ANNALS OF MATHEMATICS

Sparse equidistribution problems, period bounds and subconvexity

By Akshay Venkatesh

SECOND SERIES, VOL. 172, NO. 2
September, 2010

ANMAAH

Annals of Mathematics, **172** (2010), 989–1094

Sparse equidistribution problems, period bounds and subconvexity

By AKSHAY VENKATESH

Abstract

We introduce a "geometric" method to bound periods of automorphic forms. The key features of this method are the use of equidistribution results in place of mean value theorems, and the systematic use of mixing and the spectral gap. Applications are given to equidistribution of sparse subsets of horocycles and to equidistribution of CM points; to subconvexity of the triple product period in the level aspect over number fields, which implies subconvexity for certain standard and Rankin-Selberg L-functions; and to bounding Fourier coefficients of automorphic forms.

Contents

1. **Introduction**

1.1. *General introduction.* Let $\Gamma \subset G$ be a lattice in an S-arithmetic group. Let $Y \subset \Gamma\backslash G$ be a subset endowed with a probability measure ν, and f a function on $\Gamma\backslash G$. Fixing a basis $\{\psi_j^{(Y)}\}$ for $L^2(Y, \nu)$, we shall refer to the numbers

$\int f\psi_j^{(Y)}\,d\nu$, as the *periods* of f along Y. Evidently, the periods depend heavily on the choice of basis for $L^2(Y,\nu)$. They play a major role in the theory of automorphic forms, in significant part because they often express information about L-functions.

The present paper is centered around a geometric method yielding upper bounds for these periods. It is applicable, roughly speaking, when considering the periods of a fixed function f along a sequence of subsets (Y_i,ν_i), with the property that the Y_i are becoming equidistributed; that is to say, the ν_i approach weakly the G-invariant measure on $\Gamma\backslash G$. The key inputs of this method are, firstly, the *equidistribution* of the ν_i, and secondly, the *mixing* properties of certain auxiliary flows. More precisely, we shall need these properties in a quantitative form; in the cases we consider, this will follow eventually from an appropriate spectral gap.

This situation might seem rather restrictive. However, it arises often in many natural equidistribution questions ("sparse equidistribution problems," as we discuss below) as well as in the analytic theory of automorphic forms (especially, subconvexity results for L-functions). There are applications besides those discussed in the present paper; our aim has not been to give an exhaustive discussion, but rather just to present a representative sample of interesting cases. We shall explain the method abstractly in Section 1.3 and will carry out, in the body of the paper, one example of each of the following cases: Y_i is the orbit of a unipotent, a semisimple, and a toral subgroup of G.

In the present paper, we have focused mostly on the case of PGL_2 and GL_2 over number fields. All our results pertain to this setting, except for Theorem 3.2, which applies to a general semisimple group. The geometric methods of this paper are general and we hope to analyze further higher rank examples in a future paper.

Throughout the present methods we have tried to use "soft" techniques as a substitute for explicit spectral expansions. However, there still seem to be instances where the explicit spectral expansions are important. In a future paper [28], joint with P. Michel, we shall combine ideas drawn from this paper with ideas from Michel's paper [27]; in that paper, we shall make much more explicit use of spectral decomposition.[1]

We shall use the term "sparse equidistribution problems" to describe questions of the following flavor: Suppose $Z_i \subset Y_i$ is a subset endowed with a measure ν_i^Z, and we would like to prove that the ν_i^Z are becoming equidistributed. In other words, we wish to deduce the equidistribution of the "sparse" subset Z_i from the known equidistribution of Y_i. Examples of this type of question are Shah's conjecture [38, §1, end] (where the Z_i are discrete subsets of Y_i, a full horocycle orbit) as well as results of Michel, Harcos-Michel on subsets of Heegner points [27], [17] (where the Z_i are subsets of the Y_i, the set of all Heegner points). The connection to period integrals is as follows: one can spectrally expand the measure ν_i^Z in

[1] Added in proof: the article [29] contains a further discussion of this idea.

terms of the basis for $L^2(Y_i, \nu_i)$. Using our results for periods along Y_i, it will sometimes be possible to deduce the equidistribution of ν_i^Z.

We now briefly summarize our results.

(1) Section 3 considers where the Y_is are orbits, or pieces of orbits, of unipotent groups. The mixing flow is the horocycle flow along Y_i.

In Theorem 3.1 (p. 1015) we show that certain *sparse subsets* of horocycles on compact quotients of $\mathrm{SL}_2(\mathbb{R})$ become equidistributed. This is progress towards a conjecture of N. Shah. In Theorem 3.2 (p. 1019) we give a fairly general bound (in the context of an arbitrary semisimple group) on the Fourier coefficients of automorphic forms. In the case of $G = \mathrm{SL}_2(\mathbb{R})$ it recovers results of Good [16] and Sarnak [37], which resolved a problem of Selberg. The present proof is more direct, avoiding in particular the triple product bounds for eigenfunctions.

(2) Section 4 considers the case when $G = \mathrm{PGL}_2(F \otimes \mathbb{R})$, where F is a number field, and Γ is a congruence subgroup thereof. The Y_i are a sequence of closed diagonal G-orbits on $\Gamma\backslash G \times \Gamma\backslash G$. The mixing flow (after lifting to the adeles) is the diagonal action of $\mathrm{PGL}_2(\mathbb{A}_{F,f})$, where $\mathbb{A}_{F,f}$ is the ring of finite adeles of F.

Propositions 4.1 and 4.2 give period bounds in this context. Proposition 4.1 yields subconvexity for the triple product L-function, in the level aspect as one factor varies (this is conditional on a property of p-adic integrals, Hypothesis 11.1; this has apparently been established in the time since the paper was submitted, see the remark following Hypothesis 11.1). In Theorem 5.1 (p. 1028) it is shown that these results yield subconvex bounds, in the level aspect, for standard and Rankin-Selberg L-functions attached to PGL_2.

The results on standard and Rankin-Selberg L-functions generalize results of Duke-Friedlander-Iwaniec [12] and Kowalski-Michel-Vanderkam [24] from the case $F = \mathbb{Q}$.[2] The third result, concerning subconvexity of the triple product period in the level aspect, was not known even over $\mathbb{Q}$; however, Bernstein and Reznikov [3] have shown subconvexity for the triple product period in the *eigenvalue aspect*.

(3) Section 6 considers the case when Y_i is a certain family of noncompact torus orbits on $\Gamma\backslash G$, where (Γ, G) is as in Section 4. (In fact, the Y_i are obtained by taking a fixed noncompact torus orbit, and translating by a p-adic unipotent, where p varies.) The mixing flow is the action of the *adelic* points of the torus.

---

[2]We have not attempted to address the issue of varying the central character. This, in a sense, is the most subtle point, as is shown by Michel's recent work on Rankin-Selberg convolutions. Our aim in the present paper has been to show that one can derive a coherent theory for PGL_2 from the triple product bound of Proposition 4.1. The case of varying central character will be discussed in a future paper with Michel.

We establish in Theorem 6.1 (p. 1034) subconvexity for character twists of GL(2) in the level aspect. This was first established for $F = \mathbb{Q}$ by Duke-Friedlander-Iwaniec [11], and the special case where F is totally real and the form holomorphic at all infinite places was treated by Cogdell, Piatetski-Shapiro and Sarnak. In particular, (6.2) gives a subconvex bound for Grössencharacter L-functions over F, in the level aspect; this was known over $\mathbb{Q}$ by work of Burgess [6] and some special cases were known in the general case, e.g., [41].

(4) In Section 7 we consider the case where Y_i is a (union of) compact torus orbits on $\Gamma\backslash G$, where (Γ, G) are as in Section 4. The equidistribution of such Y_i will amount to the equidistribution of Heegner points, and we deduce it from Theorem 6.1 in Theorem 7.1 (p. 1042). This result generalizes work of Duke over $\mathbb{Q}$ and was proven, conditionally on GRH, by Zhang [47], Cohen [9], and Clozel-Ullmo [8] (independently). The present work makes this result unconditional.

Applying mixing properties of the adelic torus flow, we obtain in Theorem 7.2 (p. 1043), under a condition of splitting of enough small primes, the equidistribution of certain sparse subsets of Heegner points. In the case $F = \mathbb{Q}$, an *unconditional* result of this nature is due to Michel and Harcos-Michel.[3]

In the context of L-functions, one pleasing feature of the present method is that it is *geometric*: it proceeds not via Fourier coefficients but via the integral representation. In practice, this means that there is no difference between Maass or holomorphic forms, nor between $\mathbb{Q}$ and an arbitrary base field. Moreover, we do not make use of either the trace formula or the Kuznetsov formula; indeed, we make no explicit use of families.

The recent work of Bernstein-Reznikov [3] is of a similar flavor. They establish a "subconvex" bound for the triple product when the eigenvalue of one factor varies, whereas we have treated the case where the level of one factor varies. Their method is also geometric in nature, and moreover their result applies to a nonarithmetic group. By contrast, the level aspect question is not well-posed if one leaves the arithmetic setting.

Throughout the paper we have not attempted to optimize the results. The input to our method is an equidistribution result. As far as possible we have tried to establish these results by relatively "geometric" methods, deriving in the end from the mixing properties of a certain flow. Of course, it is in many contexts better to use spectral methods, but this would involve departing from the geometric method

[3] Our method is different to those mentioned: we do not deduce our result from results on Rankin-Selberg convolutions, and indeed it is possible to deduce a subconvexity result from ours. However, there seem to be some curious parallels between the methods. In fact, the method of Theorem 7.2 is even more closely related — as Michel has pointed out to me — to the work [13] of Duke, Friedlander and Iwaniec. In that paper they amplify class group L-functions but obtain only a conditional result for precisely the same reason that Theorem 7.2 fails to be unconditional, namely, one cannot guarantee unconditionally the existence of enough small split primes.

that is intended to be the central theme of this paper. As remarked, we will pursue such "spectral" approaches in a forthcoming paper with P. Michel [28]; some of the results of this have been discussed in [29].

Finally, implicit in various parts of the paper is "adelic analysis", i.e. the analytic theory of functions on adelic quotients, in the quantitative sense needed for analytic number theory. There seems to be considerable scope to develop this theory fully.

1.2. *Other applications.* The method of this paper has other applications not elaborated here. We discuss some of them here.

There are other subconvexity results that are naturally approached by the same method: for instance, a subconvex estimate for $L(\pi, 1/2+it)$ where t varies and π is a fixed cuspidal representation of GL(2) over a number field F. In such a context it is natural to use the fact that the horocycle flow is (quantifiably) weakly k-mixing, for certain $k > 1$; the use of this higher order mixing is closely related to Weyl's "successive squaring" approach to $\zeta(1/2+it)$. Of course, this particular instance of subconvexity is approachable by standard methods also; an intriguing question in the subconvexity context is how to combine the present methods with those such as Bernstein-Reznikov.

There are certain applications to effective equidistribution theorems: for instance, it is perhaps possible to establish some new effective cases of Ratner's theorem by the same ideas. The question of giving such "nontrivial" cases was raised by Margulis in his talk at the American Institute of Mathematics, June 2004. Unfortunately, the cases to which our method might apply are very artificial.

One can give certain analytic applications: let Γ be a cocompact subgroup of $\mathrm{SL}(2,\mathbb{R})$, and let $\pi \subset L^2(\Gamma\backslash\mathrm{SL}(2,\mathbb{R}))$ be an irreducible $\mathrm{SL}(2,\mathbb{R})$-subrepresentation. For $m \in \mathbb{Z}$, let e_m be the mth weight vector in π, if defined; i.e., a vector which transforms under the character $\left(\begin{smallmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{smallmatrix}\right) \mapsto e^{2\pi i m\theta}$. We normalize it (up to a complex scalar of absolute value 1) by requiring that $\|e_m\|_{L^2} = 1$. Bernstein and Reznikov proved the bound $\|e_m\|_{L^\infty} \ll (1+|m|)^{1/2}$, and asked [2, Remark 2.5(4)] if any improvement of the exponent $1/2$ is possible. It is quite easy to deduce from Lemma 3.1 such a bound; indeed, the analytic properties of the e_m, as $|m| \to \infty$, is connected with the long time behavior of the horocycle flow in the same fashion that the analytic behavior of Laplacian eigenfunctions are connected to the long time behavior of the geodesic flow. In the time during which this paper was being revised for submission, Reznikov has proven independently a result of this type [34]. Since the result he obtains is most likely sharper than that obtained by the technique indicated above, we will not pursue this further, noting only that an advantage of the method we have indicated above is that it is likely to generalize to higher rank.

Moving slightly away from the main subject of the present paper, the idea of using equidistribution theorems to produce mean value results for L-functions seems capable of application in a variety of settings. In particular, equidistribution results are readily available on $\mathrm{GL}(n)$, owing to Ratner's work, whereas trace formulae are extremely unwieldy for $n > 2$. It would be interesting to see what mean-value statements can be deduced from Ratner-type equidistribution results.

Historically, one application of such results has been to nonvanishing results; here the most spectacular results (e.g., [42]) have been achieved through the so-called mollifier technique. It would be quite interesting to understand if there is a geometric interpretation of the mollifier technique.

1.3. *Discussion of method*: *equidistribution*, *mixing*, *and periods.* We now turn to a discussion of the specifics of the method used in this paper. This method itself is quite easy to describe. It consists in essence of two simple steps (see (1.2) and (1.3) below).

We also remark that the discussion that follows is a relatively faithful rendition of the method of the paper. The body of this paper does not really utilize any new ideas beyond the ones indicated below. Most of the bulk consists of the technical details necessary to connect periods with other objects of interest (e.g., equidistribution questions or L-functions), as well as setting up the machinery to quantify some standard equidistribution results. As much as possible, we have tried to give a self-contained treatment of all these technical details in Sections 8–11.

We hope the ensuing discussion serves as a unifying thread for the rest of the paper. We explain the method first in an abstract setting (§1.3.1). We then explain (§§1.3.2 and 1.3.3) these ideas in a more down-to-earth fashion, emphasizing the parallel with the analytic techniques for studying L-functions. Finally, Section 1.3.4 illustrates these ideas in a simple example – that of Fourier coefficients of modular forms.

1.3.1. *Abstract setting.* Let $G_2 \subset G_1$ be locally compact groups, $\Gamma \subset G_1$ a lattice, $X = \Gamma\backslash G_1$. Let $x_i \in X$ and put $Y_i = x_i G_2$. We shall suppose that there exists a G_2-invariant probability measure ν_i on Y_i. (This does not precisely cover all the contexts we consider — at some points we will consider Y_i which are "long pieces" of a G_2-orbit rather than a single G_2-orbit, but the ideas in that case will be identical to those discussed here.)

Let f be a function on X and ψ_i a function on Y_i such that $\int_{Y_i} |\psi_i|^2 d\nu_i = 1$. We will give a bound for the period $\int_{Y_i} f \psi_i d\nu_i$.

In words, the idea will be to find certain correlations between the values of ψ_i at different points; and then show that the values of f at these same points are "uncorrelated," in some quantifiable sense. Putting these together will show that the period must be small. The "hard" ingredient here is some version of the spectral gap, i.e., quantitative mixing, which is what will show the "uncorrelatedness" property of f.

We will suppose that there exists σ, a measure on G_2, such that

$$\psi_i \star \sigma = \lambda_i \psi_i, \tag{1.1}$$

for some $\lambda_i \in \mathbb{C}$. Here $\star\sigma$ denotes the action of σ by right convolution. Let $\check{\sigma}$ be the image of σ by the involution $g \mapsto g^{-1}$ of G_2. Then

$$(1.2)\quad \left|\int f \cdot \psi_i \, d\nu_i\right|^2 = \left|\lambda_i^{-1}\int_{Y_i} f \cdot (\psi_i \star \sigma)\, d\nu_i\right|^2 = \left|\lambda_i^{-1}\int_{Y_i} (f \star \check{\sigma}) \cdot \psi_i \, d\nu_i\right|^2 \leq |\lambda_i|^{-2}\int_{Y_i} |f \star \check{\sigma}|^2 d\nu_i,$$

where we have applied Cauchy-Schwarz at the final step. Now, we are assuming that the Y_i are becoming equidistributed, and so $\nu_i \to \nu$, the G_1 invariant measure on $\Gamma\backslash G_1$. Thus

$$(1.3)\quad \int_{Y_i} |f \star \check{\sigma}|^2 d\nu_i \approx \int_X |f \star \check{\sigma}|^2 d\nu = \int_{g,g' \in G_2} \langle g g'^{-1} \cdot f, f\rangle_{L^2(X)} \, d\sigma(g) \, d\sigma(g'),$$

where $gg'^{-1} \cdot f$ denotes the right translate of f by gg'^{-1}.

If the G_2-action on X is mixing in a quantifiable way — i.e., one has strong bounds on the decay of matrix coefficients — one obtains good upper bounds on the right-hand side of (1.3); in combination with (1.2) this gives an upper bound for the period $|\int_{Y_i} f \psi_i \, d\nu_i|$.

The strength of the information required about the mixing varies. In the cases we study where G_2 is amenable, any nontrivial information will suffice. In the one case where G_2 is semisimple, a strong bound towards Ramanujan is needed. For instance, in the case of triple products, we need any improvement of the bound that the pth Hecke eigenvalue of a cusp form on GL(2) is bounded in absolute value by $p^{1/4} + p^{-1/4}$. (In this normalization, the trivial bound is $p^{1/2} + p^{-1/2}$.)

In the rest of this paper, we shall merely apply this argument many times, with various different choices for Γ, G_1, G_2. The part of the argument which will vary is quantifying the equidistribution of the ν_i, i.e. keeping track of the error in the first approximation of (1.3). Thus we make heavy use of Sobolev norms (§8), which are an efficient method of bounding this error.

In each instance, the proof of the equidistribution result $\nu_i \to \nu$ will always be rather straightforward, *except* for the result of Section 7. The equidistribution result needed for the proof of Theorem 7.2 is essentially equivalent to the subconvexity result proved in Section 6. A rather striking point is that a similar logical dependence (although manifested very differently) is present in the work of Michel. The meaning of this is unclear to the author.

In certain specific cases, the above technique is quite familiar. When G_2 is a one-parameter real group, the above argument is quite closely related to standard techniques of analytic number theory.[4] On the other hand, when G_2 is an adelic

[4]For example, in certain contexts when G_2 is abelian, one can push this method further by squaring multiple times, that is to say, considering $|\int f \psi_i \, d\nu_i|^4$, $|\int f \psi_i \, d\nu_i|^8$ and so forth. In this

group, and σ a measure on G_2 that corresponds to the action of Hecke operators (this is carried out in §4, for instance), the above argument will be essentially "amplification" in the sense of Friedlander-Iwaniec [15].

In the following two sections, we shall attempt to explain more colloquially the main idea that is at work here, and also discuss how the method described above fits into the framework of analytic number theory. Modern proofs of subconvexity, following the path-breaking work of Friedlander-Iwaniec [15], have roughly speaking consisted of a mean-value theorem and an amplification step. We shall discuss how the proof indicated above may be viewed as geometrizing this strategy, where the mean-value step is replaced by an equidistribution theorem, and the amplification step is controlled using mixing.

Note, in particular, that in the work of Friedlander-Iwaniec, families of L-functions play a central role, whereas the method above has in a certain sense eliminated the family. Although in the discussion below we rephrase matters so as to make clear the connection with the work of Friedlander-Iwaniec, it seems that from the perspective of the present paper the phrasing in terms of families is rather artificial.

1.3.2. *Connection with analytic number theory*: *equidistribution*, *and mean-value theorem for periods.* Follow the notation of the previous section. We choose an orthonormal basis $\{\psi_{i,j}\}_{j=1}^{\infty}$ for $L^2(Y_i, \nu_i)$ so that $\psi_{i,1} := \psi_i$.

By Plancherel's formula, $\sum_{j=1}^{\infty} \left|\int f\psi_{i,j}\, d\nu_i\right|^2 = \int |f|^2 d\nu_i$. Since $\nu_i \to \nu$ weakly, and we are holding f fixed, it follows that

$$\sum_{j=1}^{\infty} \left| \int f\psi_{i,j}\, d\nu_i \right|^2 \to \int_{\Gamma\backslash G} |f|^2 d\nu, \tag{1.4}$$

as $i \to \infty$. Thus the equidistribution property of ν_i underlies a mean-value theorem for the Y_i-periods.

In many cases involving automorphic forms, the periods will essentially be special values of L-functions and (1.4) amounts to a mean-value theorem for L-functions. This is fairly well-known; for example, the mean-value theorem

$$\int_{-T}^{T} |\zeta(1/2+it)|^4 dt \sim T \log(T)^4$$

is rather closely connected with the equidistribution properties of the cycle

$$\{(1+i/T)x, x \in \mathbb{R}\},$$

when projected to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. A more striking example is Vatsal's use of equidistribution to prove nonvanishing results [45]. In general, it seems that there are many interesting mean value theorems for L-functions that are connected to equidistribution results.

context, one replaces the mixing property of the G_2 action with results about higher order mixing of the G_2-flow. Although we will not carry this out in the present paper, this seems rather closely connected to Weyl's proof of subconvexity for $\zeta(1/2+it)$.

In any case, (1.4) is not unrelated to the standard methods of obtaining such results; however, its primary advantage is that it is often technically much simpler, for example when working over a number field.

1.3.3. *Connection with analytic number theory* (II): *mixing*, *and bounds for a single period.* We now wish to pass from (1.4) to nontrivial upper bounds for a single period. It is clear that (1.4) implies at once — by omitting all terms but one — that $|\int f\psi_{i,j}\,d\nu_i| \lesssim \|f\|_{L^2(X)}$; we shall refer to an improvement of this bound as *nontrivial*. It is evident that one must have some further information about $\{\psi_{i,j}\}$ in order to do this; otherwise one could simply take $\psi_{i,1}$ to be a multiple of $f|_{Y_i}$.

In the context of analytic number theory, this is often carried out by "shortening the family," that is to say: proving a sharp mean-value theorem of the form of (1.4), but over some subfamily of $\{\psi_{i,j}\}_{j=1}^{\infty}$; then omitting all terms but $\psi_{i,1} = \psi_i$ will often give a nontrivial upper bound. In the work of Friedlander-Iwaniec, a weighted mean-value theorem is derived, which has the same effect as shortening the family.

Such a weighted mean-value theorem is also implicit in our context. Following the notation of Section 1.3.1, suppose that there is a *fixed* measure σ on G_2 such that for all i, j, we have $\psi_{i,j} \star \sigma = \lambda_{i,j}\psi_{i,j}$ (some $\lambda_{i,j} \in \mathbb{C}$). Then, by Plancherel's formula, and using the fact $\nu_i \to \nu$, we conclude:

$$\sum_{j=1}^{\infty} |\lambda_{i,j}|^2 \left|\int f \cdot \psi_{i,j}\,d\nu_i\right|^2 \to \langle f \star \check{\sigma}, f \star \check{\sigma}\rangle_{L^2(X)}. \tag{1.5}$$

This gives a weighted mean value theorem, which for appropriate choices of σ amounts to shortening the effective range of summation in (1.4). Moreover, the mixing of the G_2-flow bounds the right-hand side of (1.5). In this phrasing, it becomes clear that the measure σ has played the role of an "amplifier" and the orthonormal basis for $L^2(Y_i, \nu_i)$ has played the role of the family.

Having now explained the method in an abstract context and indicated its equivalence with other methods, we now indicate more informally the source of cancellation in periods that is at the center of our results.

In many natural situations, one obtains a basis for $L^2(Y_i, \nu_i)$ by diagonalizing a geometrically defined algebra of operators on Y_i. The result of this process is that the functions $\{\psi_j\}$ exhibit correlations between their values at different points of Y (= the relevant Y_i). For instance (for example when G_2 is semisimple), it often will occur that there is a correspondence $\mathscr{C} : Y \mapsto Y$ such the value of each ψ_j at $P \in Y$ and at the collection of points $\mathscr{C}(P)$ are correlated in some way. On the other hand (and we shall now speak quite imprecisely), if the correspondence $\mathscr{C}$ "extends" to a correspondence $\widetilde{\mathscr{C}} : X \mapsto X$, then one can often show, using mixing properties of $\widetilde{\mathscr{C}}$, that the values of f at P and $\widetilde{\mathscr{C}}(P)$ will be *uncorrelated*, at least if P is chosen at random with respect to the uniform measure on X.

However, *since the Y_i are becoming equidistributed*, it amounts to almost the same thing to choose P at random with respect to ν_i and with respect to the uniform measure on X. Thus, for ν_i-typical $P \in Y_i$, the values of f at P and $\mathscr{C}(P)$ are uncorrelated, whereas the values of ψ_j at P and $\mathscr{C}(P)$ are correlated. One can then play these phenomena against each other to obtain cancellation in the period integral $\int f \psi_j d\nu_i$.

1.3.4. *A concrete example.* We shall now discuss how to bound Fourier coefficients of a modular form by the methods just described. Although the material below is essentially redone - with $\mathrm{SL}(2,\mathbb{R})$ replaced by a general group — in Section 3, the example below was very important in motivating the author's intuition, and it seems worthwhile to include it in the introduction.

Let $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ be a lattice containing the element $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Let $f(z)$ be a holomorphic form of weight 2 with respect to Γ, which we write in a Fourier expansion $f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$. Hecke proved the bound $|a_n| \le Cn$, a bound which was only improved (for a general — possibly nonarithmetic — Γ) much later, to $|a_n| \le Cn^{5/6}$, by A. Good [16]. We shall sketch a simple proof of a nontrivial bound $|a_n| \le Cn^{1-\delta}$ along the lines just indicated; for further details, we refer the reader to Section 3.2, where the procedure outlined is implemented for a general semisimple group.

We note that the ideas that will enter here are exactly those that will enter into the proof of equidistribution of sparse subsets of horocycles (see §3.1), or for the nontrivial bound for L^∞-norms in the weight aspect that is discussed in Section 1.2. The proof below also works for Maass forms (in that case the result is due to Sarnak [37]).

The Fourier expansion implies that

$$a_n = e^{2\pi} \int_{x \in \mathbb{R}/\mathbb{Z}} f\left(x + \frac{i}{n}\right) e^{-2\pi i n x} dx. \tag{1.6}$$

In words, the idea is as follows: the function $e^{-2\pi i n x}$ takes the same values at $x, x + \frac{1}{n}, x + \frac{2}{n}, \dots$. On the other hand, the values of the function f at these points are (in a quantifiable sense) uncorrelated, as we shall deduce from the mixing properties of the horocycle flow. Playing these two properties against each other will yield an improvement of the Hecke bound for $|a_n$.[5]

Let $\tilde{f}$ be the lift of f to $\Gamma\backslash\mathrm{SL}(2,\mathbb{R})$; that is to say,

$$\tilde{f} : \Gamma \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto f\Big(\frac{ai+b}{ci+d}\Big)(ci+d)^{-2}.$$

Let

$$x_n = \Gamma \begin{pmatrix} n^{-1/2} & 0 \\ 0 & n^{1/2} \end{pmatrix},$$

[5]Underlying this is the usual "van der Corput" trick: to bound $\sum_{k=1}^{K} c_k$ it suffices to bound correlations $\sum_{k=1}^{K} c_k c_{k+h}$; in effect we apply this with $c_k = f(\frac{k+i}{n})$, $K = n$.)

and put $n(t) = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)$. Then the definitions show that $\tilde{f}(x_n n(t)) = n^{-1} f\left(\frac{i+t}{n}\right)$; consequently, we see that

$$a_n = e^{2\pi} \int_{t=0}^{n} \tilde{f}(x_n n(t)) e^{-2\pi i t} \, dt. \tag{1.7}$$

Equation (1.7) expresses the nth Fourier coefficient of f as the integral of $\tilde{f}$ over a closed horocycle of length n. Moreover, (1.7) falls into the pattern of Section 1.3.1, with $G_1 = \mathrm{SL}_2(\mathbb{R})$, $G_2 = \{n(t) : t \in \mathbb{R}\}$, $Y_n = \{x_n n(t) : t \in \mathbb{R}\}$, and $\psi_n : Y_n \to \mathbb{C}$ the function given by $x_n n(t) \mapsto e^{-2\pi i t}$. The fact that the Y_n are becoming equidistributed amounts to the "equidistribution of low horocycles"; cf. [35]. In the language of Section 1.3.1, we will take σ to be the measure on $G_2 \cong \mathbb{R}$ that is a sum of point masses δ_i, for integers $i = 0, \ldots, K-1$. We now carry out the procedure of Section 1.3.1 in an explicit fashion in the paragraphs that follow.

Let T be the operation of right translation by $n(1)$ on $C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$: that is to say, for $F \in C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$, we put $TF(g) = F(gn(1))$.

The value of the right-hand side of (1.7) remains unchanged if we replace $\tilde{f}$ by $T\tilde{f}$; consequently, for any integer $K \geq 1$, we have

$$a_n = \frac{e^{2\pi}}{K} \int_{t=0}^{n} \left(\sum_{i=0}^{K-1} T^i \tilde{f}(x_n n(t) \right) e^{-2\pi i t} \, dt.$$

Applying the Cauchy-Schwarz inequality we deduce that

$$|a_n|^2 \leq \frac{n}{e^{-4\pi} K^2} \int_{t=0}^{n} \left| \sum_{i=0}^{K-1} T^i \tilde{f}(x_n n(t)) \right|^2 dt. \tag{1.8}$$

We now use come to the equidistribution part of the argument. The *equidistribution of long closed horocycles* asserts that the closed horocycle $\{x + iy : 0 \leq x \leq 1\}$ becomes equidistributed in $\Gamma\backslash\mathbb{H}$ as $y \to 0$. Quantitatively, for any $F \in C^\infty(\Gamma\backslash\mathbb{H})$, we have

$$\left| \int_0^1 F(x+iy) dx - \int_{\Gamma\backslash\mathbb{H}} F \frac{dx\,dy}{y^2} \right| \leq C_F y^\delta, \tag{1.9}$$

for some C_F depending on F, and some δ depending only on Γ. This assertion, originally proved by Sarnak [35] by spectral methods, can be deduced quite easily from the mixing properties of the geodesic flow; this is done, in a somewhat more general context, in Lemma 9.7.

We note that — a special case of the discussion in Section 1.3.2 — the equidistribution statement (1.9) above reflects a mean-value theorem for periods. Indeed, if one applies it to $F = y^2|f|^2$, one deduces the asymptotic for $\sum_{n<X} |a_n|^2$.

In any case, what will be more useful is the version of (1.9) that is lifted to $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$. This asserts that for any $F \in C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$, we have

$$\left| \frac{1}{n} \int_{t=0}^{n} F(x_n n(t)) dt - \int_{\Gamma\backslash\mathrm{SL}_2(\mathbb{R})} F(g) dg \right| \leq C_F n^{-\delta}, \tag{1.10}$$

where δ is an explicit constant depending only on Γ, and C_F is a constant depending on F. From (1.8) and (1.10) we conclude that

$$|a_n|^2 \le \frac{e^{4\pi}n^2}{K^2}\left(\left\| \sum_{i=0}^{K-1} T^i \tilde{f} \right\|^2_{L^2(\Gamma\backslash \mathrm{SL}_2(\mathbb{R}))} + C_{f,K}n^{-\delta}\right). \tag{1.11}$$

On the other hand, the explicit derivation of (1.10) shows that C_F may be bounded by a Sobolev norm of F, and consequently the constant $C_{f,K}$ that appears in (1.11) is bounded by $O_f(K^A)$ for some $A > 0$. Thus

$$|a_n|^2 \ll_f n^2K^{-2}\left(\left\| \sum_{i=0}^{K-1} T^i \tilde{f} \right\|^2_{L^2(\Gamma\backslash \mathrm{SL}_2(\mathbb{R}))} + K^An^{-\delta}\right).$$

We now use the fact that the horocycle flow is mixing, in a quantifiable way; see [32]. This amounts to the assertion that there is an explicit $\delta' > 0$ and constant C'_f such that, for $i \in \mathbb{Z}$, $|\langle T^i \tilde{f}, \tilde{f}\rangle| \ll C'_f(1+|i|)^{-\delta'}$. We may assume without loss that $\delta' < 1$; this being so, it follows:

$$\left\| \sum_{i=0}^{K-1} T^i \tilde{f} \right\|^2_{L^2} \ll_f K^{2-\delta'}.$$

We conclude that $|a_n| \ll n(K^{-\delta'/2} + K^{A/2-1}n^{-\delta/2})$. Taking K to be a sufficiently small power of n, we conclude that a_n is bounded by $n^{1-\delta''}$ for some $\delta'' > 0$ depending only on Γ.

Clearly δ'' depends only on the spectral gap of $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$. Of course, this dependence does not arise in the "spectral" methods. It can be removed in the above method, but this seems to require some extra input, e.g., the finite-dimensionality of the space of functionals on an irreducible $\mathrm{SL}_2(\mathbb{R})$-representation that are invariant under the subgroup $\{n(t) : t \in \mathbb{R}\}$.

1.3.5. *Two other viewpoints on the method of Section* 1.3.4. There are two other viewpoints which might be helpful in thinking about the previous section. Both of these viewpoints do not literally generalize to the other situations we consider (e.g., triple products) but may be helpful for intuition.

(1) The first is based on the following simple principle: suppose that T is a measure-preserving transformation of the probability space (Y, ν), and that T is *ergodic*. If μ_1, μ_2 are two T-invariant probability measures with average $\frac{\mu_1+\mu_2}{2} = \nu$, then $\mu_1 = \mu_2 = \nu$; this follows because ν is an *extreme point* of the convex set of T-invariant probability measures. More generally, given any family of probability measures averaging to ν, they must almost all *equal* ν.

   We will apply this to $Y = \mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R})$ and T the operation of translation by $n(1)$.

   Let n be large; for $t \in \mathbb{R}/\mathbb{Z}$, let μ_t be the probability measure that corresponds to normalized counting measure on $\{x_n n(t+k) : k \in \mathbb{Z}, 0 \le k < n\}$. Here notation is as prior to (1.7).

Then $\int_0^1 \mu_t$ is the measure on the closed horocycle $\{x_n n(t) : 0 \leq t \leq n\}$. Thus the family of measures μ_t averages to the measure on a long closed horocycle which, as we remarked earlier (see (1.10)) approximates the $\mathrm{SL}_2(\mathbb{R})$-invariant measure dg on $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$. But this latter measure dg is ergodic with respect to T. So applying a more quantitative form of the principle discussed above shows that, for almost all $t \in [0, 1]$, μ_t must be close to dg. It is simple to see that one can use this to deduce bounds for the Fourier coefficients, via (1.7).

(2) We will phrase the second rather imprecisely. Consider (1.7). Our strategy of proof can be rephrased as: The function $t \mapsto \widetilde{f}(x_n n(t))$ is weak-mixing, whereas the function $t \mapsto e^{2\pi i t}$ is periodic, and a weak-mixing function cannot correlate with a periodic function.

To explain this statement, we need to explain what it means for a function on the real line to be *weak-mixing*. Consider instead the case of a function $h : \mathbb{Z} \to \mathbb{R}$. Furstenberg's correspondence principle asserts that one can (loosely speaking) associate to this a dynamical system (Y, ν, T) in such a way that (again loosely speaking) h arises by sampling a function on Y along a generic trajectory $y_0, T(y_0), T^2(y_0), \ldots$. We then say that h is weak-mixing if the system (Y, ν, T) is so. The fact that our function $t \mapsto \widetilde{f}(x_n n(t))$ (say, when restricted to integer times) is weak-mixing follows from the equidistribution of long closed horocycles together with the fact that the horocycle flow is, itself, mixing.

For more on this point of view, see e.g., [44, §4] and [44, Lemma 5.2] for a version of the statement that weak-mixing functions cannot correlate with periodic ones.

1.4. *Connection to existing methods.* The following comments pertain to the results of the present paper that concern subconvex bounds for L-functions. As we have emphasized above, the methods presented here are, upon examination, seen to be closely related to existing methods: in particular, "Sarnak's spectral method," which gave the only hitherto known instance of subconvexity over a base other than $\mathbb{Q}$.

Indeed, as we have already indicated, the equidistribution step of our method can be seen as the geometric version of a mean value theorem, and the rest of the method can be seen as an amplification step (or "shortening the family"). Nevertheless, the key features of the present method are that it is essentially geometric (in that it avoids Fourier coefficients) and adelic (which allows us to import much from the modern theory of automorphic forms); it also does not use families in any explicit way. Once the notation is established — admittedly a nontrivial overhead — the method allows for very considerable technical simplification.

It is perhaps also noteworthy that the method given here does not require any exponential decay information for triple products. Although such exponential decay information is central only to subconvexity in the eigenvalue aspect, it has thus far entered as a technical device even in treatments of the level aspect.

In a sense, the present method bears the same relation to existing methods as adelic methods do to classical methods in the theory of automorphic forms. The classical situation has the advantage of concreteness, and whatever can be carried out in the adelic setting can be (in principle) carried out in the classical setting. However there is often a considerable technical and conceptual advantage in working adelically.

As we have discussed, the connection between equidistribution results and mean-value theorems for periods — implicitly exploited throughout this paper — appears in the work of V. Vatsal.

D. Hejhal considered ideas similar to that of Section 3.2 in the context of proving bounds towards Fourier coefficients; see [18]. In the language of this paper, his method used a measure σ (notation of §1.3) with much larger support, and consequently he was unable to get unconditional results.

Finally, as was remarked in Section 1.1, the main result of Section 4.1 is the analogue in the level aspect of a recent result of Bernstein-Reznikov [3]: they establish a "subconvex bound" on triple products as the eigenvalue of one factor varies. Their methods also are geometric, avoiding the use of Fourier coefficients.

1.5. *Acknowledgements.* This paper grew out of my proof of Theorem 3.1. The original proof was significantly more complicated, and I am indebted to Elon Lindenstrauss for his insistence that Theorem 3.1 should amount to nothing more than equidistribution and mixing. It was thus his intuition that led to a simplification of the proof and an important step in my understanding. The idea that the methods for Theorem 3.1 might be applicable in a more general setting arose during conversations with Andreas Strömbergsson, who also made many valuable suggestions about an early version of this paper. I thank them both for their significant contributions.

I am very grateful to Gergely Harcos and Philippe Michel for their encouragement of this project. Philippe read carefully an early draft of this paper and pointed out many points where the argument and results could be significantly improved. I am also grateful to Peter Sarnak, from whom I learned much of what I know about this subject.

I have also benefited from several conversations with Joseph Bernstein and André Reznikov. I thank them for their generosity in sharing and discussing their elegant ideas.

I have learned many of the methods that appear here from the work of others. I mention in particular Peter Sarnak's paper [36], which uses the idea of changing the test vector; the Friedlander-Iwaniec idea of amplification and the geometric version of it that appears in Bourgain-Lindenstrauss [5]; and the recent work of Bernstein-Reznikov [1], in particular their elegant use of Sobolev norms.

This paper suffered a considerable delay before submission. I would like to thank Philippe Michel for his encouragement and insistence that it be revised and submitted, without which the delay would have likely been considerably longer. I also thank Nicolas Bergeron and Marina Ratner for comments that improved the exposition and correctness.

The ideas of this paper were worked out during the workshop "Emerging applications of measure rigidity," AIM, San Francisco and at the Isaac Newton Institute.

I was supported by the Clay Mathematics Institute during much of the writing of the paper, and I thank them for their generous support. I also thank the Institute for Advanced Study for providing excellent working conditions during the academic year 2005-2006. I was also partially supported by NSF grants DMS-0111298 and DMS-0245606.

Last of all, I would like to thank the referees; in particular, one of the reports was extremely detailed and improved the paper very substantively. I am very grateful for the time they spent on it.

1.6. *Structure of paper.* The logical structure of this paper is as follows: Section 2 introduces all necessary notation. The heart of the paper are Section 3 (unipotent periods), Section 4 (the triple product period), Sections 6 and 7 (torus periods). The remaining Sections 8–11 are of a technical nature, proving various technical results required in the main text; at a first reading (or even later) they should perhaps be referred to only as necessary.

The two examples that best convey the flavor of the paper are Theorem 3.1 and Proposition 4.1. The proofs of these results are relatively self-contained, and we advise that the reader start with them.

2. Notation

2.1. *General notation.* We use the symbol $\ll$ as is standard in analytic number theory: namely, $A \ll B$ means that there exists a constant c such that $A \leq cB$. The notation $A \ll_{f,g,h} B$ means that the constant c may depend on the quantities f, g, h; the notation $A \ll_{\epsilon} B$ or $A \ll_{\varepsilon} B$ will mean, unless otherwise indicated, that the stated bound holds for all ϵ or $\varepsilon > 0$. In general, we will never explicate the dependence of implicit constants on the number field over which we work; and, by an abuse of terminology, we will sometimes use the phrase "absolute constant" to mean a constant that depends only on this number field.

If Z is a space we denote by δ_z the point measure at $z \in Z$, i.e., $\delta_z(f) = f(z)$ for f a continuous function on Z.

Now let Z be a right G-space. For f a function on Z and $g \in G$, we write $g \cdot f$ for the right translate of f by g, i.e., $g \cdot f(z) = f(zg)$. If μ is a measure on Z — we allow signed or complex-valued measures in what follows — we define the translate $g \cdot \mu$ by the rule $g \cdot \mu(g \cdot f) = \mu(f)$. In particular, if $\mu = \delta_z$ is the point mass at $z \in Z$, then $g \cdot \mu = \delta_{zg^{-1}}$ is the point mass at zg^{-1}.

If σ is a compactly supported measure on G, we set $f \star \sigma \stackrel{\text{def}}{=} \int_g (g \cdot f) d\sigma(g)$, i.e., $f \star \sigma(z) = \int_{g \in G} f(zg) d\sigma(g)$. In particular, if δ_{g_0} is the point-mass at g_0, then $f \star \delta_{g_0} = g_0 \cdot f$ is the right translate of f by g_0.

If σ_1, σ_2 are two compactly supported measures on G, we define the convolution $\sigma_1 \star \sigma_2$ to be the pushforward to G of $\sigma_1 \times \sigma_2$ on $G \times G$, under the multiplication map $(g_1, g_2) \in G \times G \mapsto g_1 g_2$. Notation as above, one has the (somewhat unfortunate) compatibility relation $(f \star \sigma_2) \star \sigma_1 = f \star (\sigma_1 \star \sigma_2)$.

For σ a measure on a group G, we denote by $\check{\sigma}$ the image of σ by the involution $g \mapsto g^{-1}$, and by $\|\sigma\|$ the total variation of σ. If G is a Lie group, then we denote by $\mathrm{Ad}(g)$ the endomorphism "$X \to gXg^{-1}$" of its Lie algebra. If $B \subset A$ is a finite index subgroup of the group A, then we denote by $[A : B]$ the index of B in A.

If h is an entire function, then the notation $\int_{\Re(s)=\sigma} h(s) ds$ denotes the line integral along the line $\Re(s) = \sigma$ from $\sigma - i\infty$ to $\sigma + i\infty$. The notation $\int_{\Re(s) \gg 1} h(s) ds$ denotes $\int_{\Re(s)=\sigma} h(s) ds$ for sufficiently large σ; in the contexts where we use this notation, the answer will be constant when σ is sufficiently large.

2.2. *Classical modular forms.* As usual $\mathbb{H}$ denotes the upper half-plane, i.e., $\{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. It admits the usual action of $\mathrm{SL}(2, \mathbb{R})$ by fractional linear transformations.

2.3. *Number fields and associated notation.* Let F be a number field. Throughout the paper we shall regard F as fixed: that is to say, we allow implicit constants in $\ll, \gg$ to depend on F without explicit statement.

We set $F_\infty = F \otimes \mathbb{R}$, $\mathbb{A}_F$ the ring of adeles of F, $\mathbb{A}_{F,f}$ the ring of finite adeles. Thus $\mathbb{A}_F = F_\infty \times \mathbb{A}_{F,f}$. We will fix *once and for all* an additive character $e_F : \mathbb{A}_F / F \to \mathbb{C}$, and denote by e_{F_v} the induced additive character of F_v.

For each place v we have a canonical "absolute value" $x \mapsto |x|_v$ on $F_v^\times$, namely, $|x|_v = \mathrm{meas}(xS)/\mathrm{meas}(S)$ for any Haar measure, meas, on $F_v^\times$, and any subset S of positive measure.

The same definition defines a character $\mathbb{A}_F^\times / F^\times \to \mathbb{R}_{>0}$, which we denote by $a \mapsto |a|_{\mathbb{A}}$, or simply by $a \mapsto |a|$ if it is clear from context. We denote by $\mathbb{A}_F^1$ the subgroup of $\mathbb{A}_F^\times$ consisting of adeles of norm 1; then the quotient $\mathbb{A}_F^1 / F^\times$ is compact. For a finite place v of F, we denote by $\mathfrak{o}_{F_v}$ the maximal compact subring of the completion F_v, by $\mathfrak{q}_v$ the maximal ideal of $\mathfrak{o}_{F_v}$, and by q_v the cardinality of the residue field.

We shall generally denote ideals of $\mathfrak{o}_F$ by gothic letters $\mathfrak{l}, \mathfrak{q}, \mathfrak{n}$, etc. If $\mathfrak{f}$ is an integral ideal of $\mathfrak{o}_F$, we set $\mathrm{N}(\mathfrak{f}) := |\mathfrak{o}_F / \mathfrak{f}|$ to be its norm. Moreover, we shall denote $\mathfrak{o}_\mathfrak{f} := \prod_{\mathfrak{q} | \mathfrak{f}} \mathfrak{o}_\mathfrak{q}$. Here $\mathfrak{o}_\mathfrak{q}$ denotes the completed ring, not the localized ring, i.e., $\mathfrak{o}_\mathfrak{f}$ is the inverse limit of the rings $\mathfrak{o}_F / \mathfrak{f}^N$.

We denote by $\mathfrak{d}$ the different of the character e_F, i.e. $\mathfrak{d}$ is a fractional ideal so that $\mathfrak{d}_v^{-1}$ is, for every finite place v, the largest $\mathfrak{o}_{F_v}$-submodule of F_v upon which e_F is trivial.

2.4. *Adele groups and their function spaces.* Let $\mathbf{G}$ be a connected reductive algebraic group over a number field F, and let $\mathbf{Z}$ be its center. Denote by $\mathbb{A}_{F,f}$ the ring of finite adeles, and fix for each finite place v a maximal open compact subgroup $K_{v,\mathbf{G}} \subset \mathbf{G}(F_v)$ with the property that $K_{\max,\mathbf{G}} := \prod_{v \text{ finite}} K_{v,\mathbf{G}}$ is a maximal open compact subgroup of $\mathbf{G}(\mathbb{A}_{F,f})$. Put $\mathbf{X}_\mathbf{G} = \mathbf{G}(F)\backslash\mathbf{G}(\mathbb{A}_F)$, $\mathbf{X}_{\mathbf{G},\mathrm{ad}} = \mathbf{Z}(\mathbb{A}_F)\mathbf{G}(F)\backslash\mathbf{G}(\mathbb{A}_F)$. Then $\mathbf{X}_{\mathbf{G},\mathrm{ad}}$ has finite volume with respect to any $\mathbf{G}(\mathbb{A}_F)$-invariant measure.

Let $\omega : \mathbf{Z}(\mathbb{A}_F) \to \mathbb{C}^\times$ be a unitary character. We define the space $C^\infty_\omega(\mathbf{X}_\mathbf{G})$ to be the space of functions on $\mathbf{X}_\mathbf{G}$ whose stabilizer in $K_{\max,\mathbf{G}}$ has finite index, which transform under $\mathbf{Z}(\mathbb{A}_F)$ by ω, and so that the function $g \mapsto f(xg)$ is a C^∞ function of $g \in \mathbf{G}(F_\infty)$, for each $x \in \mathbf{X}_\mathbf{G}$. Similarly one defines an L^2-space $L^2_\omega(\mathbf{X}_\mathbf{G})$, or simply L^2 if the central character ω is clear from context, by completing the space of compactly supported functions in $C^\infty_\omega(\mathbf{X}_\mathbf{G})$ with respect to the Hilbert norm $\|f\|_2 := \left(\int_{\mathbf{X}_{\mathbf{G},\mathrm{ad}}} |f(g)|^2 dg\right)^{1/2}$.

For $\psi \in C^\infty_\omega(\mathbf{X}_\mathbf{G})$, we denote by $K_{v,\psi}$ the stabilizer of ψ in $K_{v,\mathbf{G}}$, and put

$$K_\psi = \prod_{v \text{ finite}} K_{v,\psi}. \tag{2.1}$$

We note that K_ψ is, in general, a proper subgroup of the stabilizer of ψ in $K_{\max,\mathbf{G}}$.

For $\psi \in C^\infty_\omega(\mathbf{X}_\mathbf{G})$, we define the finite set of places $\mathrm{Supp}(\psi)$ to be those finite v for which $K_{v,\mathbf{G}}$ does not fix ψ, i.e.

$$\mathrm{Supp}(\psi) \overset{\mathrm{def}}{=} \{v : K_{v,\mathbf{G}} \neq K_{v,\psi}\}. \tag{2.2}$$

It is convenient to introduce some notions of "size" on $\mathbf{G}(\mathbb{A}_F)$. Let $\mathfrak{g}$ be the Lie algebra of $\mathbf{G}(F_\infty)$. It is a finite dimensional real vector space; fix an arbitrary norm on it. For $g_\infty \in \mathbf{G}(F_\infty)$, we denote by $\|g_\infty\|$ the operator norm of the adjoint endomorphism $\mathrm{Ad}(g_\infty^{-1}) : \mathfrak{g} \to \mathfrak{g}$. If v is a finite place of F and $g_v \in \mathbf{G}(F_v)$, then we set $\|g_v\| = [K_{v,\mathbf{G}} g_v K_{v,\mathbf{G}} : K_{v,\mathbf{G}}]$, i.e. the number of right- $K_{v,\mathbf{G}}$ cosets in $K_{v,\mathbf{G}} g_v K_{v,\mathbf{G}}$. For $g_f = (g_v)_{v \text{ finite}} \in \mathbf{G}(\mathbb{A}_{F,f})$ we put $\|g_f\| = \prod_v \|g_v\|$. Finally, for $g_\mathbb{A} = (g_\infty, g_f) \in \mathbf{G}(F_\infty) \times \mathbf{G}(\mathbb{A}_{F,f})$, set $\|g_\mathbb{A}\| = \|g_\infty\| \cdot \|g_f\|$.

We remark that $\|g_\infty\|, \|g_f\|, \|g_\mathbb{A}\|$ are all invariant by the center of $\mathbf{G}$.

