COMPOSITIO MATHEMATICA

Theta functions, fourth moments of eigenforms, and the sup-norm problem I

Ilya Khayutin and Raphael S. Steiner



Compositio Math. **160** (2024), 2916–2969.

doi:10.1112/S0010437X24007437

FOUNDATION
COMPOSITIO
MATHEMATICA



LONDON
MATHEMATICAL
SOCIETY
EST. 1865



https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Compositio Math. 160 (2024) 2916–2969
doi:10.1112/S0010437X24007437

Theta functions, fourth moments of eigenforms, and the sup-norm problem I

Ilya Khayutin and Raphael S. Steiner

Abstract

We give sharp point-wise bounds in the weight-aspect on fourth moments of modular forms on arithmetic hyperbolic surfaces associated to Eichler orders. Thereby, we strengthen a result of Xia and extend it to co-compact lattices. We realize this fourth moment by constructing a holomorphic theta kernel on $\mathbf{G} \times \mathbf{G} \times \mathbf{SL}_2$, for $\mathbf{G}$ an indefinite inner form of $\mathbf{SL}_2$ over $\mathbb{Q}$, based on the Bergman kernel, and considering its L^2-norm in the Weil variable. The constructed theta kernel further gives rise to new elementary theta series for integral quadratic forms of signature $(2, 2)$.

Contents

1 Introduction **2917**
2 General method **2919**
3 The Weil representation and theta series **2922**
3.1 Inner forms of SL2 . . . 2922
3.2 Normalization of Haar measures . . . 2923
3.3 Local Weil representation . . . 2923
3.4 The Archimedean Weil representation . . . 2925
3.5 The non-Archimedean Weil representation . . . 2927
3.6 Local uniformity . . . 2933
3.7 Global Weil representation and theta series . . . 2934
4 Theta kernels for Eichler orders **2935**
4.1 Weil action on Eichler orders . . . 2935
4.2 Theta series for Eichler orders . . . 2937
4.3 The Weil L^2-norm of the theta kernel . . . 2938
5 The theta lift **2940**
5.1 Cuspidal theta series . . . 2940
5.2 Unfolding . . . 2940
5.3 The theta lift . . . 2944
6 The Bergman kernel **2948**
6.1 The Bergman Archimedean test function . . . 2948
7 The spectral expansion **2951**
8 The geometric expansion **2954**

Received 7 November 2022, accepted in final form 16 May 2024, published online 3 April 2025.
2020 Mathematics Subject Classification 11F72 (primary), 11F11, 11F27, 11F70, 58J50 (secondary).
Keywords: sup-norm, holomorphic forms, theta function, theta correspondence, fourth moment.

© The Author(s), 2025. The publishing rights in this article are licensed to Foundation Compositio Mathematica under an exclusive licence.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

9 Second-moment count of quaternions by norm **2957**
9.1 Second-moment bound for the split matrix algebra 2957
9.2 Second-moment bound for division algebras 2960
10 Proof of the main theorem **2964**
10.1 Proof of the main theorem for the split matrix algebra 2964
10.2 Proof of the main theorem for division algebras 2966
Acknowledgements **2966**
References **2967**

1. Introduction

The study of distributional aspects of automorphic forms has enjoyed ample consideration in the past couple of decades, in particular questions related to the quantum unique ergodicity conjecture, various bounds for L^p-norms, and restriction problems. In this paper, we are mainly concerned with the L^∞-norm of holomorphic Hecke eigenforms on arithmetic hyperbolic surfaces in the large weight limit, though our method also gives essentially sharp results for moments of L^4-norms.

The sup-norm problem asks for the best possible bound on the sup-norm of a Hecke eigenform in terms of the analytic conductor. Specifically, one often seeks a non-trivial bound on the sup-norm separately with respect to the weight, Laplace eigenvalue, or level aspect. It is analogous and closely related to the Lindelöf Hypothesis for automorphic L-functions. The go-to method for the majority of previous work on this problem is amplification. It was first used in this context by Iwaniec and Sarnak in the pioneering paper [IS95], though the idea of an amplifier goes back to Selberg [Sel42]. Iwaniec and Sarnak showed the bound

$$\|\varphi\|_\infty \ll_{\Gamma,\varepsilon} (1+|\lambda_\varphi|)^{5/24+\varepsilon}\|\varphi\|_2 \tag{1}$$

for a Hecke–Maass form $\varphi\colon \Gamma\backslash\mathbb{H} \to \mathbb{C}$, where the lattice $\Gamma < \mathbf{SL}_2(\mathbb{R})$ is the unit norm elements of an Eichler order in a quadratic division algebra. Here and henceforth, we have adopted Vinogradov's notation. Their result marked the first time a power of $1+|\lambda_\varphi|$ was saved over what holds for a general Riemannian surface. Indeed, (1) has been known to hold with exponent $\frac{1}{4}$ for a general compact Riemannian surface, without any further assumptions of arithmetic nature (cf. [Sog88]). The amplifying technique has been used heavily due to its versatility. In the context of automorphic forms on arithmetic hyperbolic surfaces, Blomer and Holowinsky [BH10], Templier [Tem10, Tem15], Harcos and Templier [HT12, HT13], Saha [Sah17a, Sah17b, Sah20], Hu and Saha [HS20], and Kıral [Kır14] have used it to show subconvex bounds in various level aspects; Das and Sengupta [DS15] and Steiner [Ste17] have used it to show subconvex bounds in the weight aspect. Blomer, Harcos, and Milićević [BHM16] and Blomer, Harcos, Maga, and Milićević [BHMM20] applied it to a more general setting over number fields, which corresponds to products of hyperbolic 2- and 3-spaces. The most general $\mathbf{PGL}_2$ result is due to Assing [Ass17]. Moreover, the technique has also been adopted to arithmetic 2-spheres by Vanderkam [Van97] and products of 2- and 3-spheres by Blomer and Michel [BM11, BM13], and generalized to higher rank, e.g., by Blomer and Pohl [BP16] for $\mathbf{Sp}_4$, Blomer and Maga [BM15, BM16] for $\mathbf{PGL}_n$ ($n \geq 4$), and Marshall [Mar14] for semisimple split Lie groups over totally real fields and their totally imaginary quadratic extensions, to name a few examples.

In this paper, we employ a different tool, namely the theta correspondence. The theta correspondence was first used by the second named author [Ste20] to tackle sup-norm problems. It has been used previously by Nelson to answer questions regarding quantum unique ergodicity

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

and quantum variance [Nel16, Nel17, Nel19, Nel20], and give Fourier-like expansions for forms living on compact spaces [Nel15]. The main advantage of this approach is that instead of looking at an amplified second moment, we are able to bound a fourth moment sharply. Another advantage is that it works for co-compact lattices equally well as it does for non-co-compact lattices. Our main theorem and its corollary read as follows.

Theorem 1.1. *Let the arithmetic lattice $\Gamma < \mathbf{SL}_2(\mathbb{R})$ be the unit norm elements of an Eichler order in an indefinite quaternion algebra over $\mathbb{Q}$ and $\{f_j\}_j \subset S_m^{\mathrm{new}}(\Gamma)$ be an orthonormal[1] basis of Hecke newforms of weight $m > 4$. Then, there is a constant $A \geq 1$, such that for any $\varepsilon > 0$, there is a constant C_ε for which we have*

$$\sum_j y^{2m}|f_j(z)|^4 \leq C_\varepsilon \operatorname{covol}(\Gamma)^A m^{1+\varepsilon}\big(1 + m^{-1/2}\mathrm{ht}_\Gamma(z)^2\big), \tag{2}$$

where $\mathrm{ht}_\Gamma(z) = 1$ if Γ is co-compact and

$$\mathrm{ht}_\Gamma(z) = \max_{\gamma \in \mathbf{SL}_2(\mathbb{Z})} \Im(\gamma z)$$

if $\Gamma < \mathbf{SL}_2(\mathbb{Z})$. Furthermore, we have

$$\sum_j \|f_j\|_4^4 \leq C_\epsilon \operatorname{covol}(\Gamma)^A m^{1+\epsilon}. \tag{3}$$

Corollary 1.2. *Let $\Gamma < \mathbf{SL}_2(\mathbb{R})$ be as above with the additional assumption of being co-compact ($\Leftrightarrow B(\mathbb{Q})$ is non-split) and $f \in S_m^{\mathrm{new}}(\Gamma)$ a Hecke newform of weight $m > 4$. Then, there is a constant $A \geq 1$, such that for any $\varepsilon > 0$, there is a constant C_ε for which we have*

$$\sup_{z \in \mathbb{H}} y^{m/2}|f(z)| \leq C_\varepsilon \operatorname{covol}(\Gamma)^A m^{1/4+\varepsilon}\|f\|_2. \tag{4}$$

The first half of Theorem 1.1 marks a significant improvement over what has been known previously. It shows that the L^∞-norm of the fourth moment of holomorphic newforms of weight m is, essentially, as small as it can be, meaning that they enjoy a stronger 'orthogonality' relation than what was known previously. Remarkably, our proof does not rely on any deep results from arithmetic geometry such as Deligne's bound for the Hecke eigenvalues, but rather a sharp bound for a second-moment matrix count as we shall explain in further detail in § 2. The second half of Theorem 1.1 is a simple consequence of the first half if Γ is co-compact and otherwise it follows in conjunction with [BKY13, Theorem 1.8], which says that the mass of the fourth norm is concentrated in the domain $\{z \in \Gamma\backslash\mathbb{H} : \mathrm{ht}_\Gamma(z) \leq m^{1/4}\}$. Following Sarnak and Watson [Sar03], Inequality (3), through the use of Watson's formula [WaT08, Theorem 3] or, more generally, Ichino's formula [Ich08] (cf. [Nel11, § 4]), may be reformulated as a Lindelöf on average statement about degree-eight L-functions. In particular, assuming that the product of the reduced discriminant D_B of B and the level q of Γ is square-free, one deduces

$$\frac{1}{m}\sum_f \frac{1}{2m}\sum_g L\left(f \times f \times g, \frac{1}{2}\right) \leq C_\varepsilon \operatorname{covol}(\Gamma)^A m^\varepsilon, \tag{5}$$

where $f \in S_m^{\mathrm{new}}(\Gamma)$ runs through a basis of newforms of weight m for Γ and $g \in S_{2m}(\Gamma)$ runs through an orthonormal set of newforms of weight $2m$ for Γ with Hecke eigenvalues equal to 1 for all primes $p \mid D_B$ and Atkin–Lehner eigenvalues equal to -1 for all primes $p \mid q$. This should be compared with the result of Sun and Ye [SY19] who considered the double average of the degree-six L-function $L(\mathrm{Sym}^2 f \times g, \frac{1}{2})$, where f, g are Hecke eigenforms of weight m, respectively

[1] With respect to the invariant probability measure.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

$2m$, for $\mathbf{SL}_2(\mathbb{Z})$. Note that $L(f \times f \times g, \frac{1}{2}) = L(\mathrm{Sym}^2 f \times g, \frac{1}{2})L(g, \frac{1}{2})$. One should also mention a result of Khan [Kha14], who managed to show an asymptotic formula for the left-hand side of (3) for $\Gamma = \mathbf{SL}_2(\mathbb{Z})$ with an extra (smooth) average over the weight m. Khan's result matches up with conjectures concerning the asymptotics of the L^4-norm in the large weight aspect. We refer to [BKY13] for details regarding these conjectures. In the future, we plan to address the question whether one can upgrade the second half of Theorem 1.1 to an asymptotic without any extra average over the weight. We shall also mention the strongest individual bound for the L^4-norm of a Hecke eigenform f of weight m on $\mathbf{SL}_2(\mathbb{Z})$ which is due to Blomer, Khan, and Young [BKY13]. They managed to show $\|f\|_4 \ll_\varepsilon m^{1/12+\varepsilon}\|f\|_2$.

The convex or trivial bound in the context of Corollary 1.2 is $\ll \mathrm{covol}(\Gamma)^{1/2}m^{1/2}$ and the first non-trivial bound in the weight aspect $\ll_\varepsilon m^{1/2-\delta+\varepsilon}$ for a small $\delta > 0$ was achieved by Das–Sengupta[2] [DS15] through the use of an amplifier. The previous best bound in the weight aspect is due to Ramacher and Wakatsuki [RW21] who established a subconvex bound for the sup-norm in significant generality.

The analogue of Corollary 1.2 for non-uniform lattices is much easier to establish, because one can use the Fourier expansion at a cusp and then apply Deligne's bound for the Fourier coefficients. This was observed by Xia [Xia07], who worked out the case $\Gamma = \mathbf{SL}_2(\mathbb{Z})$. In the same fashion, a sharp hybrid bound for holomorphic forms of minimal type was derived by Hu, Nelson, and Saha [HNS19]. We would also like to thank Paul Nelson for pointing out to us the relation between our technique and [Nel15, Theorem 3.1]. Nelson uses an explicit (non-holomorphic) version of Shimizu's theta kernel [Shi72] to construct an expansion of $y^m|f(z)|^2$, where f is an arithmetically normalized newform on a compact arithmetic surface, that resembles a Fourier expansion.

Finally, we shall mention that we did not attempt to optimize the dependence on the covolume or level in Theorem 1.1 in this first paper. Due to our method requiring sharp bounds for a second-moment matrix count of length comparable to the conductor, any such undertaking must necessarily address the inability of pre-existing matrix counting techniques in the non-split case, such as [Tem10], to deal with large determinants. Furthermore, a strategy needs to be devised to incorporate the dependence on the reduced discriminant of the indefinite quaternion algebra. All of this shall be addressed in a sequel joint with Nelson [KNS22].

As far as the structure of this paper goes, in the following section, we shall briefly explain the main concept of the proof as well as mentioning an alternative approach using L-functions instead of a theta kernel. Sections 3 and 4 deal with local and global properties of the Weil representation and their consequences to the associated theta series. The action of the Hecke algebra on the theta kernel is computed in § 5. In §§ 6 and 7, we show that the Bergman kernel satisfies the required assumptions in the construction of the theta kernel and compute its spectral expansion. In § 8, we reduce a bound on the L^2-norm of the theta kernel to matrix counts. In § 9, we prove the essentially sharp second-moment matrix count. The main theorem is then established in § 10.

2. General method

In this section, we shall briefly explain two essentially equivalent strategies that lead to Theorem 1.1. We shall first lay out the approach which is conceptually closer to that of

[2] In the published version $\delta = 1/32$ appears, though this has been corrected to $\delta = 1/64$ in a recent revision on the arXiv [DS20, Remark 4.1].

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

an amplifier. For simplicity, we shall assume everything is unramified, i.e. $\Gamma = \mathbf{SL}_2(\mathbb{Z})$, which is the set of determinant one elements of the maximal order $\mathcal{R} = \mathrm{Mat}_{2x2}(\mathbb{Z})$ inside the quaternion algebra $\mathrm{Mat}_{2x2}(\mathbb{Q})$. Let $\mathcal{R}_n$ denote the elements of $\mathcal{R}$ of norm n, such that $\Gamma = \mathcal{R}_1$. We begin with a Bergman kernel (also known as a reproducing kernel) on $S_m(\Gamma)$, the space of weight m holomorphic cusp forms on Γ,

$$B(z,w) = \sum_j \Im(z)^{m/2} f_j(z) \Im(w)^{m/2} \overline{f_j(w)}, \tag{6}$$

where $\{f_j\}_j$ is an orthonormal basis of Hecke eigenforms of the space $S_m(\Gamma)$. The amplified counterparts to the Bergman kernel are

$$B_n(z,w) = \sum_j \lambda_j(n) \Im(z)^{m/2} f_j(z) \Im(w)^{m/2} \overline{f_j(w)}, \tag{7}$$

where $\lambda_j(n)$ is the n-th Hecke eigenvalue of the newform f_j. We normalize the Hecke operators so that Deligne's bound reads $|\lambda_j(n)| \leq d(n)$, $d(n)$ is the divisor function. The kernels B_n are roughly of the shape

$$B_n(z,w) \approx \frac{m}{\sqrt{n}} \sum_{\substack{\alpha \in \mathcal{R}_n \\ u(\alpha z, w) \leq 1/m}} 1, \tag{8}$$

where $u(z,w) = |z-w|^2/(4\Im(z)\Im(w))$. Instead of taking a suitable linear combination of (7) as one would do for an amplifier, we consider

$$\int_0^1 \left| \sum_{n \leq m} B_n(z,z) e(nt) \right|^2 dt = \sum_{i,j} \Im(z)^{2m} |f_j(z)|^2 |f_i(z)|^2 \sum_{n \leq m} \lambda_j(n) \overline{\lambda_i(n)}. \tag{9}$$

To the latter, or more precisely a smooth version thereof, one may apply Voronoï summation. If we set aside any intricacies stemming from Riemann zeta factors and smoothing, we pick up main terms for $i = j$ corresponding to the poles of $L(f_i \times f_j, s)$ at $s = 1$ for $i = j$ and a dual sum of length m^2/m. Thus, we find that (9) is approximately

$$m \sum_j \Im(z)^{2m} |f_j(z)|^4 + \sum_{\substack{i,j \\ i \neq j}} \Im(z)^{2m} |f_j(z)|^2 |f_i(z)|^2 \sum_{n \leq m^2/m} \lambda_j(n) \overline{\lambda_i(n)}. \tag{10}$$

We see that the *new* dual sum is once again of the shape (9) and we may replace it with its geometric counterpart. Through rearranging and the use of the approximation of the Bergman kernel (8), one arrives at

$$\sum_j \Im(z)^{2m} |f_j(z)|^4 \ll m \sum_{n \leq m} \frac{1}{n} \sum_{\substack{\alpha_1, \alpha_2 \in \mathcal{R}_n \\ u(\alpha_i z, z) \leq 1/m,\ i=1,2}} 1. \tag{11}$$

We see that we end up with a second-moment matrix count. Before we discuss the latter further, we shall describe how to arrive at the same inequality in an alternate fashion by using a theta kernel.

At its core, one wishes to find a kernel[3] $\vartheta : \Gamma \backslash \mathbf{SL}_2(\mathbb{R}) \times \Gamma \backslash \mathbf{SL}_2(\mathbb{R}) \times \Lambda \backslash \mathbf{SL}_2(\mathbb{R}) \to \mathbb{C}$, such that

$$\left\langle \vartheta(z,w;\bullet), (\Im \bullet)^{m/2} \tilde{f} \right\rangle = \Im(z)^{m/2} f(z) \Im(w)^{m/2} \overline{f(w)} \cdot \|\tilde{f}\|_2^2, \tag{12}$$

[3] Here, $\Gamma < \mathbf{G}(\mathbb{Q})$ is a congruence lattice in an indefinite inner form $\mathbf{G}$ of $\mathbf{SL}_2$ and $\Lambda < \mathbf{SL}_2(\mathbb{Q})$ is a lattice in the split form $\mathbf{SL}_2$ that arises in Shimizu's explicit Jacquet–Langlands transfer of modular forms on $\Gamma \backslash \mathbf{G}(\mathbb{R})$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

for an L^2-normalized newform f and $\tilde{f}$ an arithmetically normalized newform in the Jacquet–Langlands transfer to $\mathbf{GL}_2$ of the automorphic representation generated by f. It immediately follows that

$$\sum_f \|\tilde{f}\|_2^2 \cdot \Im(z)^{2m}|f(z)|^4 \leq \|\vartheta(z,z;\bullet)\|_2^2 \tag{13}$$

by Bessel's inequality. For $\Gamma = \mathbf{SL}_2(\mathbb{Z})$, such a kernel may be given by

$$\vartheta(z,w;\zeta) = \Im(\zeta)^{m/2}\sum_{n=1}^{\infty} B_n(z,w)n^{(m-1)/2}e(n\zeta). \tag{14}$$

This may be used to recover (11) upon using the Hoffstein–Lockhart bound for $\|\tilde{f}\|_2$ (see [HL94]) and standard bounds for the incomplete Gamma function.

We prefer to employ the latter approach as it avoids translating spectral data back into geometric terms. Specifically, in (10), we have been able to replace the dual sum by the integral of the same amplified Bergman kernel on the left-hand side of (9). This step cannot be reproduced verbatim in the ramified cases. Instead, one would need to express the dual sum in terms of Fourier expansions of amplified Bergman kernels associated to various levels and different cusps. The approach using the theta correspondence avoids these issues altogether.

Whilst the constructions of theta kernels in great generality have been known for a while, see [Shi72] or [Nel15, § 5 & Appendix B] for an explicit example, they are unfortunately generally not in L^2. An attempt to rectify this, would be to project such a theta kernel to $S_m(\Gamma)$. Formulae for such projections are given by Gross and Zagier [GZ86, § IV.5]. However, we follow a different path. Motivated by the simplicity of the kernel ϑ in the case $\Gamma = \mathbf{SL}_2(\mathbb{Z})$ (14), we modify the general construction of a theta kernel to mirror a classical Bergman kernel of weight m. In order to show that the novel theta kernel behaves in the prescribed fashion, we use a method of Vignéras [Vig77] at the infinite place and compute the Fourier–Whittaker expansion in the ζ-variable. We compare the latter with Shimizu's explicit form of the Jacquet–Langlands correspondence [Shi72]. As a corollary, we derive a new elementary theta series for indefinite quadratic forms of signature $(2,2)$.

THEOREM 2.1. *Let $\mathcal{R}$ be an Eichler order of level q in an indefinite division quaternion algebra over $\mathbb{Q}$ of reduced discriminant D_B. Denote by $\mathcal{R}^+$ the subset of elements of positive norm and by Γ the subset of elements of norm equal to one. Furthermore, let $f \in S_m(\Gamma)$ be a cusp form of weight $m > 2$. Then, for each $z \in \mathbb{H}$, the function $\mathscr{F}_f(z;\bullet)$, given by*

$$\mathscr{F}_f(z;\zeta) = \sum_{\alpha \in \Gamma\backslash\mathcal{R}^+} \mathrm{Nr}(\alpha)^{m/2-1}(f|_m\alpha)(z)e(\mathrm{Nr}(\alpha)\zeta), \tag{15}$$

is a cusp form of weight m for $\Gamma_0(qD_B)$. Moreover, we have $\mathscr{F}_{T_n f}(z;\bullet) = (T_n\mathscr{F}_f)(z;\bullet)$ for $(n, qD_B) = 1$.

Returning to the second-moment matrix count, we see that upon using partial summation we need to bound the number of solutions to

$$\alpha_1, \alpha_2 \in \mathcal{R} : 1 \leq \mathrm{Nr}(\alpha_1) = \mathrm{Nr}(\alpha_2) \leq N, \quad u(\alpha_i z, z) \leq \frac{1}{m}, \quad i = 1,2. \tag{16}$$

Consider z fixed for the moment. Then, we are given a quadratic equation in eight variables all of size $N^{1/2}$ with four additional linear inequalities of density $m^{-1/2}$. Heuristics suggest that we should have on the order of $(N^{1/2})^8 \cdot N^{-1} \cdot (m^{-1/2})^4 = N^3m^{-2}$ solutions for N large. We see that for $N \leq m$, $N^3m^{-2} \leq N$, which is the bound we are aiming for. Moreover, by considering

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

the order $\mathrm{Mat}_{2\times 2}(\mathbb{Z})$ and the special point $z = i$, we see that the matrices of the shape $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ with $1 \leq a^2 + b^2 \leq N$ satisfy the conditions in (16) and give rise to a lower bound of size N. Likewise, we should expect that such subvarieties with exceptionally many solutions exist also for other special points and orders under consideration. Hence, the general estimate we seek is at the edge of what is achievable. This is in stark contrast to the classical approach of an amplifier, where one may consider matrices of reduced norm up to only a small power of m in order to get a non-trivial result. However, the difficulty of the task at hand is rewarded with a sharp fourth moment estimate. In order to achieve the required bound, we rely on geometry of numbers arguments, which have been successful in the past for first moments (cf. [HT13]), in particular with regards to uniformity in the varying point z. To account for the additional quadratic equation, we decompose each matrix α_i into two parts: a multiple of the identity and a traceless part α_i^0. To the traceless parts α_i^0 we apply the geometry of numbers arguments. The quadratic equation now reads

$$\mathrm{Nr}\,\alpha_1 = \frac{\mathrm{Tr}(\alpha_1)^2}{4} + \mathrm{Nr}\,\alpha_1^0 = \mathrm{Nr}\,\alpha_2 = \frac{\mathrm{Tr}(\alpha_2)^2}{4} + \mathrm{Nr}\,\alpha_2^0$$

and we may use the divisor bound to bound the number of possibilities for the traces. This gives the required bound if and only if the traces are not equal in absolute value. The latter case needs to be dealt with separately. We do so by showing that there are essentially only a constant number of matrices $\alpha \in \mathscr{R}$ satisfying $u(\alpha z, z) \leq 1/m$ of a given trace and reduced norm $\leq m$.

As a final remark, we address the natural question, whether the method lends itself to further amplification. Albeit it being straightforward to produce amplified versions of (11), the problem lies within the matrix count, where there is no further space for savings as all of the savings stemming from $u(\alpha z, z) \leq 1/m$ are used up by the fact that we already have to consider matrices of determinant $\leq m$. Any additional increase in the size of the determinant will thus automatically increase the bound on the matrix count and subsequently the geometric side of (11) by a considerable amount.

3. The Weil representation and theta series

3.1 Inner forms of SL2

Let B be a quadratic central simple algebra over $\mathbb{Q}$ and for each place v denote $B_v \coloneqq B \otimes \mathbb{Q}_v$. We define the affine algebraic group $\mathbf{G}$ over $\mathbb{Q}$ as representing the group functor

$$\mathbf{G}(L) = \{x \in B \otimes L \mid \mathrm{Nr}(x) = 1\}$$

for all $\mathbb{Q}$-algebras L. The group $\mathbf{G}$ is an inner form of $\mathbf{SL}_2$, and all inner forms of $\mathbf{SL}_2$ over $\mathbb{Q}$ arise this way.

Fix a maximal order $R \subset B$ and define R_v to be the completion of R in B_v. For each finite place v the order R_v is maximal in B_v. For $v < \infty$ set $K_v \coloneqq R_v^1 < \mathbf{G}(\mathbb{Q}_v)$ to be the group of norm 1 elements in R_v. If B splits over $\mathbb{R}$ we fix once and for all an isomorphism $B_\infty \simeq \mathrm{Mat}_{2\times 2}(\mathbb{R})$ and use it to identify the two spaces. We then set $K_\infty = \mathbf{SO}_2(\mathbb{R})$ if B splits over $\mathbb{R}$ and $K_\infty = \mathbf{G}(\mathbb{R})$ otherwise. For almost all v the group K_v is a hyperspecial maximal compact subgroup of $\mathbf{G}(\mathbb{Q}_v)$. We define $\mathbf{G}(\mathbb{A})$ as the restricted direct product

$$\mathbf{G}(\mathbb{A}) :- \left\{(g_v)_v \in \prod_{v\leq\infty} \mathbf{G}(\mathbb{Q}_v) \mid g_v \in K_v \text{ for almost all } v\right\}.$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

3.2 Normalization of Haar measures

For a linear algebraic group $\mathbf{L}$ defined over $\mathbb{Q}$ we denote $[\mathbf{L}(\mathbb{A})] \coloneqq \mathbf{L}(\mathbb{Q})\backslash\mathbf{L}(\mathbb{A})$. Assume $[\mathbf{L}(\mathbb{A})]$ is of finite volume. We shall always integrate with respect to the probability Haar measure on $[\mathbf{L}(\mathbb{A})]$. Let $U < \mathbf{L}(\mathbb{A}_f)$ be a compact open subgroup. Then, $\mathbf{L}(\mathbb{R})$ acts on $[\mathbf{L}(\mathbb{A}))]_U \coloneqq \mathbf{L}(\mathbb{Q})\backslash\mathbf{L}(\mathbb{A})/U$ with finitely many orbits [Bor63], and $[\mathbf{L}(\mathbb{A})]_U \simeq \bigsqcup_i \Gamma_i\backslash\mathbf{L}(\mathbb{R})$ with $\Gamma_i < \mathbf{L}(\mathbb{R})$ finitely many lattices. On $[\mathbf{L}(\mathbb{A})]_U$ we integrate with respect to the push-forward of the probability Haar measure on $[\mathbf{L}(\mathbb{A})]$. This measure is evidently an $\mathbf{L}(\mathbb{R})$-invariant probability measure. If $[\mathbf{L}(\mathbb{A})]_U \simeq \Gamma\backslash\mathbf{L}(\mathbb{R})$ is a single $\mathbf{L}(\mathbb{R})$-orbit, then this measure is the probability Haar measure on $\Gamma\backslash\mathbf{L}(\mathbb{R})$.

On $\mathbf{SL}_2(\mathbb{R})/\mathbf{SO_2}(\mathbb{R})$ and $\mathbf{Spin}_3(\mathbb{R})/\mathbf{SO}_2(\mathbb{R})$ we fix the standard Haar measures corresponding to the volume form of Gaussian curvature ± 1 on the hyperbolic plane and the 2-sphere. We fix the unique Haar measures on $\mathbf{SL}_2(\mathbb{R})$ and $\mathbf{Spin}_3(\mathbb{R})$ whose push-forward to the symmetric space coincides with the measure above.

On $\mathbf{SL}_2(\mathbb{Q}_p)$ and $\mathbf{PGL}_2(\mathbb{Q}_p)$ we fix the Haar measure giving volume 1 to $\mathbf{SL}_2(\mathbb{Z}_p)$ and $\mathbf{PGL}_2(\mathbb{Z}_p)$, respectively. Let $\mathbb{D}_p$ be the unique ramified quaternion algebra over $\mathbb{Q}_p$ with ring of integers $\mathcal{O}(\mathbb{D}_p)$. Denote by $\mathbb{D}_p^{(1)}$ the group of norm 1 element in $\mathbb{D}_p$. We fix the Haar measures on $\mathbb{D}_p^{(1)}$, $\mathbb{Q}_p^\times\backslash\mathbb{D}_p^\times$ that give volume 1 to the compact open subgroups $\mathcal{O}(\mathbb{D}_p)\cap\mathbb{D}_p^{(1)}$, $\mathbb{Z}_p^\times\backslash\mathcal{O}(\mathbb{D}_p)^\times$ respectively. These choices fix a Haar measure $m_{\mathbf{G}(\mathbb{Q}_p)}$ on $\mathbf{G}(\mathbb{Q}_p)$ for all primes p.

The product of the local Haar measures $m_{\mathbf{G}(\mathbb{Q}_p)}$ at all primes p induce a Haar measure on $\mathbf{G}(\mathbb{A}_f) = \prod_p' \mathbf{G}(\mathbb{Q}_p)$, which we call the *unnormalized* Haar measure on $\mathbf{G}(\mathbb{A}_f)$. Similarly, we call the product of the fixed Haar measure on $\mathbf{G}(\mathbb{R})$ with the unnormalized Haar measure on $\mathbf{G}(\mathbb{A}_f)$, the *unnormalized* Haar measure on $\mathbf{G}(\mathbb{A})$. The unnormalized Haar measure on $\mathbf{G}(\mathbb{A})$ is necessarily proportional to the covolume 1 measure, but they are *not* equal. Our local measure normalization forces $m_{\mathbf{G}(\mathbb{Q}_p)}(K_p) = 1$ for all primes p, hence the volume of $[\mathbf{G}(\mathbb{A})]$ with respect to the unnormalized measure is not 1, but rather the sum of the volumes $m_{\mathbf{G}(\mathbb{R})}(R_i^{(1)}\backslash\mathbf{G}(\mathbb{R}))$ for orders $R_i \subset B$ representing all the classes in the class set[4] of R, where R is the maximal order from above. Denote by $\varrho_{\mathbf{G}}$ the volume of $[\mathbf{G}(\mathbb{A})]$ with respect to the unnormalized measure. In the indefinite case, the class number is 1 and the volume is $\varrho_{\mathbf{G}} = (\pi/3)\varphi(D_B)$ (see [Voi21, Theorem 39.1.2]). Exactly the same formula holds in the definite case, due to the Eichler mass formula [Voi21, Theorem 25.1.1]. We henceforth fix the Haar measure $m_{\mathbf{G}(\mathbb{A}_f)}$ on $\mathbf{G}(\mathbb{A}_f)$ to be the measure induced by $\varrho_{\mathbf{G}}^{-1}\prod_p m_{\mathbf{G}(\mathbb{Q}_p)}$. The product $m_{\mathbf{G}(\mathbb{R})} \times m_{\mathbf{G}(\mathbb{A}_f)}$ is the co-volume 1 Haar measure. The same discussion applies *mutatis mutandi* to $\mathbf{SL}_2$.

Note that we have several normalizations of the Haar measure on $\mathbf{G}(\mathbb{R})$. When integrating over a quotient by a lattice $\Gamma\backslash\mathbf{G}(\mathbb{R})$ we always use the co-volume 1 Haar measure. When integrating over $\mathbf{G}(\mathbb{R})$ we use the standard measure $m_{\mathbf{G}(\mathbb{R})}$ which is not a co-volume 1 measure in general. The discrepancy is accounted for by the factor $\varrho_{\mathbf{G}}^{-1}$ in the Haar measure of $\mathbf{G}(\mathbb{A}_f)$. The same discussion applies to $\mathbf{SL}_2$.

