Forum of Mathematics, Pi (2024), **Vol. 12**:e11 1–53
doi:10.1017/fmp.2024.9

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

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

Ilya Khayutin[1], Paul D. Nelson[2] and Raphael S. Steiner[3]

[1]Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208, USA.
[2]Department of Mathematics, Aarhus University, Ny Munkegade 118, Building 1530, room 417, 8000 Aarhus C, Denmark;
E-mail: paul.nelson@math.au.dk.
[3]Computing Systems Lab, Huawei Zurich Research Center, Thurgauerstrasse 80, 8050 Zurich, Switzerland;
E-mail: raphael.steiner.academic@gmail.com (corresponding author).

Received: 19 January 2023; **Revised:** 4 February 2024; **Accepted:** 5 April 2024

2020 Mathematics Subject Classification: *Primary* – 11F12; *Secondary* – 11F27, 11F70, 11F72, 11D45, 11N75, 14G35

Abstract

Let f be an L^2-normalized holomorphic newform of weight k on $\Gamma_0(N)\backslash\mathbb{H}$ with N squarefree or, more generally, on any hyperbolic surface $\Gamma\backslash\mathbb{H}$ attached to an Eichler order of squarefree level in an indefinite quaternion algebra over $\mathbb{Q}$. Denote by V the hyperbolic volume of said surface. We prove the sup-norm estimate

$$\|\Im(\cdot)^{\frac{k}{2}} f\|_\infty \ll_\varepsilon (kV)^{\frac{1}{4}+\varepsilon}$$

with absolute implied constant. For a cuspidal Maaß newform φ of eigenvalue λ on such a surface, we prove that

$$\|\varphi\|_\infty \ll_{\lambda,\varepsilon} V^{\frac{1}{4}+\varepsilon}.$$

We establish analogous estimates in the setting of definite quaternion algebras.

Contents

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

1. Introduction

Let $\Gamma\backslash\mathbb{H}$ be a finite volume hyperbolic surface. A basic problem in quantum chaos is to understand the limiting behavior of L^2-normalized Laplace eigenfunctions φ on $\Gamma\backslash\mathbb{H}$. This behavior can be quantified through weak limits of L^2-masses ('quantum ergodicity'), bounds for L^p-norms and so forth. We consider in this paper the *sup-norm problem*, which consists of bounding the supremum or L^∞-norm of an L^2-normalized eigenfunction φ with respect to the eigenvalue λ_φ and/or the geometry of the underlying manifold $\Gamma\backslash\mathbb{H}$. A general bound in this direction, due to Bérard [Bér77], asserts that

$$\|\varphi\|_\infty \ll_\Gamma (1+|\lambda_\varphi|)^{\frac{1}{4}}/\log(2+|\lambda_\varphi|). \tag{1.1}$$

Here and henceforth, $A \ll B$ means that there is a constant C such that $|A| \le CB$; we allow C to depend on any subscripts of $\ll$ and write ε for an arbitrary, but sufficiently small, positive constant, which may change from line to line.

Stronger bounds have been established in the *arithmetic* case that

- $\Gamma\backslash\mathbb{H}$ is an arithmetic manifold, such as the modular surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ or a congruence cover, and
- φ is a *Hecke–Maaß form*, that is, an eigenfunction not only of the Laplacian but also of the Hecke operators.

The pioneering result in that case is due to Iwaniec–Sarnak [IS95], who showed for congruence lattices Γ that

$$\|\varphi\|_\infty \ll_{\Gamma,\varepsilon} (1+|\lambda_\varphi|)^{\frac{5}{24}+\epsilon}. \tag{1.2}$$

The above estimates depend in an unspecified manner upon the underlying manifold. Consider, for instance, the case that Γ is the Hecke congruence subgroup $\Gamma_0(N) = \mathrm{SL}_2(\mathbb{Z}) \cap \left(\begin{smallmatrix}\mathbb{Z} & \mathbb{Z}\\ N\mathbb{Z} & \mathbb{Z}\end{smallmatrix}\right)$ so that $\Gamma\backslash\mathbb{H}$ is an arithmetic manifold of volume $N^{1+o(1)}$. We suppose that N is squarefree. A direct quantification of the Iwaniec–Sarnak argument (see [BH10, §10]) gives the estimate

$$\|\varphi\|_\infty \ll_\varepsilon N^{\frac{1}{2}+\varepsilon}(1+|\lambda_\varphi|)^{\frac{5}{24}+\epsilon}, \tag{1.3}$$

where we normalize φ to have L^2-norm one with respect to the *hyperbolic probability measure*, that is, the multiple of the hyperbolic measure having total volume one. The *level aspect* case of the sup-norm problem is to improve the dependence of the bound (1.3) upon N. The first improvement in the exponent was a major breakthrough of Blomer–Holowinsky [BH10], achieved 13 years after the work of Iwaniec–Sarnak. For a Hecke–Maaß newform φ of eigenvalue λ_φ, they managed to show

$$\|\varphi\|_\infty \ll_{\lambda_\varphi} N^{\frac{1}{2}-\frac{1}{37}} \tag{1.4}$$

(with explicit polynomial dependence upon λ_φ). Subsequently, Templier [Tem10] and Harcos–Templier [HT12, HT13] established several improved bounds, culminating in

$$\|\varphi\|_\infty \ll_{\lambda_\varphi,\epsilon} N^{\frac{1}{3}+\epsilon}. \tag{1.5}$$

The estimate (1.5) is comparable in strength to the Weyl bound for the Riemann zeta function and has long been regarded as a natural limit for the sup-norm problem in the squarefree level aspect [HT13, Remarks (i)]. It has been extended to number fields [BHM16, BHMM20, Ass24] and to more general vectors than newforms [HNS19, Ass21]. For levels that are not squarefree (e.g., powers of a fixed prime), the flavor of the problem is quite different (see Remark 1.4), and stronger estimates have been achieved in [Sah17, Mar16, Sah20, Com21, HS20].

In this work, we bring new methodology to bear on the sup-norm problem in the squarefree level aspect. By obtaining optimal solutions to the technical problems that arise in applying that methodology, we deduce the following improvement of Equation (1.5).

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Theorem 1.1. *Let N be a squarefree natural number. Let φ be a cuspidal Hecke–Maaß newform for $\Gamma_0(N)$ with trivial (central) character. Suppose that φ is L^2-normalized with respect to the hyperbolic probability measure on $\Gamma_0(N)\backslash\mathbb{H}$. Then*

$$\|\varphi\|_\infty \ll_{\lambda_\varphi,\epsilon} N^{\frac{1}{4}+\epsilon}.$$

Our main results apply not only to $\Gamma_0(N)\backslash\mathbb{H}$ but also to compact arithmetic quotients. In general, such a manifold is of the shape $\Gamma\backslash\mathbb{H}$, where Γ is commensurable with a lattice attached to a maximal order in a quaternion algebra B over a totally real field F, with B split at exactly one Archimedean place. We are content here to consider the case $F = \mathbb{Q}$ so that B is an indefinite quaternion algebra, characterized up to isomorphism by its reduced discriminant d_B. For each natural number N coprime to d_B, we denote by $\Gamma_0^B(N)$ the group of proper (i.e., norm one) units arising from an Eichler order of level N in B (see Section 2.1 for details). For example, if $B = \mathrm{Mat}_{2\times 2}(\mathbb{Q})$, then we could take $\Gamma_0^B(N) = \Gamma_0(N)$. We prove the following theorem.

Theorem 1.2. *Let $\Gamma = \Gamma_0^B(N)$ be as above with the level N being squarefree. Let φ be a cuspidal Hecke–Maaß newform for Γ with trivial (central) character, L^2-normalized with respect to the hyperbolic probability measure on $\Gamma\backslash\mathbb{H}$. Then, with $V = (d_B N)^{1+o(1)}$ the covolume of Γ,*

$$\|\varphi\|_\infty \ll_{\lambda_\varphi,\epsilon} V^{\frac{1}{4}+\epsilon}. \tag{1.6}$$

Theorem 1.2 specializes to Theorem 1.1 upon taking $B = \mathrm{Mat}_{2\times 2}(\mathbb{Q})$. It improves upon (the $F = \mathbb{Q}$ case of) Templier's result [Tem10], which gave the nontrivial bound $V^{\frac{1}{2}-\frac{1}{24}+\varepsilon}$. We emphasize that the estimate (1.6) is uniform in the quaternion algebra B, hence gives a strong saving in the 'discriminant aspect'; the first nontrivial results in that aspect (for B indefinite, as we have assumed) were established only very recently by Toma [Tom23], updating an earlier preprint, giving (among other things) the bound $V^{\frac{1}{2}-\frac{1}{30}+\varepsilon}$. Our method applies equally in the setting of *definite* quaternion algebras, where we improve the exponent $\frac{1}{3}$ of Blomer–Michel [BM11, BM13] down to $\frac{1}{4}$ in analogy with Theorem 1.2 (see Section §2.3 for details).

Remark 1.3. The dependence on the eigenvalue in Equation (1.6) that follows from our proof is of exponential nature. With some finer Archimedean considerations, it seems likely that one could show $\|\varphi\|_\infty \ll_\varepsilon \lambda_\varphi^{\frac{1}{4}+\varepsilon} V^{\frac{1}{4}+\varepsilon}$; indeed, by comparison, we obtain such an estimate for the definite analogue of Equation (1.6) (see Corollary 2.3). Such a refinement of Equation (1.6) seems to require lengthy Archimedean calculations that we feel would distract from the primary novelties of this paper concerning the level aspect.

Remark 1.4. We have noted already that we focus in this paper on the case of squarefree levels. The opposite case is the depth aspect, where the level is a power $N = p^n$ of a fixed prime p. In that case, local arguments give the bound $\|\varphi\|_\infty \ll_{p,d_B,\varepsilon} (\lambda_\varphi N)^{1/4+\varepsilon}$ [Mar16], which has been improved to $\|\varphi\|_\infty \ll_{\lambda_\varphi,p,d_B,\varepsilon} N^{5/24+\varepsilon}$ [HS20] via arithmetic amplification and refined local analysis.

Remark 1.5. In a function field setting analogous to that of Theorem 1.1, Sawin [Saw21] has used geometric techniques to establish (among other things) the sup-norm bound $\ll N^{\frac{1}{4}+\alpha_q}$, where $\alpha_q > 0$ tends to zero as the cardinality q of the underlying finite field tends to ∞. We do not see any obstruction to adapting the techniques of this paper to the function field setting, where we expect they would give the improved bound $\ll_\varepsilon N^{\frac{1}{4}+\varepsilon}$.

By combining the arguments of this paper with those of the prequel [KS20] concerning the weight aspect for holomorphic forms, we obtain the following uniform hybrid bound in the weight and level aspects.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Theorem 1.6. *Let $\Gamma = \Gamma_0^B(N)$ be as in Theorem 1.2. Let f be a cuspidal holomorphic newform for Γ with trivial (central) character and weight $k \geq 2$. Suppose f is L^2-normalized with respect to the hyperbolic probability measure on $\Gamma\backslash\mathbb{H}$. Then*

$$\|\Im(\cdot)^{\frac{k}{2}} f\|_\infty \ll_\epsilon (kV)^{\frac{1}{4}+\epsilon},$$

where $V = (d_B N)^{1+o(1)}$ denotes the covolume of Γ.

1.1. Selected applications

A straightforward application of these improved sup-norms is to L^p-norms for $2 \leq p \leq \infty$ by means of interpolation. We state here only the split holomorphic case, as in this case, strong L^4-bounds were given by Buttcane–Khan [BK15] with subconvexity input from [You17].

