ with p prime, and that Γ comes from a (necessarily anisotropic) \mathbb{Q} -form \mathbf{G} of the algebraic group \mathbf{SL}_p . By the discussion in §6.4, the maximal flat submanifold comes from a maximal torus $\mathbf{H} \subset \mathbf{G}$. If $g_\infty^{-1}\gamma g_\infty \notin M'A$ but it lies in some $M'L$, then the subvariety of \mathbf{H} that is sent to \mathbf{H} by the conjugacy action of γ is a proper subtorus defined over \mathbb{Q} . This is not possible because p is prime, in fact, the algebraic group \mathbf{G} has no nontrivial proper closed connected subgroups other than its maximal tori [20, Proposition 4.1, Corollary 4.2]. \square

Remark 3.6. In general, even when a locally symmetric space $X = \Gamma \backslash G/K$ is compact, it can happen that $g^{-1}\Gamma g$ intersects the groups $L \in \mathcal{L} - \{G, MA\}$ (see §2.4) nontrivially. When for example $G = \mathrm{PGL}_4(\mathbb{R})$ and Γ is an arithmetic lattice coming from a degree 4 division algebra D over \mathbb{Q} , then \mathcal{F} corresponds to a quartic totally real field $F \subset D$. Because F is quartic, it has a quadratic subfield E . It gives rise to a 1-dimensional subgroup $H' \subset H$ and the centralizer $Z_D(E)$ gives rise to a hyperbolic plane $\mathcal{H} \subset G/K$ and an infinite discrete subgroup of $\Gamma \cap Z_G(H')$ that acts co-compactly on \mathcal{H} .

Lemma 3.7. *Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\mathrm{gen}}$ be compact and let k_ν for $\nu \in \mathfrak{a}^*$ be as in Lemma 3.3. Uniformly for $\nu \in D_{\mathfrak{a}^*}$ and $t \geq 1$ we have*

$$\sum_j \widehat{k}_{t\nu}(\nu_j) |\mathcal{P}_{\mathcal{F}}(f_j)|^2 \asymp \beta(t\nu) \cdot (1+t)^{-r}.$$

Proof. With the analytic results from §4 and §5, this follows from standard properties of the Plancherel density and our assumptions on k_ν . We apply Lemma 3.2 to $k = k_{t\nu}$. Let $b_0 \in C_c^\infty(K \backslash G/K)$ with the property that $b_0(g) = 1$ when $g \in \bigcup_{\nu \in \mathfrak{a}^*} \mathrm{supp}(k_\nu)$; this exists by Lemma 3.3. Using (15) and Fubini the first term in the right-hand side of Lemma 3.2 can be written as

$$\mathrm{Vol}(\mathcal{F}) \cdot |N_\Gamma(H)/\Gamma_H| \int_{\mathfrak{a}^*} \widehat{k}_{t\nu}(\mu) \beta(\mu) \int_A b_0(a) \varphi_\mu(a) da d\mu.$$

The contribution of $\mu \in \mathfrak{a}^*$ with $\|\mu - t\nu\| \geq \sqrt{t}$ (say) can be bounded trivially using the rapid decay of $\widehat{k}_{t\nu}$ (Lemma 3.3), the polynomial growth of $\beta(\mu)$ [25, §IV.7, Proposition 7.2] and the bound $\varphi_\mu(g) \ll 1$ which follows for example from the fact that φ_μ is positive definite [25, Exercise IV.B.9]. For the remaining $\mu \in \mathfrak{a}^*$ we have $\mu \in (\mathfrak{a}^*)^{\mathrm{gen}}$ when t is sufficiently large, and we may apply the asymptotic from Proposition 4.1 to the inner integral. Note that for such μ we have $\beta(\mu) \asymp \beta(t\nu)$ by the almost-polynomial behavior of β [14, Lemma 3.11]. Finally, using the positivity of $\widehat{k}_{t\nu}$ and the lower bound from Lemma 3.3 we get that the double integral is $\asymp \beta(t\nu)(1+t)^{-r}$.

The other terms in Lemma 3.2 are finite in number by the support condition on $k_{t\nu}$. By our assumptions on G and X , Lemma 3.5 says that $g^{-1}\gamma g \notin \bigcup_{L \in \mathcal{L}-\{G\}} M'L$ for such terms. Using Proposition 5.1 and entirely similar arguments as for the diagonal term, one shows that they are $\ll \beta(t\nu)(1+t)^{-r-\delta}$ for some $\delta > 0$. \square

Proof of Theorem 1.3. We wish to replace the weight $\widehat{k}_{t\nu}$ in Lemma 3.7 by a sharp cutoff. This can be done by applying the lemma to various $t\nu \in \mathfrak{a}^*$; see for example [10, Lemma 4.5]. \square

4. ARCHIMEDEAN MODEL INTEGRALS

The main result of this section is the following proposition, whose proof is in §4.4. The notations for Lie groups and Lie algebras are as in §2.1.

Proposition 4.1. *Let G be a semisimple Lie group and $b \in C_c^\infty(A)$ with $b(1) > 0$. Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\text{gen}}$ be a compact set. Then*

$$\int_A \varphi_{it\nu}(a)b(a)da \asymp (1+t)^{-r},$$

uniformly for $t \in \mathbb{R}$ and $\nu \in D_{\mathfrak{a}^*}$.

4.1. Setup. We fix $b \in C_c^\infty(A)$ with $b(1) > 0$ and we will not indicate the dependence on b in our notations. Take $\nu \in \mathfrak{a}^*$ and let $H_0 \in \mathfrak{a}$ be the corresponding element under the isomorphism given by the Killing form. Define

$$I(H_0) = \int_A \varphi_{i\nu}(a)b(a)da.$$

Inserting Harish-Chandra's formula for the spherical function (13) yields the oscillatory integral

$$(19) \quad I(H_0) = \int_A \int_K \exp(i\phi_{H_0}(a, k)) b'(a, k) dk da,$$

with phase function

$$(20) \quad \phi_{H_0}(a, k) = \langle H_0, H(ka) \rangle$$

and with an amplitude function $b' \in C_c^\infty(A \times K)$ satisfying $b'(1, k) = b(1) > 0$. We will determine the critical points of ϕ_{H_0} and obtain Proposition 4.1 as an application of the stationary phase method.

4.2. Structure of the critical set. By Lemma 2.2 we have for $H \in \mathfrak{a}$ that

$$(21) \quad (D\phi_{H_0}(\cdot, k))_a(H) = \langle H_0, \text{Ad}_{\kappa(ka)}(H) \rangle.$$

Define the set

$$(22) \quad \mathcal{C}(G, H_0) = \{k \in K : \text{Ad}_{k^{-1}}(H_0) \perp \mathfrak{a}\}.$$

Lemma 4.2. *When $H_0 \in \mathfrak{a}^{\text{gen}}$, the set of critical points of ϕ_{H_0} is $\{1\} \times \mathcal{C}(G, H_0)$.*

Proof. Assume (a, k) is a critical point of ϕ_{H_0} . By [15, Proposition 5.4], criticality in k is equivalent to

$$k \in Z_K(H_0)M'Z_K(a) = M'Z_K(a),$$

where we have used that $H_0 \in \mathfrak{a}^{\text{reg}}$, so that $Z_K(H_0) = M$. By (21) and the definition (22), criticality in a is equivalent to $\kappa(ka) \in \mathcal{C}(G, H_0)$. Because $k \in M'Z_K(a)$ we have $\kappa(ka) = k$, so that $k \in \mathcal{C}(G, H_0)$. It remains to show that $a = 1$.

If not, we would have $k \in M'L$ with $L = Z_G(a) \in \mathcal{L}$ a Levi subgroup different from G . Writing $k = m\ell$, the condition that $\text{Ad}_{k^{-1}}(H_0) \perp \mathfrak{a}_L$ implies $H_0 \perp \text{Ad}_{m^{-1}}(\mathfrak{a}_L)$ and so $H_0 \in \mathfrak{a}^{m^{-1}Lm}$. This contradicts the hypothesis that $H_0 \in \mathfrak{a}^{\text{gen}}$. \square

Remark 4.3. When more generally $H_0 \in \mathfrak{a}^{\text{reg}}$, by being more careful in the proof of Lemma 4.2 one shows that the critical set of ϕ_{H_0} is

$$\bigcup_{L \in \mathcal{L}} \text{Ad}_{M'}(A_L) \times \mathcal{C}(L, H_0),$$

where the sets $\mathcal{C}(L, H_0)$ are defined analogously by

$$\mathcal{C}(L, H_0) = \{k \in K \cap L : \text{Ad}_{k^{-1}}(H_0) \perp \mathfrak{a}\},$$

and only the sets $\mathcal{C}(L, H_0)$ with $H_0 \in \mathfrak{a}^L$ are nonempty.

Define the map

$$(23) \quad \begin{aligned} f_{H_0} : K &\rightarrow \mathfrak{a} \\ k &\mapsto E_{\mathfrak{a}}(\text{Ad}_{k^{-1}}(H_0)), \end{aligned}$$

which has the properties that $\mathcal{C}(G, H_0) = f_{H_0}^{-1}(0)$ and

$$(24) \quad (D\phi_{H_0}(\cdot, k))_{\mathfrak{a}}(H) = \langle f_{H_0}(\kappa(ka)), H \rangle.$$

Remark 4.4. The set $\mathcal{C}(G, H_0)$ has the geometric interpretation that it consists of the $k \in K$ for which the maximal flat $kA \subset G/K$ is orthogonal to H_0 at the base point $k \cdot 1$. The geometric picture will come more fully into its own right in §5.

The main properties of $\mathcal{C}(G, H_0)$ are established in [38].

Lemma 4.5. *When $H_0 \in \mathfrak{a}^{\text{gen}}$, the function f_{H_0} is a submersion at the points of $\mathcal{C}(G, H_0)$. The set $\mathcal{C}(G, H_0)$ is a nonempty smooth submanifold of K of codimension $\dim(A)$ that varies smoothly with $H_0 \in \mathfrak{a}^{\text{gen}}$. It is disjoint from the sets $M'L$ with $L \in \mathcal{L} - \{G\}$.*

Proof. That f_{H_0} is a submersion at $\mathcal{C}(G, H_0)$ is [38, Corollary 4.9]. The second sentence is [38, Proposition 4.19]. The third sentence is [38, Lemma 4.7], although the (short) argument for this already appeared in the proof of Lemma 4.2 above. \square

Note that the smoothness and the codimension of $\mathcal{C}(G, H_0)$ are immediate consequences of the fact that f_{H_0} is a submersion, but the fact that it is nonempty, is a major result established in [38].

4.3. Stationary phase. Recall the map $f_{H_0} : K \rightarrow \mathfrak{a}$ defined in (23).

Lemma 4.6. *For $k \in \mathcal{C}(G, H_0)$ we have*

$$T_k(\mathcal{C}(G, H_0)) = \ker((Df_{H_0})_k) = \{X \in \mathfrak{k} : \text{Ad}_k(X) \perp [H_0, \text{Ad}_k(\mathfrak{a})]\}.$$

Proof. The first statement holds because the defining map f_{H_0} for $\mathcal{C}(G, H_0)$ is a submersion on $\mathcal{C}(G, H_0)$ (Lemma 4.5). Indeed, it is a general fact that when a submanifold is the level set of a submersion, then its tangent spaces are the kernels of the differential of the defining map [32, Proposition 5.38]. The differential may be computed using (12):

$$(Df_{H_0})_k(X) = -E_{\mathfrak{a}}([X, \text{Ad}_k^{-1}(H_0)]).$$

Therefore $X \in \ker((Df_{H_0})_k)$ if and only if $[X, \text{Ad}_k^{-1}(H_0)] \perp \mathfrak{a}$. Using associativity and Ad_k -invariance of the Killing form, this is equivalent to $\text{Ad}_k(X) \perp [H_0, \text{Ad}_k(\mathfrak{a})]$. \square

The first equality in Lemma 4.6 will be used in the following lemma. The second will be used in the proof of Lemma 5.21.

When $M \subset N$ are Riemannian manifolds and $m \in M$, a symmetric bilinear form on $T_m N$ is called transversely nondegenerate to M if its radical is contained in $T_m M$. Recall the phase function ϕ_{H_0} defined by (20).

Lemma 4.7. *When $H_0 \in \mathfrak{a}^{\text{gen}}$, the Hessian of ϕ_{H_0} is transversely nondegenerate to the critical set of ϕ_{H_0} . Moreover, its signature on the critical set equals $(n_0, n_+, n_-) = (\dim(K) - r, r, r)$, where $r = \dim(A)$.*

Proof. The Hessian of ϕ_{H_0} at a critical point (a, k) is given simply by

$$(X, Y) \mapsto L_X L_Y \phi_{H_0}(a, k).$$

By Lemma 4.2 the critical set of ϕ_{H_0} is $\{1\} \times \mathcal{C}(G, H_0)$. Let $(1, k)$ be a critical point of ϕ_{H_0} . Let $L : \mathfrak{a} \oplus \mathfrak{k} \rightarrow \mathfrak{a} \oplus \mathfrak{k}$ be the unique and self-adjoint linear map such that

$$(\text{Hess}_{(1,k)} \phi_{H_0})(X, Y) = \langle LX, Y \rangle_{\theta}$$

for all $X, Y \in \mathfrak{a} \oplus \mathfrak{k}$. We must show that $\ker L \subset \{0\} \oplus T_k \mathcal{C}(G, H_0)$. Write

$$L = \begin{pmatrix} L_{\mathfrak{a}\mathfrak{a}} & L_{\mathfrak{a}\mathfrak{k}} \\ L_{\mathfrak{k}\mathfrak{a}}^* & L_{\mathfrak{k}\mathfrak{k}} \end{pmatrix}$$

relative to the natural decomposition of $\mathfrak{a} \oplus \mathfrak{k}$. Because $\phi_{H_0}(1, k) = 0$ for all $k \in K$, it is clear that $L_{\mathfrak{k}\mathfrak{k}} = 0$. To compute $L_{\mathfrak{a}\mathfrak{k}} : \mathfrak{k} \rightarrow \mathfrak{a}$, let $X \in \mathfrak{k}$ and $H \in \mathfrak{a}$. From (24) we have $L_H \phi_{H_0}(1, k) = \langle f_{H_0}(k), H \rangle$. Therefore

$$\begin{aligned} \text{Hess}_{(1,k)} \phi_{H_0}(X, H) &= L_{XH} \phi_{H_0}(a, k) \\ &= L_X \langle f_{H_0}(k), H \rangle \\ &= \langle (Df_{H_0})_k(X), H \rangle, \end{aligned}$$

so that $L_{\mathfrak{a}\mathfrak{k}} = (Df_{H_0})_k$. From Lemma 4.5 it follows that $L_{\mathfrak{a}\mathfrak{k}}$ is surjective, so that the adjoint $L_{\mathfrak{k}\mathfrak{a}}^*$ is injective. This implies

$$\ker L = \{0\} \oplus \ker L_{\mathfrak{a}\mathfrak{k}} = \{0\} \oplus \ker((Df_{H_0})_k) = \{0\} \oplus T_k \mathcal{C}(G, H_0),$$

where the last equality is Lemma 4.6. This proves the first part of the statement.

We can compute the signature at a critical point $(1, k)$ as follows. Let V be a complement of $T_k \mathcal{C}(G, H_0)$ in \mathfrak{k} . Relative to the decomposition $\mathfrak{a} \oplus V \oplus T_k \mathcal{C}(G, H_0)$, the map L has the form

$$L = \begin{pmatrix} L_{\mathfrak{a}\mathfrak{a}} & L_{\mathfrak{a}V} & 0 \\ L_{\mathfrak{a}V}^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with $L_{\mathfrak{a}V}$ invertible, because $\ker(L_{\mathfrak{a}\mathfrak{k}}) = T_k \mathcal{C}(G, H_0)$. A self-adjoint map this form has signature $(n_0, n_+, n_-) = (\dim(\mathcal{C}(G, H_0)), r, r)$. Indeed, to show this we must show that a real symmetric $2r \times 2r$ block matrix of the form

$$M = \begin{pmatrix} A & B \\ B^T & 0 \end{pmatrix}$$

with B invertible, has signature (r, r) . Doing as if the blocks are numbers, define

$$\Lambda_{\pm} = \frac{1}{2}(A \pm \sqrt{A^2 + 4BB^T}),$$

where the square root is the symmetric positive definite one. Because “square roots are monotone on symmetric positive definite matrices”, Λ_+ is positive definite and Λ_- is negative definite. Now one can either write down explicit matrices to show that M is congruent to $\text{diag}(\Lambda_+, \Lambda_-)$, or observe that if (v_i^{\pm}) are bases of eigenvectors for Λ_{\pm} , then the vectors $(\Lambda_+ v_i^+, B^T v_i^+)$ and $(\Lambda_- v_i^-, B^T v_i^-)$ form a basis of eigenvectors for M , with the same eigenvalues as those of Λ_{\pm} . \square

We are ready to prove Proposition 4.1. We use the stationary phase theorem with critical manifolds of positive dimension, which we now recall.

Theorem 4.8. [11, Théorème 4.1] *Let M be a smooth Riemannian manifold, $b \in C_c^\infty(M)$ and $\phi \in C^\infty(X)$ real-valued. Assume that the set of critical points of ϕ intersects the support of b in a smooth closed submanifold $W \subset M$ of codimension d , and that the Hessian of ϕ is transversely nondegenerate to W . Denote for $w \in W$ by $\text{Hess}_{w,\perp} \phi$ the restriction of the Hessian to the orthogonal complement of $T_w W$. Assume that ϕ is constant on W and that $\text{Hess}_{\perp} \phi$ has constant signature (n_+, n_-) on W . Then as $t \rightarrow +\infty$,*

$$\int_M e^{it\phi(x)} b(x) dx = \left(\frac{2\pi}{t}\right)^{d/2} e^{itf(W) + \pi i(n_+ - n_-)/4} \int_W \frac{b(w)}{|\det(\text{Hess}_{w,\perp} \phi)|^{1/2}} dw + O(t^{-d/2-1}),$$

where the integral over W is with respect to the induced metric.

4.4. Proof of Proposition 4.1.

Proof of Proposition 4.1. Take $\nu_0 \in (\mathfrak{a}^*)^{\text{gen}}$ and let $H_0 \in \mathfrak{a}^{\text{gen}}$ correspond to it via the identification $\mathfrak{a} \cong \mathfrak{a}^*$ given by the Killing form. Via the reduction in the beginning of §4 we must bound the oscillatory integrals $I(tH_0)$ given by (19):

$$I(tH_0) = \int_A \int_K \exp(it\phi_{H_0}(a, k)) b'(a, k) dk da.$$

By Lemma 4.7 the Hessian of ϕ_{H_0} has signature $(n_+, n_-) = (r, r)$ transversely to its critical set, which has codimension $2 \dim(A) = 2r$ by Lemma 4.2 and Lemma 4.5, and ϕ_{H_0} takes the value 0 there because $H(K) = 0$. The stationary phase theorem implies

$$I(tH_0) = t^{-r} \cdot (2\pi)^r e^{\pi i(n_+ - n_-)/4} \int_{M'C(G, H_0)} \frac{b'(1, k)}{\sqrt{|\det(\mathbf{H}_k)|}} dk + O(t^{-r-1})$$

as $t \rightarrow +\infty$, where $\mathbf{H}_k = \text{Hess}_{(1,k),\perp} \phi_{H_0}$ is the Hessian of ϕ_{H_0} restricted to the orthogonal complement of the tangent space $T_{(1,k)}(A \times M'C(G, H_0))$, the determinant is taken in an orthonormal basis and the integral is with respect to the induced metric. Moreover, this bound is uniform for H_0 in compact subsets of $\mathfrak{a}^{\text{gen}}$ because of the smooth dependence in Lemma 4.5. We have $b'(1, k) > 0$ (see the beginning of §4), so that the constant in the main term is strictly positive. \square

Remark 4.9. A possible way to remove the condition $\nu \in \mathfrak{a}^{\text{gen}}$ in Proposition 4.1 (or at least to replace $\mathfrak{a}^{\text{gen}}$ by $\mathfrak{a}^{\text{reg}}$) would be to use the more precise statement about the critical set in Remark 4.3 and to resolve the singularities of f_{H_0} and lift the integration in (19) to a blowup of $\mathfrak{a} \times A \times K$ (the \mathfrak{a} -factor corresponding to H_0).

5. BOUNDS FOR ORBITAL INTEGRALS

In this section we prove the following proposition. The notations for Lie groups and Lie algebras are as in §2.1.