3.3 Local Weil representation

In this section, the field $F = \mathbb{Q}_v$ is a completion of $\mathbb{Q}$ at a place v, then B_v is a quadratic central simple algebra over F, i.e. $B_v = \mathrm{Mat}_{2\times 2}(F)$ or B_v is the unique quadratic division algebra over F. Denote by $x \mapsto {}^\iota x$ the canonical involution on B_v. In the split case, the involution sends a matrix to its adjugate. Denote the reduced norm on B_v by Nr and the reduced trace by Tr. We shall also fix a unitary additive character $\psi_v \colon F \to \mathbb{C}^\times$. In this section, we recall the construction and elementary properties of the Weil representation.

[4] That is, all orders everywhere locally conjugate to R by a norm 1 element, where two orders are equivalent if they are globally conjugate by a rational norm 1 element.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

The vector space B_v is endowed with an additive Haar measure. For an integrable function $M\colon B_v \to \mathbb{C}$, we define the Fourier transform by

$$\mathscr{F}M(x) = \int M(y)\psi_v(\langle x, y\rangle)\, dy,$$

where the bilinear form $\langle\,,\rangle$ is defined by

$$\langle x, y\rangle := \mathrm{Tr}(x^\iota y).$$

Note that this is twice the polarization of the norm quadratic form, i.e. $\langle x, x\rangle = 2\,\mathrm{Nr}\, x$. We normalize the measure on B_v so that it is Fourier self-dual, i.e. $\mathscr{F}^2M(x) = M(-x)$ for a Schwartz function M.

If v is non-archimedean denote by Ω_v the space of Schwartz–Bruhat functions on B_v, i.e. locally constant functions of compact support. At the archimedean place, we need to consider a space that differs from the space of Schwartz functions because the Bergman kernel does not arise from a Schwartz function. To construct Ω_∞, we will start first with a larger space $L^2(B_\infty)$ and then restrict the Weil representation to a subspace Ω_∞ to be defined later.

The Weil representation of $\mathbf{SL}_2(F)$ on Ω_v, $L^2(B_\infty)$ satisfies

$$\rho\left(\begin{pmatrix}1 & \sigma\\ 0 & 1\end{pmatrix}\right)M(x) = \psi_v\big(\sigma\,\mathrm{Nr}(x)\big)M(x),$$

$$\rho\left(\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}\right)M(x) = |\lambda|_v^2 M(\lambda x),$$

$$\rho\left(\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\right)M(x) = \gamma\mathscr{F}M(x),$$

where $\gamma = 1$ if B_v is split and $\gamma = -1$ otherwise. For a proof that this defines a representation, see [JL70, § 1.1].

Note that the representation depends on the choice of an additive character ψ_v. We will usually suppress this dependence in the notation, but when we need to keep track of the character we shall write ρ_{ψ_v}. Because $\mathbb{Q}_v$ is Fourier self-dual, all non-trivial additive characters are of the form ${}^{\varpi}\psi_v(a) = \psi_v(a\varpi)$ for some $\varpi \in \mathbb{Q}_v^\times$. We see that

$$\rho_{{}^{\varpi}\psi_v}(g) = \rho_{\psi_v}(\mathrm{diag}(\varpi, 1)g\,\mathrm{diag}(\varpi, 1)^{-1}). \tag{17}$$

Lemma 3.1. *Let $\mathrm{O}(B_v, \mathrm{Nr})$ be the group of linear transformations preserving the norm form, this group acts on functions by $u.M(x) = M(u^{-1}x)$. The action of the orthogonal group $\mathrm{O}(B_v, \mathrm{Nr})$ commutes with the action of $\mathbf{SL}_2(F)$ via ρ.*

Proof. It is sufficient to verify the claim for each of the formulae above. The action of the upper triangular matrices commutes with the action of any linear transformation that preserves the norm. The Fourier transform intertwines the action of $L \in \mathrm{GL}(B_v)$ with the action of $|\det L|_v^{-1}{}^tL^{-1}$. Hence, it commutes with orthogonal transformations. □

Corollary 3.2. *The ρ action of $\mathbf{SL}_2(F)$ commutes with the right and left actions of $\mathbf{G}(F)$ by multiplication. Moreover, the ρ action commutes with the $B^\times$-action by conjugation.*

Proof. The actions of $B^\times$ and $\mathbf{G}(F)$ preserve the norm form, hence they factor through the orthogonal group. □

Lemma 3.3. *The Weil representation is a continuous unitary representation of $\mathbf{SL}_2(F)$ on Ω_v, $L^2(B_\infty)$.*

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Proof. This is established by Weil [Wei64] for the space of Schwartz or Schwartz–Bruhat functions. The same proof works for $L^2(B_\infty)$. □

3.4 The Archimedean Weil representation

To construct the appropriate subspace $\Omega_\infty \subset L^2(B_\infty)$, we will use a method based on the work of Vignéras [Vig77]. We define the Laplacian Δ on B_∞ as the Fourier multiplier operator with symbol $-4\pi^2 \operatorname{Nr}$. Write the archimedean additive character as $\psi_\infty(a) = \exp(2\pi i a\varpi)$ and consider the partial differential equation (PDE)

$$-\Delta M(x) + \omega^2 \operatorname{Nr}(x) M(x) = \frac{\omega m}{\varpi} M(x), \tag{18}$$

where $m \in \mathbb{Z}$ and $\omega > 0$. This is nothing but the PDE for energy eigenstates of four independent quantum harmonic oscillators with total energy $(\omega\varpi)m$ and angular frequency $\omega\varpi$. We call m the *quantum number* of the equation and we denote by $V_{m,\omega}$ the L^2-closure of the space of Schwartz solutions to (18). Note that unlike the standard harmonic oscillator, the individual oscillators may have either positive or negative energy depending on the signature of the quadratic form Nr.

We fix henceforth $\psi_\infty(a) = \exp(2\pi i a)$, i.e. $\varpi = 1$. Consider the densely defined linear operator $L_\omega \colon L^2(B_\infty) \to L^2(B_\infty)$ given by $L_\omega[M] = -\Delta M + \omega^2 \operatorname{Nr}(x) \cdot M$ with the domain of Schwartz functions $D(L_\omega) = \mathcal{S}(B_\infty)$. Then, L_ω is real, i.e. $\langle L_\omega[M], M\rangle \in \mathbb{R}$ for all $M \in \mathcal{S}(B_\infty)$. Hence, L_ω is symmetric. For explicitness, we state the following classical linear algebra lemma.

LEMMA 3.4. *The spaces $\{V_{m,\omega}\}_{m\in\mathbb{Z}}$ are mutually orthogonal.*

Proof. It is enough to show that if M, M' are Schwartz solutions to (18) with quantum numbers $m \neq m'$, then $\langle M, M'\rangle = 0$. Because the operator L_ω is symmetric, we have $\langle \omega m M, M'\rangle = \langle L[M], M'\rangle = \langle M, L[M']\rangle = \langle M, \omega m' M'\rangle$. We deduce that $\langle M, M'\rangle = 0$ in the usual fashion. □

LEMMA 3.5. *Let $k_\theta := \left(\begin{smallmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{smallmatrix}\right) \in \mathbf{SO}_2(\mathbb{R})$ and set $\psi_\infty(a) = \exp(2\pi i a)$. Then, for every $M \in V_{m,2\pi}$ we have*

$$\big(\rho(k_\theta).M\big)(x) = e^{im\theta} M(x).$$

Moreover, $L^2(B_\infty) = \bigoplus_{m\in\mathbb{Z}} V_{m,2\pi}$. Therefore, $V_{m,2\pi}$ is the $\big(\rho(\mathbf{SO}_2(\mathbb{R})), e^{im\theta}\big)$-isotypic subspace of $L^2(B_\infty)$.

The idea to use the one-dimensional Hermite functions in the proof of the lemma has been suggested to us by J. Wunsch.

Proof. We establish first the direct sum decomposition. Recall that we need the Laplacian in (18) to be defined consistently as having Fourier symbol $-4\pi^2 \operatorname{Nr}$. Choose a coordinate system $x_1, \ldots, x_4$ for B_∞ such that $\operatorname{Nr}(x) = \sum_{i=1}^4 \epsilon_i x_i^2$ with $\epsilon_i \in \{\pm 1\}$. The Laplacian for our fixed character can be written in this coordinate system as

$$\frac{1}{4} \sum_{i=1}^4 \epsilon_i \frac{\partial^2}{\partial x_i^2}.$$

The $1/4$ factor appears because the Fourier transform is defined with respect to the bilinear form $\sum_{i=1}^4 2\epsilon_i x_i y_i$. The space of solutions to the one-dimensional quantum harmonic oscillator with angular frequency 4π, $n \in \mathbb{Z}_{\geq 0}$,

$$-\tfrac{1}{4} f''(x) + 4\pi^2 x^2 f(x) = (2n+1)\pi f(x)$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

is one-dimensional and the L^2-normalized solution is

$$f_n(x) \coloneqq \frac{1}{\sqrt{2^{n-1}n!}} \exp\left(-2\pi x^2\right) H_n(2\sqrt{\pi}x),$$

where H_n are the physicist's Hermite polynomials. Moreover, these solutions form an orthonormal basis of the Hilbert space $L^2(\mathbb{R})$. Define for every $\underline{n} = (n_1, n_2, n_3, n_4) \in \mathbb{Z}^4_{\geq 0}$ the function $M_{\underline{n}} \colon B_\infty \to \mathbb{C}$ by

$$M_{\underline{n}}(x) = \prod_{j=1}^{4} f_{n_j}(x_j) = \prod_{j=1}^{4} \frac{1}{\sqrt{2^{n_j-1}n_j!}} \exp\left(-2\pi x_j^2\right) H_{n_j}(2\sqrt{\pi}x_j),$$

where $x_1, \ldots, x_4$ are the normal form coordinates for the quadratic form $\mathrm{Nr}(x)$. Because $L^2(B_\infty) \simeq L^2(\mathbb{R})^{\otimes 4}$, we deduce that the functions $M_{\underline{n}}$ form an orthonormal basis of $L^2(B_\infty)$. These are Schwartz functions, and a separation of variables computation shows that $M_{\underline{n}}$ solves (18) with[5] $2m = \sum_{j=1}^4 \epsilon_j(2n_j + 1)$. This and Lemma 3.4 establish that $L^2(B_\infty) = \bigoplus_{m\in\mathbb{Z}} V_{m,2\pi}$ as claimed.

We need to prove that if $M \in V_{m,2\pi}$ then $\rho(k_\theta).M = e^{im\theta}M$ for all $\theta \in [0, 2\pi)$. By continuity of the Weil representation it is enough to establish this for Schwartz functions. Because Schwartz functions are smooth vectors for the Weil representation it is sufficient to show $(d/d\theta)\rho(k_\theta).M = im\big(\rho(k_\theta).M\big)$. Because the group $\mathbf{SO}_2(\mathbb{R})$ is abelian it is enough to verify this ordinary differential equation (ODE) at $\theta = 0$. The formula $k_\theta = \exp(\theta w)$ for $w = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ implies that the ODE at $\theta = 0$ is equivalent to

$$d\rho(w).M = imM, \tag{19}$$

where $d\rho$ is the Lie algebra representation of $\mathfrak{sl}_2(\mathbb{R})$ on $\mathcal{S}(B_\infty)$ differentiated from the Weil representation of $\mathbf{SL}_2(\mathbb{R})$. Using the definition of the Weil action for upper diagonal unipotent matrices one easily computes that $\big(d\rho\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right).M\big)(x) = 2\pi i\,\mathrm{Nr}(x)M(x)$. The formula $\left(\begin{smallmatrix} 0 & 0 \\ -1 & 0 \end{smallmatrix}\right) = w^{-1}\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)w$, then implies that $d\rho\left(\begin{smallmatrix} 0 & 0 \\ -1 & 0 \end{smallmatrix}\right).M = (1/2\pi i)\Delta M$, and $(d\rho(w).M)(x) = (1/2\pi i)\Delta M(x) + 2\pi i\,\mathrm{Nr}(x)M(x)$. Thus, (19) is equivalent to (18). □

COROLLARY 3.6. *Let $M \in V_{m,\omega}$ for arbitrary $\omega > 0$ and fix $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathbf{SL}_2(\mathbb{R})$. Then,*

$$(\rho(g).M)(x) = \frac{2\pi}{\omega}\frac{1}{D^2}\psi_\infty\left(\frac{bd\frac{2\pi}{\omega} + ac\frac{\omega}{2\pi}}{D^2}\,\mathrm{Nr}\,x\right)\left(\sqrt{\frac{2\pi}{\omega}}\frac{d}{D} - i\sqrt{\frac{\omega}{2\pi}}\frac{c}{D}\right)^m M\left(\sqrt{\frac{2\pi}{\omega}}\frac{x}{D}\right),$$

where $D = \sqrt{c^2(\omega/2\pi) + d^2(2\pi/\omega)}$.

Proof. We use the notation $a \coloneqq \mathrm{diag}\left(\sqrt{2\pi/\omega}, \sqrt{\omega/2\pi}\right)$ and write $\rho(g).M = \rho(ga^{-1})\rho(a).M$. The Iwasawa decomposition of ga^{-1} is

$$ga^{-1} = \begin{pmatrix} 1 & \frac{bd\frac{2\pi}{\omega} + ac\frac{\omega}{2\pi}}{D^2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{D} & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} \sqrt{\frac{2\pi}{\omega}}\frac{d}{D} & -\sqrt{\frac{\omega}{2\pi}}\frac{c}{D} \\ \sqrt{\frac{\omega}{2\pi}}\frac{c}{D} & \sqrt{\frac{2\pi}{\omega}}\frac{d}{D} \end{pmatrix}.$$

The function

$$(\rho(a).M)(x) = \frac{2\pi}{\omega}M\left(\sqrt{\frac{2\pi}{\omega}}x\right)$$

is a solution of (18) with angular frequency 2π. Hence, we can apply Lemma 3.5 to $\rho(ga^{-1})\rho(a).M$ and the Iwasawa decomposition of ga^{-1}. □

[5] The sum of four odd numbers $\sum_{j=1}^4 \epsilon_j(2n_j + 1)$ is always even.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

DEFINITION 3.7. Fix $m \in \mathbb{Z}$. We are now ready to define the subspace $\Omega_\infty < L^2(B_\infty)$. This space will depend on a choice of m. Recall that $V_{m,\omega}$ is the L^2-completion of the space of solutions of the quantum harmonic oscillator, (18), for a fixed $m \in \mathbb{Z}$ and $\omega > 0$. Define

$$\Omega_\infty \coloneqq \operatorname{Span}_{\mathbb{C}} \big\{ \psi_\infty(\sigma \operatorname{Nr}(x)) M(x) \mid \sigma \in \mathbb{R}, \exists \omega > 0 \colon M \in V_{m,\omega} \\ \text{and } \exists \delta > 0 \colon |M(x)| \ll (1 + \|x\|)^{-4-\delta} \big\}.$$

The span allows only for *finite* linear combinations. In other words, Ω_∞ is the space generated by orbits of functions in $\bigcup_{\omega>0} V_{m,\omega}$ satisfying a decay condition at infinity under the Weil action of unipotent matrices. The decay condition implies that any function in Ω_∞ is in $L^p(B_\infty)$ for all $p \geq 1$.

PROPOSITION 3.8. *The space Ω_∞ is invariant under the Weil representation and the action of $\mathrm{O}(B_\infty, \mathrm{Nr})$.*

Proof. The space $V_{m,\omega}$ is invariant under $\mathrm{O}(B_\infty, \mathrm{Nr})$ because (18) commutes with orthogonal transformations. Also, the decay condition is invariant under orthogonal transformations. Invariance under the Weil action follows from Corollary 3.6. □

Remark. We note that we may assume the functions in Ω_∞ to be continuous, since we may replace them with the Fourier inverse of its Fourier transform as it converges absolutely uniformly on compacta due to the decay condition.

3.5 The non-Archimedean Weil representation

We now describe the interaction between the Weil representation and Eichler orders in B_v for $v < \infty$. In this section, we fix a prime p and write $F = \mathbb{Q}_p$ and set v to be the associated place. For clarity of notation, we will write $B_p \coloneqq B_v$. We assume that $\psi_p = \psi_v$ is an unramified character.

DEFINITION 3.9. Let $\mathcal{R} \subset B_p$ be an order. Then, the dual lattice $\widehat{\mathcal{R}}$ is defined as

$$\widehat{\mathcal{R}} = \big\{ x \in B_p \mid \forall x \in \mathcal{R} \colon \operatorname{Tr}(x^\iota y) \in \mathbb{Z}_p \big\}.$$

We begin by discussing maximal orders.

DEFINITION 3.10. Set $U_0(p^n) < \mathbf{SL}_2(\mathbb{Z}_p)$ to be the congruence subgroup defined by

$$U_0(p^n) \coloneqq \begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ p^n\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix} \cap \mathbf{SL}_2(\mathbb{Z}_p).$$

LEMMA 3.11. *Let $\mathcal{R} \subset B_p$ be a maximal order. If B_p is split, then $\rho\big(\mathbf{SL}_2(\mathbb{Z}_p)\big).\mathbb{1}_{\mathcal{R}} = \mathbb{1}_{\mathcal{R}}$. If B_p is ramified, then $\rho\big(U_0(p)\big).\mathbb{1}_{\mathcal{R}} = \mathbb{1}_{\mathcal{R}}$.*

Remark. This lemma is closely related to Lemmata 7 and 10 of [Shi72].

Proof. All maximal orders in B_p are conjugate to each other by an element of $B_p^\times$. Because the Weil action commutes with conjugation, it is enough to prove the claim for a specific maximal order. Moreover, the group $\mathbf{SL}_2(\mathbb{Z}_p)$ is generated by the subgroup $\mathcal{P} < \mathbf{SL}_2(\mathbb{Z}_p)$ of upper triangular integral matrices and the involution w. The fact that $\mathbb{1}_{\mathcal{R}}$ is invariant under $\mathcal{P}$ follows because we have assumed ψ_v is unramified. If B_p is split, fix an isomorphism $B_p \simeq \mathrm{Mat}_{2\times 2}(\mathbb{Q}_p)$ and we need only show that $\mathbb{1}_{\mathrm{Mat}_{2\times 2}(\mathbb{Z}_p)}$ is invariant under the Fourier transform. This follows from the fact that $\mathbb{1}_{\mathbb{Z}_p}$ is invariant under the Fourier transform on $\mathbb{Q}_p$ with an unramified character.

If B_p is a division algebra, we need only show invariance under the element $\left(\begin{smallmatrix} 1 & 0 \\ p & 1 \end{smallmatrix}\right) = -w \left(\begin{smallmatrix} 1 & -p \\ 0 & 1 \end{smallmatrix}\right) w$. This element and the upper triangular integral matrices generate $U_0(p)$. Because of the duality of the Fourier transform, this is equivalent to showing that $\rho\left(\left(\begin{smallmatrix} 1 & -p \\ 0 & 1 \end{smallmatrix}\right)\right).\mathscr{F}\mathbb{1}_{\mathcal{R}} = \mathscr{F}\mathbb{1}_{\mathcal{R}}$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Let $E = F(\sqrt{a})/F$ be the unique unramified quadratic extension and write B_p as the cyclic algebra $\left(\frac{a,p}{\mathbb{Q}_p}\right)$ with the standard generators i, j, k and $i^2 = a$, $j^2 = p$ and $ij = -ji = k$. As usual, we identify E with the sub-ring $\mathbb{Q}_p + i\mathbb{Q}_p < B_p$. Denote by $\mathcal{O}_E$ the maximal order of E. Then, the unique maximal order of B_p is $\mathcal{R} \simeq \mathcal{O}_E + j\mathcal{O}_E$. The Fourier transform on B_p descends to the standard Fourier transform on E with an unramified character. Identifying $B_p \simeq E \times E$ via $a + jb \mapsto (a, b)$, we can write the Fourier self-dual measure on B_p in these coordinates as $p^{-1} m_E \times m_E$. The p^{-1} factor normalizes the measure to be self-dual.

The Fourier transform on E satisfies $\mathcal{F}\mathbb{1}_{\mathcal{O}_E} = \mathbb{1}_{\mathcal{O}_E}$. An explicit computation with the Fourier self-dual measure implies

$$\mathcal{F}\mathbb{1}_{\mathcal{O}_E + j\mathcal{O}_E} = p^{-1}\mathbb{1}_{\mathcal{O}_E + j^{-1}\mathcal{O}_E}.$$

Hence, $\operatorname{Nr} x \in p^{-1}\mathbb{Z}_p$ for all $x \in \operatorname{supp} \mathcal{F}\mathbb{1}_{\mathcal{R}}$, from which we deduce $\left(\rho\left(\left(\begin{smallmatrix} 1 & -p \\ 0 & 1 \end{smallmatrix}\right)\right).\mathcal{F}\mathbb{1}_{\mathcal{R}}\right)(x) = \psi_v(-p \operatorname{Nr} x)\left(\mathcal{F}\mathbb{1}_{\mathcal{R}}\right)(x) = \left(\mathcal{F}\mathbb{1}_{\mathcal{R}}\right)(x)$ and the claim follows. □

LEMMA 3.12. *Assume B_p is ramified and let $\mathcal{R} \subset B_p$ be the unique maximal order. Then, there is an isomorphism of finite abelian additive groups*

$$\jmath_v \colon \widehat{\mathcal{R}}/\mathcal{R} \to \mathbb{F}_{p^2}$$

such that $-p \operatorname{Nr} x \bmod p\mathbb{Z}_p \equiv \operatorname{Nr} \jmath_v(x)$ for all $x \in \widehat{\mathcal{R}}$. The norm on the right-hand side is the field norm $\mathbb{F}_{p^2} \to \mathbb{F}_p$.

Moreover, if j is a uniformizer of $\mathcal{R}$, then we can choose $\jmath_v$ so that the composite map

$$\mathcal{R}/j\mathcal{R} \xrightarrow{x \mapsto j^{-1}x} \widehat{\mathcal{R}}/\mathcal{R} \xrightarrow{\jmath_v} \mathbb{F}_{p^2}$$

is a field isomorphism.

Note that there are exactly two field isomorphisms $\mathcal{R}/j\mathcal{R} \to \mathbb{F}_{p^2}$ and they differ by post-composition with the Frobenius, i.e. by the action of the Galois group. If $f \colon \mathcal{R}/j\mathcal{R} \to \mathbb{F}_{p^2}$ is such an isomorphism, then $f(jxj^{-1}) = f(x)^p$, i.e. conjugation by j is intertwined with the Frobenius. Hence, the composition of $\mathcal{R}/j\mathcal{R} \xrightarrow{x \mapsto xj^{-1}} \widehat{\mathcal{R}}/\mathcal{R} \xrightarrow{\jmath_v} \mathbb{F}_{p^2}$ is necessarily also a field isomorphism differing from the original one by post-composition with the Frobenius.

Proof. Let j be a uniformizer in $\mathcal{R}$. The field norm on $\mathcal{R}/j\mathcal{R} \simeq \mathbb{F}_{p^2}$ coincides with the reduction mod$p$ of the reduced norm in $\mathcal{R}$. This can be seen by taking a subfield $E \subset B_p$, such that E is an unramified quadratic extension of F that splits B_p. In these coordinates, we can write

$$B_p = E + jE, \quad \mathcal{R} = \mathcal{O}_E + j\mathcal{O}_E, \quad \widehat{\mathcal{R}} = \mathcal{O}_E + j^{-1}\mathcal{O}_E. \tag{20}$$

Now, the restriction of the reduced norm in B_p to E coincides with the field norm on E and $\mathcal{R}/j\mathcal{R} = \mathcal{O}_E/p\mathcal{O}_E \simeq \mathbb{F}_{p^2}$.

Observe that $\widehat{\mathcal{R}} = j^{-1}\mathcal{R}$ and $\widehat{\mathcal{R}}/\mathcal{R} \simeq \mathcal{R}/j\mathcal{R} \simeq \mathbb{F}_{p^2}$. The last isomorphism is a field isomorphism and, thus, it commutes with taking norms. The first isomorphism is via the map $x \mapsto jx$ and $\operatorname{Nr}(jx) = -p \operatorname{Nr} x$. This establishes the claimed formula for norms. To summarize, the map $\jmath_v$ may thus be given explicitly as

$$\jmath_v(a + j^{-1}b) = b \mod p\mathcal{O}_E, \tag{21}$$

for $a + j^{-1}b \in \widehat{\mathcal{R}}$ with $a, b \in \mathcal{O}_E$. □

PROPOSITION 3.13. *Assume B_p is ramified and let $\mathcal{R} \subset B_p$ be the unique maximal order. Then,*

$$\rho(\mathbf{SL}_2(\mathbb{Z}_p)).\mathbb{1}_{\mathcal{R}} = \left\{ -p^{-1}\psi\left(\frac{\operatorname{Nr} x}{t}\right)\mathbb{1}_{\widehat{\mathcal{R}}} \mid 0 < t < p \right\} \cup \left\{ \mathbb{1}_{\mathcal{R}}, -p^{-1}\mathbb{1}_{\widehat{\mathcal{R}}} \right\}.$$

Moreover, each of the functions above corresponds to a single coset in $\mathbf{SL}_2(\mathbb{Z}_p)/U_0(p)$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Remark 3.14. Because $\operatorname{Nr} x \in p^{-1}\mathbb{Z}_p$ and ψ is unramified, we can rewrite the result above as

$$\rho(\mathbf{SL}_2(\mathbb{Z}_p)).\mathbb{1}_{\mathcal{R}} = \{\mathbb{1}_{\mathcal{R}}\} \cup \left\{-p^{-1}\psi(t \operatorname{Nr} x)\mathbb{1}_{\widehat{\mathcal{R}}} \mid t \in {}^{\mathbb{Z}}/_{p\mathbb{Z}}\right\}.$$

Proof. As in the previous lemma, we put coordinates on B_p corresponding to the cyclic algebra $\left(\frac{a,p}{\mathbb{Q}_p}\right)$ where $E = F(\sqrt{a})/F$ is the unique unramified quadratic extension, see (20). We shall make use of the explicit map $\jmath_v$ given by (21). It is an isomorphism of abelian groups $\widehat{\mathcal{R}}/\mathcal{R} \to \mathbb{F}_{p^2}$ and $-p \operatorname{Nr} x \bmod p\mathbb{Z}_p \equiv \operatorname{Nr} \jmath_v(x)$ for all $x \in \widehat{\mathcal{R}}$, where the norm on the right-hand side is the field norm $\mathbb{F}_{p^2} \to \mathbb{F}_p$.

For each $\alpha \in \mathbb{F}_{p^2}$, fix a representative $x_\alpha \in \jmath_v^{-1}(\alpha)$. Then,

$$\widehat{\mathcal{R}} = \bigsqcup_{\alpha \in \mathbb{F}_{p^2}} (x_\alpha + \mathcal{R}). \tag{22}$$

Explicitly, for each $\alpha \in \mathbb{F}_{p^2}$, we take $x_\alpha = j^{-1}\check{\alpha}$, where $\check{\alpha} \in \mathcal{O}_E$ satisfies $\check{\alpha} \bmod p\mathcal{O}_E = \alpha$. The duality between $\mathcal{R}$ and $\widehat{\mathcal{R}}$ implies that $\operatorname{Nr}(x_\alpha + \mathcal{R}) \subset \operatorname{Nr} x_\alpha + \mathbb{Z}_p$.

Recall that $\mathbb{1}_{\mathcal{R}}$ is $\rho(U_0(p))$-invariant. Hence, we need only calculate the action of each representative of $\mathbf{SL}_2(\mathbb{Z}_p)/U_0(p)$ on $\mathbb{1}_{\mathcal{R}}$. A set of representatives is given by w, $\left(\begin{smallmatrix}1 & 0\\ t & 1\end{smallmatrix}\right)$, $0 \le t < p$. The action of $\rho(w)$ is the Fourier transform (multiplied by $\gamma = -1$) and we have already seen in the proof of Lemma 3.11 that

$$\rho(w).\mathbb{1}_{\mathcal{R}} = -p^{-1}\mathbb{1}_{\widehat{\mathcal{R}}}.$$

Write $x = a + j^{-1}b$ for $x \in \widehat{\mathcal{R}}$. Now, we compute the action of $\left(\begin{smallmatrix}1 & 0\\ t & 1\end{smallmatrix}\right) = -w\left(\begin{smallmatrix}1 & -t\\ 0 & 1\end{smallmatrix}\right)w$ using[6] (22)

$$\begin{aligned}\left(\rho\left(\begin{pmatrix}1 & 0\\ t & 1\end{pmatrix}\right).\mathbb{1}_{\mathcal{R}}\right)(x) &= -p^{-1}\rho(w).\big(\psi(-t \operatorname{Nr} y)\mathbb{1}_{\widehat{\mathcal{R}}}(y)\big)(x)\\ &= -p^{-1}\rho(w).\left(\sum_{\alpha \in \mathbb{F}_{p^2}} \psi(-t \operatorname{Nr} x_\alpha)\mathbb{1}_{x_\alpha + \mathcal{R}}\right)(x)\\ &= p^{-2}\mathbb{1}_{\widehat{\mathcal{R}}}(x) \sum_{\alpha \in \mathbb{F}_{p^2}} \psi(-t \operatorname{Nr} x_\alpha + \langle x, x_\alpha\rangle).\end{aligned}$$

If $t \neq 0$, then the sum above becomes

$$\begin{aligned}\sum_{\alpha \in \mathbb{F}_{p^2}} \psi\left(\frac{\operatorname{Nr} x - \operatorname{Nr}(x - tx_\alpha)}{t}\right) &= \psi\left(\frac{\operatorname{Nr} x}{t}\right) \sum_{\alpha \in \mathbb{F}_{p^2}} \psi\left(\frac{\operatorname{Nr}(b - t\alpha)}{tp}\right) = \psi\left(\frac{\operatorname{Nr} x}{t}\right) \sum_{\alpha_0 \in \mathbb{F}_{p^2}} \psi\left(\frac{\operatorname{Nr} \alpha_0}{tp}\right)\\ &= \psi\left(\frac{\operatorname{Nr} x}{t}\right)\left((p+1) \sum_{\beta \in \mathbb{F}_p^\times} \psi\left(\frac{\beta}{p}\right) + 1\right) = -p\psi\left(\frac{\operatorname{Nr} x}{t}\right).\end{aligned}$$

We have used the fact that every element of $\mathbb{F}_p^\times$ is the norm of exactly $p+1$ elements in $\mathbb{F}_{p^2}^\times$. This establishes the claim for $0 < t < p$. For $t = 0$, the sum becomes

$$\sum_{\alpha \in \mathbb{F}_{p^2}} \psi\left(-\frac{\operatorname{Tr}(\alpha b)}{p}\right) = \begin{cases} p^2, & b \equiv 0 \bmod p\mathcal{O}_E,\\ 0, & \text{otherwise.}\end{cases}$$

In addition, $\mathbb{1}_{\widehat{\mathcal{R}}}(x)\delta_{b \equiv 0 \bmod p\mathcal{O}_E}(x) = \mathbb{1}_{\mathcal{R}}(x)$. Of course, the case of $t = 0$ is actually trivial to compute because it corresponds to the identity representative.

[6] Note that $\rho(-1).M(x) = M(-x)$ and that our function is symmetric, so -1 acts trivially.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

□

Lemma 3.15. *Let $M\colon B_p \to \mathbb{C}$ be a Schwartz–Bruhat function. Then, there is an open subgroup $U_M < \mathbf{SL}_2(\mathbb{Z}_p)$ such that $\rho(U_M).M = M$. In particular, $\rho\big(\mathbf{SL}_2(\mathbb{Z}_p)\big).M$ is a finite set.*

Proof. Fix a maximal order $\mathcal{R} \subset B_p$. We first claim that for every Schwartz–Bruhat function $M\colon B_p \to \mathbb{C}$ there is some diagonal matrix $a \in \mathbf{SL}_2(\mathbb{Q}_p)$ such that $\rho(a).M$ is a linear combination of translates of $\mathbb{1}_{\mathcal{R}}$. Equivalently $\rho(a).M(x+\mathcal{R}) = \rho(a).M(x)$. Because M is Schwartz–Bruhat, there is a small neighborhood of the origin $\mathcal{V} \subset B_p$, such that $M(x+\mathcal{V}) = M(x)$. There is $k \geq 1$ such that $p^k\mathcal{R} \subset \mathcal{V}$. The function $x \mapsto M(p^k x)$ is invariant under translations by $\mathcal{R}$. Set $a = \operatorname{diag}(p^k, p^{-k})$, then $\rho(a).M$ is a linear combination of translates of $\mathbb{1}_{\mathcal{R}}$ as claimed.

Fix $b \in B_p$ and consider the group

$$\mathcal{A}(b) = \big\{\operatorname{diag}(u,u^{-1})\colon u \in \mathbb{Z}_p^\times \text{ and } bu - b \in \mathcal{R}\big\}.$$

Then, $\mathcal{A}(b)$ is an open subgroup of the diagonal group in $\mathbf{SL}_2(\mathbb{Z}_p)$ and $\rho(\mathcal{A}(b)).\mathbb{1}_{b+\mathcal{R}} = \mathbb{1}_{b+\mathcal{R}}$. Taking a finite intersection of such subgroups we find an open subgroup $\mathcal{A}_0$ of the diagonal group of $\mathbf{SL}_2(\mathbb{Z}_p)$, such that $\rho(\mathcal{A}_0 a).M = \rho(a).M$. Hence, M is invariant under $\rho(\mathcal{A}_0)$.