Corollary 1.7. *Let q denote an odd prime and f a cuspidal holomorphic newform for $\Gamma_0(q)$ with trivial (central) character and weight k. Suppose f is L^2-normalized with respect to the hyperbolic probability measure on $\Gamma_0(q)\backslash\mathbb{H}$. Then, for $2 \leq p \leq \infty$ and any $\eta > 0$, we have*

$$\|\Im(\cdot)^{\frac{k}{2}} f\|_p \ll_{k,\eta} \begin{cases} q^{\frac{1}{6}-\frac{1}{3p}+\eta}, & 2 \leq p \leq 4, \\ q^{\frac{1}{4}-\frac{2}{3p}+\eta}, & 4 \leq p \leq \infty, \end{cases}$$

for k sufficiently large in terms of η.

Further applications of sup-norm bounds include shifted convolution problems and subconvexity results for L-functions; see, for example, [Har03, HM06, HC19, HS20, Nor21]. Often, such applications would be obtained from a uniform version of Wilton's estimate. By applying the arguments of [HM06, §2.7] with our improved sup-norm bound, we derive the following corollary.

Corollary 1.8. *Let $\lambda(m)$, $m \in \mathbb{N}$, denote the Hecke eigenvalues, normalized so that the Ramanujan conjecture reads $|\lambda(m)| \ll_\varepsilon m^\varepsilon$, of either a cuspidal Hecke–Maaß newform or a cuspidal holomorphic newform of weight k on $\Gamma_0(N)$ with trivial (central) character, where N is squarefree. Then, for any $\alpha \in \mathbb{R}$, one has*

$$\sum_{m \leq M} \lambda(m) e(m\alpha) \ll_\epsilon M^{\frac{1}{2}+\epsilon} \cdot \begin{cases} N^{\frac{1}{4}+\epsilon}, & \text{in the Maaß case}, \\ N^{\frac{1}{4}+\epsilon} k^{\frac{1}{2}+\epsilon}, & \text{in the holomorphic case}, \end{cases}$$

where the implied constant in the Maaß case further depends on the eigenvalue of the form.

As a consequence, we may, for example, improve the main theorem in [HC19].

Corollary 1.9. *Let φ either be a cuspidal Hecke–Maaß newform or a cuspidal holomorphic newform on $\Gamma_0(q)$, with q prime. Let χ be a primitive Dirichlet character of modulus m with $(m, q) = 1$. Suppose that $q = m^\eta$ with $0 < \eta < 2$. Then, we have*

$$L(\varphi \otimes \chi, \tfrac{1}{2}) \ll_\epsilon \mathcal{C}^{\frac{1}{4}+\epsilon} \left(\mathcal{C}^{-\frac{\eta}{4(2+\eta)}} + \mathcal{C}^{-\frac{2-\eta-4\vartheta}{8(2+\eta)}} \right),$$

where the implied constant depends on the eigenvalue respectively weight of φ, $\mathcal{C} = qm^2$ is the conductor of the L-function and ϑ is the current best bound towards the generalized Ramanujan conjecture if φ is a Maaß form and 0 if φ is holomorphic.

1.2. The fourth moment and further applications

The method underlying most previous works on this problem, including the work of Harcos–Templier giving the bound $\ll_\epsilon N^{1/3+\epsilon}$, is based on the amplification method introduced in the original paper of Iwaniec–Sarnak. Recently, Steiner [Ste20] and Khayutin–Steiner [KS20] introduced a new method

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

based on analysis of fourth moments over families. The key observation of these papers was that such a fourth moment naturally arises as the L^2-norm of a theta kernel. Alternatively, Blomer *et al.* [BHMM22] have demonstrated that one may use Voronoï summation for Rankin–Selberg convolutions in place of a theta kernel. Prior to the application to fourth moments, theta kernels have played similar roles in the study of quantum variance [Nel16, Nel17, Nel19, Nel20], numerical computations [Nel15] and in the proof of Waldspurger's formula [Wal85]. In each of these earlier works, theta kernels apparently served as a substitute for parabolic Fourier expansions, giving a tool for establishing analogues on compact quotients (where such expansions are not available) of results known already for noncompact quotients. The present work differs in that our main result is new even for the noncompact quotients $\Gamma_0(N)\backslash\mathbb{H}$.

In this paper, we follow generally the theta kernel strategy of the prequel [KS20] and prove a fourth moment bound from which one may deduce the Theorems 1.1, 1.2 and 1.6 after some additional analysis near any cusps. In what follows, we let $\Gamma = \Gamma_0^B(N)$ be a lattice as in Theorem 1.2 and denote by $V = (d_B N)^{1+o(1)}$ the volume of $\Gamma\backslash\mathbb{H}$.

The formulation of our results requires some quantification of the closeness of a point $z \in \Gamma\backslash\mathbb{H}$ to the cusps. If $\Gamma\backslash\mathbb{H}$ is noncompact (i.e., $d_B = 1$), then we may assume that $\Gamma = \Gamma_0(N)$, and we set

$$H(z) = \max_{\gamma \in A_0(N)} \Im(\gamma z),$$

where $A_0(N)$ denotes the lattice of Atkin–Lehner operators for $\Gamma_0(N)$ (see Section §2.2 for another formulation of the definition of H). If $\Gamma\backslash\mathbb{H}$ is compact, then we set $H(z) = 0$.

Theorem 1.10. *Let $\Gamma = \Gamma_0^B(N)$ be as in Theorem 1.2. Fix $\Lambda > 0$, and let $(\varphi_i)_i$ be an orthonormal set of cuspidal Hecke–Maaß newforms with trivial (central) character and Laplace-eigenvalue bounded by Λ on the hyperbolic surface $\Gamma\backslash\mathbb{H}$ equipped with the hyperbolic probability measure. Then, for any two points $z, w \in \Gamma\backslash\mathbb{H}$, we have*

$$\sum_i \left(|\varphi_i(z)|^2 - |\varphi_i(w)|^2\right)^2 \ll_{\epsilon,\Lambda} V^{1+\epsilon}\left(1 + V[H(z)^2 + H(w)^2]\right). \tag{1.7}$$

Similarly, for an orthonormal set $(f_i)_i$ of cuspidal holomorphic newforms for Γ of weight k and trivial (central) character with respect to the hyperbolic probability measure on $\Gamma\backslash\mathbb{H}$, we have

$$\sum_i \left(|\Im(z)^{\frac{k}{2}} f_i(z)|^2 - |\Im(w)^{\frac{k}{2}} f_i(w)|^2\right)^2 \ll_\epsilon (Vk)^{1+\epsilon}\left(1 + V^{\frac{1}{2}}[H(z) + H(w)] + Vk^{-\frac{1}{2}}[H(z)^2 + H(w)^2]\right)$$

for any two points $z, w \in \Gamma\backslash\mathbb{H}$.

In the case that the hyperbolic surface $\Gamma\backslash\mathbb{H}$ is compact, we may integrate z and w over the whole surface and get an essentially sharp bound on the fourth moment of fourth norms in the level aspect, thereby extending a result of Blomer [Blo13] to the case of cocompact lattices Γ.

Corollary 1.11. *With notation and assumptions as in Theorem 1.10 and assuming further that $\Gamma\backslash\mathbb{H}$ is compact, we have*

$$\sum_i \|\varphi_i\|_4^4 \ll_{\epsilon,\Lambda} V^{1+\epsilon},$$

$$\sum_i \|\Im(\cdot)^{\frac{k}{2}} f_i\|_4^4 \ll_\epsilon (Vk)^{1+\epsilon}.$$

This result may also be recast as a double average of triple L-functions by means of Watson's formula [Wat08, Theorem 3].

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

The final application of Theorem 1.10 we mention is to the diameter of compact arithmetic hyperbolic surfaces $\Gamma\backslash\mathbb{H}$ [Ste23]. Here, one may use the sharp bound on the 'fourth moment' of exceptional eigenforms, together with a strong density estimate for the exceptional eigenvalues, to get an optimal estimate on the almost diameter and an estimate on the diameter of the same strength as if one were to assume the Selberg eigenvalue conjecture.

1.3. The added complexity of the level aspect

Compared to the weight aspect treated in the prequel, the level aspect requires many new ideas. Here, we tacitly restrict to the case of *squarefree* level; the general case would require a more nuanced discussion. In some sense, the level aspect may be understood as intermediate in difficulty between the holomorphic and eigenvalue aspects. Indeed, relative to known techniques, the difficulty in the sup-norm problem is reflected in the essential support of the matrix coefficient of the automorphic form being bounded. In the weight, (squarefree) level and eigenvalue aspects, the matrix coefficient concentrates on a space of dimension one, two and three, respectively.

We now briefly recall the main idea of the theta approach and discuss some of the new challenges that arise in the level aspect. We focus first on the case of Hecke–Maaß forms on $\Gamma_0(N)\backslash\mathbb{H}$, as in Theorem 1.1. Take $R = \left(\begin{smallmatrix} \mathbb{Z} & \mathbb{Z} \\ N\mathbb{Z} & \mathbb{Z} \end{smallmatrix}\right)$ so that that the set of proper units of R is precisely $\Gamma_0(N)$. For $\ell \mid N$, let $R(\ell) = \left(\begin{smallmatrix} \mathbb{Z} & \mathbb{Z}/\ell \\ N\mathbb{Z}/\ell & \mathbb{Z} \end{smallmatrix}\right)$ denote the partially dualized lattices of the order R. Let $\sigma_z \in \mathrm{SL}_2(\mathbb{R})$ be any matrix taking i to $z \in \mathbb{H}$. Let φ be an *arithmetically normalized* cuspidal Hecke–Maaß newform. The theta identity at the heart of the argument then reads

$$\langle \theta(z, w; \cdot), \varphi \rangle = \frac{1}{V} \varphi(z) \overline{\varphi(w)}, \tag{1.8}$$

where V denotes the covolume of $\Gamma_0(N)$ and the theta function is given by

$$\theta(z, w; s) = \Im(s) \sum_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \sigma_z^{-1} R \sigma_w} e^{-\pi(a^2+b^2+c^2+d^2)\Im(s)} e^{2\pi i(ad-bc)\Re(s)}. \tag{1.9}$$

By Bessel's inequality, the left-hand side of Equation (1.7) is in essence captured by the L^2-norm of the difference of the theta kernels $\theta(z, z; \cdot) - \theta(w, w; \cdot)$. From here, one may then proceed as in the prequel by covering a fundamental domain by Siegel sets and making use of the orthogonality relations in the unipotent direction. One ends up with a weighted sum over matrices $\gamma_1, \gamma_2 \in R(\ell)$ satisfying $\det(\gamma_1) = \det(\gamma_2)$ and for which the entries of $\sigma_z^{-1}\gamma_i\sigma_z$, $i = 1, 2$, satisfy certain bounds (and similarly for w). The bounds imposed on these entries depend crucially upon the precise choice of Siegel domains, so it is important that we make a good choice. Like in the prequel, we split the count according to whether $\mathrm{tr}(\gamma_1) = \mathrm{tr}(\gamma_2)$ or not.