2.5. *The groups* $\mathbf{G} = \mathrm{GL}(2)$ *and* $\mathbf{G} = \mathrm{PGL}(2)$ *and some of their subgroups.* We will deal most often with the cases of $\mathbf{G} = \mathrm{GL}(2)$ (resp. $\mathbf{G} = \mathrm{PGL}(2)$). In that setting we shall write $\mathbf{X}_{\mathrm{GL}(2)}$ (resp. $\mathbf{X}$) for $\mathbf{X}_\mathbf{G}$.

We will make use of the following algebraic subgroups of GL_2, which we will often also regard as algebraic subgroups of PGL_2 in the obvious way:

$$N = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}, \quad A = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}, \quad Z = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}.$$

If R is any ring and $x \in R, y \in R^{\times}$, then we denote[6]

$$n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \quad \bar{n}(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}, \quad a(y) = \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix},$$
$$a'(y) = \begin{pmatrix} 1 & 0 \\ 0 & y \end{pmatrix}, \quad z(y) = \begin{pmatrix} y & 0 \\ 0 & y \end{pmatrix}, w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.3}$$

all elements of $\mathrm{GL}_2(R)$.

If v is a place of F and $x \in F_v, y \in F_v^{\times}$, then we denote by $n_v(x)$ (resp. $a_v(y)$) the element $n(x)$ (resp. $a(y)$) considered as an element of $\mathrm{GL}_2(\mathbb{A}_F)$ via the natural inclusion $\mathrm{GL}_2(F_v) \hookrightarrow \mathrm{GL}_2(\mathbb{A}_F)$.

For each place v, we let K_v be the standard maximal compact subgroup of $\mathrm{GL}_2(F_v)$, i.e., K_v is the stabilizer of the norm on F_v^2 given by

$$\sqrt{|x|_v^{2/\deg(v)} + |y|_v^{2/\deg(v)}}$$

if v is archimedean, where $\deg(v)$ is the degree[7] of F_v over $\mathbb{R}$; and $\max(|x_v|, |y_v|)$ if v is nonarchimedean. Thus, in particular, $K_v = \mathrm{GL}_2(\mathfrak{o}_{F_v})$ if v is nonarchimedean. We put $K_{\max} = \prod_{v \text{ finite}} K_v$. K_v (respectively $K_{\max}$) is a maximal compact subgroup of $\mathrm{GL}_2(F_v)$ (respectively $\mathrm{GL}_2(\mathbb{A}_{F,f})$), and (by projection) can also be regarded as a maximal compact subgroup of $\mathrm{PGL}_2(F_v)$ (respectively $\mathrm{PGL}_2(\mathbb{A}_{F,f})$). Similarly $K_{\max} \times K_{\infty}$ is a maximal compact subgroup of $\mathrm{GL}_2(\mathbb{A}_F)$, and may also be regarded as a maximal compact subgroup of $\mathrm{PGL}_2(\mathbb{A}_F)$.

For $\mathfrak{q}$ a finite prime of F, we denote by $\varpi_{\mathfrak{q}} \in F_{\mathfrak{q}}$ a uniformizer, and by $[\varpi_{\mathfrak{q}}]$ the element of $\mathbb{A}_F^{\times}$ that is the image of $\varpi_{\mathfrak{q}}$ under the natural inclusion $F_{\mathfrak{q}}^{\times} \hookrightarrow \mathbb{A}_F^{\times}$.

Let $\mathfrak{q}$ be a finite prime of F. It will be convenient to define certain open compact subgroups of $K_{\mathfrak{q}}$. For each $e_{\mathfrak{q}} > 0$, we define $K[\mathfrak{q}^{e_{\mathfrak{q}}}] \subset K_{\mathfrak{q}}$ (resp. $K_0[\mathfrak{q}^{e_{\mathfrak{q}}}] \subset K_{\mathfrak{q}}$) to be the kernel of $\mathrm{GL}_2(\mathfrak{o}_{F_{\mathfrak{q}}}) \to \mathrm{GL}_2(\mathfrak{o}_{F_{\mathfrak{q}}}/\varpi_{\mathfrak{q}}^{e_{\mathfrak{q}}}\mathfrak{o}_{F_{\mathfrak{q}}})$ (resp. the preimage, under this map, of the upper triangular matrices). Thus

$$K_0[\mathfrak{q}^{e_{\mathfrak{q}}}] = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, d \in \mathfrak{o}_{F_{\mathfrak{q}}}, c \in \mathfrak{q}^{e_{\mathfrak{q}}}, ad - bc \in \mathfrak{o}_{F_{\mathfrak{q}}}^{\times} \right\}.$$

Now let $\mathfrak{f}$ be a fractional ideal, not necessarily prime, of F. Factorize $\mathfrak{f} = \prod_{\mathfrak{q}} \mathfrak{q}^{e_{\mathfrak{q}}}$ into prime ideals. We define elements $[\mathfrak{f}] \in \mathbb{A}_F^{\times}, a([\mathfrak{f}]), n([\mathfrak{f}]) \in \mathrm{GL}_2(\mathbb{A}_F)$ via:

$$[\mathfrak{f}] = \prod_{\mathfrak{q}|\mathfrak{f}} [\varpi_{\mathfrak{q}}]^{-e_{\mathfrak{q}}}, \quad n([\mathfrak{f}]) := \prod_{\mathfrak{q}|\mathfrak{f}} n_{\mathfrak{q}}\big(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}}\big), \quad a([\mathfrak{f}]) = \prod_{\mathfrak{q}|\mathfrak{f}} a_{\mathfrak{q}}\big(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}}\big). \tag{2.4}$$

[6] In Section 3.1 alone, we will use slightly different notation for $a(y)$ to accommodate the fact that we deal with SL_2 rather than GL_2. We make the relevant notation clear in that section.

[7] Recall that $|x|_v$, for a complex place v and $x \in F_v$, is the *square* of the usual absolute value on $\mathbb{C}$!

Suppose $\chi : \mathbb{A}_F^\times / F^\times \to \mathbb{C}^\times$ is a character. We define

$$\chi(\mathfrak{f}) = \begin{cases} 0, & \chi \text{ ramified at any place dividing } \mathfrak{f}, \\ \prod_{\mathfrak{q} \mid \mathfrak{f}} \chi([\varpi_{\mathfrak{q}}])^{e_{\mathfrak{q}}}, & \text{else.} \end{cases}$$

2.6. *Measures.* The choice of measure is not especially important, as we are only interested in upper bounds; thus, so long as we are consistent, the precise selection does not matter. We choose a "standard" set of measures here; at times in the text, especially when carrying out equidistribution arguments, it will be more convenient to use probability measures, and we will indicate when this is the case.

We denote by $\mu_{\mathbf{X}}$ the $\mathrm{PGL}_2(\mathbb{A}_F)$-invariant probability measure on $\mathbf{X}$. We shall sometimes simply denote it by dx.

Let v be a finite place of F. Unless explicitly stated otherwise, the measures on $\mathrm{GL}_2(F_v)$, $\mathrm{PGL}_2(F_v)$, F_v and $F_v^\times$ are the Haar measure which assigns $\mathrm{GL}_2(\mathfrak{o}_{F_v})$ (resp. $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$, $\mathfrak{o}_{F_v}$, $\mathfrak{o}_{F_v}^\times$) the total mass 1.

For v archimedean, endow F_v with a multiple of Lebesgue measure $c_v dx$, where the constants c_v are fixed arbitrarily in such a way that the induced product measure on F_∞ satisfies $\mathrm{vol}(F_\infty/\mathfrak{o}_F) = 1$; equivalently, the product measure on $\mathbb{A}_F$ satisfies $\mathrm{vol}(\mathbb{A}_F/F) = 1$. In particular, this product measure on $\mathbb{A}_F$ is self-dual with respect to e_F. We endow $F_v^\times$ with the measure $d^\times x = \frac{dx}{|x|_v}$, where dx is Lebesgue measure.

These choices induce a Haar measure on $N(F_v)$, by means of the identification $x \mapsto n(x)$; similarly, the identifications $(y, y') \mapsto a(y)a'(y')$ and $y \mapsto z(y)$ induce Haar measures on $A(F_v)$ and $Z(F_v)$. Equip K_v with the measure of mass 1, and give $\mathrm{GL}_2(F_v)$ the measure arising from the Iwasawa decomposition $N(F_v) \times A(F_v) \times K_v$. Equip $\mathrm{PGL}_2(F_v) = \mathrm{GL}_2(F_v)/Z(F_v)$ with the "quotient" measure.

We then take the measures on $\mathrm{GL}_2(\mathbb{A}_F), \mathrm{PGL}_2(\mathbb{A}_F), \mathbb{A}_F, \mathbb{A}_F^\times$ to be the corresponding product measures.

The measure on any discrete group (e.g., $\mathrm{PGL}_2(F)$, considered as a subgroup of $\mathrm{PGL}_2(\mathbb{A}_F)$) will be counting measure.

Usually (indeed, unless otherwise specified) we shall use the $\mathrm{PGL}_2(\mathbb{A}_F)$-invariant probability measure on $\mathbf{X} = \mathrm{PGL}_2(F)\backslash\mathrm{PGL}_2(\mathbb{A}_F)$. This does not coincide with the quotient measure induced from $\mathrm{PGL}_2(\mathbb{A}_F)$, but they differ by some constant depending only on F. On the few occasions we shall have occasion to use the latter measure, we will indicate this.

2.7. *Projection onto locally constant functions.* For equidistribution questions it is usually convenient to deal with the constant function and its orthogonal complement separately. Some minor complications arise in our case since the ambient spaces are not connected. In fact: The space $C^\infty(\mathbf{X}_{\mathbf{G}})$ is a direct limit of function spaces $C^\infty(\mathbf{X}_{\mathbf{G}}/K)$ where $K \subset K_{\max,\mathbf{G}}$ has finite index. Unless $\mathbf{G}$ is simply connected, the manifolds $\mathbf{X}_{\mathbf{G}}/K$ need not be connected.

Of course, to deal with this, one can (if $\mathbf{G}$ is semisimple) simply replace the notion of *constant* function by *locally constant* function. However, in the general case of $\mathbf{G}$ reductive, matters are slightly complicated by the necessity of dealing with central characters.

Since we will only use this definition when $\mathbf{G}$ is a product of GL(2)s, we restrict ourselves to that setting. First suppose that $\mathbf{G} = \mathrm{GL}(2)$. We define a projection $\mathcal{P} : C_\omega^\infty(\mathbf{X}_{\mathrm{GL}(2)}) \to C_\omega^\infty(\mathbf{X}_{\mathrm{GL}(2)})$ via

$$\mathcal{P} f(x) = \int_{h \in \mathrm{SL}_2(F)\backslash \mathrm{SL}_2(\mathbb{A}_F)} f(hx)dh = \sum_{\chi^2=\omega} \chi(x) \int_{\mathbf{X}} f(y)\overline{\chi(y)}dy, \tag{2.5}$$

where dh is the $\mathrm{SL}_2(\mathbb{A}_F)$-invariant probability measure, dy the $\mathrm{GL}_2(\mathbb{A}_F)$-invariant probability measure on $\mathbf{X}$, χ ranges over characters of $\mathbb{A}_F^\times/F^\times$ with square ω, $\chi(y)$ the function on $\mathbf{X}_{\mathrm{GL}(2)}$ defined by $g \mapsto \chi(\det(g))$, and the second equality is easily verified. We note, in particular, that the χ sum is finite (any χ for which the corresponding term is nonvanishing must be unramified outside $\mathrm{Supp}(f)$).

Then $\|\mathcal{P} f\|_{L^\infty} \le \|f\|_{L^\infty}$, as is clear from the first equality of (2.5), and $\mathcal{P}$ is a self-adjoint projection with respect to L^2, as is clear from the second equality. We say a function f is *totally nondegenerate* if $\mathcal{P} f = 0$.

If $\mathbf{G} = \mathrm{GL}(2)\times\mathrm{GL}(2)$, and $\omega = (\omega_1, \omega_2)$ is a character of the center $\mathbf{Z}(\mathbb{A}_F) = \mathbb{A}_F^\times \times \mathbb{A}_F^\times$, then we denote by $\mathcal{P}_1$ the operator on $C_\omega^\infty(\mathbf{X}_{\mathbf{G}})$ given by

$$\begin{aligned}\mathcal{P}_1 f(x_1, x_2) &= \int_{h \in \mathrm{SL}_2(F)\backslash \mathrm{SL}_2(\mathbb{A}_F)} f(hx_1, x_2)dh \\ &= \sum_{\chi^2=\omega_1} \chi(x_1) \int_{\mathbf{X}} f(y, x_2)\overline{\chi(y)}dy.\end{aligned}$$

We define $\mathcal{P}_2$ similarly, interchanging the role of the first and second coordinate. The operators $\mathcal{P}_j$ for $j = 1, 2$ commute, satisfy $\|\mathcal{P}_j f\|_{L^\infty} \le \|f\|_{L^\infty}$ and are commuting self-adjoint projections on L^2. We say that a function f is *totally nondegenerate* if $\mathcal{P}_1 f = \mathcal{P}_2 f = 0$.

LEMMA 2.1. *Let v be a place of F. The projection $\mathcal{P}$ acts by the identity on the subspace $W \subset L_\omega^2(\mathbf{X}_{\mathbf{G}})$ spanned by one-dimensional representations of $\mathrm{GL}_2(F_v)$ occurring in $L_\omega^2(\mathbf{X}_{\mathbf{G}})$.*

Similarly, $\mathcal{P}_1$ (resp. $\mathcal{P}_2$) acts by the identity on the space W_1 (resp. W_2) spanned by one-dimensional representations of $\mathrm{GL}_2(F_v)$ occurring in $L^2(\mathbf{X}\times\mathbf{X})$ for the action on the first (resp. second) factor.

Proof. This follows from the spectral decomposition for GL(2). For instance, it is known that the space W is precisely the span of functions of the form $g \mapsto \chi(\det(g))$, where χ ranges over characters of $\mathbb{A}_F^\times/F^\times$ satisfying $\chi^2 = \omega$. □

2.8. *Hecke operators and bounds towards the Ramanujan conjecture.* Let $\mathfrak{l}$ be a prime ideal of $\mathfrak{o}_F$ and r an integer ≥ 1. Let $F_{\mathfrak{l}}$ be the completion of F at the prime $\mathfrak{l}$. Take the Haar measure on $\mathrm{GL}_2(F_\mathfrak{l})$ so that it assigns mass 1 to $\mathrm{GL}_2(\mathfrak{o}_{F_\mathfrak{l}})$. Define the measure $\mu^*_{\mathfrak{l}^r}$ on $\mathrm{GL}_2(F_\mathfrak{l})$ to the restriction of Haar measure to the set $\mathrm{GL}_2(\mathfrak{o}_{F_\mathfrak{l}}) \cdot \left(\begin{smallmatrix} \varpi_\mathfrak{l}^r & 0 \\ 0 & 1 \end{smallmatrix}\right) \mathrm{GL}_2(\mathfrak{o}_{F_\mathfrak{l}})$, so that the total mass of $\mu^*_{\mathfrak{l}^r}$ is $\mathrm{N}(\mathfrak{l})^{r-1}(\mathrm{N}(\mathfrak{l})+1)$.

Moreover, set

$$\mu_{\mathfrak{l}^r} = \frac{1}{\mathrm{N}(\mathfrak{l})^{r/2}} \sum_{k \leq \frac{r}{2}} \mu^*_{r-2k}, \quad \overline{\mu}_{\mathfrak{l}^r} := \frac{\mu_{\mathfrak{l}^r}}{\|\mu_{\mathfrak{l}^r}\|}, \tag{2.6}$$

where $\|\cdot\|$ denotes total variation. Thus $\overline{\mu}_{\mathfrak{l}^r}$ is a probability measure. Via the natural inclusion of $\mathrm{GL}_2(F_\mathfrak{l})$ in $\mathrm{GL}_2(\mathbb{A}_{F,f})$, we may regard $\mu_{\mathfrak{l}^r}$ as a compactly supported measure on $\mathrm{GL}_2(\mathbb{A}_{F,f})$; by abuse of notation, we will not introduce a different symbol for this measure. If $\mathfrak{n}$ is an integral ideal of $\mathfrak{o}_F$, then factorize $\mathfrak{n} = \prod_i \mathfrak{l}_i^{r_i}$ and put $\mu_\mathfrak{n} = \prod \mu_{\mathfrak{l}_i^{r_i}}$, $\overline{\mu}_\mathfrak{n} = \prod \overline{\mu}_{\mathfrak{l}_i^{r_i}}$. Here $\prod$ is taken to mean convolution of measures on $\mathrm{GL}_2(\mathbb{A}_{F,f})$.

Convolution by $\mu_\mathfrak{n}$ on $L^2(\mathbf{X})$ corresponds to the $\mathfrak{n}$th Hecke operator; in this normalization the Ramanujan conjecture corresponds to it having eigenvalues ≤ 2 in absolute value.

The adelic measures $\mu_\mathfrak{n}$ satisfy the usual multiplication laws, appropriately interpreted: if $\mathfrak{n}$ and $\mathfrak{m}$ are ideals, then

$$\int_{\mathrm{PGL}_2(\mathbb{A}_F)} h(x) d(\mu_\mathfrak{n} \star \mu_\mathfrak{m})(x) = \sum_{\mathfrak{d} | (\mathfrak{m},\mathfrak{n})} \int_{\mathrm{PGL}_2(\mathbb{A}_F)} h(x) d\mu_{\mathfrak{n}\mathfrak{m}\mathfrak{d}^{-2}}(x), \tag{2.7}$$

whenever h is a function on $\mathrm{PGL}_2(\mathbb{A}_F)$ that is invariant under $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$ for all $v | \mathfrak{n}\mathfrak{m}$.

Definition 2.1. Let α be a bound towards Ramanujan for GL_2 over F, i.e., α is so that $\mu_\mathfrak{l}$ acts on any $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{l}})$-invariant cuspidal eigenfunction by an eigenvalue $\leq \mathrm{N}(\mathfrak{l})^\alpha + \mathrm{N}(\mathfrak{l})^{-\alpha}$ in absolute value.

Thus $\alpha = 0$ corresponds to the Ramanujan conjecture, $\alpha = 1/2$ the trivial bound. By work of Kim and Kim-Sarnak [22], we can take $\alpha = 3/26$. For our applications, any value of α less than $1/4$ would suffice.

Note that we shall slightly vary this notation (but in a reasonably compatible way) in Sections 3.1 and 9.3.1. In those parts, we shall deal with a (not necessarily arithmetic) quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$, and α will denote a number so that $L^2(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$ does not contain any complementary series with parameter $> \alpha$. (Here the complementary series is understood to be parametrized by $(0, 1/2)$). This is compatible with the above notation; however, e.g., if Γ were a congruence subgroup, $\alpha = 3/26$ would again be admissible.

2.9. *Sobolev-type norms on real and adelic quotients.*

2.9.1. *General comments.* Let M be a real manifold. Recall that the Sobolev norm on $C^\infty(M)$ controls, roughly speaking, the L^p-norm of a function together with the L^p-norms of certain derivatives. These norms will be tremendously useful throughout the paper to control equidistribution rates. We shall use both the relatively simple definition when $M = \Gamma\backslash G$ and an adelic variant.

First let us remark on the use of L^p-Sobolev norms for $p > 2$. This is solely to do with noncompactness. If we were to deal only with *compact* quotients, then the L^2-Sobolev theory would always suffice. However, in the noncompact case, the L^2-Sobolev norms do not (e.g.) give good bounds on the size of a function high in a cusp. There are, of course, various ways to rectify this; for example we could include weights that measure the height into the cusp. We have chosen instead to use L^p-norms with $p > 2$, which is technically very simple, but has some disadvantages (e.g., it does not induce a Hilbert space structure).

Note that we will allow our seminorms and norms to take the value ∞. Thus a seminorm on a complex vector space V will be a function from V to $\mathbb{R}_{\geq 0} \cup \{\infty\}$ satisfying

(1) $\|\lambda v\| = |\lambda|\|v\|$, for any $v \in V$ such that $\|v\| < \infty$;

(2) $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ if both $\|v_1\|$ and $\|v_2\|$ are not infinite.

It is a norm if additionally $\|v\| = 0$ implies $v = 0$. Note that giving such a seminorm on V is equivalent to giving a subspace $V_f \subset V$ together with a finite-valued seminorm on V_f. Indeed take $V_f = \{v \in V : \|v\| < \infty\}$, equipped with the restriction of $\|\cdot\|$.

We remark that we do not require that our norms be complete.

2.9.2. *Nonadelic setting.* Suppose $\Gamma \subset G$ is a lattice in a connected semisimple Lie group. Fix for all time a basis $\mathcal{B}$ for the Lie algebra $\mathfrak{g}$ of G and a norm $\|\cdot\|$ on $\mathfrak{g}$. For $g \in G$, we denote by $\|g\|$ the operator norm of $\mathrm{Ad}(g^{-1}) : \mathfrak{g} \to \mathfrak{g}$, i.e., the map $X \mapsto g^{-1}Xg$.

For $f \in C^\infty(\Gamma\backslash G)$, and $1 \leq p \leq \infty$, we put

$$S_{p,d} = \sum_{\mathrm{ord}(\mathcal{D}) \leq d} \|\mathcal{D} f(g)\|_{L^p(\Gamma\backslash G)}. \tag{2.8}$$

Here $\mathcal{D}$ ranges over all monomials in $\mathcal{B}$ of order $\leq d$, and $\mathcal{D}$ acts on f by right differentiation. (For example, $X \in \mathfrak{g}$ acts on f via $Xf(g) = \frac{d}{dt} f(ge^{tX})|_{t=0}$.)

Changing $\mathcal{B}$ only distorts $S_{p,d}$ by a bounded factor. (That is to say, if $S'_{p,d}$ is the norm obtained by replacing $\mathcal{B}$ by another basis, then there are positive reals c_1, c_2, possibly depending on d, such that $c_1 S_{p,d} \leq S'_{p,d} \leq c_2 S_{p,d}$.)

We will often use the following simple remark: Fix a Riemannian metric $d(\cdot,\cdot)$ on G and suppose $g \in G$ belongs to some fixed compact set. Then, for $f \in C^\infty(\Gamma\backslash G), x \in \Gamma\backslash G$, we have $|f(xg) - f(x)| \ll S_{\infty,1}(f) d(g,1)$. Indeed, we may assume that g is close to the identity and write $g = \exp(X)$, with $X \in \mathfrak{g}$; now apply the mean value theorem to $t \mapsto f(xe^{tX})$.

Moreover, the following elementary properties are easily verified (we only need them in the case $p = \infty$).

LEMMA 2.2. *Let $f_1, f_2 \in C^\infty(\Gamma\backslash G)$ and $g \in G$. Then*

$$S_{\infty,d}(f_1 f_2) \ll_d S_{\infty,d}(f_1) S_{\infty,d}(f_2), \qquad (2.9)$$
$$S_{\infty,d}(g \cdot f_1) \ll_d \|g\|^d S_{\infty,d}(f_1).$$

We remark that, in the case $p = 2$, the rule (2.8) also defines a system of Sobolev norms on any unitary G-representation; the case discussed above corresponds to the unitary representation $L^2(\Gamma\backslash G)$.

2.9.3. *Adelic Sobolev norms.* Let us first describe what the point is intended to be (evidently there are many ways of implementing it; cf. Remark 2.1). We would like to put a norm on the adelic function space, suitable for controlling, e.g., period integrals. Consider $C^\infty_\omega(\mathbf{X_G})$ in the case of $\mathbf{G} = \mathrm{SL}_2$, $F = \mathbb{Q}$, $\omega = 1$ as a direct limit of spaces $C^\infty(\Gamma_i\backslash \mathrm{SL}_2(\mathbb{R}))$, where Γ_i ranges over some class of congruence subgroups of $\Gamma_0 := \mathrm{SL}_2(\mathbb{Z})$. We equip each quotient $\Gamma_i\backslash \mathrm{SL}_2(\mathbb{R})$ with the $\mathrm{SL}_2(\mathbb{R})$ invariant probability measure. Then, on each space $C^\infty(\Gamma_i\backslash \mathrm{SL}_2(\mathbb{R}))$ we have the norm $S_{p,d}$ defined in the previous section. On the other hand, typical bounds on automorphic forms have an implicit dependence on the "level", i.e., the index $[\Gamma : \Gamma_i]$, so one would like to have a norm that increases with the level. The most naive candidate is, fixing a real number $\beta > 0$, to define the "norm" of $f \in C^\infty(\Gamma_i\backslash \mathrm{SL}_2(\mathbb{R}))$ to be $[\Gamma : \Gamma_i]^\beta S_{p,d}(f)$. This unfortunately does not quite make sense when we pass to the direct limit: however, we can "force it to make sense" by considering the *maximal norm* on the direct limit whose restriction to each $C^\infty(\Gamma_i\backslash \mathrm{SL}_2(\mathbb{R}))$ is bounded above by $[\Gamma : \Gamma_i]^\beta S_{p,d}(f)$. This will suffice for our purposes.

Let us formalize these ideas. In what follows we return to the setting of $\mathbf{G}$ a reductive group over F. The adelic Sobolev norms will be a family of norms on $C^\infty_\omega(\mathbf{X_G})$ indexed by a triple (p, d, β). The d and β indicate, approximately speaking, how stringently one should "penalize" rapid variation at the infinite and finite places respectively.

Let $p \geq 1, k \in \mathbb{N}, \beta \geq 0$. Fix a basis $\mathcal{B} = \{X_i\}$ for the real Lie group $\mathrm{Lie}(\mathbf{G}(F_\infty))$. Recalling the definition of K_ψ from (2.1), we define the pre-Sobolev functions $PS_{p,d,\beta}$ on $C^\infty_\omega(\mathbf{X_G})$ via:

$$\begin{aligned} PS_{p,d,\beta}(\psi) &= [K_{\max,\mathbf{G}} : K_\psi]^\beta \sum_{\mathrm{ord}(\mathcal{D}) \leq d} \|\mathcal{D}\psi\|_{L^p(\mathbf{X}_{\mathbf{G},\mathrm{ad}})} \\ &= \prod_{v \text{ finite}} [K_{v,\mathbf{G}} : K_{v,\psi}]^\beta \sum_{\mathrm{ord}(\mathcal{D}) \leq d} \|\mathcal{D}\psi\|_{L^p(\mathbf{X}_{\mathbf{G},\mathrm{ad}})}, \end{aligned} \qquad (2.10)$$

where the sum ranges over $\mathcal{D}$ that are monomials in $\mathcal{B}$ of order $\leq d$.

The function $PS_{p,d,\beta}$ does not satisfy the triangle inequality. We define the (p, d, β)-Sobolev norm $S_{p,d,\beta}$ to be the maximal seminorm on $C^\infty_\omega(\mathbf{X_G})$ satisfying

$S_{p,d,\beta}(\psi) \leq PS_{p,d,\beta}(\psi)$. Explicitly,

$$S_{p,d,\beta}(\psi) = \inf\{\sum_{i=1}^{n} PS_{p,d,\beta}(\psi_j) : \sum_{i=1}^{n} \psi_j = \psi,\ \psi_j \in C_\omega^\infty(\mathbf{X_G})\}. \tag{2.11}$$

In fact, it is clear that the right-hand side of (2.11) defines a seminorm that is dominated by $PS_{p,d,\beta}$ (take the collection $\{\psi_i\}$ to consist of $\{\psi\}$ alone); moreover, it is evidently maximal in the class of such seminorms. Finally, as $PS_{p,d,\beta}(\psi) \geq \|\psi\|_{L^p}$, the $S_{p,d,\beta}$ are in fact norms on $C_\omega^\infty(\mathbf{X_G})$.

It will often be useful to omit the argument β and set it to a "default" value of $1/p$. We therefore define $S_{p,d} := S_{p,d,1/p}$, for $p \neq 0$, and $S_{\infty,d} := S_{\infty,d,\mathbf{0}}$.

Notational convention. We will very often have cause to bound linear functionals L on $C_\omega^\infty(\mathbf{X_G})$ by Sobolev norms. In writing statements of the form $|L(f)| \ll S_{p,d,\beta}(f)$, we will always allow the implicit constant to depend on p, d and β without explicitly saying so.

2.10. *Adelic Sobolev norms — a slight generalization.* The notation of this section will only be required in Section 7. We recommend it be omitted at a first reading.

In the discussion at the start of Section 2.9.3, we did not address what class of subgroups Γ_i to consider (should we take *all* finite index subgroups of a fixed Γ_0 or some subclass?). Implicitly, such a choice was made in defining the Sobolev norms of the previous section. The Sobolev norms introduced in the previous section are good for most of our purposes. However, roughly speaking, they have the following defect: they only measure the index of a stabilizer of a function $f \in C_\omega^\infty(\mathbf{X_G})$.

That this definition might lead to some peculiar results can be already seen in the case $\mathbf{G} = \mathrm{PGL}(2)$, $F = \mathbb{Q}$. Let χ_p be the character of $\mathbb{A}_\mathbb{Q}^\times/\mathbb{Q}^\times$ that corresponds to the quadratic Dirichlet character of $\mathbb{Q}$ with conductor p, a prime number. Then the function $g \mapsto \chi_p(\det(g))$ descends to a function f on $\mathbf{X}$, and it is easy to check that $[K_{\max} : K_f] = 2$, for any p. Thus the index of this stabilizer does not reflect the conductor of the underlying representation (which, by any reasonable definition of conductor, should grow as p increases). In this section we shall introduce a slight modification of the definitions which avoids this problem. (This problem would not occur for SL(2)).

This is a purely technical matter, and it seems there is much scope for giving better and more natural definitions. We restrict ourselves to the case $\mathbf{G} = \mathrm{GL}(n)$. For a finite place $\mathfrak{q}$ and $m \geq 0$, we put $K[\mathfrak{q}^m] \stackrel{\text{def}}{=} \ker(\mathrm{GL}(n, \mathfrak{o}_{F_\mathfrak{q}}) \to \mathrm{GL}(n, \mathfrak{o}_{F_\mathfrak{q}}/\varpi_\mathfrak{q}^m \mathfrak{o}_{F_\mathfrak{q}})$, where $\varpi_\mathfrak{q}$ is a uniformizer in $F_\mathfrak{q}$. Now, for $\psi \in C_\omega^\infty(\mathbf{X_G})$, put $K_{\mathfrak{q},\psi}^*$ to be the largest subgroup $K[\mathfrak{q}^m]$ which stabilizes ψ, and put $K_\psi^* = \prod_\mathfrak{q} K_{\mathfrak{q},\psi}^*$.

We define the $\star$-pre-Sobolev norm $PS_{p,d,\beta}^*$ by the rule

$$PS_{p,d,\beta}^\star(\psi) = [K_{\max,\mathbf{G}} : K_\psi^*]^\beta \sum_{\mathrm{ord}(\mathscr{D}) \leq d} \|\mathscr{D}\psi\|_{L^p(\mathbf{X}_{\mathbf{G},\mathrm{ad}})}$$

and we define the $\star$-Sobolev norm $S^*_{p,d,\beta}$ to be the maximal seminorm dominated by $PS^*_{p,d,\beta}$. Clearly, $S^*_{p,d,\beta} \geq S_{p,d,\beta}$.

The eventual purpose of this is that $S^*_{p,d,\beta}$ (unlike $S_{p,d,\beta}$) will never be "too small" on an automorphic representation whose conductor is large. This can be quantified, although we do not do so in the present document.

Remark 2.1. Evidently the definitions of this section and the previous are not the only "sensible" way of defining a notion of adelic Sobolev norms. The results of this paper do not require any more sophisticated definition, although this would certainly be of help in optimizing the results.

However, it would be interesting to impose a system of Sobolev-type norms in a less *ad hoc* fashion. Moreover, it would be pleasant if the system of norms had nice interpolation properties (this often is very helpful for getting sharp results). For example, it would be nice if as one varied β one got a family of interpolation spaces.

We remark on a simple way of defining Hilbertian norms which seems (more) appropriate to the adelic context. Let $K[\mathfrak{q}^m]$ be as above, and let $E_{\mathfrak{q}^m}$ be the averaging projection onto the $K[\mathfrak{q}^m]$-fixed vectors; i.e. $E_{\mathfrak{q}^m}(v) = \int_{k \in K[\mathfrak{q}^m]} k \cdot v dk$, where the measure is the Haar probability measure. Then $e_{\mathfrak{q}^m} := E_{\mathfrak{q}^m} - E_{\mathfrak{q}^{m-1}}$ is a projection. If $\mathfrak{f} = \prod_i \mathfrak{q}_i^{m_i}$ is an arbitrary integral ideal, put $e_{\mathfrak{f}} := \prod_i e_{\mathfrak{q}_i^{m_i}}$. Now put $P(s) = \sum_{\mathfrak{f}} e_{\mathfrak{f}} \mathrm{N}(\mathfrak{f})^s$. Then $f \mapsto \sum_{\mathrm{ord}(\mathcal{D}) \leq d} \|\mathcal{D} \cdot P(s) \cdot f\|_{L^2}$ defines a Hilbert norm which seems to have reasonably pleasant formal properties. In fact it is majorized (up to constants) by a norm of the type described above,.

J. Bernstein has a more canonical notion of norms on representation spaces of p-adic groups, and he has informed me that these norms have adelic analogues. I do not know the relation. The norms arising from his constructions are Hilbertian.

2.11. *Some properties and uses of the Sobolev norms.* We briefly summarize certain results that will be used in the text. Detailed proofs are given in Section 8.

For general $\mathbf{G}, \omega$ we have:

$$S_{p,d,\beta}(F_1 F_2) \ll_d S_{2p,d,\beta}(F_1) S_{2p,d,\beta}(F_2), \tag{2.12}$$

$$S_{p,d,\beta}(g \cdot F) \ll \|g_\infty\|^d \|g_f\|^\beta S_{p,d,\beta}(F). \tag{2.13}$$

Equation (2.12), proved in Lemma 8.1, and (2.13), proved in Lemma 8.2, give some basic stability properties of Sobolev norms.

Now we specialize to some results for GL(2) and PGL(2). Let $F \in C^\infty(\mathbf{X} \times \mathbf{X})$, let $\mathfrak{q}$ be a prime ideal of $\mathfrak{o}_F$, and suppose F is invariant by $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{q}}) \times \mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{q}})$. Then for $p > 2, d \gg 1$:

$$\left| \int_{\mathbf{X}} F(x, xa([\mathfrak{q}]))dx - \sum_{\chi^2=1} \chi([\mathfrak{q}]) \int_{\mathbf{X}} F(x,y)\chi(x)\chi(y) d\mu_{\mathbf{X}}(x) d\mu_{\mathbf{X}}(y) \right| \ll_\epsilon \mathrm{N}(\mathfrak{q})^{\frac{2\alpha-1}{p}+\epsilon} S_{p,d}(F). \tag{2.14}$$

Equation (2.14), proved in Lemma 9.9, quantifies Hecke equidistribution. To understand the relation, take F to be a pure tensor: $F(x, y) = f_1(x) f_2(y)$. Then (2.14) in effect bounds the inner product $\langle T_{\mathfrak{q}} f_1, f_2 \rangle$, where $T_{\mathfrak{q}}$ is the Hecke operator corresponding to $\mathfrak{q}$.

2.12. *Cusp forms, L-functions and the analytic conductor.* As a general remark on notation — and a mild abuse of notation — by *cuspidal representation* we shall always mean *unitary cuspidal representation*. This is automatic for PGL(2) but not for GL(2).

2.12.1. *L-functions.* Let $\pi = \otimes_v \pi_v$ be an automorphic cuspidal representation of GL(n) over F. We denote by $L_v(s, \pi_v)$ the local L-factor of the representation π_v; when it causes no confusion, we will sometimes abbreviate this to $L(s, \pi_v)$.

We write $L(s, \pi) := \prod_{v \text{ finite}} L_v(s, \pi_v)$ for the (finite part of) the global L-function attached to π, and $\Lambda(s, \pi) := \prod_v L_v(s, \pi_v)$ for the (completed) L-function attached to π.

2.12.2. *The analytic conductor of Iwaniec-Sarnak.* We recall the definition in the context where it will arise. Let $\pi = \otimes \pi_v$ be a cuspidal representation of GL(n) over F.

For each finite place v we denote by $\mathrm{Cond}_v(\pi)$ the conductor, in the sense of Jacquet, Piatetski-Shapiro, and Shalika, of π_v; thus $\mathrm{Cond}_v(\pi) = q_v^{m_v}$, where m_v is the smallest nonnegative integer such that π_v possesses a fixed vector under the subgroup of $\mathrm{GL}_n(\mathfrak{o}_{F_v})$ consisting of matrices whose bottom row is congruent to $(0, 0, \dots, 0, 1)$ modulo ϖ_v^m.

For each infinite place v, let $\Gamma_v(s) = \pi^{-s/2}\Gamma(s/2)$ or $(2\pi)^{-s}\Gamma(s)$ according to whether v is real or complex respectively, and put $\deg(v) = [F_v : \mathbb{R}]$. Let $\mu_{j,v} \in \mathbb{C}$ satisfy $L(s, \pi_v) = \prod \Gamma_v(s + \mu_{j,v})$, and put $\mathrm{Cond}_v(\pi) = \prod_v (1 + |\mu_{j,v}|)^{\deg(v)}$. We then put $\mathrm{Cond}(\pi) = \prod_v \mathrm{Cond}_v(\pi)$ (this is within a constant factor of the Iwaniec-Sarnak definition). Moreover, we put $\mathrm{Cond}_\infty(\pi) = \prod_{v \text{ infinite}} \mathrm{Cond}_v(\pi)$ and $\mathrm{Cond}_f(\pi) = \prod_{v \text{ finite}} \mathrm{Cond}_v(\pi)$ (the "infinite" and "finite" parts of the conductor).

We will occasionally refer to the "finite conductor" of π as the *ideal* $\prod_v \mathfrak{q}_v^{m_v}$, where $\mathfrak{q}_v$ is the prime ideal corresponding to the finite place v; then $\mathrm{Cond}_f(\pi)$ is the norm of this ideal. Hopefully the distinction between the two usages will be clear from context.

Remark 2.2 (Explication for GL(1) in the archimedean case). Let us be slightly more explicit in the case of a unitary character ω of $\mathbb{A}_F^\times / F^\times$. If v is real, then there is $t \in \mathbb{R}$ such that $\omega_v(x) = x^{it}$ for $x > 0$; then $\mathrm{Cond}_v(\omega) \asymp (1 + |t|)$. If v is complex, then there is $t \in \mathbb{R}, N \in \mathbb{Z}$ such that $\omega_v(re^{i\theta}) = r^{it} e^{iN\theta}$; then $\mathrm{Cond}_v(\omega) \asymp (1 + |t| + N)^2$.

We can heuristically summarize this: in the real case, ω is approximately constant in a neighborhood of the identity of size $\mathrm{Cond}_v(\omega)^{-1}$; in the complex, case ω is approximately constant in a disc around the identity of area $\mathrm{Cond}_v(\omega)^{-1}$.

2.12.3. *Cusp forms.* If π is a cuspidal representation of $\mathrm{GL}_2(\mathbb{A}_F)$ or $\mathrm{PGL}_2(\mathbb{A}_F)$, it will be convenient to denote by π_∞ the archimedean representation (of $\mathrm{GL}_2(F_\infty)$ or $\mathrm{PGL}_2(F_\infty)$) that corresponds to π.

By the dual $\widehat{\mathrm{GL}_2(F_\infty)}$ or $\widehat{\mathrm{PGL}_2}(F_\infty)$, we shall mean the space of irreducible, admissible representations. We say a subset of this dual is *bounded* if the corresponding set of Langlands parameters is bounded. We may define, in an evident way, the conductor $\mathrm{Cond}(\pi_\infty)$ for $\pi_\infty \in \widehat{\mathrm{GL}_2(F_\infty)}$; with this definition, a subset is bounded exactly when Cond takes bounded values on it.

In a similar fashion, we define the notion of a bounded subset of $\widehat{\mathrm{GL}_2(F_v)}$ or $\widehat{\mathrm{PGL}_2(F_v)}$ for any place v, where, again $\widehat{\mathrm{GL}_2(F_v)}$ denotes the set of irreducible, admissible representations.

3. **Unipotent periods**

In this section, we systematically use the (nonadelic) Sobolev norms $S_{\infty,d}$ on homogeneous spaces $\Gamma\backslash G$. In rough terms, $S_{\infty,d}$ controls the L^∞-norm of the first d derivatives (see §2.9.2). Note in particular that $S_{\infty,0}$ is just the L^∞-norm.

3.1. *Equidistribution of sparse subsets of horocycles.* Let $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$ be a cocompact lattice. For this section alone, we will use mildly different notation to that of Section 2.5, to accommodate the fact we deal with SL_2 and not with PGL_2. For $x \in \mathbb{R}$, put

$$n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, a(x) = \begin{pmatrix} x^{1/2} & 0 \\ 0 & x^{-1/2} \end{pmatrix}, \bar{n}(x) = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}. \tag{3.1}$$

We denote by $C(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$ (resp. $C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$) the space of continuous (resp. smooth) functions on the compact real manifold $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$. We denote by dg the measure on $\mathrm{SL}_2(\mathbb{R})$ that descends to a probability measure on the quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$. Finally, we denote by $\mathbb{H}$ the usual upper half-plane $\{z : \Im(z) > 0\}$ with the standard action of $\mathrm{SL}_2(\mathbb{R})$.

THEOREM 3.1. *There exists $\gamma_{\max} > 0$, depending on Γ, such that $\{x_0 n(j^{1+\gamma}) : j \in \mathbb{N}\}$ is equidistributed, for any $x_0 \in \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$ and any $0 \le \gamma < \gamma_{\max}$. In other words, for any $f \in C(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$,*

$$\lim_{N\to\infty} \frac{\sum_{j=1}^N f(x_0 n(j^{1+\gamma}))}{N} = \int_{\Gamma\backslash\mathrm{SL}_2(\mathbb{R})} f(g)dg.$$

If λ_1 is the smallest nonzero eigenvalue of the Laplacian on $\Gamma\backslash\mathbb{H}$, put

$$\alpha = \begin{cases} 0, & \lambda_1 \ge 1/4 \\ \sqrt{1/4 - \lambda_1}, & \textit{else.} \end{cases}$$

Then we can take $\gamma_{\max} = \frac{(1-2\alpha)^2}{16(3-2\alpha)}$.

This result represents (extremely modest) progress towards a conjecture of N. Shah, which asserts that the statement should remain valid for *any* $\gamma > 0$. The method is not restricted to sequences of the specific type in Theorem 3.1, and we have also not optimized the maximal value for $\gamma_{\max}$. Nevertheless the method is fundamentally limited. As it presently stands, it does not seem capable of achieving even $\gamma = 1$.

The dependence of $\gamma_{\max}$ on Γ can likely be removed, but this seems to require using further input (cf. last paragraph of §1.3.4).

The proof follows the line of Section 1.3.1, with $G_1 = \mathrm{SL}_2(\mathbb{R})$, $G_2 = \{n(x) : x \in \mathbb{R}\}$. The Y_i are not quite closed G_2-orbits, but rather long pieces of general G_2-orbits. The basis $\{\psi_{i,j}\}$ for Y_i will correspond to additive characters of $G_2 \cong \mathbb{R}$.

Let $f \in C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$, and let α be as in the statement of Theorem 3.1. Let $T \geq 1$. Let ψ be a fixed nontrivial character of the additive group of $\mathbb{R}$. Let g be a fixed smooth function of compact support on $\mathbb{R}$ satisfying $\int_{-\infty}^{\infty} g(x)dx = 1$. We denote by $\langle\cdot,\cdot\rangle_{L^2(\Gamma\backslash G)}$ the inner product in the Hilbert space $L^2(\Gamma\backslash G)$.

We set:

$$\nu_T(f) = \frac{1}{T}\int_0^T f(x_0 n(t))dt,\ \mu_{T,\psi}(f) = \frac{1}{T}\int_0^T \psi(t) f(x_0 n(t))dt. \tag{3.2}$$

Note first that the measures ν_T are equidistributed as $T \to \infty$, in the following quantitative sense: for $f \in C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$,

$$\left|\nu_T(f) - \int_{\Gamma\backslash\mathrm{SL}_2(\mathbb{R})} f(g)dg\right| \ll T^{-\kappa_1} S_{\infty,1}(f), \tag{3.3}$$

for any $\kappa_1 < \frac{1/2-\alpha}{2}$. This is proven in Lemma 9.5, without taking any pains to optimize the exponent. (We prove it to keep the paper self-contained. However, we emphasize that neither result nor proof is new; see [31] and [33]. A precise analysis of the equidistribution of long horocycles is carried out in [14].)

LEMMA 3.1. *Suppose* $\int_{\Gamma\backslash\mathrm{SL}_2(\mathbb{R})} f(g)dg = 0$. *Then*:

$$|\mu_{T,\psi}(f)| \ll T^{-b} S_{\infty,1}(f), \tag{3.4}$$

whenever $b < \frac{(1-2\alpha)^2}{8(3-2\alpha)}$ *and the implicit constant is independent of* ψ.

Note that if ψ is wildly oscillatory, cancellation in $\mu_{T,\psi}$ can be proved directly by integration by parts; on the other hand, if ψ is almost constant, the cancellation in $\mu_{T,\psi}$ arises from the equidistribution of the horocycle x_0G_2. It is therefore the intermediate case in which (3.4) is of interest.

Proof. Let $H \geq 1$, and let σ_H be the measure on $N(\mathbb{R})$ defined by $\sigma_H(g) = \frac{1}{H}\int_0^H \psi(x)g(n(x))dx$, for g a function on $N(\mathbb{R})$.

For $f \in C^\infty(\Gamma\backslash G)$, we denote by $f \star \sigma_H$ the right convolution of f by σ_H. Then it is easy to verify that $\left|\mu_{T,\psi}(f) - \mu_{T,\psi}(f \star \sigma_H)\right| \ll \frac{H}{T} S_{\infty,0}(f)$. Here $\star\sigma_H$ denotes right convolution by σ_H. On the other hand, by Cauchy-Schwarz,

$|\mu_{T,\psi}(f \star \sigma_H)|^2 \le \nu_T(|f \star \sigma_H|^2)$. Thus, expanding $|f \star \sigma_H|^2$ and applying (3.3), we conclude

$$|\mu_{T,\psi}(f)| \ll \frac{H}{T} S_{\infty,0}(f) + \left(\frac{1}{H^2}\int_{(h_1,h_2)\in[0,H]^2} \left|\nu_T(n(h_1)f \cdot \overline{n(h_2)f})\right| dh_1 dh_2\right)^{1/2} \tag{3.5}$$
$$\ll \frac{H}{T} S_{\infty,0}(f) + \left(\frac{1}{H^2}\int_{(h_1,h_2)\in[0,H]^2} \left|\langle n(h_1-h_2)f, f\rangle_{L^2(\Gamma\backslash G)}\right| dh_1 dh_2\right)^{1/2} + \left(T^{-\kappa_1} \sup_{(h_1,h_2)\in[0,H]^2} S_{\infty,1}(n(h_1)f\cdot\overline{n(h_2)f})\right)^{1/2}.$$

Utilizing bounds towards matrix coefficients — see Section 9.1.2, especially (9.7) — and basic properties of Sobolev norms (see[8] Lemma 2.2), we note:

$$\langle n(h)f, f\rangle \ll_\epsilon (1+|h|)^{2\alpha-1+\epsilon} S_{\infty,1}(f)^2, \tag{3.6}$$
$$|S_{\infty,1}(n(h_1)f\cdot\overline{n(h_2)f})| \ll (1+|h_1|+|h_2|)^2 S_{\infty,1}(f)^2.$$

Thus, $|\mu_{T,\psi}(f)| \ll_\epsilon \left(\frac{H}{T} + H^{\alpha-1/2+\epsilon} + T^{-\kappa_1/2}H\right)S_{\infty,1}(f)$. Choose H so that $H^{\alpha-1/2} = HT^{-\kappa_1/2}$ to obtain the claimed result. □

Proof of Theorem 3.1. Given Lemma 3.1, the theorem follows quite readily by Fourier-expanding the measure on $\mathbb{R}$ that is a sum of point masses at j^γ, for $j \in \mathbb{N}$. The argument that follows formalizes a minor variant of this argument (we first consider instead a sum of point masses along arithmetic progressions which approximate $\{j^\gamma : j \in \mathbb{N}\}$).

Let $x_0 \in \Gamma\backslash\mathrm{SL}_2(\mathbb{R})$, $f \in C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$. We first claim that, if b is as in the previous lemma, f so that $\int_{\Gamma\backslash\mathrm{SL}_2(\mathbb{R})} f(g)dg = 0$, and $K \ge 1$, then

$$\frac{\sum_{0\le j<K^{1/b-1}} f(x_0 n(Kj))}{K^{1/b-1}} \to 0, \tag{3.7}$$

as $K \to \infty$. In other words, $K^{1/b-1}$ points, distributed along a horocycle with spacing K, become equidistributed.

This follows from Lemma 3.1: put $g_\delta(x) = \max(\delta^{-2}(\delta-|x|), 0)$, a function on $\mathbb{R}$. For $\lambda \in \mathbb{R}$, write $a_\lambda = K^{-1}\int_{\mathbb{R}} \exp(-2\pi i K^{-1}\lambda t) g_\delta(t)dt$. Then we have $\sum_{j\in\mathbb{Z}} g_\delta(t+Kj) = \sum_{k\in\mathbb{Z}} \exp(2\pi i K^{-1}kt)a_k$. Moreover, a simple computation shows that $\sum_{k\in\mathbb{Z}} |a_k| \ll \delta^{-1}$.

[8]Lemma 2.2 would actually give the exponent $(1+|h_1|+|h_2|)^4$ in the latter inequality, but it is easy to see directly the stronger result.

Choose $\varepsilon > 0$ so that $b+\varepsilon$ still satisfies the inequality of Lemma 3.1. By Lemma 3.1,

(3.8)
$$\left|\int_{t=0}^{T} dt\Big(\sum_{j\in\mathbb{Z}} g_\delta(t+Kj)\Big) f(x_0 n(t))\right| \ll_f T^{1-b-\varepsilon}\sum_k |a_k| \ll T^{1-b-\varepsilon}\delta^{-1}.$$

Now, g_δ has integral 1 and is supported in a δ-neighborhood of 0; in particular, the left-hand side of (3.8) differs from $\sum_{j\in\mathbb{Z},0\le Kj\le T} f(x_0 n(Kj))$ by an error that is $\ll_f (1+TK^{-1}\delta)$. Thus

$$\left|\sum_{j\in\mathbb{Z},0\le Kj\le T} f(x_0 n(Kj))\right| \ll_f (1+TK^{-1}\delta+T^{1-b-\varepsilon}\delta^{-1})$$

from which (3.7) readily follows.