In a similar fashion, let $k_b \geq 0$ such that $p^{k_b} b \in \mathcal{R}$ and define $\mathcal{N}(b) = \left(\begin{smallmatrix} 1 & p^{k_b}\mathbb{Z}_p \\ 0 & 1 \end{smallmatrix}\right)$. Then, $\mathcal{N}(b)$ is an open subgroup of the upper-triangular unipotent group of $\mathbf{SL}_2(\mathbb{Z}_p)$ and $\rho(\mathcal{N}_b).\mathbb{1}_{b+\mathcal{R}} = \mathbb{1}_{b+\mathcal{R}}$. Taking a finite intersection of such subgroups we can find an open subgroup $\mathcal{N}_1'$ of the integral upper-triangular unipotent subgroup, such that $\rho(\mathcal{N}_1' a).M = \rho(a).M$. Set $\mathcal{N}_1 = a^{-1}\mathcal{N}_1' a \cap \mathbf{SL}_2(\mathbb{Z}_p)$. Then, $\rho(\mathcal{N}_1).M = M$ and $\mathcal{N}_1$ is an open subgroup of the upper unipotent integral group. Replacing M by $\rho(w).M$, we can find $\mathcal{N}_2$ such that $\rho(w^{-1}\mathcal{N}_2 w).M = M$, and $w^{-1}\mathcal{N}_2 w$ is an open subgroup of the lower-triangular integral unipotent group.

Set now U_M to be the group generated by $\mathcal{A}_0, \mathcal{N}_1, w^{-1}\mathcal{N}_2 w$. Then, U_M is an open subgroup of $\mathbf{SL}_2(\mathbb{Z}_p)$ and satisfies the requirements of the claim. □

Assume now $B_p \simeq \mathrm{Mat}_{2\times 2}(\mathbb{Q}_p)$ is split. Maximal orders in $\mathrm{Mat}_{2\times 2}(\mathbb{Q}_p)$ are endomorphism rings of lattices in $\mathbb{Q}_p^2$ and they are in one-to-one correspondence with the vertices of the Bruhat–Tits tree of $\mathbf{SL}_2(\mathbb{Q}_p)$.

Definition 3.16. An Eichler order in B_p of level p^n is an intersection of two maximal orders corresponding to two vertices in the Bruhat–Tits tree with distance n between them.

Lemma 3.17. *Let $\mathcal{R} \subset B_p$ be an Eichler order of level p^n. Then, $\rho\big(U_0(p^n)\big).\mathbb{1}_{\mathcal{R}} = \mathbb{1}_{\mathcal{R}}$.*

Proof. The action of $B_p^\times$ on the vertices of the Bruhat–Tits tree is transitive on pairs of vertices of a fixed distance,[7] thus it acts transitively by conjugation on the set of Eichler orders of a fixed level p^n. Because the conjugation action commutes with the Weil representation, it is enough to consider a single Eichler order of the form

$$\mathcal{R} = \begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ p^n\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix}.$$

Indeed, invariance of $\mathbb{1}_{\mathcal{R}}$ under upper-triangular integral matrices is immediate and it is enough to check invariance under the element $\left(\begin{smallmatrix} 1 & 0 \\ p^n & 1 \end{smallmatrix}\right) = -w\left(\begin{smallmatrix} 1 & -p^n \\ 0 & 1 \end{smallmatrix}\right)w$. Equivalently, we need to show

[7] This follows from the facts that the action of $\mathbf{SL}_2(\mathbb{Q}_p)$ is strongly transitive, i.e. it is transitive on pairs $(\mathcal{C}, \mathcal{A})$ where $\mathcal{C}$ is a chamber in the apartment $\mathcal{A}$, and that $\mathbf{PGL}_2(\mathbb{Q}_p)$ has an element which inverts the orientation of a single chamber.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

$\rho\left(\left(\begin{smallmatrix}1 & -p^n\\ 0 & 1\end{smallmatrix}\right)\right).\mathscr{F}\mathbb{1}_{\mathscr{R}} = \mathscr{F}\mathbb{1}_{\mathscr{R}}$. We can compute the Fourier transform of $\mathbb{1}_{\mathscr{R}}$ explicitly and arrive at

$$\mathscr{F}\mathbb{1}\begin{pmatrix}\mathbb{Z}_p & \mathbb{Z}_p\\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix} = p^{-n}\mathbb{1}\begin{pmatrix}\mathbb{Z}_p & p^{-n}\mathbb{Z}_p\\ \mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix} = p^{-n}\mathbb{1}_{\widehat{\mathscr{R}}}.$$

In particular, for all $x \in \operatorname{supp}\mathscr{F}\mathbb{1}_{\mathscr{R}}$, we have $\operatorname{Nr} x = \det x \in p^{-n}\mathbb{Z}_p$ and

$$\left(\rho\left(\begin{pmatrix}1 & -p^n\\ 0 & 1\end{pmatrix}\right).\mathscr{F}\mathbb{1}_{\mathscr{R}}\right)(x) = \psi_v(-p^n \operatorname{Nr} x)\big(\mathscr{F}\mathbb{1}_{\mathscr{R}}\big)(x) = \big(\mathscr{F}\mathbb{1}_{\mathscr{R}}\big)(x)$$

as necessary. □

LEMMA 3.18. *Let $\mathscr{R} = \mathscr{R}_1 \cap \mathscr{R}_2$ be an Eichler order of level p^n, where $\mathscr{R}_1$ and $\mathscr{R}_2$ are maximal orders. There is an isomorphism of additive abelian groups $\jmath_v\colon \widehat{\mathscr{R}}/\mathscr{R} \to \big(\mathbb{Z}/p^n\mathbb{Z}\big)^2$ such that*

$$\jmath_v(\mathscr{R}_1) = {}^{\mathbb{Z}}\!/_{p^n\mathbb{Z}} \times \{0\},$$
$$\jmath_v(\mathscr{R}_2) = \{0\} \times {}^{\mathbb{Z}}\!/_{p^n\mathbb{Z}},$$
$$\forall x \in \widehat{\mathscr{R}}\colon\ -p^n \operatorname{Nr}(x) \equiv \jmath_v(x)_1 \cdot \jmath_v(x)_2 \mod p^n.$$

Moreover, the isomorphism $\jmath_v$ is unique up to post-composition with the map $(b,c) \mapsto (bu, cu^{-1})$ for $u \in (\mathbb{Z}/p^n\mathbb{Z})^\times$.

Note that the isomorphism $\jmath_v$ depends not only on $\mathscr{R}$ but on an ordered choice of $\mathscr{R}_1$ and $\mathscr{R}_2$.

Proof. Because all local Eichler orders of fixed level are conjugate, it is enough to verify the claim for

$$\mathscr{R}_1 = \begin{pmatrix}\mathbb{Z}_p & \mathbb{Z}_p\\ \mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix}, \quad \mathscr{R}_2 = \begin{pmatrix}\mathbb{Z}_p & p^{-n}\mathbb{Z}_p\\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix}.$$

In this case, we have

$$\mathscr{R} = \begin{pmatrix}\mathbb{Z}_p & \mathbb{Z}_p\\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix}, \quad \widehat{\mathscr{R}} = \begin{pmatrix}\mathbb{Z}_p & p^{-n}\mathbb{Z}_p\\ \mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix}.$$

We define the abelian homomorphism $\jmath_v\colon \widehat{\mathscr{R}} \to \big(\mathbb{Z}/p^n\mathbb{Z}\big)^2$ explicitly as

$$\begin{pmatrix}a & b/p^n\\ c & d\end{pmatrix} \mapsto (c,b) \bmod p^n\mathbb{Z}_p.$$

A direct computation shows that this homomorphism has kernel $\mathscr{R}$ and that it satisfies the claimed properties.

This isomorphism is unique up to post-composition with a linear automorphism of the first and second coordinate of $\big(\mathbb{Z}/p^n\mathbb{Z}\big)^2$, i.e. a transformation of the form $(b,c) \mapsto (bu_1, cu_2)$ for $u_1, u_2 \in (\mathbb{Z}/p^n\mathbb{Z})^\times$. The requirement that the quadratic form $-p^n \operatorname{Nr}$ descends to the product form $(b,c) \mapsto b \cdot c$ forces $u_2 = u_1^{-1}$. □

Remark 3.19. The previous lemma implies that given two maximal orders $\mathscr{R}_1$, $\mathscr{R}_2$ the map $x \mapsto (\operatorname{ord}_p \jmath_v(x)_1, \operatorname{ord}_p \jmath_v(x)_2)$ from $\widehat{\mathscr{R}}$ to $\{0,1,\dots,n\}^2$ is uniquely defined.

DEFINITION 3.20. Let $\mathscr{R} \subset B_p$ be an Eichler order of level p^n. For every $p^k \mid p^n$ define the lattice

$$\widehat{\mathscr{R}}^{(p^k)} = \big\{x \in \widehat{\mathscr{R}} \mid \jmath_v(x) \equiv (0,0) \bmod p^k\big\}.$$

The definition of $\widehat{\mathscr{R}}^{(p^k)}$ does not depend on the choice of $\jmath_v$. Note that $\widehat{\mathscr{R}}^{(1)} = \widehat{\mathscr{R}}$ and $\widehat{\mathscr{R}}^{(p^n)} = \mathscr{R}$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Moreover, for $x \in \widehat{\mathcal{R}}^{(p^k)}$ define[8]

$$\nu_{p^k}(x) \coloneqq -p^{-k}\jmath_v(x)_1 \cdot p^{-k}\jmath_v(x)_2 \in {}^{\mathbb{Z}}/_{p^{n-k}\mathbb{Z}}.$$

Note that $\nu_{p^0}(x) \equiv p^n \operatorname{Nr} x \bmod p^n$.

Proposition 3.21. *Let $\mathcal{R} \subset B_p$ be an Eichler order of level p^n. Then,*

$$\rho\big(\mathbf{SL}_2(\mathbb{Z}_p)\big).\mathbb{1}_{\mathcal{R}} = \big\{\mathbb{1}_{\widehat{\mathcal{R}}}(x) \cdot p^{-n}\psi_v(-pt_0 \operatorname{Nr} x) \mid 0 < t_0 \le p^{n-1}\big\}$$
$$\cup \left\{\mathbb{1}_{\widehat{\mathcal{R}}\left(p^{\operatorname{ord}_p t}\right)}(x) \cdot p^{-(n-\operatorname{ord}_p t)}\psi_v\left(\frac{\nu_{p^{\operatorname{ord}_p t}}(x)}{p^{n-2\operatorname{ord}_p t}t}\right) \mid 0 < t \le p^n\right\}.$$

Moreover, each of the functions above corresponds to a single coset of $\mathbf{SL}_2(\mathbb{Z}_p)/U_0(p^n)$. For $t = p^n$ above, the phase is 1, hence the representative for $t = p^n$ is simply $\mathbb{1}_{\mathcal{R}}$.

Remark 3.22. Because ψ is unramified, we can rewrite the result above as

$$\rho\big(\mathbf{SL}_2(\mathbb{Z}_p)\big).\mathbb{1}_{\mathcal{R}} = \bigcup_{0<k\le n}\left\{\mathbb{1}_{\widehat{\mathcal{R}}^{(p^k)}}(x) \cdot p^{-(n-k)}\psi_v\left(\frac{u \cdot \nu_{p^k}(x)}{p^{n-k}}\right) \mid u \in \left({}^{\mathbb{Z}}/_{p^{n-k}\mathbb{Z}}\right)^{\times}\right\}$$
$$\cup \left\{\mathbb{1}_{\widehat{\mathcal{R}}}(x) \cdot p^{-n}\psi_v\big(t \operatorname{Nr} x\big) \mid t \in {}^{\mathbb{Z}}/_{p^n\mathbb{Z}}\right\}.$$

The set on the second line is analogous to the $k = 0$ case missing in the first line, but requires t to traverse the whole congruence subgroup, not just the units.

Proof. Again, as all Eichler orders are conjugate, the claim reduces to the case of

$$\mathcal{R} = \begin{pmatrix}\mathbb{Z}_p & \mathbb{Z}_p\\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix}, \quad \widehat{\mathcal{R}} = \begin{pmatrix}\mathbb{Z}_p & p^{-n}\mathbb{Z}_p\\ \mathbb{Z}_p & \mathbb{Z}_p\end{pmatrix},$$

and

$$\jmath_v\left(\begin{pmatrix}a & b/p^n\\ c & d\end{pmatrix}\right) = (c, b) \mod p^n\mathbb{Z}_p.$$

Because of Lemma 3.17, it is enough to compute $\rho(s).\mathbb{1}_{\mathcal{R}}$ for each of the representatives of $\mathbf{SL}_2(\mathbb{Z}_p)/U_0(p^n)$. To find these representatives, we first write representatives for $\mathbf{SL}_2(\mathbb{Z}_p)/U_0(p)$

$$\mathbf{SL}_2(\mathbb{Z}_p) = wU_0(p) \sqcup \bigsqcup_{0<t\le p}\begin{pmatrix}1 & 0\\ t & 1\end{pmatrix}U_0(p).$$

This decomposition follows from the fact that $U_0(p)$ is the stabilizer of an edge in the apartment of the diagonal torus in the Bruhat–Tits tree of $\mathbf{SL}_2(\mathbb{Q}_p)$ and the representatives above permute the $p+1$ neighbors of the vertex stabilized by $\mathbf{SL}_2(\mathbb{Z}_p)$.

Next, we find representatives for $U_0(p)/U_0(p^n)$ using their definition as congruence subgroups

$$U_0(p) = \bigsqcup_{0<t_0\le p^{n-1}}\begin{pmatrix}1 & 0\\ pt_0 & 1\end{pmatrix}U_0(p^n).$$

By combining, we arrive at

$$\mathbf{SL}_2(\mathbb{Z}_p) = w\bigsqcup_{0<t_0\le p^{n-1}}\begin{pmatrix}1 & 0\\ pt_0 & 1\end{pmatrix}U_0(p^n) \sqcup \bigsqcup_{0<t\le p^n}\begin{pmatrix}1 & 0\\ t & 1\end{pmatrix}U_0(p^n).$$

[8] This definition does not depend on the choice of $\jmath_v$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

We now compute explicitly the action of all representatives. We need to decompose $\widehat{\mathcal{R}}$ into $\mathcal{R}$-cosets

$$\widehat{\mathcal{R}} = \bigsqcup_{0\le\alpha,\beta<p^n} \begin{pmatrix} 0 & \alpha/p^n \\ \beta & 0 \end{pmatrix} + \mathcal{R}.$$

To simplify notation, we denote $x_{\alpha,\beta} \coloneqq \left(\begin{smallmatrix} 0 & \alpha/p^n \\ \beta & 0 \end{smallmatrix}\right)$, with $\jmath_v(x_{\alpha,\beta}) = (\beta, \alpha)$. The duality between $\widehat{\mathcal{R}}$ and $\mathcal{R}$ implies $\mathrm{Nr}(x_{\alpha,\beta} + \mathcal{R}) = \mathrm{Nr}\, x_{\alpha,\beta} + \mathbb{Z}_p = -\alpha\beta/p^n + \mathbb{Z}_p$. Write $\left(\begin{smallmatrix} 1 & 0 \\ t & 1 \end{smallmatrix}\right) = -w \left(\begin{smallmatrix} 1 & -t \\ 0 & 1 \end{smallmatrix}\right) w$ and $x = \left(\begin{smallmatrix} a & b/p^n \\ c & d \end{smallmatrix}\right)$. Then,

$$\begin{aligned}
\rho\left(\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}\right).\mathbb{1}_{\mathcal{R}}(x) &= p^{-n}\rho(w)\rho\left(\begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix}\right).\mathbb{1}_{\widehat{\mathcal{R}}}(x) = p^{-n}\rho(w).\big(\psi_v(-t\,\mathrm{Nr}\, x)\mathbb{1}_{\widehat{\mathcal{R}}}(x)\big) \\
&= p^{-n}\rho(w).\left(\sum_{0\le\alpha,\beta<p^n} \psi_v(t\alpha\beta/p^n)\mathbb{1}_{x_{\alpha,\beta}+\mathcal{R}}(x)\right) \\
&= \mathbb{1}_{\widehat{\mathcal{R}}}(x)\cdot p^{-2n} \sum_{0\le\alpha,\beta<p^n} \psi_v(t\alpha\beta/p^n + \langle x_{\alpha,\beta}, x\rangle) \\
&= \mathbb{1}_{\widehat{\mathcal{R}}}(x)\cdot p^{-2n} \sum_{0\le\alpha,\beta<p^n} \psi_v\left(\frac{-\alpha c - \beta b + t\alpha\beta}{p^n}\right).
\end{aligned}$$

Let $k = \mathrm{ord}_p\, t$. Then, summing first over α we deduce $p^k \mid c$ and summing first over β we see that $p^k \mid b$. Using this input, we can sum first over α and then over β to arrive at

$$\begin{aligned}
&\rho\left(\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}\right).\mathbb{1}_{\mathcal{R}}(x) \\
&\quad = \mathbb{1}_{\widehat{\mathcal{R}}}(x) p^{-(n-k)} \begin{cases} \psi_v\left(-\dfrac{(c/p^k)(b/p^k)}{(t/p^k)p^{n-k}}\right) = \psi_v\left(\dfrac{\nu_{p^k}(x)}{tp^{n-2k}}\right), & \jmath_v(x) \equiv (0,0) \bmod p^k, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

We need only deal now with representatives of the form $w\left(\begin{smallmatrix} 1 & 0 \\ pt_0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & -pt_0 \\ 0 & 1 \end{smallmatrix}\right) w$ which are easier to compute

$$\rho\left(w\begin{pmatrix} 1 & 0 \\ pt_0 & 1 \end{pmatrix}\right).\mathbb{1}_{\mathcal{R}}(x) = p^{-n}\psi_v(-pt_0\,\mathrm{Nr}\, x)\mathbb{1}_{\widehat{\mathcal{R}}}(x). \qquad \square$$

3.6 Local uniformity

As a preparation for the global theory, we will need the following proposition that uniformly controls the decay of test functions for the Weil representation.

PROPOSITION 3.23. *Let $M \in \Omega_v$, $s \in \mathbf{SL}_2(F)$ and $L \in \mathrm{O}(B_v, \mathrm{Nr})$. If $v = \infty$, then there is $\delta > 0$ such that the inequality*

$$|(\rho(s)M)(Lx)| \ll (1 + \|x\|)^{-4-\delta}$$

holds uniformly on compact sets in $\mathbf{SL}_2(F) \times \mathrm{O}(B_v, \mathrm{Nr})$. If $v < \infty$, then for every compact subset of $\mathscr{K} \subset \mathbf{SL}_2(F) \times \mathrm{O}(B_v, \mathrm{Nr})$ there is a compact subset $\mathscr{C} \subset B_v$ such that

$$\forall (s, L) \in \mathscr{K} \colon |(\rho(s)M)(Lx)| \ll_{\mathscr{K}} \mathbb{1}_{\mathscr{C}}.$$

Proof. The claim for $v = \infty$ follows immediately from Corollary 3.6. Fix now $v < \infty$. We may cover the set $\mathscr{K}$ by the product of its projections onto $\mathbf{SL}_2(F)$ and $\mathrm{O}(B_v, \mathrm{Nr})$, which is compact. Hence, we may assume without loss of generality that $\mathscr{K} = \mathscr{K}_0 \times \mathscr{K}_1$ is a compact product set.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Note that it is enough to show that there is some $\mathscr{C}_0 \subset B_v$ such that $|\rho(s).M| \ll \mathbb{1}_{\mathscr{C}_0}$ for $s \in \mathscr{K}_0$. In particular, the compact set $\mathscr{C} = \bigcup_{L \in \mathscr{K}_1} L^{-1}\mathscr{C}_0$ will satisfy the claimed properties. Using the Iwasawa decomposition, we can find a compact subset C_P of the group of upper triangular matrices such that $\mathscr{K}_0 \subset C_P \mathbf{SL}_2(\mathbb{Z}_p)$. Recall from Lemma 3.15 that the $\rho\big(\mathbf{SL}_2(\mathbb{Z}_p)\big)$-orbit of M is finite and a uniform bound on $\rho(\mathscr{K}_0).M$ will follow from a uniform bound on $\rho(C_P).M'$ for every M' in $\rho\big(\mathbf{SL}_2(\mathbb{Z}_p)\big).M$. A uniform bound on $|\rho(C_P).M'|$ follows immediately from the formulae for the Weil action of diagonal and upper unipotent matrices. □

3.7 Global Weil representation and theta series

Fix an additive character $\psi\colon \mathbb{A}/\mathbb{Q} \to \mathbb{C}$ such that $\psi = \prod_v \psi_v$ and ψ_v is unramified for all finite v. This is possible for the adèle ring of $\mathbb{Q}$ and the standard character with $\psi_\infty(a) = \exp(-2\pi i a)$ will do. We consider henceforth always the local Weil representations on Ω_v to be with respect to ψ_v.

Set

$$\Omega := \bigotimes{}'\Omega_v = \mathrm{Span}_{\mathbb{C}}\left\{\prod_v M_v \mid M_v \in \Omega_v, \forall \text{ a.e. } v\colon M_v = \mathbb{1}_{R_v}\right\}.$$

A pure tensor $M = \prod_v M_v \in \Omega$ is called a *standard test function*. The Weil representations for each Ω_v define in the standard way a representation of $\mathbf{SL}_2(\mathbb{A})$ on Ω. To see that this action is well-defined we need to check that for almost every v we have $\rho\big(\mathbf{SL}_2(\mathbb{Z}_v)\big).\mathbb{1}_{R_v} = \mathbb{1}_{R_v}$, and this follows from Lemma 3.11. The complex vector space Ω also carries actions of $\mathbf{G}(\mathbb{A})$ by left and right multiplication because for every $v < \infty$ the function $\mathbb{1}_{R_v}$ is invariant under left and right multiplication by elements of K_v.

Definition 3.24. For $M \in \Omega$ define the theta kernel $\Theta_M\colon \mathbf{G}(\mathbb{A}) \times \mathbf{G}(\mathbb{A}) \times \mathbf{SL}_2(\mathbb{A}) \to \mathbb{C}$ by

$$\Theta_M(l, r; s) := \sum_{\xi \in B} \big(\rho(s)M\big)(l^{-1}\xi r).$$

The uniform decay from Proposition 3.23 is sufficient for the theta series to converge absolutely and uniformly on compact sets in $\mathbf{G}(\mathbb{A}) \times \mathbf{G}(\mathbb{A}) \times \mathbf{SL}_2(\mathbb{A})$. In particular, the theta series is a well-defined continuous function on its domain.

The theta kernel is obviously $\mathbf{G}(\mathbb{Q})$-invariant on the left in the first two coordinates by virtue of its definition. Less obvious, yet well-known, is that it is also $\mathbf{SL}_2(\mathbb{Q})$ left-invariant in the third coordinate. A simple way to verify this is by first showing invariance under upper-triangular matrices by an elementary calculation and then establishing invariance under the involution $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ using the Poisson summation formula. The decay conditions we have imposed on functions in Ω_∞ are sufficient for the Poisson summation formula to hold [SW71, p. 252, Corollary 2.6].

We now recall the Fourier–Whittaker decomposition of a continuous function $\varphi\colon [\mathbf{SL}_2(\mathbb{A})] \to \mathbb{C}$. For any $\alpha \in \mathbb{Q}$, define the Whittaker function

$$W_\varphi(g, \alpha) = \int_{[\mathbf{N}(\mathbb{A})]} \varphi\left(\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} g\right)\psi(-\alpha n)\, dn.$$

We have the following standard properties of the Whittaker function

$$\forall n \in \mathbb{A}\colon W_\varphi\left(\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} g, \alpha\right) = \psi(\alpha n) W_\varphi(g, \alpha),$$

$$\forall \lambda \in \mathbb{Q}^\times\colon W_\varphi\left(\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} g, \alpha\right) = W_\varphi(g, \lambda^2 \alpha).$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Because our function φ is defined on $[\mathbf{SL}_2(\mathbb{A})]$ and not $[\mathbf{PGL}_2(\mathbb{A})]$, we cannot reduce the dependence on α to the two cases 0 and 1. Pontryagin duality for the compact abelian group $[\mathbf{N}(\mathbb{A})]$ implies that the following equality

$$\varphi(g) = \sum_{\alpha \in \mathbb{Q}} W_\varphi(g, \alpha) \tag{23}$$

holds pointwise as long as the right-hand side is absolutely convergent.[9]

PROPOSITION 3.25. *Fix $M \in \Omega$. Then, the Fourier–Whittaker coefficients of $\Theta_M(l, r; s)$ in the s-variable are*

$$W_{\Theta_M(l,r;\bullet)}(s, \alpha) = \sum_{\substack{\xi \in B \\ \operatorname{Nr} \xi = \alpha}} \big(\rho(s)M\big)(l^{-1}\xi r).$$

Because the theta series in Definition 3.24 converges absolutely, an immediate corollary is that the Fourier–Whittaker expansion (23) holds pointwise for $\Theta_M(l, r; \bullet)$.

Proof. Because the theta series converges absolutely, we may exchange summation and integration in the definition of $W_{\Theta_M(l,r;\bullet)}$ and write

$$\begin{aligned} W_{\Theta_M(l,r;\bullet)}(s, \alpha) &= \sum_{\xi \in B} \int_{[\mathbf{N}(\mathbb{A})]} \left(\rho\left(\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} s\right) M\right)(l^{-1}\xi r)\psi(-\alpha n)\, dn \\ &= \sum_{\xi \in B} \int_{[\mathbf{N}(\mathbb{A})]} \big(\rho(s)M\big)(l^{-1}\xi r)\psi(\operatorname{Nr}\xi \cdot n - \alpha n)\, dn \\ &= \sum_{\substack{\xi \in B \\ \operatorname{Nr} \xi = \alpha}} \big(\rho(s)M\big)(l^{-1}\xi r). \end{aligned}$$

□

4. Theta kernels for Eichler orders

4.1 Weil action on Eichler orders

We first introduce the notation D_B for the (reduced) discriminant of B, explicitly

$$D_B = \prod_{p : B_p \text{ is ramified}} p.$$

DEFINITION 4.1. An Eichler order $\mathcal{R} \subset B$ is an intersection of two maximal orders $\mathcal{R}_1$ and $\mathcal{R}_2$. The completion $\mathcal{R}_v$ of $\mathcal{R}$ at any finite place v is a local Eichler order in B_v. We say that $\mathcal{R}$ is ramified at v if $\mathcal{R}_v$ is non-maximal. If B is ramified at v then $\mathcal{R}_v$ is unramified at v because B_v has a unique maximal order.

For almost all places, the local orders $\mathcal{R}_{1,v}$ and $\mathcal{R}_{2,v}$ coincide[10] and $\mathcal{R}_v$ is a maximal order, i.e. the level of $\mathcal{R}_v$ at these places is 1. We define the level of $\mathcal{R}$ as

$$q = \prod_{v < \infty} \text{level of } \mathcal{R}_v.$$

The reader may recall Definition 3.16, where we defined the level of a local Eichler order.

[9] For a fixed $g \in \mathbf{PGL}_2(\mathbb{A})$ this is the Fourier expansion of the function $n \mapsto \varphi(ng)$ evaluated at $n = e$. The function $n \mapsto \varphi(ng)$ is a continuous function on the compact Abelian group $[\mathbf{N}(\mathbb{A})]$. If the Fourier transform of a continuous functions converges absolutely, then it coincides with the function everywhere.

[10] This happens for any two lattices in a rational vector space.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

The dual lattice to $\mathscr{R}$ is

$$\widehat{\mathscr{R}} \coloneqq \left\{x \in B \mid \forall y \in \mathscr{R}\colon \operatorname{Tr}(x^\iota y) \in \mathbb{Z}\right\}.$$

Dualization commutes with localization, i.e. $(\widehat{\mathscr{R}})_v = \widehat{\mathscr{R}_v}$. Using the decomposition

$$\widehat{\mathscr{R}}/\mathscr{R} = \prod_{v<\infty} \widehat{\mathscr{R}}_v/\mathscr{R}_v$$

and Lemmata 3.18 and 3.12, we see the existence of an isomorphism of abelian groups

$$\jmath\colon \widehat{\mathscr{R}}/\mathscr{R} \to \left(\mathbb{Z}/q\mathbb{Z}\right)^2 \times \prod_{p\mid D_B} \mathbb{F}_{p^2}.$$

The map $\jmath$ fibers through the local maps $\jmath_v$ and satisfies the properties inherited from Lemma 3.18:

$$\jmath(\mathscr{R}_1) = \mathbb{Z}/q\mathbb{Z} \times \{0\} \times \prod_{p\mid D_B} \mathbb{F}_{p^2},$$

$$\jmath(\mathscr{R}_2) = \{0\} \times \mathbb{Z}/q\mathbb{Z} \times \prod_{p\mid D_B} \mathbb{F}_{p^2},$$

$$\forall x \in \widehat{\mathscr{R}}\colon\ -q \operatorname{Nr} x \equiv \jmath(x)_1 \cdot \jmath(x)_2 \mod q,$$

$$\forall x \in \widehat{\mathscr{R}}, p \mid D_B\colon\ -p \operatorname{Nr} x \equiv \operatorname{Nr} \jmath(x)_3 \mod p.$$

For $m \in \mathbb{N}$ and $x \in \mathbb{F}_{p^2}$ define

$$x \bmod m = \begin{cases} x, & p \mid m, \\ 0, & p \nmid m. \end{cases}$$

Similarly, set $x \bmod m = x \bmod \gcd(m, q)$ for all $x \in \mathbb{Z}/q\mathbb{Z}$. We extend this definition elementwise to a map

$$\left(\mathbb{Z}/q\mathbb{Z}\right)^2 \times \prod_{p\mid D_B} \mathbb{F}_{p^2} \xrightarrow{\bmod m} \left(\mathbb{Z}/\gcd(m,q)\mathbb{Z}\right)^2 \times \prod_{p\mid D_B} \mathbb{F}_{p^2}.$$

Definition 4.2. Let $\mathscr{R} = \mathscr{R}_1 \cap \mathscr{R}_2 \subset B$ be an Eichler order of level q. For every $m \mid qD_B$ define

$$\widehat{\mathscr{R}}^{(m)} = \left\{x \in \widehat{\mathscr{R}} \mid \jmath(x) \equiv 0 \mod m\right\}.$$

Note that $\widehat{\mathscr{R}}^{(1)} = \widehat{\mathscr{R}}$ and $\widehat{\mathscr{R}}^{(qD_B)} = \mathscr{R}$ and

$$\left[\widehat{\mathscr{R}} : \widehat{\mathscr{R}}^{(m)}\right] = m^2.$$

Moreover, the definition of $\widehat{\mathscr{R}}^{(m)}$ does not depend on the choices involved in the definition of $\jmath$.

We also define for $x \in \widehat{\mathscr{R}}^{(m)}$

$$\nu_m(x) \coloneqq \prod_{p\mid qD_B} \begin{cases} p^{1-\operatorname{ord}_p m} \operatorname{Nr}(x) \mod p^{\operatorname{ord}_p D_B - \operatorname{ord}_p m}, & p \mid D_B, \\ \nu_{p^{\operatorname{ord}_p m}}(x), & p \mid q, \end{cases} \in \left(\mathbb{Z}/(qD_B/m)\mathbb{Z}\right).$$

The lattices $\widehat{\mathscr{R}}^{(m)}$ will feature prominently in the description of the action of the Weil representation. In classical terms, they will appear in the Fourier expansion of a theta series at different cusps. We will use the following notation for the completion of an integral lattice at all finite places.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

DEFINITION 4.3. If $L \subset B$ is a lattice, then define $\mathbb{1}_{L_f} \colon B_f \to \mathbb{C}$ to be $\mathbb{1}_{L_f} \coloneqq \prod_p \mathbb{1}_{L_p}$, where $L_p \subset B_p$ is the p-adic closure of L.

Our goal now is to describe the $\rho\big(\mathbf{SL}_2(\widehat{\mathbb{Z}})\big)$-action on $\mathbb{1}_{\mathcal{R}_f}$. The first step is to identify the stabilizer of $\mathbb{1}_{\mathcal{R}_f}$.

DEFINITION 4.4. Define the compact-open subgroup $U_{\mathcal{R}} = \prod_{p<\infty} U_p < \mathbf{SL}_2(\mathbb{A}_f)$ by

$$U_p = \begin{cases} U_0(p), & B \text{ is ramified at } p, \\ U_0(p^n), & \mathcal{R}_p \text{ has level } p^n, \\ \mathbf{SL}_2(\mathbb{Z}_p), & \text{otherwise.} \end{cases}$$

From Lemmata 3.11 and 3.17, we know that $\rho(U_{\mathcal{R}}).\mathbb{1}_{\mathcal{R}_f} = \mathbb{1}_{\mathcal{R}_f}$.