In the case of nonequal trace, the naïve choice of Siegel domains consisting of $\Gamma_0(N)\backslash \mathrm{SL}_2(\mathbb{Z})$-translates of the standard Siegel domain for $\mathrm{SL}_2(\mathbb{Z})$ leads to a rather challenging counting problem. In order to get a sharp bound on Equation (1.7), one faces the challenge of counting, for each divisor ℓ of N and each T with $\ell^{-1/2} \ll T \ll 1$, the sextuples of integers $(a_1, b_1, c_1, a_2, b_2, c_2)$ satisfying

$$(c_i y N/\ell)^2 + 2(a_i - c_i x N/\ell)^2 + y^{-2}(2a_i x + b/\ell - c_i x^2 N/\ell)^2 \le T^2, \quad i = 1, 2, \tag{1.10}$$

$$b_1 c_1 \equiv b_2 c_2 \pmod{\ell^2/N}. \tag{1.11}$$

We would need to know that the number of such sextuples is roughly $O(\ell T^2)$ in the range $N^{-1} \ll y \ll N^{-1/2}$ and $|x| \le \frac{1}{2}$. We do not know how to establish such a bound directly when, for instance, $\ell = N$. On the other hand, when $\ell = 1$, the congruence condition is void and, using arguments of Harcos–Templier, we can prove the required bound with some room to spare, namely, for T up to $N^{1/2}$. Our solution to this dichotomy is thus to decrease the size of the Siegel domains associated to

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

larger ℓ at the expense of increasing those associated to smaller ℓ. This solution may be implemented most simply by applying an Atkin–Lehner involution to the covering of $\Gamma_0(N)\backslash\mathbb{H}$ by $\mathrm{SL}_2(\mathbb{Z})$-translates of the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$. With this maneuver, we reduce to considering the range $T \ll N^{\frac{1}{2}}\ell^{-1}$. We are then able to prove the required bound by forgoing the congruence condition, reducing the problem to counting triples of integers (a_i, b_i, c_i) satisfying Equation (1.10), which we carry out using geometry of numbers techniques. We refer subsequently to this type of counting problem, where we count traceless matrices $\gamma \in R(\ell)^0$ with a bound on the entries of $\sigma_z^{-1}\gamma\sigma_z$, as 'Type I'.

In the case of equal trace, we need to count sextuples of integers $(a_1, b_1, c_1, a_2, b_2, c_2)$ satisfying Equation (1.10) and

$$a_1^2 + b_1c_1\tfrac{N}{\ell^2} = a_2^2 + b_2c_2\tfrac{N}{\ell^2}. \tag{1.12}$$

We need to bound this count by $O(\ell T)$ in the same ranges as before. We refer to this type of counting problem as 'Type II'. The key observation is that (a_1, b_1, c_1) turns out to determine (a_2, b_2, c_2) up to a small number of possibilities. This allows us to reduce Type II estimates to Type I estimates.

The above arguments suffice for noncompact quotients, that is, for the proof of Theorem 1.1. They rely on the use of matrix coordinates $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with respect to which the lattices $\Gamma_0(N)$ are described by the simple congruence condition $c \equiv 0 \ \ (N)$. We were unable to find an analogously straightforward way to separate the variables in the compact setting (e.g., using fixed quadratic subalgebras of B). In the case that B is definite, the Type I counts were treated in a coordinate-free way by Blomer–Michel [BM11, BM13], who controlled the successive minima of the ternary quadratic lattice underlying $\Gamma_0^B(N)$ in terms of only the content, level and discriminant of that lattice. We extend their arguments to the case that B is indefinite by defining analogous Archimedean quantities that control the disparity of the reduced norm and a majorant, such as the square of the Frobenius norm of $\sigma_z^{-1}\gamma\sigma_z$ for $\gamma \in R(\ell)^0$.

Following the same strategy as in the noncompact case, it remains then only to reduce Type II estimates to Type I estimates. This reduction is perhaps the most subtle part of our counting arguments. It requires us to establish the analogue in the compact setting of the key observation noted following Equation (1.12). For example, in case that B is definite, writing R for an Eichler order of level N, we need to show that for each $n \ll V$, the number of elements $\gamma \in R$ with trace 0 and norm n is essentially $O(1)$, uniformly in N and B. We eventually managed to do so through a delicate argument involving commutators and representations of binary quadratic forms.

1.4. Organization of the paper

The complete statements of our results may be found in Section §2. In Section §3, we reduce the proofs to those of two auxiliary collections of results:

- those concerning matrix counting, and
- those reducing the required estimates for theta functions to matrix counting.

The latter, including the appropriate splicing of a fundamental domain into Siegel sets, may be found in Section §4. In Section §5, we summarize the required properties of the theta functions. The proofs of said properties are deferred to Appendix A.

Sections §7 and §8 are dedicated to the anisotropic extension of the lattice counting argument of Blomer–Michel, which we subsequently apply to the Type I counting problem in Section §9.

The final section, §10, treats the crucial Type II counting problem.

2. Statement of results

2.1. Setup

Let B be a quaternion algebra over $\mathbb{Q}$. We denote by d_B its reduced discriminant, or equivalently, the product of the primes at which B ramifies. We write G for the linear algebraic group over $\mathbb{Q}$ given by

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$G(L) = L^\times\backslash(B \otimes L)^\times$ for any $\mathbb{Q}$-algebra L. Then G is an inner form of PGL_2, and all rational forms of PGL_2 arise in this way. Denote by $[G]$ the adelic quotient $G(\mathbb{Q})\backslash G(\mathbb{A})$. We fix the probability Haar measure on $[G]$. Let K_∞ be a compact maximal torus of $G(\mathbb{R})$. We assume that K_∞ comes equipped with a choice of isomorphism $\kappa : \mathbb{R}/\pi\mathbb{Z} \xrightarrow{\sim} K_\infty$. In the split case $B = \mathrm{Mat}_{2\times 2}(\mathbb{Q})$, we identify $G = \mathrm{PGL}_2$ and set $\kappa(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$.

Let R be an *Eichler order* in B, that is, an intersection of two maximal orders. We denote by N the *level* of R. It is a natural number, coprime to d_B, characterized as follows: For each prime $p \nmid d_B$, there is an isomorphism $B_p := B \otimes \mathbb{Q}_p \cong \mathrm{Mat}_{2\times 2}(\mathbb{Q}_p)$ under which $R_p := R \otimes_{\mathbb{Z}} \mathbb{Z}_p$ maps to the order $\begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ N\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix}$. We may then identify $G(\mathbb{Q}_p)$ with $\mathrm{PGL}_2(\mathbb{Q}_p)$ and the image of $R_p^\times$ with a finite index subgroup of $\mathrm{PGL}_2(\mathbb{Z}_p)$. We assume that N is squarefree so that $d_B N$ is likewise squarefree. We denote by K_R the compact open subgroup of $G(\mathbb{A}_f) = \prod_p' G(\mathbb{Q}_p)$ given by the image of $\prod_p R_p^\times$.

Fix $k \in 2\mathbb{Z}$. Let $\mathcal{A}$ denote the set of cusp forms $\varphi : [G] \to \mathbb{C}$ having the following properties:

- $\varphi(g\kappa(\theta)) = e^{ik\theta}\varphi(g)$ for all θ.
- φ is an eigenfunction for some fixed Casimir operator for $G(\mathbb{R})$, with eigenvalue λ_φ. For the sake of concreteness, we scale the Casimir operator such that it agrees with the standard Laplace operator on the locally symmetric space $G(\mathbb{R})/K_\infty$, which identifies with either $\mathbb{H}$ or S^2.
- φ is K_R-invariant: $\varphi(gk) = \varphi(g)$ for $k \in K_R$.
- φ belongs to the *newspace* for R, that is, K_R is the largest subgroup of $G(\mathbb{A}_f)$ keeping φ invariant. Equivalently, φ is orthogonal the space of $K_{R'}$-invariant cusp forms for every Eichler order R' strictly containing R.
- φ is an eigenform for almost all Hecke operators.

If $k \geq 2$, then we write $\mathcal{A}^{\mathrm{hol}} \subseteq \mathcal{A}$ for the subspace of automorphic lifts of holomorphic forms or, equivalently, the kernel of the raising (resp. lowering) operator attached to K_∞ if B is definite (resp. indefinite).

Denote by $\mathcal{F}$ a maximal orthonormal subset of $\mathcal{A}$. Analogously, we define $\mathcal{F}^{\mathrm{hol}} \subseteq \mathcal{A}^{\mathrm{hol}}$ if $k \geq 2$. Because of the multiplicity-one theorem for GL_2 and its inner forms, the bases $\mathcal{F}, \mathcal{F}^{\mathrm{hol}}$ are unique up to rescaling each element by a scalar of unit magnitude. We note that the sets $\mathcal{A}$, $\mathcal{A}^{\mathrm{hol}}$, $\mathcal{F}$ and $\mathcal{F}^{\mathrm{hol}}$ depend on k; while we suppress this dependence from the notation, k is one of the main parameters of interest.

We will consider several subfamilies of $\mathcal{F}$ and $\mathcal{F}^{\mathrm{hol}}$. Here, a minus sign in the exponent signifies the indefinite case, a plus sign the definite case.

- If B is indefinite and $k = 0$, then we take $\mathcal{F}^- := \mathcal{F}$ and let $\mathcal{F}^-_\lambda$ (resp. $\mathcal{F}^-_{\leq L}$) denote the subsets defined by taking the Casimir eigenvalue equal to $-\lambda$ (respectively at most L in magnitude).
- If B is indefinite and $k \geq 2$, then we take $\mathcal{F}^{-,\mathrm{hol}} := \mathcal{F}^{\mathrm{hol}}$.
- If B is definite and $k = 0$, then we let $\mathcal{F}^+_m \subset \mathcal{F}$ be the subset of forms, whose associated automorphic representation at infinity is isomorphic to the unique irreducible unitary representation of $\mathrm{SU}_2(\mathbb{C})$ of degree $m+1$. In other words, their eigenvalue with respect to the Casimir operator equals to $-m(m+1)$.
- If B is definite and $k \geq 2$, then we let $\mathcal{F}^{+,\mathrm{hol}} = \mathcal{F}^{\mathrm{hol}}$.

2.2. The split case

Assume for the moment that B is split. We may suppose then that

$$B = \mathrm{Mat}_{2\times 2}(\mathbb{Q}), \quad G = \mathrm{PGL}_2, \quad R = \begin{pmatrix} \mathbb{Z} & \mathbb{Z} \\ N\mathbb{Z} & \mathbb{Z} \end{pmatrix}, \tag{2.1}$$

$$K_\infty = \mathrm{PSO}_2(\mathbb{R}), \quad \kappa(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{2.2}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

and may identify

$$[G]/K_\infty K_R \cong \Gamma_0(N)\backslash\mathbb{H}.$$

We define

$$H : [G]/K_\infty K_R \to \mathbb{R}_{>0},$$

as follows. Let $A_0(N) < \mathrm{GL}_2(\mathbb{Q})^+$ denote the group generated by $\Gamma_0(N)$ and all Atkin–Lehner operators. If $g \in [G]/K_\infty K_R$ identifies with $z \in \Gamma_0(N)\backslash\mathbb{H}$, then we set

$$H(g) = H(z) := \max_{\gamma \in A_0(N)} \Im(\gamma z).$$

Since the Atkin–Lehner operators constitute scaling matrices for the various cusps of $\Gamma_0(N)$ (cf. §4.3.1), the function H may be understood as a normalized height or as quantifying closeness to the cusps. Let $\mathfrak{a} \in P^1(\mathbb{Z})$ be a cusp of $\Gamma_0(N)$, and let $\sigma_\mathfrak{a} \in \mathrm{SL}_2(\mathbb{Z})$ such that $\sigma_\mathfrak{a}\infty = \mathfrak{a}$. Then,

$$H(z) = \max_{\mathfrak{a}} \frac{\Im(z_\mathfrak{a})}{w_\mathfrak{a}}, \tag{2.3}$$

where $\mathfrak{a}$ runs over all cusps of $\Gamma_0(N)$, $z_\mathfrak{a} = \sigma_\mathfrak{a}^{-1} z$ and $w_\mathfrak{a}$ is the cusp width of $\mathfrak{a}$.

2.3. *Results on forms*

We adopt the following asymptotic notation $\preccurlyeq$:

$$A_1 \preccurlyeq A_2 \quad \Longleftrightarrow \quad A_1 \ll_\varepsilon (d_B N(1+k)(1+\mu))^\varepsilon A_2,$$

where μ is a quantity relating to the eigenvalues with respect to the Casimir operator of the automorphic forms of relevance to the inequality. Concretely, when talking about the families $\mathcal{F}^-_\lambda, \mathcal{F}^-_{\leq L}, \mathcal{F}^+_m, \mathcal{F}^{\pm,\mathrm{hol}}$ we mean $\mu = |\lambda|, L, m, k$, respectively.