Proposition 5.1. *Let G be semisimple and $b \in C_c^\infty(A \times A)$. Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\text{gen}}$ be compact and $D_G \subset G$ be compact. Then there exist $\delta > 0$ and $N > 0$ such that*

$$\int_{A \times A} \varphi_{i\nu}(a_1^{-1}ga_2)b(a_1, a_2)da_1da_2 \\ \ll (1+t)^{-r} \cdot \left(1 + t \cdot d \left(g, \bigcup_{L \in \mathcal{L} - \{G\}} M'L \right)^N \right)^{-\delta},$$

uniformly for $t \in \mathbb{R}$, $\nu \in D_{\mathfrak{a}^*}$ and $g \in D_G$.

Remark 5.2. The dependence on g in the above result is not needed in the proof of Theorem 1.3, because only finitely many $\gamma \in \Gamma$ contribute to the sum in Lemma 3.2. We will need this dependence in the proof of Theorem 1.2.

5.1. Setup and phase functions. We fix $b \in C_c^\infty(A \times A)$ and we will not incorporate it in the notations. Take $\nu \in \mathfrak{a}^*$ and let $H_0 \in \mathfrak{a}$ be the corresponding element under the isomorphism given by the Killing form. For $g \in G$ define

$$(25) \quad J(H_0, g) = \int_{A \times A} \varphi_{i\nu}(a_1^{-1}ga_2)b(a_1, a_2)da_1da_2.$$

By invariance of $\varphi_{i\nu}$ under the action of the Weyl group on ν , we have for $e \in M'$ that

$$(26) \quad J(H_0, g) = J(\text{Ad}_e(H_0), g).$$

It is therefore no restriction to assume that H_0 lies in a Weyl chamber of our choice.

Inserting Harish-Chandra's formula for the spherical function (13) yields the oscillatory integral

$$J(H_0, g) = \int_{A \times A} \int_K \exp\left(i\tilde{\phi}_{H_0, g}(a_1, a_2, k)\right) b_g(a_1, a_2, k) dk da_1 da_2,$$

with phase function

$$(27) \quad \tilde{\phi}_{H_0, g}(a_1, a_2, k) = \langle H_0, H(ka_1^{-1}ga_2) \rangle$$

and with amplitude $b_g \in C_c^\infty(A \times A \times K)$ depending smoothly on g and with support bounded independently of g , which incorporates the real exponential factor in (13). Following [34], we now bring this in a form that makes the A -derivatives more manageable.

For $h \in G$ define the map

$$(28) \quad \begin{aligned} \Theta_h : K &\rightarrow K \\ k &\mapsto \kappa(kh). \end{aligned}$$

By smoothness of the Iwasawa decomposition it is a smooth map, with smooth inverse $\Theta_{h^{-1}}$, and therefore a diffeomorphism. For $k \in K$ and $y, z \in G$ we have that (see [34, Lemma 6.2])

$$H(ky^{-1}z) = H(\Theta_{y^{-1}}(k)z) - H(\Theta_{y^{-1}}(k)y).$$

Applying this with $y = a_1$ and $z = ga_2$ and making the change of variables $k \leftarrow \Theta_{a_1}(k)$ gives

$$(29) \quad J(H_0, g) = \int_K \int_{A \times A} \exp(i\phi_{H_0, g}(a_1, a_2, k)) b'_g(a_1, a_2, k) da_1 da_2 dk$$

with phase function

$$(30) \quad \begin{aligned} \phi_{H_0, g}(a_1, a_2, k) &:= \tilde{\phi}_{H_0, g}(a_1, a_2, \Theta_{a_1}(k)) \\ &= \langle H_0, H(kga_2) \rangle - \langle H_0, H(ka_1) \rangle \end{aligned}$$

and some amplitude $b'_g \in C_c^\infty(A \times A \times K)$ depending smoothly on g .

Remark 5.3. The expression (27) will be useful when computing derivatives of $\phi_{H_0, g}$ with respect to k , and (30) will be useful when computing derivatives with respect to a_1 and a_2 .

The expression (30) separates the variables a_1 and a_2 . Our strategy, inspired by [34], is to first apply the stationary phase theorem in the variables a_1 and a_2 , leaving us with an oscillatory integral over K , and then to apply the van der Corput lemma to this integral.

5.2. Extremal points on maximal flats. In view of equation (30), we are naturally led to study the critical points of the ‘height’ functions

$$\begin{aligned} h_{H_0, g} : A &\rightarrow \mathbb{R} \\ a &\mapsto \langle H_0, H(ga) \rangle \end{aligned}$$

with $g \in G$, which allow us to write

$$(31) \quad \phi_{H_0, g}(a_1, a_2, k) = h_{H_0, kg}(a_2) - h_{H_0, k}(a_1).$$

The critical points of $h_{H_0, g}$ are studied in [38]. Many of the results concerning them require that $H_0 \in \mathfrak{a}^{\text{gen}, +}$, so that this assumption is propagated throughout most of the analysis in this section. We summarize the results as follows. Recall the set $\mathcal{C}(G, H_0) \subset K$ defined in (22).

Lemma 5.4. *Let $H_0 \in \mathfrak{a}^{\text{gen}, +}$. Then $h_{H_0, g}$ has at most one critical point. A critical point $a \in A$ is characterized by the condition that $\kappa(ga) \in \mathcal{C}(G, H_0)$. The Hessian at a critical point is negative definite and (as a quadratic form) given by*

$$\begin{aligned} \mathfrak{a} &\rightarrow \mathbb{R} \\ H &\mapsto \langle [H_0, \text{Ad}_c(H)], E_{\mathfrak{n}}(\text{Ad}_c(H)) \rangle, \end{aligned}$$

where $c = \kappa(ga)$.

Proof. The uniqueness and the fact that the Hessian is negative definite are contained in [38, Theorem 1.2]. The characterization is [38, Lemma 4.1]. The Hessian is computed in [38, Proposition 4.16] when $g \in K$, but it does not depend on the triangular part of g by [38, equation (12)]. \square

We require some additional facts about the dependence of the critical point of $h_{H_0, g}$ on the parameters H_0 and g . Take $H_0 \in \mathfrak{a}^{\text{gen}, +}$. Define $\mathcal{R}_{H_0} \subset G$ to be the set of elements g for which the function $h_{H_0, g}$ has a critical point. Define also

$$(32) \quad \mathcal{R} = \bigcup_{H_0 \in \mathfrak{a}^{\text{gen}, +}} \{H_0\} \times \mathcal{R}_{H_0}.$$

When $g \in \mathcal{R}_{H_0}$, by Lemma 5.4 there is a unique critical point of $h_{H_0, g}$. Define the function

$$(33) \quad \xi_{H_0} : \mathcal{R}_{H_0} \rightarrow A$$

that sends g to the critical point of $h_{H_0, g}$, and define

$$\begin{aligned} \xi : \mathcal{R} &\rightarrow A \\ (H_0, g) &\mapsto \xi_{H_0}(g). \end{aligned}$$

Define also the \mathcal{C} -projection

$$\begin{aligned} c_{H_0} : \mathcal{R}_{H_0} &\rightarrow \mathcal{C}(G, H_0) \\ g &\mapsto \kappa(g \xi_{H_0}(g)) \end{aligned}$$

which is guaranteed to take values in $\mathcal{C}(G, H_0)$ by Lemma 5.4.

Lemma 5.5. *The set $\mathcal{R} \subset \mathfrak{a} \times G$ is open, and $\xi_{H_0}(g)$ and $c_{H_0}(g)$ are real analytic in $(H_0, g) \in \mathcal{R}$.*

Proof. Take $(H_0, g) \in \mathcal{R}$. By Lemma 5.4 the critical points of $h_{H_0, g}$ are nondegenerate, which we may reformulate by saying that the map

$$\begin{aligned} A &\rightarrow \mathfrak{a}^* \\ a &\mapsto (Dh_{H_0, g})_a \end{aligned}$$

has invertible differential at the level set above 0, which is the singleton $\{\xi_{H_0}(g)\}$. By the implicit function theorem applied to this real analytic map with parameter $(H_0, g) \in \mathfrak{a}^{\text{gen}, +} \times G$, it follows that \mathcal{R} is open in $\mathfrak{a}^{\text{gen}, +} \times G$, and that $\xi_{H_0}(g)$ is real analytic in (H_0, g) . Consequently, $c_{H_0}(g)$ is also real analytic in (H_0, g) . \square

The following coordinate system for $\mathcal{R}_{H_0} \cap K$ will be useful when dealing with expressions involving the ξ_{H_0} .

Lemma 5.6. *Let $H_0 \in \mathfrak{a}^{\text{gen}, +}$. The map*

$$\begin{aligned} \mathcal{C}(G, H_0) \times A &\rightarrow \mathcal{R}_{H_0} \cap K \\ (c, a) &\mapsto \kappa(ca^{-1}) \end{aligned}$$

is a real analytic isomorphism whose inverse is

$$k \mapsto (c_{H_0}(k), \xi_{H_0}(k)).$$

Proof. The two maps are real analytic by Lemma 5.5, and the fact that they are mutual inverses follows from the definitions of $\xi_{H_0}(k)$ and $c_{H_0}(k)$. \square

Example 5.7. When $G = \text{PSL}(2, \mathbb{R})$ with the standard choice of Iwasawa decomposition, Lemma 5.6 can be visualized as follows. Identify G with the unit tangent bundle of \mathbb{H} and K -projections with unit tangent vectors. We have $\mathcal{R}_{H_0} \cap K = K - M'$, which has two connected components, corresponding to the directions

pointing east or west. (The north and south directions are excluded as they come from M' .) Accordingly, the set $\mathcal{C}(G, H_0)$ has two points:

$$c_{\pm} = \begin{pmatrix} \cos(\pi/4) & \pm \sin(\pi/4) \\ \mp \sin(\pi/4) & \cos(\pi/4) \end{pmatrix}.$$

When a runs through A , the direction of c_+a runs through all directions pointing east, and the direction of c_-a runs through all directions pointing west.

5.3. Reduction to an integral over K . We seek to apply the stationary phase theorem to evaluate the inner integral in (29). The main result of this subsection is Proposition 5.10. Given the results of §5.2, for which most of the work was done in [38], the proof is parallel to that of [34, Lemma 7.10] when $G = \mathrm{PSL}_2(\mathbb{R})$. Some care must be taken to obtain uniformity in H_0 .

Define the ‘‘parameter space’’

$$\mathcal{P} = \mathfrak{a}^{\mathrm{gen},+} \times G \times K$$

and let $\mathcal{R}' \subset \mathcal{P}$ be the set of parameters (H_0, g, k) for which the phase function $\phi_{H_0, g}(\cdot, \cdot, k)$ has a critical point. Denote by $\mathcal{R}'_{H_0, g} \subset K$ the fiber of \mathcal{R}' above (H_0, g) .

Lemma 5.8. *The set \mathcal{R}' is open in \mathcal{P} . When $(H_0, g, k) \in \mathcal{R}'$, the function $\phi_{H_0, g}(\cdot, \cdot, k)$ has a unique critical point (a_1, a_2) given by $(\xi_{H_0}(k), \xi_{H_0}(kg))$.*

Proof. Recall from (31) that

$$\phi_{H_0, g}(a_1, a_2, k) = h_{H_0, kg}(a_2) - h_{H_0, k}(a_1),$$

so that (a_1, a_2) is a critical point of $\phi_{H_0, g}(\cdot, \cdot, k)$ if and only if a_1 is a critical point of $h_{H_0, k}$ and a_2 is a critical point of $h_{H_0, kg}$. Lemma 5.4 gives the uniqueness, and the last part is the definition of ξ_{H_0} (see (33)). We show that \mathcal{R}' is open. Let \mathcal{R} be as in (32); it is open in $\mathfrak{a} \times G$ by Lemma 5.5, and it suffices to note that \mathcal{R}' is the preimage of $\mathcal{R} \times \mathcal{R}$ under the continuous map

$$\begin{aligned} \mathcal{P} &\rightarrow (\mathfrak{a} \times G) \times (\mathfrak{a} \times G) \\ (H_0, g, k) &\mapsto ((H_0, k), (H_0, kg)). \end{aligned} \quad \square$$

Define a function on K by

$$(34) \quad \psi_{H_0, g}(k) = \begin{cases} \phi_{H_0, g}(\xi_{H_0}(k), \xi_{H_0}(kg), k) & \text{when } k \in \mathcal{R}'_{H_0, g}, \\ 0 & \text{otherwise.} \end{cases}$$

When $(H_0, g, k) \in \mathcal{R}'$, denote by $d_{H_0, g}(k)$ the Hessian determinant of the function $\phi_{H_0, g}(\cdot, \cdot, k)$ (for fixed k) at its unique critical point. Let b'_g be as in (29) and define

$$b''_{H_0, g}(k) = \begin{cases} b'_g(\xi_{H_0}(k), \xi_{H_0}(kg), k) \frac{(2\pi)^r}{\sqrt{|d_{H_0, g}(k)|}} & \text{when } k \in \mathcal{R}'_{H_0, g}, \\ 0 & \text{otherwise,} \end{cases}$$

and call $b'' : \mathcal{P} \rightarrow \mathbb{C}$ the corresponding function of (H_0, g, k) .

Lemma 5.9. *The function b'' is smooth and $\mathrm{supp}(b'') \subset \mathcal{R}'$.*

Proof. Roughly speaking, this is because $\xi_{H_0}(g)$ diverges as (H_0, g) approaches the boundary of \mathcal{R} . We must only show that every point of $\mathcal{P} - \mathcal{R}'$ has a neighborhood on which b'' is zero. Suppose not, then there exists a sequence $(H_{0, n}, g_n, k_n) \in \mathcal{R}'$ on which b'' is nonzero and which converges to a point (H_0, g, k) of $\mathcal{P} - \mathcal{R}'$. Because $b''(H_{0, n}, g_n, k_n) \neq 0$ and b'_g has support bounded independently of g (even

independent of g altogether) the sequences $\xi_{H_{0,n}}(k_n)$ and $\xi_{H_{0,n}}(k_n g_n)$ are bounded. We may then extract a subsequence on which $\xi_{H_{0,n}}(k_n)$ and $\xi_{H_{0,n}}(k_n g_n)$ converge, say to $\xi_1, \xi_2 \in A$. By continuity we then have

$$(Dh_{H_0,k})_{\xi_1} = \lim_{n \rightarrow \infty} (Dh_{H_{0,n},k_n})_{\xi_{H_{0,n}}(k_n)} = 0,$$

and similarly for ξ_2 , so that $(H_0, g, k) \in \mathcal{R}'$. This is a contradiction. \square

Proposition 5.10. *Let $D_{\mathfrak{a}} \subset \mathfrak{a}^{\text{gen},+}$ be compact and $D_G \subset G$ be compact. Define $J(H_0, g)$ by (25). Then*

$$J(tH_0, g) = t^{-r} \int_K e^{it\psi_{H_0,g}(k)} b''_{H_0,g}(k) dk + O(t^{-r-1})$$

as $t \rightarrow +\infty$, uniformly for $H_0 \in D_{\mathfrak{a}}$ and $g \in D_G$.

Proof. We prove a uniform asymptotic for the double A -integral in (29), which we then integrate over K . Let $\mathcal{P}_0 = \text{supp}(b'') \cap (D_{\mathfrak{a}} \times D_G \times K)$. It is closed in \mathcal{P} and contained in \mathcal{R}' by Lemma 5.9. We distinguish two cases depending on where (H_0, g, k) lies. When $(H_0, g, k) \in \mathcal{R}'$, the phase function $\phi_{H_0,g}(\cdot, \cdot, k)$ has a unique and nondegenerate critical point $(\xi_{H_0}(k), \xi_{H_0}(kg))$, where the Hessian has signature (r, r) by (31) and Lemma 5.4. The stationary phase theorem [50, §VIII.2] implies that

$$(35) \quad \int_{A \times A} e^{it\phi_{H_0,g}(a_1, a_2, k)} b'_{H_0,g}(a_1, a_2, k) da_1 da_2 = t^{-r} e^{it\psi_{H_0,g}(k)} b''_{H_0,g}(k) + O(t^{-r-1})$$

as $t \rightarrow +\infty$, uniformly for (H_0, g, k) in compact subsets of \mathcal{R}' . When $(H_0, g, k) \in \mathcal{P} - \mathcal{P}_0$, the phase function $\phi_{H_0,g}(\cdot, \cdot, k)$ has no critical points in the support of $b'_{H_0,g}$, and the Van der Corput lemma [50, §VIII.2] implies that

$$\int_{A \times A} e^{it\phi_{H_0,g}(a_1, a_2, k)} b'_{H_0,g}(a_1, a_2, k) da_1 da_2 \ll_N t^{-N}$$

as $t \rightarrow +\infty$, uniformly for (H_0, g, k) in compact subsets of $\mathcal{P} - \mathcal{P}_0$. Because b'' is by definition zero on $\mathcal{P} - \mathcal{P}_0$, the estimate (35) also holds in this case, uniformly in compact subsets of $\mathcal{P} - \mathcal{P}_0$. Because $\mathcal{P} - \mathcal{P}_0$ and \mathcal{R}' are open in \mathcal{P} (Lemma 5.8) and they cover \mathcal{P} , we may find compact subsets of \mathcal{R}' and $\mathcal{P} - \mathcal{P}_0$ that cover the compact set $\mathcal{P} \cap (D_{\mathfrak{a}} \times D_G \times K)$. Therefore (35) holds uniformly for all $(H_0, g, k) \in \mathcal{P} \cap (D_{\mathfrak{a}} \times D_G \times K)$. The statement follows then by integrating (35) over K . \square

5.4. Critical points of $\psi_{H_0,g}$. This subsection is concerned with the critical points of $\psi_{H_0,g}$ (defined in (34)) when g lies in the dense open set $G - \bigcup_{L \in \mathcal{L} - \{G\}} M'L$. The main result is Proposition 5.11, which will imply Proposition 5.1.

Proposition 5.11. *Let $H_0 \in \mathfrak{a}^{\text{gen},+}$ and $g \in G - \bigcup_{L \in \mathcal{L} - \{G\}} M'L$. Then $\psi_{H_0,g}$ is nowhere locally constant in $\mathcal{R}'_{H_0,g}$.*

The proof of Proposition 5.11 is at the end of this subsection.

Remark 5.12. The proof differs from the proof in [34] for $\text{PSL}_2(\mathbb{R})$. Namely, we do not prove that the Hessian of $\psi_{H_0,g}$ is nondegenerate at critical points. In fact, to obtain a useful expression for the Hessian of $\psi_{H_0,g}$ in order to generalize to proof for $\text{PSL}_2(\mathbb{R})$, we must know something about the derivatives of the ξ_{H_0} (see (33)), which are determined implicitly by their image under an injective linear map $\mathfrak{a} \rightarrow \mathfrak{k}$

involving the Hessian of $h_{H_0, g}$. When $\dim(\mathfrak{a}) = \dim(\mathfrak{k}) = 1$, all of this can be made relatively explicit, but in general it seems very hard to say anything about the $D\xi_{H_0}$. As a replacement for this computation, we will use the change of variables from Lemma 5.6, through which one of the occurrences of ξ_{H_0} simplifies greatly. The second step in the proof is to eliminate the other occurrence of $\xi_{H_0}(g) \in A$, by using the fact that its adjoint action on \mathfrak{a} is trivial; see Lemma 5.17.

We first deduce Proposition 5.1 using very general principles.

Proof of Proposition 5.1 assuming Proposition 5.11. By (26) we may assume that $H_0 \in \mathfrak{a}^+$. Proposition 5.10 then reduces the statement to showing that for some $N > 0$ and $\delta > 0$,

$$(36) \quad \int_K e^{it\psi_{H_0, g}(k)} b''_{H_0, g}(k) dk \ll \left(1 + t \cdot d \left(g, \bigcup_{L \in \mathcal{L} - \{G\}} M'L \right)^N \right)^{-\delta},$$

with uniformity in H_0 . When g lies at positive distance from $\bigcup_{L \in \mathcal{L} - \{G\}} M'L$, this follows directly from the Van der Corput lemma [50, §VIII.2, Proposition 5], because by Proposition 5.11 there is no $k \in \text{supp}(b''_{H_0, g})$ at which $\psi_{H_0, g}$ is somewhere constant, meaning that some higher k -derivative is bounded away from zero around every $k \in \text{supp}(b''_{H_0, g})$.