We define the arithmetic function $\rho(a \mid qD_B)$ as

$$\rho(a \mid qD_B) \coloneqq \prod_{p \mid \gcd(qD_B/a, a)} (1 - p^{-1}).$$

Note that $(\log\log(10qD_B))^{-1} \ll \rho(a \mid qD_B) \leq 1$.

PROPOSITION 4.5. *Let $\mathcal{R} \subset B$ be an Eichler order. Then,*

$$\rho\big(\mathbf{SL}_2(\widehat{\mathbb{Z}})\big).\mathbb{1}_{\mathcal{R}_f} = \left\{ \frac{a(-1)^{\omega(D_B/\gcd(a,D_B))}}{qD_B} \psi\left(\frac{t \cdot \nu_a(x)}{qD_B/a}\right) \cdot \mathbb{1}_{\widehat{\mathcal{R}}_f^{(a)}} \,\middle|\, a \mid qD_B,\ t \in \mathbb{Z}/\tfrac{qD_B}{a}\mathbb{Z},\ \gcd(t,a) = 1 \right\}.$$

Moreover, each function on the right-hand side corresponds to a single coset of $\mathbf{SL}_2(\widehat{\mathbb{Z}})/U_{\mathcal{R}}$.

Remark 4.6. For every $a \mid qD_B$, the characteristic function $\mathbb{1}_{\widehat{\mathcal{R}}^{(a)}}$ appears above exactly $(qD_B/a)\rho(a \mid qD_B)$ times with different phase functions.

Proof. This follows from combining the local contributions as calculated in Propositions 3.13 and 3.21 and the remarks following these propositions. □

4.2 Theta series for Eichler orders

We fix once and for all an Eichler order $\mathcal{R} = \mathcal{R}_1 \cap \mathcal{R}_2 \subset B$ of level q. In this section, we unwind the adelic definition of a theta series for the case of Eichler orders into classical terms.

Denote $K_{\mathcal{R}} = \prod_{v<\infty} \big(\mathbf{G}(\mathbb{Q}_v) \cap \mathcal{R}_v\big)$. Strong approximation for the simply connected group $\mathbf{G}$ implies that the double quotient

$$\mathbf{G}(\mathbb{Q}) \backslash \mathbf{G}(\mathbb{A}) / K_{\mathcal{R}}$$

is a single orbit of $\mathbf{G}(\mathbb{R})$. The stabilizer of the identity double coset in $\mathbf{G}(\mathbb{R})$ is

$$\Gamma \coloneqq K_{\mathcal{R}} \cap \mathbf{G}(\mathbb{Q}) = \big\{ x \in \mathcal{R} \mid \operatorname{Nr} x = 1 \big\}.$$

Hence, there is a canonical quotient map

$$\pi_\Gamma \colon \mathbf{G}(\mathbb{Q}) \backslash \mathbf{G}(\mathbb{A}) \to \Gamma \backslash \mathbf{G}(\mathbb{R}).$$

Each fiber of this map is a torsor for $K_{\mathcal{R}}$. The quotient map π_Γ induces a natural isomorphism

$$\pi_\Gamma^* \colon \operatorname{Map}\left(\Gamma \backslash \mathbf{G}(\mathbb{R}), \mathbb{C}\right) \to \operatorname{Map}\left([\mathbf{G}(\mathbb{A})], \mathbb{C}\right)^{K_{\mathcal{R}}},$$

where the right-hand side is the set of all $K_{\mathcal{R}}$-invariant maps.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Set $\Lambda = U_{\mathscr{R}} \cap \mathbf{SL}_2(\mathbb{Q}) < \mathbf{SL}_2(\mathbb{R})$. Explicitly, $\Lambda = \Gamma_0(qD_B)$ where D_B is the product of the primes ramified in B and q is the level of $\mathscr{R}$. Again, the double quotient

$$\mathbf{SL}_2(\mathbb{Q})\backslash\mathbf{SL}_2(\mathbb{A})/U_{\mathscr{R}}$$

is a single orbit of $\mathbf{SL}_2(\mathbb{R})$ and the stabilizer of the identity is Λ. This induces a map

$$\pi_\Lambda\colon \mathbf{SL}_2(\mathbb{Q})\backslash\mathbf{SL}_2(\mathbb{A}) \to \Lambda\backslash\mathbf{SL}_2(\mathbb{R}) \tag{24}$$

and, furthermore, a natural isomorphism

$$\pi_\Lambda^*\colon \operatorname{Map}\left(\Lambda\backslash\mathbf{SL}_2(\mathbb{R}), \mathbb{C}\right) \to \operatorname{Map}\left([\mathbf{SL}_2(\mathbb{A})], \mathbb{C}\right)^{U_{\mathscr{R}}}.$$

We can now write the adelic theta kernel in classical terms.

Proposition 4.7. *Fix $M = M_\infty \cdot \prod_{v<\infty} M_v \in \Omega$ such that $M_v = \mathbb{1}_{\mathscr{R}_v}$ for all finite v. Let $l_\infty, r_\infty \in \mathbf{G}(\mathbb{R})$ and $s_\infty \in \mathbf{SL}_2(\mathbb{R})$. Then,*

$$\Theta_M\left(l_\infty K_{\mathscr{R}}, r_\infty K_{\mathscr{R}}; s_\infty U_{\mathscr{R}}\right) = \sum_{\xi\in\mathscr{R}} (\rho(s_\infty)M_\infty)(l_\infty^{-1}\xi r_\infty).$$

Hence, Θ_M defines a classical theta kernel on $\Gamma\backslash\mathbf{G}(\mathbb{R}) \times \Gamma\backslash\mathbf{G}(\mathbb{R}) \times \Lambda\backslash\mathbf{SL}_2(\mathbb{R})$.

Proof. This follows from the discussion above, Lemma 3.17, and the local-to-global principle for lattices that implies

$$\mathscr{R} = \bigcap_{v<\infty} \mathscr{R}_v,$$

where the intersection is taken in B. □

Definition 4.8. Fix $M_\infty \in \Omega_\infty$ and define $\vartheta_{M_\infty}\colon \Gamma\backslash\mathbf{G}(\mathbb{R}) \times \Gamma\backslash\mathbf{G}(\mathbb{R}) \times \Lambda\backslash\mathbf{SL}_2(\mathbb{R})$ by

$$\vartheta_{M_\infty}(l_\infty, r_\infty; s_\infty) = \Theta_M(l_\infty K_{\mathscr{R}}, r_\infty K_{\mathscr{R}}; s_\infty U_{\mathscr{R}}).$$

4.3 The Weil L^2-norm of the theta kernel

Our method relies heavily on bounding the L^2-norm of $\Theta_M(l, r; s)$ in the s-variable. We use the Fourier–Whittaker decomposition from Proposition 3.25 to bound the L^2-norm by a second-moment count of rational matrices. Unfortunately, the classical representation above is not well adapted to this calculation because of the many cusps of $\Lambda\backslash\mathbf{SL}_2(\mathbb{R})$. Instead, we compute adelically the L^2-norm. This is easier to execute because the adelic quotient $[\mathbf{SL}_2(\mathbb{A})] = \mathbf{SL}_2(\mathbb{Q})\backslash\mathbf{SL}_2(\mathbb{A})$ has a single cusp.

Proposition 4.9. *Fix $M_\infty \in \Omega_\infty$. Then,*

$$\begin{aligned}&\frac{1}{\operatorname{covol}(\Lambda)} \int_{\Lambda\backslash\mathbf{SL}_2(\mathbb{R})} |\vartheta_{M_\infty}(l_\infty, r_\infty; s_\infty)|^2 \, ds_\infty \\ &\quad\le (qD_B)^{-1} \sum_{a|qD_B} \sum_{\alpha\in\mathbb{Q}} \int_{\sqrt{3}/2}^\infty \int_{\mathbf{SO}_2(\mathbb{R})} \sum_{\substack{\xi\in\widehat{\mathscr{R}}^{(a)} \\ \operatorname{Nr}\xi=\alpha}} \left|\left(\rho\left(\operatorname{diag}(y^{1/2}, y^{-1/2})k\right).M_\infty\right)(l_\infty^{-1}\xi r_\infty)\right|^2 dk \frac{dy}{y^2},\end{aligned}$$

where the measure on $\mathbf{SO}_2(\mathbb{R})$ is normalized to be a probability measure, and the left-hand side is independent of the measure normalization on $\mathbf{SL}_2(\mathbb{R})$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Proof. Fix $M = M_\infty \cdot \prod_{v<\infty} M_v \in \Omega$ such that $M_v = \mathbb{1}_{\mathcal{R}_v}$ for all finite v. Then, Proposition 4.7 and the isomorphism (24) imply

$$\frac{1}{\operatorname{covol}(\Lambda)} \int_{\Lambda\backslash \mathbf{SL}_2(\mathbb{R})} |\vartheta_{M_\infty}(l_\infty, r_\infty; s_\infty)|^2 \, ds_\infty = \int_{[\mathbf{SL}_2(\mathbb{A})]} |\Theta_M(l_\infty K_{\mathcal{R}}, r_\infty K_{\mathcal{R}}; s)|^2 \, ds.$$

We proceed to bound the adelic integral by expanding the domain of integration to a Siegel set.

Denote by $\mathbf{N} < \mathbf{SL}_2$ the algebraic subgroup of upper triangular matrices. We have $\mathbf{N} \simeq \mathbb{G}_a$ and a fundamental domain for the action of $\mathbf{N}(\mathbb{Q})$ on $\mathbf{N}(\mathbb{A})$ is

$$\mathcal{N} = \begin{pmatrix} 1 & [0,1) \\ 0 & 1 \end{pmatrix} \cdot \prod_{v<\infty} \begin{pmatrix} 1 & \mathbb{Z}_v \\ 0 & 1 \end{pmatrix}.$$

Set $A^> \coloneqq \{\operatorname{diag}(y^{1/2}, y^{-1/2}) \colon y > \sqrt{3}/2\} \subset \mathbf{SL}_2(\mathbb{R})$. A Siegel set for the action of $\mathbf{SL}_2(\mathbb{Q})$ on $\mathbf{SL}_2(\mathbb{A})$ is given by

$$\mathcal{S} = \mathcal{N} \cdot A^> \cdot \mathbf{SO}_2(\mathbb{R})\mathbf{SL}_2(\hat{\mathbb{Z}}).$$

Denote $l = (l_\infty, e, e, \ldots)$ and similarly $r = (r_\infty, e, e, \ldots)$. Because the Siegel set contains a fundamental domain for the action of the lattice, we can write

$$\begin{aligned}
&\int_{[\mathbf{SL}_2(\mathbb{A})]} |\Theta_M(l_\infty K_{\mathcal{R}}, r_\infty K_{\mathcal{R}}; s)|^2 \, ds \\
&\quad \le \int_{\mathcal{S}} |\Theta_M(l_\infty K_{\mathcal{R}}, r_\infty K_{\mathcal{R}}; s)|^2 \, ds \\
&\quad = \varrho_{\mathbf{SL}_2}^{-1} \int_{[\mathbf{N}(\mathbb{A})]} \int_{\sqrt{3}/2}^\infty \int_{\mathbf{SL}_2(\hat{\mathbb{Z}})} \int_{\mathbf{SO}_2(\mathbb{R})} \left| \sum_{\xi \in B} (\rho(n)\rho(\operatorname{diag}(y^{1/2}, y^{-1/2})kk_f).M)(l^{-1}\xi r) \right|^2 dn \frac{dy}{y^2} \, dk_f \, dk \\
&\quad = \varrho_{\mathbf{SL}_2}^{-1} \sum_{\alpha \in \mathbb{Q}} \int_{\sqrt{3}/2}^\infty \int_{\mathbf{SL}_2(\hat{\mathbb{Z}})} \int_{\mathbf{SO}_2(\mathbb{R})} \left| \sum_{\substack{\xi \in B \\ \operatorname{Nr} \xi = \alpha}} (\rho(\operatorname{diag}(y^{1/2}, y^{-1/2})k)\rho(k_f).M)(l^{-1}\xi r) \right|^2 dk_f \frac{dy}{y^2} \, dk \\
&\quad \le \varrho_{\mathbf{SL}_2}^{-1} \sum_{\alpha \in \mathbb{Q}} \int_{\sqrt{3}/2}^\infty \int_{\mathbf{SL}_2(\hat{\mathbb{Z}})} \int_{\mathbf{SO}_2(\mathbb{R})} \left| \sum_{\substack{\xi \in B \\ \operatorname{Nr} \xi = \alpha}} \left| (\rho(\operatorname{diag}(y^{1/2}, y^{-1/2})k)\rho(k_f).M)(l^{-1}\xi r) \right| \right|^2 dk_f \frac{dy}{y^2} \, dk.
\end{aligned}$$

The last equality follows from the computation of the Fourier coefficients in the unipotent variable in Proposition 3.25 and the orthogonality of characters. We normalize the measure on $\mathbf{SL}_2(\mathbb{A}_f)$ so that $\mathbf{SL}_2(\widehat{\mathbb{Z}})$ has volume 1. Then, the global measure normalization constant $\varrho_{\mathbf{SL}_2}$ is equal to the volume of $\mathbf{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ with respect to the standard hyperbolic measure $dx\,dy/y^2$, i.e. $\varrho_{\mathbf{SL}_2} = \operatorname{covol}(\mathbf{SL}_2(\mathbb{Z})) = \pi/3$.

In the last line we have inserted the absolute value into the sum using the triangle inequality, and we continue to evaluate the integral along $\mathbf{SL}_2(\hat{\mathbb{Z}})$. From Lemmata 3.11 and 3.17, we know that the integrand is invariant under the finite index subgroup $U_{\mathcal{R}} < \mathbf{SL}_2(\hat{\mathbb{Z}})$. We decompose the integral into $\big[\mathbf{SL}_2(\hat{\mathbb{Z}}) \colon U_{\mathcal{R}}\big]$-integrals along the different cosets of $U_{\mathcal{R}}$ in $\mathbf{SL}_2(\hat{\mathbb{Z}})$. We have an exact expression for the integrand on each coset due to Proposition 4.5. The phases of the form $\psi\big(t \cdot \nu_a(x)/qD_B/a\big)$ that appear in each element in the $\rho\big(\mathbf{SL}_2(\widehat{\mathbb{Z}})\big)$-orbit are discarded because of the innermost absolute value. Hence, each $U_{\mathcal{R}}$-coset reduces to a sum over elements in a lattice $\widehat{\mathcal{R}}^{(a)}$ for $a \mid qD_B$. From Remark 4.6, we deduce for any $a \mid qD_B$ that the weight of the sum over $\mathbb{1}_{\widehat{R}^{(a)}}$ is $\rho(a \mid qD_B)\varrho_{\mathbf{SL}_2}^{-1}\big[\mathbf{SL}_2(\widehat{\mathbb{Z}}) \colon U_{\mathcal{R}}\big]^{-1}$. Because $\Lambda = U_{\mathcal{R}} \cap \mathbf{SL}_2(\mathbb{A}_f)$ and $\mathbf{SL}_2(\mathbb{Z}) = \mathbf{SL}_2(\widehat{\mathbb{Z}}) \cap \mathbf{SL}_2(\mathbb{A}_f)$, the index satisfies $\varrho_{\mathbf{SL}_2}\big[\mathbf{SL}_2(\widehat{\mathbb{Z}}) \colon U_{\mathcal{R}}\big] = \operatorname{covol}(\Lambda)$. Because $\Lambda = \Gamma_0(qD_B)$, we

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

see that the weight factor is equal to

$$\operatorname{covol}(\Lambda)^{-1}\rho(a \mid qD_B) = \frac{\prod_{p|\gcd(qD_B/a,a)}(1-p^{-1})}{(\pi/3)qD_B\prod_{p|qD_B}(1+p^{-1})} \leq (qD_B)^{-1}. \qquad \square$$

5. The theta lift

In this section we discuss the pairing between a theta kernel and an automorphic form. This subject is well studied in the literature, we need to review and revisit several results because of the explicit form that we require.

5.1 Cuspidal theta series

Definition 5.1. We say that a test function $M\colon B_{\mathbb{A}} \to \mathbb{C}$ is cuspidal if $(\rho(s)M)(l^{-1}\xi r) = 0$ for all $l, r \in \mathbf{G}(\mathbb{A})$, $s \in \mathbf{SL}_2(\mathbb{A})$, and $\xi \in B$ with $\operatorname{Nr}\xi = 0$.

For example, M is cuspidal if $M = \prod_v M_v$ and there is a place v such that $\rho(s_v).M_v(x_v) = 0$ for every $s_v \in \mathbf{SL}_2(\mathbb{Q}_v)$, and $x_v \in B_v$ with $\operatorname{Nr} x_v = 0$. The importance of cuspidal test functions is that their theta series, when well-defined, is a cuspidal function of s on $[\mathbf{SL}_2(\mathbb{A})]$. This follows from Proposition 3.25. Note that the cuspidality condition for M is very restrictive if $\mathbf{G}$ is split. For example, if $\mathbf{G} = \mathbf{SL}_2$, then the test function $\exp(-D\operatorname{Tr}(x_\infty {}^t x_\infty))P(x_\infty)\prod_{v<\infty}\mathbb{1}_{\mathcal{R}_v}$, for a polynomial $P\colon \operatorname{Mat}_{2\times 2}(\mathbb{R}) \to \mathbb{C}$ and $D > 0$, is used in [Shi72]. This test function is generally *not* cuspidal.

5.2 Unfolding

Lemma 5.2. *If $\mathcal{R}_p$ is an Eichler order, then $\operatorname{Nr}\mathcal{R}_p^\times = \mathbb{Z}_p^\times$.*

Proof. This is simple to verify if B is split at p by conjugating $\mathcal{R}_p$ to $\left(\begin{smallmatrix}\mathbb{Z}_p & \mathbb{Z}_p\\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{smallmatrix}\right)$. If B is ramified at p, then $\mathcal{R}_p^\times = \mathcal{O}_{B_p}^\times$ is the unit group of the ring of algebraic integers in B_p. The algebra B_p contains an unramified quadratic extension $E/\mathbb{Q}_p$, hence $\operatorname{Nr}\mathcal{R}_p^\times$ contains $\operatorname{Nr}\mathcal{O}_E^\times = \mathbb{Z}_p^\times$. $\square$

Lemma 5.3. *Let $M_f = \prod_p \mathbb{1}_{p^{-k_p}\mathcal{R}_p}\colon B\otimes\mathbb{A}_f \to \mathbb{C}$, where $k_p \in \mathbb{Z}_{\geq 0}$ for all p and $k_p = 0$ for almost every prime p. Denote $N = \prod_p p^{2k_p} \in \mathbb{N}$ and fix $\xi \in B^\times$. Then,*

$$\int_{\mathbf{G}(\mathbb{A}_f)} M_f(l^{-1}\xi)\,dl \ll_\varepsilon \varrho_{\mathbf{G}}^{-1}N|\operatorname{Nr}\xi|q^\varepsilon,$$

where we recall that q denotes the level of $\mathcal{R}$, and this integral vanishes unless $\operatorname{Nr}\xi \in N^{-1}\mathbb{Z}$.

Proof. The integral decomposes into a product of local integrals $\varrho_{\mathbf{G}}^{-1}\prod_p\int_{\mathbf{G}(\mathbb{Q}_p)}\mathbb{1}_{p^{-k_p}\mathcal{R}_p}(l_p^{-1}\xi)\,dl_p$. All elements of $p^{-k_p}\mathcal{R}_p$ have norms in $p^{-2k_p}\mathbb{Z}_p$. Hence, the local integral vanishes if $\operatorname{Nr}(l_p^{-1}\xi) = \operatorname{Nr}(\xi) \notin p^{-2k_p}\mathbb{Z}_p$. Because $\operatorname{Nr}\xi \in \mathbb{Q}$, the non-vanishing conditions at all primes p imply that $\int_{\mathbf{G}(\mathbb{A}_f)}\cdots$ vanishes if $\operatorname{Nr}\xi \notin N^{-1}\mathbb{Z}$.

Fix now p and assume $\operatorname{Nr}\xi \in p^{-2k_p}\mathbb{Z}_p$. Then, the local integral is equal to $\int_{\mathbf{G}(\mathbb{Q}_p)}\mathbb{1}_{\mathcal{R}_p}(l_p^{-1}p^{k_p}\xi)\,dl_p$. The integrand is right-invariant under K_p. Denote by $\mathcal{R}_p(\alpha)$ the set of elements in $\mathcal{R}_p$ of norm $\alpha \in \mathbb{Q}_p^\times$. Of course, $\mathcal{R}_p(\alpha) = \emptyset$ if $\alpha \notin \mathbb{Z}_p$. The set $\mathcal{R}_p(\alpha)$ is left-invariant under multiplication by $K_p := \mathcal{R}_p(1)$, and

$$\int_{\mathbf{G}(\mathbb{Q}_p)}\mathbb{1}_{\mathcal{R}_p}(l_p^{-1}p^{k_p}\xi)\,dl_p = m_{\mathbf{G}(\mathbb{Q}_p)}(K_p)\cdot\#\Big(K_p\backslash\mathcal{R}_p\big(p^{2k_p}\operatorname{Nr}\xi\big)\Big).$$

We have $m_{\mathbf{G}(\mathbb{Q}_p)}(K_p) = 1$ if $p \not| q$ and $m_{\mathbf{G}(\mathbb{Q}_p)}(K_p) = (p+1)^{-1}p^{-n+1}$ if $p^n \parallel q$ with $n > 0$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

We now estimate $\#\big(K_p\backslash\mathcal{R}_p(\alpha)\big)$. Define

$$\mathcal{R}_p(\alpha)^\dagger = \big\{x \in \mathcal{R}_p \colon |\operatorname{Nr} x|_p = |\alpha|_p\big\},$$

evidently $\mathcal{R}_p(\alpha) \subset \mathcal{R}_p(\alpha)^\dagger$. The set $\mathcal{R}_p(\alpha)^\dagger$ is invariant under left multiplication by $\mathcal{R}_p^\times$. Because $\operatorname{Nr}\mathcal{R}_p^\times = \mathbb{Z}_p^\times$, each coset of $\mathcal{R}_p^\times\backslash\mathcal{R}_p(\alpha)^\dagger$ contains exactly one coset of $K_p\backslash\mathcal{R}_p(\alpha)$. Thus, $\#\big(K_p\backslash\mathcal{R}_p(\alpha)\big) = \#\big(\mathcal{R}_p^\times\backslash\mathcal{R}_p(\alpha)^\dagger\big)$.

If B_p ramifies, then the fact that $\operatorname{ord}_p(\operatorname{Nr}\bullet)$ is a valuation on B_p implies that $\mathcal{R}(\alpha)^\dagger$ is a single coset of $\mathcal{R}_p^\times$ if $\alpha \in \mathbb{Z}_p$. In the split case, we can assume $B_p = \operatorname{Mat}_{2\times 2}(\mathbb{Q}_p)$ and $\mathcal{R}_p = \left(\begin{smallmatrix}\mathbb{Z}_p & \mathbb{Z}_p\\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{smallmatrix}\right)$, where $p^n \parallel q$. Let $\widetilde{\mathcal{R}}_p^\times$ be the image of $\mathcal{R}_p^\times$ in $\mathbf{PGL}_2(\mathbb{Q}_p)$. Then, the map $\mathcal{R}_p^\times\backslash\mathcal{R}_p(\alpha)^\dagger \to \widetilde{\mathcal{R}}_p^\times\backslash\mathbf{PGL}_2(\mathbb{Q}_p)$ is injective because $Z_{\mathbf{GL}_2}(\mathbb{Q}_p)\cap\{g\in\mathbf{GL}_2(\mathbb{Q}_p)\colon |\det g|_p = 1\} \subset \mathcal{R}_p^\times$. Hence, it is enough to find an upper bound for the number of $\widetilde{\mathcal{R}}_p^\times$ cosets in the image of $\mathcal{R}_p(\alpha)^\dagger$ in $\mathbf{PGL}_2(\mathbb{Q}_p)$.

The group $\mathbf{PGL}_2(\mathbb{Z}_p)$ is the stabilizer of a vertex v_0 in the Bruhat–Tits tree of $\mathbf{PGL}_2(\mathbb{Q}_p)$ and $\operatorname{diag}(p^{-n},1)\mathbf{PGL}_2(\mathbb{Z}_p)\operatorname{diag}(p^n,1)$ is a stabilizer of a vertex v_n, with $\operatorname{dist}(v_0,v_n) = n$. Hence, $\widetilde{\mathcal{R}}_p^\times$, which is the intersection of the two, is the stabilizer of the geodesic path of length n connecting v_0 and v_n in the tree. Because $\mathbf{PGL}_2(\mathbb{Q}_p)$ acts strongly transitively on its Bruhat–Tits tree, it acts transitively on the set of geodesic paths of length n. Hence, the map $g_p \mapsto (g_p^{-1}.v_0, g_p^{-1}.v_n)$ is a bijection between $\widetilde{\mathcal{R}}_p^\times\backslash\mathbf{PGL}_2(\mathbb{Q}_p)$ and the set of oriented geodesic paths of length n in the tree. We need to find an upper bound on the number of paths that correspond to the image of $\mathcal{R}_p(\alpha)^\dagger$.

Denote $\operatorname{ord}_p(\alpha) = m$. If $g_p \in \mathcal{R}_p(\alpha)^\dagger$, then the existence of the Smith normal form for $g_p \in M_2(\mathbb{Z}_p)$ implies that $g_p \in \mathbf{GL}_2(\mathbb{Z}_p)\operatorname{diag}(p^{m_1},p^{m_2})\mathbf{GL}_2(\mathbb{Z}_p)$ for some $m_2 \ge m_1 \ge 0$, with $m_2+m_1 = m$. Then, $\operatorname{dist}(g_p^{-1}.v_0, v_0) = m_2 - m_1 \le m$. Hence, the number of possibilities for the first vertex of the n-path is at most the number of vertices in a ball of radius m, that is $1 + ((p+1)/(p-1))(p^m-1) \ll p^m$. Because the length of the path is n, the number of possibilities for the final vertex, after the first vertex has been fixed, is at most $\lfloor (p+1)p^{n-1}\rfloor = m_{\mathbf{G}(\mathbb{Q}_p)}(K_p)^{-1}$. We conclude that if B_p is split, then

$$m_{\mathbf{G}(\mathbb{Q}_p)}(K_p)\cdot\#\Big(K_p\backslash^{\mathcal{R}_p(p^{2k_p}\operatorname{Nr}\xi)}\Big) \ll p^{2k_p+\operatorname{ord}_p(\operatorname{Nr}\xi)} \ll |N\operatorname{Nr}\xi|_p^{-1}.$$

Multiplying the contributions from all primes p we arrive at the claimed bound. □

PROPOSITION 5.4. *Let $M\colon B_{\mathbb{A}} \to \mathbb{C}$ be a finite linear combination of standard test functions such that the component at infinity satisfies the decay condition of Proposition 3.23. Assume that M is cuspidal. Fix $\varphi,\varphi' \in L^\infty([\mathbf{G}(\mathbb{A})])$ and let $\xi_\alpha \in B$ be an arbitrary element of norm $\alpha \in \mathbb{Q}^\times$. Then,*

$$\int_{[\mathbf{G}(\mathbb{A})]}\int_{[\mathbf{G}(\mathbb{A})]}\sum_{\xi\in B} M(l^{-1}\xi r)\varphi(l)\varphi'(r)\,dl\,dr = \sum_{\alpha\in\mathbb{Q}^\times}\int_{[\mathbf{G}(\mathbb{A})]}\varphi'(r)\int_{\mathbf{G}(\mathbb{A})} M(l^{-1}\xi_\alpha r)\varphi(l)\,dl\,dr,$$

where we make the convention that if there is no $\xi_\alpha \in B$ of norm $\alpha \in \mathbb{Q}^\times$, then the summand pertaining such α is to be omitted.

Remark 5.5. The assumption that M is cuspidal is crucial here. Otherwise, there will be an additional contribution from the norm-zero elements of B. This contribution may in general diverge.

Proof. The theta series Θ_M can be rewritten as a sum over $B^\times$ due to the vanishing condition for norm-zero elements. *A priori*, we do not even know that the left-hand side is integrable. Thus,

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

we proceed with the following computation as formal operations which hold for positive-valued functions. We will then use the positive-valued case to show absolute convergence which will justify these operations in general.

Unfold first the integral along the l variable to rewrite the left-hand side above as

$$\int_{[\mathbf{G}(\mathbb{A})]} \varphi'(r) \sum_{\xi \in \mathbf{G}(\mathbb{Q}) \backslash B^\times} \int_{\mathbf{G}(\mathbb{A})} M(l^{-1}\xi r)\varphi(l)\, dl\, dr.$$

Two elements in $B^\times$ are in the same left $\mathbf{G}(\mathbb{Q})$-orbit exactly when they have the same norm. Thus, the equality in question holds if we can establish absolute integrability. To show absolute integrability, we first bound

$$\int_{[\mathbf{G}(\mathbb{A})]} \int_{[\mathbf{G}(\mathbb{A})]} \sum_{\xi \in B} \left| M(l^{-1}\xi r)\varphi(l)\varphi'(r) \right| dl\, dr \leq \|\varphi\|_\infty \|\varphi'\|_\infty \sum_{\alpha \in \mathbb{Q}^\times} \int_{[\mathbf{G}(\mathbb{A})]} \int_{\mathbf{G}(\mathbb{A})} \left| M(l^{-1}\xi_\alpha r) \right| dl\, dr.$$

By expanding the function M into finite summands of standard test functions, we reduce to the case that $M = \prod_v M_v$. Furthermore, because we are only interested in upper bounds, we can further reduce to the case that in any finite place $v = p$ the function M_v is a multiple of the characteristic function of $p^{-k_p}\mathcal{R}_p$, where $k_p = 0$ for almost all p. Taking into account this reduction, the function M is right- and left-invariant under $K_{\mathcal{R}}$ and we can apply Lemma 5.3 above. First we deduce that the integral over l vanishes unless $\alpha \in N^{-1}\mathbb{Z}$ for some fixed integer N depending only on M. In addition, by using the bound from Lemma 5.3, we can write

$$\begin{aligned}
&\int_{[\mathbf{G}(\mathbb{A})]} \int_{[\mathbf{G}(\mathbb{A})]} \sum_{\xi \in B} \left| M(l^{-1}\xi r)\varphi(l)\varphi'(r) \right| dl\, dr \\
&\ll_{\mathbf{G},\varphi,\varphi',M,\varepsilon} Nq^{\varepsilon} \sum_{0 \neq \alpha \in N^{-1}\mathbb{Z}} |\alpha| \int_{\Lambda \backslash \mathbf{G}(\mathbb{R})} \int_{\mathbf{G}(\mathbb{R})} \left| M_\infty(l^{-1}\xi_\alpha r) \right| dl\, dr \\
&= Nq^{\varepsilon} \sum_{0 \neq \alpha \in N^{-1}\mathbb{Z}} |\alpha| \int_{\Lambda \backslash \mathbf{G}(\mathbb{R})} \int_{\mathbf{G}(\mathbb{R})} \left| M_\infty(l^{-1}\xi_\alpha) \right| dl\, dr \\
&\ll_{\mathbf{G},\Lambda} Nq^{\varepsilon} \sum_{0 \neq \alpha \in N^{-1}\mathbb{Z}} |\alpha| \int_{\mathbf{G}(\mathbb{R})} \left| M_\infty(l^{-1}\xi_\alpha) \right| dl \\
&\ll Nq^{\varepsilon} \sum_{0 \neq \alpha \in N^{-1}\mathbb{Z}} |\alpha| \int_{\mathbf{G}(\mathbb{R})} \left(1 + \left\| l^{-1}\xi_\alpha \right\|\right)^{-4-\delta} dl. \qquad (25)
\end{aligned}$$

In the second line, we have made a change of variable $l \mapsto \xi_\alpha r \xi_\alpha^{-1} l$. Note that we can take here ξ_α to be any *real* matrix of determinant α, choose $\xi_\alpha = \operatorname{diag}\left(\sqrt{|\alpha|}, \operatorname{sign}(\alpha)\sqrt{|\alpha|}\right)$. The last integral in (25) can be computed using the formula for the Haar measure in Cartan coordinates

$$\begin{aligned}
&\int_{\mathbf{G}(\mathbb{R})} \left(1 + \sqrt{|\alpha|}\left\| l^{-1} \operatorname{diag}(1, \operatorname{sign}(\alpha)) \right\|\right)^{-4-\delta} dl \\
&\ll \int_0^\infty \left(1 + \sqrt{2|\alpha| \cosh(t)}\right)^{-4-\delta} \sinh(t)\, dt \\
&\ll |\alpha|^{-2-\delta/2} \int_0^\infty \frac{\sinh(t)}{\cosh(t)^{2+\delta/2}}\, dt \\
&= |\alpha|^{-2-\delta/2} \left[\frac{-1}{\cosh(t)^{1+\delta/2}(1+\delta/2)} \right]_0^\infty \ll |\alpha|^{-2-\delta/2}.
\end{aligned}$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Finally, we see that the expression in (25) is bounded from above by

$$N^{2+\delta/2}q^{\varepsilon}\sum_{0\neq n\in\mathbb{Z}}\frac{1}{|n|^{1+\delta/2}}<\infty. \qquad \square$$

PROPOSITION 5.6. *Let $M=\prod_v M_v\in\Omega$ be cuspidal and assume $M_\infty\in V_{m,2\pi}$ for $m\in\mathbb{Z}$. Fix $\varphi,\varphi'\in L^\infty([\mathbf{G}(\mathbb{A})])$. Denote*

$$F(s)=\int_{[\mathbf{G}(\mathbb{A})]}\int_{[\mathbf{G}(\mathbb{A})]}\Theta_M(l,r;s)\varphi(l)\varphi'(r)\,dl\,dr.$$

Then, $F(s)\in L^2([\mathbf{SL}_2(\mathbb{A})])$.