We now deduce the theorem from (3.7). Let $T_0\in\mathbb{N}$ be large. Then, for t small, we have $(T_0+t)^{1+\gamma} = T_0^{1+\gamma}+(1+\gamma)T_0^{\gamma}t+O(t^2T_0^{\gamma-1})$. In particular, $(T_0+t)^{1+\gamma}$ is well-approximated by the linear function $T_0^{1+\gamma}+(1+\gamma)T_0^{\gamma}t$ in the range where $|t|\ll T_0^{\frac{1-\gamma}{2}}$.

The claim of Theorem 3.1 follows from (3.7) as long as $\frac{1-\gamma}{2\gamma} > 1/b-1$; in particular, any $\gamma < b/2$ will do. □

3.2. *Fourier coefficients of automorphic forms.* In the notation of Section 1, if we take for Y_i the closed orbits of a unipotent group, the resulting periods are so-called "Fourier coefficients of automorphic forms." We shall give a general nontrivial bound in that context. (The word "nontrivial" must be interpreted with care; see the discussion at the end of this section.) Our methods are restricted to the case of horospherical unipotent subgroups.

We have made no effort to optimize the exponents of the results, nor even to state a result of maximal generality. In fact, one can considerably increase the scope of Theorem 3.2, since we deal in the present section only with *closed orbits of horospherical subgroups*, one can profitably apply spectral theory. We do not carry this out, instead using [23] to give equidistribution statements in a fairly soft fashion.

Let G be a connected semisimple real Lie group, $\Gamma\subset G$ a lattice, $K\subset G$ the maximal compact subgroup, $\mathfrak{g}$ the Lie algebra of G, and $H\in\mathfrak{g}$ a semisimple element. Fix a norm $\|\cdot\|$ on the real vector space $\mathfrak{g}$. Let $\exp:\mathfrak{g}\to G$ be the exponential map. Fix a Haar measure on G so that $\Gamma\backslash G$ has volume 1. Let $\mathfrak{u}$ be the sum of all negative root spaces for H and let $U=\exp(\mathfrak{u})\subset G$. Let $x_0\in\Gamma\backslash G$ be so that x_0U is compact. Let $x_t = x_0\exp(tH)$, and let Δ_t be the stabilizer of x_t in U. We denote by $\langle\cdot,\cdot\rangle_{L^2(\Gamma\backslash G)}$ the inner product in the Hilbert space $L^2(\Gamma\backslash G)$.

We shall analyze periods of a fixed function along x_tU as t varies. The proofs follow Section 1.3.1 with $G_1=G, G_2=U$, $Y_i=x_tU$, and $\psi_{i,j}$ corresponding to characters of U.

Let $T > 0$ and let ψ be any character of U trivial on Δ_T. Recall that Δ_T is defined to be $\exp(-TH)\Delta_0 \exp(TH)$. We define $\nu_T, \mu_{T,\psi}$ in a closely analogous fashion to (3.2):

$$\nu_T(f) = \frac{\int_{\Delta_T \backslash U} f(x_T u) du}{\mathrm{vol}(\Delta_T \backslash U)}, \quad \mu_{T,\psi}(f) = \frac{\int_{\Delta_T \backslash U} f(x_T u)\psi(u) du}{\mathrm{vol}(\Delta_T \backslash U)}. \tag{3.9}$$

Let $f, g \in C^\infty(\Gamma\backslash G)$. Let $E \in \mathfrak{u}$ have unit length with respect to the fixed norm $\|\cdot\|$ on $\mathfrak{g}$. It is proven by Kleinbock-Margulis in [23] — see also Lemmas 9.6 and 9.7 — that there are $\kappa_1, \kappa_2 > 0$ such that

$$\left| \langle \exp(sE) \cdot f, g \rangle_{L^2(\Gamma\backslash G)} - \int_{\Gamma\backslash G} f \int_{\Gamma\backslash G} g \right| \ll (1+|s|)^{-\kappa_1} S_{\infty,\dim(K)}(f) S_{\infty,\dim(K)}(g), \tag{3.10}$$

$$|\nu_T(f) - \int_{\Gamma\backslash G} f| \ll e^{-\kappa_2 T} S_{\infty,\dim(K)}(f). \tag{3.11}$$

Equations (3.10) and (3.11) assert, respectively, quantitative mixing of the flow generated by U on $\Gamma\backslash G$, and the equidistribution of the orbit $x_T U$ as $T \to \infty$. We may assume that $\kappa_1 < 1$ (since making κ_1 smaller does not change the truth of (3.10)). This will ease the notation in the proof of the theorem.

THEOREM 3.2. *There exists $\kappa_3 > 0$ such that, for any $f \in C^\infty(\Gamma\backslash G)$ satisfying $\int_{\Gamma\backslash G} f = 0$, we have*:

$$|\mu_{T,\psi}(f)| \ll \exp(-T\kappa_3) S_{\infty,\dim(K)}(f) \tag{3.12}$$

for all $T \geq 0$, and for all characters ψ of U trivial on Δ_T.

Indeed, if o is the order of the polynomial map $\mathbb{R} \to \mathrm{End}(\mathfrak{g})$ defined by $s \mapsto \mathrm{Ad}(\exp(sH))$, then any $\kappa_3 < \frac{\kappa_1\kappa_2}{2(2o\dim(K)+\kappa_1)}$ is admissible, $\kappa_{1,2}$ being as in (3.10) *and* (3.11).

The relevance to this to "Fourier coefficients" in the classical sense may not be immediately clear; after the proof, we give the example of $\mathrm{SL}_2(\mathbb{R})$ to illustrate.

Also, observe that the estimate (3.12) is uniform in ψ. In fact, just as in Lemma 3.1, the case when ψ is constant amounts to (3.11), whereas the case where ψ is highly oscillatory could be handled by integration by parts. It is, again, the intermediate case where (3.12) has content.

Proof. We first remark that, other than being in a slightly more general setting, the proof is almost exactly the same as the proof of Lemma 3.1.

The signed measure $\mu_{T,\psi}$ satisfies $\mu_{T,\psi}(u \cdot f) = \overline{\psi(u)}\mu_{T,\psi}(f)$, for $u \in U$.

Take $E \in \mathfrak{u}$ of unit length with respect to the norm $\|\cdot\|$ on $\mathfrak{g}$. Fix $H \geq 1$. Let σ be the measure on U defined via the rule

$$\sigma(h) = \frac{1}{H} \int_{s=0}^{H} \psi(\exp(sE)) h(\exp(sE)) ds,$$

for h any continuous compactly supported function on U. Then $\mu_{T,\psi}(f) = \mu_{T,\psi}(f \star \sigma)$; thus:

$$|\mu_{T,\psi}(f)|^2 = |\mu_{T,\psi}(f \star \sigma)|^2 \leq \nu_T(|f \star \sigma|^2) \tag{3.13}$$
$$\ll \int_{\Gamma\backslash G} |f \star \sigma|^2 + e^{-\kappa_2 T} S_{\infty,\dim(K)}(|f \star \sigma|^2)$$
$$\ll \int_{\Gamma\backslash G} |f \star \sigma|^2 + H^{2o\dim(K)} \exp(-\kappa_2 T) S_{\infty,\dim(K)}(f)^2,$$

where we have applied Cauchy-Schwarz followed by (3.11), noting that by Lemma 2.2, we have that

$$S_{\infty,\dim(K)}(|f \star \sigma|^2) \ll S_{\infty,\dim(K)}(f \star \sigma)^2 \ll H^{2o\dim(K)} S_{\infty,\dim(K)}(f)^2.$$

Here o is chosen as in the statement of the theorem.

By (3.10), we see that

$$\int_{\Gamma\backslash G} |f \star \sigma|^2 \ll \left(\frac{1}{H}\int_0^H (1+|t|)^{-\kappa_1} dt\right) S_{\infty,\dim(K)}(f)^2 \tag{3.14}$$
$$\ll H^{-\kappa_1} S_{\infty,\dim(K)}(f)^2.$$

Indeed, this follows simply by expanding the leftmost expression. Thus

$$|\mu_{T,\psi}(f)|^2 \ll (H^{-\kappa_1} + H^{2o\dim(K)} \exp(-\kappa_2 T)) S_{\infty,\dim(K)}(f)^2.$$

We choose H so that $H^{2o\dim(K)+\kappa_1} = \exp(\kappa_2 T)$ to conclude. □

Remark 3.1. We now explain, when we specialize $G = \mathrm{SL}_2(\mathbb{R})$, why this recovers Sarnak's result [37], which was the first improvement of the Hecke bound for nonarithmetic groups. Take $H = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \mathfrak{sl}_2$, so that $U = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$. Let $\Gamma \subset G$ be a nonuniform lattice so that $\Gamma \cap U = \left\{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right), n \in \mathbb{Z}\right\}$, and take $x_0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Let $f(x+iy) = \sum_{n\neq 0} a_n \sqrt{y} K_{i\nu}(2\pi n y) e^{2\pi i n x}$ be a Maass cusp form of eigenvalue $1/4 + \nu^2$ on $\Gamma\backslash\mathbb{H}$, where $\mathbb{H}$ denotes the upper half-plane; it lifts to a function on $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$, viz. $g \mapsto f(g.i)$.

Then the theorem implies (in concrete language) that there exists $\delta > 0$ such that, for any $y \leq 1$ and any $n \in \mathbb{Z}$,

$$\int_{0\leq x\leq 1} f(x+iy) e(nx) dx \ll y^{\delta}. \tag{3.15}$$

Taking $y \asymp n^{-1}$ in (3.15), one easily deduces that the Fourier coefficients a_n satisfy the "nontrivial" bound $|a_n| \ll n^{1/2-\delta}$.

Remark 3.2. The bound Theorem 3.2 is nontrivial in that it improves, as $t \to \infty$ on the trivial bound:

$$\left| \frac{\int_{\Delta_t\backslash U} f(xu)\psi(u) du}{\mathrm{vol}(\Delta_t\backslash U)} \right| \ll_f 1,$$

which follows from Cauchy-Schwarz.

However, if there is another interpretation for the Fourier coefficients, it is not always the case that Theorem 3.2 improves on "trivial" bounds arising from that interpretation.

For instance, Fourier coefficients of cusp forms on GL_n admit a spectral interpretation, that is to say, they are connected to the eigenvalues of Hecke operators. In that case, Theorem 3.2 does not give anything even approaching the bounds of Jacquet-Piatetski-Shalika. Another example is when $G = \widetilde{\mathrm{SL}_2}(\mathbb{R})$, and one takes for f the Shimura lift of a cusp form of integral weight. In that case the (absolute values of the squares of) square-free Fourier coefficients of f are given by special values of a twisted L-function; but the estimate above does not even recover the convexity bound (in fact, the method as indicated *cannot* recover this bound, even under optimal assumptions.)

It seems as though, in these cases, there is extra cancellation in the unipotent integrals for subtle arithmetic reasons. The crude methods indicated above do not detect this.

Remark 3.3. We remark that, in the proof just given, the constant κ_3 depends on the spectral gap for $\Gamma\backslash G$. This dependence can very likely be removed in many cases, including the case of $G = \mathrm{SL}_2(\mathbb{R})$, but we do not carry this out; again, cf. the last paragraph of Section 1.3.4. In the higher rank case, if G has property (T), one has in any case a uniform spectral gap and this point becomes irrelevant.

It seems worthwhile to remark that, whereas the proof above is clearly not unrelated to that of Sarnak [37], it does not require any information on the decay of triple products (in particular, the deep "exponential decay" results proved by Sarnak and Bernstein-Reznikov).

We also remark that the proof indicated above, although it can be optimized in various ways, probably does not lead to as good an exponent as the work of Good [16], and the later refinement of Sarnak's result due to Bernstein-Reznikov [1]. Its advantage lies, rather, in its robustness and general applicability.

4. **Semisimple periods: triple products in the level aspect**

In this section, we will give bounds for the triple product period on PGL_2 over a number field F. We will use the notation of Section 2; in particular, $\mathbb{A}_F$ is the adele ring of F and $\mathbf{X} = \mathrm{PGL}_2(F)\backslash\mathrm{PGL}_2(\mathbb{A}_F)$.

In Section 4.1, we will give a special case of the triple product bound (Proposition 4.1) which does not require Sobolev norms to state. For some applications we will require a slight generalization, which will require the Sobolev norms of Section 2.9.3. This will be carried out in Section 4.2 (see Proposition 4.2).

4.1. *Period bound for triple products.* We now give a period bound for triple products on PGL_2. Although it is unfortunately somewhat disguised in the adelic language, the situation and method corresponds to that of Section 1.3.1 with $G_1 = \mathrm{PGL}_2(F_S) \times \mathrm{PGL}_2(F_S)$, $G_2 = \mathrm{PGL}_2(F_S)$ embedded diagonally. Here S is a set of places of F containing all infinite places, and $F_S = \prod_{v\in S} F_v$.

For $1 \le p \le \infty$ we will write L^p for $L^p(\mathbf{X})$. Thus, e.g., $\|f_1\|_{L^4}$ denotes $\left(\int_{\mathbf{X}} |f_1(x)|^4 dx\right)^{1/4}$.

PROPOSITION 4.1 ("Subconvexity for the triple product period"). *Let π be an automorphic cuspidal representation of $\mathrm{PGL}_2(\mathbb{A}_F)$ with prime finite conductor $\mathfrak{p}$. Let $f_1, f_2 \in C^\infty(\mathbf{X})$ be totally nondegenerate*[9] *and such that f_1, f_2 are $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}})$-invariant. Let $\varphi \in \pi$ and suppose*[10] *that there exists $b \in \mathbb{R}$ such that*[11]

$$\prod_{\mathfrak{q} \in \mathrm{Supp}(\varphi) \cup \mathrm{Supp}(f_1) \cup \mathrm{Supp}(f_2)} \mathrm{N}(\mathfrak{q}) \le \mathrm{N}(\mathfrak{p})^b. \tag{4.1}$$

Put $I(\varphi) = \int_{\mathbf{X}} f_1(g) f_2(ga([\mathfrak{p}]))\varphi(g)dg$, where $a([\mathfrak{p}])$ is as in (2.4) *and dg is the $\mathrm{PGL}_2(\mathbb{A}_F)$-invariant probability measure. Then*

$$|I(\varphi)| \ll_{b,\epsilon,F} \|f_1\|_{L^4} \|f_2\|_{L^4} \|\varphi\|_{L^2} \mathrm{N}(\mathfrak{p})^{\epsilon - \frac{(1-4\alpha)(1-2\alpha)}{4(3-4\alpha)}}. \tag{4.2}$$

We refer to Proposition 4.1 as *subconvexity for the triple product period*; cf. first assertion of Theorem 5.1. We note that, with $\alpha = 3/26$ (Kim's bound) we have $\frac{(1-4\alpha)(1-2\alpha)}{4(3-4\alpha)} > 1/26$. As we will see, Proposition 4.1 is a very strong result that implies many subconvexity results on PGL(2).

First let us explain the content of Proposition 4.1 in a classical setting, and how it can be regarded as the type of period bound discussed in the introduction. Suppose $F = \mathbb{Q}$; let $p \ge 1$ and let

$$\Gamma_0(p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}) : p | c \right\},$$

and put $Y(p) = \Gamma_0(p) \backslash \mathrm{PGL}_2(\mathbb{R})$. Then there is an embedding $Y(p) \to Y(1) \times Y(1)$ which corresponds to the graph of the pth Hecke correspondence on $Y(1)$; the image is a certain closed orbit of the diagonally embedded copy of $\mathrm{PGL}_2(\mathbb{R})$. More precisely, this embedding is given by $\Gamma_0(p)g \mapsto \left(\Gamma_0(1)g, \Gamma_0(1)\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)g\right)$, and the image of $Y(p)$ projects to each factor $Y(1)$ with degree $p+1$.

Let f_1, f_2 be fixed functions on $Y(1)$ and φ a Maass form on $Y_0(p)$. Then the function $f_1 \times f_2 : (x_1, x_2) \mapsto f_1(x_1) f_2(x_2)$ is a function on $Y(1) \times Y(1)$, and we can construct its restriction $f_1 \times f_2|_{Y(p)}$ by means of the embedding indicated above. Then, translating from adelic to classical, one finds that (4.2) furnishes precisely an estimate for $\int_{Y(p)} (f_1 \times f_2)|_{Y(p)} \varphi$. When we vary p, φ and hold (f_1, f_2) fixed, such an estimate falls precisely into the pattern described in the introduction: we are computing the periods of the fixed function $f_1 \times f_2$ along the varying sequence of

---

[9] See Section 2.7; equivalent to "orthogonal to locally constant functions" in this case.

[10] Recall that a finite place v belongs to the support of $f \in C^\infty(\mathbf{X})$ exactly when $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$ does not fix f.

[11] This assumption (4.1) is purely technical and the reader may safely assume that φ is spherical at all places away from $\mathfrak{p}$ and f_1, f_2 are everywhere spherical without losing the gist of the argument. It is not used in the present document, but will probably be of use in establishing polynomial dependence of subconvex bounds. (4.1) ensures, among other things, that there are many places where all of f_1, f_2, φ are unramified, so that we can use the Hecke operators at those places.

sets $Y(p)$. The fact that the $Y(p)$ become equidistributed in $Y(1)\times Y(1)$ is precisely equivalent to the equidistribution of p-Hecke orbits on $Y(1)$. Moreover, the key property of φ that is used is the fact that φ is an eigenfunction of many Hecke operators; this is used to construct the measure σ, in the notation of Section 1.3.

Prior to beginning the proof, we make some comments about applications and generalizations; for details, we refer to Section 5. The implicit constant of (4.2) is independent of f_1, f_2. Taking f_1, f_2 to be a pair of cusp forms, Proposition 4.1 implies (conditional on some computations of p-adic integrals that we state as Hypothesis 11.1) subconvexity for certain triple product L-functions. Similarly, taking f_1, f_2 to be a cusp form and an Eisenstein series, resp. a pair of Eisenstein series, Proposition 4.1 implies subconvexity for Rankin-Selberg convolutions and standard L-functions.

The latter applications are rather delicate because Eisenstein series are not in L^4. To get around this we will eventually replace the Eisenstein series by an appropriate wave-packet (cf. proof of Theorem 5.1).

Proof. Clearly we may assume that $\|\varphi\|_{L^2} = \|f_1\|_{L^4} = \|f_2\|_{L^4} = 1$. It follows that $\|f_1\|_{L^2} \leq 1$ and $\|f_2\|_{L^2} \leq 1$.

We shall moreover assume, for simplicity, that φ is spherical at all finite places $v \neq \mathfrak{p}$. The reader may verify that the proof carries through to the more general situation of Proposition 4.1 without modification.

Put $q = \mathrm{N}(\mathfrak{p})$. Let σ be a (signed real) measure on $\mathrm{PGL}_2(\mathbb{A}_F)$ such that $\varphi \star \check{\sigma} = \lambda\varphi$, for some $\lambda \in \mathbb{C}$. We shall assume that $\mathrm{supp}(\sigma)$ commutes with $\mathrm{PGL}_2(F_\mathfrak{p})$; we will choose σ later. Set further $\Psi(x) = f_1(x) f_2(xa([\mathfrak{p}])) \in C^\infty(\mathbf{X})$. Then

$$\lambda \cdot I(\varphi) = \int_{\mathbf{X}} \Psi(x)\cdot(\varphi\star\check{\sigma})(x)dx = \int_{\mathbf{X}} (\Psi\star\sigma)(x)\cdot\varphi(x)dx \leq \left(\int_{\mathbf{X}} |\Psi\star\sigma|^2 dx\right)^{1/2} \tag{4.3}$$

$$= \left(\int_{\mathbf{X}}\int_{(g,g')\in \mathrm{PGL}_2(\mathbb{A}_F)^2} (g\cdot\Psi)\,\overline{(g'\cdot\Psi)}d\sigma(g)d\sigma(g')dx\right)^{1/2}$$

$$= \left(\int_{\mathbf{X}}\int_{(g,g')\in \mathrm{PGL}_2(\mathbb{A}_F)^2} f_1(xg)f_2(xga([\mathfrak{p}]))\overline{f_1(xg')f_2(xg'a([\mathfrak{p}]))}d\sigma(g)d\sigma(g')dx\right)^{1/2}$$

$$= \left(\int_{\mathbf{X}}\int_{(g,g')\in \mathrm{PGL}_2(\mathbb{A}_F)^2} f_1(xg)f_2(xa([\mathfrak{p}])g)\overline{f_1(xg')f_2(xa([\mathfrak{p}])g')}d\sigma(g)d\sigma(g')dx\right)^{1/2}.$$

In the last step, we have used the fact that $\mathrm{PGL}_2(F_\mathfrak{p})$, and thus $a([\mathfrak{p}])$, commutes with $\mathrm{supp}(\sigma)$.

For any two functions h_1, h_2 on $\mathbf{X}$, both right invariant by $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}})$, the assumed bound on Ramanujan (Definition 2.1) implies:

$$\left|\int_{\mathbf{X}} h_1(x)h_2(xa([\mathfrak{p}]))dx - \sum_{\chi^2=1}\chi([\mathfrak{p}])\int_{\mathbf{X}} h_1(x)\chi(x)dx\int_{\mathbf{X}} h_2(x)\chi(x)dx\right| \tag{4.4}$$
$$\leq 2q^{\alpha-1/2}||h_1||_{L^2}||h_2||_{L^2}.$$

Here χ ranges over all characters of $\mathbb{A}_F^\times/F^\times$ such that $\chi^2 = 1$, and $\chi(x)$ denotes the function $g \mapsto \chi(\det(g))$ on $\mathrm{PGL}_2(F)\backslash\mathrm{PGL}_2(\mathbb{A}_F)$. Indeed, to see (4.4), we note that the quantity inside the absolute value on the left-hand side of (4.4) equals $\langle h_1 - \mathcal{P}h_1, a([\mathfrak{p}]) \cdot (h_2 - \mathcal{P}h_2)\rangle_{L^2}$, where $\mathcal{P}$ is as in Section 2.7 (in this case, the orthogonal projection onto the locally constant functions). But the L^2-orthogonal projection $\mathrm{Id} - \mathcal{P}$ kills all one-dimensional $\mathrm{PGL}_2(F_\mathfrak{p})$ representations, from which the result follows easily.

The functions $h_j(x) = f_j(xg)\overline{f_j(xg')}$ $(j = 1, 2)$ are $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}})$-invariant for $g, g' \in \mathrm{supp}(\sigma)$ since $\mathrm{supp}(\sigma)$ commutes with $\mathrm{PGL}_2(F_\mathfrak{p})$ and $\mathfrak{p} \notin \mathrm{Supp}(f_1) \cup \mathrm{Supp}(f_2)$. Moreover, $\|h_j\|_{L^2} \le \|f_j\|_{L^4}^2$. Apply (4.4) to these h_j, and substitute in (4.3). One obtains:

$$
\begin{aligned}
(4.5)\quad |\lambda \cdot I(\varphi)|^2 &\ll q^{\alpha-1/2}\|\sigma\|^2 \\
&+ \sum_{\chi^2=1} \int_{(g,g')} |\langle g^{-1}g' \cdot f_1, f_1 \otimes \chi\rangle| \cdot |\langle g^{-1}g' \cdot f_2, f_2 \otimes \chi\rangle| \, d|\sigma|(g)\, d|\sigma|(g'),
\end{aligned}
$$

where $|\sigma|$ is the total variation measure associated to σ, $\|\sigma\| = |\sigma|(\mathbf{X})$ is the total variation of σ, $f_i \otimes \chi$ is the function $x \mapsto f_i(x)\chi(\det(x))$, and brackets $\langle\cdot,\cdot\rangle$ denote inner product in the Hilbert space $L^2(\mathbf{X})$; we will suppress the reference to $L^2(\mathbf{X})$ here and in the rest of the argument.

Put $\sigma^{(2)} = |\check{\sigma}| \star |\sigma|$. We may rewrite the previous result as

$$
\begin{aligned}
(4.6)\quad |\lambda|^2 |I(\varphi)|^2 &\ll q^{\alpha-1/2}\|\sigma\|^2 \\
&+ \left(\int_g \sum_{\chi^2=1} |\langle g \cdot f_1, f_1 \otimes \chi\rangle| \cdot |\langle g \cdot f_2, f_2 \otimes \chi\rangle| \, d\sigma^{(2)}(g) \right).
\end{aligned}
$$

We shall take σ in (4.6) to be a linear combination of Hecke operators. We follow the notation introduced in Section 2.8. For $\mathfrak{n} \notin \mathrm{Supp}(\varphi)$, we denote by $\lambda(\mathfrak{n})$ the $\mathfrak{n}$th Hecke eigenvalue of φ, i.e., $\varphi \star \mu_\mathfrak{n} = \lambda(\mathfrak{n})\varphi$. With our normalizations, the Ramanujan conjecture amounts to $|\lambda(\mathfrak{l})| \le 2$ for $\mathfrak{l}$ prime.

Let $a_\mathfrak{n}$ be a sequence of complex numbers indexed by integral ideals of $\mathfrak{o}_F$. Assume moreover that $a_\mathfrak{n} = 0$ whenever $\mathfrak{n}$ is divisible by any place in $\mathrm{Supp}(\varphi) \cup \mathrm{Supp}(f_1) \cup \mathrm{Supp}(f_2)$ and whenever $\mathrm{N}(\mathfrak{n}) \ge q$. Let σ be the measure on $\mathrm{PGL}_2(\mathbb{A}_{F,f})$ defined by $\sum_\mathfrak{n} a_\mathfrak{n}\mu_\mathfrak{n}$. Then σ is symmetric under $g \mapsto g^{-1}$, and $|\sigma| = \sum_\mathfrak{n} |a_\mathfrak{n}|\mu_\mathfrak{n}$. Moreover, $\varphi \star \sigma = \lambda\varphi$, where $\lambda = \sum_\mathfrak{n} a_\mathfrak{n}\lambda(\mathfrak{n})$. From the assumed bound on Ramanujan (see §9.1, esp. equation (9.1))[12], an elementary computation shows

$$
(4.7)\quad \left| \int_{g \in \mathrm{PGL}_2(\mathbb{A}_F)} |\langle g \cdot f_1, f_1 \otimes \chi\rangle \langle g \cdot f_2, f_2 \otimes \chi\rangle| \, d\mu_\mathfrak{n}(g) \right| \ll_\epsilon \mathrm{N}(\mathfrak{n})^{2\alpha - 1/2 + \epsilon}.
$$

[12] A small caution here is that the vectors $f_1 \otimes \chi$ and $f_2 \otimes \chi$ need not be invariant by $\mathrm{GL}_2(\mathfrak{o}_{F_\mathfrak{p}})$, if $\mathfrak{p}|\mathfrak{n}$, because χ may be ramified. However, $\mathrm{GL}_2(\mathfrak{o}_{F_\mathfrak{p}})$ always fixes the line spanned by either of these vectors, and the bound of (9.1) depends only on the dimension of the $\mathrm{GL}_2(\mathfrak{o}_{F_\mathfrak{p}})$-span of the vectors in question.

Moreover, for fixed $g \in \mathrm{Supp}(\mu_{\mathfrak{n}})$, the inner product $\langle g f_1, f_1 \otimes \chi \rangle$ is nonvanishing only if χ is unramified at those places at all places not in $\mathrm{Supp}(f_1)$ and not dividing $\mathfrak{n}$. The number of such quadratic characters is $O_\epsilon(\mathrm{N}(\mathfrak{n})^\epsilon q^\epsilon)$, where the implicit constant (as always) is allowed to depend on the base field F. Thus

$$\left| \sum_{\chi^2=1} \int_g |\langle g \cdot f_1, f_1 \otimes \chi \rangle \langle g \cdot f_2, f_2 \otimes \chi \rangle| \, d\mu_{\mathfrak{n}}(g) \right| \ll_\epsilon q^\epsilon \mathrm{N}(\mathfrak{n})^{2\alpha - 1/2 + \epsilon}. \tag{4.8}$$

The total variation of σ may be computed:

$$\|\sigma\| \ll_\epsilon \sum_{\mathfrak{n}} \mathrm{N}(\mathfrak{n})^{1/2+\epsilon} |a_{\mathfrak{n}}|. \tag{4.9}$$

Using (2.7) to compute $\sigma^{(2)}$, and combining (4.6), (4.7) and (4.9), we conclude:

$$|I(\varphi)| \ll_\epsilon q^\epsilon \frac{\left(\left(\sum_{\mathfrak{n}} \mathrm{N}(\mathfrak{n})^{1/2+\epsilon} |a_{\mathfrak{n}}| \right)^2 q^{\alpha - 1/2} + \sum_{\mathfrak{n},\mathfrak{m}} \sum_{\mathfrak{d} | (\mathfrak{n},\mathfrak{m})} \left(\mathrm{N}\left(\frac{\mathfrak{n}\mathfrak{m}}{\mathfrak{d}^2} \right) \right)^{2\alpha - 1/2} |a_{\mathfrak{n}}||a_{\mathfrak{m}}| \right)^{1/2}}{\left| \sum_n a_{\mathfrak{n}} \lambda(\mathfrak{n}) \right|}, \tag{4.10}$$

where we have absorbed various epsilons into the q^ϵ at the start.

The choice of $a_{\mathfrak{n}}$ follows an idea of Iwaniec; we slightly modify the standard choice so that we do not need to appeal to Ramanujan on average.[13] Fix K with $q^{1/1000} \le K \le q^{1000}$. Let S be the set of prime ideals $\mathfrak{l}$ such that $\mathrm{N}(\mathfrak{l}) \in [K, 2K]$ and $\mathfrak{l} \notin \mathrm{Supp}(f_1) \cup \mathrm{Supp}(f_2) \cup \mathrm{Supp}(\varphi)$. In view of the assumptions, $|S| \gg_{b,\epsilon} K^{1-\epsilon}$.

For $z \in \mathbb{C}$ we put $\mathrm{sign}(z) = z/|z|$ for $z \neq 0$ and $\mathrm{sign}(0) = 1$. Put

$$a_{\mathfrak{n}} = \begin{cases} \overline{\mathrm{sign}(\lambda(\mathfrak{n}))}, & \mathfrak{n} \in S \\ \overline{\mathrm{sign}(\lambda(\mathfrak{n}^2))}, & \mathfrak{n} = \mathfrak{l}^2, \mathfrak{l} \in S \\ 0, & \text{else.} \end{cases} \tag{4.11}$$

Then $\left| \sum_n a_{\mathfrak{n}} \lambda(\mathfrak{n}) \right| \gg_{b,\epsilon,F} K^{1-\epsilon}$, $(\sum_{\mathfrak{n}} \mathrm{N}(\mathfrak{n})^{1/2+\epsilon} |a_{\mathfrak{n}}|) \ll_\epsilon K^{2+\epsilon}$, and

$$\sum_{\mathfrak{n},\mathfrak{m}} \sum_{\mathfrak{d} | (\mathfrak{n},\mathfrak{m})} \left(\mathrm{N}\left(\frac{\mathfrak{n}\mathfrak{m}}{\mathfrak{d}^2} \right) \right)^{2\alpha - 1/2} |a_{\mathfrak{n}}||a_{\mathfrak{m}}| \ll K^{4\alpha+1}. \tag{4.12}$$

We deduce from (4.10) that

$$|I(\varphi)| \ll_{\epsilon,F} (qK)^\epsilon \frac{\left(K^4 q^{\alpha - 1/2} + K^{1+4\alpha} \right)^{1/2}}{K}. \tag{4.13}$$

Taking $K = q^{\frac{1/2-\alpha}{3-4\alpha}}$, we obtain $|I(\varphi)| \ll_{\epsilon,F} q^{\epsilon + \frac{(2\alpha-1/2)(1/2-\alpha)}{3-4\alpha}}$. □

[13] The argument that follows was improved by a suggestion of P. Michel.

4.2. *A technical generalization.* For certain applications, we shall require a slight generalization of Proposition 4.1 in which the role of $g \mapsto f_1(g) f_2(ga([\mathfrak{p}]))$ is replaced by $g \mapsto F(g, ga([\mathfrak{p}]))$, where F is a function on $\mathbf{X} \times \mathbf{X}$ that is not necessarily of product type. Although the method of proof is identical to Proposition 4.1 the details are slightly more technical; in particular, to state the result we will have need of the adelic Sobolev norms discussed in Section 2.9.3. We shall also use the notion of totally nondegenerate for functions on $\mathbf{X} \times \mathbf{X}$: see Section 2.4.

PROPOSITION 4.2. *Suppose $F \in C^\infty(\mathbf{X} \times \mathbf{X})$ is totally nondegenerate and invariant under $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}}) \times \mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}})$. Suppose moreover that there is $b \in \mathbb{R}$ with*

$$\prod_{\mathfrak{q} \in \mathrm{Supp}(F) \cup \mathrm{Supp}(\varphi)} \mathrm{N}(\mathfrak{q}) \leq \mathrm{N}(\mathfrak{p})^b. \tag{4.14}$$

Let π be a cuspidal representation of PGL_2 *over F, with conductor $\mathfrak{p}$, and put $I(\varphi) = \int_{\mathbf{X}} F(x, xa([\mathfrak{p}]))\varphi(x)dx$, for $\varphi \in \pi$. Then, for any $p > 4$, $d \gg 1$,*

$$|I(\varphi)| \ll_{b,\epsilon} \mathrm{N}(\mathfrak{p})^{-\beta+\epsilon} \|\varphi\|_{L^2} S_{p,d,2/p}(F),$$

where $\beta = \frac{(1-2\alpha)(1-4\alpha)}{p(7-4\alpha)}$.

With Kim's bound $\alpha = 3/26$, we obtain $\frac{(1-2\alpha)(1-4\alpha)}{7-4\alpha} > 1/17$.

Proof. The proof follows closely the proof of Proposition 4.1; the only difference is that we apply (2.14) (proved in Lemma 9.9) in place of (4.4). Again, we may freely assume that $\|\varphi\|_{L^2} = 1$; again we put $q = \mathrm{N}(\mathfrak{p})$.

Let notation be as in the proof of Proposition 4.1; in particular, σ is a signed real measure on $\mathrm{PGL}_2(\mathbb{A}_{F,f})$ whose support commutes with $\mathrm{PGL}_2(F_\mathfrak{p})$, and $\lambda \in \mathbb{C}$ satisfies $\varphi \star \check{\sigma} = \lambda\varphi$. Proceeding as in that proof, and in particular as in (4.3), we obtain

$$|\lambda I(\varphi)|^2 \leq \int_{\mathbf{X}} \int_{(g,g') \in \mathrm{PGL}_2(\mathbb{A}_{F,f})^2} F((x,x)(g,g)(1,a([\mathfrak{p}]))) \times \overline{F((x,x)(g',g')(1,a([\mathfrak{p}])))} d\sigma(g) d\sigma(g') i. \tag{4.15}$$

Set $F_{g,g'}(x_1,x_2) = F((x_1,x_2)(g,g))\overline{F((x_1,x_2)(g',g'))}$ for any $g, g' \in \mathrm{PGL}_2(\mathbb{A}_F)$. Then $F_{g,g'}$ is invariant by $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}}) \times \mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{p}})$ for $g, g' \in \mathrm{supp}(\sigma)$. By Hecke equidistribution in the form of (2.14) (proved in Lemma 9.9), we see that for $p > 2$, $d \gg 1$:

(4.16)

$$\left| \int_{\mathbf{X}} F_{g,g'}((x,x)(1,a([\mathfrak{p}])))dx - \sum_{\chi^2=1} \chi(\mathfrak{p}) \int_{\mathbf{X}\times\mathbf{X}} F_{g,g'}(x_1,x_2)\chi(x_1)\chi(x_2)dx_1dx_2 \right| \ll_\epsilon q^{(2\alpha-1)/p+\epsilon} S_{p,d}(F_{g,g'}).$$

Here, as before, $\chi(x)$ denotes the function on $\mathbf{X}$ defined by $g \mapsto \chi(\det(g))$. By definition

$$\left| \int_{\mathbf{X}\times\mathbf{X}} F_{g,g'}((x_1, x_2))\chi(x_1)\chi(x_2) \right| = \left| \langle (g'^{-1}g, g'^{-1}g)F, F \otimes (\chi,\chi) \rangle_{L^2(\mathbf{X}\times\mathbf{X})} \right|.$$

By the basic properties of adelic Sobolev norms ((2.12) and (2.13), proofs in Lemmas 8.1 and 8.2)

(4.17)

$$\begin{aligned} S_{p,d}(F_{g,g'}) := S_{p,d,1/p}(F_{g,g'}) &\ll S_{2p,d,1/p}((g,g)\cdot F) S_{2p,d,1/p}((g',g')\cdot F) \\ &\leq \|g\|^{2/p}\|g'\|^{2/p} S_{2p,d,1/p}(F)^2. \end{aligned}$$

Let us remark that the factors $\|g\|^{2/p}$ and $\|g'\|^{2/p}$ arises in the following way: Lemma 8.2 actually gives a factor $\|(g,g)\|^{1/p}$, where the norm $\|\cdot\|$ (as in Section 2.4) is computed in $\mathrm{PGL}_2(\mathbb{A}_{F,f})^2$; this equals $\|g\|^{2/p}$ where the norm is computed in $\mathrm{PGL}_2(\mathbb{A}_{F,f})$.

Choose σ as in the proof of Proposition 4.1 (e.g. paragraph before (4.7)) and choose the coefficients $a_{\mathfrak{n}}$ as in that proof (see (4.11)). In particular, $\|\sigma\| \ll_\epsilon K^{2+\epsilon}$. Since F is totally nondegenerate, the matrix coefficients $\langle (g'^{-1}g, g'^{-1}g)F, F\rangle$ satisfy bounds that are of the same quality as in the proof of Proposition 4.1; in particular, as in (4.8),

$$\sum_{\chi^2=1} \int_g |\langle (g,g)F, F\otimes(\chi,\chi)\rangle|\, d\mu_{\mathfrak{n}}(g) \ll q^\epsilon \mathrm{N}(\mathfrak{n})^{2\alpha-1/2+\epsilon}\|F\|^2.$$

Finally, $\|g\| \ll_\epsilon K^{2+\epsilon}$ for all $g \in \mathrm{supp}(\sigma)$. Proceeding just as in the previous proof,

$$|I(\varphi)| \ll_{b,\epsilon} (qK)^\epsilon \frac{\left(K^4 q^{(2\alpha-1)/p} K^{8/p} S_{2p,d,1/p}(F)^2 + K^{1+4\alpha}\|F\|^2_{L^2(\mathbf{X}\times\mathbf{X})}\right)^{1/2}}{K}.$$

Consequently, for any $p > 2$,

$$|I(\varphi)| \ll_{b,\epsilon} (qK)^\epsilon \frac{\left(K^8 q^{(2\alpha-1)/p} S_{2p,d,1/p}(F)^2 + K^{1+4\alpha}\|F\|^2_{L^2}\right)^{1/2}}{K}.$$

To conclude, choose $K = q^{\frac{1-2\alpha}{p(7-4\alpha)}}$ and replace p by $p/2$ (thus, e.g., $p > 2$ becomes $p > 4$). □

5. Application to L-functions

We now present the first applications to subconvexity. The rough idea is simply that certain L-functions are expressed as period integrals of the type that are bounded by Propositions 4.1 and 4.2. There is one significant issue in implementing this (rather evident) idea: namely, the integral representation that we use for

Rankin-Selberg and the standard L-functions involve Eisenstein series, which are not in L^2; this causes problems in applying Proposition 4.1!

Thus we need to *regularize*. Two natural ways of doing this are to replace an Eisenstein series by a "wave-packet"; or to use a suitable form of truncation in the defining integrals. In the present paper we will use the wave-packet technique; in the paper [28] we shall also use truncation.

Let us briefly describe the wave packet technique in a classical language. Roughly speaking, we can express the Rankin-Selberg L-function of two classical forms f, g via an integral of the form $L(s, f \times g) = \int_z f(z)g(z)E(s,z)$, for some Eisenstein series $E(s)$. We now regularize, replacing $E(s,z)$ by a wave packet. Let $h(s)$ be any holomorphic function: then

$$\int_{Re(s)=1/2} h(s)L(s, f \times g)ds = \int_z f(z)g(z) \int_{\Re(s)=1/2} h(s)E(s,z). \tag{5.1}$$

We wish to eventually recover an upper bound for $L(1/2, f \times g)$ (say) from the left-hand side, so we take $h(s) = \overline{L(1-\overline{s}, f \times g)}$. Then $h(s)L(s, f \times g)$ is positive along $\Re(s) = 1/2$. To apply Proposition 4.1 to the right-hand side of (5.1), we shall moreover need to control the behavior of the regularized Eisenstein series $E_h = \int_{\Re(s)=1/2} h(s)E(s,z)$; this type of analysis is carried out in Sections 10.2 and 10.3, the main point being that the divergence of the Eisenstein series comes entirely from the constant term.

It is worth remarking that Iwaniec's bounds for the L-function near 1 enter rather crucially in this analysis: in effect, we bound E_h by an easy argument involving shift of contours; this necessitates that h be estimated on a line $\Re(s) = -\varepsilon$, which amounts to estimating $L(s, f \times g)$ for $\Re(s) = 1 + \varepsilon$.

In what follows we have not attempted to obtain polynomial dependence in all parameters. This is not hard to do — and, at its essence, a statement that one can find *analytically* suitable test vectors in a Rankin-Selberg integral; but we have not done so here. On the other hand, we give full details of this procedure in the proof of Theorem 6.1 (in which the polynomial dependence is particularly useful for applications).

THEOREM 5.1. *Let π_1, π_2 be fixed automorphic cuspidal representations of* PGL_2 *over F; fix $t \in \mathbb{R}$. Let π be an automorphic cuspidal representation with conductor* $\mathfrak{p}$, *a prime ideal that is prime to the conductors of π_1 and π_2.*

Then, assuming Hypothesis 11.1,

$$L\Big(\frac{1}{2}, \pi_1 \otimes \pi_2 \otimes \pi\Big) \ll_{\pi_\infty} \mathrm{N}(\mathfrak{p})^{1-\frac{1}{13}} \tag{5.2}$$

and, unconditionally,

$$\Big|L\Big(\frac{1}{2} + it, \pi_1 \otimes \pi\Big)\Big|^2 \ll_{\pi_\infty} \mathrm{N}(\mathfrak{p})^{1-\frac{1}{100}}, \tag{5.3}$$

$$\Big|L\Big(\frac{1}{2} + it, \pi\Big)\Big|^4 \ll_{\pi_\infty} \mathrm{N}(\mathfrak{p})^{1-\frac{1}{600}}. \tag{5.4}$$

In these statements, the notation $\ll_{\pi_\infty}$ *indicates an implicit constant that depends continuously on the local archimedean representation* π_∞ *of* $\mathrm{GL}_2(F_\infty)$ *underlying* π.

Note we make no claim about the dependency of the implicit constant on t, π_1, π_2; as remarked above, this dependence could be made polynomial in the conductors, but this would require more careful analysis of the archimedean integrals.[14]

We remark that we have used H. Kim's bound $\alpha = 3/26$; any value of α less than $1/4$ would give subconvexity and under Ramanujan one obtains for (5.2) the exponent $5/6$. The exponents for (5.3) and (5.4) can be improved; e.g., the present proof does not take into account the fact that unitary Eisenstein series satisfy Ramanujan!

5.1. *Results relating periods and integral representations.* For the convenience of the reader, we summarize here the results that relate periods and integral representations (proved in later sections). Roughly speaking, any integral representation for an L-function expresses it as a period integral with certain test vectors belonging to the space of an automorphic cuspidal representation.

A delicate point, which is quite relevant to issues of polynomial dependence in auxiliary parameters, is precisely *which* test vectors. In principle, the proofs of results about integral representation give explicit test vectors. In practice, it is tedious to extract these explicit test vectors. Our policy throughout this paper is the lazy one: to deduce results, as far as possible, by formal arguments and without choosing explicit test vectors. The price of this is that we will obtain not quite the L-function, but rather a holomorphic function that differs from the L-function by some harmless factors.

More precisely, the content of the proposition (Proposition 5.1) that follows is that one can write down an integral representation $I(s)$ for the L-functions of interest, so that:

(1) $I(1/2)$ is not too much smaller than $L(1/2)$ — or with $1/2$ replaced by the point of interest — so that a bound for $I(1/2)$ gives a bound for $L(1/2)$.

(2) $I(s)$ is not too much bigger than $L(s)$ for *any* s. This type of control will be useful in shifting contours.

One might prefer to get $I(s) = L(s)$ but we do not need this stronger statement.

As is discussed at length in Section 10, to a Schwartz function Ψ on $\mathbb{A}_F^2$ is associated a family of Eisenstein series $E_\Psi(s, g)$ on $\mathbf{X}$, which varies meromorphically in the parameter $s \in \mathbb{C}$.

[14]It is important to note, however, that this is an entirely local problem; it is intended that this will be carried out in a more general context in [28]. Both for applications and to illustrate procedure, we have carried out this type of analysis for the results on subconvexity of character twists in Section 6. Those results are proved with polynomial dependence on all parameters.

PROPOSITION 5.1. *Let* $s_0, t_0, t'_0 \in \mathbb{C}$. *Let* π_1 *be a fixed automorphic cuspidal representation of* $\mathrm{PGL}_2(\mathbb{A}_F)$ *and* π *an automorphic cuspidal representation of prime conductor* $\mathfrak{p}$; *assume that the finite conductors of* π, π_1 *are coprime.*

Denoting by π_∞ *the representation of* $\mathrm{PGL}_2(F_\infty)$ *corresponding to* π, *suppose that* $\mathrm{Cond}(\pi_\infty)$ *is bounded above*; *equivalently*, π_∞ *belongs to a bounded subset*[15] *of the dual* $\widehat{\mathrm{PGL}_2}(F_\infty)$ (*in what follows the implicit constants may depend on these bounds*).

There exists a fixed finite set $\mathcal{F}$ *of Schwartz Bruhat functions on* $\mathbb{A}_F^2$ *and a real number* $C > 0$ *so that*:

There exist vectors $\varphi_1 \in \pi_1, \varphi \in \pi$ *and* $\Psi \in \mathcal{F}$ *so that*

$$\Phi(s) := \mathrm{N}(\mathfrak{p})^{1-s} \frac{\int_{\mathbf{X}} \varphi(g)\varphi_1(ga([\mathfrak{p}]))E_\Psi(s,g)dg}{\Lambda(s, \pi_1 \otimes \pi)}$$

is holomorphic and satisfies:

(1) $|\Phi(s_0)| \gg 1$ *and* $|\Phi(s)| \ll C^{|\Re(s)|}(1+|s|)^C$.

(2) *At any nonarchimedean place* v *such that* π_1 *is unramified*, φ_1 *and* Ψ *are invariant by* $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$; *at any nonarchimedean place* v *such that* π_1 *and* π *are both unramified*, φ, φ_1 *are both invariant by* $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$.

(3) $\|\varphi_1\|_{L^\infty} \ll 1$ *and* $\|\varphi\|_{L^2(\mathbf{X})} \ll_\epsilon \mathrm{N}(\mathfrak{p})^\epsilon$.

Moreover, *there exist vectors* $\varphi \in \pi$, $\Psi_1, \Psi_2 \in \mathcal{F}$ *so that*:

$$\Phi(t,t') = \mathrm{N}(\mathfrak{p})^{1/2-t} \frac{\int_{\mathbf{X}} \varphi(g)E_{\Psi_1}(g, \frac{1}{2}+t)E_{\Psi_2}(ga([\mathfrak{p}]), \frac{1}{2}+t')dg}{\Lambda(\frac{1}{2}+t+t', \pi)\Lambda(\frac{1}{2}+t-t', \pi)} \tag{5.5}$$

is holomorphic and satisfies:

(1) $|\Phi(t_0, t'_0)| \gg 1$ *and* $|\Phi(t,t')| \ll C^{|\Re(t)|+C|\Re(t')|}(1+|t|+|t'|)^C$.

(2) *For any nonarchimedean place* v, *each* Ψ_1 *and* Ψ_2 *is invariant by* $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$; *for each place at which* π *is unramified*, *the same is true of* φ.

(3) $\|\varphi\|_{L^2(\mathbf{X})} \ll_\epsilon \mathrm{N}(\mathfrak{p})^\epsilon$.

Proof. Lemmas 11.4 and 11.5. □

In effect, we could achieve "$\Phi = 1$" in Proposition 5.1 by a more careful choice of local data; but this is irrelevant for the purpose of global estimation.

5.2. *Proof of Theorem* 5.1.

Proof of (5.2). It follows from Hypothesis 11.1 and Proposition 4.1.

[15] See Section 2.12.3 for the definition.

Proof of (5.3). The basic idea is that the Rankin-Selberg convolution is a triple product, with one factor being an Eisenstein series. However, one cannot naively apply Proposition 4.1 since Eisenstein series do not belong to $L^4(\mathbf{X})$. To avoid this, we will use a wave-packet of Eisenstein series.

First, we can assume from the start that π_∞ belongs to a bounded subset of the dual $\widehat{\mathrm{PGL}_2(F_\infty)}$. The implicit constants in the proof that follow depend on this subset. We denote by Λ the completed L-function. We begin by remarking that since $\mathrm{N}(\mathfrak{p}) \to \infty$ we may assume that π_1 is not isomorphic to π, or to any quadratic twist of π. In particular, we are free to assume that $\Lambda(s, \pi_1 \otimes \pi)$ has no poles. Moreover, the finite conductor of $\Lambda(s, \pi_1 \otimes \pi)$ differs from $\mathrm{N}(\mathfrak{p})^2$ by an absolutely bounded constant.

