Math. Proc. Camb. Phil. Soc. (2024), **176**, 517–545
doi:10.1017/S0305004124000021
First published online 12 March 2024

The sup-norm problem beyond the newform

BY EDGAR ASSING[†]
University of Bonn, Mathematical Institute, Endenicher Allee 60, D-53115 *Bonn, Germany*
e-mail: assing@math.uni-bonn.de

(*Received* 23 *February* 2022*; revised* 11 *October* 2023*; accepted* 04 *October* 2023)

Abstract

In this paper we take up the classical sup-norm problem for automorphic forms and view it from a new angle. Given a twist minimal automorphic representation π we consider a special small $\mathrm{GL}_2(\mathbb{Z}_p)$-type V in π and prove global sup-norm bounds for an average over an orthonormal basis of V. We achieve a non-trivial saving when the dimension of V grows.

2020 Mathematics Subject Classification: 11F03 (Primary); 11F70, 11F85, 22E50 (Secondary)

1. *Introduction*

It is a classical problem in analysis and mathematical physics, more precisely Quantum Chaos, to bound the L^∞-norm of certain eigenfunctions on manifolds. In the most basic situation one considers a Riemann surfaces X of finite volume and eigenfunctions ϕ of the Laplace–Beltrami operator Δ_X. A sup-norm bound in the spectral aspect is then an estimates of the form

$$\frac{\|\phi\|_\infty}{\|\phi\|_2} \ll_X (1+|t_\phi|)^{\frac{1}{2}-\delta+\epsilon}, \tag{1}$$

where $\lambda_\phi = 1/4 + t_\phi^2$ is the Laplace–Beltrami eigenvalue of ϕ. The local bound corresponds to $\delta = 0$ and is known in great generality. The sup-norm problem asks for improved bounds featuring some $\delta > 0$. The sup-norm problem has only been solved for very special surfaces X and is hopeless in general. Indeed there is a well-known obstruction to the sup-norm problem coming from large eigenspaces V_λ given by the inequality

$$\dim_{\mathbb{C}} V_\lambda \ll_X \sup_{\phi\in V_\lambda} \frac{\|\phi\|_\infty}{\|\phi\|_2}.$$

This observation is enough to establish the well-known fact, that the local bound (i.e. (1) with $\delta = 0$) can not be improved for the sphere $X = S^2$. So far we have only described the most basic version of the sup-norm problem which, is already very interesting on its own. In addition it admits many variations which have been studied throughout the years. An example for such a variation is the so called level aspect where the base manifold changes in some convenient family $X_1, X_2, \ldots$ and one keeps track of this change in the sup-norm

[†]Supported by the Germany Excellence Strategy grant EXC-2047/1-390685813 and partially supported by TRR 358.

bound (1) using a suitable parameter called the level. Another generalisation that should be mentioned allows X to be a manifold of higher dimension and rank.

Essentially any progress that has been made towards the sup-norm problem as introduced above relies on the arithmeticity of X. The basic idea introduced in the monumental paper [14] is to employ additional symmetries (in the form of Hecke operators) to build a spectral projector that is sharper than the one constructed with only the Laplace–Beltrami operator at hand. Morally this might be thought of as forcing a multiplicity one situation even if the Laplace–Beltrami eigenspaces can not be rigorously controlled. The result of this method is a bound as in (1) with $\delta = 1/12$ for compact quotients $X = \Gamma\backslash\mathbb{H}$ constructed from maximal orders in quaternion algebras.

Since its appearance the method from [14] has been tweaked, modified and generalised, see for example [1, 6, 7, 21, 23] and the references within. Much work is concerned with congruence quotients $X = \Gamma_0(N)\backslash\mathbb{H}$ on which so called Hecke–Maaß newforms are considered. Since these newforms enjoy a nice multiplicity one property they are natural candidates for the sup-norm problem. In this paper we are going beyond the case of newforms and consider situations where the dimension of the underlying eigenspace grows. In other words, we solve the sup-norm problem in the dimension aspect. This aspect is a new facet of the sup-norm problem which seems extremely interesting and is not yet well studied. While our result is the first in p-adic setting it is only preceded by [5] where an archimedean version of this aspect is discussed.

To explain our result and its connection to the work of Blomer, Harcos, Maga and Milićević it will be most convenient to leave the classical world of Hecke–Maaß-newforms behind and work in the language of automorphic forms and automorphic representations.

The sup-norm problem we will consider is connected to small $\mathrm{GL}_2(\mathbb{Z}_p)$-types in cuspidal automorphic representations π, where $p > 3$ is prime. Comparing this to the recent work [5] we are replacing the archimedean place ∞ by a finite place p and the minimal $U(2)$-type of some automorphic representation by a suitably chosen $\mathrm{GL}_2(\mathbb{Z}_p)$-type. Note that in order to afford interesting K-types at the archimedean place it is necessary to work over fields admitting complex places or in higher rank. In the p-adic world we already meet interesting cases when working with automorphic forms for GL_2 over $\mathbb{Q}$.

1·1. *Set-up and main result*

Before we continue our discussion we need to fix some notation. Let $G(R) = \mathrm{GL}_2(R)$ for some ring R and let $\mathbb{A}$ be the adele ring over $\mathbb{Q}$. We will be working with cuspidal automorphic representations π of $G(\mathbb{A})$ with unitary central character ω_π. Abusing notation we will write $\pi \subset L_0^2(G(\mathbb{Q})\backslash G(\mathbb{A}), \omega_\pi)$ assuming that π acts on an irreducible subspace of cuspidal automorphic forms by right translation. Given a compact subgroup H we write π^H for the space of H-invariant elements in π.

Set $K_\infty = SO(2)$ and $K_l = \mathrm{GL}_2(\mathbb{Z}_l)$ for primes l. Combining these we get the compact subgroup $K = \prod_v K_v \subset G(\mathbb{A})$. Given a prime $p > 3$ and $m > 0$ we consider the smaller compact subgroup

$$K(p^m) = K_\infty \times K_p(m) \times \prod_{l \neq p} K_l \text{ for } K_p(p^m) = 1 + p^m \cdot \mathrm{Mat}_{2\times 2}(\mathbb{Z}_p) \subset K_p.$$

Note that $K(p^m)$ is normal and of finite index in K.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Throughout we restrict ourselves to the situation where π is unramified (i.e. spherical) away from p. In particular it is spherical at ∞ and one associates the spectral parameter t_π. Set $T = 1 + |t_\pi|$. Further, we have

$$m_\pi = \min\{m \in \mathbb{N} \colon \pi^{K(p^m)} \neq \{0\}\} < \infty. \tag{2}$$

We set $V = \pi^{K(p^{m_\pi})}$ and observe that $\pi|_K$ endows V with the structure of a K-module. It turns out that, if π is twist minimal, V is irreducible (see Lemma 2·2 below). Set $d = \dim_{\mathbb{C}} V$, choose an orthonormal basis $\phi_1, \ldots, \phi_d$ for V with respect to the $L_0^2(G(\mathbb{Q})\backslash G(\mathbb{A}), \omega_\pi)$ inner product. Define

$$\Phi(g) = \left(\sum_{i=1}^{d} |\phi_i(g)|^2\right)^{\frac{1}{2}},$$

which is independent of the choice of the orthonormal basis $\phi_1, \ldots, \phi_d$. We are concerned with the sup-norm of $\Phi(g)$ and obtain the following theorem which is a close analogue to [**5**, theorem 1].

THEOREM 1·1. *Let $p > 3$ be prime and suppose π is twist minimal. In the notation above we have*

$$\|\Phi\|_\infty \ll T^{\frac{1}{2}+\epsilon} d^{\frac{11}{12}+\epsilon}.$$

If the (arithmetic)-conductor of π is a perfect square (i.e. the exponent-conductor of the p-component π_p of π is even) or the p-component π_p of π is not supercuspidal, then we have the better bound

$$\|\Phi\|_\infty \ll T^{\frac{1}{2}+\epsilon} d^{\frac{5}{6}+\epsilon}. \tag{3}$$

While in the spectral aspect (i.e. the T-aspect in our statement) we only recover the local bound, the key feature of our theorem is the sub-local exponent in the dimension aspect d. Given the obstruction to the sup-norm problem coming from growing eigenspaces the aspect under consideration may seem counter intuitive. However, we are letting the dimension of the eigenspace vary in a controlled manner and manage to show that one can still achieve a considerable power saving in d on average over any orthonormal basis.

Note that the sup-norm bound given in the theorem holds globally. Thus, unlike the one in [**5**, theorem 1], no restriction to a compact domain is necessary here. As usual when proving global sup-norm bound the argument consists of two steps. First, a bound via the Whittaker expansion takes care of the regions close to the cusps. This part of the argument is fairly standard but requires some new computations of ramified Whittaker vectors. Second, a bound obtained from the amplified pre-trace inequality is used to handle the bulk. At this point it becomes crucial that we are only treating the average function Φ. Indeed, this allows us to identify the test function on the geometric side as a character of a finite group. The analysis of this character is carried out in Lemma 4·4 below and relies on character tables given in [**15**]. This is the only place where the assumption $p > 3$ is used.

To end this section let us briefly discuss the numerology of the exponents in the d-aspect. For simplicity we restrict this discussion to the cases in which our result gives the strong bound (3). Let us start by talking about the local- (not to say trivial-) bound (in the bulk). To obtain this we can follow Marshall's strategy (see [**18**]) which leads to the following. Let F

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

be any cuspidal automorphic form so that the translates $\phi(\cdot\, k)$, $k\in K$, generate an irreducible K-module W_F. Then choosing certain K-matrix-coefficients as test functions in the pre-trace inequality yields

$$\frac{\|F|_\Omega\|_\infty}{\|F\|_2} \ll \dim_{\mathbb{C}}(W_F)^{\frac{1}{2}}. \tag{4}$$

Applying this to Φ upon noting that $\|\Phi\|_2 = d^{\frac{1}{2}}$ suggests the local bound

$$\|\Phi\|_\infty \ll d^{1+\epsilon}.$$

(The same bound can also be obtained from the Whittaker expansion coupled with a suitable generating domain.) Thus amplification allows us to improve the exponent from the local bound by 1/6, which should be an familiar exponent. More suggestively we can write our main result as

$$\frac{\|\Phi\|_\infty}{\|\Phi\|_2} \ll d^{\frac{1}{2}-\frac{1}{6}+\epsilon}.$$

One could say that Theorem 1·1 implies $\|\phi_i\|_\infty \ll d^{\frac{1}{3}}$ on average. Note that if p^{m_π} agrees with the arithmetic conductor p^{n_π} of π, then this result is not very interesting. Indeed, in this case we can generate the elements $\phi_1,\ldots,\phi_d$ in V directly from the newform $\phi_\circ$ in π. By now there are very good bounds for this newform (and thus also for the ϕ_i's) known in the literature. See [23] if $m_\pi = n_\pi = 1$ or [9] in general. However, in the remaining cases (since π is assumed to be twist minimal these correspond to the situation where π is supercuspidal at p) our result provides new information in the sup-norm problem. Indeed one can still generate V from a translate of the newfom $\phi_\circ$. (This is precisely the strategy used in [18, 21] to derive local bounds for the newform of arbitrary level using (4).) Translated into the level-aspect our result now essentially says that the sup-norm of the ϕ_i's is bounded by $p^{\frac{1}{3}\lceil\frac{n_\pi}{2}\rceil}$ on average. To the best of our knowledge this can not be derived from any known sup-norm results on the newform $\phi_\circ$.

Finally we want to compare our result to the guiding archimedean example [5, theorem 1]. Recall that we need to replace the K-module V by some irreducible $U(2)$ representation W. This representation W will occur as the minimal $U(2)$-type in some cuspidal automorphic π of $G(\mathbb{A}_{\mathbb{Q}(i)})$. Note that if $\dim_{\mathbb{C}} W \asymp l$ we can think of π (or rather π_∞) having spectral density $\asymp l^2$. This explains the local bounds

$$\frac{\|\Phi\|_\infty}{\|\Phi\|_2} \ll l^{1+\epsilon} \text{ or } \|\Phi\|_\infty \ll l^{\frac{3}{2}+\epsilon},$$

where Φ is constructed as an average over some suitable basis of W similar to our construction above. As result of an amplification process the authors of [5] arrive at

$$\frac{\|\Phi|_\Omega\|_\infty}{\|\Phi\|_2} \ll (l^2)^{\frac{1}{2}-\frac{1}{12}+\epsilon}.$$

Our notation suggests that in the result from [5] the number l^2 playes the role of our d. This can be explained via the spectral density of π_∞ and respectively π_p. Indeed while in the archimedean situation the spectral density is roughly l^2 in our case the spectral density is linearly related to d. Thus in both cases the square root of the spectral density seems to determine the trivial bound. (This is only reasonable because we are considering minimal or

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

close to minimal K-types in both cases.) Note that the quality of the saving $1/6$ in the p-adic versus $1/12$ in the archimedean case comes from slightly different behaviour of the spectral transform.