Proof. By Proposition 5.4, we know that $F(s)$ is well-defined pointwise and can be rewritten as

$$F(s)=\sum_{\alpha\in\mathbb{Q}^\times}\int_{[\mathbf{G}(\mathbb{A})]}\varphi'(r)\int_{\mathbf{G}(\mathbb{A})}\big(\rho(s)M\big)(l^{-1}\xi_\alpha r)\varphi(l)\,dl\,dr.$$

To calculate $\int_{[\mathbf{SL}_2(\mathbb{A})]}|F(s)|^2\,ds$, we will bound the integral over $[\mathbf{SL}_2(\mathbb{A})]$ by an integral over a Siegel set $\mathcal{S}=\mathcal{N}\cdot A^{>}\cdot\mathbf{SO}_2(\mathbb{R})\mathbf{SL}_2(\hat{\mathbb{Z}})$ as in the proof of Proposition 4.9. Because M belongs to Ω and $M_\infty\in V_{m,2\pi}$, the function M has a finite $\rho\big(\mathbf{SL}_2(\hat{\mathbb{Z}})\big)$-orbit and $\rho\big(\mathbf{SO}_2(\mathbb{R})\big)$-isotypic. Hence, it is enough to bound $\int_{\mathcal{N}\cdot A^{>}}F_1(z)\overline{F_2(z)}\,dz$ where F_1, F_2 are defined in the same manner as F but with M replaced by test functions M_1, M_2 in the $\rho\big(\mathbf{SO}_2(\mathbb{R})\mathbf{SL}_2(\hat{\mathbb{Z}})\big)$-orbit of M. Denote $a(y)=\operatorname{diag}(y^{1/2},y^{-1/2})\in\mathbf{SL}_2(\mathbb{R})$. Using the orthogonality relation of additive characters and the sup-norm bound on φ, φ', we arrive at

$$\begin{aligned}
&\left|\int_{\mathcal{N}\cdot A^{>}}F_1(z)\overline{F_2(z)}\,dz\right|\\
&\quad\ll_{\varphi,\varphi'}\sum_{\alpha\in\mathbb{Q}^\times}\int_{[\mathbf{G}(\mathbb{A})]}\int_{\mathbf{G}(\mathbb{A})}\int_{[\mathbf{G}(\mathbb{A})]}\int_{\mathbf{G}(\mathbb{A})}\\
&\qquad\int_{\sqrt{3}/2}^{\infty}\left|\rho_\infty(a(y)).M_1(l_1^{-1}\xi_\alpha r_1)\overline{\rho_\infty(a(y)).M_2(l_2^{-1}\xi_\alpha r_2)}\right|\frac{dy}{y^2}\,dl_2\,dr_2\,dl_1\,dr_1\\
&\quad=\sum_{\alpha\in\mathbb{Q}^\times}\int_{\mathbf{G}(\mathbb{A})}\int_{\mathbf{G}(\mathbb{A})}\int_{\sqrt{3}/2}^{\infty}\left|\rho_\infty(a(y)).M_1(l_1^{-1}\xi_\alpha)\overline{\rho_\infty(a(y)).M_2(l_2^{-1}\xi_\alpha)}\right|\frac{dy}{y^2}\,dl_2\,dl_1. \qquad (26)
\end{aligned}$$

As in the proof of Proposition 5.4, we apply Lemma 5.3 to the integrals in the l_1 and l_2 variables. The integral vanishes unless $\alpha\in N^{-1}\mathbb{Z}$ for some integer $N>0$ depending only on M. For every $\varepsilon>0$, we can bound (26) from above by

$$\begin{aligned}
&\varrho_{\mathbf{G}}^{-2}\sum_{0\neq\alpha\in N^{-1}\mathbb{Z}}|N\alpha|^2q^{\varepsilon}\int_{\mathbf{G}(\mathbb{R})}\int_{\mathbf{G}(\mathbb{R})}\int_{\sqrt{3}/2}^{\infty}\left|\rho_\infty(a(y)).M_1(l_1^{-1}\xi_\alpha)\overline{\rho_\infty(a(y)).M_2(l_2^{-1}\xi_\alpha)}\right|\frac{dy}{y^2}\,dl_2\,dl_1\\
&\quad=\varrho_{\mathbf{G}}^{-2}\sum_{0\neq\alpha\in N^{-1}\mathbb{Z}}|N\alpha|^2q^{\varepsilon}\int_{\mathbf{G}(\mathbb{R})}\int_{\mathbf{G}(\mathbb{R})}\int_{\sqrt{3}/2}^{\infty}\left|M_1(\sqrt{y}l_1^{-1}\xi_\alpha)\overline{M_2(\sqrt{y}l_2^{-1}\xi_\alpha)}\right|dy\,dl_2\,dl_1\\
&\quad\ll_{\mathbf{G},M}\sum_{0\neq\alpha\in N^{-1}\mathbb{Z}}|N\alpha|^2q^{\varepsilon}\int_{\mathbf{G}(\mathbb{R})}\int_{\mathbf{G}(\mathbb{R})}\int_{\sqrt{3}/2}^{\infty}\|\sqrt{y}l_1^{-1}\xi_\alpha\|^{-4-\delta}\|\sqrt{y}l_2^{-1}\xi_\alpha\|^{-4-\delta}\,dy\,dl_2\,dl_1\\
&\quad\ll\sum_{0\neq\alpha\in N^{-1}\mathbb{Z}}|N\alpha|^2q^{\varepsilon}\int_{\mathbf{G}(\mathbb{R})}\int_{\mathbf{G}(\mathbb{R})}\|l_1^{-1}\xi_\alpha\|^{-4-\delta}\|l_2^{-1}\xi_\alpha\|^{-4-\delta}\,dl_2\,dl_1.
\end{aligned}$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Take $\xi_\alpha = \mathrm{diag}(\sqrt{|\alpha|}, \mathrm{sign}(\alpha)\sqrt{|\alpha|})$ and bound the last integral from above in the same manner as in the proof of Proposition 5.4 by a multiple of $|\alpha|^{-4-\delta}$. It follows that

$$\int_{[\mathbf{SL}_2(\mathbb{A})]} |F(s)|^2\, ds \ll_{\varphi,\varphi',G,M,\varepsilon} N^{4+\delta} q^{\varepsilon} \sum_{0\neq n\in\mathbb{Z}} \frac{1}{|n|^{2+\delta}} < \infty. \qquad \square$$

5.3 The theta lift

Definition 5.7. Let $\varphi, \varphi' \in L^\infty([\mathbf{G}(\mathbb{A})])$ and fix $M \in \Omega$ cuspidal with $M_\infty \in V_{m,2\pi}$ for some $m \in \mathbb{Z}$. Define $(\varphi \otimes \overline{\varphi'})_M \colon [\mathbf{SL}_2(\mathbb{A})] \to \mathbb{C}$ by

$$(\varphi \otimes \overline{\varphi'})_M(s) \coloneqq \int_{[\mathbf{G}(\mathbb{A})]} \int_{[\mathbf{G}(\mathbb{A})]} \Theta_M(l, r; s)\varphi(l)\overline{\varphi'(r)}\, dr\, dl.$$

We call $(\varphi \otimes \overline{\varphi'})_M$ the *theta lift* of $\varphi \otimes \overline{\varphi'}$.

In the case when $\varphi = \varphi'$, we simply let $\varphi_M \coloneqq (\varphi \otimes \overline{\varphi})_M$ and call φ_M the *theta lift* of φ.

For any $\alpha \in \mathbb{Q}^\times$, we also define

$$T_\alpha^M \varphi(r) \coloneqq \begin{cases} \displaystyle\int_{\mathbf{G}(\mathbb{A})} M(l^{-1}\xi_\alpha r)\varphi(l)\, dl, & \alpha \in \mathrm{Nr}\, B^\times, \\ 0, & \text{otherwise.} \end{cases}$$

Assume $M = M_\infty M_f$ with $M_f = \prod_p M_p$. It would be useful to separate the finite and the archimedean parts in the integral above. This motivates the definition

$$T_\alpha^{M_f} \varphi(r) \coloneqq \int_{\mathbf{G}(\mathbb{A}_f)} M_f(l_f^{-1}\xi_\alpha r_f)\varphi\left(\left(\frac{\xi_\alpha}{\sqrt{|\alpha|}}\right)_\infty r_\infty \epsilon_\infty^{(1-\mathrm{sign}\,\alpha)/2} \cdot l_f\right) dl_f,$$

where $\epsilon_\infty \in B \otimes \mathbb{R}$ normalizes K_∞ and satisfies[11] $\mathrm{Nr}\,\epsilon_\infty = -1$, $\epsilon_\infty^2 = 1$. Using the change of variable $\left(\xi_\alpha/\sqrt{|\alpha|}\right)^{-1} l_\infty \epsilon_\infty^{(1-\mathrm{sign}(\alpha))/2} \mapsto l_\infty$, we arrive at

$$\begin{aligned} T_\alpha^M \varphi(r) &= \int_{\mathbf{G}(\mathbb{R})} M_\infty\left(\sqrt{|\alpha|}\epsilon_\infty^{(1-\mathrm{sign}\,\alpha)/2} l_\infty^{-1} r_\infty\right) T_\alpha^{M_f}\varphi(l_\infty r_f)\, dl_\infty \\ &= \left(T_\alpha^{M_f}\varphi \star_{\mathbf{G}(\mathbb{R})} M_\infty\left(\sqrt{|\alpha|}\epsilon_\infty^{(1-\mathrm{sign}\,\alpha)/2} \cdot \bullet\right)\right)(r). \end{aligned} \tag{27}$$

Note that by Propositions 5.4 and 5.6 the theta lift φ_M is well-defined pointwise and belongs to $L^2([\mathbf{SL}_2(\mathbb{A})])$. The proof of Proposition 5.6 implies that $T_\alpha^M \varphi$ is a square-integrable function on $[\mathbf{G}(\mathbb{A})]$ and that

$$\varphi_M(s) = \sum_{\alpha\in\mathbb{Q}^\times} \langle T_\alpha^{\rho(s)M}\varphi, \varphi\rangle \tag{28}$$

not only in an L^2-sense, but in fact pointwise.

Proposition 5.8. *Let φ and M be as in Definition 5.7. Then, for all $\alpha \in \mathbb{Q}^\times$,*

$$W_{\varphi_M}(s;\alpha) = \langle T_\alpha^{\rho(s)M}\varphi, \varphi\rangle.$$

More generally, fix $\varphi, \varphi' \in L^\infty([\mathbf{G}(\mathbb{A})])$. Then, $W_{(\varphi\otimes\overline{\varphi'})_M}(s;\alpha) = \langle T_\alpha^{\rho(s)M}\varphi, \varphi'\rangle$, and

$$(\varphi \otimes \overline{\varphi'})_M(s) = \sum_{\alpha\in\mathbb{Q}^\times} \langle T_\alpha^{\rho(s)M}\varphi, \varphi'\rangle. \tag{29}$$

[11] Such an element does not exist if B is ramified at infinity, but then there are also no elements $\xi_\alpha \in B$ of negative norm.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Proof. We only establish the second claim as it immediately implies the first. Proposition 5.6 implies $(\varphi \otimes \overline{\varphi'})_M(s) \in L^2([\mathbf{SL}_2(\mathbb{A})])$. We then apply Propositions 5.4 to deduce (29).

We use the notation $u_n \coloneqq \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Fubini's theorem and the orthogonality of characters imply for all $\alpha, \beta \in \mathbb{Q}^\times$, $s \in \mathbf{SL}_2(\mathbb{A})$, and $x \in \mathbf{G}(\mathbb{A})$

$$\int_{[\mathbf{N}(\mathbb{A})]} T_\alpha^{\rho(u_n s)M} \varphi(x) \psi(-\beta n)\, dn = \begin{cases} T_\alpha^{\rho(s)M}\varphi(x), & \alpha = \beta, \\ 0, & \alpha \neq \beta. \end{cases}$$

The claim follows from substituting this expression in the definition of the Whittaker function applied to (29). □

5.3.1 *Hecke operators.* We would like to describe the relation between the Fourier–Whittaker expansion of φ_M and the Hecke translates of φ. A minor difficulty is that the Hecke algebra of $\mathbf{G}(\mathbb{A})$ is not rich enough and we would prefer to work with the Hecke algebra of the adjoint group $\mathbf{G}^{\mathrm{adj}}(\mathbb{A})$. To that end, we lift a $K_{\mathscr{R}}$-invariant function on $[\mathbf{G}(\mathbb{A})]$ to $[\mathbf{G}^{\mathrm{adj}}(\mathbb{A})]$. An alternative more conceptual approach is to work with a $\mathbf{PGL}_2$ Weil representation, cf. [Wal85, § I.3] and [Nel17, § 2.2.5].

Let us recall that the adjoint group is the affine algebraic group over $\mathbb{Q}$ representing the functor

$$\mathbf{G}^{\mathrm{adj}}(L) \coloneqq {}_{L^\times}\backslash^{(B \otimes L)^\times}$$

for any $\mathbb{Q}$-algebra L, where $L^\times$ is embedded centrally in $(B \otimes L)^\times$. We will also use the algebraic group $\mathbf{B}^\times(L) = (B \otimes L)^\times$, i.e. $\mathbf{G}^{\mathrm{adj}} = Z_{\mathbf{B}^\times}\backslash\mathbf{B}^\times$.

DEFINITION 5.9. For each finite place v denote by $\widetilde{K}_v$ the image of $\mathscr{R}_v^\times$ in $\mathbf{G}^{\mathrm{adj}}(\mathbb{Q}_v) = \mathbb{Q}_v^\times\backslash B_v^\times$. We further let $\widetilde{K}_{\mathscr{R}} \coloneqq \prod_{v<\infty} \widetilde{K}_v$.

PROPOSITION 5.10. *The natural map*

$$\mathbf{G}(\mathbb{Q})\backslash^{\mathbf{G}(\mathbb{A})}/_{K_{\mathscr{R}}} \to \mathbf{G}^{\mathrm{adj}}(\mathbb{Q})\backslash^{\mathbf{G}^{\mathrm{adj}}(\mathbb{A})}/_{\widetilde{K}_{\mathscr{R}}}$$

is a measure-preserving bijection. In particular, we have a Hilbert space isomorphism between $L^2([\mathbf{G}(\mathbb{A})])^{K_{\mathscr{R}}}$ *and* $L^2([\mathbf{G}^{\mathrm{adj}}(\mathbb{A})])^{K_{\mathscr{R}}}$.

Proof. Denote by $h\colon \mathbf{G} \to \mathbf{G}^{\mathrm{adj}}$ the standard isogeny. The image is a normal subgroup and the quotient is abelian. The kernel of the map h is the center $\mathbf{Z} < \mathbf{G}$. The center is isomorphic to the group of second-order roots of unity μ_2. The reduced norm map then completes the exact sequence

$$1 \to \mu_2 \to \mathbf{G} \xrightarrow{h} \mathbf{G}^{\mathrm{adj}} \xrightarrow{\mathrm{Nr}} {}_{\mathbb{G}_m^2}\backslash^{\mathbb{G}_m} \to 1.$$

For a local field or a number field F the image of $\mathrm{Nr}((B \otimes F)^\times)$ in $F^\times$, is determined by the Hasse–Schilling–Maass theorem. In particular, $\mathrm{Nr}((B \otimes F)^\times) = F^\times$ if $F = \mathbb{Q}_p$, or $F = \mathbb{R}$ and B is indefinite. If B is definite, then $\mathrm{Nr}((B \otimes \mathbb{R})^\times) = \mathbb{R}_{>0}$. Finally, $\mathrm{Nr}((B \otimes \mathbb{Q})^\times)$ is $\mathbb{Q}^\times$ if B is indefinite and $\mathbb{Q}_{>0}$ otherwise. It follows that $h(\mathbf{G}(\mathbb{A}))\backslash\mathbf{G}^{\mathrm{adj}}(\mathbb{A}) \xrightarrow[\mathrm{Nr}]{\sim} \mathbb{A}^{\times 2}\backslash\mathbb{A}^\times$ if B is indefinite and

$$h(\mathbf{G}(\mathbb{A}))\backslash\mathbf{G}^{\mathrm{adj}}(\mathbb{A}) \xrightarrow[\mathrm{Nr}]{\sim} \mathbb{A}^{\times 2}\backslash\left(\mathbb{R}_{>0} \times \mathbb{A}_f^\times\right) \simeq \mathbb{A}_f^{\times 2}\backslash\mathbb{A}_f^\times$$

if B is definite.

Injectivity. Assume $h(g') = \tilde{\gamma} h(g) \tilde{k}$ for some $g, g' \in \mathbf{G}(\mathbb{A})$, $\tilde{\gamma} \in \mathbf{G}^{\mathrm{adj}}(\mathbb{Q})$, and $\tilde{k} \in \widetilde{K}_{\mathscr{R}}$. We need to show $[g] = [g']$ in $\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})/K_{\mathscr{R}}$. To show that $\tilde{\gamma} \in h(\mathbf{G}(\mathbb{Q}))$ we establish that $\mathrm{Nr}(\tilde{\gamma})$ is a square in $\mathbb{Q}^\times$, this can be checked locally at all places. Examining the archimedean component

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

of the equality, we arrive at $\tilde{\gamma} = h(g'_\infty g_\infty^{-1})$. Hence, $\operatorname{Nr}\tilde{\gamma}$ is positive. Similarly, at all finite places $p < \infty$ we can write $\tilde{\gamma} = h(g'_p g_p^{-1})\tilde{k}_p^{-1}$ and $|\operatorname{Nr}\tilde{\gamma}|_p = |h(g'_p g_p^{-1})|_p \in p^{2\mathbb{Z}}$. Thus, $\tilde{\gamma} = h(\gamma)$ for some $\gamma \in \mathbf{G}(\mathbb{Q})$. We can now write $h(g') = h(\gamma g)\tilde{k}$. Hence, $\operatorname{Nr}\tilde{k} = 1$ as well, and $\tilde{k} = h(k)$ for some $k \in K_{\mathscr{R}}$. We deduce $h(g') = h(\gamma g k)$, and $[g'] \in [\mathbf{Z}(\mathbb{A})g]$ in $\mathbf{G}(\mathbb{Q})\backslash\mathbf{G}(\mathbb{A})/K_{\mathscr{R}}$.

To conclude $[g'] = [g]$ it is enough to show that $\mathbf{Z}(\mathbb{Q})\backslash\mathbf{Z}(\mathbb{A})/\prod_p \left(K_p \cap \mathbf{Z}(\mathbb{Q}_p)\right)$ is a trivial group. Because K_p contains $\pm\operatorname{Id}$ for all p, this group is $\pm 1\backslash\mathbb{A}^\times[2]/\prod_p \mathbb{Z}_p^\times[2] \simeq 1$ as required.

Surjectivity. Using the norm map, it is enough to demonstrate that $\mathbb{A}^{\times 2}\mathbb{Q}^\times\backslash\mathbb{A}^\times/\prod_p \operatorname{Nr}\widetilde{K}_p$ is trivial. Lemma 5.2 implies that $\operatorname{Nr}\widetilde{K}_p = \mathbb{Z}_p^{\times 2}\backslash\mathbb{Z}_p^\times$ for all p. Because $\mathbb{Q}$ has class number 1, the double quotient is isomorphic to $\mathbb{R}^{\times 2}\backslash\mathbb{R}^\times/\mathbb{Z}^\times \simeq 1$.

Measure preservation. Strong approximation implies that the group $\mathbf{G}(\mathbb{R})$ acts transitively on the left-hand side in the claimed equality. Hence, it acts transitively on the right-hand side as well because the map is equivariant. The Haar measure on both spaces is a $\mathbf{G}(\mathbb{R})$-invariant probability measure on a locally compact homogeneous $\mathbf{G}(\mathbb{R})$-space. Uniqueness of Haar measure implies that the map is measure preserving. □

DEFINITION 5.11. Let $\varphi\colon [\mathbf{G}(\mathbb{A})] \to \mathbb{C}$ be $K_{\mathscr{R}}$-invariant. Denote by $\widetilde{\varphi}\colon [\mathbf{G}^{\mathrm{adj}}(\mathbb{A})] \to \mathbb{C}$ its unique lift to a $\widetilde{K}_{\mathscr{R}}$-invariant function on $[\mathbf{G}^{\mathrm{adj}}(\mathbb{A})]$.

We use the notation $\widetilde{\Gamma} = \mathbf{G}^{\mathrm{adj}}(\mathbb{Q}) \cap \widetilde{K}_{\mathscr{R}}$, which is a lattice in $\mathbf{G}^{\mathrm{adj}}(\mathbb{R})$. Equivalently, $\widetilde{\Gamma}$ is the image of $\mathscr{R}^\times$ in $\mathbf{G}^{\mathrm{adj}}(\mathbb{R})$. Proposition 5.10 implies that $\mathbf{G}^{\mathrm{adj}}(\mathbb{R})$ acts transitively on $\mathbf{G}^{\mathrm{adj}}(\mathbb{Q})\backslash\mathbf{G}^{\mathrm{adj}}(\mathbb{A})/\widetilde{K}_{\mathscr{R}}$ and $\mathbf{G}^{\mathrm{adj}}(\mathbb{Q})\backslash\mathbf{G}^{\mathrm{adj}}(\mathbb{A})/\widetilde{K}_{\mathscr{R}} \simeq \widetilde{\Gamma}\backslash\mathbf{G}^{\mathrm{adj}}(\mathbb{R})$. We introduce Hecke operators adapted to the order $\mathscr{R}$.

DEFINITION 5.12. Let $\alpha \in \mathbb{Q}$ and $f\colon \widetilde{\Gamma}\backslash\mathbf{G}^{\mathrm{adj}}(\mathbb{R}) \to \mathbb{C}$ continuous. Set

$$\mathscr{R}(\alpha) = \{x \in \mathscr{R} \mid \operatorname{Nr} x = \alpha\}, \mathscr{R}(\alpha)^\dagger = \{x \in \mathscr{R} \mid |\operatorname{Nr} x|_\infty = |\alpha|_\infty\}$$

and define

$$T_\alpha f(g) = \sum_{[\delta^+] \in \mathscr{R}(1)\backslash\mathscr{R}(\alpha)} f\left(\delta^+ g \epsilon_\infty^{(1-\operatorname{sign}\alpha)/2}\right) = \sum_{[\delta] \in \mathscr{R}^\times\backslash\mathscr{R}(\alpha)^\dagger} f\left(\delta g \epsilon_\infty^{(1-\operatorname{sign}\alpha)/2}\right).$$

The two expressions are equal because $\mathscr{R}^\times$ contains an element of determinant -1 if B is indefinite. These operators coincide with the classical Hecke operators for $\alpha > 0$ co-prime to qD_B. Note that if $\alpha \notin \operatorname{Nr}\mathscr{R}$ then $T_\alpha = 0$.

LEMMA 5.13. *Let $\alpha \in \mathbb{Q}^\times$ and $f\colon [\mathbf{G}^{\mathrm{adj}}(\mathbb{A})] \to \mathbb{C}$ continuous and $\widetilde{K}_{\mathscr{R}}$-invariant. Set*

$$\mathscr{R}_f(\alpha)^\dagger = \prod_p \left\{x_p \in \mathscr{R}_p \mid |\operatorname{Nr} x_p|_p = |\operatorname{Nr}\alpha|_p\right\}.$$

Then, for every $g \in \mathbf{G}^{\mathrm{adj}}(\mathbb{R})$,

$$\left(T_\alpha f(\bullet\widetilde{K}_{\mathscr{R}})\right)(g) = \left(f \star \mathbb{1}_{\mathscr{R}_f(\alpha)^\dagger}\right)(g\epsilon_\infty^{(1-\operatorname{sign}\alpha)/2}),$$

where the convolution takes place in $\mathbf{B}^\times(\mathbb{A}_f)$ with the measure normalization $m_{\mathbf{B}^\times(\mathbb{A}_f)}(\mathscr{R}_f^\times) = 1$.

Proof. The right $\widetilde{K}_{\mathscr{R}}$-invariance of f and the left $\mathscr{R}_f^\times$-invariance of $\mathbb{1}_{\mathscr{R}_f(\alpha)^\dagger}$ imply

$$\left(f \star \mathbb{1}_{\mathscr{R}_f(\alpha)^\dagger}\right)(g) = \int_{\mathscr{R}_f(\alpha)^\dagger} f(gh_f^{-1})\, dh_f = \sum_{[a_f] \in \mathscr{R}_f^\times\backslash\mathscr{R}_f(\alpha)^\dagger} f(ga_f^{-1}).$$

There is a natural map $\mathscr{R}^\times\backslash\mathscr{R}(\alpha)^\dagger \to \mathscr{R}_f^\times\backslash\mathscr{R}_f(\alpha)^\dagger$. Strong approximation implies that this map is surjective. To show this map is also injective, we observe that if $\delta \equiv \delta' \bmod \mathscr{R}_f^\times$ for $\delta, \delta' \in \mathscr{R}(\alpha)^\dagger$,

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

then $\delta\delta'^{-1}, \delta'\delta^{-1} \in B \cap \mathcal{R}_f^\times \subset \mathcal{R}$, and $\delta\delta'^{-1} \in \mathcal{R}^\times$. By choosing a rational representative for each coset in $\mathcal{R}_f^\times \backslash \mathcal{R}_f(\alpha)^\dagger$ and using the left $\mathbf{G}^{\mathrm{adj}}(\mathbb{Q})$-invariance of f, we arrive at the claim. □

PROPOSITION 5.14. *Let $0 \neq \alpha \in \mathbb{Q}^\times$ and assume $M_f = \prod_p \mathbb{1}_{\mathcal{R}_p}$. Then, for φ and $\widetilde{\varphi}$ as in Definition 5.11, we have*

$$\widetilde{T_\alpha^{M_f}\varphi}(r) = T_\alpha \widetilde{\varphi}(r).$$

Proof. Assume $\alpha \in \operatorname{Nr}\mathcal{R}$, otherwise the claim is trivial. Although we claim the equality for all $r \in [\mathbf{G}^{\mathrm{adj}}(\mathbb{A})]$, because of the uniqueness of the lift in Definition 5.11, it is enough to verify the claim for $r \in [\mathbf{Z}\backslash\mathbf{G}(\mathbb{A})]$. We apply Lemma 5.13 and evaluate the convolution by decomposing the Haar measure on $\mathbf{B}^\times(\mathbb{A}_f)$ into fibers over $\mathbf{G}(\mathbb{A}_f)\backslash\mathbf{B}^\times(\mathbb{A}_f)$, this is possible because $\mathbf{B}^\times(\mathbb{A}_f)$ and $\mathbf{G}(\mathbb{A}_f)$ are unimodular. For consistent measure normalization, we set $m_{\mathbf{B}^\times(\mathbb{A}_f)}(\mathcal{R}_f^\times) = m_{\mathbf{G}(\mathbb{A}_f)}(K_f) = m_{\operatorname{Nr} B_{\mathbb{A}_f}^\times}(\widehat{\mathbb{Z}}^\times) = 1$:

$$\begin{aligned}
\widetilde{\varphi} \star \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(r) &= \int_{\mathbf{B}^\times(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(l_f^{-1} r_f)\widetilde{\varphi}(r_\infty l_f)\, dl_f \\
&= \int_{\mathbf{G}(\mathbb{A}_f)\backslash\mathbf{B}^\times(\mathbb{A}_f)} \int_{\mathbf{G}(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(\lambda^{-1} l_f^{-1} r_f)\widetilde{\varphi}(r_\infty l_f \lambda)\, dl_f\, d\lambda \\
&= \int_{\mathbf{G}(\mathbb{A}_f)\backslash\mathbf{B}^\times(\mathbb{A}_f)/\mathcal{R}_f^\times} \int_{\mathbf{G}(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(\lambda^{-1} l_f^{-1} r_f)\widetilde{\varphi}(r_\infty l_f \lambda)\, dl_f\, d\lambda.
\end{aligned}$$

In the last line, we have used the fact that $\mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(l_f^{-1} r_f)$ is left $\mathcal{R}_f^\times$-invariant and $\widetilde{\varphi}$ is right $\widetilde{K}_f$-invariant. Fix $\xi_\alpha \in B^\times$ with $\operatorname{Nr}\xi_\alpha = \alpha$. Because $\operatorname{Nr} l_f^{-1} r_f = 1$ and $\operatorname{Nr}\mathcal{R}_f(\alpha)^\dagger = \hat{\mathbb{Z}}^\times \alpha$, the inner integral vanishes unless $\lambda \equiv \xi_\alpha^{-1} \bmod \mathbf{G}(\mathbb{A}_f)\backslash\mathbf{B}^\times(\mathbb{A}_f)/\mathcal{R}_f^\times$. We conclude that

$$\begin{aligned}
\widetilde{\varphi} \star \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(r) &= \int_{\mathbf{G}(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(\xi_\alpha l_f^{-1} r_f)\widetilde{\varphi}(r_\infty l_f (\xi_\alpha)_f^{-1})\, dl_f \\
&= \int_{\mathbf{G}(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(l_f^{-1} \xi_\alpha r_f)\widetilde{\varphi}(r_\infty (\xi_\alpha)_f^{-1} l_f)\, dl_f \\
&= \int_{\mathbf{G}(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(l_f^{-1} \xi_\alpha r_f)\widetilde{\varphi}((\xi_\alpha)_\infty r_\infty l_f)\, dl_f \\
&= \int_{\mathbf{G}(\mathbb{A}_f)} \mathbb{1}_{\mathcal{R}_f(\alpha)^\dagger}(l_f^{-1} \xi_\alpha r_f)\widetilde{\varphi}\left(\left(\frac{\xi_\alpha}{\sqrt{|\alpha|}}\right)_\infty r_\infty l_f\right) dl_f \\
&= T_\alpha^{M_f}\varphi(r\epsilon_\infty^{(1-\operatorname{sign}\alpha)/2}),
\end{aligned}$$

where in the first line we have used the change of variables $\xi_\alpha l_f \xi_\alpha^{-1} \mapsto l_f$ and in the second line we have applied the left $\mathbf{G}^{\mathrm{adj}}(\mathbb{Q})$-invariance of $\widetilde{\varphi}$. □

COROLLARY 5.15. *Let $\varphi, \varphi' \in L^\infty([\mathbf{G}(\mathbb{A})])^{K_\mathcal{R}}$ and $\alpha \in \mathbb{Q}^\times$. Assume $T_\alpha\varphi = \lambda(\alpha)\varphi$, $M_f = \prod_p \mathbb{1}_{\mathcal{R}_p}$, and $M_\infty \in V_{m,2\pi}$ for some $m \in \mathbb{Z}$. Then, the function*

$$F(s) = \int_{[\mathbf{G}(\mathbb{A})]} \int_{[\mathbf{G}(\mathbb{A})]} \Theta_M(l, r; s)\varphi(l)\overline{\varphi'(r)}\, dr\, dl$$

satisfies

$$W_F(s_\infty U_\mathcal{R}; \alpha) = \lambda(\alpha)\left\langle \varphi \star \left(\rho(s_\infty).M_\infty\left(\sqrt{|\alpha|}\epsilon_\infty^{(1-\operatorname{sign}\alpha)/2} \cdot \bullet\right)\right), \varphi'\right\rangle$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

for all $s_\infty \in \mathbf{SL}_2(\mathbb{R})$, *where the convolution takes place in* $\mathbf{G}(\mathbb{R})$ *and where we recall* $U_{\mathscr{R}}$ *from Definition 4.4.*

Proof. Recall $\rho(U_{\mathscr{R}}).\mathbb{1}_{\mathscr{R}_f} = \mathbb{1}_{\mathscr{R}_f}$ from the sentence immediately below Definition 4.4. Propositions 5.8, 5.14, and (27) then imply that

$$\begin{aligned} W_F(s_\infty U_{\mathscr{R}}; \alpha) &= \big\langle T_\alpha^{M_f} \varphi \star \big(\rho(s_\infty).M_\infty\big(\sqrt{|\alpha|}\epsilon_\infty^{(1-\operatorname{sign}\alpha)/2} \cdot \bullet\big)\big), \varphi' \big\rangle \\ &= \lambda(\alpha)\big\langle \varphi \star \big(\rho(s_\infty).M_\infty\big(\sqrt{|\alpha|}\epsilon_\infty^{(1-\operatorname{sign}\alpha)/2} \cdot \bullet\big)\big), \varphi' \big\rangle. \end{aligned}$$ □

6. The Bergman kernel

6.1 The Bergman Archimedean test function

From now on, we shall assume that B is split over $\mathbb{R}$. Recall that we have fixed an isomorphism $B_\infty \simeq \mathrm{Mat}_{2\times 2}(\mathbb{R})$ and have used it to identify the two spaces. We construct a theta series whose Fourier–Whittaker coefficients coincide with the Bergman kernel. For this endeavor, we will use the following archimedean test function. We fix the global character ψ as in § 3.4.