To get control over the dependence on g , consider for every $n \geq 1$ the real analytic map

$$\begin{aligned} D_n : \mathcal{R}' &\rightarrow (\mathfrak{k}^*)^{\otimes n} \\ (H_0, g, k) &\mapsto ((D^{(n)}\psi_{H_0, g})_k \end{aligned}$$

where we denote $D^{(n)}$ for the higher derivatives. Call $Z \subset \mathcal{R}'$ the joint zero locus of the D_n .

Proposition 5.11 says that

$$Z \subset \mathfrak{a} \times \left(\bigcup_{L \in \mathcal{L} - \{G\}} M'L \right) \times K.$$

Locally around a fixed (H_0, g, k) we may embed \mathcal{R}' as a real submanifold of a complex manifold Ω on which $\psi_{H_0, g}(k)$ is complex analytic in all three variables. The maps D_n are derivatives of $\psi_{H_0, g}(k)$ and therefore also extend analytically to Ω . By [55, §3.9 Theorem 9 C], their joint zero locus is locally the zero locus of only finitely many D_n , say of D_1, \dots, D_J , on a neighborhood $U \subset \mathcal{R}'$ of (H_0, g, k) . Consider now the Taylor polynomial map

$$T_J : U \rightarrow (\mathfrak{k}^*)^{\otimes 1} \oplus \dots \oplus (\mathfrak{k}^*)^{\otimes J}$$

obtained by pairing D_1, \dots, D_J , and equip the right-hand side with a fixed norm. By Lojasiewicz's inequality, for every compact subset $V \subset U$ there exists $N > 0$ such that for $(H'_0, g', k') \in V$ we have

$$\|T_J(H'_0, g', k')\| \gg d((H_0, g, k), Z)^N \gg d \left(g, \bigcup_{L \in \mathcal{L} - \{G\}} M'L \right)^N.$$

Because such a bound holds locally in \mathcal{R}' , it holds on compact subsets of \mathcal{R}' but with possibly bigger values of J and N . This gives a lower bound for a k -derivative order at most J for $\psi_{H_0, h}$. To apply this, take $D \subset \mathcal{R}'$ to be any compact neighborhood of $\text{supp}(b'') \cap (D_{\mathfrak{a}} \times D_G \times K)$. This is possible by Lemma 5.9 and because \mathcal{R}' is open in \mathcal{P} . The bound (36) then follows from the Van der Corput lemma [50, §VIII.2, Proposition 5], with the corresponding value of N and with $\delta = 1/J$. \square

The following lemma is an immediate application of the chain rule.

Lemma 5.13. *Let $H_0 \in \mathfrak{a}^{\text{gen},+}$ and $g \in G$. When $k \in \mathcal{R}'_{H_0, g}$, we have that k is a critical point of $\psi_{H_0, g}$ if and only if $(\xi_{H_0}(k), \xi_{H_0}(kg), k)$ is a critical point of $\phi_{H_0, g}$. \square*

For $g \in G$, define $\Theta_g : K \rightarrow K$ by (28).

Lemma 5.14. *Let $H_0 \in \mathfrak{a}^{\text{reg}}$ and $g \in G$. A point (a_1, a_2, k) is a critical point of $\phi_{H_0, g}$ if and only if*

$$\begin{aligned} (37) \quad & \kappa(ka_1) \in \mathcal{C}(G, H_0), \\ (38) \quad & \kappa(kga_2) \in \mathcal{C}(G, H_0), \\ (39) \quad & n(\Theta_{a_1}(k)a_1^{-1}ga_2) = 1. \end{aligned}$$

Proof. Let (a_1, a_2, k) be a critical point of $\phi_{H_0, g}$. By (31) and Lemma 5.4, criticality in a_1 and a_2 is equivalent to (37) and (38). In view of (30), k is a critical point of $\phi_{H_0, g}(a_1, a_2, \cdot)$ if and only if $\Theta_{a_1}(k)$ is a critical point of $\tilde{\phi}_{H_0, g}(a_1, a_2, \cdot)$. Using the expression (27), by [15, Lemma 5.3] this in turn is equivalent to

$$\Theta_{a_1}(k)a_1^{-1}ga_2 \in Z_N(H_0)AK = AK,$$

where we have used that $H_0 \in \mathfrak{a}^{\text{reg}}$. This is (39). \square

Remark 5.15. The condition (39) can also be written as $n(ka_1) = n(kga_2)$. Indeed, writing $ka_1 = n'a'k'$, condition (39) says that

$$k'a_1^{-1}ga_2 = a'^{-1}n'^{-1}kga_2 \in AK.$$

We may write this as $kga_2 \in n'AK$, which is what we claimed. This is analogous to the criticality condition in [34, Lemma 7.1], where it can be interpreted geometrically as saying that two geodesics in the upper half plane are concentric.

At this point it is helpful to make the change of variables given by the diffeomorphism in Proposition 5.6. Define therefore

$$(40) \quad \begin{aligned} \tilde{\psi}_{H_0, g} : \mathcal{C}(G, H_0) \times A &\rightarrow \mathbb{R} \\ (c, a) &\mapsto \psi_{H_0, g}(\kappa(ca^{-1})). \end{aligned}$$

For $H_0 \in \mathfrak{a}$ and $g \in G$, define an open subset of $\mathcal{C}(G, H_0) \times A$ by

$$\mathcal{R}''_{H_0, g} = \{(c, a) \in \mathcal{C}(G, H_0) \times A : (H_0, g, \kappa(ca^{-1})) \in \mathcal{R}'\}.$$

Thanks to this change of variables, $\tilde{\psi}_{H_0, g}$ has a more manageable expression. Using that $\xi_{H_0}(\kappa(ca^{-1})) = a$ and using the definitions (40), (34), (30), (27) we have for

$(c, a) \in \mathcal{R}''_{H_0, g}$ that

$$\begin{aligned} \tilde{\psi}_{H_0, g}(c, a) &= \psi_{H_0, g}(\kappa(ca^{-1})) \\ &= \phi_{H_0, g}(a, \xi_{H_0}(\kappa(ca^{-1})g), \kappa(ca^{-1})) \\ &= \tilde{\phi}_{H_0, g}(a, \xi_{H_0}(\kappa(ca^{-1})g), c) \\ &= \langle H_0, H(ca^{-1}g\xi_{H_0}(\kappa(ca^{-1})g)) \rangle \\ &= \langle H_0, H(ca^{-1}g\xi_{H_0}(ca^{-1}g)) \rangle. \end{aligned}$$

In the last equality we have used that ξ_{H_0} is invariant under left multiplication by NA . For $(c, a) \in \mathcal{R}''_{H_0, g}$ define

$$(41) \quad \gamma_{H_0, g}(c, a) = ca^{-1}g\xi_{H_0}(ca^{-1}g),$$

so that $\tilde{\psi}_{H_0, g}(c, a) = \langle H_0, H(\gamma_{H_0, g}(c, a)) \rangle$. By Lemma 5.4 and the definition of ξ_{H_0} (33) we have for all $(c, a) \in \mathcal{R}''_{H_0, g}$ that

$$(42) \quad \kappa(\gamma_{H_0, g}(c, a)) \in \mathcal{C}(G, H_0).$$

Lemma 5.16. *Let $H_0 \in \mathfrak{a}^{\text{gen}, +}$ and $g \in G$. Then $(c, a) \in \mathcal{R}''_{H_0, g}$ is a critical point of $\tilde{\psi}_{H_0, g}$ if and only if*

$$n(\gamma_{H_0, g}(c, a)) = 1.$$

Proof. A point $(c, a) \in \mathcal{R}''_{H_0, g}$ is a critical point of $\tilde{\psi}_{H_0, g}$ if and only if $\kappa(ca^{-1})$ is a critical point of $\psi_{H_0, g}$. Using Lemma 5.13 and Lemma 5.14, this is seen to be equivalent to $n(\gamma_{H_0, g}(c, a)) = 1$. \square

Lemma 5.17. *If $\tilde{\psi}_{H_0, g}$ is critical at $(c, a) \in \mathcal{R}''_{H_0, g}$, then*

$$\langle H_0, \text{Ad}_{ca^{-1}g}(\mathfrak{a}) \rangle = 0.$$

Proof. By Lemma 5.16 and equation (42), criticality implies that $\gamma_{H_0, g}(c, a) \in AC(G, H_0)$, so that

$$ca^{-1}g \in AC(G, H_0)A.$$

For $k \in \mathcal{C}(G, H_0)$ we have by definition that $\langle H_0, \text{Ad}_k(\mathfrak{a}) \rangle = 0$ (see (22)). By Ad_G -invariance of the Killing form, this gives

$$\langle H_0, \text{Ad}_{ca^{-1}g}(\mathfrak{a}) \rangle = 0,$$

as claimed. \square

Remark 5.18. (1) In Lemma 5.17 we write $\langle H_0, \cdot \rangle$ and not $H_0 \perp \cdot$, because $\text{Ad}_{ca^{-1}g}(\mathfrak{a})$ is not in \mathfrak{p} in general.

(2) Lemma 5.17 gets rid of the (complicated) function $\xi_{H_0, g}$ by considering the adjoint action on \mathfrak{a} . This makes differentiating much easier, at the cost of a (possible) loss of information about $\gamma_{H_0, g}(c, a)$.

Lemma 5.19. *If $\tilde{\psi}_{H_0, g}$ is constant around a point $(c, a) \in \mathcal{R}''_{H_0, g}$, then*

$$[H_0, \text{Ad}_c(T_c\mathcal{C}(G, H_0))] \perp E_{\mathfrak{p}}(\text{Ad}_{cAg}(\mathfrak{a})),$$

where $E_{\mathfrak{p}} = \frac{1}{2}(\text{id} - \theta)$ denotes the orthogonal projection onto \mathfrak{p} .

Proof. By the locally constant hypothesis implies, the conclusion of Lemma 5.17 is true for all nearby c and nearby a . In fact, by analytic continuation it is valid for all $a' \in A$. Differentiating it with respect to c gives

$$\langle H_0, [\text{Ad}_c(T_c\mathcal{C}(G, H_0)), \text{Ad}_{ca'g}(\mathfrak{a})] \rangle = 0,$$

for all $a' \in A$. Now we use associativity of the Killing form:

$$\langle [H_0, \text{Ad}_c(T_c\mathcal{C}(G, H_0))], \text{Ad}_{ca'g}(\mathfrak{a}) \rangle = 0.$$

And finally, we use that the left member lies in $[\mathfrak{p}, \mathfrak{k}] = \mathfrak{p}$. \square

Lemma 5.20. *For $H_0 \in \mathfrak{a}^{\text{gen}}$ we have that*

$$\mathfrak{m} + \text{Ad}_c(\mathfrak{m}) \subset \text{Ad}_c(T_c\mathcal{C}(G, H_0)).$$

Proof. This is easily seen from Lemma 4.6. Associativity of the Killing form gives that $\mathfrak{m} \perp [H_0, \mathfrak{g}]$, which implies that $\mathfrak{m} \subset \text{Ad}_c(T_c\mathcal{C}(G, H_0))$. Associativity also gives that $\text{Ad}_c(\mathfrak{m}) \perp [\text{Ad}_c(\mathfrak{a}), \mathfrak{g}]$, which implies $\text{Ad}_c(\mathfrak{m}) \subset \text{Ad}_c(T_c\mathcal{C}(G, H_0))$. \square

Define $\mathfrak{k}^{\perp\mathfrak{m}} = \mathfrak{k} \cap \bigoplus_{\alpha} \mathfrak{g}_{\alpha}$ and $\mathfrak{p}^{\perp\mathfrak{a}} = \mathfrak{p} \cap \bigoplus_{\alpha} \mathfrak{g}_{\alpha}$. They are the ‘‘root space parts’’ of \mathfrak{k} and \mathfrak{p} respectively. Define $E_{\mathfrak{k}^{\perp\mathfrak{m}}}$ and $E_{\mathfrak{p}^{\perp\mathfrak{a}}}$ to be the orthogonal projections onto these spaces.

The following is nothing but a useful reformulation of Lemma 4.6.

Lemma 5.21. *Let $H_0 \in \mathfrak{a}^{\text{gen}}$ and $c \in \mathcal{C}(G, H_0)$. The orthogonal complement of the subspace $[H_0, \text{Ad}_c(T_c\mathcal{C}(G, H_0))]$ in \mathfrak{p} equals $\mathfrak{a} \oplus \text{Ad}_c(\mathfrak{a})$.*

Proof. Call V the orthogonal complement of $[H_0, \text{Ad}_c(T_c\mathcal{C}(G, H_0))]$ in \mathfrak{p} . Certainly $\mathfrak{a} \subset V$, because $\mathfrak{a} \perp [\mathfrak{a}, \mathfrak{k}]$. By Lemma 4.6, the space

$$\text{Ad}_c(T_c\mathcal{C}(G, H_0)) \subset \mathfrak{k}$$

is the orthogonal complement of $[H_0, \text{Ad}_c(\mathfrak{a})]$. It follows that $\text{Ad}_c(\mathfrak{a}) \subset V$. Therefore

$$\mathfrak{a} \oplus \text{Ad}_c(\mathfrak{a}) \subset V.$$

That this is an equality, follows from dimension considerations. Consider the projection $E_{\mathfrak{k}^{\perp\mathfrak{m}}}(\text{Ad}_c(T_c\mathcal{C}(G, H_0)))$ to $\mathfrak{k}^{\perp\mathfrak{m}}$. Because $\text{Ad}_c(T_c\mathcal{C}(G, H_0))$ has codimension $\dim(\mathfrak{a})$ in \mathfrak{k} by Lemma 4.5, and because $\text{Ad}_c(T_c\mathcal{C}(G, H_0)) \supset \mathfrak{m}$ by Lemma 5.20, this projection has codimension $\dim(\mathfrak{a})$ in $\mathfrak{k}^{\perp\mathfrak{m}}$. Now the action of $[H_0, \cdot]$ on $\mathfrak{k}^{\perp\mathfrak{m}}$ gives a bijection with $\mathfrak{p}^{\perp\mathfrak{a}}$ because H_0 is regular. Therefore $[H_0, \text{Ad}_c(T_c\mathcal{C}(G, H_0))]$ has codimension $\dim(\mathfrak{a})$ in $\mathfrak{p}^{\perp\mathfrak{a}}$, and the orthogonal complement in this space must be $E_{\mathfrak{p}^{\perp\mathfrak{a}}}(\text{Ad}_c(\mathfrak{a}))$, which has the correct dimension because $\text{Ad}_c(\mathfrak{a})$ intersects trivially with \mathfrak{a} by Lemma 4.5 and Lemma 2.1. \square

The idea of the argument is the following. Lemma 5.21 gives the precise orthogonal complement of $[H_0, \text{Ad}_c(T_c\mathcal{C}(G, H_0))]$ in \mathfrak{p} , and Lemma 5.19 also gives elements in \mathfrak{p} orthogonal to it, under the assumption that $\tilde{\psi}_{H_0, g}$ is constant around a point $(c, a) \in \mathcal{R}''_{H_0, g}$. As soon as the latter lemma gives one new element, we have a contradiction. We may also formulate this in the following way.

Lemma 5.22. *If $\tilde{\psi}_{H_0, g}$ is constant around a point $(c, a) \in \mathcal{R}''_{H_0, g}$, then*

$$E_{\mathfrak{p}}(\text{Ad}_{cAg}(\mathfrak{a})) \subset \mathfrak{a} \oplus \text{Ad}_c(\mathfrak{a}).$$

Proof. Immediate from Lemma 5.19 and Lemma 5.21. \square

Lemma 5.23. *If $\tilde{\psi}_{H_0, g}$ is constant around a point $(c, a) \in \mathcal{R}''_{H_0, g}$, then*

$$E_{\mathfrak{p}^\perp \mathfrak{a}}(\mathrm{Ad}_{Ag}(\mathfrak{a})) \subset E_{\mathfrak{p}^\perp \mathfrak{a}}(\mathrm{Ad}_{c^{-1}}(\mathfrak{a})).$$

Proof. We use Lemma 5.22 and apply $\mathrm{Ad}_{c^{-1}}$, which gives that

$$E_{\mathfrak{p}}(\mathrm{Ad}_{Ag}(\mathfrak{a})) \subset \mathrm{Ad}_{c^{-1}}(\mathfrak{a}) \oplus \mathfrak{a}.$$

Here we have used that Ad_K commutes with $E_{\mathfrak{p}}$. Projecting further to $\mathfrak{p}^\perp \mathfrak{a}$ we may discard the factor \mathfrak{a} in the right-hand side, and the claim follows. \square

If we can show that the left-hand side in Lemma 5.23 is of dimension $\dim(\mathfrak{a})$, then the inclusion becomes an equality; in particular, an inclusion in the other direction. The following lemmas show that it is indeed of that dimension, provided that g is generic.

Lemma 5.24. *Let $X = (X_\alpha) \in \bigoplus_{\alpha \in \Sigma} \mathfrak{g}_\alpha$. The span of $\mathrm{Ad}_A(X)$ is $\bigoplus_{\alpha \in \Sigma} \mathbb{R}X_\alpha$.*

Proof. Clearly $\mathrm{Ad}_A(X) \subset \bigoplus_{\alpha \in \Sigma} \mathbb{R}X_\alpha$.

We must show that the vectors $v_a = (\alpha(a))_{\alpha \in \Sigma} \in \mathbb{R}^\Sigma$ with $a \in A$ span the entire space \mathbb{R}^Σ . This is equivalent to the statement that the roots of A are linearly independent, and it is a general fact that the characters of an abelian group are linearly independent. More precisely, suppose they don't span everything, then they lie on a hyperplane and there exist $c_\alpha \in \mathbb{R}$, not all zero such that

$$\forall a \in A : \sum_{\alpha \in \Sigma} c_\alpha \alpha(a) = 0.$$

This is saying that the $\alpha : A \rightarrow \mathbb{R}^\times$ are linearly dependent. \square

Lemma 5.25. *If $g \notin \bigcup_{L \in \mathcal{L} - \{G\}} M'L$, then*

$$\dim(\mathrm{span}(E_{\mathfrak{p}^\perp \mathfrak{a}}(\mathrm{Ad}_{Ag}(\mathfrak{a})))) \geq \dim(\mathfrak{a}).$$

Proof. Let V be the span of $\mathrm{Ad}_{Ag}(\mathfrak{a})$. By Lemma 5.24, it is of the form $V_0 \oplus \bigoplus_{\alpha \in \Sigma} V_\alpha$ with $V_0 \subset \mathfrak{a} + \mathfrak{m}$ and $V_\alpha \subset \mathfrak{g}_\alpha$. Let $S = \{\alpha : V_\alpha \neq 0\}$. We claim that S spans \mathfrak{a}^* . Suppose not, then there exists a nonzero $H \in \bigcap_{\alpha \in S} \ker(\alpha)$, meaning that $[H, V] = 0$. Then $[H, \mathrm{Ad}_g(\mathfrak{a})] = 0$, or equivalently, $[\mathrm{Ad}_{g^{-1}}(H), \mathfrak{a}] = 0$. Thus $\mathrm{Ad}_{g^{-1}}(H) \in \mathfrak{m} \oplus \mathfrak{a}$. By Lemma 2.1 this implies that $g \in Z_G(H)M'$, contradicting our hypothesis on g . Therefore S spans \mathfrak{a}^* .

In particular, there are at least $\dim(\mathfrak{a})$ positive roots $\alpha \in \Sigma^+$ with $V_\alpha \neq 0$, and in particular with $E_{\mathfrak{p}^\perp \mathfrak{a}}(V_\alpha) \neq 0$. Because the projections $E_{\mathfrak{p}^\perp \mathfrak{a}}(\mathfrak{g}_\alpha)$ for different $\alpha \in \Sigma^+$ are orthogonal, the projection $E_{\mathfrak{p}^\perp \mathfrak{a}}(V)$ has dimension at least $\dim(\mathfrak{a})$. \square

Lemma 5.26. *Suppose $\tilde{\psi}_{H_0, g}$ is constant around a point $(c, a) \in \mathcal{R}''_{H_0, g}$, and suppose $g \notin \bigcup_{L \in \mathcal{L} - \{G\}} M'L$. Then*

$$\mathrm{span}(E_{\mathfrak{p}^\perp \mathfrak{a}}(\mathrm{Ad}_{Ag}(\mathfrak{a}))) = E_{\mathfrak{p}^\perp \mathfrak{a}}(\mathrm{Ad}_{c^{-1}}(\mathfrak{a})).$$

Proof. Lemma 5.23 says that the left-hand side is contained in the right-hand side. The left-hand side is of dimension at least $\dim(\mathfrak{a})$ by Lemma 5.25, and the right-hand side is of dimension $\dim(\mathfrak{a})$ because $\mathrm{Ad}_c(\mathfrak{a})$ intersects trivially with \mathfrak{a} by Lemma 4.5 and Lemma 2.1. Therefore, equality must hold. \square

Proof of Proposition 5.11. Assume $\psi_{H_0, g}$ is constant around some point $k \in \mathcal{R}'_{H_0, g}$. Then $\tilde{\psi}_{H_0, g}$ (defined in (40)) is constant around some point $(c, a) \in \mathcal{R}''_{H_0, g}$.