Finally, let us remark that if the exponent conductor of π_p is odd and π_p is supercuspidal, then our bounds for the spectral transform, which in this case are linked to certain badly-behaved characters of GL_2 over finite rings, are comparable to those used in [5]. This explains that in this case we have matching numerology and obtain only a saving of $1/12$ in the final exponent. Translated to the level aspect our result states that on average the ϕ_i's are bounded by $p^{\frac{5(n_\pi+1)}{24}}$. Bounds of this quality are known for newforms only in the compact setting, see [13].

Remark 1·2. Questions of these type should be even more interesting when considered in higher rank. The reason is that in higher rank the analogously defined small K-types can not be generated from translates of the newform. For example if one considers a depth-zero supercuspidal representation π_p of $\mathrm{GL}_3(\mathbb{Q}_p)$, then it has (arithmetic)-conductor p^3 and the space $\pi_p^{K_p^{(3)}(1)}$, where $K_p^{(3)}(1)$ is the principal congruence subgroup modulo p in $\mathrm{GL}_3(\mathbb{Z}_p)$, is non-zero. However, it seems impossible to find a translate of the newform that generates $\pi_p^{K_p^{(3)}(1)}$. Indeed this would mean finding $g \in \mathrm{GL}_3(\mathbb{Q}_p)$ with

$$K_p^{(3)}(1) \subset g^{-1} \begin{bmatrix} \mathbb{Z}_p & \mathbb{Z}_p & \mathbb{Z}_p \\ \mathbb{Z}_p & \mathbb{Z}_p & \mathbb{Z}_p \\ p\mathbb{Z}_p & p\mathbb{Z}_p & 1+p\mathbb{Z}_p \end{bmatrix} g.$$

However, the question treated in this paper still makes sense and trying to answer it is work in progress.

2. *Preliminary considerations*

In this section we are putting in some ground work on which the following sections will rely.

Recall that π was a cuspidal automorphic representation. Since we are assuming that π_v is unramified for $v \neq p$, the (arithmetic)-conductor of π is p^{n_π} for $n_\pi \in \mathbb{N} \cup \{0\}$. When $n_\pi = 0$ we have $d = 1$ and our theorem reduces to the local bound in the spectral aspect, so that without loss of generality we can assume $n_\pi \geq 1$ throughout. By Flath's factorisation theorem we can fix an isomorphism $\pi \cong \bigotimes \pi_v$. Note that also the central character of π factors as $\omega_\pi = \bigotimes_v \omega_{\pi_v}$ where ω_{π_v} is the central character of π_v. For $v \neq p$ we can fix a spherical (i.e. K_v-invariant vector) $\phi_v^\circ \in \pi_v^{K_v}$. This vector is unique up to scaling. Recall that ϕ_i, $i = 1, \ldots, d$ forms an orthonormal basis of $V = \pi^{K(p^{m_\pi})}$. Thus there is $\phi_p^{(i)}$ so that we can identify

$$\phi_i = \phi_p^{(i)} \otimes \bigotimes_{v \neq p} \phi_v^\circ.$$

Since the spherical functions ϕ_p° are well understood much of our work boils down to understanding properties of an orthogonal basis

$$\mathrm{span}\{\phi_p^{(1)}, \ldots, \phi_p^{(d)}\} = \pi_p^{K_p(m_\pi)}.$$

This a purely local problem, which we investigate in the following subsection.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

2·1. *Local considerations*

We now focus on properties of the local representation π_p. We start by recalling the classification of local representations. But before we do so we need some more notation. Given a (quasi)-character $\chi : \mathbb{Q}_p^\times \to \mathbb{C}^\times$ we write $a(\chi)$ for the (exponent)-conductor. Further write

$$I_0(p) = \begin{bmatrix} \mathbb{Z}_p^\times & \mathbb{Z}_p \\ p\mathbb{Z}_p & \mathbb{Z}_p^\times \end{bmatrix} \subset K_p$$

for an Iwahori subgroup. Let $K_p' = N_{G(\mathbb{Q}_p)}(I_0(p))$ be the normaliser of $I_0(p)$ in $G(\mathbb{Q}_p)$. We also need the filtration

$$K_p'(m) = 1 + \begin{bmatrix} p\mathbb{Z}_p & \mathbb{Z}_p \\ p\mathbb{Z}_p & p\mathbb{Z}_p \end{bmatrix}^m$$

of K_p' by normal subgroups. Finally given two quasi characters $\chi_1, \chi_2 : \mathbb{Q}_p^\times \to \mathbb{C}^\times$ we form the (normalised) induced representation on $\mathrm{Ind}_B^{G(\mathbb{Q}_p)}(\chi_1 \otimes \chi_2)$ as usual. If this representation is irreducible, then we denote the so obtained representation by $\chi_1 \boxplus \chi_2$. We write St for the Steinberg representation which we may identify with the unique irreducible subspace of $\mathrm{Ind}_B^{G(\mathbb{Q}_p)}(|\cdot|^{\frac{1}{2}} \otimes |\cdot|^{-\frac{1}{2}})$. We are now ready to recite the following well-known classification.

LEMMA 2·1. *The representation π_p falls into one of the following three cases*:

(i) Case 1 (Principal series): there are (quasi)-characters $\chi_i : \mathbb{Q}_p^\times \to \mathbb{C}^\times$ such that $\chi_1\chi_2 = \omega_{\pi_p}$, $a(\chi_1) + a(\chi_2) = n_\pi$ and $\pi_p = \chi_1 \boxplus \chi_2$.

(ii) Case 2 (special): there is a (quasi)-character $\chi : \mathbb{Q}_p^\times \to \mathbb{C}^\times$ with $n_\pi = 2a(\chi)$ if $a(\chi) > 0$ or $n_\pi = 1$ otherwise, $\chi^2 = \omega_{\pi_p}$ and $\pi_p = \chi \otimes \mathrm{St}$.

(iii) Case 3 (Supercuspidal): the representation π_p is supercuspidal. In this case we can write $\pi_p = \chi \cdot \pi_p'$ for a (quasi)-character $\chi : \mathbb{Q}_p^\times \to \mathbb{C}^\times$ and some twist-minimal representation π_p' of conductor n_π' which is constructed in one of the following two ways:

(a) Case 3.1 (n_π' even): there is an irreducible representation τ of $Z \cdot K_p$ with $\tau|_Z = \chi^{-2} \cdot \omega_{\pi_p}$ which factors through $K_p(n_\pi'/2)$ so that $\pi_p' = c - \mathrm{Ind}_{ZK_p}^{G(\mathbb{Q}_p)} \tau$.

(b) Case 3.2 (n_π' odd): there is an irreducible representation τ of K_p' which is invariant by $K_p'(n_\pi' - 1)$ with $\tau|_Z = \chi^{-2} \cdot \omega_{\pi_p}$ such that $\pi_p' = c - \mathrm{Ind}_{K_p'}^{G(\mathbb{Q}_p)} \tau$.

With this classification at hand we continue to study the subspaces V in more detail.

LEMMA 2·2. *Suppose π_p is twist minimal and let m_π be as in (2), then $V = \pi_p^{K(m_\pi)}$ is irreducible as K_p-module and we have*:

(i) the invariant m_π is given by

$$m_\pi = \begin{cases} n_\pi & \text{if } \pi_p \text{ is Case~1}, \\ \lfloor \frac{n_\pi + 1}{2} \rfloor & \text{if } \pi_p \text{ is Case~2,3}; \end{cases}$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

(ii) the dimension of V is given by

$$d = p^{m_\pi} \cdot \begin{cases} (1+\frac{1}{p}) & \text{if } \pi_p \text{ is Case~1,} \\ 1 & \text{if } \pi_p \text{ is Case~2,} \\ (1-\frac{1}{p}) & \text{if } \pi_p \text{ is Case~3·1 and} \\ (1-\frac{1}{p^2}) & \text{if } \pi_p \text{ is Case~3·2;} \end{cases}$$

Proof. This is not new and we only have to ensemble the pieces appropriately. Let us proceed case by case.

First, if π_p is in Case 1, then twist-minimality implies that χ_2 (or similarly χ_1) is unramified. Thus we have $n_\pi = a(\chi_1)$ and the results on d and m_π follow from [**20**, proposition 4.3]. Irreducibility can be seen by direct computation.

Second, if π_p is in Case 2 and twist minimal, then $\pi_p = \mathrm{St}$ and $n_\pi = 1$. The results on d and m_π follow again from [**20**, proposition 4.3]. In this case irreducibility follows from [**8**, theorem 1].

Finally, if π_p belongs to Case 3, then the full statement is given in [**17**, theorem 3·5]. (See also [**20**, lemma 4·5, corollary 4·7] for the computation of m_π and d.)

2·2. *A generating domain*

We now switch to the global picture again and aim to produce a suitable set $\mathcal{F} \subset G(\mathbb{A})$ which reduces our problem to studying

$$\mathcal{S}(\Phi, \mathcal{F}) = \sup_{g \in \mathcal{F}} |\Phi(g)|.$$

Let $\mathcal{F}$ be the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, which we identify with a subset of $\mathrm{GL}_2(\mathbb{R})$ by identifying $z = x + iy \in \mathbb{H}$ with $n(x)a(y) \in B(\mathbb{R}) \subset \mathrm{GL}_2(\mathbb{R})$. Here

$$n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \text{ and } a(y) = \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}.$$

We further view $\mathcal{F}$ as a subset of $G(\mathbb{A})$ by identifying it with its image under the usual embedding $G(\mathbb{R}) \to G(\mathbb{A})$. The same series of identifications allows us to write $\Phi(z)$ for $z \in \mathbb{H}$.

LEMMA 2·3. *We have*

$$\|\Phi\|_\infty = \mathcal{S}(\Phi, \mathcal{F}).$$

Proof. First we take $g \in G(\mathbb{A})$ and observe that by strong approximation we can write

$$g = \gamma z b k \text{ with } \gamma \in G(\mathbb{Q}),\ z \in Z(\mathbb{R}),\ b \in \mathcal{F} \text{ and } k \in K.$$

We directly obtain $|\phi_i(g)| = |\phi_i(bk)|$ by automorphy and the action of Z via a unitary character. However, we now observe that if $\phi_1, \ldots, \phi_d$ forms an orthonormal basis of V, then so does $\pi(k)\phi_1, \ldots, \pi(k)\phi_d$. Let us write $\Phi^{(k)}$ for the average constructed from the latter basis. Recall that Φ was independent of the choice of the underlying orthonormal basis, so that we have $\Phi^{(k)} = \Phi$. We conclude that

$$\Phi(g) \leq \mathcal{S}(\Phi^{(k)}, \mathcal{F}) = \mathcal{S}(\Phi, \mathcal{F}).$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

3. *The Whittaker bound*

We will now start the process of deriving a first bound for Φ which will be valid (high) up in the cusp. This is done by estimating Φ using the Whittaker expansions of the ϕ_i's. Throughout we will be working with an arbitrary orthogonal basis $\phi_1, \ldots, \phi_d$ and consider only $g \in \mathcal{F}$.

3·1. *Reduction to a local problem*

Let $\phi = \phi_i$ for some $i = 1, \ldots, d$. The global Whittaker period is given by

$$W_\phi(g) = \int_{\mathbb{Q}\backslash\mathbb{A}} \phi(n(x)g)\psi_{\mathbb{A}}(x)^{-1}dx,$$

where $\psi_{\mathbb{A}}$ is the standard character of $\mathbb{Q}\backslash\mathbb{A}$ which has a factorisation $\psi_{\mathbb{A}} = \bigotimes_v \psi_v$ for $\psi_\infty(x_\infty) = e(x_\infty)$ and ψ_l unramified for all primes l. Note that $W_\phi(\cdot)$ is right $K(p^{m_\pi})$-invariant and transforms with respect to $\psi_{\mathbb{A}}$ when acted on by $N(\mathbb{A})$ from the left. Thus a standard trick shows that $W_\phi(a(q)g_\infty) = 0$ unless $0 \neq q \in \frac{1}{p^{m_\pi}}\mathbb{Z}$. Indeed, for any x with $n(x) \in K(p^{m_\pi})$, one computes

$$W(a(q)g_\infty) = W(a(q)g_\infty n(x)) = W(n(xq)a(q)g_\infty) = \psi_{\mathbb{A}}(xq)W(a(q)g_\infty).$$

we conclude that, if $W(a(q)g_\infty) \neq 0$, then we have $\psi_{\mathbb{A}}(xq) = 1$ for all such x. This gives precisely the condition $q \in 1/p^{m_\pi}\mathbb{Z}$.

This observation leads to the Whittaker expansion

$$\phi(g_\infty) = \sum_{n\in\mathbb{Z}\backslash\{0\}} W_\phi\left(a(\frac{n}{p^{m_\pi}})g_\infty\right).$$

We need to exploit the factorisation of the Whittaker function W_ϕ. To do so we first observe that we have the factorisation of Whittaker models

$$\mathcal{W}(\pi, \psi_{\mathbb{A}}) = \bigotimes_v \mathcal{W}(\pi_v, \psi_v).$$

Using the factorisation of ϕ will now determine distinguished elements in the local Whittaker models as follows. Starting at $v = \infty$ we set

$$W_v(n(x)a(y)) = \frac{|y|^{\frac{1}{2}}K_{it_\pi}(2\pi|y|)}{2|\Gamma(\frac{1}{2}+it_\pi)\Gamma(\frac{1}{2}-it_\pi)|^{\frac{1}{2}}}e(x),$$

where t_π is the spectral parameter of π_∞. Of course W_v is the spherical Whittaker function and is normalised so that

$$\int_{\mathbb{R}^\times} |W_v(a(y))|^2\frac{dy}{|y|} = 1,$$

where dy is the normal Lebesgue measure.