Definition 6.1. Fix a weight $m \geq 2$ and define

$$M_\infty^{(m)}(x) = \exp(-2\pi \operatorname{Nr} x) \begin{cases} \dfrac{\operatorname{Nr}(x)^{m-1}}{\left(\frac{(b-c)+i(a+d)}{2i}\right)^m}, & \operatorname{Nr} x > 0, \\ 0, & \operatorname{Nr} x \leq 0, \end{cases}$$

for $x = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Note that $M_\infty^{(m)}({}^\iota x) = \overline{M_\infty^{(m)}}(x)$.

Set $\mu\colon \mathbf{PGL}_2(\mathbb{R}) \to \mathbb{C}$

$$\mu(x) = \begin{cases} \dfrac{2i\sqrt{\operatorname{Nr} x}}{(b-c)+i(a+d)}, & \operatorname{Nr} x > 0, \\ 0, & \operatorname{Nr} x \leq 0. \end{cases}$$

Then, we can write $M_\infty^{(m)}(x) = \exp(-2\pi \operatorname{Nr} x) \operatorname{Nr}(x)^{m/2-1}\mu(x)^m$.

Lemma 6.2. *Let* $k_\theta := \left(\begin{smallmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{smallmatrix}\right) \in \mathbf{SO}_2(\mathbb{R})$. *Then, for every* $g \in \mathbf{PGL}_2(\mathbb{R})$,

$$\mu(gk_\theta) = \mu(g)e^{i\theta}. \tag{30}$$

Proof. We assume $\operatorname{Nr} g > 0$ as the claim is trivial for non-positive determinants. We verify the claim by a direction computation. Let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $gk_\theta = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right)$. We have

$$\begin{aligned} (b'-c') + i(a'+d') &= \big(a\sin\theta + b\cos\theta - c\cos\theta + d\sin\theta\big) + i\big(a\cos\theta - b\sin\theta + c\sin\theta + d\cos\theta\big) \\ &= \big((b-c)+i(a+d)\big)\big(\cos\theta - i\sin\theta\big). \end{aligned}$$

Hence, $\mu(gk_\theta) = \mu(g)e^{i\theta}$. □

Corollary 6.3. *For every weight* $m \geq 2$ *and* $k_{\theta_1}, k_{\theta_2} \in \mathbf{SO}_2(\mathbb{R})$,

$$M_\infty^{(m)}(k_{\theta_1} x k_{\theta_2}) = e^{im(\theta_2+\theta_1)} M_\infty^{(m)}(x).$$

Proof. Apply the previous lemma to $M_\infty^{(m)}(x) = \exp(-2\pi \operatorname{Nr} x) \operatorname{Nr}(x)^{m/2-1}\mu(x)^m$ and use the identity $M_\infty^{(m)}(x^\iota) = \overline{M_\infty^{(m)}(x)}$. □

Lemma 6.4. *If* $m \geq 2$, *then*

$$|M_\infty^{(m)}(x)| \ll_m (1 + \|x\|)^{-m}.$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Proof. This is trivial if $\operatorname{Nr} x \le 0$, hence we assume the determinant is positive. Denote $r = \|x\| = \sqrt{\operatorname{Tr}(x^t x)}$. A simple calculation shows that

$$|\mu(x)|^{-2} = \frac{r^2 + 2\operatorname{Nr}(x)}{4\operatorname{Nr}(x)},$$

$$|M_\infty^{(m)}(x)| = \exp(-2\pi \operatorname{Nr} x)\frac{2^m \operatorname{Nr}(x)^{m-1}}{(r^2 + 2\operatorname{Nr}(x))^{m/2}}.$$

If $r \le 1$, then we write

$$|M_\infty^{(m)}(x)| \ll_m \exp(-2\pi \operatorname{Nr} x)\operatorname{Nr}(x)^{m/2-1} \ll 1.$$

The last equality holds for all $\operatorname{Nr}(x) > 0$ because we have assumed $m \ge 2$. Otherwise, if $r > 1$, then

$$|M_\infty^{(m)}(x)| \le \exp(-2\pi \operatorname{Nr} x)2^m \operatorname{Nr}(x)^{m-1} r^{-m} \ll_m r^{-m},$$

where we have used the fact that the real function $\exp(-2\pi t)t^{m-1}$ is bounded for $t \in [0, \infty)$ and $m \ge 1$. □

Up until this point, we have established that $M_\infty^{(m)}$ satisfies the decay condition in Definition 3.7 if $m > 4$. We now turn to show that it also belongs to the space $V_{m,2\pi}$ by checking that it solves the quantum harmonic oscillator, see (18).

LEMMA 6.5. *If $m \ge 6$, then the function $M_\infty^{(m)}$ from Definition 6.1 belongs to $V_{m,2\pi}$.*

Proof. Lemma 6.4 above implies that $M_\infty^{(m)} \in L^2(B_\infty)$ if $m > 4$. To prove $M_\infty^{(m)} \in V_{m,2\pi}$ we will show that $M_\infty^{(m)} \perp V_{m',2\pi}$ for all $m' \ne m$. It is enough to establish $\langle M_\infty^{(m)}, M' \rangle = 0$ for any Schwartz solution M' of (18) with quantum number $m' \ne m$ and $\omega = 2\pi$.

Define $F(x) = \exp(-2\pi \operatorname{Nr}(x))N(x)$ and $N(x) = \operatorname{Nr}(x)^{m-1}(2i)^m\big((b-c) + i(a+d)\big)^{-m}$. Then, $F(x)$ is a well-defined continuous function on the open subset $\mathcal{U} = B_\infty \setminus \left\{ \left(\begin{smallmatrix} a & b \\ b & -a \end{smallmatrix}\right) : a, b \in \mathbb{R} \right\}$. Moreover, $M_\infty^{(m)} = \mathbb{1}_{\operatorname{Nr}(x)>0} \cdot F$. Define $V = \{x \in B_\infty \mid \operatorname{Nr} x \ge 0\}$ and $V_R = V \cap B(0,R) \setminus B(0, R^{-1})$ for $R > 1$, where $B(0,r)$ is a closed ball of radius r centered at the origin. Note that $V_R \subset \mathcal{U}$.

We claim that $F(x)$ solves the PDE (18) on $\mathcal{U}$ with $\omega = 2\pi$ and quantum number m. The PDE (18) with $\omega = 2\pi$, $\varpi = 1$ for F is equivalent to the following PDE for N:

$$-\Delta N + 2\pi\langle x, \nabla\rangle N = 2\pi(m-2)N, \tag{31}$$

where ∇ denotes the gradient operator and the bilinear form $\langle x_1, x_2\rangle$ is the twisted trace form $\operatorname{Tr}(x_1{}^\iota x_2)$ as before. Using the definition of the Laplace operator as the Fourier multiplier with symbol $-4\pi^2 \operatorname{Nr}$ and the definition of the gradient, we arrive at

$$\Delta = \frac{\partial^2}{\partial a \partial d} - \frac{\partial^2}{\partial b \partial c},$$

$$\langle x, \nabla\rangle = a\frac{\partial}{\partial a} + d\frac{\partial}{\partial d} + b\frac{\partial}{\partial b} + c\frac{\partial}{\partial c}.$$

Substituting the definition $N(x) = (2i)^m((ad-bc)^{m-1}/((b-c)+i(a+d))^m)$ into the formulae above we see that

$$\Delta N = 0,$$

$$\langle x, \nabla\rangle N = (m-2)N.$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

These and (31) show that F is a solution with quantum number m. Because the PDE (18) is local, this establishes that[12] $L_{2\pi}[M_\infty^{(m)}] = 2\pi m M_\infty^{(m)}$.

Fix a Schwartz function $M' \colon B_\infty \to \mathbb{C}$. We want to show now that $\langle L_{2\pi}[M_\infty^{(m)}], M'\rangle = \langle M_\infty^{(m)}, L_{2\pi}[M']\rangle$. The equality $\langle \mathrm{Nr}(x) M_\infty^{(m)}, M'\rangle = \langle M_\infty^{(m)}, \mathrm{Nr}(x) M'\rangle$ is obvious. We need only show

$$\langle \Delta M_\infty^{(m)}, M'\rangle = \langle M_\infty^{(m)}, \Delta M'\rangle. \tag{32}$$

Note that the integrals defining these individual inner products are absolutely convergent because M' and $\Delta M'$ are Schwartz, and $M_\infty^{(m)}$, $\mathrm{Nr}(x) M_\infty^{(m)}$, $\Delta M_\infty^{(m)}$ have at most polynomial growth. To establish (32) we use the equality

$$\langle \Delta M_\infty^{(m)}, M'\rangle = \lim_{R\to\infty} \int_{V_R} \Delta F(x) \overline{M'(x)}\, dx,$$

and the analogous formula for $\langle M_\infty^{(m)}, \Delta M'\rangle$. These follow from the dominated convergence theorem. Denote

$$W = \begin{pmatrix} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/2 \\ 0 & 0 & -1/2 & 0 \end{pmatrix}$$

and write $\Delta = \nabla \cdot (W\nabla)$ with respect to the coordinates (a, d, b, c). Using the symmetry of the matrix W and the divergence theorem we arrive at

$$\begin{aligned} \int_{V_R} \Delta F(x) \overline{M'(x)}\, dx &= \int_{V_R} \nabla \cdot \left(\overline{M'(x)} W \nabla F\right) dx - \int_{V_R} (W\nabla F) \cdot (\nabla \overline{M'(x)})\, dx \\ &= \int_{V_R} \nabla \cdot \left(\overline{M'(x)} W \nabla F - F W \nabla \overline{M'(x)}\right) dx + \int_{V_R} F(x) \Delta \overline{M'(x)}\, dx \\ &= \int_{\partial V_R} \left(\overline{M'(x)} W \nabla F - F W \nabla \overline{M'(x)}\right) \cdot \hat{n}\, dA(x) + \int_{V_R} F(x) \Delta \overline{M'(x)}\, dx. \end{aligned}$$

The conditions of the divergence theorem are satisfied because F and M are smooth in an open neighborhood of the closure of V_R, which is compact, and the boundary ∂V_R is piecewise smooth. A direct computation, as in the proof of Lemma 6.4, shows that F and ∇F vanish on the boundary of the cone V, except perhaps the origin where they remain bounded. It remains to consider the contributions from the surfaces $S_R = \partial V_R \cap B(0, R)$ and $s_R = \partial V_R \cap B(0, 1/R)$. The area of S_R is bounded from above by the area of a 3-sphere of radius R, thus $\mathrm{Area}(S_R) \ll R^3$. On the other hand, because M' is Schwartz and $F, \nabla F$ are bounded on V we have that

$$\left| \int_{S_R} \left(\overline{M'(x)} W \nabla F - F W \nabla \overline{M'(x)}\right) \cdot \hat{n}\, dA(x) \right| \ll_N R^{-N} R^3 \to_{R\to\infty} 0.$$

Similarly, $\mathrm{Area}(s_R) \ll R^{-3}$ and $\overline{M'(x)} W \nabla F - F W \nabla \overline{M'(x)}$ is bounded on $B(0,1) \cap V$, hence the surface integral over s_R converges to 0 as $R \to \infty$.

Now let $M' \in V_{m',2\pi}$ be a Schwartz function and assume $m' \neq m$, then we have

$$2\pi m \langle M_\infty^{(m)}, M'\rangle = \langle L_{2\pi}[M_\infty^{(m)}], M'\rangle = \langle M_\infty^{(m)}, L_{2\pi}[M']\rangle = 2\pi m' \langle M_\infty^{(m)}, M'\rangle$$

and we deduce $\langle M_\infty^{(m)}, M'\rangle = 0$ as necessary. □

[12] The function $M_\infty^{(m)}$ is not necessarily in $C^2(B_\infty)$ and we extend the definition of $L_{2\pi}$ to this function in the naive way, in particular this equality does not need to be well-defined on the cone $\mathrm{Nr}\, x = 0$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

COROLLARY 6.6. *The Bergman test function of weight $m \geq 6$ belongs to Ω_∞.*

Proof. Lemma 6.4 implies that this test function satisfies the decay condition in the definition of Ω_∞ and Lemma 6.5 above implies that the Bergman test function transforms under $\rho(\mathbf{SO}_2(\mathbb{R}))$ by a character. □

7. The spectral expansion

Fix a global Eichler order $\mathcal{R} = \mathcal{R}_1 \cap \mathcal{R}_2 \subset B$ and a weight $m > 2$. We focus henceforth on the test function $M = M_\infty^{(m)} \cdot \prod_p \mathbb{1}_{\mathcal{R}_p} \in \Omega$. This test function is cuspidal as $M_\infty^{(m)}(x) = 0$ if $\operatorname{Nr} x = 0$ and we denote the classical theta series attached to the test function M by Proposition 4.7 by

$$\vartheta^{(m)}(l, r; s) = \sum_{\xi \in \mathcal{R}} \big(\rho(s) M_\infty^{(m)}\big)(l^{-1} \xi r)$$

for $s \in \Lambda \backslash \mathbf{SL}_2(\mathbb{R})$, $l, r \in \Gamma \backslash \mathbf{G}(\mathbb{R})$. In this section, we prove the main theorem about the spectral expansion of $\vartheta^{(m)}$.

DEFINITION 7.1. Denote by $S_m(\Gamma)$ the space of Γ-modular weight m modular forms on $\mathbb{H}$. Write $S_m(\Gamma) = S_m^{\mathrm{old}}(\Gamma) \oplus S_m^{\mathrm{new}}(\Gamma)$ for the direct sum decomposition into new and old forms. The decomposition is orthogonal with respect to the Petersson inner product.

If $f \in S_m(\Gamma)$, we denote by $f^\sharp \colon \Gamma \backslash \mathbf{G}(\mathbb{R}) \to \mathbb{C}$ the automorphic lift of f. Specifically, if $g = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) \left(\begin{smallmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{smallmatrix}\right) k_\theta$, then $f^\sharp(g) = y^{m/2} e^{im\theta} f(x + iy)$. Following the discussion in § 4.2, we shall also consider $f^\sharp$ as a $K_{\mathcal{R}}$-invariant function on $[\mathbf{G}(\mathbb{A})]$.

Similarly, we decompose $S_m(\Lambda) = S_m^{\mathrm{old}}(\Lambda) \oplus S_m^{\mathrm{new}}(\Lambda)$, and denote by $f^{*\sharp} \colon \Lambda \backslash \mathbf{SL}_2(\mathbb{R}) \to \mathbb{C}$ the automorphic lift of $f^* \in S_m(\Lambda)$. Moreover, we shall also consider $f^{*\sharp}$ as a $U_{\mathcal{R}}$-invariant function on $[\mathbf{SL}_2(\mathbb{A})]$.

THEOREM 7.2. *Fix an orthonormal basis $\mathcal{B}_m$ for $S_m(\Gamma)$ of Hecke eigenforms for all Hecke operators T_n with $(n, qD_B) = 1$. For two Hecke eigenforms $f, f' \in \mathcal{B}_m$, write $f \sim f'$ if and only if their Hecke eigenvalues agree for $(n, qD_B) = 1$. Denote by $(f^\sharp \otimes \overline{f'^\sharp})_M$ the theta lift of $f^\sharp \otimes \overline{f'^\sharp}$. Then,*

$$\vartheta^{(m)}(l, r; s) = \frac{1}{\operatorname{covol}(\Gamma)} \frac{8\pi}{m-1} \sum_{\substack{f, f' \in \mathcal{B}_m \\ f \sim f'}} (f^\sharp \otimes \overline{f'^\sharp})_M(s) \cdot \overline{f^\sharp(l)} f^\sharp(r)$$

for all $s \in \Lambda \backslash \mathbf{SL}_2(\mathbb{R})$, $l, r \in \Gamma \backslash \mathbf{G}(\mathbb{R})$. Furthermore, $(f^\sharp \otimes \overline{f'^\sharp})_M$ is an automorphic lift of some cusp form $(f \otimes \overline{f'})_M \in S_m(\Lambda)$.

Suppose further that $f \in S_m^{\mathrm{new}}(\Gamma)$ is a newform. Let $\lambda_f(\alpha)$ be the T_α-eigenvalue of $f^\sharp$, then $f_M^\sharp = (f^\sharp \otimes \overline{f^\sharp})_M$ is the automorphic lift of a cusp form $f_M \in S_m(\Lambda)$ with Fourier expansion

$$f_M(\zeta) = \sum_{n > 0} n^{m/2 - 1} \lambda_f(n) \exp(2\pi i n \zeta).$$

Remark 7.3. The operator T_1 acts as the identity on $K_{\mathcal{R}}$-invariant functions on $[\mathbf{G}(\mathbb{A})]$, i.e. functions on $\Gamma \backslash \mathbf{G}(\mathbb{R})$. Hence, $\lambda_f(1) = 1$ and the theta lift f_M is an arithmetically normalized cusp form.

The case where $\Gamma \backslash \mathbf{G}(\mathbb{R}) = \mathbf{SL}_2(\mathbb{Z}) \backslash \mathbf{SL}_2(\mathbb{R})$ is already contained in [Zag77, § 2, Proposition 1], see also (14). For the general case, the proof will bootstrap from the fact that the convolution operator $\star_{\mathbf{G}(\mathbb{R})} \mu^m$ acting on $\Gamma \backslash \mathbf{G}(\mathbb{R})$ coincides with the Bergman kernel on $\Gamma \backslash \mathbb{H}$. This dates back at least to Selberg [Sel56], though geometric expressions for the Bergman kernel in terms of

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Poincaré series were already known to Petersson [Pet40, Pet41]. The particular expression for the Bergman kernel suitable for our needs may be found in [Zag76] and [Zag77, § 2, Proposition 1], [Ste16, Theorem 3], or [DS15, § 2.3]. The first three references each contain the split case and the latter the non-split case. There does, however, appear to be an error in the constant in [DS15]. Compare with the computation in [Zag76, Ste16], whose proofs also apply to co-compact lattices. The statement is as follows.

Proposition 7.4. *Set*

$$k^{(m)}(l,r) \coloneqq \sum_{\gamma\in\Gamma} \mu^m(l^{-1}\gamma r).$$

The function $k^{(m)}$ is the kernel of the convolution operator $\star\mu^m$ acting on $L^2(\Gamma\backslash\mathbf{G}(\mathbb{R}))$, where the convolution takes place in $\mathbf{G}(\mathbb{R})$. Fix an orthonormal basis $\mathscr{B}_m$ for $S_m(\Gamma)$. Then, for all $l, r \in \mathbf{G}(\mathbb{R})$,

$$k^{(m)}(l,r) = \frac{1}{\operatorname{covol}(\Gamma)}\frac{8\pi}{m-1}\sum_{f\in\mathscr{B}_m} \overline{f^\sharp(l)} f^\sharp(r).$$

In particular, the operator $\star\mu^m$ annihilates all the continuous, residual, and cuspidal spectrum, whose archimedean component is not discrete series.

Proof of Theorem 7.2. Let $\zeta = \sigma + i\tau \in \mathbb{H}$ and fix $s = \left(\begin{smallmatrix}1 & \sigma\\ 0 & 1\end{smallmatrix}\right)\left(\begin{smallmatrix}\tau^{1/2} & 0\\ 0 & \tau^{-1/2}\end{smallmatrix}\right)k_\theta \in \mathbf{SL}_2(\mathbb{R})$. The definition of the Weil action, Definition 6.1, and Lemmata 3.5 and 6.5 imply for $n > 0$, $g \in \mathbf{G}(\mathbb{R})$

$$\big(\rho(s)M_\infty^{(m)}\big)(\sqrt{n}g) = \tau^{m/2}n^{m/2-1}\exp(2\pi i n\zeta + im\theta)\mu^m(g). \tag{33}$$

We will establish that $\vartheta^{(m)}(l,r;s)$ coincides with the spectral expansion in the claim by showing equality in $L^2(\Gamma\backslash\mathbf{G}(\mathbb{R})\times\Gamma\backslash\mathbf{G}(\mathbb{R}))$. Pointwise identity then follows because both sides are continuous.

The Bergmann test function $M_\infty^{(m)}$ vanishes on the null-cone $\{x \in B_\infty\colon \operatorname{Nr} x = 0\}$, thus it follows from Corollary 3.6 that M is cuspidal. For any $\varphi, \varphi' \in L^\infty(\Gamma\backslash\mathbf{G}(\mathbb{R})) \subset L^2(\Gamma\backslash\mathbf{G}(\mathbb{R}))$, we can use Proposition 5.8, (27), and (33) to write the Fourier expansion

$$\begin{aligned}(\varphi\otimes\overline{\varphi'})_M(s) &= \int_{[\mathbf{G}(\mathbb{A})]}\int_{[\mathbf{G}(\mathbb{A})]} \Theta_M(l,r;s)\varphi(l)\overline{\varphi'(r)}\,dl\,dr\\ &= \sum_{n>0}\Big\langle (T_n^{M_f}\varphi)\star\big(\big(\rho(s)M_\infty\big)(\sqrt{n}\cdot\bullet)\big), \varphi'\Big\rangle\\ &= \sum_{n>0}\tau^{m/2}n^{m/2-1}\exp(2\pi i n\zeta + im\theta)\big\langle (T_n^{M_f}\varphi)\star\mu^m, \varphi'\big\rangle. \end{aligned}\tag{34}$$

Because $T_n^{M_f}$ is a convolution operator, the maps $\varphi \mapsto T_n^{M_f}\varphi$, $\varphi \mapsto (T_n^{M_f}\varphi)\star\mu^m$ preserve the continuous and the discrete spectra. Proposition 7.4 then implies that (34) vanishes whenever φ or φ' is a bounded function in the continuous spectrum. Using pseudo-Eisenstein series we can construct a dense set of bounded vectors in the continuous spectrum of $L^2([\mathbf{G}(\mathbb{A})])$, hence $\Theta_M(l,r;s) \in L^2_{\text{discrete}}([\mathbf{G}(\mathbb{A})]\times[\mathbf{G}(\mathbb{A})])$. Moreover, $\Theta_M(l,r;s)$ is $K_{\mathscr{R}}\times K_{\mathscr{R}}$-invariant. There is an orthonormal basis of $L^2_{\text{discrete}}([\mathbf{G}(\mathbb{A})])^{K_{\mathscr{R}}}$ consisting of bounded Hecke eigenforms for all Hecke operators T_n with $(n, qD_B) = 1$.

Furthermore, $\varphi \mapsto T_n^{M_f}\varphi$, $\varphi \mapsto (T_n^{M_f}\varphi)\star\mu^m$ preserve adelic automorphic representations. Hence, Proposition 7.4 implies that (34) vanishes unless both φ and φ' are lifts of weight m modular forms with the same Hecke eigenvalues for $(n, qD_B) = 1$. The claimed spectral expansion follows as the automorphic lifts of $\mathscr{B}_m(\Gamma)\times\mathscr{B}_m(\Gamma)$ can be completed to an orthogonal basis of

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

$L^2_{\text{discrete}}([\mathbf{G}(\mathbb{A})] \times [\mathbf{G}(\mathbb{A})])$ consisting of bounded Hecke eigenforms for all Hecke operators T_n with $(n, qD_B) = 1$. The Whittaker expansion (34) further shows that the Whittaker function of $(\varphi \otimes \overline{\varphi'})_M$ at infinity agrees with the Whittaker functions of a weight m holomorphic form, is furthermore clear, following the discussion in § 4.2, that it is an automorphic lift of a holomorphic form in $S_m(\Lambda)$.

If φ is a bounded Hecke newform with eigenvalues $\lambda(\bullet)$, we can use Corollary 5.15 to rewrite (34) as

$$\int_{[\mathbf{G}(\mathbb{A})]} \int_{[\mathbf{G}(\mathbb{A})]} \Theta_M(l, r; s)\varphi(l)\overline{\varphi'(r)}\, dl\, dr = \sum_{n>0} \lambda(n)\tau^{m/2} n^{m/2-1} \exp(2\pi i n\zeta + im\theta)\langle \varphi \star \mu^m, \varphi' \rangle. \tag{35}$$

The formula for the Fourier–Whittaker expansion of f_M follows from (35) with $\varphi = \varphi' = f^\sharp$. □

A careful local analysis, following Shimizu [Shi72] shows that if f is a newform, then $f_M(s)$ is the unique arithmetically normalized new modular form in the Jacquet–Langlands transfer of the automorphic representation generated by $\widetilde{f^\sharp}$ (to be defined momentarily). We will need only a weaker result. To discuss the Jacquet–Langlands transfer we need to lift functions from $[\mathbf{SL}_2(\mathbb{A})]$ to $[\mathbf{PGL}_2(\mathbb{A})]$. Define $\widetilde{U}_{\mathscr{R}}$ to be the image of $\left\{ g \in \left(\begin{smallmatrix} \widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \\ qD_B\widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \end{smallmatrix}\right) : \det g \in \widehat{\mathbb{Z}}^\times \right\}$ in $\mathbf{PGL}_2(\mathbb{A}_f)$. Then, $\widetilde{U}_{\mathscr{R}}$ is a compact open subgroup, and an argument identical to Proposition 5.10 proves that

$$\mathbf{SL}_2(\mathbb{Q})\backslash \mathbf{SL}_2(\mathbb{A}) / U_{\mathscr{R}} \to \mathbf{PGL}_2(\mathbb{Q})\backslash \mathbf{PGL}_2(\mathbb{A}) / \widetilde{U}_{\mathscr{R}}$$

is a measure-preserving bijection. Hence, we have a unique lift $\varphi \mapsto \widetilde{\varphi}$ from $L^2([\mathbf{SL}_2(\mathbb{A})])^{U_{\mathscr{R}}}$ to $L^2([\mathbf{PGL}_2(\mathbb{A})])^{\widetilde{U}_{\mathscr{R}}}$.

DEFINITION 7.5. Let $f \in S_m(\Gamma)$ be a Hecke eigenform for all Hecke operators T_n with $(n, qD_B) = 1$. If f is a newform denote by $f_{\text{JL}} \in S_m^{\text{new}}(\Lambda)$ the unique arithmetically normalized holomorphic newform such that $\widetilde{f^\sharp_{\text{JL}}}$ belongs to the Jacquet–Langlands transfer of the automorphic representation generated by $\widetilde{f^\sharp}$. That such a vector exists and is unique follows from [JL70, Shi72]. If f is an oldform, then it factors through a newform with respect to a lattice arising from an Eichler order $\mathscr{R}' \supsetneq \mathscr{R}$ with level $q' \mid q$. In this case, we denote by f_{JL} the lift of the Jacquet–Langlands transfer, defined as above, from $S_m^{\text{new}}(\Gamma_0(q'D_B))$ to $S_m^{\text{old}}(\Lambda = \Gamma_0(qD_B))$.

In both cases, the modular form f_{JL} is an eigenform of all the classical Hecke operators corresponding to n co-prime to qD_B, and its n-Fourier coefficient coincides with the T_n Hecke eigenvalue of $f^\sharp$ if $\gcd(n, qD_B) = 1$.

LEMMA 7.6. *Let $f \in S_m(\Gamma)$ be a Hecke eigenform for all Hecke operators T_n with $(n, qD_B) = 1$. If f is a newform, then the orthogonal projection of f_M onto $S_m^{\text{new}}(\Lambda)$ is equal to f_{JL}. If $f \sim f'$ are oldforms, then $(f \otimes \overline{f'})_M$ is an oldform as well.*

Proof. Theorem 7.2 implies that the Fourier coefficients of f_M and f_{JL} coincide for all n co-prime to qD_B, which is the level of $\Lambda\backslash\mathbb{H}$. Theorem 1 of [AL70] then implies that $f_M - f_{\text{JL}}$ is an oldform. Because f_{JL} is a newform if f is, the claim holds for newforms. If $f \sim f'$ are oldforms, then so is $f_{\text{JL}} = f'_{\text{JL}}$. Hence, in this case $(f \otimes \overline{f'})_M$ is a sum of oldforms and is an oldform by itself. □

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Corollary 7.7. *Let $l, r \in \mathbf{G}(\mathbb{R})$ and set $z_1 = l.i$, $z_2 = r.i \in \mathbb{H}$. Fix an orthonormal basis $\mathscr{B}_m^{\mathrm{new}}$ of Hecke eigenforms for $S_m^{\mathrm{new}}(\Gamma)$. Then,*

$$\begin{aligned}
&\frac{1}{\mathrm{covol}(\Lambda)}\|\vartheta^{(m)}(l,r;\bullet)\|^2_{L^2(\Lambda\backslash\mathbf{SL}_2(\mathbb{R}))}\\
&\quad\geq \left(\frac{1}{\mathrm{covol}(\Gamma)}\frac{8\pi}{m-1}\right)^2 (\Im(z_1)\Im(z_2))^m \sum_{f\in\mathscr{B}_m^{\mathrm{new}}} \frac{\|f_{\mathrm{JL}}\|_2^2}{\mathrm{covol}(\Lambda)}|f(z_1)|^2|f(z_2)|^2\\
&\quad\gg_\varepsilon \frac{1}{\mathrm{covol}(\Gamma)^2}(mqD_B)^{-\varepsilon}\frac{\Gamma(m)}{(4\pi)^m m^2}(\Im(z_1)\Im(z_2))^m \sum_{f\in\mathscr{B}_m^{\mathrm{new}}} |f(z_1)|^2|f(z_2)|^2.
\end{aligned}$$

Proof. Define $\vartheta_\sharp(\zeta) = \tau^{-m/2}\vartheta^{(m)}(l,r;s)$, where $\zeta = \sigma + i\tau \in \mathbb{H}$ and $s = \left(\begin{smallmatrix}1 & \sigma\\ 0 & 1\end{smallmatrix}\right)\left(\begin{smallmatrix}\tau^{1/2} & 0\\ 0 & \tau^{-1/2}\end{smallmatrix}\right)$. Because $\vartheta^{(m)}(l,r;\bullet)$ is K_∞-isotypic we have $\vartheta_\sharp{}^\sharp(s) = \vartheta^{(m)}(l,r;s)$. Theorem 7.2 and Lemma 7.6 above imply that the orthogonal project of $\vartheta_\sharp$ onto $S_m^{\mathrm{new}}(\Lambda)$ is equal to

$$\vartheta_\sharp^{\mathrm{new}}(\zeta) = \frac{1}{\mathrm{covol}(\Gamma)}\frac{8\pi}{m-1}\sum_{f\in\mathscr{B}_m^{\mathrm{new}}} f_{\mathrm{JL}}(\zeta)\overline{f^\sharp(l)}f^\sharp(r). \tag{36}$$

Because oldforms are orthogonal to newforms $\|\vartheta^{(m)}(l,r;\bullet)\|^2_{L^2(\Lambda\backslash\mathbf{SL}_2(\mathbb{R}))} \geq \|\vartheta_\sharp^{\mathrm{new}}\|_2^2$. The first claimed inequality follows from (36) and the orthogonality relations of Hecke eigenforms. The second inequality follows from the Hoffstein and Lockhart [HL94] bound on the L^2-norm of an arithmetically normalized holomorphic Hecke newform f of level N

$$\|f\|_2^2 \gg_\varepsilon \frac{\Gamma(m)}{(4\pi)^m}(mN)^{-\varepsilon}.$$

This bound holds when the Petersson inner product is normalized with respect to the probability measure on $\Lambda\backslash\mathbf{SL}_2(\mathbb{R})$. □

At this point, we shall note that we have also proven Theorem 2.1. Indeed, if we lift adelically $f \in S_m(\Gamma)$ to $f^\sharp$, then we find

$$\Im(z)^{m/2}\Im(\zeta)^{m/2}\mathscr{F}_f(z;\zeta) = c\sum_{n>0} T_n^{M_f} f^\sharp \star \left(\rho(s_\infty).M_\infty(\sqrt{n}\cdot\bullet)\right)(r_\infty) = c\int_{[\mathbf{G}(\mathbb{A})]} \Theta_M(l,r;s)f^\sharp(l)\,dl$$

by (34) and Propositions 5.14 and 7.4, where $c = \mathrm{covol}(\Gamma)(m-1)/(8\pi)$ and $r_\infty = \left(\begin{smallmatrix}1 & x\\ 1\end{smallmatrix}\right)\left(\begin{smallmatrix}y^{1/2} & \\ & y^{-1/2}\end{smallmatrix}\right)$ and r_p being the identity for all finite places, and similarly for s (with respect to ζ). Thus, $\mathscr{F}_f(z;\zeta)$ is the classical holomorphic modular form associated to $\int \Theta_M(l,r;s)f^\sharp(l)\,dl$, from which the theorem follows.

8. The geometric expansion

We have now established in Corollary 7.7 a lower bound on $\|\vartheta^{(m)}(l,r;\bullet)\|^2_{L^2(\Lambda\backslash\mathbf{SL}_2(\mathbb{R}))}$ in terms of a fourth moment of Hecke eigenforms of weight m. In this section, our goal is to establish an upper bound in terms of a count of quaternions by norm. In the next section, we will establish a sharp upper-bound for this count.