Because $\text{Ad}_{c^{-1}}(H_0) \perp \mathfrak{a}$ (by definition of $\mathcal{C}(G, H_0)$), Lemma 5.26 gives that

$$\text{Ad}_{c^{-1}}(H_0) \in \text{span}(E_{\mathfrak{p} \perp \mathfrak{a}}(\text{Ad}_{Ag}(\mathfrak{a}))).$$

On the other hand, by the hypothesis of being locally constant and by analytic continuation in a , Lemma 5.17 says that

$$\langle \text{Ad}_{c^{-1}}(H_0), \text{Ad}_{Ag}(\mathfrak{a}) \rangle = 0,$$

so that

$$\text{Ad}_{c^{-1}}(H_0) \perp \text{span}(E_{\mathfrak{p} \perp \mathfrak{a}}(\text{Ad}_{Ag}(\mathfrak{a}))).$$

Therefore $\text{Ad}_{c^{-1}}(H_0) = 0$, which is a contradiction. \square

Remark 5.27. When $g \in M'L$ with $L \in \mathcal{L}$ chosen to be minimal, an argument analogous to that used to prove Lemma 5.25 shows that

$$\dim(\text{span}(E_{\mathfrak{p} \perp \mathfrak{a}}(\text{Ad}_{Ag}(\mathfrak{a})))) \geq \dim(\mathfrak{a}^L).$$

Moreover, the span in the left-hand side is a direct sum of subspaces of the spaces $\mathfrak{p} \cap (\mathfrak{g}_\alpha + \mathfrak{g}_{-\alpha})$. If $\tilde{\psi}_{H_0, g}$ is constant around a point $(c, a) \in \mathcal{R}''_{H_0, g}$, this gives a very strong restriction on the space $E_{\mathfrak{p} \perp \mathfrak{a}}(\text{Ad}_{c^{-1}}(\mathfrak{a}))$ (which contains this span), but we have not been able to obtain a contradiction from this, even when exploiting the fact that the inclusion is true for all nearby $c' \in \mathcal{C}(G, H_0)$.

6. PRELIMINARIES ON ALGEBRAIC GROUPS

6.1. Algebraic groups. Let \mathbf{G}/\mathbb{Q} be a linear connected anisotropic algebraic group which is almost simple over \mathbb{R} . Let $\mathbf{H} \subset \mathbf{G}$ be a maximal torus. The base fields of groups are indicated by subscripts if they are not clear from the context. We use the same subscript notation for base change and when E/F is a finite field extension we denote $\text{Res}_{E/F}$ for Weil restriction.

Fix a closed embedding $\rho : \mathbf{G} \rightarrow \mathbf{SL}_d$ for some $d > 0$. Equipping \mathbf{SL}_d with the standard schematic structure and taking the schematic closures of $\rho(\mathbf{G})$ and $\rho(\mathbf{H})$ we obtain integral models that we continue to denote by \mathbf{G} and \mathbf{H} .

Remark 6.1. Eventually \mathbf{G} and \mathbf{H} will be as in Theorem 1.2. That is, \mathbf{G} an anisotropic form of \mathbf{PGL}_3 and \mathbf{H} maximal split over \mathbb{R} . But we allow some generality in order to be able to make remarks about other groups, and because for most results there is no reason to be restrictive. The condition that \mathbf{G} is almost simple over \mathbb{R} will only be used to apply a Weyl law-type result from [10] (see Proposition 7.7) but is of course automatic for forms of $\mathbf{PGL}_{3, \overline{\mathbb{Q}}}$.

6.2. Forms of \mathbf{PGL}_3 . By a form of a group \mathbf{G} over a field E we shall mean a form of the base change to the algebraic closure, $\mathbf{G}_{\overline{E}}$. The following proposition gives a concrete list of all groups \mathbf{G} that are allowed in Theorem 1.2. Proposition 6.3 below says what the corresponding Lie groups are.

Proposition 6.2. *The \mathbb{Q} -forms \mathbf{G} of \mathbf{PGL}_3 , and the anisotropic ones among them, are given by the following constructions.*

- (1) (Inner forms) Let D/\mathbb{Q} be a central simple algebra of degree 3 and define $\mathbf{G} = \mathbf{GL}_{1, D}/\mathbb{G}_m$. Then \mathbf{G} is anisotropic if and only if D is a division algebra.

- (2) (Outer forms) Let F/\mathbb{Q} be a quadratic field, (V, q) a 3-dimensional non-degenerate Hermitian space over F and define $\mathbf{G} = \mathbf{PU}(V)$. Then \mathbf{G} is anisotropic if and only if (V, q) is.
- (3) (Outer forms) Let F/\mathbb{Q} be a quadratic field, D/F be a central division algebra of degree 3 over F with an anti-involution $\tau \in \text{End}_{\mathbb{Q}}(D)$ that acts nontrivially on F , (V, q) a 1-dimensional Hermitian space over D with $q \neq 0$, and define $\mathbf{G} = \mathbf{PU}(V)$. Concretely, there is a nonzero $d \in D$ fixed by τ such that $\mathbf{G}(\mathbb{Q}) = \{g \in D^\times : gd\tau(g) = d\}/F^1$, where $F^1 \subset F^\times$ denotes the norm 1 subgroup. Then \mathbf{G} is anisotropic.

Proof. That this list is exhaustive can be proven as in [42, §18.4], or can be deduced from [42, Theorem 18.4.1] by using that $\mathbf{PGL}_{n, \overline{\mathbb{Q}}}$ and $\mathbf{SL}_{n, \overline{\mathbb{Q}}}$ have the same automorphism groups. The statements about anisotropy can be seen as follows:

- (1) By Wedderburn's theorem, D is either a division algebra or isomorphic to $M_3(\mathbb{Q})$. When D is a matrix algebra, clearly \mathbf{G} is split. Conversely, let D be a division algebra, $\mathbf{T} \subset \mathbf{G}$ be a maximal torus and χ a character of \mathbf{T} . We may lift \mathbf{T} to a maximal torus $\mathbf{T}' \subset \mathbf{GL}_{1, D}$ [9, Proposition 10.6]. It is well known that such tori are obtained from the multiplicative groups of étale subalgebras of D . Because D is division, $\mathbf{T}' = \text{Res}_{F/\mathbb{Q}} \mathbb{G}_{m, F}$ for a field extension F/\mathbb{Q} , and it follows that χ is a power of the norm character. Because χ is trivial on $\mathbb{G}_{m, \mathbb{Q}}$, it is trivial altogether. Therefore \mathbf{T} is anisotropic.
- (2) If q represents 0, then V contains a hyperbolic plane H , and we find a rank 1 split torus in $\mathbf{U}(V)$, necessarily noncentral, that acts diagonally on H in an isotropic basis and acts trivially on H^\perp . Conversely, if $\mathbf{T} \subset \mathbf{G}$ is a split torus then we may lift it to a torus of $\mathbf{U}(V)$ by [9, Proposition 10.6], which contains a nontrivial split subtorus \mathbf{T}' by [7, §II.8.14]. A vector $v \in V$ in a weight space with respect to \mathbf{T}' for a nonzero weight then satisfies $q(v) = 0$. (Such a vector exists because the split torus $\mathbf{T}' \subset \text{Res}_{F/\mathbb{Q}} \mathbf{GL}(V) \subset \mathbf{GL}_6$ is diagonalizable.)
- (3) This can be proven as in the second case: A split torus in \mathbf{G} would give rise to a nonzero vector $v \in V \cong D$ satisfying $vd\tau(v) = 0$. Because D is a division algebra, this is impossible. \square

Proposition 6.3. *Let \mathbf{G} be as in Proposition 6.2. With the same numbering, its group of real points is as follows:*

- (1) In the first case, $\mathbf{G}(\mathbb{R}) \cong \mathbf{PGL}_3(\mathbb{R})$.
- (2) In the second case: If F is real quadratic we have $\mathbf{G}(\mathbb{R}) \cong \mathbf{PGL}_3(\mathbb{R})$. If F is imaginary quadratic we have $\mathbf{G}(\mathbb{R}) \cong \mathbf{PU}(3)$ or $\mathbf{PU}(2, 1)$ depending on the signature of (V, q) over \mathbb{R} .
- (3) In the third case, $\mathbf{G}(\mathbb{R}) \cong \mathbf{PGL}_3(\mathbb{R})$ when F is real quadratic and $\mathbf{G}(\mathbb{R}) \cong \mathbf{PU}(3)$ or $\mathbf{PU}(2, 1)$ otherwise.

Proof. (1) By a classical theorem of Frobenius there is no degree 3 division algebra over \mathbb{R} , so that $D \otimes_{\mathbb{Q}} \mathbb{R} \cong M_3(\mathbb{R})$ by Wedderburn's theorem. Therefore $\mathbf{G}(\mathbb{R}) \cong \mathbf{PGL}_3(\mathbb{R})$.

- (2) When F is real quadratic we have $F \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R} \times \mathbb{R}$, and the induced automorphism permutes the two factors. Using this isomorphism we have $\mathbf{GL}(V)_{\mathbb{R}} \cong \mathbf{GL}_3(\mathbb{R}) \times \mathbf{GL}_3(\mathbb{R})$. The hermitian matrix defining q takes the form (A, A^T) under this isomorphism, and $A \in \mathbf{GL}_3(\mathbb{R})$ can be assumed

diagonal. Now $\mathbf{U}(V)(\mathbb{R}) \cong \{(B, C) \in \mathrm{GL}_3(\mathbb{R})^2 : BAC^T = CA^T B^T = A\}$. It is clear that for every $B \in \mathrm{GL}_3(\mathbb{R})$ there is a unique element $(B, C) \in \mathbf{U}(V)(\mathbb{R})$, and this gives an isomorphism $\mathbf{U}(V)(\mathbb{R}) \cong \mathrm{GL}_3(\mathbb{R})$. The conclusion follows.

When F is imaginary we have $F \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{C}$ and it follows that $\mathbf{U}(V)(\mathbb{R})$ is a classical unitary group, necessarily of signature $(3, 0)$ or $(2, 1)$.

- (3) This can be shown using similar arguments as in the first and second case. When F is real quadratic, using the isomorphism $F \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R} \times \mathbb{R}$ we have $D \otimes_{\mathbb{Q}} \mathbb{R} \cong M_3(\mathbb{R})^2$. Call σ the anti-involution of $M_3(\mathbb{R})^2$ that swaps the two factors and transposes them. Then $\tau \circ \sigma$ is an automorphism of $M_3(\mathbb{R})^2$. (In fact, it preserves the center $\mathbb{R} \times \mathbb{R}$, therefore preserves both factors and is inner by Skolem-Noether.) Applying an automorphism of $M_3(\mathbb{R})^3$ we may assume that $\sigma = \tau$. A similar argument as in the second case now gives an isomorphism $\mathbf{U}(V)(\mathbb{R}) \cong \mathrm{GL}_3(\mathbb{R})$.

When F is imaginary we have $F \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{C}$ and $D \otimes_{\mathbb{Q}} \mathbb{R} \cong M_3(\mathbb{C})$. By Skolem-Noether, we may assume that the automorphism induced by τ is conjugate transpose, subsequently that d is diagonal, and it follows that $\mathbf{U}(V)(\mathbb{R})$ is a classical unitary group. \square

We could state a result similar to Proposition 6.3 with \mathbb{R} replaced by any \mathbb{Q}_p : We would still distinguish cases based on whether p is split or inert in F , together with the ramified case. The main difference is that there exist degree 3 division algebras over \mathbb{Q}_p [45, §17.10] (and exactly two of them), which would be reflected in the classification. However, the only result that we will require is the following.

Lemma 6.4. *Let \mathbf{G} be a \mathbb{Q} -form of \mathbf{PGL}_3 and suppose p is a prime such that $G_{\mathbb{Q}_p}$ is split. Then $G_{\mathbb{Q}_p} \cong \mathbf{PGL}_{3, \mathbb{Q}_p}$.*

Proof. Both $G_{\mathbb{Q}_p}$ and $\mathbf{PGL}_{3, \mathbb{Q}_p}$ are split forms of $\mathbf{PGL}_{3, \overline{\mathbb{Q}_p}}$. Because a reductive group has only one split form over any field [9, Théorème 2.13], they are isomorphic. \square

6.3. Locally symmetric spaces. Because \mathbf{G} is connected, the Lie group $G = \mathbf{G}(\mathbb{R})$ is as in §2.1, and the identity component G^0 is semisimple. (Recall that we require a semisimple group by definition to be connected.) Let $K_{\infty} \subset G$ be a maximal compact subgroup. To G we associate data as in §2.1 with respect to the maximal compact K_{∞} : The Cartan involution is denoted by θ , the Iwasawa decomposition by NAK_{∞} and the restricted root system by (α^*, Σ) . (In §6.7 we shall use different notation for the roots of a maximal split torus over an algebraic closure, but that will be only relevant to define Hecke algebras.) We let $d(\cdot, \cdot)$ be a distance function as in §2.2.

Remark 6.5. Contrary to [10] we do not, and cannot, assume that the involution θ is defined over \mathbb{Q} . Indeed, when \mathbf{G} arises from a degree 3 division algebra over \mathbb{Q} (see §6.2), such θ would have to come from an anti-involution of the division algebra, and that does not exist [30, Corollary 2.8]. This remark will only play a role in Proposition 7.7.

Denote by $\mathbb{A}_{\mathbb{Q}}$ the ring of adèles over \mathbb{Q} and by \mathbb{A}_f the finite adèles. When S is any set of places of \mathbb{Q} , define \mathbb{A}_S to be the restricted product $\prod_{v \in S} \mathbb{A}_{\mathbb{Q}}$. Equip the groups $\mathbf{G}(\mathbb{Q}_p)$, $\mathbf{G}(\mathbb{A}_{\mathbb{Q}})$, $\mathbf{G}(\mathbb{A}_S)$ with their natural topologies, and similarly for \mathbf{H} .

Let $\rho : \mathbf{G} \rightarrow \mathbf{SL}_d$ and the integral models be as in §6.1. Choose compact subgroups $K_p \subset \mathbf{G}(\mathbb{Q}_p)$ for all primes p with the following properties:

- $\rho(K_p) \subset \mathbf{SL}_d(\mathbb{Z}_p)$;
- the compact subgroup

$$K_f := \prod_p K_p$$

- is open in $\mathbf{G}(\mathbb{A}_f)$;
- the center $\mathbf{Z}(\mathbf{G})(\mathbb{Q})$ intersects K_f trivially;

The first two conditions can be satisfied for example by taking $K_p = \mathbf{G}(\mathbb{Z}_p)$, and the third can be obtained by replacing any single K_p by a smaller subgroup. Define $K = K_\infty \times K_f \subset \mathbf{G}(\mathbb{A}_\mathbb{Q})$, and for any set S of places of \mathbb{Q} define $K_S = \prod_{v \in S} K_v$. Define the automorphic spaces

$$[\mathbf{G}] = \mathbf{G}(\mathbb{Q}) \backslash \mathbf{G}(\mathbb{A}_\mathbb{Q}) \quad \text{and} \quad [\mathbf{H}] = \mathbf{H}(\mathbb{Q}) \backslash \mathbf{H}(\mathbb{A}_\mathbb{Q}) \subset [\mathbf{G}]$$

and the locally symmetric space

$$X = [\mathbf{G}]/K.$$

It is a finite disjoint union of locally symmetric spaces as in §2.7 [46, Theorem 5.1] and is compact because \mathbf{G} is anisotropic [8]. Explicitly, let $(g_i)_{i \in I} \subset \mathbf{G}(\mathbb{A}_f)$ be a system of representatives for the double quotient $[\mathbf{G}]/(G \times K_f)$ and define $\Gamma_i = \mathbf{G}(\mathbb{Q}) \cap g_i K_f g_i^{-1}$. Then

$$(43) \quad [\mathbf{G}] = \bigsqcup_{i \in I} \Gamma_i \backslash (G \times g_i K_f)$$

and X is the disjoint union of the compact locally symmetric spaces

$$(44) \quad X_i = \Gamma_i \backslash G / K_\infty.$$

(Here we view $\mathbf{G}(\mathbb{Q})$ inside $\mathbf{G}(\mathbb{A}_f)$ when taking the intersection with $g_i K_f g_i^{-1}$, inside $\mathbf{G}(\mathbb{A}_\mathbb{Q})$ in (43) and inside G in (44). Notice that the lattices Γ_i do not depend on the choice of representatives g_i .)

6.4. Compact torus orbits. We explain now how to relate tori to maximal flat submanifolds of X . Let A as in §6.3 be the group given by the choice of Iwasawa decomposition.

Assume $\mathbf{H} \subset \mathbf{G}$ has maximal split rank over \mathbb{R} . Then there exists $g_\infty \in G$ such that $\mathbf{H}(\mathbb{R})$ contains $g_\infty A g_\infty^{-1}$, so that $\mathbf{H}(\mathbb{R}) \subset Z_G(g_\infty A g_\infty^{-1}) = g_\infty M A g_\infty^{-1}$. To \mathbf{H} we then associate the compact maximal flat submanifold $\mathcal{F} \subset X$ that is the image of $\mathbf{H}(\mathbb{A}_\mathbb{Q}) g_\infty$. Relative to the decomposition (44), it is the disjoint union of some (but not necessarily all) of the maximal flats $g_\infty A \subset X_i$. More precisely, \mathcal{F} intersects X_i nontrivially if and only if $\mathbf{H}(\mathbb{A}_f)$ intersects $g_i K_f$ nontrivially. Note that \mathcal{F} does not depend on the choice of g_∞ , because g_∞ is determined up to multiplication on the right by $N_G(A) = M'A$. Likewise, the definition does not depend on the choice of A , which can be seen by using that two choices of A are conjugate by an element of K_∞ .

Conversely, one can show that every compact maximal flat in X_i is obtained in this way. Assume $\mathcal{F} \subset X_i$ is a compact maximal flat submanifold. As in §2.8 there exists $g_\infty \in G$ such that \mathcal{F} is the image of $g_\infty A$ in X_i . The lattice Γ_i intersects $g_\infty A g_\infty^{-1}$ in a lattice, and define \mathbf{T} to be the Zariski closure of $\Gamma_i \cap g_\infty A g_\infty^{-1}$. Then

\mathbf{T} is a torus that is maximal split over \mathbb{R} , and we may take $\mathbf{H} \subset \mathbf{G}$ to be any maximal torus containing \mathbf{T} .

Remark 6.6. The correspondence between tori and maximal flats that we outlined above, is slightly awkward because X can have different components and because \mathbf{H} (in our notation) is always a maximal torus, not a maximal \mathbb{R} -split torus. When X is connected, one does obtain a clean bijection between compact maximal flats $\mathcal{F} \subset X$ and maximal \mathbb{R} -split tori $\mathbf{T} \subset \mathbf{G}$. Compare also [17, §2].

6.5. Integration. For every prime p we choose the Haar measures $d\mu_{\mathbf{G},p}$ and $d\mu_{\mathbf{H},p}$ on $\mathbf{G}(\mathbb{Q}_p)$ and $\mathbf{H}(\mathbb{Q}_p)$ respectively, for which K_p and $K_p \cap \mathbf{H}(\mathbb{Q}_p)$ have volume 1. We choose Haar measures on Lie groups as in §2.6. The abelian Lie group $\mathbf{H}(\mathbb{R})$ factors as a conjugate of A times a compact group, and we equip it with the measure that is the product of the measure coming from A with the volume 1 measure on the compact group. We form the product measures $d\mu_{\mathbf{G}} = \prod_v d\mu_{\mathbf{G},v}$ on $\mathbf{G}(\mathbb{A}_{\mathbb{Q}})$ and likewise the product measure $d\mu_{\mathbf{H}}$ on $\mathbf{H}(\mathbb{A}_{\mathbb{Q}})$. When S is any set of places of \mathbb{Q} , we similarly define measures $\mu_{\mathbf{G},S}$ and $\mu_{\mathbf{H},S}$ on $\mathbf{G}(\mathbb{A}_S)$ and $\mathbf{H}(\mathbb{A}_S)$ respectively.