Theorem 2.1. *Let $g_1, g_2 \in [G]$. If B is indefinite, then*

$$\sum_{\varphi \in \mathcal{F}^-_{\leq L}} (|\varphi(g_1)|^2 - |\varphi(g_2)|^2)^2 \preccurlyeq_L d_B N \left(1 + d_B N \left[H(g_1)^2 + H(g_2)^2\right]\right), \tag{2.4}$$

for $L > 0$, and

$$\sum_{\varphi \in \mathcal{F}^{-,\mathrm{hol}}} (|\varphi(g_1)|^2 - |\varphi(g_2)|^2)^2 \preccurlyeq d_B N k \left(1 + (d_B N)^{\frac{1}{2}} \left[H(g_1) + H(g_2)\right] + d_B N k^{-\frac{1}{2}} \left[H(g_1)^2 + H(g_2)^2\right]\right), \tag{2.5}$$

for $k \geq 2$ even. In both cases, the term involving $H(g_{1,2})$ is only present if B is split.

If B is definite, then

$$\sum_{\varphi \in \mathcal{F}^+_m} (|\varphi(g_1)|^2 - |\varphi(g_2)|^2)^2 \preccurlyeq d_B N (m+1)^2, \tag{2.6}$$

for $m \in \mathbb{N}_0$, and

$$\sum_{\varphi \in \mathcal{F}^{+,\mathrm{hol}}} |\varphi(g_1)|^4 \preccurlyeq d_B N k, \tag{2.7}$$

for $k \in 2\mathbb{N}$.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Remark 2.2. In the indefinite holomorphic case (2.5), one may have the same bound for the fourth moment rather than the squared difference under the assumption that the weight satisfies $k \gg_\eta (d_B N)^\eta$ for some $\eta > 0$, in which case the implied constant also depends on η and the implied constant in the assumed lower bound for the weight.

Corollary 2.3. *For $k \geq 2$ and $\varphi \in \mathcal{F}^{\mathrm{hol}}$, we have*

$$\|\varphi\|_\infty \preccurlyeq (d_B N k)^{\frac{1}{4}}.$$

For $k = 0$ and $\varphi \in \mathcal{F}$, we have

$$\|\varphi\|_\infty \preccurlyeq_{\lambda_\varphi} (d_B N)^{\frac{1}{4}}.$$

If B is definite, then we have more precisely

$$\|\varphi\|_\infty \preccurlyeq (d_B N)^{\frac{1}{4}} (1 + |\lambda_\varphi|)^{\frac{1}{4}}.$$

By a well-known procedure, these statements may be translated into the classical language, thus giving rise to the theorems in the introduction. For further details; see, for example, [Bum97, §3.2 & §3.6] for the indefinite case and [BM13] for the definite case.

2.4. Counting problems: setup

2.4.1. Lattices locally dual to R

Let ℓ be a divisor of the squarefree number $d_B N$. We denote by $R(\ell)$ the lattice in B whose local components $R(\ell)_p$ are given

- for p dividing ℓ, by the lattice $R_p^\vee \subseteq B_p$ dual to R_p, and
- otherwise, by R_p.

2.4.2. Reduced trace and norm

We denote by tr and det the reduced trace and reduced norm on B, and also on its completions. We use a superscripted 0, as in R^0 or $R(\ell)^0$, to denote the kernel of the reduced trace.

2.4.3. Coordinates tailored to K_∞

Define $B_\infty := B \otimes \mathbb{R}$. If B is indefinite, then $B_\infty \cong \mathrm{Mat}_{2\times 2}(\mathbb{R})$ is split; otherwise, B_∞ is isomorphic to the real Hamilton quaternions. The exponential series identifies B_∞^0 with the Lie algebra of $G(\mathbb{R})$. We write $\mathbf{i} \in B_\infty^0$ for the derivative at the identity of κ so that $\kappa(\theta) = \exp(\theta \mathbf{i})$. Then, $\mathbf{i}^2 = -1$. We may find $\mathbf{j} \in B_\infty^0$ with $\mathbf{j}^2 = \pm 1$ (+1 if B is indefinite, -1 if B is definite) so that $B_\infty = \mathbb{R}(\mathbf{i}) \oplus \mathbb{R}(\mathbf{i})\mathbf{j}$. We note that $\mathbf{j}$ is not uniquely determined, but any two choices differ by multiplication by a norm one element of $\mathbb{R}(\mathbf{i})$. We set $\mathbf{k} = \mathbf{ij}$. Then, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ give an $\mathbb{R}$-basis of B_∞^0. For real numbers a, b, c, we set $[a, b, c] := a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$. A general element of B_∞ may then be written $[a, b, c] + d$, where we identify the real number d with a scalar element of B_∞. In these coordinates,

$$\mathrm{tr}([a, b, c] + d) = 2d, \quad \det([a, b, c] + d) = a^2 \mp (b^2 + c^2) + d^2. \tag{2.8}$$

Example 2.4. Suppose that $B_\infty = \mathrm{Mat}_{2\times 2}(\mathbb{R})$ and that κ is as in Equation (2.2). Then, with suitable choices,

$$[a, b, c] + d = \begin{pmatrix} d + c & b + a \\ b - a & d - c \end{pmatrix}.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

2.4.4. Archimedean regions

For $T > 0$ and $\delta \in (0, 1]$, we denote by $\Omega(\delta, T)$ the set of all elements $[a, b, c] + d$ of B_∞ for which

$$a^2 + b^2 + c^2 + d^2 \leq T^2, \quad b^2 + c^2 \leq \delta T^2.$$

With $\Omega^\star(\delta, T)$, we denote the subset of *nonzero* elements of $\Omega(\delta, T)$. Likewise, for $T > 0$ and $\delta \in (0, 1]$, we let $\Psi(\delta, T)$ denote the set of all elements $[a, b, c] + d$ of B_∞ for which

$$a^2 + b^2 + c^2 + d^2 \leq T^2, \quad a^2 + d^2 \leq \delta T^2,$$

and $\Psi^\star(\delta, T)$ its subset consisting of *nonzero* elements.

2.5. Counting problems: results

We adopt the following asymptotic notation for counting estimates (compare with the notation $\preccurlyeq$ introduced in §2.3):

$$A_1 \prec A_2 \quad \Longleftrightarrow \quad A_1 \ll_\varepsilon (d_B N(1+T))^\varepsilon A_2.$$

Recall from §2.2 the height function H defined in the split case. In the nonsplit case, we adopt the convention in the following results that any terms involving H (in minima or sums) should be omitted.

Theorem 2.5 (Type I estimates). *Let $g \in G(\mathbb{R})$. Then, the first successive minima (see Definition 6.1) of $g^{-1}R(\ell)^0 g$ with respect to $\Omega(\delta, 1) \cap B_\infty^0$ is $\gg \min\left\{\ell^{-\frac{1}{2}}, \ell^{-1}\delta^{-\frac{1}{2}}H(g)^{-1}\right\}$. Furthermore, we have*

$$|g^{-1}R(\ell)^0 g \cap \Omega(\delta, T)| \prec 1 + \left(\ell^{\frac{1}{2}} + \ell\delta^{\frac{1}{2}}H(g)\right)T + \left(\frac{\ell^{\frac{3}{2}}\delta^{\frac{1}{2}}}{(d_B N)^{\frac{1}{2}}} + \ell\delta H(g)\right)T^2 + \frac{\ell^2\delta}{d_B N}T^3.$$

If B is nonsplit, we further have that the first successive minima of $g^{-1}R(\ell)^0 g$ with respect to $\Psi(\delta, 1) \cap B_\infty^0$ is at least $\gg \ell^{-\frac{1}{2}}$ and

$$|g^{-1}R(\ell)^0 g \cap \Psi(\delta, T)| \prec 1 + \ell^{\frac{1}{2}}T + \frac{\ell^{\frac{3}{2}}}{(d_B N)^{\frac{1}{2}}}T^2 + \frac{\ell^2\delta^{\frac{1}{2}}}{d_B N}T^3.$$

Theorem 2.6 (Type II estimates). *Let $g \in G(\mathbb{R})$ and $n \in \frac{1}{\ell}\mathbb{Z}$. We have*

$$|g^{-1}R(\ell)^0 g \cap \Omega(\delta, T) \cap \det^{-1}(\{n\})| \prec 1 + \ell\delta^{\frac{1}{2}}H(g)T + \frac{\ell^2}{d_B N}\delta T^2.$$

The proof of these results occupies §7 onwards. In §3, we explain how these results imply our main fourth moment bound, Theorem 2.1.

3. Division and reduction of the proof

3.1. Traversing the genus

Recall that K_R is defined as the image of the subgroup $\prod_p R_p^\times$ in $G(\mathbb{A}_f)$; it is a compact open subgroup of $G(\mathbb{A}_f)$. In due course, we will consider the conjugated sets $h_f K_R h_f^{-1}$, for $h_f \in G(\mathbb{A}_f)$. These are precisely the compact open subgroups $K_{R'}$ associated to the Eichler orders R' in the genus of R. We note that R' has the same level as R and may be given explicitly by the following intersection:

$$R' = h_f(R \otimes \widehat{\mathbb{Z}})h_f^{-1} \cap B(\mathbb{Q}),$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

where $\widehat{\mathbb{Z}}$ denotes the closure of $\mathbb{Z}$ inside $\mathbb{A}_f$. We further note that the action of $G(\mathbb{A}_f)$ on the genus of R commutes with partial dualization in the sense that

$$R'(\ell) = h_f\,(R(\ell) \otimes \widehat{\mathbb{Z}})h_f^{-1} \cap B(\mathbb{Q}).$$

This observation permits us to formulate the required L^2-estimates for our differences of theta kernels in terms of integration over Archimedean, rather than adelic, arguments. To that end, we introduce the notation

$$R(\ell; h) \coloneqq h_\infty^{-1} R'(\ell) h_\infty = h_\infty^{-1}(h_f\,(R(\ell) \otimes \widehat{\mathbb{Z}})h_f^{-1} \cap B(\mathbb{Q}))h_\infty$$

for $h = (h_\infty, h_f) \in G(\mathbb{A})$. We note that for $h \in G(\mathbb{R})$ (i.e., $h_f = 1$), the set $R(\ell; h)$ is just $h^{-1}R(\ell)h$. Since taking the trace commutes with conjugation, we may extend the notation to kernels of the reduced trace without concern for confusion regarding the order of operation, that is,

$$R(\ell; h)^0 = (h_\infty^{-1} R'(\ell) h_\infty)^0 = h_\infty^{-1} R'(\ell)^0 h_\infty = h_\infty^{-1}(h_f\,(R(\ell)^0 \otimes \widehat{\mathbb{Z}})h_f^{-1} \cap B(\mathbb{Q}))h_\infty.$$

If B is split, then the class number of R is one and we have fixed the representative as in Equation (2.1). In this case, we find for $h \in G(\mathbb{A})$ that $h^{-1}Rh = h'^{-1}Rh'$, where $h' \in G(\mathbb{R})$ has the same image under the isomorphism $[G]/K_\infty K_R \cong \Gamma_0(N)\backslash\mathbb{H}$ as h does. In particular, we have the equality of height functions (see §2.2) $H(h) = H(h')$.

Remark 3.1. By considering a right translate of $\varphi \in \mathcal{F}$ and thereby moving the maximal compact K_∞ and the Eichler order R around, one could reduce the statement of the main Corollary 2.3 to the case that g is the identity. However, in the split case, our counting arguments *do* depend on the particular order in the genus. Moreover, our method relies on a difference of theta kernels defined relative to different g. Such a reduction would thus be premature.