Definition 8.1. For $g \in \mathbf{PGL}_2(\mathbb{R})$, define

$$u(g) = \frac{\mathrm{Tr}(g^t g) - 2|\det g|}{4|\det g|}.$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Specifically, for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$

$$4u(g) = \frac{a^2+b^2+c^2+d^2}{|ad-bc|} - 2.$$

Using the fixed isomorphism $B \otimes \mathbb{R} \simeq \mathrm{Mat}_{2\times 2}(\mathbb{R})$, we extend the function u to the group $(B \otimes \mathbb{R})^\times$.

LEMMA 8.2. *For all $g \in \mathbf{PGL}_2(\mathbb{R})$ with $\det g > 0$,*

$$|\mu(g)|^2 = (1+u(g))^{-1}.$$

Proof. Fix $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $\det g > 0$. We deduce from Definition 6.1 that

$$|\mu(g)|^{-2} = \frac{(b-c)^2+(a+d)^2}{4\det g} = \frac{a^2+b^2+c^2+d^2+2\det g}{4\det g} = 1+u(g). \qquad \square$$

PROPOSITION 8.3. *We have*

$$\frac{1}{\mathrm{covol}(\Lambda)}\|\vartheta^{(m)}(l,r;\bullet)\|^2_{L^2(\Lambda\backslash\mathbf{SL}_2(\mathbb{R}))} \ll_\varepsilon (qD_B)^{1+\varepsilon}\frac{\Gamma(m-1)}{(4\pi)^m}\sum_{n>0}\frac{1}{n}\Bigg(\sum_{\substack{\xi\in\mathcal{R}\\ \mathrm{Nr}\,\xi=n}}\big(1+u(l^{-1}\xi r)\big)^{-m/2}\Bigg)^2$$
$$\cdot\begin{cases}1, & n \le (qD_B)^2 m,\\ \exp(-n/(qD_B)^2), & n > (qD_B)^2 m.\end{cases}$$

Proof. We first apply Proposition 4.9 to $\vartheta^{(m)}$ and use the fact that our choice of $M_\infty = M_\infty^{(m)}$ is K-isotypical and transforms simply under the diagonal group:

$$\frac{1}{\mathrm{covol}(\Lambda)}\int_{\Lambda\backslash\mathbf{SL}_2(\mathbb{R})}|\vartheta^{(m)}(l,r;s)|^2\,ds$$
$$\le (qD_B)^{-1}\sum_{a|qD_B}\sum_{\substack{\alpha\in\mathbb{Q}\\ \alpha>0}}\int_{\sqrt{3}/2}^\infty \tau^m\alpha^{m-2}\exp(-4\pi\alpha\tau)\Bigg(\sum_{\substack{\xi\in\widehat{\mathcal{R}}^{(a)}\\ \mathrm{Nr}\,\xi=\alpha}}\big|\mu^m(l^{-1}\xi r)\big|\Bigg)^2\frac{d\tau}{\tau^2}.$$

We bound the integral over τ, which is equivalent to the definition of the partial gamma function, in two ways. Write first

$$\int_{\sqrt{3}/2}^\infty \tau^{m-2}\alpha^{m-2}\exp(-4\pi\alpha\tau)\,d\tau = \frac{1}{\alpha}\int_{\sqrt{3}/2\cdot\alpha}^\infty x^{m-2}\exp(-4\pi x)\,dx.$$

For $\alpha \le m$, we complete the integral to deduce

$$\int_{\sqrt{3}/2\cdot\alpha}^\infty x^{m-2}\exp(-4\pi x)\,dx \le (4\pi)^{-(m-1)}\Gamma(m-1).$$

For $\alpha > m$, we argue

$$\begin{aligned}\int_{\sqrt{3}/2\cdot\alpha}^\infty x^{m-2}\exp(-4\pi x)\,dx &\le \exp(-2\pi\sqrt{3}/2\cdot\alpha)2^{m-2}\int_{\sqrt{3}/2\cdot\alpha}^\infty (x/2)^{m-2}\exp(-4\pi(x/2))\,dx\\ &\le \exp(-2\pi\sqrt{3}/2\cdot\alpha)2^{m-1}(4\pi)^{-(m-1)}\Gamma(m-1)\\ &\ll \exp\big(-\big(2\pi\sqrt{3}/2-\log 2\big)\cdot\alpha\big)(4\pi)^{-(m-1)}\Gamma(m-1)\\ &\le \exp(-\alpha)(4\pi)^{-(m-1)}\Gamma(m-1).\end{aligned}$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

We thus arrive at

$$\int_{\sqrt{3}/2}^{\infty} \tau^m \alpha^{m-2} \exp(-4\pi\alpha\tau) \Bigg(\sum_{\substack{\xi \in \widehat{\mathscr{R}}^{(a)} \\ \operatorname{Nr}\xi = \alpha}} |\mu^m(l^{-1}\xi r)| \Bigg)^2 \frac{d\tau}{\tau^2} \ll (4\pi)^{-m}\Gamma(m-1)\frac{1}{\alpha} \Bigg(\sum_{\substack{\xi \in \widehat{\mathscr{R}}^{(a)} \\ \operatorname{Nr}\xi = \alpha}} |\mu^m(l^{-1}\xi r)| \Bigg)^2 \cdot \begin{cases} 1, & \alpha \le m, \\ \exp(-\alpha), & \alpha > m. \end{cases}$$

Note that $\widehat{\mathscr{R}}^{(a)} \subset (qD_B)^{-1}\mathscr{R}$ and

$$\Bigg(\sum_{\substack{\xi \in \widehat{\mathscr{R}}^{(a)} \\ \operatorname{Nr}\xi = \alpha}} |\mu^m(l^{-1}\xi r)| \Bigg)^2 \le \Bigg(\sum_{\substack{\xi \in \mathscr{R} \\ \operatorname{Nr}\xi = (qD_B)^2\alpha}} |\mu^m(l^{-1}\xi r)| \Bigg)^2.$$

The claimed bound follows from combining these inequalities with the divisor bound and Lemma 8.2 above. □

Definition 8.4. For any $g \in \mathbf{G}(\mathbb{R})$, $n \in \mathbb{N}$, and $\delta > 0$, set

$$M(g, n; \delta) \coloneqq \{\xi \in \mathscr{R} \mid \operatorname{Nr}\xi = n, u(g^{-1}\xi g) < \delta\}.$$

Corollary 8.5. *If $m > 2$, then*

$$\begin{aligned} &\frac{1}{\operatorname{covol}(\Lambda)} \|\vartheta^{(m)}(g, g; \bullet)\|^2_{L^2(\Lambda\backslash \mathbf{SL}_2(\mathbb{R}))} \\ &\ll_\varepsilon (qD_B)^{1+\varepsilon} \frac{\Gamma(m-1)}{(4\pi)^m} \Bigg\{ \frac{m}{2} \int_0^\infty \Bigg[\sum_{n=1}^{(qD_B)^2 m} \frac{1}{n} M(g, n; \delta)^2 \Bigg]^{1/2} \\ &\quad + \Bigg[\sum_{n > (qD_B)^2 m}^{\infty} \frac{\exp(-n/(qD_B)^2)}{n} M(g, n; \delta)^2 \Bigg]^{1/2} \frac{d\delta}{(1+\delta)^{m/2+1}} \Bigg\}^2. \end{aligned}$$

Proof. Integration by parts for Riemann–Stieltjes integrals implies

$$\begin{aligned} \sum_{\substack{\xi \in \mathscr{R} \\ \operatorname{Nr}\xi = n}} \big(1 + u(g^{-1}\xi g)\big)^{-m/2} &= \int_0^\infty (1+\delta)^{-m/2}\, dM(g, n; \delta) \\ &= \lim_{\delta \to \infty} M(g, n; \delta)(1+\delta)^{-m/2} + \frac{m}{2} \int_0^\infty (1+\delta)^{-m/2-1} M(g, n; \delta)\, d\delta. \end{aligned} \tag{37}$$

The left-hand side is exactly the expression we need to bound in Proposition 8.3. Iwaniec and Sarnak in [IS95, Lemma 1.3 & Appendix 1] establish the bound

$$M(g, n; \delta) \ll_{\varepsilon, q, D_B, g} n^\varepsilon + (\delta + \delta^{1/4}) n^{1+\varepsilon}. \tag{38}$$

Thus, the first term in (37) vanishes for $m > 2$. Set $w_n = 1/n$ if $n \le (qD_B)^2 m$ and $w_n = \exp(-n/(qD_B)^2)/n$ otherwise. We apply Minkowski's integral inequality to deduce

$$\sum_{n=1}^\infty w_n \Bigg| \sum_{\substack{\xi \in \mathscr{R} \\ \operatorname{Nr}\xi = n}} (1 + u(g^{-1}\xi g))^{-m/2} \Bigg|^2 \le \Bigg(\frac{m}{2} \int_0^\infty \sqrt{\sum_{n=1}^\infty w_n M(g, n; \delta)^2}\, \frac{d\delta}{(1+\delta)^{m/2+1}} \Bigg)^2.$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

The claim follows by splitting the sum into two ranges: $1 \le n \le (qD_B)^2 m$ and $n > (qD_B)^2 m$ and using the l^2–l^1 inequality. □

9. Second-moment count of quaternions by norm

In this section, we prove our main results about the second-moment count of quaternions by norm in a small ball. This bound in combination with the results of the previous sections will lead to the proof of Theorem 1.1. To bound $\sum_{n=1}^{N} M(g,n;\delta)^2$ we can assume henceforth without loss of generality that $\mathcal{R}$ is a maximal order, otherwise we can replace the Eichler order $\mathcal{R} = \mathcal{R}_1 \cap \mathcal{R}_2$ by $\mathcal{R}_1$ and the second-moment sum will only increase.

We shall deal separately with the split case $\mathbf{G} = \mathbf{SL}_2$ and the case of anisotropic $\mathbf{G}$. The proof in both cases is very similar except that we need to track the dependence on g differently. While in the split case we shall work with the Iwasawa decomposition of g, in the anisotropic case we will use an adapted Cartan decomposition of g.

9.1 Second-moment bound for the split matrix algebra

In this section, we fix $\mathbf{G} = \mathbf{SL}_2$, i.e. $B = \mathrm{Mat}_{2\times 2}(\mathbb{Q})$ and $\mathcal{R} = \mathrm{Mat}_{2\times 2}(\mathbb{Z})$. If we write in coordinates $\xi = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then the inequalities $u(\xi) < \delta$, $0 < \det \xi < N$ imply

$$a^2 + b^2 + c^2 + d^2 < N(4\delta + 2), \tag{39}$$

$$(a-d)^2 + (b+c)^2 < 4N\delta. \tag{40}$$

For $g \in \mathbf{G}(\mathbb{R})$, write $g = nak$ with $k \in \mathbf{SO}_2(\mathbb{R})$ and

$$n = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix},$$

$$a = \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix}.$$

This is the standard Iwasawa decomposition of g.

PROPOSITION 9.1. *Let $g \in \mathbf{SL}_2(\mathbb{R})$ and write $g.i = x + iy$. Assume $|x| \le C$ and $y \ge A > 0$. Then,*

$$\sum_{n=1}^{N} M(g,n;\delta)^2 \ll_{A,C,\varepsilon} N^{3+\varepsilon}\delta^2 + N^{1+\varepsilon} + N^{1/2+\varepsilon} \min(N^{1/2}, (N\delta)^{1/2} + 1)(y^2 N\delta + 1).$$

Remark 9.2. In the end, we may restrict ourselves to g in a fundamental domain for $\mathbf{SL}_2(\mathbb{Z})$ and, hence, the restrictions on x, y will be satisfied.

Using the inequality of geometric and arithmetic means we can split the second-moment count into two cases. The first is when both matrices are upper triangular and the second is when neither is. We now prepare some preliminary results needed in the proof of Proposition 9.1.

LEMMA 9.3. *Denote by $\mathfrak{U} \subset B_\infty$ the subset of upper triangular matrices. Then,*

$$\#\Big\{(\xi_1,\xi_2) \in \big(\mathrm{Mat}_{2\times 2}(\mathbb{Z}) \cap \mathfrak{U}\big)^2 \colon u(g^{-1}\xi_1 g), u(g^{-1}\xi_2 g) < \delta,\ 0 < \det \xi_1 = \det \xi_2 < N\Big\}$$
$$\ll_\varepsilon N^{1/2+\varepsilon} \min(N^{1/2}, (N\delta)^{1/2} + 1)(y^2 N\delta + 1).$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Proof. Write $\xi_i = \left(\begin{smallmatrix} a_i & b_i \\ 0 & d_i \end{smallmatrix}\right)$. For upper triangular matrices we rewrite (40) for $g^{-1}\xi_i g$ as

$$(a_i - d_i)^2 + \frac{\big(b_i + x(a_i - d_i)\big)^2}{y^2} < 4N\delta.$$

Hence, we have at most $\ll (N\delta)^{1/2} + 1$ choices for $|a_1 - d_1|$. In addition, the condition $0 < 4 \det \xi_1 = (a_1 + d_1)^2 - (a_1 - d_1)^2 < 4N$ implies $0 < |a_1 + d_1| - |a_1 - d_1| \ll N^{1/2}$. We deduce that there are at most $\ll ((N\delta)^{1/2} + 1)N^{1/2}$ possibilities for $(|a_1 + d_1|, |a_1 - d_1|)$ and a similar statement holds for (a_1, d_1).

On the other hand, $0 < a_1 d_1 < N$ and the divisor bound implies that the number of possible pairs (a_1, d_1) is also bounded by $\ll_\varepsilon N^{1+\varepsilon}$. The number of possibilities for b_1 is now bounded above by $y(N\delta)^{1/2} + 1$. Thus, there are at most $\ll_\varepsilon N^{1/2+\varepsilon} \min\big(N^{1/2}, (N\delta)^{1/2} + 1\big)(y(N\delta)^{1/2} + 1)$ possibilities for ξ_1.

Once ξ_1 is fixed the condition $\det \xi_1 = \det \xi_2 > 0$ fixes $a_2 d_2$ and the divisor bound restricts the number of possible pairs (a_2, d_2) to $\ll_\varepsilon N^\varepsilon$. Finally, the number of possible b_2 after fixing (a_2, d_2) is at most $\ll y(N\delta)^{1/2} + 1$. □

We continue to analyze the case when neither matrix is upper triangular. We will use the direct sum decomposition $B_\infty = \mathbb{R}\,\mathrm{Id} + B_\infty^0$. This decomposition is preserved by the conjugation action. We denote by $\xi^0 = \xi - \frac{1}{2}\operatorname{Tr}(\xi)$ the traceless part of $\xi \in B_\infty$. In coordinates we write

$$\xi^0 = \begin{pmatrix} e & b \\ c & -e \end{pmatrix},$$

where $e = (a - d)/2$. If ξ satisfies (39) and (40), then ξ^0 satisfies $2e^2 + b^2 + c^2 < N(4\delta + 2)$. This leads us to define

$$B_\infty^0 \supset X := \left\{ \xi^0 = \begin{pmatrix} e & b \\ c & -e \end{pmatrix} : 4e^2 + (b + c)^2 < 4N\delta,\ 2e^2 + b^2 + c^2 < N(4\delta + 2) \right\},$$
$$B_\infty^0 \supset X^g := gXg^{-1}.$$

The set X is invariant under conjugation by K and using the Iwasawa decomposition we can write the conditions for $\left(\begin{smallmatrix} e & b \\ c & -e \end{smallmatrix}\right) \in X^g$ explicitly as

$$2(e - xc)^2 + \frac{(b + 2xe - x^2 c)^2}{y^2} + y^2 c^2 < N(4\delta + 2), \tag{41}$$

$$4(e - xc)^2 + \frac{\big(b + 2x(e - xc) + (x^2 + y^2)c\big)^2}{y^2} < 4N\delta. \tag{42}$$

LEMMA 9.4. *Assume $y \geq A > 0$ and $0 < \delta \leq 1$, then*

$$\#\big(\tfrac{1}{2}\operatorname{Mat}_{2\times 2}(\mathbb{Z})^0 \cap X^g \setminus \mathfrak{U}\big) \ll N^{3/2}\delta + N^{1/2}.$$

Proof. From (41), we learn that there are $\ll y^{-1}N^{1/2}(1 + \delta)$ options for c ($c \neq 0$ because the matrices are not upper triangular). For any fixed c, (42) describes an ellipse in the e, b plane with radii $\ll \sqrt{N\delta}, y\sqrt{N\delta}$. Hence, the number of possibilities for $(e, b) \in \frac{1}{2}\mathbb{Z} \times \frac{1}{2}\mathbb{Z}$ is bounded from above by

$$\ll yN\delta + (N\delta)^{1/2}\max(y, 1) + 1 \ll_A y\big(N\delta + (N\delta)^{1/2} + 1\big).$$

Multiplying this by the bound for possible values of c and the inequality $N\delta^{1/2} < \max(N^{3/2}\delta, N^{1/2})$ establish the claim. □

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

LEMMA 9.5. *Assume $y \geq A > 0$, $|x| < C$, and $0 < \delta \leq 1$. Then,*

$$\#\{(\xi_1^0, \xi_2^0) \in (\tfrac{1}{2}\operatorname{Mat}_{2\times 2}(\mathbb{Z})^0 \cap X^g \setminus \mathfrak{U})^2 \colon \det \xi_1^0 = \det \xi_2^0\} \ll_{A,C} N^{5/2}\delta^2 + N^{1/2}.$$

Proof. Note that the number of possible elements ξ_1^0 is bounded by Lemma 9.4 above. We fix henceforth ξ_1^0 as in the claim and count the number of possible $\xi_2^0 \notin \mathfrak{U}$ with $\det \xi_2^0 = \det \xi_1^0$. Denote

$$g^{-1}\xi_i^0 g = \begin{pmatrix} \tilde{e}_i & \tilde{b}_i \\ \tilde{c}_i & -\tilde{e}_i \end{pmatrix}.$$

We now rewrite (40) for ξ_i^0 as

$$4(e_i - c_i x)^2 + \big((\tilde{b}_i - \tilde{c}_i) + 2yc_i\big)^2 < 4N\delta. \tag{43}$$

Then, $(\tilde{b}_1 - \tilde{c}_1)$ is restricted to an interval of length $\ll \sqrt{N\delta}$. Equation (40) implies

$$(\tilde{b}_i + \tilde{c}_i)^2 + 4\tilde{e_i}^2 = (\tilde{b}_i - \tilde{c}_i)^2 - 4\det(\xi_i^0) = (\tilde{b}_i - \tilde{c}_i)^2 - 4\det(g^{-1}\xi_i^0 g) \in [0, 4N\delta]$$

and $\big|(\tilde{b_1} - \tilde{c}_1)^2 - (\tilde{b_2} - \tilde{c}_2)^2\big| \ll N\delta$. We deduce that $\big||\tilde{b_1} - \tilde{c}_1| - |\tilde{b_2} - \tilde{c}_2|\big| \ll \sqrt{N\delta}$. In particular, $(\tilde{b}_2 - \tilde{c}_2)$ is restricted to two intervals of length $\ll \sqrt{N\delta}$.

Consider (43) for ξ_2^0 with $(\tilde{b}_2 - \tilde{c}_2)$ as a varying quantity in the aforementioned intervals, thus ignoring their dependencies on e_2, c_2. It describes an ellipse in the variables e_2, c_2 with center $-(\tilde{b}_2 - \tilde{c}_2)/2y \cdot (x, 1)$. Because ξ_1^0 is fixed, the center of the ellipse is restricted to one of two intervals of length $\ll_C y^{-1}\sqrt{N\delta}$. The radii of the ellipse satisfy $\ll_C \sqrt{N\delta}, y^{-1}\sqrt{N\delta}$. We deduce that the number of possibilities for (e_2, c_2) is $\ll_{A,C} y^{-1}N\delta + (N\delta)^{1/2} + 1 \ll_{A,C} N\delta + 1$. Once ξ_1^0, e_2 and $c_2 \neq 0$ are fixed, the value of b_2 is fixed by the equality $\det \xi_2^0 = \det \xi_1^0$. Hence, the total number of pairs (ξ_1^0, ξ_2^0) is bounded from above in this case by

$$\ll_{A,C} (N^{3/2}\delta + N^{1/2})(N\delta + 1) \ll_{A,C} N^{5/2}\delta^2 + N^{3/2}\delta + N^{1/2} \ll_{A,C} N^{5/2}\delta^2 + N^{1/2}. \qquad \square$$

Proof of Proposition 9.1. Define

$$M_\star(g, n; \delta) \coloneqq \{\xi \in \operatorname{Mat}_{2\times 2}(\mathbb{Z}) \setminus \mathfrak{U} \mid \det \xi = n,\ u(g^{-1}\xi g) < \delta\},$$
$$M_u(g, n; \delta) \coloneqq \{\xi \in \operatorname{Mat}_{2\times 2}(\mathbb{Z}) \cap \mathfrak{U} \mid \det \xi = n, u(g^{-1}\xi g) < \delta\}.$$

Then, the inequality of means imply

$$\sum_{n=1}^{N} M(g, n; \delta)^2 \leq 2\sum_{n=1}^{N} M_\star(g, n; \delta)^2 + 2\sum_{n=1}^{N} M_u(g, n; \delta)^2$$

and we turn to bounding each term individually. The second term is controlled by Lemma 9.3 and is consistent with the claim.

To bound the first term we need to bound the number of pairs $(\xi_1, \xi_2) \in (\operatorname{Mat}_{2\times 2}(\mathbb{Z}) \setminus \mathfrak{U})^2$ such that $0 < \det \xi_1 = \det \xi_2 < N$ and $u(g^{-1}\xi_i g) < \delta$ for $i = 1, 2$. Assume first $\delta > 1$, we then argue as in [IS95] to show the stronger bound $M_\star(g, n; \delta) \ll_{A,C,\varepsilon} n^{1+\varepsilon}\delta$. Let $\det \xi = n$ and write $\xi = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ as usual. When $\delta > 1$, we can replace the right-hand side in inequalities (39) and (41) by $6N\delta$. If either $a = 0$ or $d = 0$, then the equation $bc = n$ and the divisor bound imply that we have at most $\ll_\varepsilon n^\varepsilon$ possibilities for (b, c). Moreover, (39) implies that there are at most $\ll (n\delta)^{1/2}$ options for $\operatorname{Tr}(\xi)$. Hence, the number of possible values of ξ in these cases is $\ll_\varepsilon n^{1/2+\varepsilon}\delta^{1/2} \ll n^{1+\varepsilon}\delta$. Assume next $a \neq 0$ and $d \neq 0$. Equation (41) implies that we have at most $\ll y^{-1}(n\delta)^{1/2}$ options for c. Likewise, we have $\ll 1 + (n\delta)^{1/2}(x^2 + y^2)y^{-1} \ll_{A,C} y(n\delta)^{1/2}$ choices for b. This may be seen by either repeating the computation for (41) using the Iwasawa decomposition with respect to the lower triangular unipotents or noting that $\left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right) g.i = (-x + iy)/(x^2 + y^2)$ and

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

$\left(\begin{smallmatrix}-1 & 1\end{smallmatrix}\right)\xi\left(\begin{smallmatrix}-1 & 1\end{smallmatrix}\right)^{-1} = \left(\begin{smallmatrix}d & -c\\ -b & a\end{smallmatrix}\right)$. Now that (b,c) is fixed, we use the equality $ad = bc+n$ and the divisor bound to see that there are at most $\ll_\varepsilon n^\varepsilon$ possibilities for (a,d). This establishes the inequality for $\delta > 1$.

Assume henceforth $\delta < 1$. We will be using the simple identity

$$\det \xi = \frac{\operatorname{Tr}(\xi)^2}{4} + \det \xi^0 \tag{44}$$

and argue in two different ways depending on whether the traces of ξ_1, ξ_2 are equal or not.

Case I: $|\operatorname{Tr}(\xi_1)| \neq |\operatorname{Tr}(\xi_2)|$. Lemma 9.4 implies that there are at most $\ll_A N^3\delta^2 + N$ options for (ξ_1^0, ξ_2^0). After fixing the traceless parts, (44) fixed $\operatorname{Tr}(\xi_1)^2 - \operatorname{Tr}(\xi_2)^2$. Because the traces are not equal in absolute value the divisor bound and the trivial bound $|\operatorname{Tr}(\xi)| \ll N$ imply there are at most $\ll_\varepsilon N^\varepsilon$ choices for the traces. This establishes the claim in this case.

Case II: $|\operatorname{Tr}(\xi_1)| = |\operatorname{Tr}(\xi_2)|$. In this case, we use the trivial bound $|\operatorname{Tr}(\xi_1)| = |\operatorname{Tr}(\xi_2)| \ll N^{1/2}$ from (39) to fix the traces and Lemma 9.5 to fix the traceless part. The final bound is consistent with the claim. □

9.2 Second-moment bound for division algebras

In the section, we assume $\mathbf{G}$ is anisotropic, i.e. B is a ramified quaternion algebra over $\mathbb{Q}$, and that B is split over $\mathbb{R}$. Fix an imaginary quadratic field $E/\mathbb{Q}$ of discriminant D_E, such that every prime dividing D_B is inert in E. Let us denote by $\mathcal{O}_E$ the ring of integers of E. By a theorem of Chinburg and Friedman [CF99], there is an optimal embedding $\mathcal{O}_E \hookrightarrow \mathscr{R}$. We henceforth identify $\mathcal{O}_E$ with its image in $\mathscr{R}$. Denote by $K_E < \mathbf{G}(\mathbb{R})$ the group of norm 1 elements in $(E\otimes\mathbb{R})^\times$. Recall that we have a fixed isomorphism $B_\infty \simeq \mathrm{Mat}_{2\times 2}(\mathbb{R})$ that induces a group isomorphism $\mathbf{G}(\mathbb{R}) \simeq \mathbf{SL}_2(\mathbb{R})$, which we use to identify the two groups. Moreover, K_∞ was defined as the image of $\mathbf{SO}_2(\mathbb{R})$ in $\mathbf{G}(\mathbb{R})$, and we define similarly A to be the image of the diagonal subgroup in $\mathbf{G}(\mathbb{R})$. The group K_E is conjugate to K_∞, and we can write $K_\infty = hK_Eh^{-1}$.

Proposition 9.6. *Assume $\mathcal{O}_E \hookrightarrow \mathscr{R}$ is an optimal embedding of the ring of integers $\mathcal{O}_E$ of an imaginary quadratic field E in the maximal order $\mathscr{R}$. Let $h \in \mathbf{G}(\mathbb{R})$ be an element conjugating K_E to $K_\infty < \mathbf{SL}_2(\mathbb{R})$. Then, for any $g \in \mathbf{G}(\mathbb{R})$, $1 > \delta > 0$,*

$$\sum_{n=1}^{N} M(g,n;\delta)^2 \ll |D_E|^{2+\varepsilon} N^\varepsilon \left[N^3\delta^2 + (\lambda+\lambda^{-1})^{2+\varepsilon}(N^{5/2}\delta^{3/2}+N)\right],$$

where we write $\sqrt{\lambda} \geq (\sqrt{\lambda})^{-1} > 0$ for the eigenvalue of the diagonal part in the $K_\infty A K_\infty$ Cartan decomposition of hg. Moreover, if $\delta \geq 1$ the bound

$$M(g,n;\delta) \ll_\varepsilon \left((\lambda+\lambda^{-1})n\delta|D_E|^{1/2}\right)^{1+\varepsilon}$$

holds for all $g \in \mathbf{G}(\mathbb{R})$ and $n \in \mathbb{N}$.

We now fix $\mathscr{R}, E, h, g$ as in the proposition above and prepare some notation and lemmata that we will use in the course of the proof. The proof is very similar to the split case, except that we track the dependence on g differently, not using its Iwasawa decomposition but rather its Cartan decomposition relative to the stabilizer of $E \hookrightarrow B$.

Because of our choice of E as optimally embedded in $\mathscr{R}$, we can find an isomorphism $B \otimes E \simeq M_2(E)$ where $\mathscr{R}$ is mapped to

$$\left\{\begin{pmatrix} a & D_B b \\ \sigma b & \sigma a\end{pmatrix} : a,b \in \widehat{\mathcal{O}_E}, a+b \in \mathcal{O}_E\right\}$$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

and after fixing a field embedding $E \hookrightarrow \mathbb{C}$ the algebra B_∞ coincides with $\left\{ \left(\begin{smallmatrix} a & D_B b \\ \sigma b & \sigma a \end{smallmatrix} \right) : a, b \in \mathbb{C} \right\}$. We denote by B_∞^0 the subspace of traceless elements, equivalently pure quaternions. There is a direct sum decomposition $B_\infty = \mathbb{R}\,\mathrm{Id} + B_\infty^0$. This decomposition is preserved by the conjugation action. In our new coordinate system, the space B_∞^0 is identified with $i\mathbb{R} \times \mathbb{C}$ and the projection map becomes $(a, b) \mapsto (a^0, b)$ where $a^0 = (a - {}^\sigma a)/2 = i\Im a$ is the traceless part of $a \in \mathbb{C}$. The space B_∞^0 is equipped with an inner product constructed as the direct sum of the standard inner product on $\mathbb{R}$ and $\mathbb{C}$, i.e. $|(a^0, b)|^2 = |a^0|^2 + |b|^2$. Let $\mathscr{R}^0$ be the projection of $\mathscr{R}$ to B_∞^0. Then, $\mathscr{R}^0 < B_\infty^0$ is a lattice of co-volume $\asymp 1$.

In this new coordinate system, we have for a quaternion of positive norm[13]

$$u\left(h \begin{pmatrix} a & D_B b \\ \sigma b & \sigma a \end{pmatrix} h^{-1}\right) = \frac{1}{2}\left[\frac{\mathrm{Nr}\, a + D_B \,\mathrm{Nr}\, b}{\mathrm{Nr}\, a - D_B \,\mathrm{Nr}\, b} - 1\right] = \frac{D_B \,\mathrm{Nr}\, b}{\mathrm{Nr}\, a - D_B \,\mathrm{Nr}\, b}.$$

Hence, if we write in coordinates

$$(gh)^{-1}\xi(gh) = \begin{pmatrix} a & D_B b \\ \sigma b & \sigma a \end{pmatrix}$$

with $\mathrm{Nr}\,\xi > 0$, then the conditions $u(g^{-1}\xi g) < \delta$, $\mathrm{Nr}\,\xi < N$ imply

$$\mathrm{Nr}\, a + D_B \,\mathrm{Nr}\, b < N(2\delta + 1), \tag{45}$$

$$D_B \,\mathrm{Nr}\, b < \delta N. \tag{46}$$

The traceless part of $(gh)^{-1}\xi(gh)$ is

$$(gh)^{-1}\xi^0(gh) = \begin{pmatrix} a^0 & D_B b \\ \sigma b & \sigma a^0 \end{pmatrix},$$

where $a^0 = i\Im a$ is the traceless part of a. Equation (45) implies that $\mathrm{Nr}\, a^0 \le N(2\delta + 1)$. Motivated by these inequalities, we denote

$$B_\infty^0 \supset X :- \left\{ x = \begin{pmatrix} a^0 & D_B b \\ \sigma b & \sigma a^0 \end{pmatrix} : D_B \,\mathrm{Nr}\, b \le \delta N, \mathrm{Nr}\, a^0 \le N(2\delta + 1) \right\},$$

$$B_\infty^0 \supset X^g :- (gh)X(gh)^{-1}.$$

We decompose gh according to a Cartan decomposition in $K_E A_E K_E$ where A_E is the orthogonal group preserving the quadratic form $(\Im a)^2 - D_B(\Im b)^2$. Equivalently, the Lie algebra of A_E is

$$\mathrm{Lie}\, A_E = \mathbb{R} \cdot \begin{pmatrix} 0 & D_B \\ 1 & 0 \end{pmatrix} < B_\infty^0.$$

Write $gh = k_2 a_E k_1$ with $k_1, k_2 \in K_E$, $a_E \in A_E$, and denote by $\sqrt{\lambda} \ge (\sqrt{\lambda})^{-1} > 0$ the eigenvalues of a_E. Then, $\sqrt{\lambda}, (\sqrt{\lambda})^{-1}$ are also the eigenvalues of the diagonal part of the regular $K_\infty A K_\infty$ Cartan decomposition of $hg = h(gh)h^{-1}$, i.e. the singular values.

The set X is invariant under conjugation by K_E, hence $X^g = (k_2 a_E)X(k_2 a_E)^{-1}$. We can write the equations defining the set $a_E X a_E^{-1}$ explicitly by decomposing the Lie algebra B_∞^0 into the weight spaces of A_E. The result of the computation is that every $x = \left(\begin{smallmatrix} a^0 & D_B b \\ \sigma b & \sigma a^0 \end{smallmatrix} \right) \in a_E X a_E^{-1}$ satisfies

$$\left(\frac{\lambda + \lambda^{-1}}{2}\Im a^0 + \sqrt{D_B}\frac{\lambda - \lambda^{-1}}{2}\Im b\right)^2 \le N(2\delta + 1), \tag{47}$$

[13] Note that the transpose operation with respect to K_E is $\left(\begin{smallmatrix} \alpha & D_B \beta \\ \sigma\beta & \sigma\alpha \end{smallmatrix} \right) \mapsto \left(\begin{smallmatrix} \sigma\alpha & D_B \beta \\ \sigma\beta & \alpha \end{smallmatrix} \right)$.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

$$\left(\frac{\lambda-\lambda^{-1}}{2}\Im a^0+\sqrt{D_B}\frac{\lambda+\lambda^{-1}}{2}\Im b\right)^2+D_B(\Re b)^2\le N\delta. \tag{48}$$

The set $X^g=k_2(a_EXa_E^{-1})k_2^{-1}$ is a rotation of $a_EXa_E^{-1}$ around the $\Im a^0$ axis. Hence, the equations defining X^g are derived from (47) and (48) by a rotation in the b-plane. Note that (47) and (48) imply that $|\Im a^0|\ll(\lambda+\lambda^{-1})N^{1/2}(1+\delta^{1/2})$. Because the axis $\Im a^0$ is invariant under conjugation by k_2 this inequality holds also for X^g.