Fixing $t_0 \in \mathbb{R}$, let $\Psi, \varphi, \varphi_1, E_\Psi(g, s), \Phi$ be as in Proposition 5.1 with $s_0 = 1/2 + it_0$, so that $|\Phi(1/2 + it_0)| \gg 1$. For simplicity we simply write $E(g, s)$ for $E_\Psi(g, s)$. Fix $\kappa > 0$. In the rest of the proof we omit the subscript κ, ϵ from $\ll$, with the understanding that all implicit constants depend on κ and ϵ. Put

$$I(s) = \mathrm{N}(\mathfrak{p})^{s-1} \Lambda(s, \pi_1 \otimes \pi) \Phi(s) = \int_{\mathbf{X}} \varphi_1(ga([\mathfrak{p}])) E(s, g) \varphi(g) dg.$$

From Iwaniec's upper bounds for L-functions [19, Th. 8.3], the functional equation for Λ, and the bounds on Φ furnished by Proposition 5.1,

$$(5.6)\quad |I(1 + \kappa + it)| \ll (1 + |t|)^{-6} \mathrm{N}(\mathfrak{p})^{\kappa+\epsilon}, |I(-\kappa + it)| \ll (1 + |t|)^{-6} \mathrm{N}(\mathfrak{p})^{\kappa+\epsilon}.$$

Put $h(s) = s(1-s)(s-\frac{1}{2})^2 \overline{I(1-\overline{s})}$. Then $h(s)$ is holomorphic in $-\kappa \le \Re(s) \le 1 + \kappa$ and $h(\frac{1}{2}) = 0$.[16] Moreover, $h(s)$ has rapid decay as $\Im(s) \to \infty$, in view of the Γ-factors of the completed L-function. Put $E_h(g) = \int_{\Re(s)=1+\kappa} h(s) E(s, g)$. It is proved in Lemma 10.6 that, for such h,

$$\|E_h(g)\|_{L^\infty} \ll \int_{-\infty}^{\infty} (|h(-\kappa + it)| + |h(1 + \kappa + it)|)\, dt.$$

Applying (5.6), we conclude that $\|E_h(g)\|_{L^\infty} \ll \mathrm{N}(\mathfrak{p})^{\kappa+\epsilon}$.

On the other hand, we see from the definition of $I(s)$ that

$$(5.7)\quad \int_{\Re(s)=1+\kappa} h(s) I(s) = \int_{\Re(s)=1+\kappa} h(s) ds \int_{\mathbf{X}} \varphi_1(ga([\mathfrak{p}])) E(s, g) \varphi(g) dg.$$

[16] The fact that we impose $h(1/2) = 0$ has a very concrete meaning in classical terms. Fix, for example, a form f and $t \in \mathbb{R}$. Consider $\sum_g |L(\frac{1}{2} + it, f \otimes g)|^2$ where the sum is taken over a basis of holomorphic Hecke eigenforms of level N and trivial Nebentypus. If $t = 0$, this has the asymptotic behavior $N \log(N)^3$. On the other hand, if $t \neq 0$, it behaves like $N \log(N)$. Forcing $h(1/2) = 0$ "counteracts" this extra singularity.

The double integral on the right-hand side of (5.7) is absolutely convergent and orders may be switched; thus

$$\int_{\Re(s)=1+\kappa} h(s)I(s) = \int_{\mathbf{X}} \varphi_1(ga([\mathfrak{p}]))E_h(g)\varphi(g)dg \tag{5.8}$$
$$= \int_{\mathbf{X}} \varphi_1(ga([\mathfrak{p}]))E_h^0(g)\varphi(g)dg,$$

where $E_h^0 := \mathscr{P}(E_h)$ is totally nondegenerate (see §2.7) and satisfies $\|E_h^0\|_{L^\infty} \ll_\epsilon \mathrm{N}(\mathfrak{p})^{\kappa+\epsilon}$.

We then deduce from Proposition 4.1 that

$$\left|\int_{\Re(s)=1+\kappa} h(s)I(s)\right| \ll \|\varphi_1\|_{L^4(\mathbf{X})}\mathrm{N}(\mathfrak{p})^{-\frac{1}{26}+\kappa+\epsilon}. \tag{5.9}$$

$I(s)$ and $h(s)$ both decay exponentially rapidly as $\Im(s) \to \infty$. It is therefore simple to justify shifting the line of integration in (5.9) to $\Re(s) = 1/2$. We deduce thereby that

$$\left|\int_{\Re(s)=\frac{1}{2}} t^2\left|I\left(\frac{1}{2}+it\right)\right|^2 dt\right| \ll \|\varphi_1\|_{L^4(\mathbf{X})}\mathrm{N}(\mathfrak{p})^{\kappa-\frac{1}{26}+\epsilon}. \tag{5.10}$$

From (5.6) we deduce bounds on I and I' inside the strip $0 \le \Re(s) \le 1$ by the maximal modulus principle. In particular,

$$\left|I'\left(\frac{1}{2}+it\right)\right| \ll_t \mathrm{N}(\mathfrak{p})^{\kappa+\epsilon}. \tag{5.11}$$

Combining (5.10) and (5.11), and recalling that κ is arbitrary, we obtain $|I(1/2+it)| \ll_t \mathrm{N}(\mathfrak{p})^{-\frac{1}{5\cdot 26}+\epsilon}$. Thus $|\Lambda(\frac{1}{2}+it_0, \pi_1 \otimes \pi)| \ll_{\epsilon,t_0} \mathrm{N}(\mathfrak{p})^{\frac{1}{2}-\frac{1}{130}+\epsilon}$. $\square$

Proof of (5.4). The proof is similar to that of (5.3), but a slightly more elaborate regularization is required, since we shall proceed from the expression (5.5) of $L(s,\pi)$ as a triple product against *two* Eisenstein series. Again we may assume from the start that π_∞ is confined to a bounded subset of $\widehat{\mathrm{PGL}_2(F_\infty)}$; the implicit constants will, again, depend on this subset.

Let $\Lambda(s,\pi)$ be the completed L-function attached to π. Fixing $t_0, t_0' \in i\mathbb{R}$, Proposition 5.1 gives the existence of Eisenstein series $E_{\Psi_1}(g,s) = E_1(g,s)$ and $E_{\Psi_2}(g,s) = E_2(g,s)$ on $\mathbf{X}$, and $\varphi \in \pi$ so that

$$\Phi(t,t') := \mathrm{N}(\mathfrak{p})^{1/2-t}\frac{\int_{\mathbf{X}} \varphi(g)E_1(g,\frac{1}{2}+t)E_2(ga([\mathfrak{p}]),\frac{1}{2}+t')dg}{\Lambda(\frac{1}{2}+t+t',\pi)\Lambda(\frac{1}{2}+t-t',\pi)}$$

satisfies $|\Phi(t_0,t_0')| \gg 1$ and $\Phi(t,t') \ll C^{|\Re(t)|+|\Re(t')|}(1+|t|+|t|'|)^C$.

We put

$$I(z_1,z_2) = \Phi(z_1,z_2)\mathrm{N}(\mathfrak{p})^{z_1-1/2}\Lambda\left(\tfrac{1}{2}+z_1+z_2,\pi\right)\Lambda\left(\tfrac{1}{2}+z_1-z_2\right) \tag{5.12}$$
$$= \Phi(z_1,z_2)\mathrm{N}(\mathfrak{p})^{\frac{z_1+z_2}{2}-1/4}\Lambda\left(\tfrac{1}{2}+z_1+z_2,\pi\right)\mathrm{N}(\mathfrak{p})^{\frac{z_1-z_2}{2}-1/4}\Lambda\left(\tfrac{1}{2}+z_1-z_2,\pi\right).$$

Then $I(z_1, z_2)$ is a holomorphic function of $(z_1, z_2) \in \mathbb{C}^2$. $I(z_1, z_2)$ has rapid decay along "vertical lines"; that is, for σ, σ' in a fixed compact set and $(t, t') \in \mathbb{R}$ we have $I(\sigma + it, \sigma' + it') \ll_N (1 + |t| + |t'|)^{-N}$.

Let $\kappa > 0$ be fixed. From (5.12), Iwaniec's bounds for L-functions near 1, and the rapid decay of I along "vertical lines," we obtain by the maximal modulus principle:

$$(1 + |z_1| + |z_2|)^N \max(|I(z_1, z_2)|, |\partial_1 I(z_1, z_2)|, |\partial_2 I(z_1, z_2)|) \ll_N \mathrm{N}(\mathfrak{p})^{\kappa},$$
$$|\Re(z_1)| + |\Re(z_2)| \leq 1/2 + \kappa,$$

where ∂_1 (resp. ∂_2) is the operator of differentiation with respect to z_1 (resp. z_2). Put

$$h(z_1, z_2) = z_1^2 z_2^2 (1/4 - z_1^2)(1/4 - z_2^2)\overline{I(-\overline{z_1}, -\overline{z_2})}. \tag{5.13}$$

Then, in the notation of Section 10.3 (esp. Definition 10.1) , h belongs to the space $\mathcal{H}^{(2)}(\kappa)$ and satisfies $\|h\|_N \ll_N \mathrm{N}(\mathfrak{p})^{\kappa}$.

Put $I = \int_{\Re(z_1)=\Re(z_2)=0} h(z_1, z_2) I(z_1, z_2) dz_1 dz_2$. Then

$$\begin{aligned} I &= \int_{\Re(z_1)=0, \Re(z_2)=0} h(z_1, z_2) dz_1 dz_2 \\ &\quad \cdot \int_{\mathbf{X}} \varphi(g) E_1(g, 1/2 + z_1) E_2(ga([\mathfrak{p}]), 1/2 + z_2) dg \\ &= \int_{\mathbf{X}} \varphi(x) E_h((x, x)(1, a([\mathfrak{p}]))) dx, \end{aligned} \tag{5.14}$$

where the function E_h on $\mathbf{X} \times \mathbf{X}$ is defined by

$$E_h(g_1, g_2) = \int_{\Re(z_1)=0, \Re(z_2)=0} h(z_1, z_2) E_1(g_1, 1/2 + z_1) E_2(g_2, 1/2 + z_2) dz_1 dz_2,$$

and the interchange of orders is justified by the (easily verified) absolute convergence of the double integral defining I. Note that $E_h(g_1, g_2)$ is totally nondegenerate (see Section 2.7 for the definition.). We now apply Proposition 4.2 to conclude that $|I| \ll_{p,d} S_{p,d,2/p}(E_h) \|\varphi\|_{L^2} \mathrm{N}(\mathfrak{p})^{-\frac{1}{17p}}$ for any $p > 4, d \gg 1$. We note at this point that the requirement $p > 4$ makes it critical that the regularized Eisenstein series E_h belong to L^4; the trivial fact that Eisenstein series belong to $L^{2-\epsilon}$ is far from sufficient.

On the other hand, by Lemma 10.9, $S_{p,d,2/p}(E_h) \ll \|h\|_N$ for some N (possibly depending on p, d) and all $p < \frac{4}{1-2\kappa}$. By Proposition 5.1, $\|\varphi\|_{L^2} \ll_\epsilon \mathrm{N}(\mathfrak{p})^{\epsilon}$.

Thus $|I| \ll_\epsilon \mathrm{N}(\mathfrak{p})^{\epsilon - 1/68}$. Now, by the definition of h (5.13) we have $I = \mathrm{N}(\mathfrak{p})^{-1} \int_{(t,t') \in \mathbb{R}^2} (1/4 + t_1^2)(1/4 + t_2^2) t_1^2 t_2^2 |I(it_1, it_2)|^2 dt_1 dt_2$. Thus we obtain

$$\int_{(t,t') \in \mathbb{R}^2} |I(it_1, it_2)|^2 t_1^2 t_2^2 dt_1 dt_2 \ll_\epsilon \mathrm{N}(\mathfrak{p})^{1 - 1/68 + \epsilon}. \tag{5.15}$$

Using (5.15), and the given properties of Φ, we deduce that

$$\left|\Lambda\Big(\frac{1}{2}+t_0+t_0'\Big)\Lambda\Big(\frac{1}{2}+t_0-t_0'\Big)\right|^2 \ll_{t_0,t_0'} \mathrm{N}(\mathfrak{p})^{1-1/600}$$

in a similar fashion to the conclusion of the proof of (5.3). We take $t_0' = 0$ to conclude. □

6. **Torus periods (I): subconvex bounds for character twists over a number field**

In this section we shall work in considerable generality; we shall derive subconvex bounds without any assumptions of prime or squarefree conductor, and obtaining polynomial dependence in all auxiliary parameters. This is useful for applications, but will involve some notational overhead. As a result, we have sacrificed good exponents for simplicity at many steps.

THEOREM 6.1. *Let* π *be a* (*unitary*) *cuspidal representation of* $\mathrm{GL}_2(\mathbb{A}_F)$, *and* χ *a unitary character of* $\mathbb{A}_F^\times/F^\times$, *with finite conductor* $\mathfrak{f}$. *Then there is* $N > 0$ *such that*

$$L\big(\tfrac{1}{2}, \pi \times \chi\big) \ll \mathrm{Cond}(\pi)^N \mathrm{Cond}_\infty(\chi)^N \mathrm{N}(\mathfrak{f})^{1/2-\frac{1}{24}}, \tag{6.1}$$

$$L\big(\tfrac{1}{2}, \chi\big) \ll \mathrm{Cond}_\infty(\chi)^N \mathrm{N}(\mathfrak{f})^{1/4-\frac{1}{200}}. \tag{6.2}$$

Note that the result also implies a corresponding statement for the L-functions evaluated at $\frac{1}{2}+it$, since one may replace χ by $\chi|\cdot|^{it}$. Over $\mathbb{Q}$ this result is due to [11]; presently the best known exponent was achieved in [4].

Since it is perhaps hidden in the proof where the polynomial dependence on conductor arises, we would like to explicate it now. If π is an automorphic cuspidal representation with analytic conductor $\mathrm{Cond}(\pi)$, there exists a vector $\psi \in \pi$ with Sobolev norms $S_{2,d,\beta}(\psi) \ll \mathrm{Cond}(\pi)^{\mathrm{const}\max(\beta,d)}$. Moreover, one can choose such a ψ to be a "good" test vector with respect to certain toral periods. Thus the analytic conductor enters precisely through the minimal Sobolev norm of a suitable vector belonging to the space of π. We note that the test vectors we choose are smooth but not K-finite at infinite places; this idea has been heavily exploited in the previous work of Bernstein and Reznikov.

Note that some cases of Theorem 6.1 — where π has trivial central character and π is quadratic — are subsumed by the previous result Theorem 5.1. Nevertheless, we have chosen to give a distinct presentation since the method is entirely different, it is simpler in the present method to deal with the case of noncuspidal π. Also, we shall consistently deal in the present section with GL_2, rather than PGL_2. Thus ω will be a unitary character of $\mathbb{A}_F^\times/F^\times$, and $C_\omega^\infty(\mathbf{X}_{\mathrm{GL}(2)})$ the space of functions on $\mathbf{X}_{\mathrm{GL}(2)} = \mathrm{GL}_2(F)\backslash\mathrm{GL}_2(\mathbb{A}_F)$ with central character ω.

In the case $F = \mathbb{Q}$, the subconvexity result (6.2) for characters is due to Burgess [6]. Burgess' method gives a much better exponent; of course there is considerable scope for improvement in the present technique also.

For the ease of the reader, we briefly explain in advance the points of our proof in classical language. The discussion that follows is not a completely faithful rendition of the proof, but it hopefully conveys the main ideas. While it follows the pattern of all the proofs in this paper, one minor complication is that we deal with integrals with respect to certain measures of infinite mass.

(1) If f is a Maass form on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, then the integral

$$\frac{1}{q}\int_{y=0}^{\infty}\sum_{x=1}^{q}\chi(x)f\Big(\frac{x}{q}+iy\Big)\frac{dy}{y} \tag{6.3}$$

equals, up to some Γ-factors, $\frac{1}{\sqrt{q}}L(\frac{1}{2},f\times\chi)$. This is an exercise in Hecke-Jacquet-Langlands theory. The version of this equality that we shall use is proved in Lemma 11.8 when f is a cusp form and Lemma 11.10 when f is Eisenstein.

(2) It will then suffice to bound $\sum_{x=1}^{q}\chi(x)f(\frac{x}{q}+iy)$ for each fixed value of y. As it turns out, the crucial range of y is around $y=q^{-1}$; the contribution of other ys are small for relatively trivial reasons (use the Fourier expansion). This is roughly a geometric form of the approximate functional equation: it says that the Fourier coefficients $a_n(f)$ with $n\asymp q$ are most important to determining the L-function. The general version of this fact is proven in Lemma 11.9.

(3) In the range when $y\asymp q^{-1}$, the set $\{\frac{x}{q}+iy\}_{\{1\le x\le q-1\}}$ is roughly equidistributed, because it is (with the exception of two points) the orbit of $iqy\in\mathbb{H}$ by the qth Hecke operator. This is easy to quantify and actually can be regarded as a statement about equidistribution of *p-adic* horocycles.The general version of this is proved in Lemma 9.11.

(4) We are now in a situation where we are trying to bound the period of f along the roughly equidistributed set $\{\frac{x}{q}+iy\}_{\{1\le x\le q-1\}}$. To do this, we apply mixing properties of the adelic torus flow, in the same fashion as the previous proofs of this paper. This shows that $\sum_{x=1}^{q}\chi(x)f(\frac{x}{q}+iy)$ is small.

The computations that underlie steps (1), (2) and (3) are fairly routine but technically complicated. We have therefore carried them out in Section 11.4. In the sections that follow, we merely quote the results and carry out what amounts to step (4).

6.1. *Relating integral representations and periods.* Let $z\in\mathbb{R}$ and let μ_z,ν_z,μ,ν be the measures on $\mathbf{X}_{\mathrm{GL}(2)}$ defined by

$$\begin{aligned}\mu_z(f)&=\int_{|y|=z}f(a(y)n([\mathfrak{f}]))\chi(y)d^{\times}y, &\mu&=\int_{z>0}\mu_z d^{\times}z,\\ \nu_z(f)&=\int_{|y|=z}f(a(y)n([\mathfrak{f}]))d^{\times}y, &\nu&=\int_{z>0}\nu_z d^{\times}z.\end{aligned} \tag{6.4}$$

In both cases, the measure $d^\times y$ is the probability measure invariant by $\mathbb{A}_F^1/F^\times$ and the measure $d^\times z$ is a Haar measure on $\mathbb{R}^\times$. Thus μ_z, ν_z are probability measures, whereas μ, ν have infinite mass. It is simple to see that the integrals defining $\mu(f), \nu(f)$ converge absolutely if f is a function decaying rapidly enough at the cusps, e.g. satisfying $|f(x)| \ll \mathrm{ht}(x)^{-\varepsilon}$ (notation of §8.2), for any $\varepsilon > 0$. Note also the analogy between these measures and those used in the analysis of unipotent periods (cf. (3.2).) Classically, $\nu_z(f)$ should be thought of the measure on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ defined by $\sum_{0\le x\le q-1} f(\frac{x}{q}+iz)$, and $\mu_z(f)$ the measure on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ defined by $\sum_{0\le x\le q-1} f(\frac{x}{q}+iz)\chi(x)$. (These statements are not to be interpreted precisely; they are for intuition only.)

Here is the proposition that formalizes (1) and (2) of the discussion above, in the cuspidal case.

PROPOSITION 6.1. *Let π be a cuspidal representation on $\mathrm{GL}_2(\mathbb{A}_F)$, χ a character of $\mathbb{A}_F^\times/F^\times$ of finite conductor $\mathfrak{f}$. Write*

$$L_{\mathrm{unr}}(s, \pi\times\chi) = \prod_{\chi_v \text{ unram.}} L_v(s, \pi\times\chi),$$

where the product is taken over all finite places at which χ is not ramified.

Let $d, \beta \ge 0$. Let g_+, g_- be positive smooth functions on $\mathbb{R}_{\ge 0}$, with g_- nonincreasing and g_+ nondecreasing, such that $g_+ + g_- = 1$, $g_+(t) = 1$ for $t \ge 2$ and $g_-(t) = 1$ for all $t \le 1/2$.

Then there exists $\varphi \in \pi$ such that, with

$$\Phi(s) = \mathrm{N}(\mathfrak{f})^{1/2}\frac{\int_z \mu_z(\varphi) z^{s-1/2} d^\times z}{L_{\mathrm{unr}}(s, \pi\times\chi)}, \tag{6.5}$$

then $\Phi(s)$ is holomorphic and satisfies:

(1) $|\Phi(s)| \ll_{\Re(s),\epsilon} \mathrm{N}(\mathfrak{f})^\epsilon$ *and* $|\Phi(\frac{1}{2})| \gg_\epsilon \mathrm{N}(\mathfrak{f})^{-\epsilon}$.

(2) *φ is new at every finite place (i.e., for each finite prime $\mathfrak{q}$ it is invariant by $K_0[\mathfrak{q}^{s_\mathfrak{q}}]$, where $s_\mathfrak{q}$ is the local conductor of π).*

(3) *The Sobolev norms of φ satisfy the bounds*

$$S_{2,d,\beta}(\varphi) \ll_\epsilon \mathrm{Cond}_\infty(\pi)^{2d+\epsilon}\mathrm{Cond}_f(\pi)^{\beta+\epsilon}\mathrm{Cond}_\infty(\chi)^{1/2+2d}. \tag{6.6}$$

(4) *The integration of (6.5) may be "truncated without significant change" to the region z around $\mathrm{N}(\mathfrak{f})^{-1}$; more formally,*

$$\left|\int_z \mu_z(\varphi) g_+(z/T) d^\times z\right| \ll (\mathrm{N}(\mathfrak{f})T)^{-1/2}(T\mathrm{Cond}(\pi)\mathrm{Cond}(\chi))^\epsilon,$$

$$\left|\int_z \mu_z(\varphi) g_-(z/T) d^\times z\right| \ll (\mathrm{N}(\mathfrak{f})T)^{1/2}(T\mathrm{Cond}(\chi))^\epsilon(\mathrm{Cond}_\infty(\chi)\mathrm{Cond}(\pi))^{1+\epsilon}.$$

Proof. Lemmas 11.8 and 11.9. □

We next give the corresponding result for the "noncuspidal case." We recall that the Eisenstein series $E_\Psi(s,g)$ associated to a Schwartz function Ψ on $\mathbb{A}_F^2$ are discussed in Section 10. The normalization is so that the functional equation interchanges s and $1-s$. $\bar{E}(s,g)$ denotes, as explained in that section (cf. (10.9)) the truncated Eisenstein series obtained by subtracting the constant term; it is a function on $B(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)$.

PROPOSITION 6.2. *Suppose that χ is ramified at least one finite place. Let $g_\pm$ be as in Proposition* 6.1. *There is an absolute $C>0$ (i.e., depending only on F) and a choice of $K_{\max}$-invariant Schwartz function Ψ (depending on χ) so that if we put*

$$\Phi(s,s') := \mathrm{N}(\mathfrak{f})^{1/2} \frac{\int_{y\in\mathbb{A}_F^\times/F^\times} \bar{E}_\Psi(s,a(y)n([\mathfrak{f}]))\chi(y)|y|^{s'}d^\times y}{L(\chi,s+s')L(\chi,1-s+s')},$$

where $\bar{E}$ is defined as in (10.9), *then the integral defining Φ is absolutely convergent when $\Re(s),\Re(s')\gg 1$. Moreover, Φ extends from $\Re(s),\Re(s')\gg 1$ to a holomorphic function on $\mathbb{C}^2$, satisfying*:

(1) $|\Phi(1/2,0)|\gg 1$ *and* $|\Phi(s,s')|\ll C^{1+|\Re(s)|+|\Re(s')|}(1+|s|+|s'|)^C$. *Moreover, given $N>0$ we have that*

$$|\Phi(s,s')|(1+|s|+|s'|)^N \ll_{\Re(s),\Re(s'),N} \mathrm{Cond}_\infty(\chi)^{N'}, \tag{6.7}$$

where N' and the implicit constant may be taken to depend continuously on $N,\Re(s),\Re(s')$.

(2) Ψ, *and so also $E_\Psi(s,g)$ is invariant by $K_{\max}$.*

(3) *Let $h\in\mathcal{H}(\kappa)$ be as in* (10.18), *and put $E_h := \int_{\Re(s)\gg 1} h(s)E_\Psi(s,g)dg$. Then, for each d,β, there is $N>0$ such that $S_{\infty,d,\beta}(E_h)\ll_\kappa \|h\|_0 \mathrm{Cond}_\infty(\chi)^N$, where the norm $\|h\|_0$ is defined in* (10.18).

(4) *We have $\mu_z(E_h)\ll_{K,\Psi,h}\min(z^K,z^{-K})$ for each*[17] $K\geq 1$. *Moreover, there is $N>0$ such that*

$$\left|\int_z \mu_z(E_h)g_+(z/T)d^\times z\right| \ll (\mathrm{N}(\mathfrak{f})T)^{-1/2}(T\mathrm{Cond}(\chi))^\epsilon \|h\|_N,$$

$$\left|\int_z \mu_z(E_h)g_-(z/T)d^\times z\right| \ll (\mathrm{N}(\mathfrak{f})T)^{1/2}(T\mathrm{Cond}(\chi))^\epsilon \mathrm{Cond}_\infty(\chi)^{1+\epsilon}\|h\|_N.$$

Proof. Lemmas 11.10 and 11.11. □

6.2. *Proof of Theorem* 6.1 *— cuspidal case.* Let χ be a character of $\mathbb{A}_F^\times/F^\times$, of varying conductor $\mathfrak{f}$. Put $q=\mathrm{N}(\mathfrak{f})$.

We need the following estimate, proved in Lemma 9.11. It amounts in essence to a statement about the equidistribution of p-adic horocycles (classically, these

[17]The implicit constants here are totally unimportant; this estimate will be used only to verify that certain integrals converge.

roughly correspond to a statement about the equidistribution of $\{\frac{x}{q}+iz\}_{0\le x\le q-1}$, if $z \asymp q^{-1}$).

For any function f that is invariant by $\prod_{\mathfrak{q}} K_0[\mathfrak{q}^{s_\mathfrak{q}}]$, we have (with $\mathfrak{m}=\prod_{\mathfrak{q}}\mathfrak{q}^{s_\mathfrak{q}}$)

(6.8)
$$\left|\nu_z(f)-\int_{\mathbf{X}} f\right| \ll_\epsilon q^{\alpha-1/2+\epsilon}\mathrm{N}(\mathfrak{m})^{3/2+\epsilon}\max\left(qz,\frac{1}{qz}\right)^{1/2} S_{2,d}(f),\ \ f\in C^\infty(\mathbf{X}).$$

Proof of Theorem 6.1 — *cuspidal case.* Choose $f\in\pi$ to be the "φ" of Proposition 6.1, so that $|\mu(f)| \gg_\epsilon \mathrm{N}(\mathfrak{f})^{-1/2-\epsilon}|L_{\mathrm{unr}}(1/2,\pi\times\chi)|$. For each ramified prime $\mathfrak{q}$ of π, let $\mathfrak{q}^{s_\mathfrak{q}}$ be the local conductor of the local representation $\pi_\mathfrak{q}$. Set $\mathfrak{m} := \prod_{\mathfrak{q}}\mathfrak{q}^{s_\mathfrak{q}}$, the finite conductor of π.

Let $K\ge 1$ be an integer satisfying $K\le \mathrm{N}(\mathfrak{f})$. Let $\mathcal{S}$ be the set of prime ideals of $\mathfrak{o}_F$, with norm lying in $[K,2K]$, and satisfying $(\mathfrak{n},\mathfrak{f})=1$ and $(\mathfrak{n},\mathrm{Supp}(f))=1$. Fix $\mathfrak{n}_0\in\mathcal{S}$. For each prime ideal $\mathfrak{n}\in\mathcal{S}$, let $\varpi_\mathfrak{n}\in F_\mathfrak{n}$ be a uniformizer. We define a measure σ on $\mathrm{GL}_2(\mathbb{A}_F)$ so that

$$\sigma = |\mathcal{S}|^{-1}\sum_{\mathfrak{n}\in\mathcal{S}}\chi([\varpi_\mathfrak{n}][\varpi_{\mathfrak{n}_0}]^{-1})\delta_{a([\varpi_\mathfrak{n}][\varpi_{\mathfrak{n}_0}]^{-1})}. \tag{6.9}$$

Clearly σ has total mass 1. Moreover, $\mu(f)=\mu(f\star\sigma)$ and $f\star\sigma$ is invariant by $K_0[\mathfrak{q}^{s_\mathfrak{q}}]$ for each $\mathfrak{q}|\mathfrak{m}$.

Choose κ "slightly smaller than 1", to be specified later. Our aim is now to cut the z integration in $\mu=\int_z \mu_z d^\times z$ into three ranges, the crucial range of which will be $q^{-2+\kappa}\ll z\ll q^{-\kappa}$; this avoids the pain of dealing with the infinite mass measure μ. Let g_+, g_- be as in Proposition 6.1. Define $h(t)$ by the rule $g_-(\frac{t}{q^{-2+\kappa}})+h(t)+g_+(\frac{t}{q^{-\kappa}})=1$. Then

(6.10)
$$|\mu(f)|^2 = |\mu(f\star\sigma)|^2 = \left|\int_z \mu_z(f\star\sigma)d^\times z\right|^2 \ll \left|\int g_-(\frac{z}{q^{-2+\kappa}})\mu_z(f\star\sigma)\right|^2 + \left|\int h(z)\mu_z(f\star\sigma)d^\times z\right|^2 + \left|\int g_+(\frac{z}{q^{-\kappa}})\mu_z(f\star\sigma)d^\times z\right|^2.$$

By Proposition 6.1, the first and last term (e.g., $\left|\int g_+(\frac{z}{q^{-\kappa}})\mu_z(f\star\sigma)d^\times z\right|$, without the square) are $\ll \mathrm{Cond}(\pi)^{1+\epsilon}\mathrm{Cond}_\infty(\chi)^{1+\epsilon}q^{\frac{\kappa-1}{2}+\epsilon}$. More explicitly, we note that

$$\mu_z(f\star\sigma) = |S|^{-1}\sum_{\mathfrak{n}\in\mathcal{S}}\mu_{\mathrm{N}(\mathfrak{n})^{-1}\mathrm{N}(\mathfrak{n}_0)z}(f). \tag{6.11}$$

Now $1/2\le \mathrm{N}(\mathfrak{n})\mathrm{N}(\mathfrak{n}_0)^{-1}\le 2$ for all $\mathfrak{n}$. This was the purpose of the factors involving $\mathfrak{n}_0$ in (6.9), and so one easily deduces a bound on $\int g_-(\frac{z}{q^{-2+\kappa}})\mu_z(f\star\sigma)d^\times z$ from the final assertion of Proposition 6.1. Similarly for the term involving g_+.

That the first and last terms of (6.10) should be less significant may be seen in the classical setting from the Fourier expansion; it should be regarded as a geometric version of the approximate functional equation.

As for the intermediate term, we note

$$\int_z h(z)\mu_z(f) = \int_{y\in\mathbb{A}_F^\times/F^\times} h(|y|)\chi(y) f(a(y)n([\mathfrak{f}]))d^\times y.$$

Applying Cauchy-Schwarz, and the fact that $\int_{\mathbb{A}_F^\times/F^\times} h(|y|)d^\times y \ll \log(q)$, we get

$$(6.12)\quad \left|\int_z h(z)\mu_z(f\star\sigma)\right|^2 \ll_\epsilon q^\epsilon \int_z h(z)\nu_z(|f\star\sigma|^2)d^\times z$$
$$\ll_\epsilon q^\epsilon \int_{\mathbf{X}} |f\star\sigma|^2 d\mu_{\mathbf{X}} + q^{\alpha-\kappa/2+\epsilon}\mathrm{N}(\mathfrak{m})^{3/2+\epsilon} S_{2,d}(|f\star\sigma|^2),$$

where we have applied (6.8). By Lemmas 8.1 and 8.2,

$$(6.13)\quad S_{2,d,\beta}(|f\star\sigma|^2) \ll S_{4,d,\beta}(f\star\sigma)^2$$
$$\ll (\sup_{g\in\mathrm{supp}(\sigma)}\|g\|)^{2\beta} S_{4,d,\beta}(f)^2 \ll K^{4\beta} S_{2,d',\beta+3/2}(f)^2,$$

where the last line holds for $d' \gg d$, and we have used Lemma 9.4 (which bounds the L^∞-norm of a cusp form in terms of L^2-norms), together with the easily verified fact that $\sup_{g\in\mathrm{supp}(\sigma)}\|g\| \ll K^2$.

By bounds towards Ramanujan (see §9.1),

$$(6.14)\qquad \|f\star\sigma\|_{L^2}^2 = \int_{g,g'} \langle g^{-1}g'f, f\rangle d\overline{\sigma}(g)d\sigma(g') \ll K^{2\alpha-1}\|f\|_{L^2}^2.$$

Thus

$$(6.15)\quad |\mu(f)| \ll_\epsilon \mathrm{Cond}(\pi)\mathrm{Cond}_\infty(\chi)(\mathrm{Cond}(\chi)\mathrm{Cond}(\pi))^\epsilon q^{\frac{\kappa-1}{2}}$$
$$+ \left(K^{\alpha-1/2}q^\epsilon + \mathrm{N}(\mathfrak{m})^{3/4+\epsilon}q^{-\kappa/4+\alpha/2+\epsilon}K\right) S_{2,d',2}(f)$$
$$\ll q^\epsilon(q^{(\kappa-1)/2} + K^{\alpha-1/2} + q^{\alpha/2-\kappa/4}K)\mathrm{Cond}_\infty(\chi)^N\mathrm{Cond}(\pi)^N,$$

for appropriate $N > 0$. We have used Proposition 6.1, (3) at the last step.

Proposition 6.1 guarantees that $|L_{\mathrm{unr}}(1/2,\pi\times\chi)| \ll_\epsilon q^{1/2+\epsilon}|\mu(f)|$. From this, optimizing κ, K, and applying trivial bounds at ramified places, we obtain the conclusion, taking $\alpha = 3/26$. □

6.3. *Proof of Theorem* 6.1 — *noncuspidal case.* We turn to the proof of (6.2). This is very similar, but we implement a mild regularization procedure to deal with the Eisenstein series, just as in the case of Rankin-Selberg L-functions.

Proof of (6.2). We may assume that χ ramifies at least at one finite place. Let Ψ be a Schwartz function on $\mathbb{A}_F^2$, $E(g,s) := E_\Psi(g,s)$ the corresponding Eisenstein series, chosen as in Proposition 6.2.

Let $\kappa' > 0$, let h be holomorphic in an open neighborhood of the vertical strip $-\kappa' \le \Re(s) \le 1+\kappa'$ and put $E_h(s) = \int_{\Re(s)=1+\kappa'} h(s)E(g,s)ds$. Then if $h(0) = h(\frac{1}{2}) = h(1) = 0$, it follows from the third assertion of Proposition 6.2 that

$$S_{\infty,d,\beta}(E_h) \ll_{\epsilon,d} \mathrm{Cond}_\infty(\chi)^N \|h\|_0$$

for appropriate $N = N(d,\beta) > 0$. Here, as in (10.18) with κ replaced by κ', the norm $\|h\|_N$ is defined to be $\int_{-\infty}^{\infty} (|h(1+\kappa'+it)| + |h(-\kappa'+it)|)(1+|t|)^N dt$.

Put, in the notation of Proposition 6.2, $I(s) = \Phi(s,0)L(\chi,s)L(\chi,1-s)$. Then

$$\int_z \mu_z(E_h)d^\times z = \mathrm{N}(\mathfrak{f})^{-1/2}\int_{\Re(s)=1/2} h(s)I(s)ds. \tag{6.16}$$

This is established in (11.31); for now, we remark that this is "almost" obvious from Proposition 6.2, the only additional point being that one can replace E by $\overline{E}$, and this is exactly where the fact that χ is ramified at a finite place comes in — to kill the constant term of the Eisenstein series.[18]

Take $h = (s-1/2)^2 s(1-s)\overline{I(1-\overline{s})}$. The "good" analytic properties of h, e.g. rapid decay along vertical lines, follow[19] from (6.7). In particular, h belongs to the function spaces $\mathcal{H}(\kappa)$ defined in (10.18) for any $\kappa > 0$, and the norms $\|h\|_N$ are all bounded by suitable powers of $\mathrm{Cond}_\infty(\chi).q$.

Then (6.16) becomes

$$\int_{t=-\infty}^{\infty} t^2(1/4+t^2)|I(\frac{1}{2}+it)|^2 = q^{1/2}\int_z \mu_z(E_h)d^\times z. \tag{6.17}$$

To bound the right-hand side, we proceed as in Section 6.2, but with f replaced by E_h. We use notation as in that section, except replacing the "h" defined before (6.10) by $1-g_- - g_+$ to avoid clashing with its alternate usage here.

One proves as in that section, that for $d \gg 1$,

$$\begin{aligned}\left|\int_z (1-g_- - g_+)\mu_z(E_h \star \sigma)\right| &\\ \ll_\epsilon q^\epsilon \left(K^{\alpha-1/2} + q^{\alpha/2-\kappa/4}(\sup_{g\in \mathrm{supp}(\sigma)}\|g\|)^{1/2}\right) S_{4,d,1/2}(E_h)&\\ \ll q^\epsilon \left(K^{\alpha-1/2} + q^{\alpha/2-\kappa/4}K\right)\|h\|_0 \mathrm{Cond}_\infty(\chi)^N&\end{aligned} \tag{6.18}$$

for some appropriate $N > 0$. At the last stage we have applied Proposition 6.2 to control the Sobolev norm.

Proposition 6.2 also guarantees that, for appropriate $N > 0$, we have

$$\begin{aligned}\left|\int g_-(\frac{z}{q^{-2+\kappa}})\mu_z(E_h\star\sigma)\right| + \left|\int g_+(\frac{z}{q^{-\kappa}})\mu_z(E_h\star\sigma)d^\times z\right| &\\ \ll \mathrm{Cond}_\infty(\chi)^{1+\epsilon} q^{\frac{\kappa-1}{2}+\epsilon}\|h\|_N.&\end{aligned} \tag{6.19}$$

[18] The classical version of this fact — see (6.3) — it is clear that the χ-sum will kill any constant term of f, as long as χ is not trivial.

[19] This point was not clear in a previous version; thanks to N. Bergeron for pointing this out.

Combining (6.17), (6.18), and (6.19), we obtain as in the previous section the bound, for sufficiently large N,

$$\int_{t=-\infty}^{\infty} t^2(\frac{1}{4}+t^2)|L(\frac{1}{2}+it,\chi)L(\frac{1}{2}-it,\chi)|^2|\Phi(1/2+it,0)|^2 \\ \ll_\epsilon \left(q^{\frac{\kappa-1}{2}}+K^{\alpha-1/2}+q^{\alpha/2-\kappa/4}K\right)\|h\|_N q^{1/2+\epsilon}\mathrm{Cond}_\infty(\chi)^N. \tag{6.20}$$

One applies the convexity bound to bound $\|h\|_N$, obtaining

$$\int_{-\infty}^{\infty} t^2|L(\frac{1}{2}+it,\chi)|L(\frac{1}{2}-it,\chi)|^2|\Phi(1/2+it,0)|^2 \ll \mathrm{Cond}_\infty(\chi)^N q^{24/25},$$

where we have increased N as necessary. From this we get

$$L\left(\frac{1}{2},\chi\right) \ll \mathrm{Cond}_\infty(\chi)^N q^{1/4-1/200}. \qquad \square$$

7. Torus periods (II): equidistribution of compact torus orbits

It has been independently shown by Zhang [47], Clozel-Ullmo [8] and P. Cohen [9] that the subconvexity result Theorem 6.1 implies the equidistribution of Heegner points over totally real fields; in particular, they pointed out that GRH implies this equidistribution. Theorem 6.1 makes this result unconditional.

The main aim of this section is to explain how one can obtain certain conditional results about equidistribution of subsets of Heegner points, and how this fits into the general framework of "sparse equidistribution questions." In particular, this approach does not rely on reducing questions about subsets of Heegner points to subconvexity, but rather approaches the equidistribution question directly.

The proofs of the results (and various supporting lemmas) will only be sketched, and we will confine ourselves for simplicity to the case of narrow class number 1; we will in any case present an unconditional approach, based on combining the ideas of this paper with the ideas of Michel, in the paper [28] (joint with P. Michel). We nevertheless feel that the ideas presented here may be of use in other contexts. Indeed, this section is of a different flavor to the other sections; it uses "adelic analysis" more genuinely.

In fact, we shall need a mild refinement of the results of [8], which will allow better control of the dependence on the test vectors. We state this refinement without proof in Theorem 7.1; the proof is an exercise in explicating some of the proofs in [8].

7.1. *Equidistribution of Heegner points.* We recall the definition of Heegner points. Let F be a totally real number field of degree d over $\mathbb{Q}$. For simplicity we shall confine ourselves to the case where the ring of integers of F has narrow class number 1. This assumption does not change any of the technical details, which are in any case carried out adelically; it simply allows us to be a little more explicit about the torus orbits we consider. Let $E = F(\sqrt{-\mathbf{d}})$ be a totally imaginary quadratic extension of F, where $\mathbf{d} \in \mathfrak{o}_F$ is totally positive and squarefree. Here "squarefree" means that it is of valuation ≤ 1 at all finite places.

Let T_E be the torus $\mathrm{Res}_{E/F}(\mathbb{G}_m)/\mathbb{G}_m$; we embed T_E in PGL_2 via (in obvious notation):

$$\iota_E : x + y\sqrt{-\mathbf{d}} \mapsto \begin{pmatrix} x & y \\ -y\mathbf{d} & x \end{pmatrix}. \tag{7.1}$$

Regard $\mathbf{d}$ as an element of $F \otimes \mathbb{R}$ via the inclusion $F \hookrightarrow F \otimes \mathbb{R}$. Since it is totally positive, it possesses a unique totally positive square root, $\sqrt{\mathbf{d}} \in F \otimes \mathbb{R}$. Set $[\mathbf{d}]_\infty = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{\mathbf{d}} \end{smallmatrix}\right) \in \mathrm{PGL}_2(F \otimes \mathbb{R})$. We define a map $\mathcal{H} : T_E(\mathbb{A}_F)/T_E(F) \to \mathbf{X}$ via

$$\mathcal{H} : x \mapsto \iota_E(x)[\mathbf{d}]_\infty, \tag{7.2}$$

where we regard $[\mathbf{d}]_\infty \subset \mathrm{PGL}_2(F \otimes \mathbb{R}) \subset \mathrm{PGL}_2(\mathbb{A}_F)$ acting by right translation on $\mathbf{X}$. Denote by $\mathrm{N}(\mathbf{d})$ the absolute norm of $\mathbf{d}$, i.e., $\mathrm{N}(\mathbf{d}) = |\mathfrak{o}_F/\mathbf{d}\mathfrak{o}_F|$.

The F-torus T_E is anisotropic, and there is a unique $T_E(\mathbb{A}_F)$-invariant probability measure on $T_E(\mathbb{A}_F)/T_E(F)$. Let ν_E be its image by the map $\mathcal{H}$.

THEOREM 7.1. *Set $E = F(\sqrt{-\mathbf{d}})$, where $\mathbf{d} \in \mathfrak{o}_F$ is totally positive and squarefree. The measures ν_E become equidistributed as* $\mathrm{N}(\mathbf{d}) \to \infty$. *Indeed, there exist $\delta > 0, d, \beta$ such that for $f \in C^\infty(\mathbf{X})$ we have*

$$\left| \int f d\nu_E - \int_{\mathbf{X}} f(x)dx \right| \ll \mathrm{N}(\mathbf{d})^{-\delta} S^*_{\infty,d,\beta}(f).$$

Recall the definition of S^* from Section 2.10. We do not give the proof; as we have remarked it can be obtained by following the computations of [8] a little more explicitly.

One recovers from Theorem 7.1 the equidistribution of certain Heegner points associated to $E = F(\sqrt{-\mathbf{d}})$ as $\mathbf{d}$ varies. Theorem 7.1 also gives an effective rate of equidistribution for Heegner points with polynomial dependence on the level and the eigenvalue of a test function. This rather innocuous polynomial dependence (in the level aspect, at least) will in fact play a crucial role in our deduction of the equidistribution of sparse subsets in the following section.

7.2. *Equidistribution of subsets of Heegner points.* We turn to certain conditional results on equidistribution of sparse subsets. F being as in Section 7.1, let $E_i = F(\sqrt{\mathbf{d}_i})$ be a sequence of distinct quadratic, totally imaginary, extensions of F. For each E_i, let $S_i \subset T_{E_i}(\mathbb{A}_F)/T_{E_i}(F)$ be a subgroup of finite index m_i. Let $\mu_{E_i}^{S_i}$ be the image of the Haar probability measure on S_i by the map $\mathcal{H}$.

The import of the next theorem is that, *if E_i has enough small split primes*, one can obtain the equidistribution of the measures $\mu_{E_i}^{S_i}$ as $i \to \infty$. This result is quite similar to the results of Duke-Friedlander-Iwaniec [13] in the case $F = \mathbb{Q}$, although the method is at least superficially rather different. One can also contrast with the striking results of Michel [27] and Harcos-Michel [17], for $F = \mathbb{Q}$, that give comparable results but without the condition on enough small split primes.

Our method is different to these, where the results are deduced from subconvexity bounds for Rankin-Selberg L-functions.[20] In the present approach, we prove the equidistribution theorem directly. In a sequel to this paper, the author and P. Michel combine the methods here with some methods developed by Michel to make the results of this section unconditional.

To quantify the existence of enough small split primes, one might impose the condition (as does Linnik [25]) that the E_i vary through a sequence of quadratic extensions that split at a fixed prime of F. We will prefer to take a more quantitative approach, which will yield a stronger result at the price of a stronger assumption. In that regard we introduce the following notation: For $\delta > 0$, we put

$$\mathrm{wt}(E_i, \delta) = \#\{\mathfrak{q} \subset \mathfrak{o}_F \text{ prime and split in } E_i, \mathrm{N}(\mathbf{d}_i)^{\delta} \le \mathrm{N}(\mathfrak{q}) \le 2\mathrm{N}(\mathbf{d}_i)^{\delta}\}.$$

THEOREM 7.2. *There exists* $\delta_1 > 0$ *such that, if* $\frac{m_i}{\min(\mathrm{N}(\mathbf{d}_i)^{\delta_1(1/2-\alpha)}, \mathrm{wt}(E_i,\delta_1)^{1/2})} \to 0$, *the sequence* $\mu_{E_i}^{S_i}$ *converges, as* $i \to \infty$, *to the invariant measure on* $\mathbf{X}$.

Proof. This is deduced from Theorem 7.1 by using the mixing properties of the $T_E(\mathbb{A}_F)$-flow. Indeed, we fix an index i and a corresponding field E_i. Let $\delta_1 > 0$ be fixed. Let $\mathcal{S}$ be the set of prime ideals of F which split in E_i and with norm in $[\mathrm{N}(\mathbf{d}_i)^{\delta_1}, 2\mathrm{N}(\mathbf{d}_i)^{\delta_1}]$. For each $\mathfrak{q} \in \mathcal{S}$, the torus $T_{E_i}(F_{\mathfrak{q}})$ is isomorphic to $F_{\mathfrak{q}}^{\times}$. Fix an isomorphism $\Upsilon_{\mathfrak{q}} : T_{E_i}(F_{\mathfrak{q}}) \to F_{\mathfrak{q}}^{\times}$, and let $\varpi_{\mathfrak{q}}$ be an element in $T_{E_i}(F_{\mathfrak{q}})$ such that $\Upsilon_{\mathfrak{q}}(\varpi_{\mathfrak{q}})$ has valuation ± 1 in $F_{\mathfrak{q}}^{\times}$.

Let χ be a character of $T_{E_i}(\mathbb{A}_F)/T_{E_i}(F)$, trivial on S_i. Let ν_{E_i} be as defined prior to Theorem 7.1, and define

$$\mu_{E_i}(f) = \int_{t \in T_{E_i}(\mathbb{A}_F)/T_{E_i}(F)} f(\mathcal{H}(t))\chi(t)\,dt,$$

where dt is the Haar probability measure on $T_{E_i}(\mathbb{A}_F)/T_{E_i}(F)$. Let σ be the probability measure $\frac{1}{|\mathcal{S}|}\sum_{\mathfrak{q}\in\mathcal{S}} \chi(\varpi_{\mathfrak{q}})\delta_{\varpi_{\mathfrak{q}}}$ on $T_{E_i}(\mathbb{A}_F)$. Then

$$\mu_{E_i}(f) = \mu_{E_i}(f \star \mathcal{H}_*\sigma),$$

where $\mathcal{H}_*\sigma$ denotes the image of σ by the map $\mathcal{H}$.

By Cauchy-Schwarz, and Theorem 7.1,

$$\begin{aligned} (7.3)\quad |\mu_{E_i}(f \star \mathcal{H}_*\sigma)|^2 &\le \nu_{E_i}(|f \star \mathcal{H}_*\sigma|^2) \\ &\le \|f \star \mathcal{H}_*\sigma\|_{L^2}^2 + O\left(\mathrm{N}(\mathbf{d}_i)^{-\delta} S_{\infty,d,\beta}^*(|f \star \mathcal{H}_*\sigma|^2)\right), \end{aligned}$$

[20] We note that the required bounds on Rankin-Selberg L-functions are considerably deeper than (5.3), for they deal with varying central character. For some speculative discussion on the "reason" that Michel's method can avoid this condition, see the last paragraph of [29].

where δ, d, β are as in Theorem 7.1. Now, appropriate variants of Lemmas 8.1 and 8.2 (for S^* instead of S) show that

$$
\begin{aligned}
(7.4)\quad S^*_{\infty,d,\beta}(|f \star \mathcal{H}_*\sigma|^2) &\ll S^*_{\infty,d,\beta}(f \star \mathcal{H}_*\sigma)^2 \\
&\ll \sup_{g \in \mathrm{supp}\mathcal{H}_*\sigma} \|g\|^{6\beta} S^*_{\infty,d,\beta}(f)^2 \ll \mathrm{N}(\mathbf{d}_i)^{6\delta_1\beta} S^*_{\infty,d,\beta}(f)^2
\end{aligned}
$$

and bounds towards Ramanujan (see §9.1) show that

$$
(7.5)\qquad \|f \star \mathcal{H}_*\sigma\|^2_{L^2} \ll \left(\mathrm{N}(\mathbf{d}_i)^{\delta_1(2\alpha-1)} + |\mathcal{S}|^{-1}\right) \|f\|^2_{L^2}.
$$

We note that (7.4) and (7.5) are very closely analogous to (6.13) and (6.14), with K replaced by $\mathrm{N}(\mathbf{d}_i)^{\delta_1}$. In the context of (6.14), the set $\mathcal{S}$ has size $K^{1-\epsilon}$; thus the term $|\mathcal{S}|^{-1}$ that appears in (7.5) could be neglected.