3.2. Estimating fourth moments via lattice sums

In §5, we introduce certain theta kernels. A spectral expansion of their L^2-norms will yield the fourth moments of interest, while a 'geometric' expansion, using Siegel domains and Fourier expansions, bounds those L^2-norms in terms of certain lattice sums. We now state the latter bounds.

Proposition 3.2. *Suppose B is indefinite. Then, for $g_1, g_2 \in [G]$, there exists $\ell | d_B N$, $g \in \{g_1, g_2\}$, and $0 < T \leqslant \frac{(d_B N(k+1))^{\frac{1}{2}}}{\ell}$ (here, the notation $\leqslant$ is as in §2.3) so that, for $k = 0$,*

$$\sum_{\varphi \in \mathcal{F}^-_{\leq L}} (|\varphi(g_1)|^2 - |\varphi(g_2)|^2)^2 \leqslant_L 1 + \frac{d_B N}{\ell T^2} \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \cap \Omega^\star(1, T): \\ \det(\gamma_1) = \det(\gamma_2)}} 1, \tag{3.1}$$

while for $k > 0$,

$$\sum_{\varphi \in \mathcal{F}^-, \mathrm{hol}} (|\varphi(g_1)|^2 - |\varphi(g_2)|^2)^2 \leqslant 1 + \frac{d_B N k}{\ell T^2} \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \cap \Omega^\star(1, T): \\ \det(\gamma_1) = \det(\gamma_2)}} 1. \tag{3.2}$$

Proposition 3.3. *Suppose B is indefinite. Let $g \in [G]$, and assume that $k \gg (d_B N)^\eta$ for some arbitrarily small $\eta > 0$. Then, there exists $\ell | d_B N$ and $0 < T \leqslant \frac{(d_B N k)^{\frac{1}{2}}}{\ell}$ so that*

$$\sum_{\varphi \in \mathcal{F}^-, \mathrm{hol}} |\varphi(g)|^4 \leqslant_\eta 1 + \frac{d_B N k}{\ell T^2} \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \cap \Omega^\star(k^{-1+\varepsilon}, T): \\ \det(\gamma_1) = \det(\gamma_2) > 0}} 1. \tag{3.3}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Proposition 3.4. *Suppose B is definite and the weight is $k = 0$. Then, for $g_1, g_2 \in [G]$ and $m \in \mathbb{N}_0$, there exists $\ell | d_B N$, $0 < T \leqslant \frac{(d_B N (m+1))^{\frac{1}{2}}}{\ell}$ and $\frac{1}{m^2+1} \leqslant \delta \leq 1$ so that*

$$\sum_{\varphi \in \mathcal{F}_m^+} \left(|\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right)^2 \leqslant 1 + \frac{d_B N}{\ell \delta^{\frac{1}{2}} T^2} \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \\ \gamma_1, \gamma_2 \in \Omega^\star(\delta, T) \cup \Psi^\star(\delta, T): \\ \det(\gamma_1) = \det(\gamma_2)}} 1. \tag{3.4}$$

Proposition 3.5. *Suppose B is definite. Then, for $g \in [G]$, there exists $\ell | d_B N$ and $0 < T \leqslant \frac{(d_B N k)^{\frac{1}{2}}}{\ell}$ so that*

$$\sum_{\varphi \in \mathcal{F}^{+,\mathrm{hol}}} |\varphi(g)|^4 \leqslant 1 + \frac{d_B N k}{\ell T^2} \left(\sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \cap \Omega^\star(k^{-1+\varepsilon}, T): \\ \det(\gamma_1) = \det(\gamma_2)}} 1 + \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \cap \Omega^\star(1, T): \\ \det(\gamma_1) = \det(\gamma_2)}} k^{-2027} \right). \tag{3.5}$$

3.3. Reduction to ternary lattices

In this section, we reduce the vital counting problem involving quaternary quadratic form to problems involving only ternary quadratic forms. The key observation is that we may orthogonally decompose the quaternion algebra B_∞ into its trace part and its traceless part B_∞^0. Thus, for any $\alpha = \frac{1}{2}\operatorname{tr}(\alpha) + \alpha^0 \in \mathbb{R} \oplus B_\infty^0$, we have

$$\det(\alpha) = \tfrac{1}{4}\operatorname{tr}(\alpha)^2 + \det(\alpha^0). \tag{3.6}$$

We further note that the trace is invariant under conjugation. Hence, we have $\operatorname{tr}(R(\ell; g)) \subseteq \mathbb{Z}$. We conclude that

$$R(\ell; g) \subseteq \tfrac{1}{2}\mathbb{Z} \oplus \tfrac{1}{2} R(\ell; g)^0 \tag{3.7}$$

is a sublattice of the direct sum of the lattices $\frac{1}{2}\mathbb{Z}$ in $\mathbb{R}$ and $\frac{1}{2}R(\ell; g)^0$ in B_∞^0. Using this decomposition, we deduce

$$\sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g) \cap \Omega^\star(\delta, T) \\ \det(\gamma_1) = \det(\gamma_2)}} 1 \ll_\varepsilon T^\varepsilon \left| R(\ell; g)^0 \cap \Omega(\delta, 2T) \right|^2 + T \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g)^0 \cap \Omega(\delta, 2T) \\ \det(\gamma_1) = \det(\gamma_2)}} 1 \tag{3.8}$$

by distinguishing the two cases of equal and nonequal trace and applying the divisor bound to the equality

$$\det(\gamma_1) = \det(\gamma_2) \Leftrightarrow \tfrac{1}{4}\operatorname{tr}(\gamma_1)^2 - \tfrac{1}{4}\operatorname{tr}(\gamma_2)^2 = \det(\gamma_2^0) - \det(\gamma_1^0).$$

We remark that we have forfeited the congruence condition $\det(\gamma_1^0) \equiv \det(\gamma_2^0) \bmod (1)$, and this forfeiture will be reflected in the suboptimality of our final counting estimates on larger scales when $\ell > 1$. We circumnavigate these larger scales by an appropriate choice of a covering domain (cf. Lemma 4.1).

Note that we may further bound the diagonal contribution by considering its largest fiber:

$$\sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g)^0 \cap \Omega(\delta, 2T) \\ \det(\gamma_1) = \det(\gamma_2)}} 1 \leq |R(\ell; g)^0 \cap \Omega(\delta, 2T)| \times \max_{\substack{n \in \frac{1}{\ell}\mathbb{Z} \\ |n| \leq 4T^2}} |R(\ell; g)^0 \cap \Omega(\delta, 2T) \cap \det^{-1}(\{n\})|. \tag{3.9}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Arguing along the same lines, we also arrive at

$$\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Psi^\star(\delta,T)\\ \det(\gamma_1)=\det(\gamma_2)}} 1 \ll_\varepsilon T^\varepsilon \left|R(\ell;g)^0\cap\Psi(\delta,2T)\right|^2 + \delta^{\frac12}T \sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)^0\cap\Psi(\delta,2T)\\ \det(\gamma_1)=\det(\gamma_2)}} 1 \tag{3.10}$$

and

$$\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)^0\cap\Psi(\delta,2T)\\ \det(\gamma_1)=\det(\gamma_2)}} 1 \le |R(\ell;g)^0\cap\Psi(\delta,2T)| \times \max_{\substack{n\in\frac{1}{\ell}\mathbb{Z}\\ |n|\le 4T^2}} |R(\ell;g)^0\cap\Omega(1,2T)\cap\det^{-1}(\{n\})|. \tag{3.11}$$

Note that in this last inequality, we have passed from $\Psi(\delta,2T)$ to the larger set $\Omega(1,2T)=\Psi(1,2T)$; the resulting bound remains adequate for us thanks to the additional saving of $\delta^{\frac12}$ in Equation (3.10).

3.4. Proof of Theorem 2.1

Theorem 2.1 is an immediate consequence of the following pair of lemmas together with Propositions 3.2 through 3.5.

Lemma 3.6. *We have*

$$\begin{aligned}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Omega^\star(\delta,T)\\ \det(\gamma_1)=\det(\gamma_2)}} 1 \prec \ell T^2\left(1+\ell^{\frac12}\delta^{\frac12}H(g)+\ell^{\frac12}\delta H(g)T+\frac{\ell^2}{d_BN}\delta T^2\right)\\ \times\left(1+\ell^{\frac12}\delta^{\frac12}H(g)+\ell^{\frac12}\delta^{\frac12}H(g)T+\frac{\ell^2}{d_BN}\delta T^2\right).\end{aligned} \tag{3.12}$$

Proof. Recall, from the discussion of Section §3.1, that we may express $R(\ell;g)^0$, for $g\in G(\mathbb{A})$, as $(g')^{-1}R'(\ell)^0g'$, where R' is an Eichler order of the same level and $g'\in G(\mathbb{R})$, with $H(g)=H(g')$ in the case that B is split. We may thus apply the results of Section §2.5.

Since $\mathrm{tr}(R(\ell;g))\subseteq\mathbb{Z}$, we find that the first successive minimum of $R(\ell;g)$ with respect to $\Omega(\delta,1)$ is at least the minimum of 1 and the first successive minimum of $R(\ell;g)^0$ with respect to $\Omega(\delta,1)\cap B_\infty^0$. The latter is $\gg\min\{\ell^{-\frac12},\ell^{-1}\delta^{-\frac12}H(g)^{-1}\}=:\Lambda$ by Theorem 2.5, where the term involving $H(g)$ is to be omitted if B is nonsplit. Thus, we find that $R(\ell;g)\cap\Omega^\star(\delta,T)$ is empty for $T\ll\Lambda$, in which case there is nothing to show. Next, assume instead that $T\gg\Lambda$. Then, by Theorem 2.5, we have

$$\begin{aligned}|g^{-1}R(\ell)^0g\cap\Omega(\delta,2T)| &\prec 1+\left(\ell^{\frac12}+\ell\delta^{\frac12}H(g)\right)T+\left(\frac{\ell^{\frac32}\delta^{\frac12}}{(d_BN)^{\frac12}}+\ell\delta H(g)\right)T^2+\frac{\ell^2}{d_BN}\delta T^3\\ &\prec \ell^{\frac12}T\left(1+\ell^{\frac12}\delta^{\frac12}H(g)+\frac{\ell}{(d_BN)^{\frac12}}\delta^{\frac12}T+\ell^{\frac12}\delta H(g)T+\frac{\ell^2}{d_BN}\delta T^2\right),\end{aligned} \tag{3.13}$$

where we have used $1\ll\ell^{\frac12}T+\ell\delta^{\frac12}H(g)T$ and $\ell^{\frac32}\le\ell^2$. Further note that the middle term in the bracket is dominated by the sum of the first and last term in the bracket. We also find by Theorem 2.6 that

$$\begin{aligned}\max_{n\in\frac{1}{\ell}\mathbb{Z}}|g^{-1}R(\ell)^0g\cap\Omega(\delta,2T)\cap\det^{-1}(\{n\})| &\prec 1+\ell\delta^{\frac12}H(g)T+\frac{\ell^2}{d_BN}\delta T^2\\ &\prec \ell^{\frac12}\left(1+\ell^{\frac12}\delta^{\frac12}H(g)T+\frac{\ell^2}{d_BN}\delta T^2\right).\end{aligned} \tag{3.14}$$

We conclude the Lemma by further appealing to the inequalities (3.8) and (3.9). □

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Lemma 3.7. *Assume that B is nonsplit. Then*

$$\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Psi^\star(\delta,T)\\ \det(\gamma_1)=\det(\gamma_2)}} 1 \prec \ell T^2\left(1+\frac{\ell}{(d_BN)^{\frac12}}T+\frac{\ell^2}{d_BN}\delta^{\frac12}T^2\right)^2. \tag{3.15}$$