We turn towards the finite places $v \neq p$ given by some prime $l \neq p$. The spherical Whittaker function in $\mathcal{W}(\pi_v, \psi_v)$ is then given by

$$W_v(a(y)) = |y|_v^{\frac{1}{2}}\lambda_\pi(l^{v_l(y)}).$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Here $\lambda_\pi(n)$ is defined by

$$L^{\{\infty,p\}}(s,\pi)=\prod_{v\neq p,\infty}L_v(s,\pi_v)=\sum_{(n,p)=1}\lambda_\pi(n)n^{-s}$$

in analytic normalisation. We have set things up so that $W_v(1)=1$.

Finally we turn towards $v=p$. Here we write $W_p^{(i)}$ for an element in the image of $\mathbb{C}\phi_p^{(i)}$ in the Whittaker model $\mathcal{W}(\pi_p,\psi_p)$ such that

$$\langle W_p^{(i)},W_p^{(i)}\rangle_{\mathcal{W}(\pi_p,\psi_p)}=\int_{\mathbb{Q}_p^\times}|W_p^{(i)}(a(y))|^2\frac{dy}{|y|}=1,$$

here dy is the Haar measure of $\mathbb{Q}_p$ normalised so that $\mathrm{Vol}\,(\mathbb{Z}_p,dy)=1$.

With these choices made there are constants $C_\pi^{(i)}\in\mathbb{C}^\times$ so that

$$\frac{W_{\phi_i}(g)}{\|\phi_i\|_2}=C_\pi^{(i)}\cdot\prod_v W_v(g_v).$$

As shown in [19, (4·16)] (see also [16, section 4]) the absolute values of these constants satisfy

$$|C_\pi^{(i)}|^2=\lim_{s\to1}\frac{\zeta^{\{p,\infty\}}(1)\zeta^{\{p,\infty\}}(2)}{L^{\{p,\infty\}}(s,\pi\otimes\breve{\pi})}.$$

Note that we choose the global measure on $Z(\mathbb{A})G(\mathbb{Q})\backslash G(\mathbb{A})$ to be the Tamagawa measure. In particular, the absolute value is independent of i and using [12] we get

$$|C_\pi^{(i)}|^2\ll_\epsilon p^{\epsilon n_\pi}\cdot(1+t_\pi)^\epsilon.$$

Combining everything we end up with

$$\frac{\phi_i(n(x)a(y))}{\|\phi_i\|_2}=C_\pi^{(i)}\sum_{k\in\mathbb{N}_0}\sum_{\substack{0\neq n\in\mathbb{Z},\\(n,p)=1}}\mathrm{sgn}\,(n)^\rho\frac{\lambda_\pi\,(n)}{\sqrt{|n|}}W_p^{(i)}(a(np^{k-m_\pi}))\cdot W_\infty\left(a\left(\frac{n}{p^{m_\pi-k}}y\right)\right)e\left(\frac{n}{p^{m_\pi-k}}x\right),$$

for $x\in\mathbb{R}$ and $y\in\mathbb{R}^+$. Here $\rho\in\{0,1\}$ depends on whether $\phi_1,\ldots,\phi_d$ are even or odd.

Let $v_1,\ldots,v_d$ be an orthogonal basis of $\pi_p^{K_p(m_\pi)}$. We fix a Whittaker functional and thus an embedding

$$v\longmapsto W_v\in\mathcal{W}(\pi_p,\psi_p).$$

Define

$$\mathcal{S}_{\pi_p}(g_p)=\sum_{i=1}^d\frac{|W_{v_i}(g_p)|^2}{\langle W_{v_i},W_{v_i}\rangle_{\mathcal{W}(\pi_p,\psi_p)}}.$$

Note that $\mathcal{S}_{\pi_p}$ is well defined as it is independent of the choice of Whittaker functional and the choice of basis $v_1,\ldots,v_d$.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

LEMMA 3·1. *For any orthonormal basis $\phi_1, \ldots, \phi_d$ we have*

$$\Phi(g) \leq (dT)^{\epsilon} \sum_{k\in\mathbb{N}_0} \sum_{\substack{0\neq n\in\mathbb{Z},\\ (n,p)=1}} \frac{|\lambda_\pi(n)|}{\sqrt{|n|}} \cdot \left| W_\infty\left(\frac{n}{p^{m_\pi-k}}y\right)\right| \cdot \mathcal{S}_{\pi_p}(a(np^{k-m_\pi}))^{\frac{1}{2}},$$

where $g = n(x)a(y) \in \mathcal{F}$.

Proof. To simplify notation we define

$$a(t) = \operatorname{sgn}(n)^{\rho} \frac{\lambda_\pi(n)}{\sqrt{|n|}} W_\infty\left(\frac{n}{p^{m_\pi-k}}y\right) e\left(\frac{n}{p^{m_\pi-k}}x\right) \text{ and } b_i(t) = W_p^{(i)}(a(np^{k-m_\pi}))$$

if $t = np^{k-m_\pi}$ for $k \in \mathbb{N}_0$ and $(n,p) = 1$, and $a(t) = 0 = b_i(t)$ otherwise. The Whittaker expansion now neatly reads

$$\phi_i(n(x)a(y)) = C_\pi^{(i)} \sum_{t\in\mathbb{Q}^\times} a(t)b_i(t).$$

With this at hand we estimate

$$\begin{aligned}
\Phi(g) &= \left(\sum_{i=1}^{d} \left| C_\pi^{(i)} \sum_{t\in\mathbb{Q}^\times} a(t)b_i(t)\right|^2\right)^{\frac{1}{2}} \\
&\leq \max_i |C_\pi^{(i)}| \cdot \left(\sum_{t_1\in\mathbb{Q}^\times}\sum_{t_2\in\mathbb{Q}^\times} a(t_1)\overline{a(t_2)} \sum_{i=1}^{d} b_i(t_1)\overline{b_i(t_2)}\right)^{\frac{1}{2}} \\
&\ll (dT)^{\epsilon} \left(\sum_{t_1\in\mathbb{Q}^\times}\sum_{t_2\in\mathbb{Q}^\times} |a(t_1)a(t_2)| \left(\sum_{i=1}^{d}|b_i(t_1)|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{d}|b_i(t_2)|^2\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \\
&= (dT)^{\epsilon} \sum_{t\in\mathbb{Q}^\times} |a(t)| \left(\sum_{i=1}^{d}|b_i(t)|^2\right)^{\frac{1}{2}}.
\end{aligned}$$

The claim follows by inserting the definitions of $a(t)$ and $b_i(t)$.

Before we can estimate this expression we need to investigate the size of the local average $\mathcal{S}_{\pi_p}(a(y))$. This is the content of the following subsection.

3·2. *Computing the local averages*

The computation of $\mathcal{S}_{\pi_p}(a(y))$ involves a case study and each case will be treated using different techniques. Finally, combining all possible cases, will lead to the bound

$$\mathcal{S}_{\pi_p}(a(p^{-m_\pi}y)) \ll d^{1+\epsilon} \cdot |y|_p. \tag{5}$$

See Lemma 3·3, 3·6 and 3·7 below.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

3·2·1. *The Steinberg representation*

Let $V=\mathrm{Ind}_B^G(|\cdot|^{\frac{1}{2}}\otimes|\cdot|^{-\frac{1}{2}})$. Then we can identify $\pi=\mathrm{St}$ with the unique irreducible generic subspace of V. Let $V^\vee=\mathrm{Ind}_B^G(|\cdot|^{-\frac{1}{2}}\otimes|\cdot|^{\frac{1}{2}})$. This is the dual space of V and the invariant bilinear pairing is given by

$$\langle f,f^\vee\rangle=\int_K f(k)f^\vee(k)dk.$$

Further $\tilde{\pi}=\mathrm{St}$ can be identified as the unique irreducible generic sub-quotient of $V^\vee$.

Next we choose a basis $v_0,\ldots,v_p$ of $V^{K_p(1)}$. (In an analogous way one constructs the dual basis $v_0^\vee,\ldots,v_p^\vee$ in $(V^\vee)^{K_p(1)}$.) This is done as follows: we first construct

$$v_p(g)=\mathrm{Vol}\,(B(\mathbb{Z}_p)K_p(1),dk)^{-\frac{1}{2}}\cdot\begin{cases}|\frac{a}{d}| & \text{if } g=\begin{pmatrix}a&b\\0&d\end{pmatrix}k\in B(\mathbb{Q}_p)K_p(1),\\ 0 & \text{else.}\end{cases}$$

Further $\gamma_i=wn(i)$ for $i=0,\ldots,p-1$. For consistency of the indices we put $\gamma_p=1$ so that we can identify

$$B(\mathbb{Z}_p)\backslash K_p/K_p(1)=\{\gamma_0,\ldots,\gamma_p\}$$

via the Bruhat decomposition of $G(\mathbb{F}_p)$. (Note that $\gamma_0=w$.) Finally define $v_i(g)=v_p(g\cdot\gamma_i^{-1})$. This is the desired basis.

Now there is an (up to scaling) unique ψ_p-Whittaker functional $\Lambda:\ V\to\mathbb{C}$ (resp. a unique ψ_p^{-1}-Whittaker functional $\Lambda^\vee:\ V^\vee\to\mathbb{C}$). As usual we set

$$W_v(g)=\Lambda(g.v)\text{ or }W_{v^\vee}(g)=\Lambda^\vee(g.v^\vee).$$

We will first consider the related average

$$\mathcal{S}_V(y)=\sum_{i=0}^{p}W_{v_i}(a(y))W_{v_i^\vee}(a(y)).$$

Note that also this is independent of the choice of the particular basis $v_0,\ldots,v_p$ as long as one considers the corresponding dual basis of $V^\vee$.

We will write $\int^{st}$ for the stable integral as defined [16, definition 2·1]. By [16, lemma 4·4 and remark 4·6] we get

$$\begin{aligned}W_{v_i}(a(y))W_{v_i^\vee}(a(y))&=\int_{\mathbb{Q}_p}^{st}\langle n(x)a(y).v_i,a(y).v_i^\vee\rangle\psi_p(x)^{-1}dx\\&=|y|_p\int_{\mathbb{Q}_p}^{st}\langle n(x)v_i,v_i^\vee\rangle\psi_p(xy)^{-1}dx.\end{aligned}$$

Knowing the exact shape of the v_i's we can compute these integrals. First, we observe that a simple change of variables yields

$$\begin{aligned}\langle n(x).v_i,v_i^\vee\rangle&=\int_K v_p(kn(x)\gamma_i^{-1})v_p^\vee(k\gamma_i^{-1})dk=\int_K v_p(k\gamma_in(x)\gamma_i^{-1})v_p^\vee(k)dk\\&=\frac{\sharp[B(\mathbb{Z}_p)/K_p(1)\cap B(\mathbb{Z}_p)]^{\frac{1}{2}}}{\mathrm{Vol}\,(K_p(1),dk)^{\frac{1}{2}}}\cdot\int_{K_p(1)}v_p(k\gamma_in(x)\gamma_i^{-1})dk.\end{aligned}$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

The case $i=p$ is somehow special and will be treated later. For now let us assume $0 \leq i < p$. In this case we have

$$\gamma_i n(x)\gamma_i^{-1} = wn(i)n(x)n(-i)w^{-1} = wn(x)w^{-1}.$$

To take advantage of the support of v_p we have to investigate

$$kwn(x)w^{-1} = b\tilde{k} \in B \cdot K_p(1).$$

In view of the Iwahori-factorisation of k we find that $n(x) \in N(\mathbb{Q}_p) \cap K_p(1)$ is necessary for the integral to be non-zero. Thus one gets

$$\langle \pi(n(x))v_i, v_i^{\vee}\rangle = \delta_{n\in N(p\mathbb{Z}_p)}.$$

With this at hand it is easy to compute

$$\int_{\mathbb{Q}_p}^{st} \langle \pi(n(x))v_i, v_i^{\vee}\rangle \psi_p(yx)^{-1}dx = \int_{p\mathbb{Z}_p} \psi(yx)^{-1}dx = p^{-1}\delta_{y\in p^{-1}\mathbb{Z}_p},$$

for $0 \leq i < p$.