Recalling the definition of μ_{E_i}, we conclude

$$
\begin{aligned}
(7.6)\quad &\left| \int_{t \in T_{E_i}(\mathbb{A}_F)/T_{E_i}(F)} f(\mathcal{H}(t))\chi(t)dt \right| \\
&\qquad \ll \left(\mathrm{N}(\mathbf{d}_i)^{3\delta_1\beta-\delta/2} + \mathrm{N}(\mathbf{d}_i)^{\delta_1(\alpha-1/2)} + |\mathcal{S}|^{-1/2}\right) S^*_{\infty,d,\beta}(f).
\end{aligned}
$$

Summing the left-hand side of (7.6) over all m_i characters χ of $T_{E_i}(\mathbb{A}_F)/T_{E_i}(F)$ that are trivial on S_i, and substituting $|\mathcal{S}| = \mathrm{wt}(E_i, \delta_1)$, we obtain

$$
\left|\mu^{S_i}_{E_i}(f)\right| \ll m_i \left(\mathrm{N}(\mathbf{d}_i)^{3\delta_1\beta-\delta/2} + \mathrm{N}(\mathbf{d}_i)^{\delta_1(\alpha-1/2)} + \mathrm{wt}(E_i, \delta_1)^{-1/2}\right) S^*_{\infty,d,\beta}(f).
$$

Choosing δ_1 sufficiently small (the exact value will depend on the value of β, δ from Theorem 7.1) we obtain the claimed conclusion. □

8. Background on Sobolev norms and reduction theory

The rest of the paper consists of technical lemmas. The sections that follow are arranged to be used as a reference, rather than to be read through.

8.1. *Formal properties of the Sobolev norms.* We begin by explicating certain formal properties of the Sobolev norms defined in Section 2.9.3.

Remark 8.1. The following properties of this definition are formal and will be repeatedly used:

(1) Translations by $K_{\max,\mathbf{G}}$ preserve $S_{p,d,\beta}$, i.e., $S_{p,d,\beta}(k \cdot f) = S_{p,d,\beta}(f)$ for $k \in K_{\max,\mathbf{G}}$.

(2) If $L : C^\infty_\omega(\mathbf{X}_\mathbf{G}) \to \mathbb{C}$ is a linear functional and $|L(\psi)| \le PS_{p,d,\beta}(\psi)$, then also $|L(\psi)| \le S_{p,d,\beta}(\psi)$. Indeed $\psi \mapsto |L(\psi)|$ is itself a seminorm on $C^\infty_\omega(\mathbf{X}_\mathbf{G})$.

(3) Suppose that $E : C^\infty_\omega(\mathbf{X}_\mathbf{G}) \to C^\infty_\omega(\mathbf{X}_\mathbf{G})$ is a linear endomorphism satisfying $PS_{p,d,\beta}(Ef) \le A \cdot PS_{p,d,\beta}(f)$, for some $A \in \mathbb{R}$. Then also $S_{p,d,\beta}(Ef) \le AS_{p,d,\beta}(f)$. Indeed, $f \mapsto A^{-1}S_{p,d,\beta}(Ef)$ is a seminorm dominated by $PS_{p,d,\beta}$.

(4) We shall need a slight variant of (3) in the case where we are studying only the space of f with some invariance property.

Suppose that $E : C_\omega^\infty(\mathbf{X_G}) \to C_\omega^\infty(\mathbf{X_G})$ is a linear endomorphism, M is a finite set of finite places, and for each $v \in M$ we are given an open compact $K_{1,v} \subset K_v$. Suppose moreover that $PS_{p,d,\beta}(Ef) \leq A \cdot PS_{p,d,\beta}(f)$ for some $A \in \mathbb{R}$ and for all f which are $\prod_{v\in M} K_{1,v}$-fixed. Then, for all f which are $\prod_{v\in M} K_{1,v}$-fixed, we have in fact

$$S_{p,d,\beta}(Ef) \leq A \prod_{v\in M} [K_v : K_{1,v}]^\beta S_{p,d,\beta}(f).$$

Indeed, put $K_{1,M} = \prod_{v\in M} K_{1,v}$ and let Π be the averaging operator

$$\int_{k\in K_{1,M}} \pi(k)dk,$$

where $K_{1,M}$ is endowed with the Haar probability measure. Then apply (3) above to the operator $f \mapsto E(\Pi f)$.

LEMMA 8.1. *Let* $F_1 \in C_{\omega_1}^\infty(\mathbf{X_G})$, $F_2 \in C_{\omega_2}^\infty(\mathbf{X_G})$. *Then*

$$S_{p,d,\beta}(F_1F_2) \ll_d S_{2p,d,\beta}(F_1)S_{2p,d,\beta}(F_2).$$

Note that $F_1F_2 \in C_{\omega_1\omega_2}^\infty(\mathbf{X_G})$.

Proof. Put $F = F_1F_2$. For any monomial $\mathscr{D}$ of degree d in $\mathscr{B}$, we can write $\mathscr{D}(F_1F_2) = \sum_{\alpha\in\mathscr{I}}(\mathscr{D}_{\alpha,1}F_1)(\mathscr{D}_{\alpha,2}F_2)$, where α ranges over an index set $\mathscr{I}$ whose size is bounded by a constant depending only on d, and the $\mathscr{D}_{\alpha,\star}$ are certain monomials in $\mathscr{B}$ satisfying $\mathrm{ord}(\mathscr{D}_{\alpha,1}) + \mathrm{ord}(\mathscr{D}_{\alpha,2}) = d$. It follows that

$$\|\mathscr{D}F\|_{L^p(\mathbf{X}_{\mathbf{G},\mathrm{ad}})} \leq \sum_{\alpha\in\mathscr{I}} \left(\int_{\mathbf{X}_{\mathbf{G},\mathrm{ad}}} |\mathscr{D}_{\alpha,1}F_1|^p |\mathscr{D}_{\alpha,2}F_2|^p\right)^{1/p}.$$

Applying Cauchy-Schwarz, we conclude

$$\|\mathscr{D}F\|_{L^p(\mathbf{X}_{\mathbf{G},\mathrm{ad}})} \leq \sum_{\alpha\in\mathscr{I}} \|\mathscr{D}_{\alpha,1}F_1\|_{L^{2p}(\mathbf{X}_{\mathbf{G},\mathrm{ad}})} \|\mathscr{D}_{\alpha,2}F_2\|_{L^{2p}(\mathbf{X}_{\mathbf{G},\mathrm{ad}})}. \tag{8.1}$$

Clearly, for each finite place v, we have $K_{v,F} \supset K_{v,F_1} \cap K_{v,F_2}$; in particular $[K_{\max,\mathbf{G}} : K_F] \leq [K_{\max,\mathbf{G}} : K_{F_1}][K_{\max,\mathbf{G}} : K_{F_2}]$. It follows that

$$\begin{aligned}
&[K_{\max,\mathbf{G}} : K_F]^\beta \sum_{\mathscr{D}} \|\mathscr{D}F\|_{L^p(\mathbf{X}_{\mathbf{G},\mathrm{ad}})} \\
&\quad \ll \left([K_{\max,\mathbf{G}} : K_{F_1}]^\beta \sum_{\mathscr{D}} \|\mathscr{D}F_1\|_{L^{2p}}\right)\left([K_{\max,\mathbf{G}} : K_{F_2}]^\beta \sum_{\mathscr{D}} \|\mathscr{D}F_2\|_{L^{2p}}\right),
\end{aligned} \tag{8.2}$$

where the implicit constant depends only on d, and in all three instances $\mathscr{D}$ varies over the set of monomials in $\mathscr{B}$ of degree $\leq d$.

That is to say, there is a constant $C = C(d)$ such that

$$PS_{p,d,\beta}(F_1F_2) \leq C \cdot PS_{2p,d,\beta}(F_1)PS_{2p,d,\beta}(F_2).$$

From (2.11) we deduce

$$S_{p,d,\beta}(F_1F_2) \leq C \cdot S_{2p,d,\beta}(F_1)S_{2p,d,\beta}(F_2),$$

as required. □

We recall the definition of $\|g\|$ for $g \in \mathbf{G}(F_\infty), \mathbf{G}(\mathbb{A}_F)$ etc. from Section 2.4.

LEMMA 8.2. *Let $F \in C^\infty(\mathbf{X_G})$ and $g = (g_\infty, g_f) \in \mathbf{G}(\mathbb{A}_F)$.*

$$S_{p,d,\beta}(g \cdot F) \ll \|g_\infty\|^d \|g_f\|^\beta S_{p,d,\beta}(F).$$

Proof. Put $F' = (g_\infty, g_f) \cdot F$, where $g_f = (g_v)_{v \text{ finite}}$. For each finite place v, we note that $K_{v,F'} \supseteq g_v K_{v,F} g_v^{-1} \cap K_{v,\mathbf{G}}$. The index $[K_{v,\mathbf{G}} : K_{v,F'}]$ is therefore bounded above by the number of cosets $xg_vK_{v,F}$ in $K_{v,\mathbf{G}}g_vK_{v,F}$. Clearly this is bounded above by the number of left $K_{v,F}$ cosets in $K_{v,\mathbf{G}}g_vK_{v,\mathbf{G}}$; but the number of such cosets is precisely $\|g_v\| \cdot [K_{v,\mathbf{G}} : K_{v,F}]$. It now follows easily from the definitions that $PS_{p,d,\beta}(F') \ll \|g_\infty\|^d \|g_f\|^\beta PS_{p,d,\beta}(F)$. Applying Remark 8.1 to the endomorphism $F \mapsto (g_\infty, g_f) \cdot F$, we obtain the claim. □

The following crude lemma is as much of interpolation as we need. It will be applied, in practice, where E is a composite of a Hecke operator and a certain L^2-projection.

LEMMA 8.3. *Let E be a linear endomorphism of $C_\omega^\infty(\mathbf{X_G})$ which commutes with $\mathbf{G}(F_\infty) \times K_{\max,\mathbf{G}}$. Suppose there are real numbers $A, B > 0$ such that for any $f \in C_\omega^\infty(\mathbf{X_G})$, we have $\|Ef\|_{L^2} \leq A\|f\|_{L^2}$, $\|Ef\|_{L^\infty} \leq B\|f\|_{L^\infty}$. Then for $2 \leq p \leq \infty$,*

$$S_{p,d,\beta}(Ev) \leq A^{2/p}B^{1-\frac{2}{p}}S_{p,d,\beta}(v).$$

(*We admit also $B = \infty$, in which case the L^∞ hypothesis should be seen as void*, *and the result becomes $S_{2,d,\beta}(Ev) \leq AS_{2,d,\beta}(v)$*.)

Proof. By interpolation, the operator norm of E with respect to the L^p norm on $C_\omega^\infty(\mathbf{X_G})$ is $\leq A^{2/p}B^{1-2/p}$. Moreover, the assumption on E shows that $K_{Ef} \supset K_f$.

It follows that for $f \in C_\omega^\infty(\mathbf{X_G})$ we have the inequality

$$PS_{p,d,\beta}(Ef) \leq A^{2/p}B^{1-2/p}PS_{p,d,\beta}(f).$$

Remark 8.1 implies the conclusion. □

8.1.1. *Computing Sobolev norms in the Kirillov model.* In the present section, let v be an archimedean place of F.

Let π_v be a generic unitary irreducible representation of $\mathrm{GL}_2(F_v)$. Recall that this means that π_v is realized in a space of functions $\mathcal{K}$ (the Kirillov model, consisting of restrictions of functions in the Whittaker model to the diagonal torus) on $F_v^\times$. Recall also the definition of the local conductor $\mathrm{Cond}_v(\pi_v)$ from Section 2.12.2.

In this model, the diagonal torus acts by translation and upper triangular matrices act through multiplication by characters: that is to say, for $f \in \mathcal{K}$, $y_1, y_2 \in F_v^\times$, $z \in F_v$ we have the rules

$$\pi(a(y_1))f : y_2 \mapsto f(y_1 y_2), \pi(n(z))f : y_2 \mapsto f(y_2)e_{F_v}(zy_2). \tag{8.3}$$

From these facts it is easy to verify that the space of smooth vectors in π_v contains all compactly supported smooth functions on $F_v^\times$. Moreover,

$$\|f\|_2^2 = \int_{F_v^\times} |f(y)|^2 d^\times y \tag{8.4}$$

defines a $\mathrm{GL}_2(F_v)$-invariant inner product on $\mathcal{K}$.

We will eventually have occasion to choose test vectors in π_v in this model, and wish to evaluate the "Sobolev norms" of the resulting vectors.

LEMMA 8.4. *Suppose $F_v \cong \mathbb{R}$. Let $f \in \mathcal{K}$ be C^∞ and compactly supported. Then*

$$\sum_{\mathrm{ord}(\mathcal{D}) \le k} \|\mathcal{D}f\|_2 \ll \mathrm{Cond}_v(\pi_v)^{2k} \left(\sum_{j=0}^{2k} \int_{\mathbb{R}^\times} (|y| + |y|^{-1})^{2k} \left| \frac{d^j f}{d^j y} \right|^2 d^\times y \right)^{1/2},$$

where the $\mathcal{D}$ sum ranges over all monomials in a fixed basis for $\mathrm{Lie}(\mathrm{GL}_2(F_v))$ *of degree $\le k$.*

Suppose $F_v \cong \mathbb{C}$, and suppose $f \in \mathcal{K}$ is C^∞ and compactly supported. Then

$$\sum_{\mathrm{ord}(\mathcal{D}) \le k} \|\mathcal{D}f\|_2 \ll \mathrm{Cond}_v(\pi_v)^{k} \left(\sum_{0 \le i+j \le 2k} \int_{\mathbb{C}^\times} (|z| + |z|^{-1})^{2k} \left| \frac{\partial^{i+j} f}{\partial^i z \partial^j \bar{z}} \right|^2 d^\times z \right)^{1/2}.$$

Proof. We prove only the case with $F_v \cong \mathbb{R}$, the complex case being similar. Let h, e, f, z be nonzero elements of the (real) Lie algebra of $\mathrm{GL}_2(\mathbb{R})$, defined via

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

These satisfy the usual commutation relations $[h, e] = 2e, [h, f] = 2f, [e, f] = h$. Let λ be the scalar by which the Casimir operator $\frac{1}{2}h^2 + ef + fe$ acts, and ν the scalar by which z acts; then $1 + |\lambda| + |\nu|^2 \ll \mathrm{Cond}_v(\pi_v)^2$.

It is easy to see how h, e act on $\mathcal{K}$: h acts by a multiple of the differential operator $c_1\nu + y\frac{d}{dy}$ and e acts by multiplication by $c_3 y$, for some constants c_1, c_2, c_3. The Casimir operator $\frac{1}{2}h^2 + ef + fe = \frac{1}{2}h^2 + 2ef - h$ acts by the scalar λ; so it follows that for $v \in \mathcal{K}$ we have $efv = \frac{1}{2}(\lambda + h - h^2)v$. In particular, f acts on any compactly supported function via the differential operator $c_1' y^{-1} + c_2' \frac{d}{dy} + c_3' y \frac{d^2}{dy^2}$, for certain constants c_1', c_2', c_3', satisfying $|c_1'|, |c_2'|, |c_3'| \ll \mathrm{Cond}_v(\pi_v)^2$. (In fact, $|c_i'| \ll \mathrm{Cond}_v(\pi_v)^{3-i}$.)

Any monomial of degree k in h, e, f, z is therefore a sum of terms $c_{\gamma\delta} y^{\gamma} \partial_y^{\delta}$, where $|c_{\gamma\delta}| \ll \mathrm{Cond}_v(\pi_v)^{2k}$, $|\gamma| \le k, \delta \le 2k$. The claimed result follows in the case $F_v \cong \mathbb{R}$.

A similar proof holds for $F_v \cong \mathbb{C}$. $\square$

8.2. *Reduction theory.* Recall that $F_\infty := F \otimes_{\mathbb{Q}} \mathbb{R}$. Let $K_\infty, K_v, K_{\max}$ be as in Section 2.5. Then $K_\infty \times K_{\max}$ is a maximal compact subgroup of $\mathrm{GL}_2(\mathbb{A}_F)$. Given $g \in \mathrm{GL}_2(\mathbb{A}_F)$ we may always write $g = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} x & 0 \\ 0 & y \end{smallmatrix}\right)k$, with $t \in \mathbb{A}_F, x, y \in \mathbb{A}_F^\times, k \in K_\infty \times K_{\max}$. We set $\mathrm{ht}(g) = |xy^{-1}|_{\mathbb{A}}$; this is well-defined, although x, y are not unique.

Then ht descends to a function $B(F)\backslash \mathrm{GL}_2(\mathbb{A}_F) \to \mathbb{R}_{>0}$. Explicitly,

$$\mathrm{ht}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{|ad-bc|_{\mathbb{A}}}{\prod_v \|(c_v, d_v)\|_v^2}, \tag{8.5}$$

where one defines $\|(c_v, d_v)\|_v = \max(|c_v|_v, |d_v|_v)$ for v finite, and

$$\|(c_v, d_v)\|_v = \left(|c_v|_v^{2/\deg(v)} + |d_v|_v^{2/\deg(v)}\right)^{\deg(v)/2} \tag{8.6}$$

for v infinite, where $\deg(v) = [F_v : \mathbb{R}]$.

Define $\mathfrak{S}(T) \subset B(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)$ to be $\mathfrak{S}(T) := \{g : \mathrm{ht}(g) \ge T\}$. Then, for all $T > 0$ the natural projection $\Pi : \mathfrak{S}(T) \to \mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)$ has finite fibers; for sufficiently large T, it is injective, and for sufficiently small T it is surjective. This is the content of reduction theory for GL_2. As a consequence, the complement of $\Pi(\mathfrak{S}(T))$ has compact closure, modulo the center, for each T.

Fix T_0 such that $\Pi : \mathfrak{S}(T_0) \to \mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_F)$ is injective. Then we define a function $\mathrm{ht} : \mathrm{GL}_2(F)\backslash \mathrm{GL}_2(\mathbb{A}_F) \to \mathbb{R}$ via the rule

$$\mathrm{ht}(g) = \begin{cases} \mathrm{ht}(g'), & \text{if } g = \Pi(g') \text{ for some } g' \in \mathfrak{S}(T_0), \\ T_0, & \text{else.} \end{cases}$$

In fact, it is clear that ht descends to a function $\mathbf{X} = \mathrm{PGL}_2(F)\backslash \mathrm{PGL}_2(\mathbb{A}_F) \to \mathbb{R}_{\ge T_0}$.

LEMMA 8.5. *Let $U \subset \mathrm{GL}_2(F_\infty)$ be compact and $x \in \mathbf{X}_{\mathrm{GL}(2)}$. The fibers of the map $U \times K_{\max} \to \mathbf{X}_{\mathrm{GL}(2)}$ defined by $(u, k) \mapsto xuk$ have size bounded by $O(\mathrm{ht}(x))$, where the implicit constant depends on U.*

Proof. Suppose $g \in \mathrm{GL}_2(\mathbb{A}_F)$ is a lift of $x \in \mathbf{X}_{\mathrm{GL}(2)}$. Consider the map $U \times K_{\max} \to \mathbf{X}_{\mathrm{GL}(2)}$ given by $(u, k) \mapsto guk$, as above. Let (u, k) belong to a fiber of maximal size. Call this size M. Then

$$\begin{aligned} M &= \#\{\gamma \in \mathrm{GL}_2(F) : guk = \gamma g u' k', \exists\ u' \in U, k' \in K_{\max}\} \\ &\le \#\{\gamma : g u'' k'' = \gamma g, \exists\ u'' \in U \cdot U^{-1}, k'' \in K_{\max}\}. \end{aligned} \tag{8.7}$$

Set $V = U \cdot U^{-1}$, a compact subset of $\mathrm{GL}_2(F_\infty)$. The definition of $\mathfrak{S}(T)$ shows that there exists a constant $c < 1$, depending on V, such that $\mathfrak{S}(T) \cdot V \cdot K_{\max} \subset$

$\mathfrak{S}(cT)$. Choose T so large that the projection $\mathfrak{S}(cT) \to \mathbf{X}_{\mathrm{GL}(2)}$ is injective. It will suffice to show, whenever $B(F)g \in \mathfrak{S}(T)$, that

$$\#\{\gamma \in \mathrm{GL}_2(F) : \gamma g \in gVK_{\max}\} \ll \mathrm{ht}(g). \tag{8.8}$$

(The $B(F)$-coset of) both g and gVK belong entirely to $\mathfrak{S}(cT)$. By the choice of T, $\gamma g \in gVK_{\max}$ implies γ in $B(F)$. Write $\gamma = a_\gamma n_\gamma$, with $a_\gamma \in A(F)$ and $n_\gamma \in N(F)$; also, write $g = n_g a_g k_g$ with $n_g \in N(\mathbb{A}_F), a_g \in A(\mathbb{A}_F), k_g \in K_\infty \times K_{\max}$. We are free to adjust g on the left by an element of $N(F)$, since doing so will not affect the cardinality of the set $\{\gamma \in \mathrm{GL}_2(F) : \gamma g \in gVK_{\max}\}$. We may thereby assume that n_g lies in a fixed compact subset of $N(\mathbb{A}_F)$. Thus we can write $g = a_g k_g'$, where $k_g' := a_g^{-1} n_g a_g k_g$ lies in a certain fixed compact subset Ω of $\mathrm{GL}_2(\mathbb{A}_F)$.

Now, $\gamma g \in gVK_{\max}$ implies that $a_g^{-1} a_\gamma n_\gamma a_g \in \Omega V K_{\max} \Omega^{-1}$. Noting that $a_g^{-1} a_\gamma n_\gamma a_g = a_\gamma a_g^{-1} n_\gamma a_g$, we deduce that a_γ lies in a fixed compact subset of $\mathrm{GL}_2(\mathbb{A}_F)$, depending only on U; thus the number of possibilities for a_γ are $\ll_U 1$. Moreover, it now follows that $a_g^{-1} n_\gamma a_g$ lies in a compact subset of $\mathbb{A}_F$ depending only on U.

Thus, if we write $a_g = \left(\begin{smallmatrix} x & 0 \\ 0 & y \end{smallmatrix}\right), n_\gamma = \left(\begin{smallmatrix} 1 & \beta \\ 0 & 1 \end{smallmatrix}\right)$, then $\beta \in xy^{-1}\Omega'$, where $\Omega' \subset \mathbb{A}_F$ is a compact subset that depends only on U. It is easy to see that the number of possibilities for β is $\ll_U 1 + |xy^{-1}|_{\mathbb{A}_F}$. But $|xy^{-1}|_{\mathbb{A}_F} = \mathrm{ht}(g)$, which is a function that is bounded away from zero, and we are done. $\square$

LEMMA 8.6. *Let notation be as in the previous Lemma* 8.5. *Consider the composite map* $U \cdot K_{\max} \xrightarrow{\Pi} \mathbf{X}_{\mathrm{GL}(2)} \to \mathbf{X}$. *Each fiber of this map may be written as the union of at most* $O(\mathrm{ht}(x))$ *sets each of the form* $yZ(\mathbb{A}_F) \cap UK_{\max}$, *where* Z *is the center of* GL_2 *and* $y \in \mathrm{GL}_2(\mathbb{A}_F)$.

Proof. Let $\bar{x}$ be the image of x in $\mathbf{X}$. Let $u, u' \in U, k, k' \in K_{\max}$. Suppose that $\bar{x}uk = \bar{x}u'k'$ in $\mathbf{X}$. Then there is $z \in \mathbb{A}_F^\times$ and $\gamma \in \mathrm{GL}_2(F)$ such that

$$xuk = \gamma x u' k' a(z,z), \text{ equality in } \mathrm{GL}_2(\mathbb{A}_F). \tag{8.9}$$

For fixed u, k and γ, the set of $u'k'$ satisfying (8.9) is visibly the intersection of $UK_{\max}$ with a fixed $Z(\mathbb{A}_F)$-coset. This coset depends only on the class of γ in $\mathrm{PGL}_2(F)$, so it suffices to show that those $\gamma \in \mathrm{GL}_2(F)$ that occur in equalities such as (8.9) for varying u, k, u', k' represent at most $O(\mathrm{ht}(x))$ distinct cosets $\gamma Z(F)$ in $\mathrm{PGL}_2(F)$.

Taking determinant followed by the norm $\mathbb{A}_F^\times / F^\times \to \mathbb{R}$, we conclude that $|z|_{\mathbb{A}}$ belongs to a compact subset of $\mathbb{R}^\times$ that depends only on U. The norm map $\mathbb{A}_F^\times / F^\times \to \mathbb{R}^\times$ being proper, it follows that z itself belongs to a compact subset $\mathbb{A}_F^\times / F^\times$ that depends only on U.

In particular, there is a compact subset $\Omega \subset F_\infty^\times$, depending only on U, and a finite subset $P \subset \mathbb{A}_F^\times$, containing 1 and also depending only on U, such that $z \in F^\times \Omega . P . \prod_{v \text{ finite}} \mathfrak{o}_{F_v}^\times$. Let $\widetilde{U} = U \cdot \{a(z_\infty, z_\infty) : z_\infty \in \Omega\}$. Given a solution to

(8.9), write $z = \delta z_\infty p o$, with $\delta \in F^\times, z_\infty \in \Omega, p \in P, o \in \prod_{v \text{ finite}} \mathfrak{o}_{F_v}^\times$. Then

$$xuk = \gamma a(\delta,\delta) x a(p,p) u' a(z_\infty, z_\infty) k' a(o,o);$$

in particular, taking $\tilde{u} = u' a(z_\infty, z_\infty) \in \widetilde{U}, k'' = k' a(o,o) \in K_{\max}$, the image of $xa(p,p)\tilde{u}k''$ in $\mathbf{X}_{\mathrm{GL}(2)}$ coincides with xuk. So the number of possibilities for the $Z(F)$-coset of γ is bounded above by the fibers of the map $P \times \widetilde{U} \times K_{\max} \to \mathbf{X}_{\mathrm{GL}(2)}$ given by $(p, \tilde{u}, k) \to xa(p,p)\tilde{u}k$. The result follows from Lemma 8.5. □

We shall now need a quantitative version of certain statements in reduction theory. The subsequent lemma is a fancier version of the following statement: the number of $\gamma \in \mathrm{SL}(2,\mathbb{Z})$ that map a fixed $z \in \mathbb{H}$ to the Siegel set $\{x + iy : 0 \le x \le 1, y \ge T\}$ is $\ll 1 + T^{-1}$.

LEMMA 8.7. *Let $g \in \mathrm{GL}_2(\mathbb{A}_F)$ and $Y > 0$ a positive real number. Then*

$$\#\{\gamma \in B(F)\backslash \mathrm{GL}_2(F) : \mathrm{ht}(\gamma g) \ge Y\} \ll_\epsilon 1 + Y^{-1-\epsilon}. \tag{8.10}$$

Here the implicit constant is independent of g. Moreover, suppose $g \in \mathfrak{S}(T)$ with $T \ge 1$. Then

$$\sup\{\mathrm{ht}(\gamma g) : \gamma \notin B(F)\} \le T^{-1}. \tag{8.11}$$

Proof. The proof of (8.10) is not difficult, generalizing in a straightforward way the proof with $F = \mathbb{Q}$. However, it is somewhat notationally tedious; the (hypothetical) reader may wish to simply work out the proof for $F = \mathbb{Q}$, where it is equivalent to the following fact: the number of primitive vectors in a unimodular sublattice of $\mathbb{R}^2$ that are contained in an R-ball is $\ll (1 + R^2)$, uniformly in the lattice. (The result can also be deduced if one admits some basic facts from the theory of Eisenstein series over F, but we wish to rather deduce these basic facts from the present lemma.) We also remark that the entire content of (8.10) lies in the uniformity in g.

Without loss of generality, we take $g \in \mathfrak{S}(T_0)$, where T_0 is sufficiently small that the map $\mathfrak{S}(T_0) \to \mathbf{X}_{\mathrm{GL}(2)}$ is surjective. So $g = \left(\begin{smallmatrix} 1 & t \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} x & 0 \\ 0 & y \end{smallmatrix}\right)k$ with $|xy^{-1}|_{\mathbb{A}} \ge T_0$. Moreover, replacing g by gz, for any $z \in Z(\mathbb{A}_F)$ does not affect the problem, so we may take $y = 1$. Then, for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(F)$, we have

$$\mathrm{ht}(\gamma g) = \frac{|x|_{\mathbb{A}}}{\prod_v \|(x_v c, ct_v + d)\|_v^2}. \tag{8.12}$$

The equivalence class of γ in $B(F)\backslash \mathrm{GL}_2(F)$ depends only on the pair $(c,d) \in F^2$, considered up to $F^\times$ equivalence (i.e., it depends only on $c/d \in F \cup \{\infty\}$.) It suffices, then, to estimate the number

$$\#\{[c:d] \in \mathbb{P}^1(F), \prod_v \|(x_v c, ct_v + d)\|_v^2 \le Y^{-1}|x|_{\mathbb{A}}\}. \tag{8.13}$$

If Ω is any fixed compact subset of $\mathbb{A}_F^\times$, then for $\omega \in \Omega$,

$$\prod_v \|(x_v \omega c, ct_v + d)\|_v \asymp_\Omega \prod_v \|(x_v c, ct_v + d)\|_v.$$

Consider $\mathbb{R}_{>0}$ as embedded in $\mathbb{A}_F^\times$ via $\mathbb{R}_{>0} \hookrightarrow \mathbb{A}_\mathbb{Q}^\times \hookrightarrow \mathbb{A}_F^\times$. Then there is a compact subset $\Omega \in \mathbb{A}_F^\times$ such that $\mathbb{A}_F^\times = F^\times \cdot \Omega \cdot \mathbb{R}_{>0}$.

The size of (8.13) is unaffected by the substitution $(x, t) \mapsto (x\tau, t\tau)$, for any $\tau \in F^\times$. In view of the above remarks we may assume — decreasing Y by a constant that depends only on F — that $x \in \mathbb{R}_{>0}$. Moreover, the size of (8.13) is also unaffected by the substitution $t \mapsto t + \tau$, for $\tau \in F$. We may therefore assume that $|t|_v \leq 1$ for all finite places v.

Fix a set of representatives $\mathfrak{J}_1, \ldots, \mathfrak{J}_h$ for the class group of $\mathfrak{o}_F$; we will assume each $\mathfrak{J}_i$ is integral. For any $[c : d] \in \mathbb{P}^1(F)$, we may find a representative (c, d) so that the ideal $c\mathfrak{o}_F + d\mathfrak{o}_F$ is one of the $\mathfrak{J}_i$; moreover, replacing $(c, d) \in \mathfrak{J}_i^2$ by $(\epsilon c, \epsilon d)$ for $\epsilon \in \mathfrak{o}_F^\times$ does not change the class $[c : d]$.

The restrictions on x, t imply that $\|(x_v c, ct_v + d)\|_v = \|(c, d)\|_v$ for all finite v. Then $\prod_{v \text{ finite}} \|(c, d)\|_v = \mathrm{N}(\mathfrak{J}_i)^{-1}$, the inverse of the norm of $\mathfrak{J}_i = c\mathfrak{o}_F + d\mathfrak{o}_F$. So it will suffice to bound, for each $1 \leq i \leq h$, the quantity

$$\#\Big\{(c, d) \in \mathfrak{J}_i^2/\mathfrak{o}_F^\times : c\mathfrak{o}_F + d\mathfrak{o}_F = \mathfrak{J}_i : \prod_{\infty | v} (\|(x_v c, ct_v + d)\|_v^2) \leq Y^{-1} |x|_\mathbb{A} \mathrm{N}(\mathfrak{J}_i)^2\Big\}.$$

Since $\mathfrak{J}_i$ belongs to a finite set, the quantity $\mathrm{N}(\mathfrak{J}_i)$ is bounded; thus, decreasing Y again as necessary, it suffices to estimate, for each $1 \leq i \leq h$,

$$\#\Big\{(c, d) \in \mathfrak{J}_i^2/\mathfrak{o}_F^\times : c\mathfrak{o}_F + d\mathfrak{o}_F = \mathfrak{J}_i, \prod_{\infty | v} \|(x_v c, ct_v + d)\|_v^2 \leq Y^{-1} |x|_\mathbb{A}\Big\}.$$

There is only one term corresponding to $c = 0$. Otherwise, (c) is a principal ideal divisible by $\mathfrak{J}_i$; let $\mathcal{P}$ be the set of integral principal ideals. Then the size of the set above is precisely

$$\sum_{(c) \in \mathcal{P}} \#\Big\{d \in \mathfrak{J}_i : (c) + d\mathfrak{o}_F = \mathfrak{J}_i, \prod_{\infty | v} \|(x_v c, ct_v + d)\|_v^2 \leq Y^{-1} |x|_\mathbb{A}\Big\}. \tag{8.14}$$

We note that the size of the inner set is independent of the choice of generator for the principal ideal (c). Moreover, the inequality of (8.14) implies that the norm $\mathrm{N}((c))$ of the principal ideal (c) satisfies $\mathrm{N}((c))^2 \leq Y^{-1} |x|_\mathbb{A}^{-1}$.

Let us estimate the number of d that can correspond to a fixed principal ideal (c) in (8.14). Recall that $|x|_\mathbb{A} \geq T_0$ and that x is in the image of the embedding $\mathbb{R}_{>0} \hookrightarrow \mathbb{A}_Q^\times \hookrightarrow \mathbb{A}_F^\times$. In particular, $|x|_v$ is bounded below at each infinite place. Moreover, since $\mathfrak{o}_F^\times$ is a cocompact subgroup of the elements of $F_\infty^\times$ with norm 1, we can choose a representative for the principal ideal (c) so the same is true of $|c|_v$. Note that (cf. (8.6)) $\|(x_v c, ct_v + d)\|_v \asymp (|x_v c|_v + |ct_v + d|_v)$.

So in fact, again decreasing Y as necessary, it will suffice to estimate

$$\sum_{(c)\in\mathcal{P}:\mathrm{N}(c)\leq Y^{-1/2}|x|_{\mathbb{A}}^{-1/2}} \#\Big\{d\in\mathfrak{J}_i:\prod_{\infty|v}(1+|ct_v+d|_v)^2\leq Y^{-1}|x|_{\mathbb{A}}\Big\}. \tag{8.15}$$

To estimate the right-hand side, first observe that if $\{M_v\}_{\infty|v}$ is any set of positive real numbers indexed by the infinite places of F, then $\#\{d\in\mathfrak{J}_i:|ct_v+d|_v\leq M_v \text{ for } \infty|v\}\ll\prod_{\infty|v}(1+M_v)$. Indeed, by subtraction, it will suffice to estimate $\#\{d\in\mathfrak{J}_i:|d|_v\leq 2M_v \text{ for } \infty|v\}$; this amounts to counting points in the lattice $\mathfrak{J}_i\subset F_\infty$ in a region that is the product of a box and a disc; the result is then clear.

Next, if $T\geq 1$, the subset $\{(y_1,\dots,y_d):\prod_i(1+y_i)\leq T\}$ in $\mathbb{R}^d_{>0}$ is contained in the union of $O_\epsilon(T^\epsilon)$ boxes $\{(y_1,\dots,y_d):y_i\leq M_i\}$, where $\prod_i(1+M_i)\ll T$. We may assume $Y^{-1}|x|_{\mathbb{A}}\geq 1$, else (8.15) has no solutions. We conclude that the number of d attached to each principal ideal (c) in (8.15) is $\ll_\epsilon (Y^{-1/2}|x|_{\mathbb{A}}^{1/2})^{1+\epsilon}$.

The number of possibilities for (c) is bounded by the number of integral ideals with norm $\leq Y^{-1/2}|x|_{\mathbb{A}}^{-1/2}$, which is $\ll_\epsilon (Y^{-1/2}|x|_{\mathbb{A}}^{-1/2})^{1+\epsilon}$. Finally, there is one class with $c=0$. We conclude that the number of pairs (c,d) up to equivalence is $\ll Y^{-1-\epsilon}+1$. This proves (8.10).

As for (8.11), suppose $g\in\mathfrak{S}(T)$, so we may write $g=\left(\begin{smallmatrix}x&z\\0&y\end{smallmatrix}\right)k$ with $k\in K_\infty\times K_{\max}$, and $|xy^{-1}|_{\mathbb{A}}\geq T$. Suppose $\gamma=\left(\begin{smallmatrix}\alpha&\beta\\\alpha'&\beta'\end{smallmatrix}\right)$. If $\gamma\notin B(F)$, then $\alpha'\neq 0$. In that case, following the notation of (8.5), we have:

$$\prod_v\|(\alpha'_v,\beta'_v)g_v\|_v\geq\prod_v|\alpha'_vx_v|_v=|x|_{\mathbb{A}}.$$

and therefore, by (8.5), $\mathrm{ht}(\gamma g)\leq|\det(g)x^{-2}|_{\mathbb{A}}\leq T^{-1}$. □

9. **Background on quantitative equidistribution results**

The aim of this section is to quantify various standard equidistribution results (equidistribution of long horocycles, Hecke points, etc.), using the adelic Sobolev norms. As such neither the results nor the methods are new; we just collect together those results we need and provide brief proofs.

As regards the origin of the ideas used here, we have drawn in particular from the work of Clozel-Ullmo, Linnik, Oh, Margulis, Ratner and Sarnak.

9.1. *Decay of matrix coefficients.*

9.1.1. *Local setting.* Our fundamental tool in establishing all these results is the spectral gap, i.e., quantitative mixing properties of real and p-adic flow. As such, we begin by recalling the basic relevant bound on matrix coefficients.

Let $0\leq\alpha\leq 1/2$. Let v be a place of F, and suppose that (V,π) is a unitary representation of $\mathrm{GL}_2(F_v)$ which does not contain, in its spectral decomposition, any complementary series with parameter $>\alpha$. (More formally: V does not *weakly* *contain* such a representation.) Thus $\alpha = 0$ corresponds to V being tempered; on the other hand, any value of $\alpha < 1/2$ implies that V contains no almost invariant vectors.

LEMMA 9.1. *For w_1, w_2 any two K_v-finite elements of V, satisfying $\langle w_1, w_1\rangle = \langle w_2, w_2\rangle = 1$, and any $x \in F_v$,*

$$(9.1)\quad \langle \pi(a(x))w_1, w_2\rangle \ll_{\epsilon,F} \dim(K_v w_1)^{1/2} \dim(K_v w_2)^{1/2}(1+|x|_v)^{\alpha-1/2+\epsilon}.$$

The implicit constant of (9.1) depends only on ϵ. Since we do not know of an available reference, we briefly sketch an argument for (9.1).

Proof. In the case where $\alpha = 0$, i.e., V is *tempered*, then (9.1) is proven in [10] (strictly this is for semisimple groups, and the generalization to reductive groups is established in [30]).

In the general case, we present an argument along the lines of that found in [39, p. 132] for the case $k = \mathbb{R}$ (that reference deals also with other rank one groups, however); this is also related to an argument presented in [10].

Let $(\sigma_{1/2-\alpha}, W)$ be the complementary series of trivial central character with parameter $1/2-\alpha$; let $v^0 \in W$ be a spherical vector of norm 1. The matrix coefficient $\langle gv^0, v^0\rangle$ is a spherical function. It is *positive* and satisfies the bound

$$(9.2)\qquad (1+|x|_v)^{-\alpha+\epsilon} \gg_\epsilon \langle a(x)v^{(0)}, v^{(0)}\rangle \gg_\epsilon (1+|x|_v)^{-\alpha-\epsilon}.$$

This (or an even stronger form) is stated and used in [39]; in the present case it can be verified by direct computation. For example, when v is real, the left-hand coefficient can be expressed explicitly as $\frac{1}{2\pi}\int_{\theta=0}^{2\pi} \frac{d\theta}{(x\cos^2(\theta)+x^{-1}\sin^2(\theta))^\alpha}$; the integrand is positive, thus the positivity; the lower bound follows by considering the contribution near $\theta = 0$, and the upper bound comes from observing that this contribution is dominant.

Then the representation $V \otimes W$ is tempered. Indeed it suffices — again by [10] — to verify that a dense set of matrix coefficients are in $L^{2+\epsilon}$, which follows from (9.2).

Now one may estimate the matrix coefficient $\langle a(x)w_1 \otimes v^0, w_2 \otimes v^0\rangle$ by appealing again to [10]. On the other hand,

$$\langle a(x)w_1 \otimes v^0, w_2 \otimes v^0\rangle = \langle a(x)w_1, w_2\rangle\langle a(x)v^0, v^0\rangle,$$

and thus (9.1) follows from the lower bound of (9.2). □

Let us record a useful further variant. Suppose v is finite. Let $K_1, K_2 \subset K_v$ be subgroups and let σ be the (K_1, K_2)-bi-invariant probability measure supported on $K_1 a(x) K_2$.

Then

$$(9.3)\qquad \|v \star \sigma\|_2 \ll [K_v : K_1]^{1/2}[K_v : K_2]^{1/2}(1+|x|_v)^{\alpha-1/2+\epsilon}\|v\|_2.$$

Indeed, for $i = 1, 2$ let Π_{K_i} be the projection operator $w \mapsto \int_{K_i} kw$ on V, where K_i is endowed with the Haar probability measure. Then

$$\begin{aligned}(9.4) \qquad \|v \star \sigma\|_2 = \sup_{w \in V} \frac{\langle v \star \sigma, w\rangle}{\|w\|_2} &= \sup_{w\in V} \frac{\langle a(x)\Pi_{K_1} v, \Pi_{K_2} w\rangle}{\|w\|_2} \\ &\leq [K_v : K_1]^{1/2}[K_v : K_2]^{1/2}(1 + |x|_v)^{\alpha - 1/2+\epsilon}\|v\|_2.\end{aligned}$$

9.1.2. *Variant for* $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$. Let $0 \leq \alpha \leq 1/2$, suppose $G = \mathrm{SL}_2(\mathbb{R})$, and let V be a unitary representation of G such that V does not weakly contain any complementary series with parameter $\geq \alpha$. The normalization is again so that $\alpha = 0$ corresponds to tempered and $\alpha = 1/2$ corresponds to V not having almost invariant vectors.

Then one has the following variant of (9.1), proved by the same method:

$$\begin{aligned}(9.5) \quad &\left\langle \pi\begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2}\end{pmatrix} w_1, w_2\right\rangle \\ &\qquad \ll_\epsilon \dim(\mathrm{SO}(2)\cdot w_1)^{1/2}\dim(\mathrm{SO}(2)\cdot w_2)^{1/2}(1 + |y|)^{\alpha - 1/2+\epsilon}.\end{aligned}$$

It is convenient to extend the validity of (9.5) beyond the K-finite space by replacing $\dim(\mathrm{SO}(2)w_i)$ by appropriate Sobolev norms. We confine ourselves to the case of main interest, where V is the orthogonal complements of the constants in $L^2(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$, where Γ is a lattice in $\mathrm{SL}_2(\mathbb{R})$. The estimates we are about to describe are, again, not new; estimates for effective mixing of geodesic and horocycle flows in this setting are contained in [32].

For our purposes it would be optimal to use fractional Sobolev norms; since we have not defined these, we shall use a rather crude form of interpolation instead.

Thus let $f_1, f_2 \in C^\infty(\Gamma\backslash\mathrm{SL}_2(\mathbb{R}))$. One expands both f_1 and f_2 into a sum of SO(2)-types and applies (9.5). Indeed, write for $i \in \{1, 2\}$ an expansion $f_i = \sum_{n=-\infty}^{\infty} f_i^{(n)}$, where $f_i^{(n)}$ transforms under the character $\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta)\end{pmatrix} \mapsto e^{in\theta}$. Expanding:

$$\begin{aligned}&(9.6)\\ &\left\langle\begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2}\end{pmatrix} f_1, f_2\right\rangle = \sum_{n,m\in\mathbb{Z}} \left\langle\begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2}\end{pmatrix} f_1^{(n)}, f_2^{(m)}\right\rangle \\ &\qquad\qquad \ll_\epsilon (1 + |y|)^{\alpha-1/2+\epsilon} \sum_{n,m} \|f_1^{(n)}\|_2 \|f^{(m)}\|_2 \\ &\qquad\qquad = (1 + |y|)^{\alpha - 1/2+\epsilon}\left(\sum_n \|f_1^{(n)}\|_2\right)\left(\sum_m \|f_2^{(m)}\|_2\right).\end{aligned}$$

Our definitions of the Sobolev norms (Section 2.9.2) are so that $S_{2,1}(f_1)^2 \gg \sum_n (1 + |n|)^2 \|f_1^{(n)}\|_2^2$, and similarly for f_2. On the other hand, it is an elementary estimate that

$$\Big(\sum_n \|f_1^{(n)}\|_2\Big)^2 \ll_\epsilon \Big(\sum_n \|f_1^{(n)}\|_2^2(1+|n|)^2\Big)^{1/2+\epsilon}\Big(\sum_n \|f_1^{(n)}\|_2^2\Big)^{1/2-\epsilon}.$$

It follows from this that for any $k, k' \in \mathrm{SO}(2)$ we have the matrix coefficient bound

$$\left|\left\langle k\begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2}\end{pmatrix}k' f_1, f_2\right\rangle\right| \\ \ll (1+|y|)^{\alpha-1/2+\epsilon}(S_{2,1}(f_1)S_{2,1}(f_2))^{1/2+\epsilon}\|f_1\|^{1/2-\epsilon}\|f_2\|^{1/2-\epsilon}, \tag{9.7}$$

at least for f_1, f_2 which are SO(2)-finite. But the general case of smooth f_1, f_2 follows from density.

Note that in (9.7) that the factor $\|f_1\|^{1/2-\epsilon}S_{2,1}(f_1)^{1/2+\epsilon}$ is a crude substitute for the fractional $(1/2+\epsilon)$- Sobolev norm of f_1.

9.2. *Pointwise bounds.* In this section, we make free use of the adelic Sobolev norms introduced in Section 2.9.3. We recall the definition $S_{p,d} := S_{p,d,1/p}$. We also recall that in statements of the form $|L(f)| \ll S_{p,d}(f)$, for certain linear functionals L, we shall allow the implicit constant of $\ll$ to depend on p and d without explicit mention.

LEMMA 9.2. *Let $f \in C_\omega^\infty(\mathbf{X}_{\mathrm{GL}(2)})$ and let $x \in \mathbf{X}_{\mathrm{GL}(2)}$. Then, for any $p \geq 2$ and $d \gg 1$,*

$$|f(x)| \ll \mathrm{ht}(x)^{1/p}S_{p,d}(f). \tag{9.8}$$

Moreover, if $F \in C^\infty(\mathbf{X}\times\mathbf{X})$, and $p > 2, d \gg 1$,

$$\int_{\mathbf{X}} |F(x,x)|dx \ll S_{p,d}(F). \tag{9.9}$$

Proof. As in (2.1), set $K_f = \prod_{v \text{ finite}} K_{v,f}$, where $K_{v,f}$ is the stabilizer of f in K_v. Fix an open neighborhood of the identity $U \subset \mathrm{GL}_2(F_\infty)$. Consider the map $\Pi : U\cdot K_f \to \mathbf{X}$ defined by $(u,k) \mapsto xuk$. By Lemma 8.6, the fibers are unions of at most $O(\mathrm{ht}(x))$ sets, each of the form $yZ(\mathbb{A}_F)\cap UK_f$. Moreover, for any $y \in U\cdot K_f$, the measure of $\{z \in \mathbb{A}_F^\times : ya(z,z) \in U\cdot K_f\}$ is bounded above by a constant depending only on U. Indeed, the set of such z is contained in a fixed compact subset of $\mathbb{A}_F^\times$ that depends only on U.

Equip $U\cdot K_f$ with the restriction of Haar measure from $\mathrm{GL}_2(\mathbb{A}_F)$. From the preceding paragraph, one easily deduces that the push-forward of this measure to $\mathbf{X}$, under $(u,k) \mapsto xuk$, is bounded above by $C\cdot\mathrm{ht}(x)$ times the measure on $\mathbf{X}$, where the constant C depends only on U.

Then:

$$\int_{u\in U}|f(xu)|^p=\mathrm{vol}(K_f)^{-1}\int_{u\in U,k\in K_f}|f(xuk)|^p\,du\,dk \tag{9.10}$$
$$\ll \mathrm{ht}(x)[K_{\max}:K_f]\int_{\mathbf{X}}|f(x)|^p\,d\mu_{\mathbf{X}}(x).$$

Equation (9.10) holds with f replaced by $\mathscr{D}f$, for $\mathscr{D}$ any fixed monomial in $\mathrm{Lie}(\mathrm{GL}_2(F_\infty))$. The standard Sobolev estimate, applied to the function $u\mapsto f(xu)$ on the real manifold U, implies that $|f(x)|\ll \mathrm{ht}(x)^{1/p}PS_{p,d,1/p}(f)$ for sufficiently large d. (Indeed, it suffices to take any $d>\dim(U)/2=2[F:\mathbb{Q}]$.) Then Remark 8.1, (2) implies the conclusion.

As for the second conclusion, we proceed in a similar fashion as above (with $\mathbf{X}$ replaced by $\mathbf{X}\times\mathbf{X}$) to obtain the estimate $|F(x,y)|\ll \mathrm{ht}(x)^{1/p}\mathrm{ht}(y)^{1/p}S_{p,d}(F)$. It is easy to see that $\int_{\mathbf{X}}\mathrm{ht}(x)^{2/p}dx<\infty$ for $p>2$, and the conclusion follows. $\square$

The next lemma quantifies the rapid decay of a cuspidal function, or more generally a truncated automorphic function, in the cusp. Recall that for $T_0>0$ we have defined the Siegel domain $\mathfrak{S}(T_0)$ in Section 8.2.

LEMMA 9.3. *Let $f\in C^\infty_\omega(\mathbf{X}_{\mathrm{GL}(2)})$. Put $f^N(g)=\int_{N(F)\backslash N(\mathbb{A}_F)}f(ng)dn$, where the measure on $N(F)\backslash N(\mathbb{A}_F)$ is the $N(\mathbb{A}_F)$-invariant probability measure. Then for $x\in\mathfrak{S}(T_0)$, $p\geq 2, k\geq 0$ and $d\gg 1$,*

$$|f(x)-f^N(x)|\ll_{T_0}\mathrm{ht}(x)^{1/p-k}S_{p,d,1/p+k}(f). \tag{9.11}$$

Proof. We may assume that $x\in N(\mathbb{A}_F)A(\mathbb{R})\Omega(K_\infty\times K_{\max})$, for some fixed compact set $\Omega\subset A(\mathbb{A}_F)$. Here $A(\mathbb{R})$ is regarded as a subset of $A(F_\infty)$ via the natural inclusion $\mathbb{R}\hookrightarrow F_\infty$. Write accordingly $x=na\omega k$, where $\omega\in\Omega$.

Consider the function on F_∞ defined by $g(t)=f(n(t)x)-f^N(x)$. Let $\Lambda=\{t\in\mathfrak{o}_F: n(t)\in \mathrm{GL}_2(F_\infty)\omega k K_f k^{-1}\omega^{-1}\}$, where K_f is again as in (2.1). Then the function $g(t)$ is invariant under Λ (thought of as a sublattice of F_∞).