LEMMA 9.7. *Assume* $0<\delta\le 1$. *Then,*

$$\#\big(\mathcal{R}^0\cap X^g\big)\ll|D_E|\big(N^{3/2}\delta+(\lambda+\lambda^{-1})(N\delta^{1/2}+N^{1/2})\big).$$

Proof. From $|\Im a^0|\ll(\lambda+\lambda^{-1})N^{1/2}$ we deduce that there are $\ll(\lambda+\lambda^{-1})(N|D_E|)^{1/2}$ possibilities for $a^0=i\Im a^0$. The second equation (48) implies that for any fixed $a^0=i\Im a^0$ the element b belongs to an ellipse with radii $\sqrt{N\delta/D_B}$, $(2/(\lambda+\lambda^{-1}))\sqrt{N\delta/D_B}$. Conjugation by k_2 amounts to rotating the set around the $\Im a^0$ axis. Hence, this observation remains valid for X^g. We deduce that for any fixed $\Im a^0$ we have

$$\ll\frac{N\delta\sqrt{|D_E|}}{D_B(\lambda+\lambda^{-1})}+\left(\frac{N\delta|D_E|}{D_B}\right)^{1/2}+1$$

possibilities for b. The claim follows by multiplying the number of possibilities for a^0 by the number of possible b for each a^0. □

LEMMA 9.8. *Assume* $0<\delta\le 1$. *Then,*

$$\#\big\{(\xi_1^0,\xi_2^0)\in\big(\mathcal{R}^0\cap X^g\big)^2\colon\ \operatorname{Nr}\xi_1^0=\operatorname{Nr}\xi_2^0\big\}\ll_\varepsilon|D_E|^{1+\varepsilon}(\lambda+\lambda^{-1})^{2+\varepsilon}N^\varepsilon\big(N^2\delta^{3/2}+N^{1/2}\big).$$

Proof. Write

$$\xi_{1,2}^0=\begin{pmatrix}a_{1,2}^0 & D_Bb_{1,2}\\ \sigma b_{1,2} & {}^\sigma a_{1,2}{}^0\end{pmatrix}$$

and assume $\xi_1^0,\xi_2^0\in X^g$ and $\operatorname{Nr}\xi_1^0=\operatorname{Nr}\xi_2^0$. Our goal is to count the number of possible pairs (ξ_1^0,ξ_2^0).

For every $\xi\in B_\infty^0$ let $\tilde a^0$ be the a^0 coordinate of $a_E^{-1}\xi a_E$. Then,

$$\Im\tilde a^0=\frac{\lambda+\lambda^{-1}}{2}\Im a^0+\sqrt{D_B}\frac{\lambda-\lambda^{-1}}{2}\Im b.$$

Moreover, $\tilde a^0$ is also the a^0 coordinate of $(gh)^{-1}\xi(gh)$ because conjugation by K_E acts trivially on the a^0-axis. By substitution, we can rewrite (48) as

$$\left(\frac{\lambda-\lambda^{-1}}{\lambda+\lambda^{-1}}\Im\tilde a^0+\sqrt{D_B}\frac{2}{\lambda+\lambda^{-1}}\Im b\right)^2+D_B(\Re b)^2\le N\delta. \tag{49}$$

Assume $\xi\in X^g$. Because (45) implies that $|\Im\tilde a^0|\ll N^{1/2}$, we see that (49) restricts $b\in X^g$ to an ellipse with radii $(N\delta/D_B)^{1/2},(\lambda+\lambda^{-1})(N\delta/D_B)^{1/2}$ and center in an interval of length $\ll N^{1/2}$. We deduce that there are at most

$$\ll(\lambda+\lambda^{-1})N\delta^{1/2}\sqrt{\frac{|D_E|}{D_B}}+(\lambda+\lambda^{-1})N^{1/2}\sqrt{\frac{|D_E|}{D_B}}+1 \tag{50}$$

choices for b if $\xi\in X^g$. Moreover, we see that necessarily $\operatorname{Nr}b\ll(\lambda+\lambda^{-1})N$

Case I: $|a_1^0|=|a_2^0|$. In this case, the condition $\operatorname{Nr}\xi_1^0=\operatorname{Nr}\xi_2^0$ implies that $\operatorname{Nr}b_1=\operatorname{Nr}b_2$. Because there are at most $\ll_\varepsilon(n|D_E|)^\varepsilon$ elements of norm n in $\widehat{\mathcal{O}}_E$ and $\operatorname{Nr}b_2\ll(\lambda+\lambda^{-1})N$ we see that

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

for any fixed ξ_1^0 there are at most $\ll_\varepsilon ((\lambda+\lambda^{-1})N|D_E|)^\varepsilon$ possibilities for ξ_2^0. We deduce from Lemma 9.7 that the number of possible pairs (ξ_1,ξ_2) with $|a_1^0|=|a_2^0|$ satisfies

$$\ll_\varepsilon ((\lambda+\lambda^{-1})N|D_E|)^\varepsilon |D_E|\big(N^{3/2}\delta+(\lambda+\lambda^{-1})(N\delta^{1/2}+N^{1/2})\big)$$

and this bound is compatible with the claim.

Case II: $|a_1^0|\neq|a_2^0|$. In this case, we will first count the number of possibilities for (b_1,b_2). We bound the number of choices for b_1 using (50) above. If $\xi\in X^g$, then (46) implies

$$(\Im\tilde{a}^0)^2-\operatorname{Nr}\xi=(\Im\tilde{a}^0)^2-\operatorname{Nr}((gh)^{-1}\xi(gh))\in[0,N\delta].$$

Thus, we deduce for $\xi_{1,2}^0$ that $\big|(\Im\tilde{a}_1^0)^2-(\Im\tilde{a}_2^0)^2\big|\le 2N\delta$ and

$$\big||\Im\tilde{a}_1^0|-|\Im\tilde{a}_2^0|\big|\ll\sqrt{N\delta}. \tag{51}$$

Once b_1 is fixed, (49) restricts $((\lambda-\lambda^{-1})/(\lambda+\lambda^{-1}))\Im\tilde{a}_1^0$ to an interval of length $\ll\sqrt{N\delta}$. Equation (51) then restricts $((\lambda-\lambda^{-1})/(\lambda+\lambda^{-1}))|\Im\tilde{a}_2^0|$ to an interval also of length $\ll\sqrt{N\delta}$.

This constraints the possibilities for the center of the ellipse in (49) for b_2 into two intervals of length $\ll\sqrt{N\delta}$. Hence, given b_1, there are at most

$$\ll(\lambda+\lambda^{-1})N\delta\sqrt{\frac{|D_E|}{|D_B|}}+(\lambda+\lambda^{-1})(N\delta)^{1/2}\sqrt{\frac{|D_E|}{|D_B|}}+1$$

options for the b_2.

After fixing b_1,b_2, we use the condition $\operatorname{Nr}\xi_1^0=\operatorname{Nr}\xi_2^0$ to fix $(\Im a_1^0)^2-(\Im a_2^0)^2$. The divisor bound and the condition $|\Im a_1^0|\neq|\Im a_2^0|$ now implies there are at most $\ll_\varepsilon((\lambda+\lambda^{-1})N|D_E|)^\varepsilon$ options for the pair (a_1^0,a_2^0).

The total number of possible pairs (ξ_1^0,ξ_2^0) in this case is thus bounded by

$$\begin{aligned}&\ll_\varepsilon((\lambda+\lambda^{-1})N|D_E|)^\varepsilon|D_E|\big((\lambda+\lambda^{-1})N\delta^{1/2}+(\lambda+\lambda^{-1})N^{1/2}\big)\\&\cdot\big((\lambda+\lambda^{-1})N\delta+(\lambda+\lambda^{-1})(N\delta)^{1/2}+1\big).\end{aligned}$$

This bound is also compatible with the claim. □

Proof of Proposition 9.6. Assume first $\delta>1$. Then, we follow [IS95] to establish the bound $M(g,n;\delta)\ll_\varepsilon\big((\lambda+\lambda^{-1})n\delta|D_E|^{1/2}\big)^{1+\varepsilon}$. We have the bounds $\operatorname{Tr}(a)=2\Re a\ll(n\delta)^{1/2}$ and $\Im a=\Im a^0\ll(\lambda+\lambda^{-1})(n\delta)^{1/2}$. After fixing a we can fix b using the equality $n=\det\xi=\operatorname{Nr}a-D_B\operatorname{Nr}b$. The divisor bound and the inequality $\operatorname{Nr}b\ll(\lambda+\lambda^{-1})^2n\delta$ imply we have at most $\ll_\varepsilon(\lambda+\lambda^{-1})^\varepsilon n^\varepsilon\delta^\varepsilon|D_E|^\varepsilon$ choices for b.

Assume next $\delta\le1$. Once again, an important role is reserved for the simple formula

$$\operatorname{Nr}x=\frac{\operatorname{Tr}(x)^2}{4}+\operatorname{Nr}x^0 \tag{52}$$

that holds for all $x\in B_\infty$ with $x^0\in B_\infty^0$ the traceless part of x. Our goal is to bound the number of pairs $(\xi_1,\xi_2)\in\mathcal{R}$ such that $0\le\operatorname{Nr}\xi_1=\operatorname{Nr}\xi_2\le N$ and $u(g^{-1}\xi_1g)=u(g^{-1}\xi_2g)<\delta$.

Case I: $|\operatorname{Tr}(\xi_1)|\neq|\operatorname{Tr}(\xi_2)|$. Lemma 9.7 implies that the number of possibilities for the pair (ξ_1^0,ξ_2^0) is bounded by

$$\ll|D_E|^2\big(N^3\delta^2+(\lambda+\lambda^{-1})^2(N^{5/2}\delta^{3/2}+N)\big).$$

For any pair $(\xi_1^0,\xi_2^0)\in B_\infty^0\times B_\infty^0$, the lifts to $B_\infty\times B_\infty$ are determined by $\big(\operatorname{Tr}(\xi_1),\operatorname{Tr}(\xi_2)\big)$.

From the formula (52), we derive $\operatorname{Tr}(\xi_1)^2-\operatorname{Tr}(\xi_2)^2=4\big(\operatorname{Nr}\xi_2^0-\operatorname{Nr}\xi_1^0\big)$. The right-hand side is bounded in absolute value by $\ll N$. The divisor bound and the assumption $|\operatorname{Tr}(\xi_1)|\neq|\operatorname{Tr}(\xi_2)|$

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

imply that for every (ξ_1^0, ξ_2^0) the number of possible pairs $\big(\mathrm{Tr}(\xi_1), \mathrm{Tr}(\xi_2)\big)$ is bounded by $\ll_\varepsilon N^\varepsilon$. The cumulative bound is consistent with the claim.

Case II: $|\mathrm{Tr}(\xi_1)| = |\mathrm{Tr}(\xi_2)|$. In this case, formula (52) implies that $\mathrm{Nr}\,\xi_1^0 = \mathrm{Nr}\,\xi_2^0$ and we can bound the total number of pairs (ξ_1^0, ξ_2^0) using Lemma 9.8. The number of pairs $\big(\mathrm{Tr}(\xi_1), \mathrm{Tr}(\xi_2)\big)$ is trivially bounded by $\ll \sqrt{N}$ because $|\mathrm{Tr}(\xi_1)| = |\mathrm{Tr}(\xi_2)|$. The resulting bound on the pairs (ξ_1, ξ_2) is consistent with the claim. □

10. Proof of the main theorem

This section is dedicated to establishing our main result, Theorem 1.1. Recall that $\mathscr{B}_m^{\mathrm{new}}$ is an orthonormal basis of Hecke newforms of weight $m > 2$. We can combine Corollaries 7.7 and 8.5 to deduce

$$\sum_{f\in\mathscr{B}_m^{\mathrm{new}}} |f^\sharp(g)|^4 \ll_\varepsilon (qD_B)^{3+\varepsilon}\frac{m^2}{m-2}\Bigg\{\frac{m}{2}\int_0^\infty \Bigg[\sum_{n=1}^{(qD_B)^2 m}\frac{1}{n}M(g,n;\delta)^2\Bigg]^{1/2} + \Bigg[\sum_{n>(qD_B)^2 m}\frac{\exp(-n/(qD_B)^2)}{n}M(g,n;\delta)^2\Bigg]^{1/2}\frac{d\delta}{(1+\delta)^{m/2+1}}\Bigg\}^2. \tag{53}$$

On the right-hand side, we have denoted by $M(g,n;\delta)$ the counting function associated to a maximal order containing $\mathscr{R}$.

10.1 Proof of the main theorem for the split matrix algebra

Let $\mathscr{F}$ be a fundamental domain for the action $\mathbf{SL}_2(\mathbb{Z})$ on $\mathbb{H}$. Recall that in this case $\Gamma = \Gamma_0(q) < \mathbf{SL}_2(\mathbb{Z})$. For $g \in \mathbf{G}(\mathbb{R}) = \mathbf{SL}_2(\mathbb{R})$, we denote

$$\mathrm{ht}_\Gamma(g) = \min\big\{y \mid \exists\gamma\in\mathbf{SL}_2(\mathbb{Z})\colon (\gamma g).i = x + iy \in \mathscr{F}\big\}.$$

We first bound the sum $\sum_{n=1}^{(qD_B)^2 m}$. Because $M(g,n;\delta)$ is the count associated to the maximal order $\mathrm{Mat}_{2\times 2}(\mathbb{Z})$, the sum is invariant under the operation of replacing g by γg for any $\gamma \in \mathbf{SL}_2(\mathbb{Z})$. In particular, we can arrange $g.i = x + iy$ with $y = \mathrm{ht}_\Gamma(g)$. We need to convert the logarithmic sum $\sum(1/n)M(g,n;\delta)^2$ to an unweighted sum. We achieve this using the general, integration-by-parts, identity

$$\sum_{n=1}^N \frac{1}{n}f(n) = \frac{1}{N}\sum_{n=1}^N f(n) + \int_1^N \frac{1}{t}\sum_{n=1}^t f(n)\frac{dt}{t},$$

which holds for any $f\colon \mathbb{N}\to\mathbb{C}$. This identity and Proposition 9.1 imply

$$\sum_{n=1}^N \frac{1}{n}M(g,n;\delta)^2 \ll_\varepsilon N^{2+\varepsilon}\delta^2 + N^\varepsilon + \mathrm{ht}_\Gamma(g)^2\begin{cases} N^{1+\varepsilon}\delta^{3/2} + N^{1/2+\varepsilon}\delta, & \delta < 1,\\ N^{1+\varepsilon}\delta, & \delta \geq 1,\end{cases} \tag{54}$$

where $N = (qD_B)^2 m$. We next need to compute the integral $\int_0^\infty \sqrt{\cdots}(d\delta/(1+\delta)^{m/2+1})$. We use the l^2–l^1 inequality to separate the terms in (54) under the square root. To compute the contribution of the term proportional to $\mathrm{ht}_\Gamma(g)$ in (54), we split the integral over δ into $\int_0^1 + \int_1^\infty$. For the former integral, we have $(1+\delta) \geq e^{\delta/2}$ for $\delta \in [0,1]$ as the function $(1+\delta)e^{-\delta/2}$ only

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

increases in the interval $[0,1]$ and is equal to 1 at 0. Hence, for $\kappa > -1$, we have[14]

$$\int_0^1 \frac{\delta^\kappa}{(1+\delta)^{m/2+1}}\,d\delta \le \int_0^\infty \delta^\kappa e^{-\delta(m/4+1/2)}\,d\delta = \frac{\Gamma(1+\kappa)}{(m/4+1/2)^{1+\kappa}} \ll_\kappa m^{-1-\kappa}.$$

On the other hand, we have for $\kappa < m/2$ that

$$\int_1^\infty \frac{\delta^\kappa}{(1+\delta)^{m/2+1}}\,d\delta \le \int_1^\infty (1+\delta)^{\kappa-m/2-1}\,d\delta = \frac{2^{-(m/2-\kappa)}}{m/2-\kappa}.$$

Hence, for the first two terms in (54), we have

$$\int_0^\infty \frac{N^{1+\varepsilon}\delta + N^\varepsilon}{(1+\delta)^{m/2+1}}\,d\delta \ll_\kappa \frac{N^{1+\varepsilon}m^{-1}+N^\varepsilon}{m-2} \ll \frac{(qD_B)^{2+\varepsilon}m^\varepsilon}{m-2},$$

and for the last term, we have

$$\begin{aligned}
&\mathrm{ht}_\Gamma(g)\int_0^\infty \begin{cases} N^{1/2+\varepsilon}\delta^{3/4}+N^{1/4+\varepsilon}\delta^{1/2}, & \delta<1,\\ N^{1/2+\varepsilon}\delta^{1/2}, & \delta\ge 1,\end{cases}\cdot\frac{d\delta}{(1+\delta)^{m/2+1}}\\
&\ll_\kappa \mathrm{ht}_\Gamma(g)N^\varepsilon\left(\frac{N^{1/2}}{m^{1+3/4}}+\frac{N^{1/4}}{m^{1+1/2}}+2^{-m/2}\frac{N^{1/2}}{m-2}\right)\\
&\ll \mathrm{ht}_\Gamma(g)(qD_B)^{1+\varepsilon}\frac{m^{-1/4+\varepsilon}}{m-2}.
\end{aligned}$$

In conclusion,

$$\int_0^\infty \left[\sum_{n=1}^{(qD_B)^2m}\frac{1}{n}M(g,n;\delta)^2\right]^{1/2}\frac{d\delta}{(1+\delta)^{m/2+1}} \ll_\varepsilon (qD_B)^{2+\varepsilon}m^\varepsilon\frac{1+\mathrm{ht}_\Gamma(g)m^{-1/4}}{m-2}.$$

The computation of the bound for the integral

$$\int_0^\infty \left[\sum_{n>(qD_B)^2m}\frac{\exp(-n/(qD_B)^2)}{n}M(g,n;\delta)^2\right]^{1/2}\frac{d\delta}{(1+\delta)^{m/2+1}}$$

uses a very similar argument, except that we need to apply the integration-by-parts identity

$$\sum_{n>Am}\frac{\exp(-n/A)}{n}f(n) = -\frac{\exp(-m)}{Am}\sum_{n=1}^{Am}f(n)+\frac{1}{A}\int_m^\infty \exp(-t)\sum_{n=1}^{At}f(n)\left(1+\frac{1}{t}\right)\frac{dt}{t},$$

that holds for any function $f\colon \mathbb{N}\to\mathbb{C}$ satisfying $\log f(n) = o(n)$ and $A, m \ge 1$. The contributions of these terms is then easily seen to be negligible.

Combining these inequalities with (53), we arrive at

$$\sum_{f\in\mathcal{B}_m^{\mathrm{new}}}|f^\sharp(g)|^4 \ll_\varepsilon (qD_B)^{7+\varepsilon}\frac{m^{4+\varepsilon}}{(m-2)^3}\big(1+\mathrm{ht}_\Gamma(g)^2m^{-1/2}\big).$$

This is consistent with the first claim in Theorem 1.1 for $m > 2$. As mentioned in the introduction, the second claim requires the additional input of [BKY13, Theorem 1.8], which says that most of the L^4-mass is concentrated on $\mathrm{ht}_\Gamma(g) \ll m^{1/4}$. Since the extension of said theorem to include a polynomial-level dependence follows their proof almost verbatim, we leave it to the reader.

[14] We thank the referee for this simplification in estimating the integral.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

10.2 Proof of the main theorem for division algebras

In this section, we use the notation of § 9.2. We follow the same arguments as for the split algebra replacing Proposition 9.1 by Proposition 9.6 to arrive at

$$\sum_{f\in\mathcal{B}_m^{\mathrm{new}}} |f^\sharp(g)|^4 \ll_\varepsilon |D_E|^{2+\varepsilon}(\lambda+\lambda^{-1})^{2+\varepsilon}(qD_B)^{7+\varepsilon}\frac{m^{4+\varepsilon}}{(m-2)^3}.$$

Recall that $\mathcal{O}_E \hookrightarrow \mathcal{R}$ is any optimal embedding of the ring of integers of an imaginary quadratic field into the fixed maximal order. By [CF99], this is always possible if any prime dividing D_B is inert in E. Using the Chinese remainder theorem, one deduces that such a discriminant D_E exists satisfying $|D_E| \ll D_B$.

Lastly, we can replace g by any γg for any $\gamma \in \Gamma$, hence $\lambda+\lambda^{-1}$ is polynomially bounded by the volume of $\Gamma\backslash\mathbf{G}(\mathbb{R})$, as follows from [CL16]. The latter has been recently improved by the second named author in [Ste23].

Acknowledgements

It is a pleasure to thank Paul Nelson for many enlightening discussions on this project and very useful comments on an earlier draft of this manuscript. We would like to deeply thank Peter Sarnak for his continuous encouragement and numerous fruitful discussions on the topic. We thank Jared Wunsch for his help with the analytic aspects of this work. We are also grateful to Valentin Blomer and Simon Marshall for their comments on a previous version of the manuscript. Last but not least, we would like to thank the referees for their thorough reading of the manuscript and their suggestions which lead to simplifications in various parts and improved the overall quality of the manuscript.

The majority of this work was conducted during a stay of the second named author at Northwestern University, and subsequently completed at the respective home institutions: Northwestern University and Institute for Advanced Study/ETH Zürich.

Conflicts of interest

None.

Financial support

The first named author has been supported by an AMS Centennial Fellowship and he would like to thank the AMS for its generosity. The second named author would like to thank Northwestern University for their hospitality, the Institute for Advanced Study, where he was supported by the National Science Foundation Grant No. DMS-1638352 and the Giorgio and Elena Petronio Fellowship Fund II, and the Institute for Mathematical Research (FIM) at ETH Zürich. During revisions for publication, the second named author was employed at the Huawei Research Center in Zürich.

Journal information

Compositio Mathematica is owned by the Foundation Compositio Mathematica and published by the London Mathematical Society in partnership with Cambridge University Press. All surplus income from the publication of *Compositio Mathematica* is returned to mathematics and higher education through the charitable activities of the Foundation, the London Mathematical Society and Cambridge University Press.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

REFERENCES

AL70 A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

Ass17 E. Assing, *On sup-norm bounds part I: ramified Maaß newforms over number fields*, J. Eur. Math. Soc. (JEMS) **26** (2024), 1559–1609.

BH10 V. Blomer and R. Holowinsky, *Bounding sup-norms of cusp forms of large level*, Invent. Math. **179** (2010), 645–681.

BHM16 V. Blomer, G. Harcos and D. Milićević, *Bounds for eigenforms on arithmetic hyperbolic 3-manifolds*, Duke Math. J. **165** (2016), 625–659.

BHMM20 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** (2020), 1–53.

BKY13 V. Blomer, R. Khan and M. Young, *Distribution of mass of holomorphic cusp forms*, Duke Math. J. **162** (2013), 2609–2644.

BM11 V. Blomer and P. Michel, *Sup-norms of eigenfunctions on arithmetic ellipsoids*, Int. Math. Res. Not. IMRN **2011** (2011), 4934–4966.

BM13 V. Blomer and P. Michel, *Hybrid bounds for automorphic forms on ellipsoids over number fields*, J. Inst. Math. Jussieu **12** (2013), 727–758.

BM15 V. Blomer and P. Maga, *The sup-norm problem for PGL(4)*, Int. Math. Res. Not. IMRN **2015** (2015), 5311–5332.

BM16 V. Blomer and P. Maga, *Subconvexity for sup-norms of cusp forms on* $\mathrm{PGL}(n)$, Selecta Math. (N.S.) **22** (2016), 1269–1287.

Bor63 A. Borel, *Some finiteness properties of adele groups over number fields*, Publ. Math. Inst. Hautes Études Sci. **16** (1963), 5–30.

BP16 V. Blomer and A. Pohl, *The sup-norm problem on the Siegel modular space of rank two*, Amer. J. Math. **138** (2016), 999–1027.

CF99 T. Chinburg and E. Friedman, *An embedding theorem for quaternion algebras*, J. Lond. Math. Soc. (2) **60** (1999), 33–44.

CL16 M. Chu and H. Li, *Small generators of cocompact arithmetic Fuchsian groups*, Proc. Amer. Math. Soc. **144** (2016), 5121–5127.

DS15 S. Das and J. Sengupta, *L^∞ norms of holomorphic modular forms in the case of compact quotient*, Forum Math. **27** (2015), 1987–2001.

DS20 S. Das and J. Sengupta, *L^∞ norms of holomorphic modular forms in the case of compact quotient*, Preprint (2020), arXiv:1301.3677v3.

GZ86 B. H. Gross and D. B. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), 225–320.

HL94 J. Hoffstein and P. Lockhart, *Coefficients of Maass forms and the Siegel zero*, Ann. of Math. (2) **140** (1994), 161–181; with an appendix by D. Goldfeld, J. Hoffstein and D. Lieman.

HNS19 Y. Hu, P. D. Nelson and A. Saha, *Some analytic aspects of automorphic forms on* $\mathrm{GL}(2)$ *of minimal type*, Comment. Math. Helv. **94** (2019), 767–801.

HS20 Y. Hu and A. Saha, *Sup-norms of eigenfunctions in the level aspect for compact arithmetic surfaces, II: newforms and subconvexity*, Compos. Math. **156** (2020), 2368–2398.

HT12 G. Harcos and N. Templier, *On the sup-norm of Maass cusp forms of large level: II*, Int. Math. Res. Not. IMRN **2012** (2012), 4764–4774.

HT13 G. Harcos and N. Templier, *On the sup-norm of Maass cusp forms of large level. III*, Math. Ann. **356** (2013), 209–216.

Ich08 A. Ichino, *Trilinear forms and the central values of triple product L-functions*, Duke Math. J. **145** (2008), 281–307.

IS95 H. Iwaniec and P. Sarnak, *L^∞ norms of eigenfunctions of arithmetic surfaces*, Ann. of Math. (2) **141** (1995), 301–320.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

JL70 H. Jacquet and R. P. Langlands, *Automorphic forms on GL(2)*, Lecture Notes in Mathematics, vol. 114 (Springer, Berlin–New York, 1970).

Kha14 R. Khan, *On the fourth moment of holomorphic Hecke cusp forms*, Ramanujan J. **34** (2014), 83–107.

Kır14 E. M. Kıral, *Bounds on sup-norms of half-integral weight modular forms*, Acta Arith. **165** (2014), 385–399.

KNS22 I. Khayutin, P. D. Nelson and R. S. Steiner, *Theta functions, fourth moments of eigenforms, and the sup-norm problem II*, Forum Math. Pi **12** (2024), e11.

Mar14 S. Marshall, *Upper bounds for Maass forms on semisimple groups*, Preprint (2014), arXiv:1405.7033.

Nel11 P. D. Nelson, *Equidistribution of cusp forms in the level aspect*, Duke Math. J. **160** (2011), 467–501.

Nel15 P. D. Nelson, *Evaluating modular forms on Shimura curves*, Math. Comp. **84** (2015), 2471–2503.

Nel16 P. D. Nelson, *Quantum variance on quaternion algebras, I*, Preprint (2016), arXiv:1601.02526.

Nel17 P. D. Nelson, *Quantum variance on quaternion algebras, II*, Preprint (2017), arXiv:1702.02669.

Nel19 P. D. Nelson, *Quantum variance on quaternion algebras, III*, Preprint (2019), arXiv:1903.08686.

Nel20 P. D. Nelson, *Bounds for twisted symmetric square L-functions via half-integral weight periods*, Forum Math. Sigma **8** (2020), e44.

Pet40 H. Petersson, *Über eine Metrisierung der automorphen Formen und die Theorie der Poincaréschen Reihen*, Math. Ann. **117** (1940), 453–537.

Pet41 H. Petersson, *Einheitliche Begründung der Vollständigkeitssätze für die Poincaréschen Reihen von reeller Dimension bei beliebigen Grenzkreisgruppen von erster Art*, Abh. Math. Semin. Hansischen Univ. **14** (1941), 22–60.

RW21 P. Ramacher and S. Wakatsuki, *Subconvex bounds for Hecke-Maass forms on compact arithmetic quotients of semisimple Lie groups*, Math. Z. **298** (2021), 1383–1424.

Sah17a A. Saha, *Hybrid sup-norm bounds for Maass newforms of powerful level*, Algebra Number Theory **11** (2017), 1009–1045.

Sah17b A. Saha, *On sup-norms of cusp forms of powerful level*, J. Eur. Math. Soc. (JEMS) **19** (2017), 3549–3573.

Sah20 A. Saha, *Sup-norms of eigenfunctions in the level aspect for compact arithmetic surfaces*, Math. Ann. **376** (2020), 609–644.

Sar03 P. Sarnak, *Spectra of hyperbolic surfaces*, Bull. Amer. Math. Soc. (N.S.) **40** (2003), 441–478.

Sel42 A. Selberg, *On the zeros of Riemann's zeta-function*, Skr. Norske Vid.-Akad. Oslo I **1942** (1942), 59.

Sel56 A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20** (1956), 47–87.

Shi72 H. Shimizu, *Theta series and automorphic forms on* GL_2, J. Math. Soc. Japan **24** (1972), 638–683.

Sog88 C. D. Sogge, *Concerning the* L^p *norm of spectral clusters for second-order elliptic operators on compact manifolds*, J. Funct. Anal. **77** (1988), 123–138.

Ste16 R. S. Steiner, *Uniform bounds on sup-norms of holomorphic forms of real weight*, Int. J. Number Theory **12** (2016), 1163–1185.

Ste17 R. S. Steiner, *Supnorm of modular forms of half-integral weight in the weight aspect*, Acta Arith. **177** (2017), 201–218.

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press

Ste20 R. S. Steiner, *Sup-norm of Hecke–Laplace eigenforms on* S^3, Math. Ann. **377** (2020), 543–553.

Ste23 R. S. Steiner, *Small diameters and generators for arithmetic lattices in* $\mathrm{SL}_2(\mathbb{R})$ *and certain Ramanujan graphs*, Ramanujan J. **62** (2023), 953–966.

SW71 E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*. Princeton Mathematical Series, No. 32 (Princeton University Press, Princeton, NJ, 1971).

SY19 H. Sun and Y. Ye, *Double first moment for* $L(\frac{1}{2}, \mathrm{Sym}^2 f \times g)$ *by applying Petersson's formula twice*, J. Number Theory **202** (2019), 141–159.

Tem10 N. Templier, *On the sup-norm of Maass cusp forms of large level*, Selecta Math. (N.S.) **16** (2010), 501–531.

Tem15 N. Templier, *Hybrid sup-norm bounds for Hecke–Maass cusp forms*, J. Eur. Math. Soc. (JEMS) **17** (2015), 2069–2082.

Van97 J. M. VanderKam, L^∞ *norms and quantum ergodicity on the sphere*, Int. Math. Res. Not. IMRN **1997** (1997), 329–347.

Vig77 M.-F. Vignéras, *Séries thêta des formes quadratiques indéfinies*, in Séminaire Delange-Pisot-Poitou, 17e année (1975/76), Théorie des nombres: Fasc. 1, Exp. No. 20 (1977), 3, http://www.numdam.org/article/SDPP_1975-1976__17_1_A18_0.pdf.

Voi21 J. Voight, *Quaternion algebras*, Graduate Texts in Mathematics, vol. 288 (Springer, Cham, 2021).

Wal85 J.-L. Waldspurger, *Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie*, Compos. Math. **54** (1985), 173–242.

WaT08 T. C. Watson, *Rankin triple products and quantum chaos*, PhD dissertation, Princeton University (2008), arXiv:0810.0425.

Wei64 A. Weil, *Sur certains groupes d'opérateurs unitaires*, Acta Math. **111** (1964), 143–211.

Xia07 H. Xia, *On* L^∞ *norms of holomorphic cusp forms*, J. Number Theory **124** (2007), 325–327.

Zag76 D. Zagier, *Appendix: The Eichler–Selberg trace formula on* $\mathrm{SL}_2(\mathbb{Z})$, in *Introduction to modular forms*. Grundlehren der mathematischen Wissenschaften, vol. 222 (Springer, Berlin–New York, 1976), 44–55.

Zag77 D. Zagier, *Modular forms whose Fourier coefficients involve zeta-functions of quadratic fields*, in Modular functions of one variable, VI (Proc. Second Internat. Conf., Univ. Bonn, Bonn, 1976), Lecture Notes in Mathematics, vol. 627 (Springer, 1977), 105–169.

Ilya Khayutin Department of Mathematics, Northwestern University, Evanston IL 60203, USA

Raphael S. Steiner raphael.steiner.academic@gmail.com
Computing Systems Lab, Huawei Zurich Research Center, Thurgauerstrasse 80,
CH-8050 Zurich, Switzerland

https://doi.org/10.1112/S0010437X24007437 Published online by Cambridge University Press