6.6. Hecke algebras. A smooth function on $\mathbf{G}(\mathbb{Q}_p)$ is defined to be a locally constant one. When S is a set of places of \mathbb{Q} , a compactly supported Schwartz-Bruhat function on $\mathbf{G}(\mathbb{A}_S)$ is a finite linear combination of functions of the form

$$\prod_{v \in S} f_v,$$

where $f_v \in C_c^\infty(\mathbf{G}(\mathbb{Q}_v))$ and $f_p = 1_{K_p}$ for almost all primes $p \in S$. We denote by $C_c^\infty(\mathbf{G}(\mathbb{A}_S))$ the set of such (complex-valued) functions. If $f \in C_c^\infty(\mathbf{G}(\mathbb{A}_S))$, we define an operator $\pi(f)$ on $L^2([\mathbf{G}])$ by the rule

$$(\pi(f)\phi)(x) = \int_{\mathbf{G}(\mathbb{A}_S)} \phi(xg)f(g)d\mu_{\mathbf{G},S}(g).$$

If we define $f^*(g) = \overline{f(g^{-1})}$, then $\pi(f)$ and $\pi(f^*)$ are adjoints. For every set S of places of \mathbb{Q} , define the algebra \mathcal{H}_S of compactly supported smooth bi- K_S invariant functions on $\mathbf{G}(\mathbb{A}_S)$, with convolution given by

$$(f_1 * f_2)(g) = \int_{\mathbf{G}(\mathbb{A}_S)} f_1(gh^{-1})f_2(h)d\mu_{\mathbf{G},S}(h).$$

The algebras \mathcal{H}_S act on $L^2(X)$ by the same rule as above. When p is a prime, the element $1_{K_p} \in \mathcal{H}_p$ is the identity.

By [52, §3.9], there exists an integer $D > 0$ with the property that for $p \nmid D$, K_p is the compact subgroup associated with a hyperspecial point of the building of $\mathbf{G}_{\mathbb{Q}_p}$. By [52, §3.3.3] this implies that the Hecke algebra \mathcal{H}_p is commutative when $p \nmid D$.

Define \mathcal{H}_f to be the algebra \mathcal{H}_S when S consists of all finite places that do not divide D , and $\mathcal{H} = \mathcal{H}_\infty \otimes \mathcal{H}_f$. It is commutative by the above fact for finite places, and by [25, Theorem IV.3.1] for the infinite place.

6.7. Truncated Hecke algebras. We will use the notion of truncated Hecke algebras that is also used in [10], which makes it more convenient to formulate bounds for orbital integrals. While we will mostly be interested in primes at which all data is split, we do provide the general definition. Let \mathbf{T} be any maximal split torus of $\mathbf{G}_{\overline{\mathbb{Q}}}$, $X^*(\mathbf{T})$ and $X_*(\mathbf{T})$ be the groups of characters and cocharacters of \mathbf{T} . Let

Δ the set of roots of \mathbf{T} in $\mathbf{G}_{\overline{\mathbb{Q}}}$, Δ^+ be a choice of positive roots and W the Weyl group with respect to \mathbf{T} . Let $\rho \in X^*(\mathbf{T}) \otimes_{\mathbb{Z}} \mathbb{Q}$ be the half-sum of positive roots and denote the natural pairing $X^*(\mathbf{T}) \times X_*(\mathbf{T}) \rightarrow \mathbb{Z}$ by $\langle \cdot, \cdot \rangle$. Define a norm on $X_*(\mathbf{T})$ by

$$\|\mu\| = \max_{w \in W} \langle w\mu, \rho \rangle \in \frac{1}{2}\mathbb{Z}.$$

Now let $p \nmid D$ be a prime. As noted above there is a maximal split torus $\mathbf{A} \subset \mathbf{G}_{\mathbb{Q}_p}$ such that K_p corresponds to a hyperspecial point of the apartment of \mathbf{A} , and by the Cartan decomposition [52, §3.3.3] the double cosets $K_p \backslash \mathbf{G}(\mathbb{Q}_p) / K_p$ are represented uniquely by the elements $\mu(p)$ when $\mu \in X_*(\mathbf{A})$ runs through the cocharacters in the closure of the positive chamber. Conjugating \mathbf{A} to \mathbf{T} over $\overline{\mathbb{Q}_p}$ yields a norm $\|\cdot\|$ on $X_*(\mathbf{A})$ induced from the one on $X_*(\mathbf{T})$, which does not depend on the choice of \mathbf{A} [10, §2.6]. For $\kappa \geq 0$ we define now the ‘‘truncated algebra’’ (which is not an algebra)

$$\mathcal{H}_p^{\leq \kappa} = \text{span}_{\mathbb{C}} \{1_{K_p \mu(p) K_p} : \mu \in X_*(\mathbf{A}), \|\mu\| \leq \kappa\}.$$

When S is a set of primes not dividing D , we define $\mathcal{H}_S^{\leq \kappa}$ to be the restricted tensor product $\bigotimes_{p \in S} \mathcal{H}_p^{\leq \kappa} \subset \mathcal{H}_S$.

Finally, we will use yet another notion of truncated Hecke algebra to accommodate the Hecke operators that we will use to prove Theorem 1.2. When S is a set of primes not dividing D , $\kappa \geq 0$ and $M \geq 1$, define the truncated algebra

$$\mathcal{H}_{S,M}^{\leq \kappa} = \bigoplus_{\substack{n \leq M \\ \text{squarefree}}} \bigotimes_{\substack{p \in S \\ p|n}} \mathcal{H}_p^{\leq \kappa}.$$

6.8. Hecke-Maass forms. The group G acts on $[\mathbf{G}]$ by translation on the right, and this induces an action of the universal enveloping algebra $U(\mathfrak{g})$ on $C^\infty([\mathbf{G}])$, which using the decomposition (43) we may view as given by left-invariant differential operators on G (§2.5). By a Hecke-Maass form on X we mean a smooth function on X that is a joint eigenfunction for the center $Z(U(\mathfrak{g}))$ and for the Hecke algebra \mathcal{H}_f from §6.6.

Let $(f_j)_{j \geq 0}$ be an orthonormal basis of $L^2(X)$ consisting of (complex-valued) Hecke-Maass forms. It may be obtained by decomposing $L^2(X)$ into eigenspaces for the Laplace-Beltrami operator, and finding in every eigenspace a basis of eigenfunctions for the commutative algebras $Z(U(\mathfrak{g}))$ and \mathcal{H}_f . When $k \in \mathcal{H}$, \mathcal{H}_∞ or \mathcal{H}_f , denote by $\widehat{k}(f_j)$ the eigenvalue of f_j under k , and define the spectral parameter ν_j as in §2.7.

Let $g_\infty \in G$ be as in §6.4 with the property that $\mathbf{H}(\mathbb{R})$ contains $g_\infty A g_\infty^{-1}$. Define the \mathbf{H} -period of f_j by

$$(45) \quad \mathcal{P}_{\mathbf{H}}(f_j) = \int_{[\mathbf{H}]} f_j(h g_\infty) d\mu_{\mathbf{H}}(h),$$

where the automorphic quotient $[\mathbf{H}]$ is as in §6.3. As in §6.4, this definition does not depend on the choice of A or g_∞ .

7. EXTREME VALUES OF TORIC PERIODS

In this section we prove Theorem 1.2. We set up a relative pre-trace formula in §7.1 and prove the theorem in §7.3. Our notations for locally symmetric spaces, operator algebras and Hecke-Maass forms are as in §6.

7.1. Adelic setup. Let \mathbf{G}/\mathbb{Q} be semisimple anisotropic as in §6.1, and let $\mathbf{H} \subset \mathbf{G}$ be a maximal torus of maximal split rank over \mathbb{R} .

The pre-trace formula for \mathbf{G} states that for any $k \in \mathcal{H}$ (with archimedean component satisfying a positivity condition as in §3.1) one has

$$\sum_j \widehat{k}(f_j) f_j(x) \overline{f_j(y)} = \sum_{\gamma \in \mathbf{G}(\mathbb{Q})} k(x^{-1}\gamma y),$$

uniformly for $x, y \in [\mathbf{G}]$. As in §6.4 let $g_\infty \in \mathbf{G}(\mathbb{R})$ be such that $\mathbf{H}(\mathbb{R}) \supset g_\infty A g_\infty^{-1}$, define the \mathbf{H} -periods $\mathcal{P}_{\mathbf{H}}$ as in (45) and define $K_{\mathbf{H},f} = K_f \cap \mathbf{H}(\mathbb{A}_f)$.

We will require the following analogue of Lemma 3.1.

Lemma 7.1 (Partitions of unity). *There exists a nonnegative Schwartz-Bruhat function $b \in C_c^\infty(\mathbf{H}(\mathbb{A}_\mathbb{Q})/K_{\mathbf{H},f})$ such that $\sum_{\gamma \in \mathbf{H}(\mathbb{Q})} b(\gamma h) = 1$ for all $h \in \mathbf{H}(\mathbb{A}_\mathbb{Q})$.*

Proof. The quotient $\mathbf{H}(\mathbb{A}_\mathbb{Q})/(\mathbf{H}(\mathbb{Q})\mathbf{H}(\mathbb{R})K_{\mathbf{H},f})$ is finite by [46, Theorem 5.1]. Take coset representatives $h_1, \dots, h_N \in \mathbf{H}(\mathbb{A}_f)$ and define $b_f \in C_c^\infty(\mathbf{H}(\mathbb{A}_f))$ by

$$b_f = \sum_{j=1}^N 1_{h_j K_{\mathbf{H},f}}.$$

The group $\mathbf{H}(\mathbb{R})$ factors uniquely as $V \times T$ with V a Euclidean space and T compact (possibly disconnected). The discrete subgroup $\Gamma_{\mathbf{H}} := \mathbf{H}(\mathbb{Q}) \cap K_{\mathbf{H},f}$ is a lattice in $\mathbf{H}(\mathbb{R})$ and therefore $\Gamma_V := \Gamma_{\mathbf{H}} \cap V$ has finite index in $\Gamma_{\mathbf{H}}$. We may construct $b_V \in C_c^\infty(V)$ as in Lemma 3.1 relative to the lattice $\Gamma_V \subset V$ and define $b_T(h) = [\Gamma_{\mathbf{H}} : \Gamma_V]^{-1}$ for $h \in T$. Define $b_\infty \in C_c^\infty(\mathbf{H}(\mathbb{R}))$ by

$$b_\infty(h) = b_V(h_V) \cdot b_T(h_T),$$

where we denote $h = (h_V, h_T) \in V \times T \cong \mathbf{H}(\mathbb{R})$. Define now $b \in C_c^\infty(\mathbf{H}(\mathbb{A}_\mathbb{Q}))$ by

$$b = b_\infty b_f.$$

That this b satisfies the requirements can be checked by writing

$$\sum_{\gamma \in \mathbf{H}(\mathbb{Q})} b(\gamma h) = \sum_{\gamma \in \mathbf{H}(\mathbb{Q})/(\mathbf{H}(\mathbb{Q}) \cap K_{\mathbf{H},f})} \sum_{\mu \in \mathbf{H}(\mathbb{Q}) \cap K_{\mathbf{H},f}} b(\gamma \mu h). \quad \square$$

Remark 7.2. Lemma 7.1 is a special case of a much more general statement about continuous functions on locally compact homogeneous spaces; see [43, §III.4].

The Weyl group $N_{\mathbf{G}}(\mathbf{H})/\mathbf{H}$ is finite [7, §IV.11.19] and therefore so is its group of rational points.

Lemma 7.3. *Let $k \in C_c^\infty(K \backslash \mathbf{G}(\mathbb{A}_\mathbb{Q})/K)$ be a Schwartz-Bruhat function. There exists $b \in C_c^\infty(\mathbf{H}(\mathbb{A}_\mathbb{Q}))$ such that*

$$\begin{aligned} & \sum_j \widehat{k}(f_j) |\mathcal{P}_{\mathbf{H}}(f_j)|^2 \\ &= \text{Vol}([\mathbf{H}]) \sum_{\gamma \in N_{\mathbf{G}}(\mathbf{H})(\mathbb{Q})/\mathbf{H}(\mathbb{Q})} \int_{\mathbf{H}(\mathbb{A}_\mathbb{Q})} k(g_\infty^{-1} \gamma h g_\infty) d\mu_{\mathbf{H}}(h) \\ &+ \sum_{\gamma \in \mathbf{G}(\mathbb{Q}) - N_{\mathbf{G}}(\mathbf{H})(\mathbb{Q})} \int_{\mathbf{H}(\mathbb{A}_\mathbb{Q})^2} b(h_1) b(h_2) k(g_\infty^{-1} h_1^{-1} \gamma h_2 g_\infty) d\mu_{\mathbf{H}}(h_1) d\mu_{\mathbf{H}}(h_2). \end{aligned}$$

Proof. Using Lemma 7.1, this is entirely analogous to the proof of Lemma 3.2 by an unfolding argument and introduction of cutoff functions. \square

Remark 7.4. The integrals in the diagonal term in Lemma 7.3 can depend on (the finite part of) γ , because it is not always true that $N_{\mathbf{G}}(\mathbf{H})(\mathbb{Q}) \subset \mathbf{H}(\mathbb{Q}) \cdot K_f$. When G is \mathbf{PGL}_2 or an inner form, this inclusion is related to the notion of “reciprocal geodesics” in [48].

7.2. Comparison of trace formulas. We now specialize to the groups \mathbf{G} in Theorem 1.2, although \mathbf{G} can still be as general as in Theorem 1.3 in this subsection. We will prove Theorem 1.2 by comparing asymptotics for an amplified trace formula and an amplified relative trace formula.

Let $\rho : \mathbf{G} \rightarrow \mathbf{SL}_d$ be as in §6.1, and when p is prime define $\|g\|_p$ for $g \in \mathbf{G}(\mathbb{Q}_p)$ as the maximal absolute value of the entries of $\rho(g)$. When $g \in \mathbf{G}(\mathbb{Q})$ we have for almost all primes p that $\rho(g) \in \mathbf{SL}_d(\mathbb{Z}_p)$ and consequently $\|g\|_p = 1$. We may therefore define $\|g\|_f = \prod_p \|g\|_p$.

Lemma 7.5. *There exists $A > 0$ such that when $\gamma \in \mathbf{G}(\mathbb{Q})$ is such that $g_\infty^{-1}\gamma g_\infty \notin \bigcup_{L \in \mathcal{L} - \{G\}} M'L$, then*

$$d \left(g_\infty^{-1}\gamma g_\infty, \bigcup_{L \in \mathcal{L} - \{G\}} M'L \right) \gg \|\gamma\|_f^{-A}.$$

Proof. This can be shown as in [10, Lemma 5.1] as a consequence of the product formula for global fields. The difference is that we must deal with varieties that are not defined over our base field \mathbb{Q} .

It suffices to prove a bound as in the statement for every individual Levi $L \in \mathcal{L} - \{G\}$. Fix such L .

Let $E \subset \mathbb{R}$ be a number field such that the extension of scalars \mathbf{H}_E has the same split rank as $\mathbf{H}_{\mathbb{R}}$. Let $\mathbf{T} \subset \mathbf{H}_E$ be the maximal split subtorus, defined over E . Because all roots of $\mathbf{H}_{\mathbb{R}}$ are defined over E , we may find an element $h \in \mathbf{T}(E)$ whose centralizer in \mathbf{G} has real points $g_\infty L g_\infty^{-1}$. Consider now the subvariety \mathbf{V} of \mathbf{G} defined over E by the equation $\text{Ad}_g(h) \in \mathbf{H}$. Replacing h by a finite power if necessary, we may assume that $h \in \mathbf{T}(\mathbb{R})^0 = g_\infty A g_\infty^{-1}$. Lemma 2.1 then implies that $\mathbf{V}(\mathbb{R}) = g_\infty M' L g_\infty^{-1}$. Thus by assumption, $\gamma \notin \mathbf{V}(E)$.

Consider now the embedding $\rho : \mathbf{G} \rightarrow \mathbf{SL}_d \subset \mathbf{A}_{\mathbb{Q}}^{d^2}$, and let p_1, \dots, p_k be polynomials defined over E whose joint zero locus is $\rho(\mathbf{V})$. Expanding their coefficients in a basis for E/\mathbb{Q} , we may find polynomials q_1, \dots, q_t defined over \mathbb{Q} whose rational joint zero locus is $\rho(\mathbf{V}(E) \cap \mathbf{G}(\mathbb{Q}))$. There exists $A > 0$ such that for all q_i and all $\gamma \in \mathbf{G}(\mathbb{Q})$ we have a bound of the form

$$\prod_p |p_i(\rho(\gamma))|_p \ll \|\gamma\|_f^A.$$

Indeed, $A > 0$ can be chosen to be $\max_i \deg(q_i)$, simply because a polynomial of degree s can only increase denominators as much as raising them to the power s .

Now assume $\gamma \in \mathbf{G}(\mathbb{Q})$ does not lie in $\mathbf{V}(E)$. Then there exists q_i such that $q_i(\rho(\gamma)) \neq 0$. By the product formula for \mathbb{Q}^\times , this means that

$$\prod_v |q_i(\rho(\gamma))|_v = 1,$$

where the product now runs over all absolute values of \mathbb{Q} . Using our bound for the product over the finite primes, this gives

$$|q_i(\rho(\gamma))|_\infty \gg \|\gamma\|_f^{-A}.$$

By smoothness of the action of $\mathbf{G}(\mathbb{R})$ on h , the left-hand side is bounded from above by a constant times $d(\gamma, \mathbf{V}(\mathbb{R}))$. This shows that

$$d(\gamma, g_\infty M' L g_\infty^{-1}) \gg \|\gamma\|_f^{-A},$$

and the lemma follows. \square

When S is a finite set of primes, define $q_S = \prod_{p \in S} p$. Let the integer D be as in §6.6.

Proposition 7.6. *Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\text{gen}}$ be compact. let $\kappa \geq 0$. There exist $A, B, \delta > 0$ such that the following holds. Let S a finite set of primes not dividing D and $k_f \in \mathcal{H}_S^{\leq \kappa}$. Let $\nu \in D_{\mathfrak{a}^*}$, $t \geq 0$ and $k_{t\nu} \in \mathcal{H}_\infty$ as in Lemma 3.3. Define $k = k_{t\nu} \otimes k_f$. Then*

$$\begin{aligned} \sum_j \widehat{k}(f_j) |\mathcal{P}_{\mathbf{H}}(f_j)|^2 &= \text{Vol}([\mathbf{H}]) \cdot \sum_{\gamma \in N_{\mathbf{G}}(\mathbf{H})(\mathbb{Q})/\mathbf{H}(\mathbb{Q})} \int_{\mathbf{H}(\mathbb{A}_{\mathbb{Q}})} k(g_\infty^{-1} \gamma h g_\infty) d\mu_{\mathbf{H}}(h) \\ &\quad + O(\beta(t\nu)(1+t)^{-r}(1+q_S^{-A}t)^{-\delta} q_S^B \|k_f\|_\infty), \end{aligned}$$

uniformly in ν and t . If $M \geq 1$ and $k_f \in \mathcal{H}_{S,M}^{\leq \kappa}$ then the same result holds with q_S replaced by M .

Proof. We start from Lemma 7.3 and must bound the off-diagonal terms. The fixed function b is a finite sum of factorizable Schwartz-Bruhat functions, so up to introducing a finite sum, we may replace the factors $b(h_1)$ and $b(h_2)$ by factorizable Schwartz-Bruhat functions $b_1(h_1)$ and $b_2(h_2)$. (In fact, we only require that they factorize as an infinite times a finite part, and this is already the case for b as constructed in Lemma 7.1.) For fixed γ , the double integral now factorizes as the integral

$$\int_{\mathbf{H}(\mathbb{R})^2} b_1(h_1) b_2(h_2) k_{t\nu}(g_\infty^{-1} h_1^{-1} \gamma h_2 g_\infty) d\mu_{\mathbf{H},\mathbb{R}}(h_1) d\mu_{\mathbf{H},\mathbb{R}}(h_2)$$

times an integral involving b_1 , b_2 and k_f over a fixed compact subset of $\mathbf{H}(\mathbb{A}_f)$. The latter can be bounded trivially by

$$\ll \|k_f\|_\infty \cdot \text{Vol}(\text{supp}(b_1)) \cdot \text{Vol}(\text{supp}(b_2)) \cdot \text{Vol}(\text{supp}(k_f)).$$

The first two volumes are bounded, and the third can be bounded by a power of q_S , as in [10, Lemma 4.4].