We turn towards $i=p$, so that $\gamma_p = 1$. Further we replace y by yp^{-1} and consider $y \in \mathbb{Z}_p$. Recall that every $k \in K_p(1)$ can be written as $k = t_k\overline{n}_k n_k \in B(\mathbb{Z}_p)N(p\mathbb{Z}_p)^t N(p\mathbb{Z}_p)$ by using the Iwahori-factorisation. We obtain

$$\begin{aligned}
&\int_{\mathbb{Q}_p}^{st} \langle \pi_p(n(x))v_p, v_p^{\vee}\rangle \psi_p(yp^{-1}x)^{-1}dx \\
&= \frac{\sharp[B(\mathbb{Z}_p)/K_p(1)\cap B(\mathbb{Z}_p)]^{\frac{1}{2}}}{\mathrm{Vol}\,(K_p(1), dk)^{\frac{1}{2}}} \cdot \int_{\mathbb{Q}_p}^{st}\int_{K_p(1)} v_p(kn(x))dk\psi_p(yp^{-1}x)^{-1}dx \\
&= \frac{\sharp[B(\mathbb{Z}_p)/K_p(1)\cap B(\mathbb{Z}_p)]^{\frac{1}{2}}}{\mathrm{Vol}\,(K_p(1), dk)^{\frac{1}{2}}} \cdot \int_{K_p(1)}\int_{\mathbb{Q}_p}^{st} v_p(\overline{n}_k n(x))\psi_p(yp^{-1}x)^{-1}dxdk.
\end{aligned}$$

Note that the integrand only depends on $\overline{n}_k$. Therefore we start by discussing a suitable measure on $K_p(1)$. Indeed using the Iwahori factorisation we can write

$$\int_{K_p(1)} f(k)dk = \frac{\mathrm{Vol}\,(K_p(1), dk)}{\mathrm{Vol}\,(B(\mathbb{Z}_p)\cap K_p(1), db)} \int_{B(\mathbb{Z}_p)\cap K_p(1)} p\int_{p\mathbb{Z}_p} f(b\overline{n}(pu))dudb.$$

If we write $\tilde{v}_p$ to be the re-normalisation of v_p with $\tilde{v}_p(1)=1$, then we have

$$\begin{aligned}
\int_{\mathbb{Q}_p}^{st} \langle \pi_p(n(x))v_p, v_p^{\vee}\rangle \psi_p(yp^{-1}x)^{-1}dx &= p\int_{\mathbb{Q}_p}^{st}\int_{p\mathbb{Z}_p} \tilde{v}_p(n(z)^t n(x))\psi_p(yp^{-1}x)^{-1}dzdx \\
&= p\int_{\mathbb{Q}_p}^{st}\int_{p\mathbb{Z}_p} \tilde{v}_p\left(\begin{pmatrix} 1 & x-z^{-1} \\ z & zx \end{pmatrix}\right)\psi_p(yp^{-1}(x-z^{-1}))^{-1}dzdx.
\end{aligned}$$

In the last step we simply made a change of variables in the x-integral. A simple matrix computation shows that

$$\begin{pmatrix} 1 & x-z^{-1} \\ z & zx \end{pmatrix} = n(\star)\begin{pmatrix} (zx)^{-1} & 0 \\ z & zx \end{pmatrix}.$$

Inserting this and using the transformation behaviour of $\tilde{v}_p$ one obtains

$$\int_{\mathbb{Q}_p}^{st} \langle \pi_p(n(x))v_p, v_p^\vee\rangle \psi_p(yp^{-1}x)^{-1}dx$$
$$= p\int_{p\mathbb{Z}_p} \psi_p(yz^{-1}p^{-1})\frac{dz}{|z|_p^2}\int_{\mathbb{Q}_p}^{st} \tilde{v}_p\left(\begin{pmatrix} 1 & 0 \\ x^{-1} & 1\end{pmatrix}\right)\psi_p(-xyp^{-1})\frac{dx}{|x|_p^2}.$$

Both integrals can now be computed quite easily. Starting from the first one we obtain

$$p\int_{p\mathbb{Z}_p} \psi_p(yz^{-1}p^{-1})\frac{dz}{|z|_p^2} = \sum_{l=1}^{\infty} p^{l+1}\int_{\mathbb{Z}_p^\times} \psi_p(zyp^{-1-l})dz$$
$$= \sum_{l=1}^{v_p(y)-1} p^{l+1}(1-p^{-1}) - \delta_{v_p(y)\geq 1}p^{v_p(y)} = -p\delta_{v_p(y)\geq 1}.$$

Turning to the other integral we find

$$\int_{\mathbb{Q}_p}^{st} \tilde{v}_p\left(\begin{pmatrix} 1 & 0 \\ x^{-1} & 1\end{pmatrix}\right)\psi_p(-xyp^{-1})\frac{dx}{|x|_p^2} = \int_{\mathbb{Q}_p\setminus\mathbb{Z}_p}^{st} \psi_p(-xyp^{-1})\frac{dx}{|x|_p^2}$$
$$= \sum_{l=1}^{\infty} p^{-l}\int_{\mathbb{Z}_p^\times} \psi_p(-xyp^{-1-l})dx$$
$$= \sum_{l=1}^{v_p(y)-1} p^{-l}(1-p^{-1}) - \delta_{v_p(y)\geq 1}p^{-v_p(y)-1}$$
$$= \delta_{v_p(y)\geq 2}(p^{-1}-p^{-v_p(y)}) - \delta_{v_p(y)\geq 1}p^{-1-v_p(y)}.$$

In particular we have

$$\int_{\mathbb{Q}_p}^{st} \langle \pi_p(n(x))v_0, v_0^\vee\rangle \psi_p(yp^{-1}x)^{-1}dx = \delta_{v_p(y)\geq 2}(p^{1-v_p(y)}-1) + \delta_{v_p(y)\geq 1}p^{-v_p(y)}.$$

Note that this can be negative, but for non-unitary representations there is no expectation for these integrals to be non-negative.

Combining the computations above and swapping back to $y\in p^{-1}\mathbb{Z}_p$ leads us to the following result.

LEMMA 3·2. *In the notation above we have*

$$\mathcal{S}_V(y) = |y|_p\left[\delta_{v_p(y)\geq -1} + \delta_{v_p(y)\geq 0}p^{-1}|y|_p + \delta_{v_p(y)\geq 1}(|y|_p-1)\right].$$

We will obtain the desired estimate by relating $\mathcal{S}_{\mathrm{St}}(a(y))$ to $\mathcal{S}_V$.

LEMMA 3·3. *For $\pi_p = St$ and $y\in\mathbb{Q}_p$ we have*

$$\mathcal{S}_{\pi_p}(a(y)) \ll |y|_p.$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Proof. Recall that the definitions of $\mathcal{S}_{\pi_p}$ and $\mathcal{S}_V$ are independent of the choice of the underlying basis. Thus we can choose an orthogonal basis $w_1, \ldots, w_p$ of $\pi_p^{K_p(1)}$. Viewing π as invariant subspace of V we can assume that the w_i's are in V. We then have

$$\mathcal{S}_\pi(a(y)) = \sum_{i=1}^{p} \frac{|W_{w_i}(a(y))|^2}{\langle W_{w_i}, W_{w_i}\rangle} = \sum_{i=1}^{p} \frac{W_{w_i}(a(y))W_{w_i^\vee}(a(y))}{\langle w_i, w_i^\vee\rangle}.$$

Finally if we choose $w_0^\vee \in (V^\vee)^{K_p(1)}$ in the annihilator of the $w_1, \ldots, w_p$ and let $w_0 \in V^{K_p(1)}$ be the dual element then after renormalising we have

$$\mathcal{S}_V(y) = \mathcal{S}_\pi(a(y)) + W_{w_0}(a(y))W_{w_0^\vee}(a(y)).$$

However, since there is a unique Whittaker functional on $V^\vee$ which descents to the unique Whittaker functional on π when viewed as a sub-quotient we must have $W_{w_0^\vee}(a(y)) = 0$. (Since the unique invariant subspace is non-generic.) Thus $\mathcal{S}_V(y) = \mathcal{S}_{\pi_p}(a(y))$ and the desired estimate follows directly from the previous lemma.

3·2·2. *Twist minimal principal series*

Turning to this case we assume that $\pi_p = \chi \cdot |\cdot|^\rho \boxplus |\cdot|^{-\rho}$ where $a(\chi) = n_\pi > 0$. Without loss of generality we can assume that $\chi(p) = 1$. (If we assume that π is unitary then it is tempered so that $\rho \in i\mathbb{R}$.) Now we can choose a basis in the induced picture essentially as above, but we need to find a suitable decomposition of $B(\mathbb{Z}_p)\backslash K_p/K_p(n_\pi)$ (since the Bruhat decomposition does not hold in $G(\mathbb{Z}_p/p^m\mathbb{Z}_p)$ if $m > 1$). First we start by defining

$$v_0(g) = \mathrm{Vol}\,(B(\mathbb{Z}_p)K_p(m_\pi), dk)^{-\frac{1}{2}} \cdot \begin{cases} \chi(a)|\frac{a}{d}|^{\frac{1}{2}+\rho} & \text{if } g = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} k \in B(\mathbb{Q}_p)K_p(m_\pi), \\ 0 & \text{else.} \end{cases}$$

From this element we can construct a basis of $\pi_p^{K_p(m_\pi)}$ as in the Steinberg case. Indeed, we fix a system of representatives $\{\gamma_j\}$ for $B(\mathbb{Z}_p)\backslash K_p/K_p(m_\pi)$ and set $v_j = \pi_p(\gamma_j^{-1})v_0$.

In order to explicate this basis we need to compute a suitable coset decomposition for $B(\mathbb{Z}_p)\backslash K_p/K_p(m_\pi)$. This is the content of the following lemma.

LEMMA 3·4. *We have*

$$K_p = \bigsqcup_{a\in\mathbb{Z}_p/p^{m_\pi}\mathbb{Z}_p} B(\mathbb{Z}_p)\gamma_{0,a}K_p(m_\pi) \sqcup \bigsqcup_{i=1}^{m_\pi} \bigsqcup_{a\in(\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p)^\times} B(\mathbb{Z}_p)\gamma_{i,a}K_p(m_\pi),$$

for

$$\gamma_{i,a} = \begin{cases} \begin{pmatrix} 0 & -1 \\ 1 & a \end{pmatrix} & \textit{if } i = 0, \\ \begin{pmatrix} 1 & 0 \\ ap^i & 1 \end{pmatrix} & \textit{if } 1 \le i \le m_\pi - 1, \\ 1_2 & \textit{if } i = m_\pi. \end{cases}$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Proof. Take $g=\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in K_p$ and set $i=v_p(c)$. We treat several cases distinguished by the value of i.

First, if $i=m_\pi$, then we have

$$g=\begin{pmatrix} a & b\\ c & d\end{pmatrix}=\begin{pmatrix} a & b\\ 0 & d\end{pmatrix}\begin{pmatrix} 1-\frac{bc}{ad} & 0\\ \frac{c}{d} & 1\end{pmatrix}\in B(\mathbb{Z}_p)K_p(m_\pi).$$

Second, for $1\le i<m_\pi$ we have

$$g=\begin{pmatrix} a-\frac{bc}{d} & b\\ 0 & d\end{pmatrix}\begin{pmatrix} 1 & 0\\ \frac{c}{d} & 1\end{pmatrix}.$$

By right multiplication with elements in $K_p(m_\pi)$ we can view $c/d\in p^i\mathbb{Z}_p^\times/p^{m_\pi}\mathbb{Z}_p$.

The critical contribution is given by the matrices with $i=0$. We can write

$$g=\begin{pmatrix} \frac{ad}{c}-b & a+(b-\frac{ad}{c})p^{m_\pi}\\ 0 & c\end{pmatrix}\begin{pmatrix} 0 & -1\\ 1 & \frac{d}{c}\end{pmatrix}\begin{pmatrix} 1+\frac{d}{c}p^{m_\pi} & \frac{d^2}{c^2}p^{m_\pi}\\ -p^{m_\pi} & 1-\frac{d}{c}p^{m_\pi}\end{pmatrix}.$$

Given $v\in\pi^{K_p(n_\pi)}$ we can compute the Jacquet Integral as follows. Without loss of generality assume $v_p(y)\ge -n_\pi$, since otherwise the Whittaker function W_v vanishes for trivial reasons. We compute

$$\begin{aligned}W_v(a(y)) &= \int_{\mathbb{Q}_p} v(wn(x)a(y))\psi_p(x)^{-1}dx\\ &= |y|_p^{\frac{1}{2}-\rho}\int_{\mathbb{Q}_p} v(wn(x))\psi_p(xy)^{-1}dx\\ &= |y|_p^{\frac{1}{2}-\rho}p^{-n_\pi}\sum_{a\in\mathbb{Z}_p/p^{n_\pi}\mathbb{Z}_p}\psi_p(ay)^{-1}v(wn(a))\\ &\quad + |y|_p^{\frac{1}{2}-\rho}\int_{\mathbb{Q}_p\setminus\mathbb{Z}_p}|x|^{-2\rho}\chi(x)^{-1}\psi(xy)^{-1}v(n(x^{-1})^t)\frac{dx}{|x|_p}.\end{aligned}$$

Note that $wn(a)=\gamma_{0,a}$. Now we will have a closer look at the remaining integral:

$$\begin{aligned}&\int_{\mathbb{Q}_p\setminus\mathbb{Z}_p}|x|^{-2\rho}\chi(x)^{-1}\psi(xy)^{-1}v(n(x^{-1})^t)\frac{dx}{|x|_p}\\ &= \int_{p\mathbb{Z}_p}|x|^{2\rho}\chi(x)\psi(x^{-1}y)^{-1}v(n(x)^t)\frac{dx}{|x|_p}\\ &= v(1)\int_{p^{m_\pi}\mathbb{Z}_p}\chi(x)|x|^{2\rho}\psi(x^{-1}y)^{-1}\frac{dx}{|x|_p}\\ &\quad+\sum_{i=1}^{m_\pi-1}\sum_{b\in(\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p)^\times}p^{-2\rho i}\chi(b)^{-1}v(n(bp^i)^t)\cdot\int_{1+p^{m_\pi-i}\mathbb{Z}_p}\chi(x)^{-1}\psi(-byxp^{-i})dx.\end{aligned}$$

Note that $n(bp^i)^t=\gamma_{i,b}$.