Proof. As in the proof of Lemma 3.6, we find that the first successive minimum of $R(\ell;g)$ with respect to $\Psi(\delta,1)$ is at least the minimum of $\delta^{-\frac12}$ and the first successive minimum of $R(\ell;g)^0$ with respect to $\Psi(\delta,1)\cap B_\infty^0$. The latter is $\gg \ell^{-\frac12}$ by Theorem 2.5. Therefore, $R(\ell;g)\cap\Psi^\star(\delta,T)$ is empty for $T\ll \ell^{-\frac12}\le 1\le \delta^{-\frac12}$, in which case there is nothing to show. If $T\gg \ell^{-\frac12}$, then, by Theorem 2.5, we have

$$\begin{aligned}|g^{-1}R(\ell)^0g\cap\Psi(\delta,2T)| &\prec 1+\ell^{\frac12}T+\frac{\ell^{\frac32}}{(d_BN)^{\frac12}}T^2+\frac{\ell^2}{d_BN}\delta^{\frac12}T^3\\ &\prec \ell^{\frac12}T\left(1+\frac{\ell}{(d_BN)^{\frac12}}T+\frac{\ell^2}{d_BN}\delta^{\frac12}T^2\right),\end{aligned} \tag{3.16}$$

where we have used $1\ll \ell^{\frac12}T$ and $\ell^{\frac32}\le \ell^2$. Furthermore, by Theorem 2.6, we have

$$\max_{n\in\frac{1}{\ell}\mathbb{Z}}|g^{-1}R(\ell)^0g\cap\Omega(1,2T)\cap\det^{-1}(\{n\})| \prec 1+\frac{\ell^2}{d_BN}T^2 \prec \ell^{\frac12}\delta^{-\frac12}\left(1+\frac{\ell^2}{d_BN}\delta^{\frac12}T^2\right), \tag{3.17}$$

where we have used $1\le \ell^{\frac12}$ and $1\le\delta^{-\frac12}$. We conclude the lemma by further appealing to the inequalities (3.10) and (3.11). □

3.5. *Proof of Corollary 2.3*

Let $\varphi\in\mathcal{F}$, respectively $\mathcal{F}^{\mathrm{hol}}$, be L^2-normalized. Assume first that B is nonsplit. Then, since $[G]/K_\infty K_R$ is compact and equipped with a probability measure, we may find g_2 in $[G]$ such that $|\varphi(g_2)|\le 1$. Hence, Corollary 2.3 follows immediately from Theorem 2.1 by positivity and the particular choice of g_2.

We now turn our attention to the case that B is split, in other words when $d_B=1$. Here, we need to supplement Theorem 2.1 with the additional information that for $H(g)\ge N^{-\frac12}$, we have

$$|\varphi(g)| \preccurlyeq_{\lambda_\varphi} N^{\frac14} \qquad \text{if } \varphi\in\mathcal{F}^-, \tag{3.18}$$

$$|\varphi(g)| \preccurlyeq (kN)^{\frac14} \qquad \text{if } \varphi\in\mathcal{F}^{-,\mathrm{hol}}. \tag{3.19}$$

The former is recorded in [Tem15, Prop. 3.1 & 3.2], for example, and the latter may be deduced from the Fourier expansion along the lines of Xia [Xia07]. We include a brief proof here for the sake of completeness.

Lemma 3.8. *Assume B is split, and let $\varphi\in\mathcal{F}^{-,\mathrm{hol}}$ be an L^2-normalized holomorphic cuspidal newform of squarefree level N and even weight $k\ge 2$. Then, we have for all $g\in[G]$,*

$$|\varphi(g)| \preccurlyeq H(g)^{o(1)}\left(k^{\frac14}H(g)^{-\frac12}+k^{-\frac14}H(g)^{\frac12}\right). \tag{3.20}$$

If $H(g)\ge\frac{k}{2\pi}$, then we have the stronger bound

$$|\varphi(g)| \preccurlyeq 1. \tag{3.21}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Proof. Suppose that g corresponds to $z = x + iy \in \Gamma_0(N)\backslash\mathbb{H}$. As $|\varphi(g)|$ is further invariant under the Atkin–Lehner operators we may further assume that z has maximal imaginary part under the action of the group $A_0(N)$ generated by the Atkin–Lehner operators and $\Gamma_0(N)$, thus $H(g) = y$. We shall subsequently make use of the Fourier expansion of φ at ∞:

$$|\varphi(g)| = \left| y^{\frac{k}{2}} \sum_{n=1}^{\infty} a_n e(n(x+iy)) \right|.$$

We may bound the Fourier coefficients by appealing to Deligne's bound for the Hecke eigenvalues [Del71, Del74]. This implies $|a_n| \ll_\varepsilon n^{\frac{k-1}{2}+\varepsilon}|a_1|$.[1] We find

$$|\varphi(g)| \ll_\varepsilon |a_1|(2\pi)^{-\frac{k}{2}} y^{\frac{1}{2}-\varepsilon} \sum_{n=1}^{\infty} (2\pi n y)^{\frac{k-1}{2}+\varepsilon} e^{-2\pi n y}.$$

The above sum, we may bound by comparison to the corresponding integral. For this manner, we note that the function $x^\alpha e^{-x}$ increases up to $x = \alpha$ and then decreases. We may also bound the first Fourier coefficient a_1 by a result of Hoffstein–Lockhart [HL94] (cf. [HM06, Eq. (31)][2]). The bound reads $|a_1| \ll_\varepsilon (Nk)^\varepsilon (4\pi)^{\frac{k}{2}} \Gamma(k)^{-\frac{1}{2}}$. We thus arrive at

$$\begin{aligned} |\varphi(g)| &\ll_\varepsilon (Nk)^\varepsilon y^{-\varepsilon} \frac{2^{\frac{k}{2}} y^{\frac{1}{2}}}{\Gamma(k)^{\frac{1}{2}}} \left(\frac{1}{y} \Gamma\left(\frac{k+1}{2}\right) + \left(\frac{k-1}{2}\right)^{\frac{k-1}{2}+\varepsilon} e^{-\frac{k-1}{2}} \right) \\ &\ll_\varepsilon (Nk)^\varepsilon y^{-\varepsilon} \left(k^{\frac{1}{4}} y^{-\frac{1}{2}} + k^{-\frac{1}{4}} y^{\frac{1}{2}} \right), \end{aligned}$$

where we have made use of Stirling's approximation. If $y \geq \frac{k}{2\pi}$, then the maximum summand occurs when $n = 1$ and we may derive the improved bound

$$\begin{aligned} |\varphi(g)| &\ll_\varepsilon (Nk)^\varepsilon y^{-\varepsilon} \frac{2^{\frac{k}{2}} y^{\frac{1}{2}}}{\Gamma(k)^{\frac{1}{2}}} \left(\frac{1}{y} \Gamma\left(\frac{k+1}{2}\right) + (2\pi y)^{\frac{k-1}{2}+\varepsilon} e^{-2\pi y} \right) \\ &\ll_\varepsilon (Nk)^\varepsilon y^{-\varepsilon}. \end{aligned}$$

□

To deduce Equation (3.19) from the lemma, we consider separately the cases $N^{-\frac{1}{2}} \leq H(g) \leq \frac{k}{2\pi}$ and $H(g) \geq \frac{k}{2\pi}$, applying Equation (3.20) in the former case and Equation (3.21) in the latter.

We may now deduce the split case of Corollary 2.3, as follows. Our task is to bound $\varphi(g_1)$ suitably for $g_1 \in [G]$. We may assume that $H(g_1) \leq N^{-\frac{1}{2}}$, as otherwise the estimates (3.18) and (3.19) are adequate. In that case, we choose another point $g_2 \in [G]$ arbitrarily with $H(g_2) = N^{-\frac{1}{2}}$ such that $|\varphi(g_2)| \ll_{\lambda_\varphi} N^{\frac{1}{4}}$, respectively $|\varphi(g_2)| \ll (kN)^{\frac{1}{4}}$, by Equation (3.18), respectively Equation (3.19). We apply Theorem 2.1 with these choices of g_1 and g_2. Upon recalling that $d_B = 1$ in the split case, we find by positivity that Equation (2.4), respectively Equation (2.5), yield

$$\begin{aligned} &\left| |\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right| \ll_{\lambda_\varphi} N^{\frac{1}{2}}, \quad \text{respectively} \\ &\left| |\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right| \ll (kN)^{\frac{1}{2}}. \end{aligned}$$

We conclude by the triangle inequality and taking square roots:

$$|\varphi(g_1)|^2 \leq \left| |\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right| + |\varphi(g_2)|^2.$$

[1] At the ramified primes $p|N$ the stronger bound $|a_{p^l}| \leq (p^l)^{\frac{k}{2}-1}|a_1|$ holds and is a far less deep result (cf. [AL70, Thm. 3]).
[2] Notice the different normalization of the measure and a_1 in Section §2.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

4. Arithmetic quotients as real manifolds

4.1. Measure normalizations

For indefinite B, we fix an isomorphism $G(\mathbb{R}) \cong \mathrm{PGL}_2(\mathbb{R})$ sending K_∞ to $\mathrm{PSO}_2(\mathbb{R})$. We fix the Haar measure $\mathrm{d}g = \frac{\mathrm{d}y\,\mathrm{d}x}{y^2}\frac{\mathrm{d}\theta}{2\pi}$ for $g = \begin{pmatrix} y^{1/2} & xy^{-1/2} \\ 0 & y^{-1/2} \end{pmatrix}\kappa(\theta)$ on $\mathrm{SL}_2(\mathbb{R})$. The push-forward of this measure to the hyperbolic plane $\mathbb{H} \cong \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$ is then the measure $\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$. The Haar measure on $\mathrm{PGL}_2(\mathbb{R})$ is fixed so that its restriction to $\mathrm{PSL}_2(\mathbb{R})$ coincides with the push-forward of the Haar measure from $\mathrm{SL}_2(\mathbb{R})$.

If B is definite, we fix an isomorphism $G(\mathbb{R}) \cong \mathrm{SO}_3(\mathbb{R})$ sending K_∞ to $\mathrm{SO}_2(\mathbb{R})$. We fix a Haar measure on $\mathrm{SO}_3(\mathbb{R})$ so that the measure of the 2-sphere $S^2 \cong \mathrm{SO}_3(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$ is 4π.

4.2. Volumes

Recall, that we fixed the measure on $[G]$ to be the probability Haar measure. Hence, the volume of the quotient $[G]/K_R$ is 1. In due course, we shall also require the volume of said quotient when viewed as a real manifold with respect to our fixed Haar measure on $G(\mathbb{R})$. More specifically, we will need the volume with respect to the measure on $G'(\mathbb{R})$, where G' is the linear algebraic group defined over $\mathbb{Q}$ whose rational points are the proper unit quaternions B^1. There is an obvious isogeny map $G' \to G$, where G' is the simply connected form and G is the adjoint one. Define $R_p^1 = R_p \cap G'(\mathbb{Q}_p)$ to be the proper unit quaternions in the local order R_p, and set $K_R^1 = \prod_p R_p^1$. Then, the map $[G']/K_R^1 \to [G]/K_R$ is a homeomorphism that pushes forward the probability Haar measure on $[G']/K_R^1$ to the probability Haar measure on $[G]/K_R$; see Lemma A.2. In general, this map is not bijective if K_R is replaced by a general compact open subgroup of $G(\mathbb{A}_f)$ and the fact that the map is indeed a homeomorphism is due to K_R being the projectivized group of units of an Eichler order.