One sees that, since ω belongs to the fixed compact Ω, the covolume bound $\mathrm{vol}(F_\infty/\Lambda)\ll[K_{\max}:K_f]$. Moreover, since Λ may be regarded as a fractional ideal of F, the homothety class of Λ lies in a fixed compact set in the space of homothety classes of lattices in F_∞. Also, $g(t)$ defines a function on F_∞/Λ, with integral 0.

Suppose now that G is a smooth function on $\mathbb{R}^d/L$, for some $d>1$ and some lattice $L\subset\mathbb{R}^d$, with integral 0. Let $\|G\|_{(i)}=\sup_{\mathscr{D}}\sup_{z\in\mathbb{R}^d/L}|\mathscr{D}G(z)|$, where $\mathscr{D}$ varies over all monomials in $\partial_1,\dots,\partial_d$ of exact order i. Then an elementary argument shows that $\|G\|_{(0)}\ll\mathrm{vol}(\mathbb{R}^d/L)^{i/d}\|G\|_{(i)}$, and the implicit constant may be taken to vary continuously with the homothety class of L.

Apply this lemma to the function g on F_∞/Λ, with $i=k[F:\mathbb{Q}]$ for some $k\geq 1$. The norm $\|g\|_{(k[F:\mathbb{Q}])}$, in the sense of the above paragraph, is bounded, by Lemma 9.2 and an elementary computation, by $\mathrm{ht}(x)^{-k}(\mathrm{ht}(x)^{1/p}PS_{p,d'}(f))$, for some $d' \gg 1$. It follows that

$$\sup |g(t)| \ll \mathrm{ht}(x)^{1/p-k} PS_{p,d'}(f)[K_{\max} : K_f]^k = \mathrm{ht}(x)^{1/p-k} PS_{p,d',1/p+k}(f).$$

Applying Remark 8.1, (2), we conclude

$$|f(x) - f^N(x)| \ll \mathrm{ht}(x)^{1/p-k} S_{p,d',1/p+k}(f). \qquad \square$$

LEMMA 9.4. *Suppose* $f \in C_\omega^\infty(\mathbf{X}_{\mathrm{GL}(2)})$ *is cuspidal. Then* $S_{\infty,d,\beta}(f) \ll S_{2,d',\beta+3/2}(f)$, *for sufficiently large* d'.

Proof. By Lemma 9.3 for f cuspidal, applied with $p = 2, k = 1$, we see that $|f(x)| \ll S_{2,d,3/2}(f)$ for $d \gg 1$. Applying this inequality to $\mathscr{D} f$, for $\mathscr{D}$ in the universal enveloping algebra of $\mathrm{GL}_2(F_\infty)$, we see that $PS_{\infty,d,\beta}(f) \ll PS_{2,d',\beta+3/2}(f)$, for d' sufficiently large. This inequality holds for cuspidal f.

Let Π be the L^2-orthogonal projection onto the space of cuspidal functions; then Π commutes with $\mathrm{GL}_2(\mathbb{A}_F)$, and it follows that

$$PS_{\infty,d,\beta}(\Pi f) \ll PS_{2,d',\beta+3/2}(\Pi f) \le PS_{2,d',\beta+3/2}(f)$$

for arbitrary $f \in C_\omega^\infty(\mathbf{X}_{\mathrm{GL}(2)})$. Now Remark 8.1, (3) (or, more precisely, a trivial modification thereof) implies the conclusion. $\square$

9.3. *Equidistribution of long horocycles and closed horospheres.* Let G be a semisimple group, $\Gamma \subset G$ a lattice, U a unipotent subgroup of G. It is well-known that one can prove, in a quantitative fashion, the equidistribution of U-orbits on $\Gamma\backslash G$ if U is a *horospherical* subgroup, i.e., the unipotent radical of a proper parabolic subgroup. We shall quantify two instances of this that will be of interest to us.

We emphasize that neither the results nor the techniques of this section are new; we have included proofs *only* to keep the present paper as self-contained as possible.

Effective estimates for equidistribution of long horocycles on quotients of $\mathrm{SL}_2(\mathbb{R})$ are already implicit in the work of Ratner [31] and [33], where the effective mixing of the horocycle flow is used. We will proceed in a closely related fashion, using the mixing property of the Cartan action; again, this is definitely not new and appears already, although in a different context, in the doctoral thesis of Margulis (reprinted in [26]).

9.3.1. *Equidistribution of long horocycles in hyperbolic* 2*-space.* Let $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$ be a lattice such that $L^2(\Gamma\backslash \mathrm{SL}(2,\mathbb{R}))$ does not contain any complementary series representation with parameter $> \alpha$, for any $0 \le \alpha < 1/2$. (That is: $\alpha \in [0, 1/2)$ is such that all nonzero eigenvalues of the hyperbolic Laplacian $-y^2(\partial_{xx} + \partial_{yy})$ on $\Gamma\backslash\mathbb{H}^2$ are bounded below by $1/4 - \alpha^2$.)

We define $n, a, \bar{n}$ as in (3.1).

The following lemma quantifies the equidistribution of long horocycles. Results of this type are already implicit in [31] and [33]. This problem is analyzed in much more detail than we go into, in [43] and [14].

LEMMA 9.5. *Assume* Γ *is cocompact, and let* $x_0 \in \Gamma\backslash \mathrm{SL}_2(\mathbb{R})$.

$$\left|\frac{1}{T}\int_{t=0}^{T} f(x_0 n(t))dt - \int_{\Gamma\backslash \mathrm{SL}_2(\mathbb{R})} f(g)dg\right| \ll_\epsilon T^{\frac{\alpha-1/2}{2}+\epsilon} S_{\infty,1}(f). \tag{9.12}$$

Proof. The idea (which is certainly not new; cf. remarks at start of §9.3) is that, upon flowing a small ball in $\Gamma\backslash G$ for a long time by the geodesic flow, it turns into a narrow neighborhood of a long horocycle. One thereby can deduce the equidistribution of the long horocycle from the mixing properties of the geodesic flow.

Let $N, A, \bar{N}$ be the images of $n, a, \bar{n}$ respectively. Let g_1 be a smooth function of compact support on the real line, with integral $\int_{-\infty}^{\infty} g_1(x)dx = 1$. It will remain fixed for all time throughout our arguments. Fix $1 > \delta > 0$ and let $g_\delta : \mathbb{R} \to \mathbb{R}$ be the convolution of the characteristic function of $[0,1]$ with $g_1(x/\delta)\delta^{-1}$; that is to say

$$g_\delta(x) = \delta^{-1}\int_{t=0}^{1} g_1\Big(\frac{x-t}{\delta}\Big)dt.$$

Then g_δ is a smooth function of integral 1, which is supported in a small interval around $[0,1]$.

Define a probability measure μ_δ on $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ via the rule

$$\mu_\delta(f) = \delta^{-1}\int_{x,y,z\in\mathbb{R}} f(x_0 n(x)a(e^y)\bar{n}(z))g_\delta(x)g_1(y/\delta)g_1(z)dxdydz.$$

In words, μ_δ is a measure supported on a small box around x_0; this box has width $O(1)$ in the N and $\bar{N}$ directions, and $O(\delta)$ in the A direction. When we flow this by A, it will become a measure supported along a box that closely approximates an N-orbit.

We observe that

$$\mu_\delta(a(T^{-1})f) = \frac{1}{\delta}\int_{x,y,z} f(x_0 a(T)^{-1}n(x)a(e^y)\bar{n}(z))g_\delta(x/T)g_1(y/\delta)g_1(Tz)dxdydz.$$

On the other hand, for any fixed $x_1 \in \Gamma\backslash \mathrm{SL}_2(\mathbb{R})$, we note that

$$\left|\delta^{-1}T\int f(x_1 a(e^y)\bar{n}(z))g_1(y/\delta)g_1(Tz)dydz - f(x_1)\right| \ll \max(T^{-1},\delta)S_{\infty,1}(f). \tag{9.13}$$

Indeed, (9.13) merely quantifies the fact that the right-hand side integral is against a probability measure supported in a very small ball (of size $\max(T^{-1},\delta)$) around x_1. Since $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})$ is assumed compact, the implicit constant of (9.13) may be taken independent of x_1.

Consequently,

$$\left|\frac{1}{T}\int_{t\in\mathbb{R}} g_\delta(t/T)f(x_0 a(T)^{-1}n(t))dt - \mu_\delta(a(T)^{-1}f)\right| \ll \max(T^{-1},\delta)S_{\infty,1}(f). \tag{9.14}$$

On the other hand, the measure μ_δ has a continuous distribution function h_δ, i.e. $\mu_\delta(f) = \int_{\Gamma\backslash \mathrm{SL}_2(\mathbb{R})} f(g)\cdot h_\delta(g)dg$, and $\mu_\delta(a(T)^{-1}f)$ may be estimated using (9.7), i.e., the decay of matrix coefficients.

A routine computation shows that $\|h_\delta\|_{L^2} \ll \delta^{-1/2}$ and $S_{2,1}(h_\delta) \ll \delta^{-3/2}$; on account of the cocompactness of $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$, both these estimates are uniform in x_0.

Using (9.7) now yields

$$\left|\mu_\delta(a(T)^{-1}f) - \int_{\Gamma\backslash\mathrm{SL}_2(\mathbb{R})} f(g)dg\right| \ll_\epsilon T^{\alpha-1/2+\epsilon} S_{2,1}(f)\delta^{-1-\epsilon}. \tag{9.15}$$

Finally, note that if $\chi_{[0,T]}$ denotes the characteristic function of $[0,T]$ in the real line, then $\frac{1}{T}\int_{t\in\mathbb{R}} |g_\delta(t/T) - \chi_{[0,T]}(t)|dt \ll \delta$. It follows that

(9.16)

$$\frac{1}{T}\left|\int_0^T f(x_0 a(T)^{-1}n(t))dt - \int_t g_\delta(t/T) f(x_0 a(T)^{-1}n(t))dt\right| \ll \delta\cdot S_{\infty,0}(f).$$

Combining (9.14), (9.15) and (9.16), and replacing x_0 by $x_0 a(T)$, we conclude that the left-hand side of (9.12) is bounded by

$$O_\epsilon\left(S_{\infty,1}(f)(\max(T^{-1},\delta) + T^{\alpha-1/2+\epsilon}\delta^{-1-\epsilon} + \delta)\right).$$

We choose $\delta^2 = T^{\alpha-1/2}$ to obtain the claimed conclusion. $\square$

9.3.2. *Equidistribution of large horospheres on higher rank groups.* We now prove quantitative equidistribution of large closed horospheres. This result is well-known and generalizes the result of Sarnak, that the closed horocycle $\{x+iy\}_{0\le x\le 1}$ is equidistributed in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, as $y \to 0$.

We shall follow the notation of Section 3.2, which we briefly reprise. Let G be a connected semisimple (real) Lie group, $\Gamma \subset G$ a lattice, $K \subset G$ the maximal compact subgroup, $\mathfrak{g}$ the Lie algebra of G, and $H \in \mathfrak{g}$ a semisimple element. Fix arbitrarily a norm $\|\cdot\|$ on $\mathfrak{g}$. We equip G with the Haar measure in which $\Gamma\backslash G$ has volume 1. Let $\exp : \mathfrak{g} \to G$ be the exponential map. Let $\mathfrak{u}$ be the sum of all negative root spaces for H, and let $U = \exp(\mathfrak{u}) \subset G$. Let $x_0 \in \Gamma\backslash G$ be so that $x_0 U$ is compact; note that the existence of such x_0 implies that $\Gamma\backslash G$ is noncompact.

Let $x_t = x_0 \exp(tH)$, and let Δ_t be the stabilizer of x_t in U. We denote by $\langle\cdot,\cdot\rangle_{L^2(\Gamma\backslash G)}$ the inner product in the Hilbert space $L^2(\Gamma\backslash G)$.

LEMMA 9.6. *There is $\kappa_1 > 0$ such that, for any $f, g \in C^\infty(\Gamma\backslash G)$ and for any $U \in \mathfrak{u}$ with unit length* (*with respect to the fixed norm $\|\cdot\|$ on $\mathfrak{g}$*) *we have*

(9.17)

$$\left|\langle\exp(tH)\cdot f, g\rangle - \int_{\Gamma\backslash G} f \int_{\Gamma\backslash G} g\right| \ll \exp(-\kappa_1|t|) S_{\infty,\dim(K)}(f) S_{\infty,\dim(K)}(g),$$

$$\left|\langle\exp(sU)\cdot f, g\rangle - \int_{\Gamma\backslash G} f \int_{\Gamma\backslash G} g\right| \ll (1+|s|)^{-\kappa_1} S_{\infty,\dim(K)}(f) S_{\infty,\dim(K)}(g).$$

Of course the constant κ_1 will depend on the choice of the norm $\|\cdot\|$.

Proof. This follows from a nice result of Kleinbock and Margulis; see [23]. (The orthogonal complement L^2_0 of the identity representation in $L^2(\Gamma\backslash G)$ is isolated, by [23, Th. 1.12], from the trivial representation in the unitary dual of $\widehat{G}$. A sufficiently high tensor power of L^2_0 is therefore tempered, whereupon one applies the bounds of [10].) Note that [23] only claims the result (in effect) with $S_{\infty,d}$ for some d. The fact that we can take $d = \dim(K)$ follows by explicating the argument just sketched; indeed, the necessary argument is substantially that presented in Section 9.1.2, with $(\mathrm{SL}_2(\mathbb{R}), \mathrm{SO}(2))$ replaced by (G, K). □

Recall the definition of ν_T from Section 3.2; i.e., $\nu_T(f) = \frac{\int_{\Delta_T\backslash U} f(x_T u)du}{\mathrm{vol}(\Delta_T\backslash U)}$. Thus ν_T is the measure supported on a closed horosphere, and this horosphere expands as $T \to \infty$. One deduces from Lemma 9.6 that the measures ν_T are equidistributed as $T \to \infty$:

LEMMA 9.7. *Set* $\kappa_2 = \frac{\kappa_1}{\dim(G)+\dim(K)+1}$, κ_1 *being as in the previous lemma. Then, for* $T \geq 0$ *and* $f \in C^\infty(\Gamma\backslash G)$,

$$\left|\nu_T(f) - \int_{\Gamma\backslash G} f\right| \ll e^{-\kappa_2 T} S_{\infty,\dim(K)}(f).$$

Proof. The idea is identical to Lemma 9.5 and we refer to the first paragraph of that proof for a description of it.

Fix a left-invariant Riemannian metric on G. This descends to a metric on $\Gamma\backslash G$. We first choose some "smoothing kernels" on G. For each $\epsilon > 0$, choose a function $k_\epsilon \in C^\infty(G)$ such that k_ϵ is positive, supported in an ϵ-neighborhood of the identity, $\int_G k_\epsilon = 1$, and so that for any $X_1, X_2, \dots, X_l \in \mathfrak{g}$ we have

$$\sup_{g\in G} |X_1 \dots X_l k_\epsilon| \ll_{X_1,\dots,X_l} \epsilon^{-l-\dim(G)}. \tag{9.18}$$

It is easy to see this is possible (for example: choose an appropriate sequence of functions on $\mathfrak{g}$ and transport to G via the exponential map).

The measure ν_0 is a U-invariant probability measure supported on the closed orbit $x_0 U$. $\nu_0 \star k_\epsilon$ is supported in an ϵ-neighborhood of $x_0 U$ and is given by integration against a C^∞ density function g_ϵ, that is: $\nu_0 \star k_\epsilon(f) = \int_{\Gamma\backslash G} f g_\epsilon$.

Moreover, it follows from (9.18) that g_ϵ satisfies the bounds $S_{\infty,l}(g_\epsilon) \ll \epsilon^{-l-\dim(G)}$, for any $l \geq 0$.

The translate of $\nu_0 \star k_\epsilon$ by $\exp(-TH)$ is supported in an ϵ-neighborhood of $x_T U$; note it is essential that $T \geq 0$ for this. (Recall that our conventions are such that the right translate of the point mass at x by $g \in G$ is the point mass at xg^{-1}; see §2.1.)

In fact, one verifies that

$$|\nu_T(f) - \nu_0 \star k_\epsilon(\exp(TH)\cdot f)| \ll \epsilon S_{\infty,1}(f). \tag{9.19}$$

Indeed, let $g \in \mathrm{supp}(k_\epsilon)$ and let δ_g be the point mass at g. It suffices to check that the identical bound holds for $\left|\nu_T(f) - \nu_0 \star \delta_g(\exp(TH)\cdot f)\right|$, which equals

$|\nu_T(f) - \nu_0(g\exp(TH)\cdot f)|$. Let $\mathfrak{b}$ the sum of nonnegative root spaces for H on $\mathfrak{g}$. If ϵ is sufficiently small, we may write $g = um$, with $u \in \exp(\mathfrak{u})$ and $m \in \exp(\mathfrak{b})$. Moreover, again if ϵ is sufficiently small, u, m lie in a $C\epsilon$-neighborhood of the identity, for some fixed constant C. Then $\exp(-TH)g\exp(TH) = u'm'$, with $u' \in \exp(\mathfrak{u})$ and where $m' \in \exp(\mathfrak{b})$ is in a $C'\epsilon$-neighborhood of the identity, for some absolute C'. Also, $\nu_0(g\exp(TH)\cdot f) = \nu_T(m'f)$. Thus it suffices to bound $|\nu_T(f) - \nu_T(m'f)|$. But the L^∞-norm of $f - m'\cdot f$ is $\ll \epsilon S_{\infty,1}(f)$.

On the other hand, by Lemma 9.6, for $T \geq 0$: we have

$$\left|\nu_0 \star k_\epsilon(\exp(TH)\cdot f) - \int_{\Gamma\backslash G} f\right| = \left|\langle \exp(TH)f, g_\epsilon\rangle_{L^2(\Gamma\backslash G)} - \int_{\Gamma\backslash G} f\right| \ll \exp(-\kappa_1 T) S_{\infty,\dim(K)}(f)\epsilon^{-\dim(G)-\dim(K)}. \tag{9.20}$$

It follows from this and (9.19) that

$$\left|\nu_T(f) - \int_{\Gamma\backslash G} f\right| \ll (\epsilon + \exp(-\kappa_1 T)\epsilon^{-\dim(G)-\dim(K)}) S_{\infty,\dim(K)}(f).$$

To conclude, take $\epsilon = \exp(-\frac{\kappa_1 T}{\dim(G)+\dim(K)+1})$. □

9.4. *The equidistribution of Hecke orbits and p-adic horocycles.* In this section, we prove some "p-adic" equidistribution statements, pertaining to the equidistribution of Hecke points and p-adic horocycles.

In the lemmas that follow, $\mathfrak{f}$ will be a prime ideal of F, $\overline{\mu}_\mathfrak{f}$, the normalized Hecke measure defined subsequent to (2.6), and $[\mathfrak{f}]$ as defined in Section 2.5.

The first lemma is an adelic version of the fact that the Hecke orbit $T_q(z)$ of a point $z \in \mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}$ is equidistributed, as $q \to \infty$.

LEMMA 9.8. *Let $f \in C^\infty_\omega(\mathbf{X}_{\mathrm{GL}(2)})$ and $\mathfrak{f}$ an ideal of F. Then, for $x_0 \in \mathbf{X}_{\mathrm{GL}(2)}$, $d \gg 1$,*

$$\left| f \star \overline{\mu}_\mathfrak{f}(x_0) - \sum_{\chi^2=\omega} \chi(\mathfrak{f})\chi(x_0)\int_{x\in\mathbf{X}} f(x)\overline{\chi(x)}d\mu_{\mathbf{X}}(x)\right| \ll \mathrm{N}(\mathfrak{f})^{\alpha-1/2+\epsilon}\mathrm{ht}(x_0)^{1/2} S_{2,d}(f). \tag{9.21}$$

Here $\chi(x)$ denotes the function $g \mapsto \chi(\det(g))$ on $\mathbf{X}_{\mathrm{GL}(2)}$.

Proof. Let $\mathcal{P}$ be the projection defined in Section 2.7. Let E be the endomorphism $f \mapsto (f - \mathcal{P}f)\star\overline{\mu}_\mathfrak{f}$ of $C^\infty_\omega(\mathbf{X}_{\mathrm{GL}(2)})$. The operator E has norm $\ll_\epsilon \mathrm{N}(\mathfrak{f})^{\alpha-1/2+\epsilon}$ with respect to the L^2-norm (this follows from Lemma 2.1 and the bounds of §9.1). By Lemma 8.3 it follows that the operator norm of E with respect to $S_{2,d,\beta}$ is also $\ll_\epsilon \mathrm{N}(\mathfrak{f})^{\alpha-1/2+\epsilon}$.

The left-hand side of (9.21) is exactly $|Ef(x_0)|$. Now apply Lemma 9.2, with $p = 2$, to conclude. □

The next lemma is an adelic version of the following (again closely connected to equidistribution of Hecke points). Let $Y(p)$ be embedded in $Y(1)\times Y(1)$ (notation of discussion after Proposition 4.1). Then $Y(p)$ is equidistributed as $p\to\infty$. The quantification of this is slightly complicated by noncompactness; in particular, we must use Sobolev norms $S_{p,d}$ for $p>2$. (cf. discussion in §2.9.1).

LEMMA 9.9. *Let $\mathfrak{q}$ be a prime ideal of $\mathfrak{o}_F$. Let $F\in C^\infty(\mathbf{X}\times\mathbf{X})$ be* $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{q}})\times\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{q}})$ *invariant. Then, for any $d\gg 1$, $p>2$,*

$$\left|\int_{\mathbf{X}}F(x,xa([\mathfrak{q}]))dx-\sum_{\chi^2=1}\chi([\mathfrak{q}])\int_{\mathbf{X}}F(x,y)\chi(x)\chi(y)d\mu_{\mathbf{X}}(x)d\mu_{\mathbf{X}}(y)\right| \ll_\epsilon \mathrm{N}(\mathfrak{q})^{\frac{2\alpha-1}{p}+\epsilon}S_{p,d}(F). \tag{9.22}$$

Proof. Let σ be the measure $\delta_1\times\overline{\mu}_\mathfrak{q}$ on $\mathrm{PGL}_2(F_\mathfrak{q})\times\mathrm{PGL}_2(F_\mathfrak{q})$, where δ_1 is the measure consisting of a point mass at the identity. Recalling (see (2.6) in the case of a prime ideal, and §2.5 for the definition of $K_\mathfrak{q}$) that $\overline{\mu}_\mathfrak{q}$ is the $K_\mathfrak{q}$-bi-invariant probability measure supported on $K_\mathfrak{q}a([\mathfrak{q}])K_\mathfrak{q}$, we note that

$$(F\star\sigma)(x,x)=\int_{k_1,k_2\in K_\mathfrak{q}}F(x,xk_1a([\mathfrak{q}])k_2)dk_1dk_2=\int_{K_\mathfrak{q}}F(xk,xka([\mathfrak{q}]))dk,$$

where we equip $K_\mathfrak{q}$ with the Haar measure of mass 1, and we use the $\mathrm{PGL}_2(\mathfrak{o}_{F_\mathfrak{q}})$-invariance of F at the second step. It follows that

$$\int_{\mathbf{X}}F(x,xa([\mathfrak{q}]))dx=\int_{\mathbf{X}}(F\star\sigma)(x,x)dx. \tag{9.23}$$

Let $\mathcal{P}_2$ be as in Section 2.7. Let E be the endomorphism of $C^\infty(\mathbf{X}\times\mathbf{X})$ defined by $E(F)=(F-\mathcal{P}_2F)\star\sigma$. Combining (9.23) and the easily verified equality

$$\int_{\mathbf{X}}(\mathcal{P}_2F\star\sigma)(x,x)dx=\sum_{\chi^2=1}\chi([\mathfrak{q}])\int_{\mathbf{X}}F(x,y)\chi(x)\chi(y)d\mu_{\mathbf{X}}(x)d\mu_{\mathbf{X}}(y),$$

we see that the left-hand side of (9.22) is precisely $|\int_{\mathbf{X}}EF(x,x)dx|$.

Since $\mathcal{P}_2$ does not increase L^∞-norms, and σ is a probability measure, it follows that the operator norm of E with respect to the L^∞-norm is ≤ 2. Moreover, the operator norm of E with respect to the L^2-norm is $\ll\mathrm{N}(\mathfrak{q})^{\alpha-1/2}$, as follows from Lemma 2.1.

Lemma 8.3 now implies that for $2\le p\le\infty$ we have the majorization $S_{p,d}(EF)\ll\mathrm{N}(\mathfrak{q})^{\frac{2\alpha-1}{p}+\epsilon}S_{p,d}(F)$. Now Lemma 9.2 shows that

$$\left|\int_{\mathbf{X}}EF(x,x)dx\right|\ll S_{p,d}(EF)\ll\mathrm{N}(\mathfrak{q})^{\frac{2\alpha-1}{p}+\epsilon}S_{p,d}(F)$$

for $p>2$, $d\gg 1$; whence the conclusion of the Lemma. $\square$

The next lemma shows the equidistribution of certain p-adic horocycle orbits, as p varies. The idea will be as follows: (speaking very loosely, in the case of

SL_2) a typical p-adic horocycle orbit, when projected to $\mathrm{SL}(2,\mathbb{Z})\backslash\mathbb{H}$, looks like $\{z+\frac{i}{p}\}_{0\leq i\leq p-1}$. This set looks very much like the image, under the p-Hecke operator, of the point pz. Thus one can deduce distribution properties of the p-adic horocycle orbit from some standard facts about Hecke operators.

This is a rather *ad hoc* argument. Let us say a few words about why this problem does not quite fit into the usual setup of such questions. We are proving statements about the distribution of e.g., p-adic horocycles *when p varies*. This does not fit easily into the usual context of such matters, where one considers, e.g., a fixed unipotent flow on an S-arithmetic homogeneous space. It would be interesting to have a more conceptual and natural way of treating such questions, in the aspect where "p varies."

LEMMA 9.10. *Let $f\in C^{\infty}(\mathbf{X})$ and let $\mathfrak{f}$ be an integral ideal of $\mathfrak{o}_F$, factorizing as $\mathfrak{f}=\prod_{\mathfrak{q}}\mathfrak{q}^{e_{\mathfrak{q}}}$. For each $\mathfrak{q}|\mathfrak{f}$, let $s_{\mathfrak{q}}\geq 0$ be a nonnegative integer, and suppose f is invariant by $\prod_{\mathfrak{q}|\mathfrak{f}}K_0[\mathfrak{q}^{s_{\mathfrak{q}}}]$. Put $\mathfrak{m}=\prod_{\mathfrak{q}|\mathfrak{f}}\mathfrak{q}^{s_{\mathfrak{q}}}$. Let $\eta_{\mathfrak{f}}$ be the Haar probability measure on $\prod_{\mathfrak{q}|\mathfrak{f}}N(\mathfrak{q}^{-e_{\mathfrak{q}}}\mathfrak{o}_{\mathfrak{q}})$ and dh the Haar probability measure on $\mathrm{SL}_2(F)\backslash\mathrm{SL}_2(\mathbb{A}_F)$.*

Then, for $y\in F^{\times}\backslash\mathbb{A}_F^{\times}$,

$$\left|f\star\eta_{\mathfrak{f}}(a(y))-\int_{h\in\mathrm{SL}_2(F)\backslash\mathrm{SL}_2(\mathbb{A}_F)}f(ha(y))dh\right| \\ \ll_{\epsilon}\mathrm{N}(\mathfrak{f})^{\alpha-1/2+\epsilon}\max\left(\mathrm{N}(\mathfrak{f})|y|,\frac{1}{\mathrm{N}(\mathfrak{f})|y|}\right)^{1/2}\mathrm{N}(\mathfrak{m})^{3/2+\epsilon}S_{2,d}(f). \tag{9.24}$$

Proof. As usual let $K_{v,f}$ be the stabilizer of f in $K_v=\mathrm{GL}_2(\mathfrak{o}_{F_v})$, so that $K_{\mathfrak{q},f}$ contains $K_0[\mathfrak{q}^{s_{\mathfrak{q}}}]$ for each $\mathfrak{q}|\mathfrak{f}$.

We now define a measure $\tilde{\eta}_{\mathfrak{q}}$ on $\mathrm{PGL}_2(F_{\mathfrak{q}})$ for each $\mathfrak{q}|\mathfrak{f}$. It will "approximate" $\eta_{\mathfrak{f}}$ but will be composed of Hecke operators.

For those $\mathfrak{q}$ such that $s_{\mathfrak{q}}=0$, put

$$\tilde{\eta}_{\mathfrak{q}}=\mathrm{N}(\mathfrak{q})^{-e_{\mathfrak{q}}/2}\delta_{a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}})}\star\mu_{\mathfrak{q}^{e_{\mathfrak{q}}}}-\mathrm{N}(\mathfrak{q})^{\frac{-e_{\mathfrak{q}}-1}{2}}\delta_{a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}-1})}\star\mu_{\mathfrak{q}^{e_{\mathfrak{q}}-1}}. \tag{9.25}$$

(We refer to §2.8 for definitions of $\mu_?$ appearing above.) For $\mathfrak{q}$ such that $s_{\mathfrak{q}}\geq 1$, we set $\sigma_{\mathfrak{q}}$ to be the unique bi-$K_0[\mathfrak{q}^{s_{\mathfrak{q}}}]$-invariant probability measure on

$$K_0[\mathfrak{q}^{s_{\mathfrak{q}}}]a(\varpi^{e_{\mathfrak{q}}})K_0[\mathfrak{q}^{s_{\mathfrak{q}}}],$$

normalized to have mass 1, and we put $\tilde{\eta}_{\mathfrak{q}}=\delta_{a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}})}\star\sigma_{\mathfrak{q}}$. Finally, set $\tilde{\eta}_{\mathfrak{f}}=\prod_{\mathfrak{q}|\mathfrak{f}}\tilde{\eta}_{\mathfrak{q}}$.

One then verifies by a direct computation that

$$f\star\eta_{\mathfrak{f}}=f\star\tilde{\eta}_{\mathfrak{f}}. \tag{9.26}$$

The intuition for this statement, in the classical setting, as follows: let $z\in\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. Then (for a prime number p) the set $\{z+i/p\}_{0\leq i\leq p-1}$ is the p-Hecke orbit of pz, with the point p^2z removed. In the case $e_{\mathfrak{q}}=1$, the first term on the right-hand side of (9.25) corresponds to the p-Hecke orbit of pz, and the second term corresponds to removing the point p^2z.

More formally, to verify (9.26), the unramified computation, at those places where $s_{\mathfrak{q}} = 0$, is easy; the ramified computation is just Hecke theory at ramified primes, see e.g., [40, Prop. 3.33].[21]

The projection $\mathcal{P}$ of Section 2.7 commutes with the action of $\mathrm{GL}_2(\mathbb{A}_F)$, and so (9.26) holds also with f replaced by $\mathcal{P}f$ or $f - \mathcal{P}f$. Moreover, $\mathcal{P}f \star \eta_{\mathfrak{f}} = \mathcal{P}f$. It follows that

$$(9.27)\qquad f \star \eta_{\mathfrak{f}}(x) = \int_{h \in \mathrm{SL}_2(F)\backslash \mathrm{SL}_2(\mathbb{A}_F)} f(hx)dh + (f - \mathcal{P}f) \star \tilde{\eta}_{\mathfrak{f}}(x).$$

Set $\bar{f} = f - \mathcal{P}f$. Then, expanding the term $\bar{f} \star \tilde{\eta}_{\mathfrak{f}}$:

$$(9.28)\quad \bar{f} \star \tilde{\eta}_{\mathfrak{f}} = \sum_{S \subset \{\mathfrak{q}|\mathfrak{f}, s_{\mathfrak{q}}=0\}} \bar{f} \star \prod_{\mathfrak{q}|\mathfrak{f}: s_{\mathfrak{q}} \geq 1} \delta_{a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}})} \star \sigma_{\mathfrak{q}} \star$$
$$\prod_{\substack{\mathfrak{q}|\mathfrak{f}:\\ s_{\mathfrak{q}}=0, \mathfrak{q}\notin S}} \left(\mathrm{N}(\mathfrak{q})^{-e_{\mathfrak{q}}/2} \delta_{a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}})} \star \mu_{\mathfrak{q}^{e_{\mathfrak{q}}}}\right) \star \prod_{\mathfrak{q} \in S} \left(-\mathrm{N}(\mathfrak{q})^{-\frac{e_{\mathfrak{q}}+1}{2}} \delta_{a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}-1})} \star \mu_{\mathfrak{q}^{e_{\mathfrak{q}}-1}}\right).$$

We now specialize to the case under consideration where $x = a(y)$ for some $y \in \mathbb{A}_F^{\times}$. For $S \subset \{\mathfrak{q}|\mathfrak{f}, s_{\mathfrak{q}} = 0\}$ set

$$\sigma_S = \prod_{\mathfrak{q}|\mathfrak{f}: s_{\mathfrak{q}} \geq 1} \sigma_{\mathfrak{q}} \prod_{\mathfrak{q}|\mathfrak{f}: s_{\mathfrak{q}}=0, \mathfrak{q}\notin S} \mathrm{N}(\mathfrak{q})^{-e_{\mathfrak{q}}/2} \mu_{\mathfrak{q}^{e_{\mathfrak{q}}}} \star \prod_{s_{\mathfrak{q}}=0, \mathfrak{q}\in S} -\mathrm{N}(\mathfrak{q})^{-\frac{e_{\mathfrak{q}}+1}{2}} \mu_{\mathfrak{q}^{e_{\mathfrak{q}}-1}}.$$

With this notation, we have

$$(9.29)\qquad \bar{f} \star \tilde{\eta}_{\mathfrak{f}}(a(y)) = \sum_{S \subset \{\mathfrak{q}|\mathfrak{f}, s_{\mathfrak{q}}=0\}} \bar{f} \star \sigma_S \left(a(y[\mathfrak{f}]) \prod_{s_{\mathfrak{q}}=0, \mathfrak{q}\in S} a([\mathfrak{q}])\right).$$

Now apply Lemma 9.2 to see that, for any $z \in \mathbb{A}_F^{\times}$ and $d \gg 1$, we have

$$(9.30)\qquad \bar{f} \star \sigma_S(a(z)) \ll \max(|z|, |z|^{-1})^{1/2} PS_{2,d}(\bar{f} \star \sigma_S),$$

where we have used the easily verified fact that $\mathrm{ht}(a(z)) \asymp \max(|z|, |z|^{-1})$.

Now, for any $f \in C^{\infty}(\mathbf{X})$, we have

$$[K_{\max} : K_{\bar{f}\star\sigma_S}] \leq \prod_{s_{\mathfrak{q}} \geq 1} [K_{\mathfrak{q}} : K_0[\mathfrak{q}^{s_{\mathfrak{q}}}]][K_{\max} : K_f] \ll_{\epsilon} \mathrm{N}(\mathfrak{m})^{1+\epsilon}[K_{\max} : K_f].$$

[21] For the unramified assertion, let $\mathcal{B} = \mathrm{PGL}_2(F_{\mathfrak{q}})/K_{\mathfrak{q}}$ and let $x_0 \in \mathcal{B}$ be the identity coset. The set $\mathcal{B}$ has the structure of the vertices of a $q_v + 1$-valent tree. Let S_1 be the set of all vertices at distance $e_{\mathfrak{q}} - 2i$ (some $i \geq 0$) from $a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}})x_0$. Let S_2 be the set of all vertices at even distance $\leq e_{\mathfrak{q}} - 1 - 2i$ (some $i \geq 0$) from $a(\varpi_{\mathfrak{q}}^{-e_{\mathfrak{q}}-1})x_0$. Then $S_2 \subset S_1$ and $S_1 - S_2$ is precisely the $n(\mathfrak{q}^{-e_{\mathfrak{q}}})$-orbit of x_0. As for the ramified case: one notes that, if $s_{\mathfrak{q}} > 0$, the Haar measure on $K_0[\mathfrak{q}^{s_{\mathfrak{q}}}]$ is just the pushforward of the Haar measure on $n(\mathfrak{o}_{\mathfrak{q}}) \times a(\mathfrak{o}_{\mathfrak{q}}^{\times}) \times \bar{n}(\mathfrak{q}^{s_{\mathfrak{q}}})$ by the product map $(n, a, \bar{n}) \mapsto na\bar{n}$.

By the bounds on matrix coefficients (9.3), and recalling that $\mathfrak{m} = \prod_{\mathfrak{q}|\mathfrak{f}} \mathfrak{q}^{s_\mathfrak{q}}$, we compute that

$$PS_{2,d}(\bar{f} \star \sigma_S) \ll_\epsilon \mathrm{N}(\mathfrak{m})^{3/2+\epsilon} (\mathrm{N}(\mathfrak{f}) \prod_{\mathfrak{q}\in S} \mathrm{N}(\mathfrak{q})^{-1})^{\alpha-1/2+\epsilon} \prod_{\mathfrak{q}\in S} \mathrm{N}(\mathfrak{q})^{-1} PS_{2,d}(f).$$

Combining this with (9.29) and (9.30), we find that for each $S \subset \{\mathfrak{q} : s_\mathfrak{q} \geq 1\}$,

$$\begin{aligned} (9.31) \quad & \left|\bar{f} \star \tilde{\eta}_\mathfrak{f}(a(y))\right| \\ & \ll_\epsilon \mathrm{N}(\mathfrak{f})^{\alpha-1/2+\epsilon} \mathrm{N}(\mathfrak{m})^{3/2+\epsilon} \max(\mathrm{N}(\mathfrak{f})|y|, \mathrm{N}(\mathfrak{f})^{-1}|y|^{-1})^{1/2} PS_{2,d}(f). \end{aligned}$$

This bound is valid for all $f \in C^\infty(\mathbf{X})$, not merely those f that are invariant by $\prod_{\mathfrak{q}|\mathfrak{f}} K_0[\mathfrak{q}^{s_\mathfrak{q}}]$. Apply Remark 8.1, (3) to the endomorphism $f \mapsto \bar{f} \star \tilde{\eta}_\mathfrak{f}$; this shows that (9.31) remains valid, for any $f \in C^\infty(\mathbf{X})$, if we replace $PS_{2,d}$ by $S_{2,d}$ on the right-hand side. Now, specialize to the case where $f \in C^\infty(\mathbf{X})$ is actually $\prod_{\mathfrak{q}|\mathfrak{f}} K_0[\mathfrak{q}^{s_\mathfrak{q}}]$-invariant and apply (9.27) to obtain the conclusion of the lemma. □

The following lemma states an adelic version of the fact that the measure on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ defined by $\nu_y := q^{-1} \sum_{0\leq x\leq q-1} \delta_{\frac{x}{q}+iy}$, approximates the uniform measure if $y \asymp q^{-1}$. More precisely we have an inequality that $\left|\nu_y(f) - \int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} f\right|$ is bounded by $\max(qy, \frac{1}{qy})^{1/2} q^{-\delta} S(f)$, where S is an appropriate Sobolev norm and $\delta > 0$.

LEMMA 9.11. *Let $f \in C^\infty(\mathbf{X})$ and let notation be as in Section* 6 (*see esp.* (6.4)). *In particular*, $\mathfrak{f}$ *is an integral ideal of* $\mathfrak{o}_F$, $q = \mathrm{N}(\mathfrak{f})$, $[\mathfrak{f}]$ *is as in* (2.4) *and*

$$\nu_z(f) = \int_{|y|=z, y\in\mathbb{A}_F^\times/F^\times} f(a(y)n([\mathfrak{f}]))d^\times y.$$

Suppose f is invariant by $K_0[\mathfrak{q}^{s_\mathfrak{q}}]$, for each $\mathfrak{q}|\mathfrak{f}$, and put $\mathfrak{m} = \prod_{\mathfrak{q}|\mathfrak{f}} \mathfrak{q}^{s_\mathfrak{q}}$. Then,

$$\begin{aligned} (9.32) \quad & \left|\nu_z(f) - \int_\mathbf{X} f(x)d\mu_\mathbf{X}(x)\right| \\ & \ll_\epsilon \mathrm{N}(\mathfrak{f})^{\alpha-1/2+\epsilon} \mathrm{N}(\mathfrak{m})^{3/2+\epsilon} \max\left(\mathrm{N}(\mathfrak{f})z, \frac{1}{\mathrm{N}(\mathfrak{f})z}\right)^{1/2} S_{2,d}(f). \end{aligned}$$

Proof. For each $\mathfrak{q}|\mathfrak{f}$ and integer $0 \leq e \in \mathbb{Z}$, let $\eta_{\mathfrak{q}^e}$ be the Haar probability measure on the group $N(\mathfrak{q}^{-e}\mathfrak{o}_\mathfrak{q})$. Then, since the assumption implies that f is right invariant by $a_\mathfrak{q}(\mathfrak{o}_\mathfrak{q}^\times)$, for each $\mathfrak{q}$ dividing $\mathfrak{f}$, we see that for any $x \in \mathbf{X}$:

$$\begin{aligned} (9.33) \quad \int_{y\in\mathfrak{o}_{F_\mathfrak{q}}^\times} f(xa(y)n_\mathfrak{q}(\varpi_\mathfrak{q}^{-e}))d^\times y &= \int_{y\in\mathfrak{o}_{F_\mathfrak{q}}^\times} f(xn_\mathfrak{q}(y\varpi_\mathfrak{q}^{-e}))d^\times y \\ &= \frac{f \star (\eta_{\mathfrak{q}^e} - \mathrm{N}(\mathfrak{q})^{-1}\eta_{\mathfrak{q}^{e-1}})(x)}{1 - \mathrm{N}(\mathfrak{q})^{-1}}. \end{aligned}$$

It follows that

$$\nu_z(f) = \prod_{\mathfrak{q}|\mathfrak{f}} (1 - \mathrm{N}(\mathfrak{q})^{-1})^{-1} \cdot \int_{y \in \mathbb{A}_F^\times / F^\times, |y| = z} f \star \prod_{\mathfrak{q}|\mathfrak{f}} (\eta_{\mathfrak{q}_\mathfrak{q}^e} - \mathrm{N}(\mathfrak{q})^{-1} \eta_{\mathfrak{q}^{e_\mathfrak{q}-1}})(a(y)).$$

We conclude by applying the previous lemma. □

10. **Background on Eisenstein series**

This section essentially develops the theory of Eisenstein series on PGL_2 over a number field. This is needed for the Rankin-Selberg method that we reprise in the next section, which in turn is used in the text to relate a period integral with an L-function.

Let Z be a topological space. In this section, we will often speak — in various contexts, often with $Z = \mathbf{X}$ or $\mathrm{GL}_2(\mathbb{A}_F)$ — of a function $F(s,z)$ on $\mathbb{C} \times Z$ being "holomorphic" or "holomorphic in s". For the purposes of this document, this can be assumed to mean that the function is jointly continuous and holomorphic for each z individually.

Note that $s \mapsto \int_Z F(s,z)dz$, if absolutely convergent and uniformly so in s, defines a holomorphic function. Indeed, it suffices to verify that its integral over a closed curve in the s-variable is zero, which follows by Fubini's theorem.

Similarly, we will say that $F(s,z)$ is meromorphic if there exists a holomorphic function $h(s)$ so that $h(s)F(s,z)$ is holomorphic.

10.1. *Construction and basic properties of the Eisenstein series.* We recall the Eisenstein series that we shall have need of and its basic properties, following Jacquet [20, §19]. We will need Eisenstein series only on PGL_2.

10.1.1. *Schwartz functions.* Let Ψ be a Schwartz-Bruhat function on $\mathbb{A}_F^2$, i.e., Ψ is a finite linear combination of functions $\prod_v \Psi_v$, where each Ψ_v is locally constant of compact support, for v finite, Ψ_v is a Schwartz function on F_v^2 for v infinite, and Ψ_v is the characteristic function of $\mathfrak{o}_{F_v}^2$ for almost all v.

If v is a real place, choose $a_v \in F_v$ so that $e_{F_v}(x) = e^{2\pi i a_v x}$, and say a Schwartz function Ψ_v on F_v^2 is *standard* if it is the product of a polynomial and $e^{-\pi |a_v|_v (|x|_v^2 + |y|_v^2)}$. If v is a complex place, choose $a_v \in F_v$ so that $e_{F_v}(x) = e^{2\pi i \mathrm{Tr}_{\mathbb{C}/\mathbb{R}}(a_v x)}$; we say that a Schwartz function Ψ_v is *standard* if it is the product of a polynomial and $e^{-2\pi |a|_v^{1/2}(|x|_v + |y|_v)}$. The significance of this normalization is twofold: a standard function is automatically K_v-finite and also the class of standard functions is self-dual under the Fourier transform corresponding to the character e_{F_v}.

If V is a real vector space, then by a *Schwartz norm* on the space of Schwartz functions on V, we shall mean a norm $\mathcal{S}$ of the form

$$\mathcal{S}(\Psi) = \sup_{\mathcal{D}} \sup_x |(1 + \|x\|)^M \mathcal{D}\Psi(x)|, \tag{10.1}$$

for some finite collection of constant-coefficients differential operators $\mathfrak{D}$ on V and some norm $\|x\|$ on V.

Put, for $g \in \mathrm{GL}_2(\mathbb{A}_F)$,

$$f_\Psi(s,g) = |\det(g)|^s \int_{t \in \mathbb{A}_F^\times} \Psi((0,t)g)|t|^{2s} d^\times t.$$

The integral converges absolutely for $\Re(s) > 1/2$ and extends to a meromorphic function of s with possible poles at most at $s = 0, 1/2$. Moreover, for all s,

$$f_\Psi\left(\begin{pmatrix} a & x \\ 0 & b \end{pmatrix} g\right) = |a/b|^s f_\Psi(g). \tag{10.2}$$

Put $E_\Psi(s,g) = \sum_{\gamma \in B(F)\backslash \mathrm{GL}_2(F)} f_\Psi(s,\gamma g)$. This converges when $\mathrm{Re}(s) > 1$, extends to a meromorphic function of s with a simple pole at $s = 0, 1$ and satisfies the functional equation

$$E_\Psi(s,g) = E_{\widehat{\Psi}}(1-s,g),$$

where $\widehat{\Psi}$ is the Fourier transform

$$\widehat{\Psi}(x_1,y_1) = \int_{\mathbb{A}_F^2} \Psi(x,y) e_F(x_1 y - y_1 x) dx dy. \tag{10.3}$$

This is not the conventional normalization; it is chosen so that the map $\Psi \mapsto \widehat{\Psi}$ is $\mathrm{SL}_2(\mathbb{A}_F)$-equivariant. Moreover, the pole at $s = 1$ is the constant function with value $c_1 \int_{\mathbb{A}_F^2} \Psi(x,y) dx dy$, and the pole at $s = 0$ is the constant function with value $c_2 \Psi(0)$, where c_1, c_2 are constants (depending only on the choice of measure). Finally, for any fixed g the function $s \mapsto s(1-s)E_\Psi(s,g)$ decays rapidly in vertical strips, i.e., $(1+|s|)^N |s(1-s)E_\Psi(s,g)|$ is bounded in any strip $A \le \Re(s) \le B$. The proof of all these properties follows from "Poisson summation" for $F^2 \subset \mathbb{A}_F^2$, and we omit them.

Moreover, the association $\Psi \mapsto E_\Psi$ is twisted-equivariant for the natural $\mathrm{GL}_2(\mathbb{A}_F)$-action on the space of Schwartz functions and on $C^\infty(\mathbf{X})$; i.e.,

$$E_{h.\Psi}(s,g) = |\det(h)|^{-s}(h \cdot E_\Psi(s,g)), \tag{10.4}$$

where $h\cdot$ denotes right translation by h.

We give an example with $F = \mathbb{Q}$ (cf. [19, (3.29)]).

Example 10.1. Suppose $F = \mathbb{Q}$, $\Psi = \prod_v \Psi_v$ where, for each finite v, Ψ_v is the characteristic function of the maximal compact of F_v, and $\Psi_\infty(x,y) = e^{-\pi(x^2+y^2)}$. Then $E_\Psi(g)$ is determined by its restriction to $\mathrm{SL}_2(\mathbb{R})$.

Moreover, $E_\Psi(s,g)$ descends from a function of $g \in \mathrm{SL}_2(\mathbb{R})$ to a function $E^*(s,z)$ on $\mathbb{H} = \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2$, where the identification is $g \mapsto g \cdot i$. In fact,

$$E^*(s,z) = \pi^{-s}\Gamma(s)\zeta(2s) \sum_{[c:d] \in \mathbb{P}^1(\mathbb{Q})} \frac{y^s}{|cz+d|^{2s}}. \tag{10.5}$$

If we put $\xi(s)=\pi^{-s/2}\Gamma(s/2)\zeta(s)$, then $E^*(s,z)$ has the Fourier expansion

$$E^*(s,z)=\xi(2s)y^s+\xi(2-2s)y^{1-s} + 4\sqrt{y}\sum_{n\in\mathbb{N}}K_{s-1/2}(2\pi ny)\cos(2\pi ny)\sum_{ab=n}\left(\frac{a}{b}\right)^{s-1/2}. \tag{10.6}$$

It satisfies the functional equation $E^*(s,z)=E^*(1-s,z)$. Moreover it is a meromorphic function of s with poles precisely at $s=0$ and $s=1$. In both cases the residue is the constant function.

Motivated by this example, the reader may find it helpful to keep in mind the "dictionary": $f_\Psi(s,g)$ corresponds to $\pi^{-s}\Gamma(s)\zeta(2s)y^s=\xi(2s)y^s$, and $E_\Psi(s,g)$ to $E^*(s,z)$ as defined in (10.5).

Remark 10.1. Suppose Ψ is invariant by $K_\infty\times K_{\max}$. Then f_Ψ is a multiple of $g\mapsto \mathrm{ht}(g)^s$, as follows from the uniqueness of spherical functions satisfying (10.2). Thus, for $\Re(s)>1$, $E_\Psi(s,g)=c(s)\sum_{\gamma\in B(F)\backslash \mathrm{GL}_2(F)}\mathrm{ht}(\gamma g)^s$.

We now proceed to establish the "standard" properties of the Eisenstein series for E_Ψ. It is convenient to first recall an explicit bound for archimedean Mellin transforms; the first part is Tate's thesis, and the second will only be needed much later.