Since the $v(1)$-contribution is easily computed we arrive at the following lemma.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

LEMMA 3·5. *For* $v_p(y)\geq -m_\pi$ *we have*

$$W_v(a(y)) = |y|_p^{\frac{1}{2}-\rho}p^{-m_\pi}\sum_{a\in\mathbb{Z}_p/p^{m_\pi}\mathbb{Z}_p}\psi_p(ay)^{-1}v(\gamma_{0,a})$$
$$+|y|_p^{\frac{1}{2}-\rho}\sum_{i=1}^{m_\pi-1}\sum_{b\in(\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p)^\times}p^{-2\rho i}\chi(b)^{-1}v(\gamma_{i,b})G_{m_\pi-i}(-byp^{-i},\chi^{-1})$$
$$+|y|_p^{\frac{1}{2}-\rho}\cdot v(\gamma_{m_\pi,0})\cdot\int_{p^{m_\pi}\mathbb{Z}_p}\chi(x)|x|^{2\rho}\psi(x^{-1}y)^{-1}\frac{dx}{|x|_p},$$

for

$$G_l(y,\chi)=\int_{1+p^l\mathbb{Z}_p}\chi(x)\psi(yx)dx,$$

This supplies us with the necessary ingredients to show the required estimate for $\mathcal{S}_{\pi_p}$.

LEMMA 3·6. *For* $\pi_p=\chi|\cdot|^\rho\boxplus|\cdot|^{-\rho}$ *unitary and* $y\in\mathbb{Q}_p$ *we have*

$$\mathcal{S}_{\pi_p}(a(p^{-m_\pi}y))\ll d^{1+\epsilon}|y|_p.$$

Proof. Note that since all v_j's are translates of v_0 their Whittaker-norm all coincides. So it suffices to compute one of these norms and it is easy to see that

$$\langle W_{v_0},W_{v_0}\rangle=\int_{\mathbb{Q}_p^\times}|W_{v_0}(a(y))|^2\frac{dy}{|y|}=(1+p^{-1}).$$

Next we observe that one can choose representatives so that $v_j=\pi_p(\gamma_{i,a}^{-1})v_0$ for some $i=i(j)$ and $a=a(j)$. In particular, we can sort the terms of the sum $\mathcal{S}_{\pi_p}(a(y))$ according to this i. We get

$$\mathcal{S}_{\pi_p}(a(y))=\sum_{i=0}^{m_\pi}\mathcal{S}_i(y)\text{ for }\mathcal{S}_i(y)=(1+p^{-1})^{-1}\sum_{a\in\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p}|W_{\pi_p(\gamma_{i,a}^{-1})v_0}(a(y))|^2.$$

Applying the previous Lemma with $v=\pi_p(\gamma_{i,a}^{-1})v_0$ and taking support properties of v_0 into account provides us with nice formulae for the $W_{\pi_p(\gamma_{i,a}^{-1})v_0}(a(y))$.

As soon as we can show that $\mathcal{S}_i(y)\ll|y|_p$ for all i we are done. We start with $i=0$. Here we have the explicit formula

$$\mathcal{S}_i(y)=(1+p^{-1})^{-1}\sum_{a\in\mathbb{Z}_p/p^{m_\pi}\mathbb{Z}_p}p^{-2m_\pi}|y|_pv_0(1)^2=\frac{(1+p^{-1})^{-1}p^{-m_\pi}|y|_p}{\mathrm{Vol}\,(B(\mathbb{Z}_p)K_p(m_\pi),dk)}=|y|_p$$

for $y\in p^{-m_\pi}\mathbb{Z}_p$.

We turn towards $1\leq i\leq m_\pi-1$. In this range we get

$$\mathcal{S}_i(y)=(1+p^{-1})^{-1}v_0(1)^2|y|_p\sum_{a\in(\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p)^\times}|G_{m_\pi-i}(-ayp^{-i},\chi^{-1})|^2.$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Thus we need to bound the integrals $G_l(z,\chi)$, which are somehow incomplete Gauß sums in the sense that one sums only over a specific congruence class. These sums were essentially computed in the proof of [2, lemma 5·8]. Indeed one extracts

$$G_l(zp^{-k},\chi)=\begin{cases}\epsilon(\frac{1}{2},\chi^{-1})\chi^{-1}(z)p^{-\frac{k}{2}} & \text{if } l\le\lfloor\frac{a(\chi)}{2}\rfloor, k=a(\chi) \text{ and } z\in b(\chi)+p^l\mathbb{Z}_p,\\ \psi_p(zp^k)p^{-l} & \text{if } l\ge\lceil\frac{a(\chi)}{2}\rceil, k=a(\chi) \text{ and } z\in -b(\chi)+p^{a(\chi)-l}\mathbb{Z}_p,\\ 0 & \text{else,}\end{cases}$$

for $z\in\mathbb{Z}_p^\times$, $k\in\mathbb{Z}$ and $b(\chi)\in\mathbb{Z}_p^\times$ is determined by χ. With this at hand we can easily evaluate $\mathcal{S}_i$. For $i\le\lfloor m_\pi/2\rfloor$ we have

$$\begin{aligned}\mathcal{S}_i(y)&=\delta_{v_p(y)=i-m_\pi}(1+p^{-1})^{-1}v_0(1)^2|y|_p\sum_{b\in(\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p)^\times}p^{2i-2m_\pi}\delta_{-byp^{-v_p(y)}\in-b(\chi)+p^i\mathbb{Z}_p}\\ &=\delta_{v_p(y)=i-m_\pi}|y|_p.\end{aligned}$$

Similarly for $i\ge\lceil m_\pi/2\rceil$ we have

$$\begin{aligned}\mathcal{S}_i(y)&=\delta_{v_p(y)=i-m_\pi}(1+p^{-1})^{-1}v_0(1)^2|y|_p\sum_{b\in(\mathbb{Z}_p/p^{m_\pi-i}\mathbb{Z}_p)^\times}p^{-m_\pi}\cdot\delta_{-byp^{-v_p(y)}\in b(\chi)+p^{m_\pi-i}\mathbb{Z}_p}\\ &=\delta_{v_p(y)=i-m_\pi}|y|_p.\end{aligned}$$

Finally consider $i=m_\pi$. We have

$$\begin{aligned}\mathcal{S}_{m_\pi}(y)&=(1+p^{-1})v_0(1)^2|y|_p\left|\int_{p^{m_\pi}\mathbb{Z}}\chi(x)|x|^{2\rho}\psi(x^{-1}y)^{-1}\frac{dx}{|x|_p}\right|^2\\ &=p^{m_\pi}|y|_p\cdot\left|\int_{p^{m_\pi}\mathbb{Z}_p}\chi(x)|x|^{2\rho}\psi(x^{-1}y)^{-1}\frac{dx}{|x|_p}\right|^2.\end{aligned}$$

Therefore it suffices to compute the remaining integral. By some basic Gauß sum evaluations one gets

$$\int_{p^{m_\pi}\mathbb{Z}_p}\chi(x)|x|^{2\rho}\psi(x^{-1}y)^{-1}\frac{dx}{|x|_p}=\delta_{y\in\mathbb{Z}_p}\epsilon(\frac{1}{2},\chi)\chi(y)p^{-\frac{m_\pi}{2}-2\rho[v_p(y)+m_\pi]}.$$

Inserting this above concludes the proof since it implies $\mathcal{S}_{m_\pi}(y)=\delta_{v_p(y)>0}|y|_p$.

3·2·3. *Supercuspidal representations*

Let $\mathfrak{X}_k$ be the set of character $\chi:\mathbb{Q}_p^\times\to\mathbb{C}^\times$ with $a(\chi)\le k$ and $\chi(p)=1$. Note that

$$\sharp\mathfrak{X}_k=p^{k-1}(p-1).$$

For $\chi\in\mathfrak{X}_k$ and $m\in\mathbb{Z}$ we will consider the functions $\xi_\chi^{(m)}\in\mathcal{C}_c^\infty(\mathbb{Q}_p^\times)$ given by

$$\xi_\chi^{(m)}(y)=\mathbb{1}_{p^{-m}\mathbb{Z}_p^\times}(y)\chi(y).$$

Given any representation π_p we write $\mathcal{K}_{\psi_p}(\pi_p)$ for the corresponding ψ_p-Kirillov model. Note that this model contains the Schwartz functions so that we have $\xi_\chi^{(m)}\in\mathcal{K}_{\psi_p}(\pi_p)$. Note that by construction of the Kirillov model we have

$$W_f(a(y))=f(y)\text{ for } f\in\mathcal{K}_{\psi_p}(\pi_p)\text{ and } y\in\mathbb{Q}_p^\times.$$

Thus we compute

$$\langle W_{\xi^{(m_1)}_{\chi_1}}, W_{\xi^{(m_2)}_{\chi_2}}\rangle = \int_{\mathbb{Q}_p^\times} \xi^{(m_1)}_{\chi_1}(y)\overline{\xi^{(m_2)}_{\chi_2}(y)}d^\times y = \delta_{m_1=m_2}\int_{p^{-m_1}\mathbb{Z}_p^\times} \chi_1(y)\chi_2^{-1}(y)d^\times y$$

$$= \delta_{\substack{m_1=m_2,\\ \chi_1=\chi_2}}.$$

This suffices to compute $S_{\pi_p}(a(y))$ for supercuspdial representations π_p.

LEMMA 3·7. *Suppose π_p is a twist minimal supercuspidal representation with (exponent)-conductor n_π. Then the following is true:*

(i) If $n_\pi = 2m_\pi$, then

$$S_{\pi_p}(a(y)) = \frac{p^{m_\pi}}{\zeta_p(1)} \cdot \delta_{y\in p^{-m_\pi}\mathbb{Z}_p^\times};$$

(ii) If $n_\pi = 2m_\pi - 1$, then

$$S_{\pi_p}(a(y)) = \frac{p^{m_\pi}}{\zeta_p(1)} \cdot \delta_{y\in p^{-m_\pi}\mathbb{Z}_p^\times} + \frac{p^{m_\pi-1}}{\zeta_p(1)} \cdot \delta_{y\in p^{1-m_\pi}\mathbb{Z}_p^\times};$$

In general we have the bound

$$S_{\pi_p}(a(p^{-m_\pi}y)) \ll d \cdot |y|_p \cdot \delta_{y\in\mathbb{Z}_p}.$$

Proof. We start with the case $n_\pi = 2m_\pi$. By [20, lemma 4·4] we find that a basis for $\pi^{K_p(m_\pi)}$ in the Kirillov model is given by

$$\{\xi^{(m_\pi)}_\chi : \chi \in \mathfrak{X}_{m_\pi}\}.$$

Note that we already took advantage of twist-minimality using that $n_{\chi\pi} = n_\pi$ for all $\chi \in \mathfrak{X}_{m_\pi}$. Our computations above show that this basis is orthonormal (with respect to the Whittaker inner product). Thus we have

$$\mathcal{S}_{\pi_p}(y) = \sum_{\chi\in\mathfrak{X}_{m_\pi}} |\xi^{(m_\pi)}_\chi(y)|^2 = \delta_{y\in p^{-m_\pi}\mathbb{Z}_p^\times} \cdot \sharp\mathfrak{X}_{m_\pi}.$$

We turn towards the second case where n_π is odd. Then we get the orthonormal basis

$$\{\xi^{m_\pi}_\chi : \chi \in \mathfrak{X}_{m_\pi}\} \cup \{\xi^{m_\pi-1}_\chi : \chi \in \mathfrak{X}_{m_\pi-1}\}.$$

It is again easy to compute the desired quantity:

$$\mathcal{S}_{\pi_p}(y) = \delta_{y\in p^{-m_\pi}\mathbb{Z}_p^\times} \cdot \sharp\mathfrak{X}_{m_\pi} + \delta_{y\in p^{1-m_\pi}\mathbb{Z}_p^\times} \cdot \sharp\mathfrak{X}_{m_\pi-1}.$$

The result follows directly.