The archimedean integral equals

$$\int_{A \times A} b_1(g_\infty a_1 g_\infty^{-1}) b_2(g_\infty a_2 g_\infty^{-1}) k_{t\nu}(a_1 g_\infty^{-1} \gamma g_\infty a_2) da_1 da_2,$$

where we have written $\mathbf{H}(\mathbb{R})$ as a product of $g_\infty A g_\infty^{-1}$ and a compact torus, where the compact torus necessarily lies in $g_\infty M g_\infty^{-1}$ (see §6.4) and we may omit it from the integration by our choice of measure on $\mathbf{H}(\mathbb{R})$ (see §6.5). By our choice of G and X , we have that

$$g_\infty^{-1} \gamma g_\infty \notin \bigcup_{L \in \mathcal{L} - \{G\}} M' L$$

for $\gamma \in \mathbf{G}(\mathbb{Q})$ as above. Using the inversion formula (15), the rapid decay of $\widehat{k}_{t\nu}$ and Proposition 5.1, this means that the archimedean integral above is bounded by

$$\ll \beta(t\nu)(1+t)^{-r} \left(1 + d \left(g_\infty^{-1} \gamma g_\infty, \bigcup_{L \in \mathcal{L} - \{G\}} M'L \right) \cdot t \right)^{-\delta}$$

for certain $\delta, N > 0$. By Lemma 7.5 we may bound the distance from below by $\|\gamma\|_f^{-A}$ for some $A > 0$, and as in [10, Lemma 4.4] the latter is again bounded from below by $q_S^{-A'}$ for some $A' > 0$. It remains to show that the number of γ that contribute to the sum is at most polynomial in q_S . Indeed, for such γ we have that the rational element $\rho(\gamma) \in \mathbf{SL}_d(\mathbb{Q})$ is bounded in the archimedean sense by the support conditions on b_1 and b_2 , and its denominators are at most of size $q_S^{A'}$ by the same argument we just used. This proves the desired bound for the off-diagonal terms.

When $k_f \in \mathcal{H}_{S,M}^{\leq \kappa}$, it is a sum of at most M elements $k_{f,n} \in \mathcal{H}_{S_n}^{\leq \kappa}$, with S_n a set of primes with $q_{S_n} \leq M$. Moreover, we may assume that the $k_{f,n}$ have disjoint supports. The result then follows by summing the different bounds for each of the $k_{f,n}$, and replacing B by $B+1$. \square

We will require the following estimates for the trace formula.

Proposition 7.7. *There are constants $\eta, A, B > 0$ and an integer multiple D' of D such that the following holds. Let $\kappa \geq 0$ and S a finite set of primes not dividing D' , and $k_f \in \mathcal{H}_S^{\leq \kappa}$. Let $D_{\mathfrak{a}^*} \subset \mathfrak{a}^* - \{0\}$ be compact, $\nu \in D_{\mathfrak{a}^*}$, $t \geq 1$ and $k_{t\nu} \in \mathcal{H}_\infty$ as in Lemma 3.3. Define $k = k_{t\nu} \otimes k_f$. Then*

$$\sum_j \widehat{k}(f_j) = \text{Vol}(X) \cdot k(1) + O\left(\beta(t\nu)(1+t)^{-\eta} q_S^{A\kappa+B} \|k_f\|_\infty\right),$$

uniformly in ν and t . If $M \geq 1$ and $k_f \in \mathcal{H}_{S,M}^{\leq \kappa}$ then the same result holds with q_S replaced by M .

Proof. The first statement follows from [10, Theorem 3.1]. Note that our Cartan involution is not assumed to be defined over \mathbb{Q} (see Remark 6.5). This is compensated for by our assumption that \mathbf{G} is almost simple over \mathbb{R} , which we can use to replace the argument in [10, Lemma 6.3] needed to show that the centralizer of G^0 in $\mathbf{G}(\mathbb{Q})$ is reduced to the center.

The statement about $k_f \in \mathcal{H}_{S,M}^{\leq \kappa}$ is shown as for Proposition 7.6. \square

Remark 7.8. The estimate [10, Theorem 3.1] is stronger than what is needed here, because we do not need uniformity with respect to a variable compact subgroup K (the level aspect) nor do we need asymptotics for spectral parameters that can be singular. But it is the only directly quotable Weyl law-type result that we have found, that covers anisotropic groups and adelic test functions. Note also that the exponents A and B are ineffective, but they will not play any role.

We will apply Propositions 7.6 and 7.7 always with the following positivity assumption on k_f :

$$(46) \quad \widehat{k}_f(f_j) \geq 0 \text{ for all } f_j, k_f(1) \geq 1 \text{ and } k_f(g) \geq 0 \text{ for } g \in G(\mathbb{A}_f).$$

The following Proposition combines the two trace formulae and reduces Theorem 1.2 to finding a suitable test function k_f .

Proposition 7.9. *Let $D_{\mathfrak{a}^*} \subset (\mathfrak{a}^*)^{\text{gen}}$ be compact. Let $\kappa \geq 0$. There exist $\delta, C > 0$ and an integer multiple D' of D such that the following holds. Let $\nu \in D_{\mathfrak{a}^*}$, S be a finite set of primes not dividing D' , $t, M \geq 1$ and $k_f \in \mathcal{H}_{S, M}^{\leq \kappa}$ satisfying (46). If $M \leq (1+t)^\delta$ and $\|k_f\|_\infty \leq (1+t)^\delta$, then*

$$\frac{\sum_{\|\nu_j - t\nu\| \leq C} \widehat{k}_f(f_j) |\mathcal{P}_H(f_j)|^2}{\sum_{\|\nu_j - t\nu\| \leq C} \widehat{k}_f(f_j)} \gg (1+t)^{-r} \cdot \frac{\int_{H(\mathbb{A}_f)} k_f}{k_f(1)},$$

uniformly in ν, t, S, M, k_f .

Proof. We first bound the main term in Proposition 7.6 from below. The integrals factorize. The archimedean component can be bounded from below by $\beta(t\nu)(1+t)^{-r}$. Indeed, using the inversion formula (15) it reduces to a similar integral involving the spherical function, which is seen to be independent of γ by making a change of variables, and we conclude using Proposition 4.1. For the finite component we may use the positivity of k_f and bound the integral trivially from below by $k_f(1) \text{Vol}(\mathbf{H}(\mathbb{A}_f) \cap K_f) \geq 1$. Given this lower bound for the main term and given the quality of the error term in Proposition 7.6, a truncation argument as in [10, Lemma 4.5] shows the asymptotic formula

$$\begin{aligned} & \sum_{\|\nu_j - t\nu\| \leq C} \widehat{k}_f(f_j) |\mathcal{P}_H(f_j)|^2 \\ & \asymp \text{Vol}([\mathbf{H}]) \sum_{\gamma \in N_{\mathbf{G}}(\mathbf{H}(\mathbb{Q}))/\mathbf{H}(\mathbb{Q})} \int_{\mathbf{H}(\mathbb{A}_{\mathbb{Q}})} k(g_\infty^{-1} \gamma h g_\infty) d\mu_{\mathbf{H}}(h), \end{aligned}$$

provided that C is large enough and that M and $\|k_f\|_\infty$ are bounded by a sufficiently small power of t so that the error terms are controlled. Similarly, the inversion formula (15) gives that $k_{t\nu}(1) \asymp \beta(t\nu)$ and a truncation argument applied to Proposition 7.7 gives that

$$\sum_{\|\nu_j - t\nu\| \leq C} \widehat{k}_f(f_j) \asymp \text{Vol}(X) k(1),$$

provided that C is large enough and that M and $\|k_f\|_\infty$ are bounded by a sufficiently small power of t .

The statement now follows by taking the quotient of those two asymptotics, using again the archimedean asymptotics, and in the case of the relative trace formula, using positivity of k_f to discard the terms corresponding to γ that are not the identity. \square

7.3. Amplification. The construction of amplifiers is our only reason to restrict to groups \mathbf{G} as in Theorem 1.2. The theorem follows by applying Proposition 7.9 with k_f as given by the following proposition, and with M a sufficiently small power of t .

Proposition 7.10. *Let G be a \mathbb{Q} -form of \mathbf{PGL}_3 and let δ be as stated in Theorem 1.2. Let $M \geq 2$ and S be the set of primes less than M that do not divide D' . There exists $k_f \in \mathcal{H}_{S, M}^{\leq 4}$ satisfying (46), such that $\|k_f\|_\infty \leq M^A$ for some $A > 0$ and*

$$\frac{\int_{\mathbf{H}(\mathbb{A}_f)} k_f}{k_f(1)} \gg (\log \log M)^{\delta + o(1)}.$$

We prove Proposition 7.10 in §8. We will also prove a result that states that the lower bound in Proposition 7.10 is optimal in a certain sense, although the exact result is not as strong as the optimality statement in [39], and we do not formally exclude the existence of amplifiers of similar quality for, say, forms of \mathbf{PGL}_n . See Remark 8.13.

8. CONSTRUCTION OF AMPLIFIERS

In this section we prove the existence of the amplifier in Proposition 7.10 (in §8.2). We also prove an optimality result in Proposition 8.12 (in §8.3).

8.1. Preliminary computations. We begin with some preliminary computations of integrals on p -adic groups. It is convenient to re-introduce some notation in order to make the key computations more self-contained. Let $n \geq 2$, let p be a prime number and denote $G_p = \mathbf{PGL}_n(\mathbb{Q}_p)$, $K_p = \mathbf{PGL}_n(\mathbb{Z}_p)$ and define $H_p \subset G_p$ to be the diagonal torus. Most of the time we will specialize to $n = 3$, but we allow some generality to be able to make remarks about other n , and because certain statements will be proven for arbitrary n . We fix Haar measures on G_p and H_p such that the compact open subgroups K_p and $H_p \cap K_p$ have volume 1. Define $\mathcal{H}_p = C_c^\infty(K_p \backslash G_p / K_p)$, a commutative algebra with convolution defined by

$$(k_1 * k_2)(g) = \int_{G_p} k_1(h)k_2(h^{-1}g)dh.$$

The adjoint of $k \in \mathcal{H}_p$ is defined by $k^*(g) = \overline{k(g^{-1})}$. This defines an algebra involution on \mathcal{H}_p .

The question we seek to answer is the following: When $k_p \in \mathcal{H}_p$ is a function of the form $k'_p * (k'_p)^*$ with $k'_p \in \mathcal{H}_p$, how big can $\int_{H_p} k_p$ be relative to $k_p(1)$?

Such k'_p is a finite linear combination of the basic double coset kernels, which are defined as follows. Let $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{Z}^n$ and denote $\mu_{\mathbf{a}}(p) = \text{diag}(p^{a_1}, \dots, p^{a_n})$. One may view $\mu_{\mathbf{a}}$ as a cocharacter of H_p . Define the function

$$\tau(\mathbf{a}, p) = 1_{K_p \mu_{\mathbf{a}}(p) K_p}.$$

We have $\tau(\mathbf{a}, p)^* = \tau(-\mathbf{a}, p)$, and $\tau(\mathbf{0}, p)$ is the identity of $C_c^\infty(K_p \backslash G_p / K_p)$. The function $\tau(\mathbf{a}, p)$ depends only on the entries of \mathbf{a} up to permutation and up to translation by a common element in \mathbb{Z} . We may therefore always reduce to the case where $a_1 \geq a_2 \geq \dots \geq a_n = 0$. The degree $\text{deg}(\mu_{\mathbf{a}}(p))$ is defined as the cardinality $\#(K_p \mu_{\mathbf{a}}(p) K_p / K_p)$. With our choice of Haar measures we have

$$(47) \quad \text{deg}(\mu_{\mathbf{a}}(p)) = \|\tau(\mathbf{a}, p)\|_2^2 = (\tau(\mathbf{a}, p) * \tau(-\mathbf{a}, p))(1).$$

We will need bounds for integrals of convolutions of these basic kernels. Specifically, the quantity we will be interested in is the off-diagonal contribution to the integral $\int_{H_p} \tau(\mathbf{a}_1, p) * \tau(\mathbf{a}_2, p)$, when $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{Z}^n$. That is, we are interested in the integral over $H_p - (H_p \cap K_p)$, or what is the same, the integral over H_p minus $(\tau(\mathbf{a}_1, p) * \tau(\mathbf{a}_2, p))(1)$.

The following is an example computation of off-diagonal contributions when $n = 2$, which to an extent can be seen geometrically in the Bruhat-Tits tree.

Example 8.1. Let $n = 2$ and define $\mathbf{a} = (1, 0)$ and $\mathbf{b} = (0, 0)$. Then $\deg(\mu_{\mathbf{a}}(p)) = p + 1$, $\tau(\mathbf{a}, p)$ is self-adjoint and

$$\left(\int_{H_p} \tau(\mathbf{a}, p) * \tau(\mathbf{a}, p) \right) - \deg(\mu_{\mathbf{a}}(p)) = (p + 3) - (p + 1) = 2,$$

$$\int_{H_p} \tau(\mathbf{a}, p) * \tau(\mathbf{b}, p) = 2.$$

That $\deg(\mu_{\mathbf{a}}(p)) = p + 1$, the cardinality of a radius 1 ball in the Bruhat-Tits tree, is a classical counting problem of double cosets. That $\tau(\mathbf{a}, p)$ is self-adjoint holds because \mathbf{a} and $-\mathbf{a}$ are \mathbb{Z} -translates. We look at the second integral first. It equals 2 because it is the intersection of a radius 1 ball in the Bruhat-Tits tree with an apartment containing the center of the ball. We can also show this more explicitly as follows. We must count how many left cosets of K_p are contained in $K_p \operatorname{diag}(p, 1) K_p$ and intersect H_p . We may lift this problem to $\operatorname{GL}_2(\mathbb{Q}_p)$ and ask how many diagonal matrices lie in $\operatorname{GL}_2(\mathbb{Z}_p) \operatorname{diag}(p, 1) \operatorname{GL}_2(\mathbb{Z}_p)$. Multiplication on the left or the right by $\operatorname{GL}_2(\mathbb{Z}_p)$ does not change the invariant factors of a matrix, so that such a diagonal matrix must have invariant factors $(p, 1)$. It is clear that this gives exactly two matrices, and the statement follows.

The first integral can be computed using a similar argument, after using the Hecke relation $\tau(\mathbf{a}, p) * \tau(\mathbf{a}, p) = \tau(2\mathbf{a}, p) + (p + 1)\tau(\mathbf{0}, p)$.

Lemma 8.2. *When $\mathbf{a}, \mathbf{b} \in \mathbb{Z}^n$ are decreasing tuples with $a_n = b_n = 0$, we have $(\tau(\mathbf{a}, p) * \tau(-\mathbf{b}, p))(1) = \delta_{\mathbf{a}, \mathbf{b}} \deg(\mu_{\mathbf{a}}(p))$.*

Proof. When $\mathbf{a} = \mathbf{b}$ this is (47). When $\mathbf{a} \neq \mathbf{b}$ the double cosets represented by $\mu_{\mathbf{a}}(p)$ and $\mu_{\mathbf{b}}(p)$ are distinct, either by the Cartan decomposition [52, §3.3.3] or by an argument using invariant factors. They are therefore disjoint, so that

$$(\tau(\mathbf{a}, p) * \tau(-\mathbf{b}, p))(1) = \langle \tau(\mathbf{a}, p), \tau(\mathbf{b}, p) \rangle_{L^2(G_p)} = 0. \quad \square$$

To do explicit computations, we translate integrations into counting problems involving lattices in \mathbb{Q}_p^n , which are the natural generalization of points in the Bruhat-Tits tree. By a lattice in \mathbb{Q}_p^n we mean a finitely generated \mathbb{Z}_p -submodule of rank n . Denote by \mathcal{R} the set of lattices in \mathbb{Q}_p^n , and by $\overline{\mathcal{R}}$ the set of homothety classes of lattices. The group $\operatorname{GL}_n(\mathbb{Q}_p)$ acts on \mathcal{R} and the group G_p acts on $\overline{\mathcal{R}}$. Denote $L_0 = \mathbb{Z}_p^n$ and $\overline{L_0}$ its homothety class; they are our base points. The stabilizers of L_0 and $\overline{L_0}$ in $\operatorname{GL}_n(\mathbb{Q}_p)$ and G_p are $\operatorname{GL}_n(\mathbb{Z}_p)$ and K_p , respectively. Acting on the base point yields bijections $\operatorname{GL}_n(\mathbb{Q}_p)/\operatorname{GL}_n(\mathbb{Z}_p) \cong \mathcal{R}$ and $G_p/K_p \cong \overline{\mathcal{R}}$.

Lemma 8.3. *Let $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{Z}^n$ with $a_1 \geq a_2 \geq \dots \geq a_n = 0$. There is a natural bijection between the following sets:*

- The lattices $L \subset L_0$ for which L_0/L has invariant factors p^{a_1}, \dots, p^{a_n} ;
- The left coset space $K_p \mu_{\mathbf{a}}(p) K_p / K_p$,

which is given as follows: To a lattice L one associates the homothety class $\overline{L} \in \overline{\mathcal{R}}$, which is identified with a left coset of K_p through the bijection $G_p/K_p \cong \overline{\mathcal{R}}$.

Proof. That the map is well-defined (meaning, that it lands in $K_p \mu_{\mathbf{a}}(p) K_p / K_p$) is the following fact: When M is a free module over a PID with a submodule N , then there exists a basis (e_i) of M that is adapted to N , meaning that N has a basis consisting of scalar multiples of the e_i . This fact shows that every lattice $L \subset L_0$

such that L_0/L has invariant factors as given, is of the form $k_p \operatorname{diag}(a_1, \dots, a_n)L_0$ with $k_p \in \operatorname{GL}_n(\mathbb{Z}_p)$. That the map is surjective is trivial, because the lattice $k_p \operatorname{diag}(a_1, \dots, a_n)L_0$ is sent to the corresponding left coset in $K_p \mu_{\mathbf{a}}(p)K_p/K_p$. \square

Denote by (e_1, \dots, e_n) the standard basis of L_0 . We call $L \in \mathcal{R}$ an adapted lattice if it has a basis of the form $(b_1 e_1, \dots, b_n e_n)$ with the $b_i \in \mathbb{Q}_p^\times$.

Lemma 8.4. *Let $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{Z}^n$ with $a_1 \geq a_2 \geq \dots \geq a_n = 0$. The bijection from Lemma 8.3 restricts to a bijection between the following sets:*

- The set of adapted lattices $L \subset L_0$ for which L_0/L has invariant factors p^{a_1}, \dots, p^{a_n} .
- $(H_p \cap K_p \mu_{\mathbf{a}}(p)K_p)/(H_p \cap K_p)$.

Proof. This can be proven as for Lemma 8.3. \square

Lemma 8.5. *Let $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$ be decreasing tuples with $a_n = b_n = 0$. Then $\int_{H_p} \tau(\mathbf{a}, p) * \tau(-\mathbf{b}, p)$ counts the pairs of lattices (L_1, L) with the following properties:*

- L_1 is adapted and $L \subset L_0 \cap L_1$.
- The invariant factors of L_0/L are given by \mathbf{a} and those of L_1/L by \mathbf{b} .

Proof. We have

$$\int_{H_p} \tau(\mathbf{a}, p) * \tau(-\mathbf{b}, p) = \int_{H_p/(H_p \cap K_p)} \int_{G_p/K_p} \tau(\mathbf{a}, p)(g) \tau(-\mathbf{b}, p)(g^{-1}h) dg dh.$$

By our convention on Haar measures, the right hand side is a sum over cosets $h(H_p \cap K_p)$ and gK_p . Take such cosets for which the integrand is nonzero (and thus equal to 1). Then $g\bar{L}_0$ is represented by a unique lattice $L \subset L_0$ for which L/L_0 has invariant factors given by \mathbf{a} , and this L does not depend on the representative g . There is a unique lift $h_1 \in \operatorname{GL}_n(\mathbb{Q}_p)$ such that $h_1^{-1}L \subset L_0$ has invariant factors given by \mathbf{b} , and this again does not depend on the representative h . Define then $L_1 = h_1 L_0$. The pair (L_1, L) satisfies the conditions of the statement. Conversely, to such a pair we can associated unique cosets $h(H_p \cap K_p)$ and gK_p on which the integrand is nonzero. \square

8.2. Lower bounds. We prove Proposition 7.10. Our construction of an amplifier k_f is inspired by the construction in [39]. When translated into the adelic language, the amplifier from [39] takes the form $k_f = \omega * \omega^*$, with

$$(48) \quad \omega = \sum_{\substack{n \leq M \\ \text{squarefree}}} \prod_{p|n} c_p \omega_p,$$

for ω_p elementary Hecke operators, which in the notation from §8.1 equal $\tau((1, 0), p)$, and with $c_p > 0$ parameters to be optimized. The equalities in Example 8.1, which are about $\tau((1, 0), p)$, appear in [39], not explicitly but through global versions thereof that are formulated in terms of divisor functions, in the proof of [39, Lemma 5].