By Borel's finiteness of class numbers [Bor63], $[G']/K_R^1$ is a finite collection of $G'(\mathbb{R})$-orbits with representatives $\delta_1, \ldots, \delta_h \in G'(\mathbb{A})$. Define $\Gamma_i = G'(\mathbb{Q}) \cap \delta_i K_R^1 \delta_i^{-1}$; the intersection is taken in $G'(\mathbb{A}_f)$ but regarded as a subset of $G'(\mathbb{Q})$ and hence also of $G'(\mathbb{R})$. In particular, Γ_i is a lattice in $G'(\mathbb{R})$. It follows that

$$[G]/K_R \cong [G']/K_R^1 = \bigsqcup_{i=1}^{h} \Gamma_i \backslash G'(\mathbb{R}).$$

This is a finite disjoint union of finite-volume homogeneous spaces for the real Lie group $G'(\mathbb{R})$. We define $\mathrm{covol}(\Gamma_i)$ to be the measure of a fundamental domain for the action of Γ_i on $G'(\mathbb{R})$ with respect to the fixed Haar measure on $G'(\mathbb{R})$, either $\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}\frac{\mathrm{d}\theta}{2\pi}$ in the indefinite case or the measure giving volume 4π to $\mathrm{SO}_3(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$ in the definite case. We finally set

$$V = V_{d_B,N} = \sum_{i=1}^{h} \mathrm{covol}(\Gamma_i).$$

If B is indefinite, then $G'(\mathbb{R}) \cong \mathrm{SL}_2(\mathbb{R})$ is noncompact and strong approximation implies that $h = 1$ and we can write $[G]/K_R \cong \Gamma \backslash \mathrm{SL}_2(\mathbb{R})$. In this case, V is the volume of the hyperbolic surface $\Gamma\backslash\mathbb{H}$ with respect to the volume form $\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$. If B is definite, then in general h can be large and $[G]/K_\infty K_R$ is a finite collection of quotients of 2-spheres by discrete rotation groups.

Recall that we have denoted by d_B the reduced discriminant of B and by N the squarefree level of the Eichler order R. The volume is given in both cases by

$$V = V_{d_B,N} = \frac{\pi}{3} d_B N \prod_{p \mid d_B}\left(1 - \frac{1}{p}\right)\prod_{p \mid N}\left(1 + \frac{1}{p}\right) = (d_B N)^{1+o(1)}. \tag{4.1}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

This follows from a corresponding mass formula; see [Voi18, Thm 39.1.8] in the indefinite case and [Voi18, Thm 25.1.1 & Thm 25.3.18] in the definite one. The space is furthermore compact if and only if B nonsplit, that is, $d_B > 1$.

4.3. Siegel domains

The main purpose of this section is to provide a specific Siegel-domain covering in order to bound the L^2-norm of the (difference of) theta kernels in §5.2. Let M be a squarefree natural number and set $U = \left(\begin{smallmatrix} \widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \\ M\widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \end{smallmatrix}\right) \cap \mathrm{SL}_2(\mathbb{A}_f)$. The theta functions of interest will turn out to be right invariant in the symplectic variable under U with $M = d_B N$, but the present discussion applies to any squarefree M.

4.3.1. Cusps and Atkin–Lehner operators

Since M is squarefree, a representative set of cusps for $\Gamma_0(M)$ is given by the ratios $\frac{\ell}{M}$, where ℓ runs through the positive divisors of M. The width of the cusp $\frac{\ell}{M}$ (understood here as with respect to the group $\Gamma_0(M)$) is given by ℓ [Iwa97, §2.4]. For each $\ell | M$, we choose an element $\tau_\ell \in \mathrm{SL}_2(\mathbb{Z})$ satisfying

$$\tau_\ell \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \bmod \ell, \quad \tau_\ell \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod M/\ell. \tag{4.2}$$

Then, the cusp $\tau_\ell \infty$ is $\Gamma_0(M)$-equivalent to the cusp $\frac{\ell}{M}$. Hence, writing $n(x) = \left(\begin{smallmatrix} 1 & x \\ & 1 \end{smallmatrix}\right)$, we see that the elements $\tau_\ell n(j)$, where $j = 0, \dots, \ell - 1$ and $\ell | M$, give a complete system of representatives for $\Gamma_0(M) \backslash \mathrm{SL}_2(\mathbb{Z})$. The normalized matrices $\tilde{\tau}_\ell := \tau_\ell a(\ell)$, where $a(y) = \mathrm{diag}(y^{\frac{1}{2}}, y^{-\frac{1}{2}})$, are scaling matrices for the respective cusps. Furthermore, the matrices $\tilde{\tau}_\ell$ are Atkin–Lehner operators for $\Gamma_0(M)$ and give a set of representatives for $A_0(M)/\Gamma_0(M)$ [AL70, Lemma 9].

4.3.2. Coverings

The basic idea of the following lemma is to apply the Fricke involution to the tiling of $\Gamma_0(M) \backslash \mathbb{H}$ by translates of the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z}) \backslash \mathbb{H}$.

Lemma 4.1. *Let $F : [\mathrm{SL}_2] \to \mathbb{R}_{\geq 0}$ be a measurable function that is right U invariant and of weight 0. Then,*

$$\int_{[\mathrm{SL}_2]} F(g) dg \leq \frac{1}{V_{1,M}} \sum_{\ell | M} \int_{\frac{\sqrt{3}}{2}}^{\infty} \int_{\frac{l^2}{M}}^{\ell} F\left((\tau_\ell)_\infty \begin{pmatrix} y^{1/2} & x y^{-1/2} \\ & y^{-1/2} \end{pmatrix} K_\infty U) \right) dx \frac{dy}{y^2}.$$

Here, $(\tau)_\infty$ denotes the image of τ in the Archimedean coordinate of $\mathrm{SL}_2(\mathbb{A})$.

Proof. Let $f : \mathbb{H} \to \mathbb{R}_{\geq 0}$ be given by

$$f(z) = F\left(\begin{pmatrix} y^{1/2} & x y^{-1/2} \\ & y^{-1/2} \end{pmatrix} K_\infty U \right).$$

Then, f is $\Gamma_0(M)$ invariant on the left and we have

$$\int_{[\mathrm{SL}_2]} F(g) dg = \frac{1}{V_{1,M}} \int_{\Gamma_0(M) \backslash \mathbb{H}} f(z) \frac{dx dy}{y^2} = \frac{1}{V_{1,M}} \int_{\Gamma_0(M) \backslash \mathbb{H}} f(\tilde{\tau}_M z) \frac{dx dy}{y^2}.$$

The standard Siegel set $\{z \in \mathbb{H} |\ 0 \leq \Re(x) \leq 1 \text{ and } \Im(z) \geq \frac{\sqrt{3}}{2}\}$ contains a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$. Using that the $\tau_\ell n(j)$, for $j = 0, \dots, \ell - 1$ and $\ell | M$, form a representative set for $\Gamma_0(M) \backslash \mathrm{SL}_2(\mathbb{Z})$ and that f is nonnegative, we may bound

$$\int_{\Gamma_0(M) \backslash \mathbb{H}} f(\tilde{\tau}_M z) \frac{dx dy}{y^2} \leq \sum_{\ell | M} \int_{\frac{\sqrt{3}}{2}}^{\infty} \int_0^{\ell} f(\tilde{\tau}_M \tau_\ell z) dx \frac{dy}{y^2}.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Since $\tilde{\tau}_M \tilde{\tau}_\ell = \gamma \tilde{\tau}_{\frac{M}{\ell}}$ for some $\gamma \in \Gamma_0(M)$, we have the identity

$$f(\tilde{\tau}_M \tau_\ell z) = f(\tilde{\tau}_{\frac{M}{\ell}} a(\ell)^{-1} z) = f(\tau_{\frac{M}{\ell}} a(\tfrac{M}{\ell}) a(\ell)^{-1} z).$$

Substituting this identity above and applying the change of variables $\frac{M}{l^2} z \mapsto z$ gives the desired result. □

5. Theta kernels and their L^2-norms

5.1. Theta kernels and lifts

In this section, we summarize the required results on theta kernels and their lifts. The necessary theory is developed in Appendix A.

5.1.1. Theta functions

The theta kernels constructed in Appendix A are modular functions on $\mathrm{O}_{\mathrm{det}}(\mathbb{A}) \times \mathrm{SL}_2(\mathbb{A})$. The group G acts by conjugation on the quadratic space $(B, \det)$, preserving the quadratic form. This gives an embedding $G \hookrightarrow \mathrm{O}_{\mathrm{det}}$. We are mainly concerned here with the pullback of the theta kernels to $G(\mathbb{A}) \times \mathrm{SL}_2(\mathbb{A})$. We denote that pullback by $\Theta(g, s)$. The function $\Theta(g, s)$ is right $K_R \times U_R^1$ invariant, where $U_R^1 := \begin{pmatrix} \widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \\ d_B N \widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \end{pmatrix} \cap \mathrm{SL}_2(\mathbb{A}_f)$ and of moderate growth. We caution that it is *not* a theta kernel for the Howe dual pair of the orthogonal group of the traceless quaternions and SL_2.

We shall require several explicit expressions of the theta kernels Θ. Define functions P, u, X on B_∞ by setting, for $\gamma = [a, b, c] + d \in B_\infty$,

$$P(\gamma) := a^2 + b^2 + c^2 + d^2, \quad u(\gamma) := b^2 + c^2, \quad X(\gamma) := d + ia. \tag{5.1}$$

In other words, by identifying $\mathbf{i} \in B_\infty$ with $i \in \mathbb{C}$, we have that

- X is the projection from $B_\infty = \mathbb{C} \oplus \mathbb{C}\mathbf{j}$ to the summand $\mathbb{C}$,
- u is the squared magnitude of the projection onto the other summand $\mathbb{C}\mathbf{j}$ and
- P is the sum of the squared magnitudes of the two projections.

Upon recalling the notation $R(\ell; g)$ from Section §3.1, we define for $g \in G(\mathbb{A})$ and $z = x + iy \in \mathbb{H}$ the theta functions

$$\theta^-_{g,\ell}(z) := y^{1+\frac{k}{2}} \sum_{\gamma \in R(\ell; g)} X(\gamma)^k e^{-2\pi y P(\gamma)} e(x \det(\gamma)), \tag{5.2}$$

if $k \geq 6$

$$\theta^{-,\mathrm{hol}}_{g,\ell}(z) := \frac{k-1}{4\pi} y^{k/2} \sum_{\substack{\gamma \in R(\ell; g) \\ \det(\gamma) > 0}} \det(\gamma)^{k-1} \overline{X(\gamma)}^{-k} e(z \det(\gamma)), \tag{5.3}$$

$$\theta^{+,m}_{g,\ell}(z) := (2m+1) y^{1+m} \sum_{\gamma \in R(\ell; g)} \det(\gamma)^m P_m\left(\frac{|X(\gamma)|^2 - u(\gamma)}{\det(\gamma)}\right) e(z \det(\gamma)), \tag{5.4}$$

$$\theta^{+,\mathrm{hol}}_{g,\ell}(z) := (k+1) y^{1+\frac{k}{2}} \sum_{\gamma \in R(\ell; g)} X(\gamma)^k e(z \det(\gamma)), \tag{5.5}$$

where P_m is the m-th Legendre polynomial. When $\ell = 1$, we abbreviate by dropping the subscript, for example, $\theta^-_g := \theta^-_{g,1}$. We are now ready to express Θ by means of strong approximation. Set

$$s_\infty = \begin{pmatrix} y^{1/2} & x y^{-1/2} \\ & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Table 1. *Families and the choice of* Θ.