3·3. *Conclusion*

We can now give a decent bound for $\Phi(z)$ using the Whittaker expansion. We will use the bound (5) and follow the standard procedure.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

LEMMA 3·8. *We have*

$$\Phi(z) \ll \left(\frac{dT}{y}\right)^{\epsilon}\left(d^{\frac{1}{2}}T^{\frac{1}{6}} + \frac{dT^{\frac{1}{2}}}{y^{\frac{1}{2}}}\right).$$

Proof. Inserting (5) into Lemma 3·1 yields

$$\Phi(g) \ll d^{\frac{1}{2}+\epsilon}T^{\epsilon}\sum_{0\neq n\in\mathbb{Z}}\frac{|\lambda_{\pi}\ (n/(n,p^{\infty}))|}{\sqrt{|n|}}|W_{\infty}\left(n\frac{y}{p^{m_{\pi}}}\right)|.$$

Estimating the remaining n-sum as for example in [23] or [21] yields the desired result.

4. *A bound via the pre-trace formula*

The next bound will be derived from the pre-trace inequality. We start by discussing the local test functions. At the archimedean place we closely follow [21, section 3.5] and fix f_{∞} so that it satisfies:

(1) $f_{\infty}(g) = 0$ unless $g \in G(\mathbb{R})^{+}$ and $u(g) \leq 1$;

(2) $\hat{f}_{\infty}(\sigma) > 0$ for all irreducible spherical unitary principal series representations σ of $G(\mathbb{R})$;

(3) $\hat{f}(\pi_{\infty}) \gg 1$;

(4) $|f_{\infty}(g)| \leq T$ and if $u(g) \geq T^{-2}$, then $|f_{\infty}(g)| \leq T^{\frac{1}{2}}u(g)^{-\frac{1}{4}}$.

(The final property is not really necessary because we are ignoring the spectral aspect for now.) Note that $\hat{f}$ is the spherical transform (also Selberg/Harish–Chandra transform) of f and $u(g)$ is the point-pair invariant on group level.

At the place $v = p$ we define multiple test functions:

$$f_p^{(i)}(g) = \mathbb{1}_{ZK_p}(g)\frac{\overline{\langle\pi(g)\phi_p^{(i)}, \phi_p^{(i)}\rangle_{\pi_p}}}{\langle\phi_p^{(i)}, \phi_p^{(i)}\rangle_{\pi_p}}.$$

LEMMA 4·1. *For every irreducible admissible unitary representation σ of $G(\mathbb{Q}_p)$ the operator $\sigma(f_p^{(i)})$ is non-negative and self-adjoint. Further we have $\pi_p(f_p^{(i)})\phi_p^{(i)} =$* $\dim_{\mathbb{C}}(\pi_p^{K_p(m_{\pi})})^{-1}\cdot\phi_p^{(i)}$.

Proof. The proof is standard and relies on Schur's orthogonality relations for irreducible representations of K_p. To apply this it will be important to keep in mind that $\pi^{K_p(m_{\pi})}$ is irreducible.

The operators are self-adjoint since $f_p^{(i)}(g^{-1}) = \overline{f_p^{(i)}(g)}$. To see non-negativity we will show the convolution identity

$$f_p^{(i)} = \dim_{\mathbb{C}}\pi^{K_{p,1}(m_{\pi})}\cdot(f_p^{(i)} \star f_p^{(i)}).$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Indeed we compute

$$
\begin{aligned}
[f_p^{(i)}\star f_p^{(i)}](h) &= \int_{Z\backslash G} f_p^{(i)}(g^{-1})f_p^{(i)}(gh)dg \\
&= \frac{\mathbb{1}_{ZK}(h)}{\langle \phi_p^{(i)},\phi_p^{(i)}\rangle_{\pi_p}^{\ 2}} \int_{K_p} \langle \pi_p(g)\phi_p^{(i)},\phi_p^{(i)}\rangle_{\pi_p} \overline{\langle \pi_p(gh)\phi_p^{(i)},\phi_p^{(i)}\rangle_{\pi_p}} dk \\
&= \dim_{\mathbb{C}}(\pi^{K_p(m_\pi)})^{-1} \frac{\mathbb{1}_{ZK}(h)}{\langle \phi_p^{(i)},\phi_p^{(i)}\rangle_{\pi_p}} \overline{\langle \pi_p(h)\phi_p^{(i)},\phi_p^{(i)}\rangle_{\pi_p}}
\end{aligned}
$$

and the claimed identity follows directly.

It remains to show the final claim. First observe that the image of $\pi_p(f_p^{(i)})\phi_p^{(i)}$ is obviously $K_p(m_\pi)$-invariant. Thus it suffices to show that

$$
\langle \pi_p(f_p^{(i)})\phi_p^{(i)}, w\rangle_{\pi_p} = \dim_{\mathbb{C}}(\pi_p^{K_p(m_\pi)})^{-1}\cdot \langle \phi_p^{(i)}, w\rangle_{\pi_p}
$$

for any $w\in\pi_p^{K_p(m_\pi)}$ But this follows again from the orthogonality relations since $\pi_p^{K_p(m_\pi)}$ is irreducible (as K_p-module) and

$$
\langle \pi_p(f_p^{(i)})\phi_p^{(i)}, w\rangle_{\pi_p} = \langle \int_{Z\backslash G} f_p^{(i)}(g)\pi_p(g)\phi_p^{(i)}dg, w\rangle = \int_{Z\backslash G} f_p^{(i)}(g)\langle \pi_p(g)\phi_p^{(i)}, w\rangle_{\pi_p}dg.
$$

Finally we define the unramified part of the test function f_{ur} by setting

$$
f_{ur} = \left(\sum_{l\in S} c_l\kappa_l\right)\star\left(\sum_{l\in S} c_l\kappa_l\right)^* + \left(\sum_{l\in S} c_{l^2}\kappa_{l^2}\right)\star\left(\sum_{l\in S} c_{l^2}\kappa_{l^2}\right)^*,
$$

$$
\text{for } c_r = \begin{cases} \frac{|\lambda_\pi(r)|}{\lambda_\pi(r)} & \text{if } r=l \text{ or } r=l^2 \text{ for } l\in S,\\ 0 & \text{else,}\end{cases}
$$

for a set of primes S (to be determined) and normalised rth Hecke-operators κ_r. This implements the usual amplification procedure. Finally we define the global test functions

$$
f^{(i)} = f_\infty\otimes f_p^{(i)}\otimes f_{ur} \text{ and } f = \sum_i f^{(i)}.
$$

We introduce

$$
M(l,g) = \{A\in M_2(\mathbb{Z}):\ \det(A) = l,\ A\equiv g \bmod p^{m_\pi}\} \text{ for } g\in \mathrm{GL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z}).
$$

Further let σ denote the irreducible representation of $\mathrm{GL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})$ through which the irreducible K_p-module $\pi^{K_p(m_\pi)}$ factors. This is a representation of a finite group and we write χ_σ for its character. Finally we define the coefficients y_r by linearising the convolutions of Hecke-operators in the definition of f_{ur}. More precisely we write

$$
f_{ur} = \sum_r y_r\kappa_r.
$$

This can be compared to the analogous expression in [21, section 7].

The following pre-trace inequality provides the transition to the counting problem.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

LEMMA 4·2. *For $z \in \mathcal{F}$ we have*

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}} \pi_p^{K_p(m_\pi)}} \cdot \Phi(z)^2 \ll \sum_r \frac{|y_r|}{\sqrt{r}} \sum_{g \in \mathrm{GL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})} |\chi_\sigma(g)| \sum_{A \in M(l,g)} |f_\infty(u(Az, z))|.$$

Proof. We start by considering the spectral expansion of the automorphic kernel $k_{f^{(i)}}$ associated to the self-adjoint operators $R(f^{(i)})$ and dropping all terms except ϕ_i. The latter is possible by positivity. We obtain

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}} \pi_p^{K_p(m_\pi)}} \cdot \phi_i(g_\infty)^2 \leq k_{f^{(i)}}(g, g) = \sum_r y_r \sum_{\gamma \in Z(\mathbb{Q}) \backslash G(\mathbb{Q})} f_p^{(i)}(\gamma)\kappa_r(\gamma) f_\infty(g_\infty^{-1}\gamma g_\infty).$$

We now sum this inequality over i to obtain

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}} \pi_p^{K_p(m_\pi)}} \Phi(g_\infty)^2 \leq \sum_r y_r \sum_{\gamma \in Z(\mathbb{Q}) \backslash G(\mathbb{Q})} \left(\sum_{i=1}^d f_p^{(i)}(\gamma) \right) \kappa_r(\gamma) f_\infty(g_\infty^{-1}\gamma g_\infty).$$

Recall that the test functions $f_p^{(i)}$ are supported in ZK_p. Thus, after choosing a suitable representative for γ modulo $Z(\mathbb{Q})$, we can assume that $\gamma \in K_p$ and write $\overline{\gamma}$ for the image of γ in $\mathrm{GL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})$. For such γ we get

$$\sum_{i=1}^d f_p^{(i)}(\gamma) = \chi_\sigma(\overline{\gamma}).$$

The rest of the argument is standard and can for example be found in [21].

By the choice of f_∞ we can already eliminate the archimedean influence from the right-hand side. (Note that we are not aiming to amplify in the T-aspect.)

COROLLARY 4·3. *For $z \in \mathcal{F}$ we have*

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}} \pi_p^{K_p(m_\pi)}} \cdot \Phi(z)^2 \ll T \sum_{g \in \mathrm{GL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})} |\chi_\sigma(g)| \sum_r \frac{|y_r|}{\sqrt{r}} \cdot \sharp M_z(r, g),$$

for

$$M_z(r, g) = \{A \in M_2(\mathbb{Z})\colon \det(A) = r,\ A \equiv g \bmod p^{m_\pi} \text{ and } u(Az, z) \leq 1\}.$$

This last corollary tells us that we need to control the character χ_σ and solve a counting problem estimating $M_z(r, g)$.

To estimate the character we need to define certain level sets.

$$K_{m,\lambda} = \{g \in G(\mathbb{Z}/p^m\mathbb{Z})\colon g \equiv 1 \bmod p^\lambda\} \text{ for } 0 \leq \lambda \leq m.$$

Note that $K_{m,m} = \{1\}$ and $K_{m,0} = G(\mathbb{Z}/p^m\mathbb{Z})$.

Before we continue let us make a remark concerning notation. We will write $A \neg B = A \cap B^c$ for the set-theoretic difference. We are not using $A \setminus B$, since it may be confused with the quotient when A and B are groups.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

LEMMA 4·4. *Suppose* $p > 3$. *Let* π *belong to Case* 1, 2 *or* 3·1. *Then, for* $0 \leq \lambda < m_\pi$ *and* $g \in Z \cdot K_{m_\pi,\lambda} \neg Z \cdot K_{m_\pi,\lambda+1}$ *we have*

$$\chi_\sigma(g) \ll p^\lambda. \tag{6}$$

If π *belongs to Case* 3·2 *and* λ *and* g *are as above, then we have the slightly weaker bound*

$$\chi_\sigma(g) \ll p^{\frac{m_\pi+\lambda}{2}}.$$

Furthermore, let $h \in G(\mathbb{Z}/p^m\mathbb{Z})$ *be a diagonal matrix such that* $G(\mathbb{Z}/p^m\mathbb{Z}) = Z \cdot \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z}) \sqcup hZ \cdot \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$. *Then the same estimates hold for* $hZ \cdot K_{m_\pi,\lambda} \neg hZ \cdot K_{m_\pi,\lambda+1}$, *where* $\det(h)$ *is not a square modulo* p^{m_π}.

The representations of GL_2 over finite rings such as $\mathbb{Z}/p^m\mathbb{Z}$ and their characters are well studied but explicit estimates for the characters as needed here seem to be hard to find. We choose to use the character tables for $\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$ computed by Kutzko [15]. This makes it necessary to pass from SL_2 to GL_2 using Mackey Theory. Note that the character values in question were calculated in [3]. However, they remain hard to extract and we hope our approach is more transparent.

Proof. As mentioned above our starting point are well-known character tables for $\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$. These are upgraded to $\mathrm{GL}_2(\mathbb{Z}/p^m\mathbb{Z})$ via a simple application of Mackey theory.

Recall that p is assumed to be odd, write ω_σ for the central character of σ and let $\tilde{\sigma}$ be an irreducible component of $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})}$. Further, fix h such that $\mathrm{GL}_2(\mathbb{Z}/p^m\mathbb{Z}) = Z\,\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z}) \cup h \cdot Z\,\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$ and write $\tilde{\sigma}^h(g) = \tilde{\sigma}(hgh^{-1})$. If $\tilde{\sigma} \not\cong \tilde{\sigma}^h$, then

$$\sigma = \mathrm{Ind}_{Z\cdot\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})}^{\mathrm{GL}_2(\mathbb{Z}/p^m\mathbb{Z})}(\omega_\sigma \cdot \tilde{\sigma}). \tag{7}$$

In this case $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})} = \tilde{\sigma} + \tilde{\sigma}^h$ and we have $\chi_\sigma(zs) = w_\sigma(z)[\chi_{\tilde{\sigma}}(s) + \chi_{\tilde{\sigma}}^h(s)]$ for $z \in Z$ and $s \in \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$. Otherwise, if $\tilde{\sigma} \cong \tilde{\sigma}^h$, then

$$\sigma \oplus \sigma' = \mathrm{Ind}_{Z\cdot\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})}^{\mathrm{GL}_2(\mathbb{Z}/p^m\mathbb{Z})}(\omega_\sigma \cdot \tilde{\sigma}), \tag{8}$$

where σ' is another irreducible representation of $\mathrm{GL}_2(\mathbb{Z}/p_\pi^m\mathbb{Z})$. Here we have $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})} = \tilde{\sigma}$ and obtain $\chi_\sigma(zs) = \omega_\sigma(z)\chi_{\tilde{\sigma}}(s)$.