One can think of the Hecke operator ω in (48) as being the formal expansion of the product $\prod(1 + c_p \omega_p)$, truncated to only include terms corresponding to sets of primes with product less than M .

Our aim is to optimize the choice of ω_p and of c_p for forms of \mathbf{PGL}_3 . We continue to use the notation from §8.1 but will soon after switch to a global setup. The following lemma is the key in the construction of an amplifier.

Lemma 8.6. *Let $n = 3$ and $\mathbf{a} = (1, 0, 0)$ or $(1, 1, 0)$. Then $\deg(\mu_{\mathbf{a}}(p)) = p^2 + p + 1$, and we have*

$$\begin{aligned} \int_{H_p} \tau(\mathbf{a}, p) * \tau(\mathbf{a}, p) &= 3(p+2), \\ \left(\int_{H_p} \tau(\mathbf{a}, p) * \tau(-\mathbf{a}, p) \right) - \deg(\mu_{\mathbf{a}}(p)) &= 6, \\ \int_{H_p} \tau(\mathbf{a}, p) &= 3. \end{aligned}$$

Proof. Using that adjugation is an algebra involution and that H_p is unimodular, it suffices to prove each identity for either $\mathbf{a} = (1, 0, 0)$ or $(1, 1, 0)$, because $\tau(\mathbf{a}, p)$ and $\tau(-\mathbf{a}, p)$ are adjoints. When $L_1, L_2 \in \mathcal{R}$ are lattices, define their generalized index by

$$[L_1 : L_2] = \frac{[L_1 : L_1 \cap L_2]}{[L_2 : L_1 \cap L_2]}.$$

It satisfies the usual transitivity relation.

For the degree, choose $\mathbf{a} = (1, 0, 0)$. We must count lattices $L \subset L_0$ of index p . Such L contain pL_0 and are determined by their quotient modulo pL_0 . We must then count index p subgroups in $(\mathbb{Z}/p\mathbb{Z})^3$, or equivalently lines in \mathbb{F}_p^3 , of which there are $p^2 + p + 1$.

For the first integral, choose $\mathbf{a} = (1, 1, 0)$. We must count pairs of lattices (L_1, L) satisfying the conditions in Lemma 8.5, with $\mathbf{b} = (1, 0, 0)$. Thus $L \subset L_0$ is a lattice with $L_0/L \cong (\mathbb{Z}/p\mathbb{Z})^2$ and $[L_1 : L] = p$. Because L_1 is adapted, contains $L \supset p\mathbb{Z}_p$ and $[L_0 : L_1] = [L_0 : L][L : L_1] = p$, it is clear that the only possibilities for L_1 are the following:

- $p^{-1}\mathbb{Z}_p \oplus p\mathbb{Z}_p \oplus p\mathbb{Z}_p$ and
- $\mathbb{Z}_p \oplus \mathbb{Z}_p \oplus p\mathbb{Z}_p$,

up to permutations of the factors. (This list would have been longer if we had switched \mathbf{a} and \mathbf{b} ; what helps us here is that \mathbf{b} has small entries.) It suffices to count pairs (L_1, L) where L_1 is one of these two lattices, and multiply the result by 3. In the first case, the condition $L \subset L_0 \cap L_1$ forces $L \subset \mathbb{Z}_p \oplus p\mathbb{Z}_p \oplus p\mathbb{Z}_p$, and by considering indices we must have equality. This L is indeed a solution. In the second case, note that L is determined by its quotient modulo pL_0 , and that $L/pL_0 \subset L_1/pL_0 \cong (\mathbb{Z}/p\mathbb{Z})^2$ has index p . There are $p+1$ such subgroups, and the corresponding L satisfy the conditions. This proves the first statement.

For the second integral, take instead $\mathbf{a} = (1, 0, 0)$ and apply Lemma 8.5. Using as before the observation that $L \supset p\mathbb{Z}_p$ and using that $[L_0 : L_1] = 1$, the possibilities for L_1 are now:

- L_0 ,
- $p^{-1}\mathbb{Z}_p \oplus \mathbb{Z}_p \oplus p\mathbb{Z}_p$ and
- $p^{-2}\mathbb{Z}_p \oplus p\mathbb{Z}_p \oplus p\mathbb{Z}_p$,

up to permutation of the factors. When $L_1 = L_0$ we obtain $\deg(\mu_{\mathbf{a}}(p))$ pairs (L_0, L) . In the second case we use that $L \subset L_0 \cap L_1$ and observe that equality must

occur, giving one solution L . In the third case the condition $L \subset L_0 \cap L_1$ forces $L \subset \mathbb{Z}_p \oplus p\mathbb{Z}_p \oplus p\mathbb{Z}_p$, and this is impossible. The second statement follows, taking into account permutations.

The computation of the last integral also follows from Lemma 8.5, with $\mathbf{b} = \mathbf{0}$. We must count the adapted lattices of index p in L_0 , and there are 3 of them. \square

We now return to the notation from the previous subsections, and let \mathbf{G} be a form of \mathbf{PGL}_3 . Let E be the minimal splitting field of \mathbf{H} and choose an isomorphism $\sigma : \mathbf{G}_E \rightarrow \mathbf{PGL}_{3,E}$ that sends \mathbf{H} to the diagonal torus. Define $\mathcal{P}_{\text{good}}$ to be the set of primes with the following properties:

- (1) $p \nmid D'$, with D' as in Proposition 7.9;
- (2) p splits in E ; equivalently, E embeds in \mathbb{Q}_p ;
- (3) $\sigma(K_p) = \mathbf{PGL}_3(\mathbb{Z}_p)$.

The condition that p splits in E is equivalent to saying that \mathbf{H} is split over \mathbb{Q}_p . In particular we have for $p \in \mathcal{P}_{\text{good}}$ that $G_{\mathbb{Q}_p}$ is split, and therefore isomorphic to $\mathbf{PGL}_{3,\mathbb{Q}_p}$, by Lemma 6.4. This is also apparent from the fact that we may extend σ from E to \mathbb{Q}_p .

Lemma 8.7. *We have*

$$\sum_{\substack{p \in \mathcal{P}_{\text{good}} \\ p \leq x}} \log p = \frac{x}{[E : \mathbb{Q}]} + o(x),$$

where E is as above.

Proof. The first condition defining $\mathcal{P}_{\text{good}}$ does not influence the asymptotic, and modulo the second, neither does the third. Meaning, the statement is that the set of primes that split in E has natural density $1/[E : \mathbb{Q}]$. Because the splitting field E is Galois, this follows from Chebotarev's density theorem with natural density [23, Theorem 4]. \square

Proof of Proposition 7.10. Let $M \geq 2$. Fix a real number $c > 0$, which we will optimize later. Let $c_1 > 0$ be a real number that we will later assume to be sufficiently small. Let $M_1 = c_1 \log M$. When p is a prime, define

$$a_p = \begin{cases} \frac{c}{p} & \text{when } p \in \mathcal{P}_{\text{good}} \text{ and } p \leq M_1, \\ 0 & \text{otherwise.} \end{cases}$$

For $p \in \mathcal{P}_{\text{good}}$ we may define the elementary Hecke operators $\tau(\mathbf{a}, p) \in \mathcal{H}_p = C_c^\infty(K_p \backslash \mathbf{G}(\mathbb{Q}_p)/K_p)$ as in §8.1, through the isomorphism σ with $\mathbf{PGL}_3(\mathbb{Q}_p)$. Define $\mathbf{a} = (1, 0, 0)$ and for $p \in \mathcal{P}_{\text{good}}$ define

$$\omega_p = a_p \tau(\mathbf{a}, p) + a_p \tau(\mathbf{a}, p)^* \in \mathcal{H}_p^{\leq 2}.$$

Define

$$(49) \quad \omega = \sum_{\substack{n \leq M \\ \text{squarefree}}} \prod_{p|n} \omega_p \in \mathcal{H}_{\mathcal{P}_{\text{good}}, M}^{\leq 2},$$

and finally

$$(50) \quad k_f = \omega * \omega^* = \sum_{\substack{n, m \leq M \\ \text{squarefree}}} \prod_{\substack{p|n \\ q|m}} \omega_p * \omega_q^* = \sum_{\substack{n \leq M \\ \text{squarefree}}} \prod_{p|n} (\omega_p + \omega_p^* + \omega_p * \omega_p^*),$$

where the second equality holds by grouping pairs (n, m) with the same least common multiple. (In fact $\omega_p^* = \omega_p$.) Clearly $k_f \in \mathcal{H}_{S, M}^{\leq 4}$.

It is clear that k_f satisfies (46). Indeed, it is a self-convolution and therefore has nonnegative eigenvalues. It takes nonnegative values because the a_p are nonnegative, and $k_f(1) \geq 1$ thanks to the term for $n = m = 1$.

We have that $\|k_f\|_\infty \ll M^A$ for some $A > 0$. This can be seen either by expanding the convolution $\omega_p \omega_p^*$ in terms of elementary Hecke operators, or using the same arguments as used in [10, Lemma 4.4].

It remains to show the lower bound in Proposition 7.10. Because $\omega_p(1) = \omega_p^*(1) = 0$, we have

$$k_f(1) = \sum_{\substack{n \leq M \\ \text{squarefree}}} \prod_{p|n} \|\omega_p\|_2^2.$$

To prove the lower bound, we begin by trivially estimating

$$(51) \quad k_f(1) \leq \prod_{p \leq M_1} (1 + \|\omega_p\|_2^2)$$

by completing the sum over n to all square-free integers. To compute $\int_{H(\mathbb{A}_f)} k_f$ we integrate (50) and use the computations from Lemma 8.6, which give

$$\int_{H(\mathbb{A}_f)} k_f = \sum_{\substack{n \leq M \\ \text{squarefree}}} \prod_{p|n} (12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2),$$

where we use that $\|\omega_p\|_2^2 = 2a_p^2 \deg(\mu_p)$. The term $\|\omega_p\|_2^2$ corresponds to the diagonal contribution, and we will want the other terms to be large relative to this. Let $\alpha > 0$. We complete the sum over n to a full product, which introduces an error term that we estimate using Rankin's trick by introducing a factor $(n/M)^\alpha$ in the resulting error terms.

$$(52) \quad \begin{aligned} \int_{H(\mathbb{A}_f)} k_f &= \prod_{p \leq M_1} (1 + 12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2) \\ &\quad + O \left(M^{-\alpha} \sum_{\substack{n > M \\ \text{squarefree}}} \prod_{p|n} p^\alpha (12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2) \right) \\ &= \prod_{p \leq M_1} (1 + 12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2) \\ &\quad + O \left(M^{-\alpha} \prod_{p \leq M_1} (1 + p^\alpha (12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2)) \right) \end{aligned}$$

Applying the inequality $\frac{1+x}{1+y} \leq 1 + (x-y) \leq \exp(x-y)$ (for $x \geq y \geq 0$), the ratio of the error term to the main term in (52) is at most

$$(53) \quad M^{-\alpha} \exp \left(\sum_{p \leq M_1} (p^\alpha - 1)(12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2) \right).$$

We want this to be strictly less than 1. By the mean value theorem, $p^\alpha - 1 \leq \alpha p^\alpha \log p$. Using the bound $a_p \ll 1/p$, we can thus bound (53) by

$$\exp \left(-\alpha \log M + O \left(\sum_{p \leq M_1} \alpha p^\alpha \log p \right) \right).$$

Now choose $\alpha = 1/\log M_1$, so that by Chebyshev's estimates this is at most

$$\exp(-\alpha(\log M + O(M_1))).$$

By choosing c_1 sufficiently small, $M_1 = c_1 \log M$ is small enough for the above expression to be at most $1/2$ (say).

It remains to find a lower bound for the ratio of the main term in (52) to the right-hand side of (51). Using the bound $1 + x \geq \exp(x + O(x^2))$, we have

$$\begin{aligned} \frac{\int_{H(\mathbb{A}_f)} k_f}{k_f(1)} &\gg \prod_{p \leq M_1} \frac{1 + 12a_p + (6p + 24)a_p^2 + \|\omega_p\|_2^2}{1 + \|\omega_p\|_2^2} \\ &\gg \exp \left(\sum_{p \leq M_1} \left(\frac{12a_p + (6p + 24)a_p^2}{1 + \|\omega_p\|_2^2} + O(1/p^2) \right) \right) \\ &\gg \exp \left(\sum_{\substack{p \leq M_1 \\ p \in \mathcal{P}_{\text{good}}}} \left(\frac{12c + 6c^2}{1 + 2c^2} \cdot \frac{1}{p} + O(1/p^2) \right) \right) \\ &\gg \exp \left(\frac{1}{[E : \mathbb{Q}]} \cdot \frac{12c + 6c^2}{1 + 2c^2} (\log \log M_1)(1 + o(1)) \right). \end{aligned}$$

In the last step we have used Lemma 8.7 together with Abel's summation formula to deduce a Mertens type result from the PNT-type result. The rational function in c that appears is maximal for $c = 1$, where it takes the value 6. If we recall that $M_1 = c_1 \log M$, this is the lower bound stated in Proposition 7.10. \square

8.3. Upper bounds. We prove that the construction in §8.2 is in a sense optimal, by giving an upper bound for the quotient

$$\frac{\int_{H(\mathbb{A}_f)} k_f}{k_f(1)}$$

modulo certain restrictions on k_f . The main result is Proposition 8.12. When $p \in \mathcal{P}_{\text{good}}$ is a prime, we may identify $\mathbf{G}_{\mathbb{Q}_p}$ with $\mathbf{PGL}_{3, \mathbb{Q}_p}$ using the isomorphism σ . To simplify the notation, we will not keep track of the set $\mathcal{P}_{\text{good}}$ here, and state the global bounds instead for the group \mathbf{PGL}_3 , for which we define truncated Hecke algebras in the same way as in §6.7. In fact, we will give statements that are valid more generally for \mathbf{PGL}_n with $n \geq 3$. In either case, \mathbf{H} denotes the diagonal torus.

We will first prove the following key bound. While it is valid for all $n \geq 3$, it is likely not the strongest possible result, see Remark 8.13. We use the local notation from §8.1.

Proposition 8.8. *Let $n \geq 3$ and let $\mathbf{a}, \mathbf{b} \in \mathbb{N}^n$ be decreasing tuples with $a_n = b_n = 0$. Then*

$$\left(\int_{H_p} \tau(\mathbf{a}, p) * \tau(-\mathbf{b}, p) \right) - \delta_{\mathbf{a}, \mathbf{b}} \deg(\mu_{\mathbf{a}}(p)) \ll \frac{1}{p} \deg(\mu_{\mathbf{a}}(p))^{1/2} \deg(\mu_{\mathbf{b}}(p))^{1/2},$$

where the implicit constant is allowed to depend on \mathbf{a} and \mathbf{b} but not on p .

The proof of Proposition 8.8 uses various arguments, one of which is the following trivial bound.

Lemma 8.9. *The conclusion of Proposition 8.8 holds when $\deg(\mu_{\mathbf{b}}(p)) \geq p^2 \cdot \deg(\mu_{\mathbf{a}}(p))$.*

Proof. We may bound the integral using Lemma 8.5, by counting pairs (L_1, L) as in the statement of the lemma. First, because \mathbf{a} and \mathbf{b} may be assumed fixed, the number of possibilities for L_1 is bounded. To be precise, the inclusions $L_1 \supset L \supset p^{a_1} L_0$ and $p^{b_1} L_1 \subset L \subset L_0$ imply that there are at most $(a_1 + b_1 + 1)^n$ possibilities for L_1 . Second, the number of possibilities for L is trivially bounded by $\deg(\mu_{\mathbf{a}}(p))$. Therefore the left-hand side in Proposition 8.8 (even without the subtraction) is bounded up to a constant by

$$\deg(\mu_{\mathbf{a}}(p)) \leq \frac{1}{p} \deg(\mu_{\mathbf{a}}(p))^{1/2} \deg(\mu_{\mathbf{b}}(p))^{1/2}. \quad \square$$

To prove Proposition 8.8, it suffices to prove an upper bound of the form

$$\left(\int_{H_p} \tau(\mathbf{a}, p) * \tau(-\mathbf{b}, p) \right) - \delta_{\mathbf{a}, \mathbf{b}} \deg(\mu_{\mathbf{a}}(p)) \ll \frac{1}{p^2} \cdot \max(\deg(\mu_{\mathbf{a}}(p)), \deg(\mu_{\mathbf{b}}(p))).$$

Indeed, by symmetry and by Lemma 8.9 we must only consider the situation where

$$\deg(\mu_{\mathbf{a}}(p)) \leq \deg(\mu_{\mathbf{b}}(p)) \leq p^2 \deg(\mu_{\mathbf{a}}(p)),$$

in which case the above bound is at least as strong as what is needed.

The other type of argument we will use is the following. Let G be any (abstract) group and $f : X \rightarrow Y$ a G -equivariant map between finite transitive G -sets. Then the preimages $f^{-1}(y)$ have the same cardinality. In particular, when $S \subset X$, a bound of the form $|f(S)| \leq \delta|Y|$ implies $|S| \leq \delta|X|$. We will apply this principle with $G = \mathrm{GL}_n(\mathbb{Z}/p^a\mathbb{Z})$, X a set of subgroups of $(\mathbb{Z}/p^a\mathbb{Z})^n$, and f reduction mod p .

Lemma 8.10. *Let $\mathbf{a} \in \mathbb{N}^n$ be a decreasing tuple with $a_n = 0$. Let $1 \leq m \leq n$ and consider the subgroups $L \subset (\mathbb{Z}/p^{a_1}\mathbb{Z})^n$ with the following properties:*

- *The quotient $(\mathbb{Z}/p^{a_1}\mathbb{Z})^n/L$ has invariant factors given by \mathbf{a} .*
- *$L \subset (p\mathbb{Z}/p^{a_1}\mathbb{Z})^m \oplus (\mathbb{Z}/p^{a_1}\mathbb{Z})^{n-m}$.*

Let $d = \#\{i : a_i = 0\}$. Then the number of such subgroups is bounded up to a constant by $p^{-md} \deg(\mu_{\mathbf{a}}(p))$, where the constant does not depend on \mathbf{a} nor p .

Proof. The group $G = \mathrm{GL}_n(\mathbb{Z}/p^{a_1}\mathbb{Z})$ acts transitively on the subgroups L whose quotient has invariant factors given by \mathbf{a} . There are precisely $\deg(\mu_{\mathbf{a}}(p))$ of those. The reduction mod p of all such L is a subspace of \mathbb{F}_p^n of dimension d . The group G acts transitively on subspaces of given dimension in a way compatible with reduction mod p . There are $\asymp p^{d(n-d)}$ such subspaces. On the other hand, the subgroups L as in the statement have reduction mod p lying in a fixed $(n-m)$ -dimensional subspace. When $n-m < d$ there is nothing to prove. When $n-m \geq d$, the number

of such subspaces is $\asymp p^{d(n-m-d)}$. We conclude that the number of subgroups we want to count, is bounded up to a constant by

$$\frac{p^{d(n-m-d)}}{p^{d(n-d)}} \deg(\mu_{\mathbf{a}}(p)) = p^{-md} \deg(\mu_{\mathbf{a}}(p)). \quad \square$$

Finally, we will use the perfect pairing on $(\mathbb{Z}/p^a\mathbb{Z})^n$, which provides a notion of duality. Specifically, denote by $\langle \cdot, \cdot \rangle$ the component-wise pairing in the standard basis. When $L \subset (\mathbb{Z}/p^a\mathbb{Z})^n$, define $L^* = \{x : \langle x, L \rangle = 0\}$. Then $(L^*)^* = L$, duality reverses inclusions and if $(\mathbb{Z}/p^a\mathbb{Z})^n/L$ has invariant factors (p^{a_i}) then $(\mathbb{Z}/p^a\mathbb{Z})^n/L^*$ has invariant factors (p^{a-a_i}) . For an explicit example, the dual of $\bigoplus p^{a-a_i}\mathbb{Z}/p^a\mathbb{Z}$ is $\bigoplus p^{a_i}\mathbb{Z}/p^a\mathbb{Z}$.