LEMMA 10.1. *Let v be archimedean and let Ψ_v be a Schwartz function on F_v. The integral $G(s):=\int_{x\in F_v^\times}\Psi_v(x)|x|^s d^\times x$ extends to a meromorphic function and*

(1) *$\frac{G(s)}{\zeta_{F,v}(s)}$ is holomorphic, where $\zeta_{F,v}(s)$ is the local factor of the Dedekind ζ-function of F at v.*
(2) *For any $N\geq 0$, the function $G_N(s):=\prod_{i=0}^N(s+i)G(s)$ is holomorphic in $\Re(s)\geq -N$, and the absolute value of $(1+|s|)^M G_N(s)$ in any strip $-N\leq \Re(s)\leq A$ is bounded by some Schwartz norm (depending on A,N,M; see (10.1) for the definition) of Ψ_v.*

Proof. The first assertion is Tate's thesis, and we leave the second (if any) to the reader. □

LEMMA 10.2. *The function $s\mapsto s(1/2-s)f_\Psi(s,g)$ extends to a holomorphic function of s. It decays rapidly along vertical lines*

$$|(1+|\Im(s)|)^N s(1/2-s)f_\Psi(s,g)|\ll_{\Psi,N}\mathrm{ht}(g)^{\Re(s)}, \tag{10.7}$$

where the implicit constant is uniform for $\Re(s)$ in a compact set.

Proof. By (10.2) and the Iwasawa decomposition, it will suffice to prove the assertions in the special case $g\in K_\infty\times K_{\max}$. So we write $g=k\in K_\infty\times K_{\max}$ and denote by k_v the component of k in $\mathrm{PGL}_2(F_v)$. Moreover, without loss of generality, we may assume Ψ is a product of Schwartz functions at each place, i.e.,

$\Psi = \prod_v \Psi_v$. Then

$$f_\Psi(s,g) = \prod_{v \text{ infinite}} \int_{F_v^\times} \Psi_v((0,t)k_v)|t|^{2s} d^\times t \prod_{v \text{ finite}} \int_{F_v^\times} \Psi_v((0,t)k_v)|t|^{2s} d^\times t.$$

By Tate's thesis, it follows that the product over finite places is of the form $\zeta_F(2s)h(s)$, where $\zeta_F(\cdot)$ is the (finite part of the) Dedekind ζ-function of the number field F and $h(s)$ is a holomorphic function with at most polynomial growth in vertical strips (indeed, a polynomial in $q^{\pm s}$ for various q). All the assertions of the lemma now follow from Lemma 10.1, and standard facts about the analytic properties of ζ_F.

In fact, if the Ψ_v for v finite are regarded as fixed, then the implicit constant in (10.7) is bounded by an appropriate Schwartz norm, depending on N and the compact set to which $\Re(s)$ is constrained, of $\prod_{v \text{ infinite}} \Psi_v$. This follows from the second assertion of Lemma 10.1. □

LEMMA 10.3. *The constant term* $E_\Psi^N(s,g) := \int_{x\in F\backslash\mathbb{A}_F} E_\Psi(s,n(x)g)dx$ *equals* $f_\Psi(s,g) + f_{\widehat{\Psi}}(1-s,g)$.

Sketch of proof. A double coset decomposition shows that, for $s \gg 1$, $E_\Psi^N(s,g) = f_\Psi(s,g) + \int_{n\in N(\mathbb{A}_F)} f_\Psi(s,wng)dn$, where w is as in (2.3). So it will suffice to show that $\int_{n\in N(\mathbb{A}_F)} f_\Psi(s,wng) = f_{\widehat{\Psi}}(1-s,g)$. The left-hand side may be expressed as

$$\begin{aligned} (10.8)\quad & |\det(g)|^s \int_{t\in\mathbb{A}_F^\times} \int_{x\in\mathbb{A}_F} \Psi((t,tx)g)|t|^{2s} d^\times t\, dx \\ &= |\det(g)|^s \int_{t\in\mathbb{A}_F^\times/F^\times} \int_{x\in\mathbb{A}_F} \sum_{\delta\in F^\times} \Psi(t(\delta,x)g)|t|^{2s} d^\times t\, dx. \end{aligned}$$

For any Schwartz function Ψ on $\mathbb{A}_F^2$, one has

$$\sum_{\alpha\in F} \int_{y\in\mathbb{A}_F} \Psi(\alpha,y) = \sum_{\beta\in F} \widehat{\Psi}(0,\beta).$$

The result follows from routine manipulation and use of Tate's functional equation. □

We set

$$\bar{E}_\Psi(s,g) = E_\Psi(s,g) - f_\Psi(s,g) - f_{\widehat{\Psi}}(1-s,g), \tag{10.9}$$

so $\bar{E}_\Psi$ defines a function on $B(F)\backslash PGL_2(\mathbb{A}_F)$. It is a "truncated" Eisenstein series where we have removed the constant term. Moreover, $\bar{E}_\Psi(s,g)$ is holomorphic in s (this follows, for example, by computing residues at each of the points $s = 0, 1/2, 1$ and seeing they are all zero). By definition, for $g \in \mathrm{GL}_2(\mathbb{A}_F)$ we have an equality

$$E_\Psi(s,g) = \bar{E}_\Psi(s,g) + f_\Psi(s,g) + f_{\widehat{\Psi}}(1-s,g). \tag{10.10}$$

LEMMA 10.4. *Let $T, N > 0$ and let $\Re(s)$ lie in a fixed compact subset of $\mathbb{R}$. Then*

$$(1+|s|)^4 \bar{E}_\Psi(s,g) \ll_{\Psi,N,T} \mathrm{ht}(g)^{-N}, \tag{10.11}$$

for $g \in \mathfrak{S}(T)$. In particular, if $\Omega \subset \mathbf{X}$ is compact, then $s(1-s)E_\Psi(s,g)$ is uniformly bounded in $|\Re(s)| \leq 2$, $g \in \Omega$.

Proof. We first claim that, for $t \in \mathbb{R}$, we have $|(1+t^4)\bar{E}_\Psi(N+1+it, g)| \ll_{\Psi,N} \mathrm{ht}(g)^{-N+\epsilon}$. Indeed, by definition,

$$\bar{E}_\Psi(s,g) = \sum_{\gamma \in B(F)\backslash \mathrm{PGL}_2(F), \gamma \notin B(F)} f_\Psi(s,\gamma g) - f_{\widehat{\Psi}}(1-s,g).$$

In view of Lemma 10.2, it will suffice to show that

$$\sum_{\gamma \in B(F)\backslash \mathrm{PGL}_2(F), \gamma \notin B(F)} \mathrm{ht}(\gamma g)^\sigma \ll_\epsilon \mathrm{ht}(g)^{1-\sigma+\epsilon}, \tag{10.12}$$

which follows from (8.10) and (8.11).

Now (10.11) follows from the functional equation $\bar{E}_\Psi(s,g) = \bar{E}_{\widehat{\Psi}}(1-s,g)$, the maximal modulus principle in the strip $|\Re(s)| \leq N+1$, and the previous lemma.[22]

The second assertion (involving Ω) follows from (10.7) and (10.11). □

We now compute the Fourier coefficients of the Eisenstein series in general. Recall that e_F is a fixed additive character of $\mathbb{A}_F/F$.

LEMMA 10.5. *Set $W_\Psi(s,g) = \int_{x \in F\backslash \mathbb{A}_F} E_\Psi(s, n(x)g) e_F(x) dx$. Then, for $\Re(s) > 1$,*

$$W_\Psi(s, a(y)) = |y|^{1-s} \int_{t \in \mathbb{A}_F^\times, x \in \mathbb{A}_F} \Psi(t, tx) e_F(xy) |t|^{2s} dx d^\times t. \tag{10.13}$$

In particular, if $\Psi = \otimes_v \Psi_v$, then $W_\Psi = \prod_v W_{\Psi_v}$, where for $\Re(s) > 1$,

$$W_{\Psi_v}(s, a(y)) = |y|_v^{1-s} \int_{t \in F_v^\times, x \in F_v} \Psi_v(t, tx) e_F(xy) |t|_v^{2s} dx d^\times t \ \text{ for } y \in F_v.$$

Finally, if $\Psi_v(x,y) = \varphi_1(x)\varphi_2(y)$, ω_v a character of $F_v^\times$, and $\Re(s') + |\Re(s)| \gg 1$,

$$\begin{aligned} &\int_{y \in F_v^\times} W_{\Psi_v}(s, a(y)) |y|^{s'} \omega_v(y) d^\times y \\ &\qquad = \int_{y \in F_v^\times} \varphi_1(y) |y|^{s'+s} \omega_v(y) d^\times y \int_{y \in F_v^\times} \widehat{\varphi_2}(y) |y|^{1+s'-s} \omega_v(y) d^\times y, \end{aligned} \tag{10.14}$$

where $\widehat{\varphi_2}$ is the Fourier transform, defined by $\widehat{\varphi_2}(y) = \int_{F_v} \varphi_2(y) e_{F_v}(yt) dt$.

[22] To apply the maximal modulus principle in this context, one needs some *a priori* decay of $\bar{E}_\Psi$, which follows easily from the corresponding properties of E_Ψ and f_Ψ.

Proof. By the Bruhat decomposition,

$$W_\Psi(s,g) = \int_{F\backslash\mathbb{A}_F} e_F(x) \sum_{\gamma\in B(F)\backslash \mathrm{PGL}_2(F)} f_\Psi(s,\gamma n(x)g) \tag{10.15}$$

$$= \int_{\mathbb{A}_F} f_\Psi(s, wn(x)g) e_F(x)\,dx.$$

Thus

$$W_\Psi(s,g) = |\det(g)|^s \int_{x\in\mathbb{A}_F}\int_{t\in\mathbb{A}_F^\times} \Psi((t,0)n(x)g)d^\times t|t|^{2s}e_F(x) \tag{10.16}$$

$$= |\det(g)|^s \int_{t\in\mathbb{A}_F, x\in\mathbb{A}_F^\times} \Psi((t,tx)g)|t|^{2s}e_F(x)\,dx\,d^\times t.$$

The claimed conclusion follows upon substituting $g = a(y)$, together with some routine computations. □

Remark 10.2. Remark that $W_\Psi(s,g)$ belongs to the Whittaker model of a certain induced representation of $\mathrm{PGL}_2(\mathbb{A}_F)$, namely the representation $\pi(s)$ induced from the character $a(y) \mapsto |y|^{s-1/2}$ of the maximal torus (unitary induction, so $\pi(s)$ is tempered for $\Re(s) = 1/2$). This representation is the tensor product of local representations $\pi_v(s)$, analogously defined; these local representations are irreducible and generic for all s.

Thus (10.13) determines W_Ψ uniquely (the theory of the Kirillov model). Similarly the condition $W_{\Psi_v,s}(1) = 1$ uniquely determines the (spherical) vector $W_{\Psi_v,s}$.

We finally remark that $W_{\Psi_v,s}$, as Ψ_v ranges over all Schwartz-Bruhat functions on F_v^2 if v is nonarchimedean, or over all standard functions if v is archimedean, exhausts the Whittaker model of $\pi(s)$. Indeed, the set of such functions $W_{\Psi_v,s}$ is a subspace of the Whittaker model of $\pi(s)$ that is stable under the action of the Hecke algebra of $\mathrm{PGL}_2(F_v)$; this action is irreducible, whence the result.

We recall that $\mathfrak{d}$ denotes the different (§2.3) and we denote by $\zeta_{F,v}(s)$ or simply $\zeta_v(s)$ the local factor of the Dedekind ζ-function of F at the place v.

COROLLARY 10.1. *Suppose v is nonarchimedean, and Ψ_v the characteristic function of $\mathfrak{o}_{F_v}^2$. Then $W_v(a(y))$ satisfies*

$$\int_{F_v^\times} W_v(a(y))|y|^{s'}d^\times y = q_v^{d_v(1+s'-s)}\zeta_v(s+s')\zeta_v(1-s+s'), \tag{10.17}$$

with $d_v = v(\mathfrak{d})$. Note that this specifies W_v, because it is K_v-invariant.

In particular, for each finite v with $v(\mathfrak{d}) = 0$, the function $W_v(g)$ is the unique spherical Whittaker function on $\mathrm{GL}_2(F_v)$ with Hecke eigenvalue $q_v^s + q_v^{1-s}$, and with $W_v(1) = 1$.

As is evident from (10.6), the Eisenstein series themselves are not bounded. They belong to $L^{2-\varepsilon}$, but not L^2. To avoid some difficulties with growth, we shall use wave-packets of Eisenstein series. We now turn to their analysis.

10.2. *Regularization of Eisenstein series on* PGL_2. Our aim in this section is to show that an appropriate "wave packet" of the Eisenstein series $E_\Psi(g,s)$ constructed in the previous section lies in L^∞.

Note that, in Example 10.1 above $E^*(s,z)$ differs from the usual unitary Eisenstein series by a factor $\xi(2s)$. This factor ensures that $E^*(s,z)$ is holomorphic, but this causes an inconvenience at $s = 1/2$, which will manifest itself in our construction of bounded wave-packets. Recall that this pole can be interpreted rather naturally; see footnote on page 1031.

Let $\kappa > 0$, and let $\mathcal{H}(\kappa)$ be the family of functions holomorphic in an open neighborhood of the strip $-\kappa \le \Re(s) \le 1+\kappa$, with rapid polynomial decay in vertical strips (i.e., $\sup_{t\in\mathbb{R}}(1+|t|)^N |h(\sigma+it)|$ is bounded, for each N, by a continuous function of σ) and satisfying $h(0) = h(\frac{1}{2}) = h(1) = 0$. For each $N \in \mathbb{Z}$ we have a norm $\|\cdot\|_N$ on $\mathcal{H}(\kappa)$ defined via

$$\|h\|_N = \int_{-\infty}^{\infty} \left(|h(1+\kappa+it)| + |h(-\kappa+it)|\right)(1+|t|)^N dt. \tag{10.18}$$

LEMMA 10.6. *Let* $h \in \mathcal{H}(\kappa)$, *and set* $E_{h,\Psi}(g) = \int_{\Re(s)=1+\kappa} h(s)E_\Psi(g,s)ds$. *Then*:

$$\|E_{h,\Psi}(g)\|_{L^\infty} \ll_{\Psi,\kappa,F} \|h\|_0.$$

Proof. In the notation of (10.10),

$$\begin{aligned} E_{h,\Psi}(g) = &\int_{\Re(s)=1+\kappa} \bar{E}_\Psi(s,g)h(s)ds \\ &+ \int_{\Re(s)=1+\kappa} h(s)f_\Psi(s,g)ds + \int_{\Re(s)=1+\kappa} h(s)f_{\widehat{\Psi}}(1-s,g)ds. \end{aligned} \tag{10.19}$$

Fix $T > 0$ so that $\mathfrak{S}(T)$ surjects onto $\mathbf{X}$ (see Section 8.2 for definitions). We will bound each term on the right-hand side of the above equation for $g \in \mathfrak{S}(T)$.

By Lemma 10.4, the first term on the right-hand side is $O_{\Psi,\kappa}(\|h\|_0)$. By Lemma 10.2, the function $f_{\widehat{\Psi}}(1-s,g)$ is uniformly bounded above in the region $\Re(s) = 1+\kappa, g \in \mathfrak{S}(T)$; thus the third term on the right-hand side is also $O_{\Psi,\kappa}(\|h\|_0)$.

As for the second term, we shift contours to the line $\Re(s) = -\kappa$. The shift of contours is justified by the rapid decay of $h(s)$ along vertical lines and Lemma 10.2. Moreover, since $h(0) = h(1/2) = 0$, the function $s \mapsto h(s)f_\Psi(s,g)$ has no poles in between the contours.

Applying Lemma 10.2 one more time to control the contour integral along $\Re(s) = -\kappa$, we conclude. □

Remark 10.3. Suppose $\Psi = \prod_v \Psi_v$, and the Ψ_v are regarded as fixed for v finite. Put $\Psi_f = \prod_{v \text{ finite}} \Psi_v$, a Schwartz function on $\mathbb{A}_{F,f}^2$, and $\Psi_\infty = \prod_{v \text{ infinite}} \Psi_v$.

Then the above argument gives the slightly more explicit bound

$$\|E_{h,\Psi}\|_{L^\infty} \ll_{\kappa,\Psi_f} \|h\|_0 \mathcal{S}(\Psi_\infty), \tag{10.20}$$

where $\mathcal{S}$ is a Schwartz norm on F_∞^2. This follows by explicating the above argument, taking into account the last sentence of the proof of Lemma 10.2. Indeed, one obtains even the corresponding bound for Sobolev norms, namely

$$S_{\infty,d,\beta}(E_{h,\Psi}) \ll_{\kappa,\Psi_f} \|h\|_0 \mathcal{S}(\Psi_\infty) \tag{10.21}$$

for an appropriate Schwartz norm of F_∞^2. One deduces this from (10.20) upon noting that, if $\mathcal{D}$ belongs to the universal enveloping algebra of $\mathrm{SL}_2(F_\infty)$, then, by (10.4), $\mathcal{D}E_\Psi(s,g) = E_{\mathcal{D}\Psi}(s,g)$, so also $\mathcal{D}E_{h,\Psi} = E_{h,\mathcal{D}\Psi}$. It is then easy to check that a Schwartz norm of $\mathcal{D}\Psi_\infty$ is bounded by a Schwartz norm of Ψ_∞.

10.3. *Regularization of Eisenstein series on* $\mathrm{PGL}_2 \times \mathrm{PGL}_2$. In this section we carry out the analogue of Lemma 10.6 in the context of $\mathrm{PGL}_2 \times \mathrm{PGL}_2$ (this amounts to regularizing the rank 2 Eisenstein series on $\mathrm{PGL}_2 \times \mathrm{PGL}_2$).

To ease the reader's path, we briefly mention what the point of this section is in classical notation: Suppose $h(s_1,s_2)$ is holomorphic in two variables inside the square $|\Re(s_1)| + |\Re(s_2)| \le 1/2 + \kappa$, and, moreover, $h(s_1,s_2)$ has zeroes along the six planes defined by any of the linear constraints $s_1 = 0$, $s_1 = 1/2, s_1 = -1/2$, $s_2 = 0$, $s_2 = -1/2$, $s_2 = 1/2$.

Define the wave-packet $E_h(z_1,z_2)$ on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H} \times \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ via

$$E_h(z_1,z_2) = \int_{t,t'\in\mathbb{R}} h(it_1,it_2)E^*(1/2+it,z_1)E^*(1/2+it',z_2)\,dt\,dt'.$$

Here E^* is as in Example 10.1. We shall show — under mild decay conditions on h — that $E_h(z_1,z_2)$ is majorized, on the product of two fundamental regions, by $A(y_1,y_2) := \frac{\sqrt{y_1y_2}}{y_1^{1/2+\kappa}+y_2^{1/2+\kappa}}$. Since $\int_{y_1\ge 1, y_2\ge 1} A(y_1,y_2)^4 \frac{dy_1dy_2}{y_1^2y_2^2}$ is finite, E_h lies in L^4, and even in $L^{4+\epsilon}$ for ϵ small.

As the reader may verify at this point, the majorization is little more than an exercise in complex integration, using the fact that the large contribution to the Eisenstein series comes from the constant term.

We will need to repeatedly shift contours in the setting of a function of two complex variables. To clarify matters, we state the following lemma, which we will use repeatedly *without explicitly invoking it.*

LEMMA 10.7. *Suppose $U \subset \mathbb{R}^2$ is an open domain and $f(z_1,z_2)$ a holomorphic function on the complex domain $\{(z_1,z_2) \in \mathbb{C}^2 : (\Re(z_1),\Re(z_2)) \in U\}$. Suppose, moreover, that there is a continuous function $M : U \to \mathbb{R}$ such that*

$$\sup_{(t_1,t_2)\in\mathbb{R}^2} |f(\sigma_1+it_1,\sigma_2+it_2)|(1+|t_1|+|t_2|)^3 \le M(\sigma_1,\sigma_2). \tag{10.22}$$

Then the function

$$(\sigma_1,\sigma_2)\mapsto \int_{\Re(z_1)=\sigma_1,\Re(z_2)=\sigma_2} f(z_1,z_2)dz_1dz_2 \tag{10.23}$$

is locally constant on U.

We omit the easy proof.

We will now introduce a family of normed spaces $\mathcal{H}^{(2)}(\kappa)$. In fact, the spaces themselves are independent of κ, but the norm depends on κ. These are spaces of holomorphic functions in two variables z_1, z_2; and the norm, roughly speaking, controls the behavior of h when the real parts of (z_1,z_2) lie in the square $|\Re(z_1)|+|\Re(z_2)|\le 1/2+\kappa$.

Definition 10.1. Let $0<\kappa<1$. Let $\mathcal{H}^{(2)}(\kappa)$ be the family of functions $h(z_1,z_2)$ in two complex variables, holomorphic in a neighborhood of $(0,0)$, and satisfying:

(1) Write $h'=\frac{h(z_1,z_2)}{z_1z_2(1/4-z_1^2)(1/4-z_2^2)}$. Then h', originally a meromorphic function in a neighborhood of 0, extends to a holomorphic function in the strip

$$\{z_1:|\Re(z_1)|\le 2\}\times\{z_2:|\Re(z_2)|\le 2\}.$$

(2) Growth condition: for every $N\ge 0$,

$$\sup_{(\sigma,\sigma')\in[-2,2]^2}\sup_{(t,t')\in\mathbb{R}^2}(1+|t|+|t'|)^N h(\sigma+it,\sigma'+it)<\infty$$

For each $N\in\mathbb{Z}$ we introduce a norm on $H^{(2)}(\kappa)$ via

$$\|h\|_N=\int_{(t,t')\in\mathbb{R}^2}\sum_{\epsilon_1,\epsilon_2\in\{\pm1\}}(1+|t|+|t'|)^N \cdot\left(|h'(\epsilon_1(1/2+\kappa)+it,it')|+|h'(it,\epsilon_2(1/2+\kappa)+it')|\right)dt\,dt'. \tag{10.24}$$

LEMMA 10.8. *For $h\in H^{(2)}(\kappa)$, put*

$$E_{h,\Psi,\Psi'}(g_1,g_2)=\int_{\Re(t)=0}\int_{\Re(t')=0}h(t,t')E_\Psi(g_1,1/2+t)E_{\Psi'}(g_2,1/2+t')\,dt\,dt'.$$

Then

$$E_{h,\Psi,\Psi'}(x_1,x_2)\ll_{\Psi,\Psi'}\|h\|_0\frac{\mathrm{ht}(x_1)^{1/2}\mathrm{ht}(x_2)^{1/2}}{\mathrm{ht}(x_1)^{1/2+\kappa}+\mathrm{ht}(x_2)^{1/2+\kappa}}.$$

Proof. We may assume that $\|h\|_0=1$. Let notation be as established prior to Lemma 10.4. We will proceed as in Lemma 10.6, expanding E_Ψ via (10.10).

It will suffice to give an upper bound, in absolute value, for each of

$$I_0(g_1,g_2)=\int_{t,t'}h(t,t')\bar{E}_{\Psi_1}(g_1,1/2+t)\bar{E}_{\Psi_2}(g_2,1/2+t')dtdt', \tag{10.25}$$

$$I_1(g_1,g_2)=\int_{t,t'}h(t,t')\bar{E}_{\Psi_1}(g_1,1/2+t)f_{\Psi_2}(g_2,1/2\pm t')dtdt', \tag{10.26}$$

$$I_2(g_1, g_2) = \int_{t,t'} h(t,t') f_{\Psi_1}(g_1, 1/2 \pm t) \bar{E}_{\Psi_2}(g_2, 1/2 + t') dt dt', \tag{10.27}$$

$$I_3(g_1, g_2) = \int_{t,t'} h(t,t') f_{\Psi_1}(g_1, 1/2 \pm t) f_{\Psi_2}(g_2, 1/2 \pm t') dt dt', \tag{10.28}$$

whenever Ψ_1, Ψ_2 are Schwartz functions on $\mathbb{A}_F^2$, and in each case the contour of integration is the surface $\Re(t) = \Re(t') = 0$. Moreover, in view of condition (1) in Definition 10.1, each integrand extends to a holomorphic function of (t, t') in the region $|\Re(t)| \leq 1/2 + \kappa, |\Re(t')| \leq 1/2 + \kappa$.

The bound $|I_0| \ll \mathrm{ht}(g_1)^{-N} \mathrm{ht}(g_2)^{-N}$ follows from Lemma 10.4, whereas the bounds $|I_1| \ll \mathrm{ht}(g_1)^{-N} \mathrm{ht}(g_2)^{-\kappa}$ and $|I_2| \ll \mathrm{ht}(g_2)^{-N} \mathrm{ht}(g_1)^{-\kappa}$ follow from moving the t' (in the case of I_1) integral to the contour $\Re(t') = \pm(1/2 + \kappa)$, applying Lemmas 10.4 and 10.2.

We now turn to I_3. We shall consider the case where both signs are $+$, the other cases being similar with appropriate interchanges of sign. Thus set

$$\begin{aligned} Z(t,t') &= h(t,t') f_{\Psi_1}(g_1, 1/2 + t) f_{\Psi_2}(g_2, 1/2 + t') \\ &= h'(t,t') t (1/4 - t^2) f_{\Psi_1}(g_1, 1/2 + t) t' (1/4 - t'^2) f_{\Psi_2}(g_2, 1/2 + t'). \end{aligned} \tag{10.29}$$

In view of Lemma 10.2, the function $Z(t,t')$ satisfies the conditions for f in Lemma 10.7. We apply Lemma 10.7 to shift the contour to $\Re(t) = -1/2 - \kappa$, $\Re(t') = 0$. Now Lemma 10.2 implies that $|\int_{\Re(t) = -1/2-\kappa, \Re(t')=0} Z(t,t')| \ll \mathrm{ht}(g_2)^{1/2} \mathrm{ht}(g_1)^{-\kappa}$. A similar bound holds with (g_1, g_2) interchanged, so in fact we have the stronger bound $|Z(t,t')| \ll \min(\mathrm{ht}(g_2)^{1/2} \mathrm{ht}(g_1)^{-\kappa}, \mathrm{ht}(g_1)^{1/2} \mathrm{ht}(g_2)^{-\kappa})$. This may also be written $|Z(t,t')| \ll \frac{\mathrm{ht}(g_1)^{1/2} \mathrm{ht}(g_2)^{1/2}}{\mathrm{ht}(g_1)^{1/2+\kappa} + \mathrm{ht}(g_2)^{1/2+\kappa}}$.

Similar considerations apply to the terms in I_3 corresponding to other choices of sign, so we conclude that $|I_3| \ll \frac{\mathrm{ht}(g_1)^{1/2} \mathrm{ht}(g_2)^{1/2}}{\mathrm{ht}(g_1)^{1/2+\kappa} + \mathrm{ht}(g_2)^{1/2+\kappa}}$. ☐

LEMMA 10.9. *Let notation be as in the previous lemma. For any $p < \frac{4}{1-2\kappa}$, any $d, \beta > 0$, there exists N such that*

$$S_{p,d,\beta}(E_{h,\Psi}) \ll_{\Psi,\kappa,p,\beta} \|h\|_N.$$

Proof. Indeed, we note that

$$\int_{y_1, y_2 \geq 1} \left(\frac{\sqrt{y_1 y_2}}{y_1^{1/2+\kappa} + y_2^{1/2+\kappa}} \right)^p \frac{dy_1 dy_2}{y_1^2 y_2^2} < \infty$$

whenever $p < 4/1 - 2\kappa$. We apply the previous lemma and reduction theory to conclude. ☐

11. Background on integral representations of L-functions

The purpose of this section is as follows. The geometric method we have explained in the text yields upper bounds for certain periods; to obtain subconvexity, we need to know that L-functions can be expressed in terms of these periods. This is the whole point of the theory of integral representations of L-functions; however, we cannot quite simply quote from that theory, as we often need, e.g., some analytic control on the choice of test vector for which there is no readily available reference.

On occasion we have only sketched proofs in this section, as they amount to simple explications of standard techniques such as the Rankin-Selberg method, and moreover they are in some sense irrelevant to the main point of this paper (which is to bound *periods*, not L-functions!).

11.1. *Cuspidal triple product L-functions.*

HYPOTHESIS 11.1. *Let π_2 and π_3 be* fixed *automorphic cuspidal representations of* PGL_2 *over F. Let π_1 be an automorphic cuspidal representation, whose finite conductor is a prime ideal, prime to the finite conductors of π_2 and π_3. Suppose that $\pi_{1,\infty}$ (the representation of* $\mathrm{PGL}_2(F_\infty)$ *underlying π_1) is restricted to a bounded set; let φ_1 be the new vector in π_1.*

Then there exists finite collections of vectors $\mathcal{F}_2 \subset \pi_2, \mathcal{F}_3 \subset \pi_3$ so that, for any such π_1, there exist $\varphi_j \in \mathcal{F}_j$ $(j = 2, 3)$ with

$$\frac{L(\frac{1}{2}, \pi_1 \otimes \pi_2 \otimes \pi_3)}{\left|\int_{\mathbf{X}} \varphi_2(xa([\mathfrak{p}]))\varphi_3(x)\varphi_1(x)dx\right|^2} \ll_{\epsilon,F,\pi_{1,\infty}} N(\mathfrak{p})^{1+\epsilon}. \tag{11.1}$$

Note that no claim is made about the dependence of the constants in (11.1) on π_2, π_3 or the bounded set containing $\pi_{1,\infty}$; presumably with enough effort one could obtain polynomial dependence on the conductors.

Remark 11.1. In the time since this paper was submitted, the above Hypothesis has been apparently established by M. Woodbury, and will appear in his PhD thesis [46]; it uses in particular the work of A. Ichino. This renders the application to the triple product L-functional unconditional.

11.2. *Rankin-Selberg convolutions.*

11.2.1. *The Rankin-Selberg integral representation.* Let π_1, π_2 be two automorphic representations, with π_2 cuspidal.

Let Ψ_v be a Schwartz-Bruhat function on F_v^2 such that, for almost all v, Ψ_v is the characteristic function of $\mathfrak{o}_{F_v}^2$. Put $\Psi = \prod_v \Psi_v$, a Schwartz function on $\mathbb{A}_F^2$. Let φ_j belong to the space of π_j for $j = 1, 2$ and put

$$I(\varphi_1, \varphi_2, \Psi, s) = \int_{\mathbf{X}} \varphi_1(g)\varphi_2(g)E_\Psi(s, g)dg. \tag{11.2}$$

Unwinding, we see that for $\Re(s) > 1$,

$$\begin{aligned} I(\varphi_1, \varphi_2, \Psi, s) &= c_F \int_{B(F)\backslash \mathrm{PGL}_2(\mathbb{A}_F)} f_\Psi(s, g)\varphi_1(g)\varphi_2(g)dg \\ &= c_F \int_{B(F)\backslash \mathrm{PGL}_2(\mathbb{A}_F)} f_\Psi(s, g) \left(\int_{n\in N(F)\backslash N(\mathbb{A}_F)} \varphi_1(ng)\varphi_2(ng)dn\right) dg. \end{aligned} \tag{11.3}$$

Here the constant c_F arises from change of measure: the measure on $\mathbf{X}$ is the $\mathrm{PGL}_2(\mathbb{A}_F)$-invariant probability measure, which is not the same as the quotient measure from $\mathrm{PGL}_2(\mathbb{A}_F)$. Note that c_F will be unimportant in our arguments, as it depends only on F and we are only interested in bounds.

Put $W_1(g) = \int_{F\backslash\mathbb{A}_F} \varphi_1(n(x)g)e_F(x)dx$, and define W_2 similarly but with e_F replaced by $\overline{e_F}$. Recall that our normalizations are so that the volume of $\mathbb{A}_F/F$ is 1. Fourier inversion shows that $\varphi_i(g) = \sum_{\alpha\in F^\times} W_i(a(\alpha)g)$ if φ_i is cuspidal. Thus, as long as one of φ_1, φ_2 is cuspidal, we see that

(11.4)

$$\begin{aligned} I(\varphi_1,\varphi_2,\Psi,s) &= c_F \int_{B(F)\backslash\mathrm{PGL}_2(\mathbb{A}_F)} f_\Psi(s,g)\left(\sum_{\alpha\in F^\times} W_1(a(\alpha)g)W_2(a(\alpha)g)\right)dg \\ &= c_F \int_{N(\mathbb{A}_F)\backslash\mathrm{PGL}_2(\mathbb{A}_F)} W_1(g)W_2(g)f_\Psi(s,g)dg. \end{aligned}$$

If φ_1, φ_2 are pure tensors, then there is a corresponding product decomposition $W_1 = \prod_v W_{1,v}$, $W_2 = \prod_v W_{2,v}$, where $W_{j,v}$ belongs to the local Whittaker model of $\pi_{j,v}$, a representation of $\mathrm{PGL}_2(F_v)$. In that case,

$$I(\varphi_1,\varphi_2,\Psi,s) = c_F \prod_v I_v(W_{1,v}, W_{2,v}, \Psi_v, s), \tag{11.5}$$

where

$$\begin{aligned} &I_v(W_{1,v}, W_{2,v}, \Psi_v, s) \\ &\quad= \int_{N(F_v)\backslash\mathrm{PGL}_2(F_v)} dg_v\ W_1(g_v)W_2(g_v)\left(|\det(g_v)|_v^s \int_{t\in F_v^\times} \Psi_v((0,t)g_v)|t|^{2s}d^\times t\right). \end{aligned}\tag{11.6}$$

We note that the bracketed quantity, defined *a priori* for $g_v \in \mathrm{GL}_2(F_v)$, descends to $\mathrm{PGL}_2(F_v)$.

Applying the Iwasawa decomposition to (11.6) yields the equivalent

$$\begin{aligned} &I_v(W_{1,v}, W_{2,v}, \Psi_v, s) \\ &\quad= \int_{y\in F_v^\times, k\in K_v} W_1(a(y)k)W_2(a(y)k)|y|^{s-1}d^\times y\, dk \left(\int_{t\in F_v^\times} \Psi((0,t)k)|t|^{2s}d^\times t\right). \end{aligned}\tag{11.7}$$

LEMMA 11.1. *Let v be nonarchimedean. Suppose $W_{1,v}, W_{2,v}$ are the new vectors associated to spherical representations $\pi_{1,v}, \pi_{2,v}$, Ψ_v is the characteristic function of $\mathfrak{o}_{F_v}^2$, and e_{F_v} is unramified. Then*

$$I_v(W_{1,v}, W_{2,v}, \Psi_v, s) = L_v(s, \pi_{1,v}\otimes\pi_{2,v}).$$

If $W_{1,v}, W_{2,v}$ are nonzero and $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$-invariant, Ψ_v as above, but e_{F_v} is possibly ramified, then $I_v(W_{1,v}, W_{2,v}, \Psi_v, s) = aq_v^{ks}L_v(s,\pi_{1,v}\times\pi_{2,v})$ where $k\in\mathbb{Z}$ is so that e_{F_v} is trivial on $\varpi_v^{-k}\mathfrak{o}_{F_v}$ but not on $\varpi_v^{-k-1}\mathfrak{o}_{F_v}$. Moreover, $a = 1$ if $W_{1,v}(\varpi_v^{-k}) = W_{2,v}(\varpi_v^{-k}) = 1$.

Suppose $W_{1,v}, W_{2,v}$ are the new vectors associated to $\pi_{1,v}$ a spherical and $\pi_{2,v}$ a Steinberg representation, and that e_{F_v} is unramified. Then, with Ψ_v the characteristic function of $\mathfrak{o}_{F_v}^2$, we have

$$I_v\left(\pi_{1,v}\begin{pmatrix}1 & 0\\ 0 & \varpi_v\end{pmatrix}W_{1,v}, W_{2,v}, \Psi_v, s\right) = \pm\frac{q_v^s}{q_v+1}L(s, \pi_{1,v}\otimes\pi_{2,v}).$$

Proof. See [20, Th. 15.9] for the first assertion. The second assertion is an easy consequence. See Section 11.3 for the final assertion. □

11.2.2. *Topologizing the space of local representations.* The results in [20] provide "good" test vectors for the functionals I_v when the local representations $\pi_{1,v}, \pi_{2,v}$ are fixed. On the other hand, we will need such results with some mild uniformity in $\pi_{1,v}$. One can certainly extract the stronger results from the proofs in [20]. For now, we will proceed by deducing the results "by continuity"; to do this, we will need to define the topology on the space of possible $\pi_{1,v}$. The considerations that follow are not really very crucial; it would be better simply to explicate the implicit dependences in [20].

Let $\mathcal{S}$ be a finite set of irreducible (continuous) representations of $K_v = \mathrm{PGL}_2(\mathfrak{o}_{F_v})$. For any representation W of K_v, we denote by $W^{\mathcal{S}}$ that subspace of W consisting of vectors whose K_v-span contains only irreducibles that belong to $\mathcal{S}$. We shall say that elements of $W^{\mathcal{S}}$ are *of type* $\mathcal{S}$.

Let $\mathcal{G}_v$ be the set of isomorphism classes of generic irreducible representations of $\mathrm{PGL}_2(F_v)$. If π is a generic irreducible representation which is a discrete series or supercuspidal, we shall define it to be isolated. Otherwise, π is induced from two quasicharacters $\mu, \nu : F_v^\times \to \mathbb{C}$. For $s \in \mathbb{C}$, let $(\pi(s), V(s))$ be the representation induced from the quasicharacters $\mu|\cdot|_v^s, \nu|\cdot|_v^{-s}$. Then $\pi(s)$ is generic for all $s \in \mathbb{C}$ and irreducible in a neighborhood of 0. We shall topologize $\mathcal{G}_v$ in such a way that sets of the form $[\pi(s)]$, for $|s| < \varepsilon$ form a basis.

If $E \subset \mathcal{G}_v$ is a closed subset that is bounded (when considered as a subset of the set of isomorphism classes of irreducible admissible representations, and bounded in the sense of §2.12.3), then E is compact, as one checks by direct verification.

For each $\pi \in \mathcal{G}$, we have a Whittaker model $\mathcal{W}(\pi)$. Consider a function $\pi \mapsto W_\pi$, that assigns to each $\pi \in \mathcal{G}_v$ an element W_π of its Whittaker model. We shall say that such an assignment $\pi \mapsto W_\pi$ is continuous if there exists a neighborhood of each π, which we may assume to be of the form, $\{\pi(s) : |s| < \varepsilon\}$, and a set $\mathcal{S}$ of irreducible representations of K_v so that

(1) $W_{\pi(s)}$ is of type $\mathcal{S}$, for each $|s| < \varepsilon$.

(2) The assignment $s \mapsto W_{\pi(s)}(g)$ is continuous for each $g \in \mathrm{PGL}_2(F_v)$, uniformly for g in any fixed compact.

It can be verified that if W_0 is an element of the Whittaker model of π_0, there exists a continuous assignment $\pi \mapsto W_\pi$ in a neighborhood of π_0 which has the value W_0 at π_0.

The requirement (2) is not very strong, as it does not impose any uniformity on all of $\mathrm{PGL}_2(F_v)$. However, in every context we shall consider, the necessary uniformity in g is automatic. Let us sketch how one can prove such results. Assume that v is finite; the infinite case is similar although more technically involved. One first observes that if $\pi \mapsto W_\pi$ is a continuous assignment on some open set, then, for a fixed character χ_v of $F_v^\times$, the quotient $\frac{\int_{y\in F_v^\times} W_\pi(a(y))\chi_v(y)|y|^{s-1/2}d^\times y}{L_v(s,\pi\otimes\chi_v)}$ is a polynomial of the form $\sum_{k=-N}^{N} c_k q_v^{ks}$; moreover, the degree N is locally bounded as π varies, and all the coefficients c_k can be taken to depend continuously on π. To verify the local boundedness of the degree — which requires only property (1) above — one just notes that there is (locally) a fixed M such that $W_\pi(a(y))$ vanishes for $|y|_v > M$; this, together with the functional equation, gives the local boundedness. To see that the coefficients vary continuously, it suffices to check that, for any fixed integer t, the integral $\int_{v(y)=t} W_\pi(a(y))\chi_v(y)|y|^{s-1/2}d^\times y$ varies continuously, which follows from the definition of continuity for the assignment $\pi \mapsto W_\pi$. The archimedean case proceeds similarly, but one replaces the role of polynomials in $q_v^{\pm s}$ by functions of the form $c^s P(s)$, where P is a polynomial and $c \in \mathbb{R}$.

In the next few pages, we will make certain claims regarding the continuity of various integrals involving W_π, if $\pi \mapsto W_\pi$ is a continuous assignment. One can reduce all the claimed continuity statements (by standard "Mellin transform" arguments) to the result just discussed. We will omit the details.

11.2.3. *Choice of test vectors.*

LEMMA 11.2. *Let notation be as above.*

(1) *The quotient*

$$\Xi_v(W_{1,v}, W_{2,v}, \Psi_v, s) := \frac{I_v(W_{1,v}, W_{2,v}, \Psi_v, s)}{L_v(s, \pi_{1,v} \otimes \pi_{2,v})}$$

is holomorphic in s. If v is nonarchimedean, Ξ_v is a polynomial in $q_v^{\pm s}$; if v is archimedean and Ψ_v is standard, then $\Xi_v |a_v|_v^{2s}$ is a polynomial in s. Here a_v is as in Section 10.1.1.

(2) *For any fixed $s_0 \in \mathbb{C}$ we may choose data $(W_{1,v}, W_{2,v}, \Psi_v)$ of the type described in* (1) *with $\Xi_v(W_{1,v}, W_{2,v}, \Psi_v, s_0) \neq 0$.*

(3) *If $\pi_{2,v}$ is regarded as fixed, and $\pi_{1,v}$ remains within a fixed compact subset of $\mathscr{G}_v$ consisting entirely of unitarizable representations, then there exists a constant C depending on the compact set so that one may choose data as in* (2) *in such a way that*:

  (a) *Ψ_v and $W_{2,v}$ may both be chosen from a finite list of size $\leq C$*;

  (b) $\int_{F_v^\times} |W_{1,v}(a(y))|^2 d^\times y \leq C$;

  (c) $|\Xi_v(s_0)| \geq 1$ *and, for all $s \in \mathbb{C}$, we have* $|\Xi_v(s)| \ll C^{|\Re(s)|}(1+|s|)^C$.

Proof. The first two assertions are [20, Ths. 14.8, 17.2]. We will only sketch the last assertion. It can be also be proved directly by *exhibiting* such data by explicating the arguments of [20]. In any case, we start by noting: Given any continuous assignment $\pi_{1,v} \mapsto W_{1,v}, \pi_{2,v} \mapsto W_{2,v}$, the functions $\Xi_v(W_{1,v}, W_{2,v}, \Psi_v, s)$ and $\int |W_{i,v}(a(y)|^2 d^\times y$ all vary continuously in $\pi_{1,v}, \pi_{2,v}$. This assertion can be deduced by the methods explained in Section 11.2.2. Here, when we speak of $\Xi_v(W_{1,v}, W_{2,v}, \Psi_v, s)$ varying continuously, we mean this in the "strong sense", i.e., the statement that Ξ_v can be expressed as a polynomial in q_v^s (nonarchimedean case) or $b^s P(s)$ where P is polynomial (archimedean case), so that all coefficients vary continuously with $\pi_{1,v}, \pi_{2,v}$.

Now, fix momentarily $\pi_{1,v}$ and $\pi_{2,v}$ and suppose that we have chosen data $(W_{1,v}, W_{2,v}, \Psi_v)$ as in (2). Extend $W_{1,v}$ to a continuous assignment $\pi \mapsto W_\pi$ in a neighborhood of $\pi_{1,v}$. By the remarks above, $(W_\pi, W_{2,v}, \Psi_v)$ will satisfy (b) and (c), for a suitable constant C, whenever π belongs to a sufficiently small neighborhood of $\pi_{1,v}$. Now a compactness argument demonstrates (3). $\square$

We emphasize again that (11.6) is valid so long as *one* of π_1, π_2 is cuspidal.

LEMMA 11.3. *Let π be an automorphic cuspidal representation in $L^2(\mathbf{X})$, and let $\varphi \in \pi$ be so that $W_\varphi := \int_{F\backslash\mathbb{A}_F} e_F(x)\varphi(n(x)g)$ factorizes as a product $\prod_v W_v(g_v)$. Then, for a certain absolute constant c,*

$$\int_{\mathbf{X}} |\varphi(g)|^2 dg = c\mathrm{Res}_{s=1}\Lambda(s, \pi\otimes\tilde{\pi}) \prod_v \frac{\int_{F_v^\times} |W_v(a(y))|^2 d^\times y}{L_v(s, \pi_v \otimes \tilde{\pi}_v)}. \tag{11.8}$$

Observe that the product on the right-hand side defines a holomorphic function of s; this follows from the prior lemmas.

Proof. This follows by taking the residue of $I(\varphi, \overline{\varphi}, \Psi, s)$ at $s = 1$. Indeed, this residue equals, up to a constant depending only on the measure normalization, $\left(\int_{\mathbf{X}} |\varphi(g)|^2 dg\right)\left(\int_{\mathbb{A}_F^2} \Psi(x, y) dx dy\right)$ (see discussion of properties of E_Ψ after (10.3)).

On the other hand, by (11.5) and (11.6) $I(\varphi, \overline{\varphi}, \Psi, s)$ may be written as a product $c_F \prod_v I_v(W_v, \overline{W_v}, \Psi_v, s)$, where each I_v is given by (11.7). The integral $\int_{F_v^\times} |W_v(a(y)k)|^2 d^\times y$ is independent of $k \in K_v$, so $I_v(W_v, \overline{W_v}, \Psi_v, 1)$ factors as the product of $\int_{y\in F_v^\times} |W_v(a(y))|^2 d^\times y$ and $\int_{t\in F_v^\times, k\in K_v} \Psi_v((0,t)k)|t|^2 d^\times t$. The latter integral differs from $\int_{F_v^2} \Psi_v(x, y) dx dy$ by a factor that depends only on the normalizations of measure; moreover, this factor equals $(1 - q_v^{-2})^{-1}$ for almost all v, so the product of these factors is convergent. The conclusion easily follows. $\square$

We now specialize to the cases of interest. Fix π_1. We vary $\pi_2 := \pi$ through a sequence of automorphic cuspidal representations with prime conductor $\mathfrak{p}$, prime to the conductor of π_1. In particular, the local constituent of π at $\mathfrak{p}$ is a special representation. We denote by π_∞ the representation of $\mathrm{GL}_2(F_\infty)$ underlying the representation π.

LEMMA 11.4. *Suppose the archimedean constituent π_∞ belongs to a bounded subset of $\widehat{\mathrm{PGL}_2(F_\infty)}$ (in what follows the implicit constants may depend on this subset) and regard π_1 as being fixed.*

Let $s_0 \in \mathbb{C}$. There exists a fixed finite set $\mathcal{F}$ of Schwartz Bruhat functions and a real number[23] $C > 0$ so that, for any such π,

There exist vectors $\varphi_1 \in \pi_1, \varphi \in \pi$ and $\Psi \in \mathcal{F}$ so that

$$\Phi(s) := \mathrm{N}(\mathfrak{p})^{1-s} \frac{I(a([\mathfrak{p}]) \cdot \varphi_1, \varphi, \Psi, s)}{\Lambda(s, \pi_1 \otimes \pi)}$$

is holomorphic and satisfies:

(1) $|\Phi(s_0)| \gg 1$ *and* $|\Phi(s)| \ll C^{|\Re(s)|}(1 + |s|)^C$;

(2) *At any nonarchimedean place v such that π_1 is unramified, φ_1 and Ψ are invariant by $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$; at any nonarchimedean place v such that π_1 and π are both unramified, both φ and φ_1 are invariant by $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$;*

(3) $\|\varphi_1\|_{L^\infty}$ *is* $O(1)$;

(4) $\|\varphi\|_{L^2(\mathbf{X})} \ll_\epsilon \mathrm{N}(\mathfrak{p})^\epsilon$.

Proof. We first choose local data. For each place where $e_{F,v}$ and π_1 are not ramified, we take W_v (resp. $W_{v,1}$) to be the new vector in the Whittaker model of π_v (resp $\pi_{1,v}$). We put Ψ_v to be the characteristic function of $\mathfrak{o}_{F_v}^2$.

Let $\mathcal{B}$ be the set of remaining v. For $v \in \mathcal{B}$, the assumptions show that π_v is restricted to a bounded set. We choose W_v, $W_{v,1}$, Ψ_v for $v \in \mathcal{B}$ according to Lemma 11.2. Finally we choose φ so that $\int_{x \in F \backslash \mathbb{A}_F} e_F(x)\varphi(n(x)g) = \prod_v W_v(g_v)$, and similarly for φ_1, and take $\Psi = \prod_v \Psi_v$. The first two assertions of the lemma are immediate (cf. Lemma 11.1).

To bound the L^2-norm of φ, use Lemma 11.2 (b), Lemma 11.3, and Iwaniec's bounds on L-functions near 1 (see [19, Th. 8.3]). As for φ_1, it in fact belongs to a fixed finite set of cusp forms, so the third assertion is immediate. □

We continue to keep π an automorphic cuspidal representation of $\mathrm{PGL}_2(\mathbb{A}_F)$ with prime conductor.

LEMMA 11.5. *Suppose π_∞ belongs to a bounded subset of $\widehat{\mathrm{PGL}_2(F_\infty)}$ (in what follows the implicit constants may depend on this bounded subset).*

Let $t_0, t_0' \in \mathbb{C}$. There exists a fixed finite set $\mathcal{F}$ of Schwartz Bruhat functions and a real number $C > 0$ so that:

There exist vectors $\varphi \in \pi, \Psi_1, \Psi_2 \in \mathcal{F}$ so that:

$$\Phi(t, t') = \mathrm{N}(\mathfrak{p})^{1/2-t} \frac{\int_{\mathbf{X}} \varphi(g) E_{\Psi_1}(g, \frac{1}{2} + t) E_{\Psi_2}(ga([\mathfrak{p}]), \frac{1}{2} + t') dg}{\Lambda(\frac{1}{2} + t + t', \pi)\Lambda(\frac{1}{2} + t - t', \pi)} \tag{11.9}$$

is holomorphic and satisfies:

[23] depending on π_1 and the choice of bounded subset of $\widehat{\mathrm{PGL}_2(F_\infty)}$.

(1) $|\Phi(t_0, t_0')| \gg 1$ *and* $|\Phi(t, t')| \ll (1 + |t| + |t'|)^C C^{|\Re(t)|+|\Re(t')|}$;

(2) *For any nonarchimedean place* v, *each* Ψ_1 *and* Ψ_2 *is invariant by* $\mathrm{PGL}_2(\mathfrak{o}_{F_v})$, *for each nonarchimedean place* v *at which* π *is unramified, the same is true of* φ;

(3) $\|\varphi\|_{L^2(\mathbf{X})} \ll_\epsilon \mathrm{N}(\mathfrak{p})^\epsilon$.

Proof. The proof is similar to that of the previous lemma; recall that (11.6) was valid as long as one of π_1, π_2 were cuspidal.