The subgroups $K_{m,\lambda}$ are all normal in $K_{m,0}$, so that the sets $Z \cdot K_{m_\pi,\lambda} \neg K_{m_\pi,\lambda+1}$ and $hZ \cdot K_{m_\pi,\lambda} \neg hZ \cdot K_{m_\pi,\lambda+1}$ can be decomposed into (disjoint) $\mathrm{GL}_2(\mathbb{Z}/p^{m_\pi})$ conjugacy classes. We focus on estimating χ_σ for

$$g = z \cdot s \in Z \cdot K_{m_\pi,\lambda} \neg K_{m_\pi,\lambda+1} \subseteq Z \cdot \mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z}).$$

The case $g \in hZ \cdot \mathrm{SL}_2(\mathbb{Z}/p^m Z)$ is similar.

Let us recall that non-trivial irreducible representations of $\mathrm{SL}_2(\mathbb{Z}/p^m\mathbb{Z})$ with $m > 1$ have dimensions $p^m(1-p^{-1})$, $p^m(1+p^{-1})$ or $1/2p^m(1-p^{-2})$. If $m = 1$, we are dealing with the representation theory of $\mathrm{SL}_2(\mathbb{F}_p)$ which was for example studied by Schur in [22]. It turns out that, besides the normal representations of dimensions $p+1$ and $p-1$, we have certain special cases. Indeed, we have the Steinberg representation ρ_{St} of dimension p and

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

two irreducible representations ρ_+ and ρ_+^h (resp. ρ_- and ρ_-^h) of dimension $1/2(p+1)$ (resp. $1/2(p-1)$).

We now treat each case appearing in Lemma 2·1 separately.

(a) Suppose π belongs to Case 1. We recall from Lemma 2·2 that in this case σ has dimension $p^{m_\pi}(1+1/p)$. We need to look into two sub cases:

(i) Suppose $m_\pi = 1$. In this case two situations can occur. If σ stays irreducible when restricted to $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, then we are in case (8) and $|\chi_\sigma(zs)| = |\chi_{\tilde{\sigma}}(s)|$ with an irreducible representation $\tilde{\sigma}$ of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ of dimension $p+1$. On the other hand, if $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})} = \rho_+ \oplus \rho_-$, then we have $|\chi_\sigma(zs)| = |\chi_{\rho_+}(s) + \chi_{\rho_+^h}(s)|$. In both cases one concludes by looking up the corresponding character values, for example in [**15**, table V]. (Alternatively one can compute the character of σ directly by observing that it can be realised as an irreducible parabolically induced representation of $\mathrm{GL}(\mathbb{F}_p)$.)

(ii) Suppose $m_\pi > 1$. By looking at the possible dimensions of non-trivial irreducible representations of $\mathrm{SL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})$ we see that $\tilde{\sigma}$ must have dimension $p^{m_\pi}(1+1/p)$ as well. Thus σ remains irreducible when restricted to $\mathrm{SL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})$ and we are in the situation of (8). We have $|\chi_\sigma(zs)| = |\chi_{\tilde{\sigma}}(s)|$ and the relevant character values for $\chi_{\tilde{\sigma}}$ can be found in [**15**, table III].

(b) Suppose π belongs to Case 2. Here we automatically have $m_\pi = 1$ and applying Lemma 2·2 yields that the dimension of σ is p. We conclude that we are in the situation of (8) and $\tilde{\sigma} = \rho$ is the Steinberg representation. We have $|\chi_\sigma(zs)| = |\chi_{\tilde{\sigma}}(s)| = 1$ as can be seen from [**15**, table V]. (Alternatively one could note that σ itself is the Steinberg representation of $\mathrm{GL}_2(\mathbb{F}_p)$, whose character is also well known.)

(c) Suppose π belongs to Case 3·1. An application of Lemma 2·2 yields that σ has dimension $p^{m_\pi}(1-1/p)$. As in Case 1 we need to distinguish two sub cases:

(i) Suppose $m_\pi = 1$. We can have $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})} = \rho_- \oplus \rho_-^h$. In this case we estimate $|\chi_\sigma(zs)| = |\chi_{\rho_+}(s) + \chi_{\rho_+^h}(s)|$. Otherwise, $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})} = \tilde{\sigma}$ for an irreducible representation $\tilde{\sigma}$ of $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ of dimension $p-1$. Here we simply have $|\chi_\sigma(zs)| = |\chi_{\tilde{\sigma}}(s)|$. The desired estimate is produced by looking up the relevant character values. These can be found in [**15**, table V] for example. (Alternatively one can observe that σ is a cuspidal representation of $\mathrm{GL}_2(\mathbb{F}_p)$. It turns out that values of characters of cuspidal representations are even known for $\mathrm{GL}_n(\mathbb{F}_p)$, see for example [**10**].)

(ii) Suppose $m_\pi > 1$. In this case we again must be in the situation of (8). In particular, $\tilde{\sigma}$ is a representation of dimension $p^{m_\pi}(1-1/p)$. The relevant character values can be found in [**15**, table III].

(d) Suppose π belongs to Case 3.2. Note that in this case we automatically have $m_\pi > 1$. Furthermore, Lemma 2·2 tells us that the dimension of σ is $p^{m_\pi}(1-1/p^2)$. Thus, taking the dimensions of (irreducible) representations of $\mathrm{SL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})$ into account, we see that $\sigma|_{\mathrm{SL}_2(\mathbb{Z}/p^{m_\pi}\mathbb{Z})}$ must be reducible, so that we are in the situation of (7). We estimate $|\chi_\sigma(zs)| \leq 2\max(|\chi_{\tilde{\sigma}}(s)|, |\chi_{\tilde{\sigma}^h}(s)|)$. Both representations $\tilde{\sigma}$ and $\tilde{\sigma}^h$ must have dimension $1/2p^{m_\pi}(1-1/p^2)$ and the corresponding character values can be found in [**15**, table IV].[1]

[1] It is maybe interesting to note that the characters appearing in Case 3.2 feature Kloosterman sums. These sums are responsible for the large values of these characters on certain conjugacy classes.

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

This exhausts all the cases and the proof is complete.

Before continuing we will discuss our choice of S. But first recall that $d = \dim_{\mathbb{C}} \pi_p^{K_p(m_\pi)} \asymp p^{m_\pi}$. Further note that $y_r = 0$ unless $r = 1, l_1, l_1 l_2, l_1^2 l_2^2$ for $l_1, l_2 \in S$. Put $\Lambda = d^{\frac{1}{3}}$ or $\Lambda = d^{\frac{1}{6}}$. This is a slight spoiler but for experts in amplification it should be no surprise that this is the optimal size of the amplifier in this setting. Let

$$S = \{l \text{ prime} : l \asymp \Lambda\}.$$

By the prime number theorem (assuming d is sufficiently large, which is no problem) we have $\sharp S \sim \Lambda / \log(\Lambda)$, but for us the following crude bound suffices

$$\sharp S \gg_\epsilon \Lambda^{1-\epsilon}.$$

Before we are ready to prove our key estimate we need to establish some counting results. Let

$$M_z^{(\lambda)}(r) = \{A \in \begin{bmatrix} \mathbb{Z} & p^\lambda \mathbb{Z} \\ p^\lambda \mathbb{Z} & \mathbb{Z} \end{bmatrix} : \det(A) = r \text{ and } u(Az, z) \leq 1\}.$$

The case $\lambda = 0$ is easily handled using existing results. For example taking $N = \delta = 1$ in [23, proposition 6·1]. We follow standard procedure and write

$$M_z^{(\lambda)}(r) = M_{z,\star}^{(\lambda)}(r) \sqcup M_{z,p}^{(\lambda)}(r) \sqcup M_{z,u}^{(\lambda)}(r).$$

Here the subscript $\star$ indicated that we are dealing with *generic* matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $c \neq 0$ and $(a+d)^2 \neq 4r$. On the other hand u stands for *unipotent* so that $A \in M_{z,u}^{(\lambda)}(r)$ if and only if $c = 0$. Finally, $A \in M_{z,p}^{(\lambda)}(r)$ if $c \neq 0$ and $(a+d)^2 = 4r$. These are the *parabolic* matrices.

Counting the contribution of generic matrices is a standard lattice point counting argument. Note that, since we can take z in the classical fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$, this argument can be simplified. Indeed, in contrast to [11, 23] we do not need to invoke Atkin–Lehner operators here.

LEMMA 4·5. *For $\lambda > 1$ and $y \geq \sqrt{3}/2$ we have*

$$\sum_{r \asymp K} [\sharp M_{z,\star}^{(\lambda)}(r) + M_{z,p}^{(\lambda)}(r)] \ll \frac{K^{\frac{3}{2}}}{p^\lambda} + \frac{K^2}{p^{2\lambda}}$$

and

$$\sum_{\substack{r \asymp K, \\ r = \square}} \sharp M_{z,\star}^{(\lambda)}(r) \ll \frac{K^{1+\epsilon}}{p^\lambda} + \frac{K^{\frac{3}{2}+\epsilon}}{p^{2\lambda}}.$$

We closely follow the argument in [23].

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

Proof. Write $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Since $c \neq 0$ and $c \ll \sqrt{r}y^{-1}$ we have $\ll K^{\frac{1}{2}}/p^{\lambda}y$ choices for c. Similarly, using the bound $|a+d| \ll \sqrt{K}$, we have $\ll K^{\frac{1}{2}}$ possibilities to choose $a+d$. Finally, we have the bound

$$|-cz^2 + (a-d)z + b|^2 \leq 4Ky^2. \tag{9}$$

Write $b = p^{\lambda}b'$ and consider the lattice $L = \langle z, p^{\lambda}\rangle$. Note that L has cocolume $\asymp p^{\lambda}y$ and first successive minima $\gg 1$ (since $y \gg 1$). Let B be the ball of radius $2\sqrt{K}y$ around $-cz^2$. Then we have

$$\sharp\{(b', a-d)\} \ll \sharp(B \cap L) \ll 1 + \sqrt{K}y + \frac{Ky}{p^{\lambda}}.$$

We have counted the number of possibilities for the admissible quadruples $(c, b', a+d, a-d)$. Since each of those quadruples uniquely determines a matrix A we have established

$$\begin{aligned}\sharp\{A \in M_{z,\star}^{(\lambda)}(r)\colon r \asymp K\} &\ll \frac{K^{\frac{1}{2}}}{p^{\lambda}y} \cdot K^{\frac{1}{2}} \cdot (1 + \sqrt{K}y + \frac{Ky}{p^{\lambda}}) \\ &\ll \frac{K}{p^{\lambda}} + \frac{K^{\frac{3}{2}}}{p^{\lambda}} + \frac{K^2}{p^{2\lambda}}.\end{aligned}$$

The case when we are only considering square matrices only needs a minor modification. Indeed, instead of counting $a+d$ trivially as earlier we observe that

$$(a-d)^2 + 4bc = (a+d)^2 - 4(\sqrt{r})^2.$$

If the matrix is parabolic the right-hand side would be 0. Thus we now consider only generic matrices. For those we can fix the left-hand side first, so that we determine $(a+d, \sqrt{r})$ essentially as solutions to a generalised Pell equation. There are at most $\ll K^{\epsilon}$ possibilities.

LEMMA 4·6. *We have*

$$\sharp M_{z,u}^{(\lambda)}(r) \ll r^{\epsilon}(1 + \sqrt{r}p^{-\lambda}y).$$

$$\sum_{l \in S} M_{z,u}^{(\lambda)}(l) \ll \Lambda + \frac{\Lambda^{\frac{3}{2}}y}{p^{\lambda}},$$

$$\sum_{l_1,l_2 \in S} M_{z,u}^{(\lambda)}(l_1 l_2) \ll \Lambda^2 + \frac{\Lambda^3 y}{p^{\lambda}} \text{ and}$$

$$\sum_{l_1,l_2 \in S} M_{z,u}^{(\lambda)}(l_1^2 l_2^2) \ll \Lambda^2 + \frac{\Lambda^4 y}{p^{\lambda}}.$$

Proof. The first estimate follows analogously to [**14**, (A·10)] using $p^{\lambda} \mid b$. Recall the bound $|b| \ll \sqrt{r}y$ used in the process.

The other bounds are derived elementary using only the fact that S contains only primes. (In contrast to [**11**, lemma 2·4] we do not need a lattice counting argument, because we

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

have an additional congruence condition on b that we can use.) We will only show the final estimate, since the others are derived similarly.