We have the following dual version of Lemma 8.10.

Lemma 8.11. *Let $\mathbf{a} \in \mathbb{N}^n$ be a decreasing tuple with $a_n = 0$. Let $1 \leq m \leq n$ and consider the subgroups $L \subset (\mathbb{Z}/p^{a_1}\mathbb{Z})^n$ with the following properties:*

- *The quotient $(\mathbb{Z}/p^{a_1}\mathbb{Z})^n/L$ has invariant factors given by \mathbf{a} .*
- *$L \supset (p^{a_1-1}\mathbb{Z}/p^{a_1}\mathbb{Z})^m \oplus \{0\}^{n-m}$.*

Let $d = \#\{i : a_i = a_1\}$. Then the number of such subgroups is bounded up to a constant by $p^{-md} \deg(\mu_{\mathbf{a}}(p))$, where the constant does not depend on \mathbf{a} nor p .

Proof. For L as in the statement, we have that L^* satisfies the conditions in Lemma 8.10. The only observation we have to make is that the tuple $\mathbf{a}^* := (a_1 - a_n, \dots, a_1 - a_2, a_1 - a_1)$ satisfies $\deg(\mu_{\mathbf{a}^*}(p)) = \deg(\mu_{\mathbf{a}}(p))$; this is just the statement that $\deg(\mu_{-\mathbf{a}}(p)) = \deg(\mu_{\mathbf{a}}(p))$. \square

With these three ingredients we are ready to prove the key bound.

Proof of Proposition 8.8. By symmetry we may assume that $\deg(\mu_{\mathbf{a}}(p)) \leq \deg(\mu_{\mathbf{b}}(p))$. To bound the integral we use Lemma 8.5. We must bound the number of pairs (L_1, L) with L_1 adapted and different from L_0 , and $L \subset L_0 \cap L_1$ for which L_0/L has invariant factors given by \mathbf{a} and those of L_1/L are given by \mathbf{b} . Write $L_1 = \bigoplus p^{t_i}\mathbb{Z}_p$ with the $t_i \in \mathbb{Z}$. As in the proof of Lemma 8.9 there are only finitely many possibilities for L_1 , so we may assume L_1 is fixed.

Suppose there exists $t_i > 0$. Then up to permutation of factors, L lies in $p\mathbb{Z}_p \oplus \mathbb{Z}_p^{n-1}$, so that the subgroup $L/p^{a_1}L_0 \subset L_0/p^{a_1}L_0 \cong (\mathbb{Z}/p^{a_1}\mathbb{Z})^n$ satisfies the conditions of Lemma 8.10 with $d, m \geq 1$. Therefore there are at most (up to a constant) $p^{-1} \deg(\mu_{\mathbf{a}}(p))$ possibilities for L in this case, and this bound is good enough because $\deg(\mu_{\mathbf{a}}(p)) \leq \deg(\mu_{\mathbf{b}}(p))$.

We may now assume that all $t_i \leq 0$, so that $L_0 \subset L_1$. If two distinct $t_i < 0$, then we may apply Lemma 8.10 again, this time to $L/p^{b_1}L_1 \subset L_1/p^{b_1}L_1$ and with $m \geq 2, d \geq 1$. We find that there are at most $p^{-2} \deg(\mu_{\mathbf{b}}(p))$ possibilities for L in this case, and this bound is good enough as remarked below Lemma 8.9.

We may now assume that a single $t_i < 0$ and all others are 0. If $b_1 \leq a_1$, then L contains $p^{a_1+t_i}e_i$ and in particular $p^{a_1-1}e_i$. We may then apply Lemma 8.11 to $L/p^{a_1}L_0 \subset L_0/p^{a_1}L_0$, with $m, d \geq 1$ and conclude that there are at most $p^{-1} \deg(\mu_{\mathbf{a}}(p))$ possibilities for L . If $b_1 > a_1$, then L_0 contains all the elements $p^{b_1-1}e_j$ with $j \neq i$. Because $n \geq 3$, there are at least two of these. We may now apply Lemma 8.11 to $L/p^{b_1}L_1 \subset L_1/p^{b_1}L_1$ with $m \geq 2$ and $d \geq 1$, and conclude that there are at most (up to a constant) $p^{-2} \deg(\mu_{\mathbf{b}}(p))$ possibilities for L in this case. Again, this is sufficient by the comment below Lemma 8.9. \square

Let \mathcal{P} denote the set of all primes.

Proposition 8.12. *Let $n \geq 3$, $\kappa > 0$ and $M \geq 3$. For every nonzero $\omega \in \mathcal{H}_{\mathcal{P}, M}^{\leq \kappa}$ the convolution operator $k_f = \omega * \omega^*$ satisfies*

$$\frac{\int_{\mathbf{H}(\mathbb{A}_f)} k_f}{k_f(1)} \ll (\log \log M)^C,$$

for a constant C that is allowed to depend on n and κ .

Proof. We may find a finite family of squarefree integers $(n_i)_{i \in I} \in [0, M]$ and for every prime $p \mid n_i$ a scalar multiple of elementary Hecke operator $\omega_{i,p} \in \mathcal{H}_p^{\leq \kappa}$, such that $\omega = \sum_{i \in I} \prod_{p \mid n_i} \omega_{i,p}$. Moreover, we may assume that the $\prod_{p \mid n_i} \omega_{i,p}$ have disjoint supports. If X_κ denotes the set of cocharacters of $\mathbf{H}_{\overline{\mathbb{Q}}}$ of norm at most κ , then $|X_\kappa|$ is bounded. If $\omega(m)$ denotes the number of prime divisors of m , then every $m \leq M$ occurs at most $|X_\kappa|^{\omega(m)}$ times as an integer n_i in the family.

Let C be the largest implicit constant in Proposition 8.8 when \mathbf{a} and \mathbf{b} run through the tuples with $\mu_{\mathbf{a}}, \mu_{\mathbf{b}} \in X_\kappa$.

When $p \nmid n_i$, define $\omega_{i,p} = 1_{K_p}$. We then have

$$k_f = \sum_{i,j \in I} \prod_{\substack{p \mid n_i \\ q \mid n_j}} \omega_{i,p} * \omega_{j,q}^* = \sum_{i,j \in I} \prod_{p \mid n_i n_j} \omega_{i,p} * \omega_{j,p}^*.$$

The disjointness of the supports of the $\prod_{p \mid n_i} \omega_{i,p}$ implies

$$k_f(1) = \sum_{i \in I} \prod_{p \mid n_i} \|\omega_{i,p}\|_2^2.$$

Using Proposition 8.8 we have

$$\begin{aligned} \int_{\mathbf{H}(\mathbb{A}_f)} k_f &= \sum_{i,j \in I} \prod_{p \mid n_i n_j} \int_{H_p} (\omega_{i,p} * \omega_{j,p}^*) \\ &\leq \sum_{i,j \in I} \prod_{p \mid n_i n_j} C p^{-1} \|\omega_{i,p}\|_2 \|\omega_{j,p}\|_2 \\ &= \sum_{i,j \in I} \text{lcm}(n_i, n_j)^{-1} C^{\omega(\text{lcm}(n_i n_j))} \prod_{p \mid n_i n_j} \|\omega_{i,p}\|_2 \|\omega_{j,p}\|_2 \\ &\leq \sum_{d \leq M} \frac{d}{C^{\omega(d)}} \left(\sum_{\substack{d \mid n_i \\ p \mid n_i}} \frac{C^{\omega(n_i)}}{n_i} \prod_{p \mid n_i} \|\omega_{i,p}\|_2 \right)^2, \end{aligned}$$

where in the last equality we have written $\text{lcm}(n_i n_j) = n_i n_j / d$ with $d = \gcd(n_i, n_j)$, and then extended the sum to run over all $d \mid n_i, n_j$. By Cauchy–Schwarz, this is bounded by

$$\sum_{d \leq M} \frac{d}{C^{\omega(d)}} \left(\sum_{\substack{d \mid n_i \\ p \mid n_i}} \frac{C^{2\omega(n_i)}}{n_i^2} \right) \left(\sum_{\substack{d \mid n_i \\ p \mid n_i}} \prod_{p \mid n_i} \|\omega_{i,p}\|_2^2 \right).$$

It follows that

$$\frac{\int_{\mathbf{H}(\mathbb{A}_f)} k_f}{k_f(1)} \leq \sup_{\substack{n \leq M \\ \square\text{-free} \\ d \mid n, n_i}} \sum_{\substack{d, n_i \\ d \mid n, n_i}} \frac{d}{C^{\omega(d)}} \frac{C^{2\omega(n_i)}}{n_i^2}.$$

Because every m occurs at most $|X_\kappa|^{\omega(m)}$ times in the family $(n_i)_{i \in I}$, this is at most

$$\begin{aligned} &\ll_{\kappa, \epsilon} \sup_{\substack{n \leq M \\ \square\text{-free}}} \sum_{\substack{d|n, m \\ m \square\text{-free}}} \frac{d}{C^{\omega(d)}} \frac{(C^2 |X_\kappa|)^{\omega(m)}}{m^2} \\ &= \sup_{\substack{n \leq M \\ \square\text{-free}}} \sum_{d|n} \frac{(C |X_\kappa|)^{\omega(d)}}{d} \sum_{m \square\text{-free}} \frac{(C^2 |X_\kappa|)^{\omega(m)}}{m^2} \\ &\ll \sup_{\substack{n \leq M \\ \square\text{-free}}} \sum_{d|n} \frac{(C |X_\kappa|)^{\omega(d)}}{d}, \end{aligned}$$

because the sum over m is convergent. The latter expression is largest when n has the smallest possible prime factors. So take $n = \prod_{p \leq x} p$ for some $x > 1$. Then $n \leq M$ implies $x \ll \log M$ by Chebyshev's estimate, and we have

$$\begin{aligned} \sum_{d|n} \frac{(C |X_\kappa|)^{\omega(d)}}{d} &= \prod_{p \leq x} \left(1 + \frac{C |X_\kappa|}{p} \right) \\ &\ll (\log x)^{C |X_\kappa|} \\ &\ll (\log \log M)^{C |X_\kappa|}, \end{aligned}$$

where we have used the Mertens' theorem in the first estimate. \square

Remark 8.13. It is likely that a stronger version of Proposition 8.8 is still true. Namely, when $n \geq 4$ we expect that a similar bound holds with the power of p in the right-hand side replaced by $p^{-3/2}$. Moreover, when $n = 3$ we expect that this can also be shown, except in the situation of Lemma 8.6. We have partial proofs of these statements, which use mostly the same arguments as in the proof of Proposition 8.8, and contains lots of casework. However a few cases remain, where we notably require bounds for specific Hall polynomials. We hope to settle this stronger version in the near future.

The stronger version would imply that the upper bound in Proposition 8.12 can be replaced by 1 (no growth at all) for $n \geq 4$. Indeed, the exponent $3/2$ is then propagated throughout the proof until the very last lines, where we may then use that $\prod_{p \leq X} (1 + p^{-3/2})$ is bounded.

Remark 8.14. When $n = 3$, Proposition 8.8 is not as strong as we would like in different ways. It would be desirable to remove or make explicit the dependence on κ and to make the power of $\log \log M$ match with the exponent in Proposition 7.10. Such bounds should follow from the stronger version of Proposition 8.8 that we expect to hold.

Acknowledgements. I am very grateful to Farrell Brumley for introducing me to the circle of questions that led us to study toric periods for \mathbf{PGL}_n , and for informing me about the period relation that put the result into context. I also thank Simon Marshall for spotting an error in an earlier version of the manuscript, and thank him together with Jasmin Matz for their comments after a careful reading.

REFERENCES

- [1] C. Aistleitner, K. Mahatab, and M. Munsch. Extreme Values of the Riemann Zeta Function on the 1-Line. *Int. Math. Res. Not.*, 2019(22):6924–6932, November 2019.

- [2] M. V. Berry. Regular and irregular semiclassical wavefunctions. *J. Phys. A*, 10(12):2083–2091, 1977.
- [3] V. Blomer and F. Brumley. Simultaneous equidistribution of toric periods and fractional moments of L -functions. preprint at <https://arxiv.org/abs/2009.07093>.
- [4] V. Blomer and J. Buttcane. On the subconvexity problem for L -functions on $GL(3)$. *Ann. Sci. Éc. Norm. Supér.*, 53(6), 2015.
- [5] V. Blomer, É. Fouvry, E. Kowalski, Ph. Michel, D. Milićević, and W. Sawin. *The second moment theory of families of L -functions*, pages 93–119. Memoirs of the American Mathematical Society. American Mathematical Society, to appear.
- [6] A. Bondarenko and K. Seip. Large greatest common divisor sums and extreme values of the Riemann zeta function. *Duke Math. J.*, 166(9):1685–1701, 2017.
- [7] A. Borel. *Linear algebraic groups*, volume 126 of *Graduate Texts in Mathematics*. Springer-Verlag, second enlarged edition, 1991.
- [8] A. Borel and Harish-Chandra. Arithmetic subgroups of algebraic groups. *Ann. of Math.*, 75(3):485–535, 1962.
- [9] A. Borel and J. Tits. *Groupes réductifs*. Publications mathématiques de l’I.H.É.S., 1965.
- [10] F. Brumley and S. Marshall. Lower bounds for Maass forms on semisimple groups. *Compos. Math.*, 156(5):959–1003, 2020.
- [11] Y. Colin de Verdière. Spectre du laplacien et longueur des géodésiques périodiques. II. *Compos. Math.*, 27(2):159–184, 1973.
- [12] R. de la Bretèche and G. Tenenbaum. Gál sums and applications (in French). *Proc. Lond. Math. Soc. (3)*, 119(3):104–134, 2019.
- [13] H. Donnelly. Exceptional sequences of eigenfunctions for hyperbolic manifolds. *Proc. Amer. Math. Soc.*, 135(5):1551–1555, 2007.
- [14] J. Duistermaat, J. Kolk, and V. Varadarajan. Spectra of compact locally symmetric manifolds of negative curvature. *Invent. Math.*, 52:27–93, 1979.
- [15] J. Duistermaat, J. Kolk, and V. Varadarajan. Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups. *Compos. Math.*, 49(3):309–398, 1983.
- [16] P. B. Eberlein. *Geometry of Nonpositively Curved Manifolds*. Chicago Lectures in Mathematics. The University of Chicago Press, 1996.
- [17] M. Einsiedler, E. Lindenstrauss, Ph. Michel, and A. Venkatesh. Distribution of periodic torus orbits on homogeneous spaces. *Duke Math. J.*, 148(1):119–174, 2009.
- [18] D. W. Farmer, S. M. Gonek, and C. P. Hughes. The maximum size of L -functions. *J. Reine Angew. Math.*, 609:215–236, August 2007.
- [19] R. Gangolli. On the Plancherel formula and the Paley-Wiener theorem for spherical functions on semisimple Lie groups. *Ann. of Math.*, 93(1):150–165, 1971.
- [20] S. Garibaldi and Ph. Gille. Algebraic groups with few subgroups. *J. London Math. Soc.*, 80(2):405–430, 2009.
- [21] A. Granville and K. Soundararajan. Extreme values of $|\zeta(1 + it)|$. In *The Riemann zeta function and related themes: papers in honour of Professor K. Ramachandra*, volume 2 of *Ramanujan Math. Soc. Lect. Notes Ser.*, pages 65–80, 2006.
- [22] Harish-Chandra. Harmonic analysis on real reductive groups I: the theory of the constant term. *J. Funct. Anal.*, 19:104–204, 1975.
- [23] H. Heilbronn. Zeta-Functions and L-Functions. In J. W. W. Cassels and A. Fröhlich, editors, *Algebraic Number Theory*. Academic Press, 1967.
- [24] S. Helgason. *Differential geometry, Lie groups and symmetric spaces*. Academic Press, 1978.
- [25] S. Helgason. *Groups and geometric analysis*. Academic Press, 1984.
- [26] H. Iwaniec and P. Sarnak. L^∞ norms of eigenfunctions on arithmetic surfaces. *Ann. of Math. (2)*, 141:301–320, 1995.
- [27] H. Iwaniec and P. Sarnak. Perspectives on the analytic theory of L -functions. *Geom. Funct. Anal.*, Special volume:705–741, 2000.
- [28] A. W. Knap. *Representation Theory of Semisimple Group*. Princeton University Press, 1989.
- [29] A. W. Knap. *Lie Groups Beyond an Introduction*. Birkhauser, 2 edition, 2002.
- [30] M.-A. Knus, A. Merkurjev, M. Rost, and J.-P. Tignol. *The Book of Involutions*. Number 44 in Colloquium publications. American Mathematical Society, 1998.
- [31] E. Lapid and O. Offen. Compact unitary periods. *Compos. Math.*, 143:323–338, 2007.

- [32] J. M. Lee. *Introduction to Smooth Manifolds*. Number 218 in Graduate Texts in Mathematics. Springer, second edition, 2013.
- [33] N. Levinson. Ω -theorems for the Riemann zeta-function. *Acta Arith.*, 20:317–330, 1972.
- [34] S. Marshall. Geodesic restrictions of arithmetic eigenfunctions. *Duke Math. J.*, 165(3), 2016.
- [35] S. Marshall. L^p norms of higher rank eigenfunctions and bounds for spherical functions. *J. Eur. Math. Soc.*, 18(7):1437–1493, 2016.
- [36] J. Mercer. Functions of positive and negative type, and their connection the theory of integral equations. *Phil. Trans. Roy. Soc. London (A)*, 209:415–446, 1909.
- [37] B. Michels. Extreme values of geodesic periods on arithmetic hyperbolic surfaces. *J. Inst. Math. Jussieu*, pages 1–36, 2020.
- [38] B. Michels. The geometry of maximal flat submanifolds of symmetric spaces. preprint, 2022.
- [39] D. Milićević. Large values of eigenfunctions on arithmetic hyperbolic surfaces. *Duke Math. J.*, 155(2):365–401, 2010.
- [40] D. Milićević. Large values of eigenfunctions on arithmetic hyperbolic 3-manifolds. *Geom. Funct. Anal.*, 21(6):1375–1418, December 2011.
- [41] H. L. Montgomery. Extreme values of the Riemann zeta function. *Comment. Math. Helv.*, 52:511–518, 1977.
- [42] D. W. Morris. *Introduction to arithmetic groups*. Deductive Press, 2015.
- [43] L. Nachbin. *The Haar Integral*. D. Van Nostrand Company, 1965.
- [44] Ł. Pańkowski and J. Steuding. Extreme values of L -functions from the Selberg-class. *Int. J. Number Theory*, 9(5):1113–1124, 2013.
- [45] R. S. Pierce. *Associative Algebras*. Number 88 in Graduate Texts in Mathematics. Springer-Verlag, 1982.
- [46] V. Platonov and A. Rapinchuk. *Algebraic Groups and Number Theory*. Academic Press, 1994.
- [47] Z. Rudnick and P. Sarnak. The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Commun. Math. Phys.*, 161:195–213, 1994.
- [48] P. Sarnak. Reciprocal geodesics. *Clay Math. Proc.*, 7:217–237, 2007.
- [49] K. Soundararajan. Extreme values of zeta and L -functions. *Math. Ann.*, 342:467–486, 2008.
- [50] E. M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Princeton Mathematical Series. Princeton University Press, 1993.
- [51] T. Tamagawa. On Selberg’s trace formula. *J. Fac. Sci. Univ. Tokyo*, pages 363–386, 1960.
- [52] J. Tits. Reductive groups over local fields. In *Automorphic forms, representations, and L -functions*, 1979.
- [53] J.-L. Waldspurger. On the values of certain automorphic L -functions and their center of symmetry (in French). *Compos. Math.*, 54:173–242, 1985.
- [54] D. Werner. *Funktionalanalysis*. Springer Spektrum, 8. auflage edition, 2018.
- [55] H. Whitney. *Complex analytic varieties*. Addison-Wesley, 1972.
- [56] S. Zelditch. Kuznecov sum formulae and Szegő limit formulae on manifolds. *Comm. Partial Differential Equations.*, 17(1–2):221–260, 1992.