Let Ψ_2' be the translate of the Schwartz function Ψ_2 by $a([\mathfrak{p}])$. Then by (10.4),

$$E_{\Psi_2'}(s, g) = \mathrm{N}(\mathfrak{p})^{-s} E_{\Psi_2}^{a([\mathfrak{p}])}(s, g).$$

Suppose Ψ_1, Ψ_2 factorize as $\prod_v \Psi_{1,v}, \prod_v \Psi_{2,v}$, and define $W_{\Psi_{1,v}}(s, g)$ and $W_{\Psi_{2,v}}(s, g)$ as in Lemma 10.5. Suppose moreover that $\int_{x \in F\backslash \mathbb{A}_F} e_F(x)\varphi(n(x)g)$ factorizes as $\prod_v W_v(g)$. Then we can express the global integral of (11.9) as a product in two different ways, depending on whether we let E_{Ψ_1} or E_{Ψ_2} play the role of π_2. Namely, as in (11.7),

$$\begin{aligned}(11.10)\quad \int_{\mathbf{X}} &\varphi(g) E_{\Psi_1}(g, 1/2 + t) E_{\Psi_2}(ga([\mathfrak{p}]), 1/2 + t') dg \\ &= c_F \mathrm{N}(\mathfrak{p})^{1/2+t'} \prod_v I_v(W_v, W_{\Psi_{1,v}}(1/2 + t, \cdot), \Psi_{2,v}', 1/2 + t') \\ &= c_F \prod_v I_v(W_v, W_{\Psi_{2,v}}(1/2 + t', \cdot)^{a([\mathfrak{p}])_v}, \Psi_{1,v}, 1/2 + t).\end{aligned}$$

Here $W_{\Psi_{2,v}}(1/2 + t', \cdot)^{a([\mathfrak{p}])_v}$ denotes the translate of $W_{\Psi_{2,v}}$ by the vth component of $a([\mathfrak{p}])$.

For v nonarchimedean (notation being similar to that of the previous lemma) we take $\Psi_{1,v}$ and $\Psi_{2,v}$ to be the characteristic function of $\mathfrak{o}_v^2$ for every finite v, and W_v to be the new vector.

For v archimedean we first apply Lemma 11.2 with $s_0 = 1/2 + t_0$, and $\pi_{2,v}$ the representation of $\mathrm{PGL}_2(F_v)$ spanned by $E_{\Psi_2}(1/2 + t_0', g)$, i.e., the representation unitarily induced from the character $a(y) \mapsto |y|_v^{it_0'}$. Lemma 11.2 provides W_v in the Whittaker model of π_v, $W_{2,v}$ in the Whittaker model of $\pi_{2,v}$, and a Schwartz function $\Psi_{1,v}$ with $|I_v(W_v, W_{2,v}, \Psi_{1,v}, 1/2 + t_0)| \geq 1$. The last comment of Remark 10.2 shows that there is a standard $\Psi_{2,v}$ so that $W_{\Psi_{2,v}}(1/2 + t_0', g_v) = W_{2,v}(g_v)$ (notation of Lemma 10.5). Moreover, Lemma 11.2 also shows that $\Psi_{1,v}$ and $W_{2,v}$ (so also $\Psi_{2,v}$) may be chosen from a fixed finite set of possibilities (depending, of course, on the original bounded set to which π_∞ belongs, as well as t_0 and t_0').

Again we put $\Psi_i = \prod_v \Psi_{i,v}$ for $i = 1, 2$ and take φ with $\int_{x \in F\backslash\mathbb{A}_F} \varphi(n(x)g) = \prod_v W_v(g)$. From (11.10) we deduce that, with our choices, $|\Phi(t_0, t_0')| \gg 1$. The assertion about $\|\varphi\|_{L^2(\mathbf{X})}$ follows as in the proof of the previous lemma. The second assertion of the Lemma (concerning invariance of Ψ_1, Ψ_2) is immediate.

It remains to prove that Φ is actually holomorphic in (t,t') and that $|\Phi(t,t')| \ll (1+|t|+|t'|)^C e^{C|\Re(t)|+C|\Re(t')|}$. Put $\Xi_v = \frac{I_v}{L_v(\frac{1}{2}+t+t',\pi_v)L_v(\frac{1}{2}+t-t',\pi_v)}$. It is simple to explicitly compute Ξ_v for nonarchimedean v, using Corollary 10.1 and Lemma 11.1. One thereby sees that it will suffice to check, by similar arguments to those used in Lemma 11.2, the following statement for v archimedean: $\Xi_v = c^s c'^{s'} P(s,s')$, where P is a polynomial, and moreover c, c', P vary continuously in π_v, if $\pi_v \mapsto W_v$ is a continuous assignment. We only sketch the proof of this. From (11.7) and the fact that $W_v, \Psi_{1,v}, \Psi_{2,v}$ are all K_v-finite, it suffices to prove the corresponding assertions for $\int_{F_v^\times} W_v(a(y))W_{\Psi_{1,v}}(s,a(y))|y|^{s'-1}d^\times y$. For this we use Barnes' formula as in [20]; see e.g., [20, Lemmas 17.3.1, 17.3.2]. $\square$

11.3. *Local Rankin-Selberg convolutions.* Let v be a nonarchimedean place of F with residue characteristic q_v. Let π_1, π_2 be generic irreducible admissible representations of $\mathrm{GL}(2, F_v)$ with trivial central character. (Since we shall work purely locally over F_v throughout the present subsection, we shall use the notation π_1 rather than, e.g., $\pi_{1,v}$.)

Then π_1, π_2 are self-dual. We assume that π_2 is unramified and π_1 has conductor q_v, and denote by $L(s,\pi_j)$ the local L-factors.

Fix once and for all an additive unramified character ψ of F_v. Let $v : F_v^\times \to \mathbb{Z}$ be the valuation, put $\mathfrak{o}_{F_v} = \{x \in F_v^\times : v(x) \geq 0\}$, and choose a uniformizer $\varpi \in F_v^\times$. Let $\mathfrak{o}_{F_v}^\times$ be the multiplicative group of units in $\mathfrak{o}_{F_v}$. Let $d^\times x, dx$ be Haar measures on $F_v^\times, F_v$ respectively, assigning mass 1 to $\mathfrak{o}_{F_v}^\times$ and $\mathfrak{o}_{F_v}$ respectively. For $x \in F_v$, put $n(x) = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$. Also let $w = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. We choose a Whittaker model for π_1 transforming by the character $n(x) \mapsto \psi(x)$, and a Whittaker model for π_2 transforming by the character $n(x) \mapsto \overline{\psi(x)}$.

Let Ψ_v be the characteristic function of $\mathfrak{o}_{F_v}^2$. Set W_1 to be the new vector in the Kirillov model of π_1, let W_2^* be the new vector in the Kirillov model of π_2, and set $W_2 = \pi_2\left(\begin{smallmatrix} 1 & 0 \\ 0 & \varpi \end{smallmatrix}\right)W_2^*$. Then both W_1, W_2 are invariant by the subgroup

$$K_0 = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,d \in \mathfrak{o}_{F_v}, c \in \varpi\mathfrak{o}_{F_v} \right\}. \tag{11.11}$$

Moreover W_2 is invariant by $n(\varpi^{-1}\mathfrak{o}_{F_v})$.

LEMMA 11.6. *Notation being as in* (11.6), *let* $L(s,\pi_1 \times \pi_2)$ *be the local L-factor. Then*

$$\frac{I_v(W_{1,v}, W_{2,v}, \Psi_v, s)}{L(s,\pi_1 \times \pi_2)} = \pm\frac{q_v^s}{q_v+1}.$$

Proof. We shall often use the following shorthand: if W is a function in the Whittaker model of $\pi \in \{\pi_1,\pi_2\}$ and for $z \in F_v^\times$, we write $W(z)$ for $W\left(\begin{smallmatrix} z & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Thus, for instance, $W_2(z) = W_2^*(z\varpi^{-1})$. The function $z \mapsto W(z)$ belongs to the Kirillov model of π.

Let $\epsilon \in \{-1,1\}$ be the local root number of π_1 (it lies in $\{-1,1\}$ since π_1 is self-dual). Then

$$\begin{aligned}(11.12)\qquad \int_{a\in F_v^\times} W_1(a)|a|^{s-1/2}d^\times a &= L(s,\pi_1),\\ \int_{a\in F_v^\times} W_2(a)|a|^{s-1/2}d^\times a &= q_v^{-(s-1/2)}L(s,\pi_2),\end{aligned}$$

as follows from defining properties of newforms and the fact $W_2(z)=W_2^*(z\varpi^{-1})$; moreover

$$\begin{aligned}(11.13)\qquad \int_{a\in F_v^\times} \pi_1(w)W_1(a)|a|^{s-1/2}d^\times a &= \epsilon q_v^{(s-1/2)}L(s,\pi_1),\\ \int_{a\in F_v^\times} \pi_2(w)W_2(a)|a|^{s-1/2}d^\times a &= q_v^{(s-1/2)}L(s,\pi_2),\end{aligned}$$

as follows from local functional equation for the standard L-function on GL(2): see [21, 2.18].[24]

Note moreover that W_1, W_2, $\pi_1(w)W_1$, and $\pi_2(w)W_2$ are all invariant by the maximal compact subgroup of the diagonal torus of GL_2. Thus (11.12) and (11.13) completely determine their restriction to the diagonal torus; we now explicate this.

Choose $\alpha\in\mathbb{C}$ so that $L(s,\pi_1)=(1-\alpha q_v^{-s})^{-1}$. In fact, $\alpha=-\epsilon q_v^{-1/2}$, by [21, Prop. 3.6]. Choose γ_1,γ_2 so that $L(s,\pi_2)=((1-\gamma_1 q_v^{-s})(1-\gamma_2 q_v^{-s}))^{-1}$. Recalling the notational convention established in the paragraph prior to (11.12), we see:

$$(11.14)\qquad W_1(\varpi^r)=\begin{cases}\alpha^r q_v^{-r/2}, & r\ge 0\\ 0, & r<0,\end{cases}$$

$$\pi_1(w)W_1(\varpi^r)=\begin{cases}\epsilon\alpha^{r+1}q_v^{-\frac{r+1}{2}}, & r\ge -1\\ 0, & r<-1,\end{cases}$$

$$W_2(\varpi^r)=\begin{cases}0, & r\le 0\\ 1, & r=1\\ (\gamma_1^{r-1}+\gamma_1^{r-2}\gamma_2+\cdots+\gamma_2^{r-1})q_v^{-\frac{r-1}{2}}, & r\ge 2,\end{cases}$$

$$\pi_2(w)W_2(\varpi^r)=\begin{cases}0, & r<-1\\ 1, & r=-1\\ (\gamma_1^{r+1}+\gamma_1^{r}\gamma_2+\cdots+\gamma_2^{r+1})q_v^{-\frac{r+1}{2}}, & r\ge 0.\end{cases}$$

[24] That is, $\int_{a\in F^\times} W(a)|a|^{s-1/2}d^\times a=\frac{L(s,\pi)}{\epsilon(s,\pi)L(1-s,\tilde{\pi})}\int_{F^\times} W(aw)|a|^{1/2-s}\omega^{-1}(a)d^\times a$, where ω denotes the central character. In particular, if π is a representation with trivial central character, and χ a character of $F^\times$, $\int_{a\in F^\times} W(a)\chi(a)d^\times a=\frac{L(\frac12,\pi\otimes\chi)}{\epsilon(\frac12,\pi\otimes\chi)L(\frac12,\tilde{\pi}\otimes\bar{\chi})}\int_{F^\times} W(aw)\chi^{-1}(a)d^\times a$.

The local integral we wish to evaluate is the right-hand side of (11.6). In the case at-hand, with N, G, Z denoting the F_v-points of the respective groups, we have (11.15)

$$I(s) := \int_{ZN\backslash G} W_1(g)W_2(g)\left(|\det(g)|^s \int_{t\in F_v^\times} \Psi_v((0,t)\cdot g)|t|^{2s}d^\times t\right)dg.$$

Using the Iwasawa decomposition, and recalling Ψ_v was the characteristic function of $\mathfrak{o}_{F_v}^2$ one finds

$$I(s) = (1-q_v^{-2s})^{-1}\int_{A\times K_v} \pi_1(k)W_1(a)\pi_2(k)W_2(a)|a|^{s-1}d^\times a\,dk,$$

where the measure dk is the Haar measure of total mass 1, and $d^\times a$ assigns mass 1 to $A\cap K_v$.

The function $k\mapsto \pi_1(k)W_1(a)\pi_2(k)W_2(a)$ is right invariant by K_0 (see (11.11) for definition) and left invariant by $N\cap K_v$. There are two $(N\cap K_v, K_0)$ double cosets in K_v, and we may therefore express $I(s)$ as a sum:

$$\begin{aligned}(11.16)\quad (1-q_v^{-2s})I(s) = {} & \frac{1}{q_v+1}\int_{a\in F_v^\times} W_1(a)W_2(a)|a|^{s-1}d^\times a \\ & + \frac{q_v}{q_v+1}\int_{a\in F_v^\times} \pi_1(w)W_1(a)\pi_2(w)W_2(a)|a|^{s-1}d^\times a.\end{aligned}$$

To evaluate $I(s)$, we use (11.14). Noting that $L(s,\pi_1\times\pi_2) = \frac{1}{(1-\alpha\gamma_1 q_v^{-s})(1-\alpha\gamma_2 q_v^{-s})}$, an easy computation shows

$$I(s) = \frac{L(s,\pi_1\times\pi_2)}{(q_v+1)(1-q_v^{-2s})}\left(\alpha q_v^{-1/2}q_v^{-(s-1)} + \epsilon q_v q_v^{s-1}\right)$$

from where we obtain $I(s) = \epsilon\frac{q_v^s}{q_v+1}L(s,\pi_1\times\pi_2)$. Note also that $I(s)$ satisfies the necessary functional equation. □

11.4. *Hecke-Jacquet-Langlands integral representations for standard L-functions.* Our goal here is to prove Propositions 6.1 and 6.2, used in the text. This amounts to explicit computations connected to Hecke-Jacquet-Langlands integral representations. Since, in the main text, we obtain subconvexity for GL(1) twists of GL(2) L-functions, *with polynomial dependence in all parameters*, we will have to be somewhat more precise than in the case of Rankin-Selberg L-functions.

Let π be a cuspidal representation of GL_2 over $\mathbb{A}_F$. Let χ be a unitary character of $\mathbb{A}_F^\times/F^\times$ of finite conductor $\mathfrak{f}$. Put $L_{\mathrm{unr}}(s,\pi\times\chi)$ to be the unramified part of the (finite) standard L-function:

$$L_{\mathrm{unr}}(s,\pi\times\chi) := \prod_{v \text{ finite}, \chi_v \text{ unramified}} L_v(s,\pi_v\times\chi_v).$$

Define μ_z as in (6.4), i.e., the measure on $\mathbf{X}_{\mathrm{GL}(2)}$ defined as

$$\mu_z(f) = \int_{|y|=z} f(a(y)n([\mathfrak{f}]))\chi(y)d^\times y.$$

We refer to Sections 2.3 and 2.5 for notation, as well as the start of Section 6 for a discussion of the meaning of μ_z in classical terms.

LEMMA 11.7. *Let v be a nonarchimedean place of F with residue characteristic q_v, and π_v an irreducible generic representation of* $\mathrm{GL}_2(F_v)$. *Let ψ_v be an unramified additive character of F_v. Let $\chi_v : F_v^\times \to \mathbb{C}$ a multiplicative character of conductor r, W_v be the new vector in the ψ_v-Whittaker model of π_v. Then*

$$\int_{y\in F_v^\times} W_v(a(y)n(\varpi_v^{-r}))\chi_v(y)|y|^{s-1/2}d^\times y = \begin{cases} L_v(s,\pi_v\times\chi_v), & r=0 \\ \theta, & r\geq 1, \end{cases}$$

where θ is a scalar of absolute value $q_v^{-r/2}(1-q_v^{-1})^{-1}$.

Proof. If $r=0$, then χ_v is unramified, and the result follows immediately from the definition of the new vector. Otherwise, χ_v is ramified, and we rewrite the integral under consideration as

$$\int_{y\in F_v^\times} W_v(a(y))\psi_v(\varpi_v^{-r}y)\chi_v(y)|y|^{s-1/2}d^\times y. \tag{11.17}$$

Now $W_v(a(y))$ vanishes when $v(y)<0$ and it is $\mathfrak{o}_{F_v}^\times$-invariant. The integral $\int_{v(y)=k}\chi_v(y)\psi_v(\varpi_v^{-r}y)d^\times y$ is nonvanishing only when $k=0$. (The vanishing for $k>0$ may be seen by considering the substitution $y\leftarrow yz, z\in 1+\varpi_v^{r-1}\mathfrak{o}_{F_v}$.)

In that case, it is a Gauss sum with absolute value $\frac{q_v^{-r/2}}{(1-q_v^{-1})}$, where the factor $(1-q_v^{-1})^{-1}$ arises from the measure normalization (cf. §2.6) namely $\int_{v(y)=0} d^\times y = 1$. The result follows. $\square$

LEMMA 11.8. *Let $d,\beta\geq 0$. Then there exists $\varphi\in\pi$ such that if*

$$\Phi(s) = \mathrm{N}(\mathfrak{f})^{1/2}\frac{\int_z \mu_z(\varphi)|z|^{s-1/2}d^\times z}{L_{\mathrm{unr}}(s,\pi\times\chi)}, \tag{11.18}$$

then $\Phi(s)$ is holomorphic and satisfies:

(1) $|\Phi(s)| \ll_{\Re(s),\epsilon} \mathrm{N}(\mathfrak{f})^\epsilon$ *and* $|\Phi(\frac{1}{2})| \gg_\epsilon \mathrm{N}(\mathfrak{f})^{-\epsilon}$.

(2) φ *is new at every finite place* (*i.e.*, *for each finite prime* $\mathfrak{q}$ *it is invariant by* $K_0[\mathfrak{q}^{s_\mathfrak{q}}]$, *where $s_\mathfrak{q}$ is the local conductor of the local constituent* $\pi_\mathfrak{q}$).

(3) *The Sobolev norms of φ satisfy the bounds* (*conductor notation as in* §2.12.2)

$$S_{2,d,\beta}(\varphi) \ll_\epsilon \mathrm{Cond}_\infty(\pi)^{2d+\epsilon}\mathrm{Cond}_f(\pi)^{\beta+\epsilon}\mathrm{Cond}_\infty(\chi)^{1/2+2d}. \tag{11.19}$$

Proof. For each infinite place w of F, denote by $\mathrm{Cond}_w(\chi)$ the contribution from w to the Iwaniec-Sarnak analytic conductor of χ (see §2.12.2.)

The map $\varphi \mapsto W_\varphi = \int_{F\backslash\mathbb{A}_F} e_F(x)\varphi(n(x)g)$ is an isomorphism between the space of π and the Whittaker model of π. For each finite v, take W_v to be a new vector in the Whittaker model of π_v. A point of caution is that e_F may not be unramified on F_v; to be absolutely concrete, we set $W_v(g) = W_{v,\mathrm{new}}(a(\varpi_v^{d_v})g)$, where $W_{v,\mathrm{new}}$ is ehe new vector in the Whittaker model of π_v taken with respect to an unramified additive character of F_v, and $d_v = v(\mathfrak{d})$ is the local valuation of the different.

Let us now choose W_v at the infinite places. Let g_1 be a smooth positive function of compact support on F_v, with support containing 0. Let $\deg(v) = 2$ if v is complex and $\deg(v) = 1$ if v is real. For $\infty|v$, define

$$W_v(y) = \mathrm{Cond}_v(\chi)g_1(\mathrm{Cond}_v(\chi)^{1/\deg(v)}(y-1)).$$

This is possible by the theory of the Kirillov model; thus W_v is a smooth (but not K_v-finite) vector. In words, if v is real, the function W_v is supported in a neighborhood of the identity of size $\mathrm{Cond}_v(\chi)^{-1}$ and takes values of size $|\mathrm{Cond}_v(\chi)|$ there; if v is complex, a similar statement holds but now W_v is supported in a disc around the identity with area $\mathrm{Cond}_v(\chi)^{-1}$.

Then there exists $\varphi \in \pi$ with $W_\varphi = \prod_v W_v$. By unfolding, it follows that for $\Re(s) \gg 1$,

(11.20)
$$\int_z \mu_z(\varphi)|z|^{s-1/2} d^\times z = c_F \prod_v \int_{y\in F_v^\times} W_v(a(y)n([\mathfrak{f}]))|y|^{s-1/2}\chi_v(y)d^\times y.$$

Here c_F is a constant depending only on F, arising from change of measure; it is entirely unimportant as we will be only interested in bounds.[25]

By Lemma 11.7, with $\Phi(s)$ as in the statement of the lemma,

$$\Phi(s) = c_F\theta' \cdot \mathrm{N}(\mathfrak{d})^{s-1/2} \prod_{\text{infinite } v} \int_{F_v^\times} W_v(a(y))|y|^{s-1/2}\chi_v(y)d^\times y, \tag{11.21}$$

where $|\theta'| = \prod_{\mathfrak{q}|\mathfrak{f}}(1-\mathrm{N}(\mathfrak{q})^{-1})^{-1}$. For this choice of φ, the second assertion if the lemma is clear, and, if we choose the support of g_1 to be small enough, the first assertion also follows easily.[26]

(11.19) follows from Lemma 8.4, together with Lemma 11.3 and the upper bound for L-functions near 1 due to Iwaniec. See [19, Chap. 8] for this bound. □

The previous lemma shows that $L(1/2, \pi \times \chi)$ may be "well-approximated" by an appropriate period integral. Unfortunately, this period integral is against a

[25] The measure μ_z is normalized as a probability measure, whereas to unfold from $\mathbb{A}_F^\times$ to $\prod_v F_v^\times$ we use the measures previously set up there (see §2.6).

[26] For the assertion concerning the lower bound for $|\Phi(1/2)|$, the point, in words, is that our choices are so that χ_v does not oscillate over the support of W_v; cf. Remark 2.2. Note how convenient it is, here and elsewhere, to use smooth vectors rather than K_∞-finite vectors; one could not, e.g., achieve W_v of compact support with K_∞-finite vectors.

measure of infinite mass, since $\mathbb{A}_F^\times/F^\times$ is of infinite volume. It is, therefore, convenient to know that the μ_z-integral of (11.18) can be truncated to a compact range without affecting the answer too much. This is, roughly speaking, the geometric equivalent of the approximate functional equation in the classical theory, and is provided by the next lemma. It says, roughly speaking, that the integral of (11.18) can be truncated to the range where z is around $\mathrm{N}(\mathfrak{f})^{-1}$.

LEMMA 11.9. *Let notation be as in Lemma* 11.8. *Let g_+, g_- be positive smooth functions on $\mathbb{R}_{\geq 0}$ such that $g_+ + g_- = 1$, $g_+(t) = 1$ for $t \geq 2$ and $g_-(t) = 1$ for all $t \leq 1/2$. Then*

$$\begin{aligned} I_+ &:= \int_z \mu_z(\varphi) g_+(z/T) d^\times z \ll_{g_+,\epsilon} (\mathrm{N}(\mathfrak{f})T)^{-1/2} (T\mathrm{Cond}(\pi)\mathrm{Cond}(\chi))^\epsilon, \\ I_- &:= \int_z \mu_z(\varphi) g_-(z/T) d^\times z \\ &\ll_{g_-,\epsilon} (\mathrm{N}(\mathfrak{f})T)^{1/2} (T\mathrm{Cond}(\chi))^\epsilon (\mathrm{Cond}_\infty(\chi)\mathrm{Cond}(\pi))^{1+\epsilon}. \end{aligned}$$

Proof. Recall the definition of μ_z from (6.4). Put $\widehat{g_\pm}(s) = \int g_\pm(x) x^{s-1} dx$, the Mellin transform of $g_\pm$; then $g_\pm$ is holomorphic in $\pm\Re(s) < 0$ and for any $M \geq 0$, $\pm\sigma < 0$ the integral $\int_{\Re(s)=\sigma} |\widehat{g_\pm}(s)|(1+|s|)^M ds$ is convergent. Then, for any $\pm\sigma > 0$, we have, by the Plancherel formula on $\mathbb{R}^\times$, that

$$\int_z \mu_z(\varphi) g_\pm(z/T) d^\times z = \frac{1}{2\pi i} \int_{\Re(s)=-\sigma} \left(\int \mu_z(\varphi) |z|^{-s} d^\times z \right) T^s \widehat{g_\pm}(s) ds.$$

So, for any $M > 0$,

$$|I_\pm| \ll_{\sigma, g_\pm, M} T^{-\sigma} \mathrm{N}(\mathfrak{f})^{-1/2} \sup_{\Re(s)=1/2+\sigma} \frac{|L_{\mathrm{unr}}(s)\Phi(s)|}{(1+|s|)^M} i,$$

where Φ is as in the previous lemma.Take $\sigma = 1/2 + \varepsilon$ in the $+$ case, $-1/2 - \varepsilon$ in the $-$ case.

Using Iwaniec's bounds on L-functions near 1 [19, Chap. 8] and the functional equation, we see that for sufficiently large M:

$$\begin{aligned} &\sup_{\Re(s)=1+\varepsilon} |L_{\mathrm{unr}}(s, \pi\times\chi)| \ll \mathrm{Cond}(\pi\otimes\chi)^\varepsilon, \qquad (11.22) \\ &\sup_{\Re(s)=-\varepsilon} \frac{|L_{\mathrm{unr}}(s,\pi\times\chi)|}{(1+|s|)^M} \ll_M \mathrm{Cond}(\pi\otimes\chi)^{1/2+\varepsilon} \\ &\qquad\qquad \times \prod_{\chi_v \text{ ramified finite}} \sup_{\Re(s)=-\varepsilon} |L_v(s, \pi_v \times \chi_v)|^{-1}. \end{aligned}$$

For each v where χ_v is ramified and $L_v(s, \pi_v \times \chi_v)$ is not identically 1, the representation π_v must also be ramified (i.e., not spherical). So one can bound the product on the second line on (11.22), using trivial bounds towards the Ramanujan conjecture, by $\mathrm{Cond}(\pi)^{1/2+2\varepsilon}$. The fact that $\mathrm{Cond}(\chi) = \mathrm{Cond}_\infty(\chi)\mathrm{N}(\mathfrak{f})$, the bound [7] $\mathrm{Cond}(\pi \otimes \chi) \ll \mathrm{Cond}(\pi)\mathrm{Cond}(\chi)^2$, and the (easily verified) analogue of this bound of [7] at archimedean places, allows one to conclude. $\square$

We now address the analogue of the previous lemmas when π is noncuspidal.[27]

LEMMA 11.10. *There is an absolute* $C > 0$ (*i.e.*, *depending only on* F) *and a Schwartz function* Ψ (*depending on* χ) *so that if we put*

$$\Phi(s,s') := \mathrm{N}(\mathfrak{f})^{1/2} \frac{\int_{y \in \mathbb{A}_F^\times / F^\times} \overline{E}_\Psi(s, a(y)n([\mathfrak{f}]))\chi(y)|y|^{s'} d^\times y}{L(\chi, s+s')L(\chi, 1-s+s')},$$

where $\overline{E}$ *is defined as in* (10.9), *then the integral defining* Φ *is absolutely convergent in a right half-plane* $\Re(s) \gg 1$. *Moreover*, Φ *extends from* $\Re(s), \Re(s') \gg 1$ *to a holomorphic function on* $\mathbb{C}^2$, *satisfying*

(1) $|\Phi(1/2, 0)| \gg 1$ *and* $|\Phi(s,s')| \ll C^{1+|\Re(s)|+|\Re(s')|}(1+|s|+|s'|)^C$.
Moreover, *given* $N > 0$ *we have that*

$$|\Phi(s,s')|(1+|s|+|s'|)^N \ll_{\Re(s),\Re(s'),N} \mathrm{Cond}_\infty(\chi)^{N'}, \tag{11.23}$$

where N' *and the implicit constant may be taken to depend continuously on* N, $\Re(s)$, $\Re(s')$.

(2) Ψ, *and so also* $E_\Psi(s,g)$ *is invariant by* $K_{\max}$.

(3) *Let* $h \in \mathcal{H}(\kappa)$ *be as in* (10.18), *and put* $E_h := \int_{\Re(s) \gg 1} h(s) E_\Psi(s,g)dg$. *For each* d, β *there is* N *such that* $S_{\infty,d,\beta}(E_h) \ll_\kappa \|h\|_0 \mathrm{Cond}_\infty(\chi)^N$.

Proof. We shall not explicitly address details of convergence. The manipulations that follow may be justified by similar reasoning to that of Lemma 10.6.

We now define a Schwartz function Ψ_v on F_v^2 for each place v. For each finite place v, let Ψ_v be the characteristic function of $\mathfrak{o}_{F_v}^2$.

For infinite v, we will first define a Schwartz function ρ_v on F_v, and then take $\Psi_v(x,y) = \rho_v(x)\widehat{\rho_v}(y)$; here $\widehat{\rho_v}$ is the inverse Fourier transform of ρ_v, satisfying $\int_{F_v} \widehat{\rho_v}(y) e_{F_v}(xy)dy = \rho_v(x)$.

Let g_1 be a smooth positive function of compact support on F_v, with support containing 0. Let $\deg(v) = 2$ if v is complex and $\deg(v) = 1$ if v is real. For $\infty | v$, define

$$\rho_v(y) = \mathrm{Cond}_v(\chi) g_1(\mathrm{Cond}_v(\chi)^{1/\deg(v)}(y-1)).$$

[27] The content of the following lemma, in classical language, is related to the following observation. Let χ be an even Dirichlet character mod q, and let $E^*(s,z)$ to be the Eisenstein series of (10.6), and $\overline{E}^*(s,z) := E^*(s,z) - \xi(2s)y^s - \xi(2(1-s))y^{1-s}$, then $\frac{1}{q}\int_0^\infty \sum_{1 \le x \le q-1} \chi(x)\overline{E}^*(s, \frac{x}{q} + iy)y^{s'} d^\times y$ coincides, up to some harmless factor, with $q^{-1/2}\Lambda(\chi, s+s')\Lambda(\chi, 1-s+s')$), where $\Lambda(\chi, s)$ is the usual Dirichlet L-function completed to include the Γ-factor at ∞. This particular expression is actually not quite suitable for our needs, because of the rapid decay of the Γ-factor swamps information about the finite L-function, and in fact the lemma uses (the equivalent of) a different test vector belonging to the automorphic representation underlying $E^*(z,s)$.

In words: in the real (resp. complex) case, ρ_v is localized in a real (resp. complex) interval (resp. disc) around 1, of length (resp. area) $\mathrm{Cond}_v(\chi)^{-1}$. Now put $\Psi_v(x,y) = \rho_v(x)\widehat{\rho_v}(y)$. The function Ψ_v is not compactly supported; however, it is of rapid decay. Indeed for each Schwartz norm $\mathcal{S}$, there is $M > 0$ such that

$$\mathcal{S}(\Psi_v) \ll \mathrm{Cond}_v(\chi)^M. \tag{11.24}$$

Define a Schwartz function on $\mathbb{A}_F^2$ via $\Psi(x,y) = \prod_v \Psi_v(x,y)$. As in Lemma 10.5, define $W_\Psi(s,g)$ as the Fourier coefficient of $E_\Psi(s,g)$. The choice of Ψ and Lemma 10.5 shows that $W_\Psi(s,g) = \prod_v W_v(g)$, where, for each finite v, W_v is given by Corollary 10.1, and satisfies

$$\int_{F_v^\times} W_v(a(y))|y|^{s'} d^\times y = q_v^{d_v(1+s'-s)} L_v(|\cdot|^s, s') L_v(|\cdot|^{1-s}, s'). \tag{11.25}$$

For infinite v, W_v satisfies (Lemma 10.5)

$$\begin{aligned}\int_{F_v^\times} & W_v(a(y))\chi_v(y)|y|^{s'} d^\times y \\ &= \int_{F_v^\times} \rho_v(x)\chi_v(x)|x|^{s+s'} d^\times x \int_{F_v^\times} \rho_v(x)\chi_v(x)|x|^{1-s+s'} d^\times x.\end{aligned} \tag{11.26}$$

By Fourier analysis and Lemma 10.3, $\bar{E}_\Psi(s,g) = \sum_{\alpha \in F^\times} W_\Psi(a(\alpha)g)$. Thus, for $\Re(s) \gg 1$,

$$\begin{aligned}\int_{y \in \mathbb{A}_F^\times / F^\times} & \bar{E}_\Psi(s, a(y)n([\mathfrak{f}]))\chi(y)|y|^{s'} d^\times y \\ &= \int_{y \in \mathbb{A}_F^\times} W_\Psi(s, a(y)n([\mathfrak{f}]))\chi(y)|y|^{s'} d^\times y \\ &= \prod_v \int_{y \in F_v^\times} W_v(a(y)n([\mathfrak{f}]))\chi_v(y)|y|^{s'} d^\times y.\end{aligned} \tag{11.27}$$

For $s \gg 1$, we use (10.17), (11.26) and Lemma 11.7 to evaluate the local factors, obtaining

$$\begin{aligned}\Phi(s,s') &= \theta' \cdot \mathrm{N}(\mathfrak{d})^{1+s'-s} \prod_{v \text{ infinite}} \int_{y \in F_v^\times} W_v(a(y))\chi_v(y)|y|^{s'} d^\times y \\ &= \theta' \cdot \mathrm{N}(\mathfrak{d})^{1+s'-s} \prod_{v \text{ infinite}} \int_{F_v^\times} \rho_v(x)\chi_v(x)|x|^{s+s'} d^\times x \int_{F_v^\times} \rho_v(x)\chi_v(x)|x|^{1-s+s'} d^\times x,\end{aligned} \tag{11.28}$$

where $|\theta'| = \prod_{\mathfrak{q}|\mathfrak{f}} (1 - \mathrm{N}(\mathfrak{q})^{-1})^{-1}$. Now, by choice of φ_v, the integral

$$I_v(s) := \int_{y \in F_v^\times} \rho_v(y)\chi_v(y)|y|^s d^\times y$$

satisfies $|I_v(1/2)| \gg 1$ and $|I_v(s)| \ll (1+|s|)^C C^{1+|\Re(s)|}$, at least when we choose the support of g_1 to be sufficiently small. It also satisfies

$$|I_v(s)|(1+|s|)^N \ll_{N,\Re(s)} \mathrm{Cond}_\infty(\chi)^{N'},$$

where N' and the implicit constant may be taken to depend continuously on $N, \Re(s)$.

The corresponding facts (i.e., the first assertion of the lemma) about Φ follow immediately. The second assertion of the lemma is immediate from our choice of Ψ. As for the third and final assertion, it follows from Remark 10.3 and (11.24). □

LEMMA 11.11. *Let notation be as in the previous lemma. Assume χ is ramified at least one finite place. Let g_+, g_- be positive smooth functions on $\mathbb{R}_{\geq 0}$ such that $g_+ + g_- = 1$, $g_+(t) = 1$ for $t \geq 2$ and $g_-(t) = 1$ for all $t \leq 1/2$.*

Then

$$\mu_z(E_h) \ll_{K,\Psi,h} \min(z^K, z^{-K}) \tag{11.29}$$

for any $K \geq 1$.

Moreover, there is an absolute $N > 0$ such that

$$I_+ := \int_z \mu_z(E_h) g_+(z/T) d^\times z \ll (\mathrm{N}(\mathfrak{f})T)^{-1/2} (T\mathrm{Cond}(\chi))^\epsilon \|h\|_N,$$
$$I_- := \int_z \mu_z(E_h) g_-(z/T) d^\times z \ll (\mathrm{N}(\mathfrak{f})T)^{1/2} (T\mathrm{Cond}(\chi))^\epsilon \mathrm{Cond}_\infty(\chi)^{1+\epsilon} \|h\|_N.$$

(Here the norms $\|\cdot\|_N$ are as in (10.18).*)*

Proof. Again, we shall leave verification of convergence to the reader. Recall that, with the relevant measure on $\mathbb{A}_F^\times/F^\times$ having mass 1,

$$\begin{aligned} \mu_z(E_h) &= \int_{y\in\mathbb{A}_F^\times/F^\times, |y|=z} E_h(a(y)n[\mathfrak{f}])\chi(y) d^\times y \\ &= \int_{y\in\mathbb{A}_F^\times/F^\times, |y|=z} \int_{\Re(s)\gg 1} h(s) E_\Psi(s, a(y)n([\mathfrak{f}]))\chi(y) d^\times y \\ &= \int_{y\in\mathbb{A}_F^\times/F^\times, |y|=z} \int_{\Re(s)\gg 1} h(s) \bar{E}_\Psi(s, a(y)n([\mathfrak{f}]))\chi(y) d^\times y. \end{aligned} \tag{11.30}$$

Here, the last equality is justified by the fact that $(E_\Psi - \bar{E}_\Psi)(s, a(y)n([\mathfrak{f}]))$ is invariant under $y \mapsto yy'$, for $y' \in \prod_v \mathfrak{o}_{F_v}^\times$. On the other hand, χ is nontrivial on $\prod_v \mathfrak{o}_{F_v}^\times$, by assumption.

Combining (11.30) with Lemma 11.10, we have

$$\begin{aligned} &\int_z \mu_z(E_h) z^{s'} d^\times z \\ &\quad = c_F \mathrm{N}(\mathfrak{f})^{-1/2} \int_{\Re(s)\gg 1} h(s) L(\chi, 1-s+s') L(\chi, s+s') \Phi(s,s') ds. \end{aligned} \tag{11.31}$$

Here c_F is an (unimportant) constant arising from measure normalization, as in (11.20).

The assertion (11.29) follows immediately from this, inverse Mellin transform, and analytic properties of the right-hand side.

Now proceed as in Lemma 11.9; it follows that (for any M)

$$|I_\pm| \ll T^{\mp(1/2+\varepsilon)} \mathrm{N}(\mathfrak{f})^{-1/2} \sup_{\Re(s')=\pm(1/2+\varepsilon)} (1+|s'|)^{-M} \times \int h(s) L(\chi, 1-s+s') L(\chi, s+s') \Phi(s,s') ds.$$

We deal with the case of I_-. In that case, we take the inner integral to be over $\Re(s) = 1/2$, and put $s' = -1/2 - \varepsilon - it'$, and it will suffice to bound

$$\int h(1/2+it) L(\chi, -\varepsilon - it - it') L(\chi, -\varepsilon + it - it') (1+|t|+|t'|)^C.$$

This is bounded, up to an implicit constant depending on ε, by

$$\mathrm{Cond}(\chi)^{1+2\varepsilon} \|h\|_{M'} (1+|t'|)^{C'}$$

for sufficiently big M', C', whence the result. □

References

[1] J. BERNSTEIN and A. REZNIKOV, Sobolev norms of automorphic functionals and Fourier coefficients of cusp forms, *C. R. Acad. Sci. Paris Sér.* I *Math.* **327** (1998), 111–116. MR 99g:11061 Zbl 0915.11030

[2] ———, Sobolev norms of automorphic functionals, *Int. Math. Res. Not.* **2002** (2002), 2155–2174. MR 2003h:11058 Zbl 1022.11025

[3] ———, Periods, subconvexity of L-functions and representation theory, *J. Differential Geom.* **70** (2005), 129–141. MR 2007b:11063 Zbl 1097.11022

[4] V. BLOMER, G. HARCOS, and P. MICHEL, Bounds for modular L-functions in the level aspect, *Ann. Sci. École Norm. Sup.* **40** (2007), 697–740. MR 2009g:11058 Zbl 05240226

[5] J. BOURGAIN and E. LINDENSTRAUSS, Entropy of quantum limits, *Comm. Math. Phys.* **233** (2003), 153–171. MR 2004c:11076 Zbl 1018.58019

[6] D. A. BURGESS, On character sums and L-series, *Proc. London Math. Soc.* **12** (1962), 193–206. MR 24 #A2570 Zbl 0106.04004

[7] C. J. BUSHNELL and G. HENNIART, An upper bound on conductors for pairs, *J. Number Theory* **65** (1997), 183–196. MR 98h:11153 Zbl 0884.11049

[8] L. CLOZEL and E. ULLMO, Équidistribution de mesures algébriques, *Compos. Math.* **141** (2005), 1255–1309. MR 2006j:22009 Zbl 1077.22014

[9] P. B. COHEN, Hyperbolic equidistribution problems on Siegel 3-folds and Hilbert modular varieties, *Duke Math. J.* **129** (2005), 87–127. MR 2006g:11090 Zbl 1155.11326

[10] M. COWLING, U. HAAGERUP, and R. HOWE, Almost L^2 matrix coefficients, *J. Reine Angew. Math.* **387** (1988), 97–110. MR 89i:22008 Zbl 0638.22004

[11] W. DUKE, J. FRIEDLANDER, and H. IWANIEC, Bounds for automorphic L-functions, *Invent. Math.* **112** (1993), 1–8. MR 94c:11043 Zbl 0765.11038

[12] ———, Bounds for automorphic L-functions. II, *Invent. Math.* **115** (1994), 219–239. MR 95a:11044 Zbl 0812.11032

[13] ———, Class group L-functions, *Duke Math. J.* **79** (1995), 1–56. MR 97h:11138 Zbl 0838.11058

[14] L. FLAMINIO and G. FORNI, Invariant distributions and time averages for horocycle flows, *Duke Math. J.* **119** (2003), 465–526. MR 2004g:37039 Zbl 1044.37017

[15] J. FRIEDLANDER and H. IWANIEC, A mean-value theorem for character sums, *Michigan Math. J.* **39** (1992), 153–159. MR 92k:11084 Zbl 0765.11037

[16] A. GOOD, Cusp forms and eigenfunctions of the Laplacian, *Math. Ann.* **255** (1981), 523–548. MR 82i:10029 Zbl 0439.30031

[17] G. HARCOS and P. MICHEL, The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points. II, *Invent. Math.* **163** (2006), 581–655. MR 2007j:11063 Zbl 1111.11027

[18] D. A. HEJHAL, On value distribution properties of automorphic functions along closed horocycles, in XVI th *Rolf Nevanlinna Colloquium* (Joensuu, 1995) (I. LAINE and O. MARTIO, eds.), de Gruyter, Berlin, 1996, pp. 39–52. MR 97j:11024 Zbl 0867.11030

[19] H. IWANIEC, *Introduction to the Spectral Theory of Automorphic Forms*, *Biblioteca de la Revista Matemática Iberoamericana.* [*Library of the Revista Matemática Iberoamericana*], Revista Matemática Iberoamericana, Madrid, 1995. MR 96f:11078 Zbl 0847.11028

[20] H. JACQUET, *Automorphic Forms on* GL(2). *Part* II, *Lecture Notes in Math.* **278**, Springer-Verlag, New York, 1972. MR 58 #27778 Zbl 0243.12005

[21] H. JACQUET and R. P. LANGLANDS, *Automorphic Forms on* GL(2), *Lecture Notes in Math.* **114**, Springer-Verlag, New York, 1970. MR 53 #5481 Zbl 0236.12010

[22] H. H. KIM, Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2, *J. Amer. Math. Soc.* **16** (2003), 139–183. MR 2003k:11083 Zbl 1018.11024

[23] D. Y. KLEINBOCK and G. A. MARGULIS, Logarithm laws for flows on homogeneous spaces, *Invent. Math.* **138** (1999), 451–494. MR 2001i:37046 Zbl 0934.22016

[24] E. KOWALSKI, P. MICHEL, and J. VANDERKAM, Rankin-Selberg L-functions in the level aspect, *Duke Math. J.* **114** (2002), 123–191. MR 2004c:11070 Zbl 1035.11018

[25] Y. V. LINNIK, *Ergodic Properties of Algebraic Fields*, *Translated from the Russian by M. S. Keane. Ergebnisse der Mathematik und ihrer Grenzgebiete* **45**, Springer-Verlag, New York, 1968. MR 39 #165 Zbl 0162.06801

[26] G. A. MARGULIS, *On Some Aspects of the Theory of Anosov Systems*, *Springer Monogr. Math.*, Springer-Verlag, New York, 2004. MR 2004m:37049 Zbl 1140.37010

[27] P. MICHEL, The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points, *Ann. of Math.* **160** (2004), 185–236. MR 2006d:11048 Zbl 1068.11033

[28] P. MICHEL and A. VENKATESH, The subconvexity problem for GL_2, in preparation.

[29] ———, Equidistribution, L-functions and ergodic theory: on some problems of Yu. V. Linnik, in *Proc. Madrid* ICM, Vol. II, 421–457, Eur. Math. Soc., Zürich, 2006. MR 2008g:11085 Zbl 1157.11019

[30] H. OH, Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants, *Duke Math. J.* **113** (2002), 133–192. MR 2003d:22015 Zbl 1011.22007

[31] M. RATNER, Rigidity of time changes for horocycle flows, *Acta Math.* **156** (1986), 1–32. MR 87h:58175 Zbl 0694.58036

[32] ———, The rate of mixing for geodesic and horocycle flows, *Ergodic Theory Dynam. Systems* **7** (1987), 267–288. MR 88j:58103 Zbl 0623.22008

[33] ———, Raghunathan's conjectures for $\mathrm{SL}(2,\mathbf{R})$, *Israel J. Math.* **80** (1992), 1–31. MR 94k:22024 Zbl 0785.22013

[34] A. REZNIKOV, Rankin-Selberg without unfolding and bounds for spherical Fourier coefficients of Maass forms, *J. Amer. Math. Soc.* **21** (2008), 439–477. MR 2008m:11106 Zbl 05234673

[35] P. SARNAK, Asymptotic behavior of periodic orbits of the horocycle flow and Eisenstein series, *Comm. Pure Appl. Math.* **34** (1981), 719–739. MR 83m:58060 Zbl 0501.58027

[36] P. SARNAK, Fourth moments of Grössencharakteren zeta functions, *Comm. Pure Appl. Math.* **38** (1985), 167–178. MR 86g:11051 Zbl 0577.10026

[37] ______, Integrals of products of eigenfunctions, *Internat. Math. Res. Notices* (1994), 251 ff., approx. 10 pp. MR 95i:11039 Zbl 0833.11020

[38] N. A. SHAH, Limit distributions of polynomial trajectories on homogeneous spaces, *Duke Math. J.* **75** (1994), 711–732. MR 95j:22022 Zbl 0818.22005

[39] Y. SHALOM, Rigidity, unitary representations of semisimple groups, and fundamental groups of manifolds with rank one transformation group, *Ann. of Math.* **152** (2000), 113–182. MR 2001m:22022 Zbl 0970.22011

[40] G. SHIMURA, *Introduction to the Arithmetic Theory of Automorphic Functions*, Publications of the Mathematical Society of Japan, Iwanami Shoten, Publishers, Tokyo and Princeton Univ. Press, 1971. MR 47 #3318 Zbl 0221.10029

[41] P. SÖHNE, An upper bound for Hecke zeta-functions with Groessencharacters, *J. Number Theory* **66** (1997), 225–250. MR 98k:11170 Zbl 0885.11062

[42] K. SOUNDARARAJAN, Nonvanishing of quadratic Dirichlet L-functions at $s = \frac{1}{2}$, *Ann. of Math.* **152** (2000), 447–488. MR 2001k:11164 Zbl 0964.11034

[43] A. STRÖMBERGSSON, On the uniform equidistribution of long closed horocycles, *Duke Math. J.* **123** (2004), 507–547. MR 2005f:11105 Zbl 1060.37023

[44] T. TAO, The ergodic and combinatorial approaches to Szemerédi's theorem, in *Additive Combinatorics*, CRM *Proc. Lecture Notes* **43**, Amer. Math. Soc., Providence, RI, 2007, pp. 145–193.

[45] V. VATSAL, Uniform distribution of Heegner points, *Invent. Math.* **148** (2002), 1–46. MR 2003j:11070 Zbl 1119.11035

[46] M. WOODBURY, Ph.D. thesis, University of Wisconsin, in preparation.

[47] S.-W. ZHANG, Equidistribution of CM-points on quaternion Shimura varieties, *Int. Math. Res. Not.* **2005** (2005), 3657–3689. MR 2007g:11067 Zbl 1096.14016

(Received August 1, 2006)

E-mail address: akshay@math.stanford.edu
DEPARTMENT OF MATHEMATICS, BUILDING 380, STANFORD UNIVERSITY,
PALO ALTO, CA 94305, UNITED STATES

ISSN 0003-486X

ANNALS OF MATHEMATICS

This periodical is published bimonthly by the Department of Mathematics at Princeton University with the cooperation of the Institute for Advanced Study. Annals is typeset in TeX by Sarah R. Warren and produced by Mathematical Sciences Publishers. The six numbers each year are divided into two volumes of three numbers each.

Editorial correspondence

Papers submitted for publication and editorial correspondence should be addressed to Maureen Schupsky, Annals of Mathematics, Fine Hall-Washington Road, Princeton University, Princeton, NJ, 08544-1000 U.S.A. The e-mail address is annals@math.princeton.edu.

Preparing and submitting papers

The Annals requests that all papers include an abstract of about 150 words which explains to the nonspecialist mathematician what the paper is about. It should not make any reference to the bibliography. Authors are encouraged to initially submit their papers electronically and in PDF format. Please send the file to: annals@math.princeton.edu or to the Mathematics e-print arXiv: front.math.ucdavis.edu/submissions. If a paper is submitted through the arXiv, then please e-mail us with the arXiv number of the paper.

Proofs

A PDF file of the galley proof will be sent to the corresponding author for correction. If requested, a paper copy will also be sent to the author.

Offprints

Authors of single-authored papers will receive 30 offprints. (Authors of papers with one co-author will receive 15 offprints, and authors of papers with two or more co-authors will receive 10 offprints.) Extra offprints may be purchased through the editorial office.

Subscriptions

The price for a print and online subscription, or an online-only subscription, is $390 per year for institutions. In addition, there is a postage surcharge of $40 for print subscriptions that are mailed to countries outside of the United States. Individuals interested in subscriptions for their own personal use should contact the publisher at the address below. Subscriptions and changes of address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 (e-mail: contact@mathscipub.org; phone: 1-510-643-8638; fax: 1-510-295-2608). (Checks should be made payable to "Mathematical Sciences Publishers".)

Back issues and reprints

Orders for missing issues and back issues should be sent to Mathematical Sciences Publishers at the above address. Claims for missing issues must be made within 12 months of the publication date. Online versions of papers published five or more years ago are available through JSTOR (www.jstor.org).

Microfilm

Beginning with Volume 1, microfilm may be purchased from NA Publishing, Inc., 4750 Venture Drive, Suite 400, PO Box 998, Ann Arbor, MI 48106-0998; phone: 1-800-420-6272 or 734-302-6500; email: info@napubco.com, website: www.napubco.com/contact.html.

Printed in U.S.A. by Sheridan Printing Company, Inc., Alpha, NJ

TABLE OF CONTENTS