There are $\ll \Lambda^2$ possible choices for $r = l_1^2 l_2^2 \asymp \Lambda^4$. Having fixed the determinant of this form we find that there are only $\ll 1$ choices for (a,d) with $ad = l_1^2 l_2^2$. Finally we observe that we can choose b in $\ll 1 + \Lambda^2 y/p^\lambda$ ways, since $p^\lambda \mid b$ and $|b| \ll \Lambda^2 y$. Putting these estimates together completes the proof.

Remark 4·7. Suppose $y \geq \sqrt{3}/2$. As in [14, (A.7)] we have the bound

$$|c| \leq \frac{\sqrt{8r}}{y} \leq \sqrt{\frac{2^5 r}{3}}. \tag{10}$$

Thus if $p^\lambda \geq 3,5 \cdot \sqrt{r}$, we must have $c = 0$. This is because $p^\lambda \mid c$. In this case we obtain

$$M_z^{(\lambda)}(r) = M_{z,u}^{(\lambda)}(r).$$

Finally we need to consider the parabolic contribution.

LEMMA 4·8. *For $\lambda \geq 1$ and $y \geq \sqrt{3}/2$ we have*

$$\sum_{\substack{r \asymp K,\\ r = \square}} \sharp M_{z,p}^{(\lambda)}(r) \ll \frac{K^{\frac{3}{2}}}{p^\lambda}.$$

Proof. This follows along the lines of [4, lemma 14]. We provide some details for the convenience of the reader. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a matrix contributing the the count $M_{z,p}^{(\lambda)}(r)$ for some $r \asymp K$ with $r = \square$. We start by recalling the bounds

$$cy \ll \sqrt{K} \text{ and } |2cx - (a-d)| \ll \sqrt{K}.$$

The first one is (10) and the second one follows from (9) when looking at the imaginary part.

Now, since A is parabolic, we must have

$$(a+d)^2 = 4(ad - bc) = 4r.$$

The latter equality can be rewritten as

$$(a-d)^2 + 4bc = 0.$$

This implies that, if c is fixed, then $a-d$ has a fixed divisor of size at least $\sqrt{|c|}$. We conclude that there are $\ll 1 + \sqrt{K}/\sqrt{|c|}$ possible choices for $a-d$. By summing over all possibilities for $c \neq 0$ and $p^\lambda \mid c$ we find that

$$\sharp\{c, a-d\} \ll \frac{K}{p^\lambda}.$$

Finally, note that $|a+d| = 2\sqrt{r} \asymp \sqrt{K}$, so that there are $\ll \sqrt{K}$ possible choices for $a+d$. Since $a-d$ and c determine b we are done after gathering all the contributions.

Table 1. *Summary of counting results*

$\sum_{r\in T}\frac{\lvert y_r\rvert}{\sqrt{r}}\cdot\sharp M_{z,\Diamond}^{(\lambda)}(r)$	$\Diamond=u$	$\Diamond=\star$	$\Diamond=p$
$T=\{1\}$	$\Lambda+\frac{\Lambda y}{p^\lambda}$	0	0
$T=\{l\colon l\in S\}$	$\Lambda^{\frac{1}{2}}+\frac{\Lambda y}{p^\lambda}$	$\frac{\Lambda}{p^\lambda}+\frac{\Lambda^{\frac{3}{2}}}{p^{2\lambda}}$	included in $\star$
$T=\{l_1l_2\colon l_1,l_2\in S\}$	$\Lambda+\frac{\Lambda^2 y}{p^\lambda}$	$\frac{\Lambda^2}{p^\lambda}+\frac{\Lambda^3}{p^{2\lambda}}$	included in $\star$
$T=\{l_1^2l_2^2\colon l_1,l_2\in S\}$	$1+\frac{\Lambda^2 y}{p^\lambda}$	$\frac{\Lambda^{2+\epsilon}}{p^\lambda}+\frac{\Lambda^{4+\epsilon}}{p^{2\lambda}}$	$\frac{\Lambda^4}{p^\lambda}$
total	$\Lambda+\frac{\Lambda^2 y}{p^\lambda}$	$\frac{\Lambda^{2+\epsilon}}{p^\lambda}+\frac{\Lambda^{4+\epsilon}}{p^{2\lambda}}$	$\frac{\Lambda^4}{p^\lambda}$

We can now prove the main estimate of this section.

LEMMA 4·9. *Assume $p>3$. Suppose π belongs to Case 1, 2 or 3·1. Then, for $\sqrt{3}/2\ll y\ll\sqrt{d}$ we have*

$$\Phi(x+iy)\ll\sqrt{T}d^{\frac{5}{6}}.$$

If π belongs to Case 3.2, then we have the weaker bound

$$\Phi(x+iy)\ll\sqrt{T}d^{\frac{11}{12}}\ \textit{for}\ \frac{\sqrt{3}}{2}\ll y\ll d^{\frac{1}{4}}.$$

Proof. We start with Cases 1, 2 or 3·1. Our starting point is Corollary 4·3. Breaking the g-sum up into pieces on which we can estimate the character using Lemma 4·4 we get

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}}\pi_p^{K_p(m_\pi)}}\cdot\Phi(z)^2\ll T\sum_{0\leq\lambda\leq m_\pi}p^\lambda\sum_r\frac{\lvert y_r\rvert}{\sqrt{r}}\cdot\sharp M_z^{(\lambda)}(r).$$

We first consider the contribution of $\lambda=0$. In this case the counting problem is independent of p and relatively easy. Indeed we have

$$\sum_r\frac{y_r}{\sqrt{r}}\sharp\{A\in M_2(\mathbb{Z})\colon\ \det(A)=r\text{ and }u(Az,z)\leq 1\}\ll\Lambda^4+\Lambda^{\frac{5}{2}}y. \tag{11}$$

This is for example [23, proposition 6·1] with $N=\delta=1$ and $z\in\mathcal{F}$ so that $y\gg 1$.

Next we assume $\lambda>0$. We summarise the results from Lemma 4·5, 4·8 and 4·6 in Table I below. Note that for the contribution of $r=1$ we have used Remark 4·7.

All together this gives a contribution of

$$\sum_r\frac{\lvert y\rvert}{\sqrt{r}}\cdot\sharp M_z^{(\lambda)}(r)\ll\Lambda+\frac{\Lambda^4}{p^\lambda}+\frac{\Lambda^2 y}{p^\lambda}.$$

Inserting these estimates in our (amplified) pre-trace inequality we get

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}}\pi_p^{K_p(m_\pi)}}\cdot\Phi(z)^2\ll T\left(\Lambda^4+\Lambda^{\frac{5}{2}}y+\sum_{1\leq\lambda\leq m_\pi}p^\lambda\left[\Lambda+\frac{\Lambda^4}{p^\lambda}+\frac{\Lambda^2 y}{p^\lambda}\right]\right).$$

https://doi.org/10.1017/S0305004124000021 Published online by Cambridge University Press

This simplifies to

$$\Phi(z)^2 \ll \frac{Td^{1+\epsilon}}{\Lambda^{2-\epsilon}}\left(\Lambda^{\frac{5}{2}}y + \Lambda^4 + \Lambda d\right).$$

Inserting $\Lambda = d^{\frac{1}{3}}$ yields

$$\Phi(z)^2 \ll Td^{1+\epsilon}(d^{\frac{2}{3}} + d^{\frac{1}{6}}y)$$

which directly implies the result for the non-exceptional cases 1, 2 and 3·1.

Finally if π belongs to Case 3·2, then the same analysis with the weaker character estimates yields

$$\frac{(\sharp S)^2}{\dim_{\mathbb{C}} \pi_p^{K_p(m_\pi)}} \cdot \Phi(z)^2 \ll T\left(d^{\frac{1}{2}}\Lambda^4 + d^{\frac{1}{2}}\Lambda^{\frac{5}{2}}y + \sum_{1\leq \lambda\leq m_\pi} p^{\frac{m_\pi+\lambda}{2}}\left[\Lambda + \frac{\Lambda^4}{p^\lambda} + \frac{\Lambda^2 y}{p^\lambda}\right]\right)$$
$$\ll T\left(d^{\frac{1}{2}}\Lambda^4 + d^{\frac{1}{2}}\Lambda^{\frac{5}{2}}y + \Lambda d^{1+\epsilon}\right).$$

This prompts the choice $\Lambda = d^{\frac{1}{6}}$ and we find

$$\Phi(z)^2 \ll Td^{1+\epsilon}(d^{\frac{5}{6}} + d^{\frac{7}{12}}y)$$

This completes the proof.

Proof. of Theorem 1·1: We are now ready to complete the proof of our main theorem. First of all note that it is enough to bound $\Phi(z)$ for $z \in \mathcal{F}$. Next, if $y \ll \sqrt{d}$ (resp. $y \ll d^{\frac{1}{4}}$), then we are done by the previous lemma. For larger y the Whittaker expansion (see Lemma 3·8) gives even better result.

Acknowledgements. I thank Prof. Dr. V. Blomer for fruitful discussions and useful comments on an earlier draft of this manuscript. I would also like to thank the anonymous referees for pointing out several oversights and inaccuracies.

REFERENCES

[1] E. ASSING. On sup-norm bounds part i: ramified maaß newforms over number fields. To appear in *J. Eur. Math. Soc.* ArXiv:1710.00362 (2017).
[2] E. ASSING. On the size of p-adic Whittaker functions. *Trans. Amer. Math. Soc.* **372**(8) (2019), 5287–5340.
[3] R. B. LEIGH, G. CLIFF and Q. WEN. Character values for GL$(2, \mathbb{Z}/p^\ell\mathbb{Z})$. *J. Algebra* **323**(5) (2010), 1288–1320.
[4] V. BLOMER, G. HARCOS, P. MAGA and D. MILIĆEVIĆ. The sup-norm problem for GL(2) over number fields, *J. Eur. Math. Soc. (JEMS)* **22**(1) (2020), 1–53.
[5] V. BLOMER, G. HARCOS, P. MAGA and D. MILIĆEVIĆ. Beyond the spherical sup-norm problem. *J. Math. Pure Appl.* **168**(2022), 1-64.
[6] V. BLOMER and R. HOLOWINSKY. Bounding sup-norms of cusp forms of large level. *Invent. Math.* **179**(3) (2010), 645–681.
[7] V. BLOMER and P. MAGA. Subconvexity for sup-norms of cusp forms on PGL(n). *Selecta Math. (N.S.)* **22**(3) (2016), 1269–1287.
[8] W. CASSELMAN. The restriction of a representation of $GL_2(k)$ to $GL_2(o)$. *Math. Ann.* **206** (1973), 311–318.

[9] F. COMTAT. Optimal sup norm bounds for newforms on GL_2 with maximally ramified central character. *Forum Math*. **33**(1) (2021), 1–16.

[10] J. A. GREEN. Discrete series characters for $GL(n, q)$. *Algebr. Represent. Theory* **2**(1) (1999), 61–82.

[11] G. HARCOS and N. TEMPLIER. On the sup-norm of Maass cusp forms of large level. III, *Math. Ann.* **356**(1) (2013), 209–216.

[12] J. HOFFSTEIN and P. LOCKHART. Coefficients of Maass forms and the Siegel zero. *Ann. of Math. (2)* **140**(1) (1994), 161–181, with an appendix by D. Goldfeld, J. Hoffstein and D. Lieman.

[13] Y. HU and A. SAHA. Sup-norms of eigenfunctions in the level aspect for compact arithmetic surfaces, II: newforms and subconvexity. *Compositio. Math*. **156**(11) (2020), 2368–2398.

[14] H. IWANIEC and P. SARNAK. L^∞ norms of eigenfunctions of arithmetic surfaces. *Ann. of Math. (2)* **141**(2) (1995), 301–320.

[15] P.C. KUTZKO, The characters of the binary modular congruence group. PhD. thesis. University of Wisconsin (1972).

[16] E. LAPID and Z. MAO. A conjecture on Whittaker-Fourier coefficients of cusp forms. *J. Number Theory* **146** (2015), 448–505.

[17] D. LOEFFLER and J. WEINSTEIN. On the computation of local components of a newform. *Math. Comp*. **81**(278) (2012), 1179–1200.

[18] S. MARSHALL. Local bounds for L^p norms of Maass forms in the level aspect. *Algebra Number Theory* **10**(4) (2016), 803–812.

[19] P. MICHEL and A. VENKATESH. The subconvexity problem for GL_2. *Publ. Math. Inst. Hautes Études Sci.* **111** (2010), 171–271.

[20] M. MIYAUCHI and T. YAMAUCHI. A remark on conductor, depth and principal congruence subgroups. ArXiv:2107.08130 (2021).

[21] A. SAHA. Hybrid sup-norm bounds for Maass newforms of powerful level, *Algebra Number Theory* **11**(5) (2017), 1009–1045.

[22] J. SCHUR. Untersuchungen über die Darstellung der endlichen Gruppen durch gebrochene lineare Substitutionen. *J. Reine Angew. Math.* **132** (1907), 85–137.

[23] N. TEMPLIER. Hybrid sup-norm bounds for Hecke–Maass cusp forms. *J. Eur. Math. Soc. (JEMS)* **17**(8) (2015), 2069–2082.