B	indefinite			definite	
$\mathcal{G}$	$\mathcal{F}^{-}$	$\mathcal{F}^{-,\mathrm{hol}}$		$\mathcal{F}_m^{+}$	$\mathcal{F}^{+,\mathrm{hol}}$
θ_g	θ_g^{-} $(k=0)$	θ_g^{-}	$\theta_g^{-,\mathrm{hol}}$	$\theta_g^{+,m}$	$\theta_g^{+,\mathrm{hol}}$
κ	0	k	k	$2m+2$	$k+2$

Then, $\Theta(g, s_\infty U_R^1) = \theta_g(z)e^{i\kappa\theta}$ for some $\kappa \in 2\mathbb{Z}$ and a choice of θ_g from Equations (5.2)–(5.5). The precise choice and value of κ depends on the family $\mathcal{G}$ under consideration and may be read off Table 1. (For our study of $\mathcal{F}^{-,\mathrm{hol}}$, the precise choice of Θ depends upon the size of k.)

Antipating the application of Lemma 4.1, we further require Fourier–Whittaker expansions of Θ at all of the cusps. They are expressable in terms of the $\theta_{g,\ell}$ from Equations (5.2)–(5.5) and a weight κ, the choice of which are given by the Table 1 as before. We have

$$\Theta(g, (\tau_\ell)_\infty s_\infty U_R^1) = \frac{\mu(\gcd(\ell, d_B))}{\ell}\theta_{g,\ell}(z)e^{i\kappa\theta}, \tag{5.6}$$

for $\ell | d_B N$ with μ the Möbius function, τ_ℓ as in Equation (4.2), and where $(\tau_\ell)_\infty$ denotes the image of τ_ℓ in the Archimedean coordinate of $\mathrm{SL}_2(\mathbb{A})$.

5.1.2. Jacquet–Langlands lifts

Set U_R to be the image of

$$\left\{ g \in \begin{pmatrix} \widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \\ d_B N\widehat{\mathbb{Z}} & \widehat{\mathbb{Z}} \end{pmatrix} : \det g \in \widehat{\mathbb{Z}}^\times \right\}$$

in $\mathrm{PGL}_2(\mathbb{A}_f)$. This is a compact open subgroup of $\mathrm{PGL}_2(\mathbb{A}_f)$. For each φ in the families $\mathcal{F}^{-}, \mathcal{F}^{-,\mathrm{hol}}, \mathcal{F}_m^{+}, \mathcal{F}^{+,\mathrm{hol}}$, we consider the Jacquet–Langlands transfer π^{JL} to $[\mathrm{PGL}_2]$ of the representation π generated by φ. In the case that G is split, we let $\pi^{\mathrm{JL}} = \pi$. The space of vectors in π^{JL} that are U_R-invariant and K_∞-isotypical of minimal nonnegative weight is one-dimensional [JL70, Cas73]. We define the *arithmetically normalized* Jacquet–Langlands lift $\varphi^{\mathrm{JL}} \in L^2([\mathrm{SL}_2])$ of φ to be the U_R^1-invariant restriction[3] to $[\mathrm{SL}_2]$ of a vector in this one-dimensional space, that has a Whittaker function at $\begin{pmatrix} y^{1/2} & xy^{-1/2} \\ & y^{-1/2} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \hookrightarrow \mathrm{SL}_2(\mathbb{A})$given by

- $2\sqrt{y}K_{it}(2\pi y)e(x)$ if $\varphi \in \mathcal{F}^{-}_{\frac{1}{4}+t^2}$, and
- $y^{\frac{\kappa}{2}}e(x+iy)$ if φ is in either of the families $\mathcal{F}^{-,\mathrm{hol}}, \mathcal{F}_m^{+}, \mathcal{F}^{+,\mathrm{hol}}$, where $\kappa = k, 2m+2, k+2$ depends on the family as before.

The bounds by Hoffstein–Lockhart [HL94] and Iwaniec [Iwa90] then imply the following bounds for the L^2-norm of the arithmetically normalized Jacquet–Langlands lift (cf. [HM06, (30), (31)][4]). One may also compare with the geometric normalization in [PY19, Thm. 6.1]. If B is indefinite and $\varphi \in \mathcal{F}^{-}_{\frac{1}{4}+t^2}$, we have

$$\|\varphi^{\mathrm{JL}}\|_2^2 = (d_B N(1+|t|))^{o(1)} \cosh(\pi t)^{-1}. \tag{5.7}$$

In the other cases, that is, when φ lies in either of the families $\mathcal{F}^{-,\mathrm{hol}}, \mathcal{F}_m^{+}$ or $\mathcal{F}^{+,\mathrm{hol}}$, we have

$$\|\varphi^{\mathrm{JL}}\|_2^2 = (d_B N\kappa)^{o(1)} \frac{\Gamma(\kappa)}{(4\pi)^\kappa}, \tag{5.8}$$

where κ depends on the family in accord with Table 1.

[3]It is more natural to consider φ^{JL} as a function on $[\mathrm{PGL}_2]$, but we allow ourselves to reduce to the restriction because the map $[\mathrm{SL}_2]/U_R^1 \to [\mathrm{PGL}_2]/U_R$ is a homeomorphism; see Appendix A.

[4]Notice the different normalization of the measure and a_1 in Section §2.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

5.1.3. Explicit theta lifting

The key identity is summarized in the following proposition.

Proposition 5.1. *Let $g \in [G]$. Let $\mathcal{G}$, Θ and κ according to Table 1. Then, for $\varphi \in \mathcal{G}$, we have*

$$\frac{\langle \Theta(g,\cdot), \varphi^{\mathrm{JL}} \rangle}{\langle \varphi^{\mathrm{JL}}, \varphi^{\mathrm{JL}} \rangle} = \frac{|\varphi(g)|^2}{V_{d_B,N}}. \tag{5.9}$$

Proof. The proof is carried out in the appendix. In short, Proposition A.16 implies that for $\varphi \in \mathcal{G}$, the theta lift φ_Φ of φ – defined in Equation (A.5), depending upon the precise family $\mathcal{G}$ – satisfies $\varphi_\Phi = (V_{d_B,N})^{-1}\varphi^{\mathrm{JL}}$. The claim then follows from Propositions A.15 and A.12. □

5.2. L^2-norms of theta kernels

5.2.1. Proofs of Propositions 3.2 through 3.5

The proofs are similar, so we discuss the first in detail and then explain the nonoverlapping parts of the rest. Recall the notation $\preccurlyeq$ from §2.3. We denote by $\Theta^-, \Theta^{-,\mathrm{hol}}, \Theta^{+,m}, \Theta^{+,\mathrm{hol}}$ the various functions 'Θ' defined as in §5.1.1.

Proof of Proposition 3.2. Let $\mathcal{G}$ denote either $\mathcal{F}^-_{\preccurlyeq L}$ or $\mathcal{F}^{-,\mathrm{hol}}$ according to whether $k = 0$ or $k \geq 2$. By Proposition 5.1, we may write

$$\frac{\|\varphi^{\mathrm{JL}}\|^2}{V_{d_B,N}} \left(|\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right) = \langle \Theta^-(g_1,\cdot) - \Theta^-(g_2,\cdot), \varphi^{\mathrm{JL}} \rangle.$$

By Bessel's inequality, it follows that

$$\sum_{\varphi \in \mathcal{G}} \frac{\|\varphi^{\mathrm{JL}}\|^2}{(V_{d_B,N})^2} \left(|\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right)^2 \leq \|\Theta^-(g_1,\cdot) - \Theta^-(g_2,\cdot)\|_2^2. \tag{5.10}$$

We now bound the right-hand side of Equation (5.10) (and in particular, verify that it is finite). Since $\Theta^-(g,\cdot)$ is K_∞-isotypical, Lemma 4.1 and Equation (5.6) give the bound

$$\ll \frac{1}{V_{1,d_B N}} \sum_{\ell | d_B N} \int_{\frac{\sqrt{3}}{2}\frac{\ell^2}{d_B N}}^{\infty} \int_0^\ell \frac{1}{\ell^2} |\theta^-_{g_1,\ell}(z) - \theta^-_{g_2,\ell}(z)|^2 dx \frac{dy}{y^2}. \tag{5.11}$$

We insert the definition (5.2) into the inner integral and evaluate, giving

$$\frac{1}{\ell^2}\int_0^\ell |\theta^-_{g_1,\ell}(z) - \theta^-_{g_2,\ell}(z)|^2 dx = \frac{1}{\ell} y^{2+k} \sum_{n \in \frac{1}{\ell}\mathbb{Z}} \left| \sum_{i=1}^{2} (-1)^i \sum_{\substack{\gamma \in R(\ell; g_i) \\ \det(\gamma) = n}} X(\gamma)^k e^{-2\pi y P(\gamma)} \right|^2.$$

Note that the sum over i kills the contribution from $\gamma = 0$, so we may omit that contribution in what follows. We separate the two sums by Cauchy–Schwarz and bound $X(\alpha)$ by $P(\alpha)^{\frac{1}{2}}$, giving

$$\frac{1}{\ell^2}\int_0^\ell |\theta^-_{g_1,\ell}(z) - \theta^-_{g_2,\ell}(z)|^2 dx \ll \sum_{i=1}^{2} \frac{1}{\ell} \sum_{n \in \frac{1}{\ell}\mathbb{Z}} \left| \sum_{\substack{0 \neq \gamma \in R(\ell; g_i) \\ \det(\gamma) = n}} P(\gamma)^{\frac{k}{2}} e^{-2\pi y P(\gamma)} \right|^2 y^{2+k}. \tag{5.12}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

We now treat the contributions from $i = 1, 2$ individually. We commence with the integral in the variable y. Let

$$Q(s,x) = \frac{1}{\Gamma(s)} \int_x^\infty t^s e^{-t} \frac{dt}{t} \leq 1$$

denote the normalized incomplete gamma function. Setting $g := g_i$, we find

$$\begin{aligned} &\frac{(4\pi)^{k+1}}{\Gamma(k+1)} \int_Y^\infty \left| \sum_{\substack{0 \neq \gamma \in R(\ell;g) \\ \det(\gamma)=n}} P(\gamma)^{\frac{k}{2}} e^{-2\pi y P(\gamma)} \right|^2 y^{2+k} \frac{dy}{y^2} \\ &\quad = \sum_{\substack{0 \neq \gamma_1, \gamma_2 \in R(\ell;g) \\ \det(\gamma_1)=\det(\gamma_2)=n}} \frac{2}{P(\gamma_1)+P(\gamma_2)} \left(\frac{2\sqrt{P(\gamma_1)P(\gamma_2)}}{P(\gamma_1)+P(\gamma_2)} \right)^k Q\left(k+1, 2\pi Y(P(\gamma_1)+P(\gamma_2))\right) \\ &\quad \leq \sum_{\substack{0 \neq \gamma_1, \gamma_2 \in R(\ell;g) \\ \det(\gamma_1)=\det(\gamma_2)=n}} \frac{2}{P(\gamma_1)+P(\gamma_2)} Q\left(k+1, 2\pi Y(P(\gamma_1)+P(\gamma_2))\right). \end{aligned} \tag{5.13}$$

Since $Q(s,x)$ is superpolynomially small in both s and x as soon as $x \gg s$, we see by dyadically partitioning $\max_i \{P(\gamma_i)^{\frac{1}{2}}\}$ that Equation (5.13) is further bounded by

$$\ll_A \sum_j \frac{1}{T_j^2} \left(1 + \frac{T_j^2 Y}{k+1}\right)^{-A} \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell;g) \cap \Omega^\star(1,T_j) \\ \det(\gamma_1)=\det(\gamma_2)=n}} 1$$

for any $A \geq 0$, where $T_j = 2^j$, $j \in \mathbb{Z}$. By putting all of these estimates together, we arrive at

$$\begin{aligned} &\frac{1}{(V_{d_B,N})^2} \sum_{\varphi \in \mathcal{G}} \|\varphi^{\mathrm{JL}}\|^2 \left(|\varphi(g_1)|^2 - |\varphi(g_2)|^2 \right)^2 \\ &\qquad \preccurlyeq_A \frac{\Gamma(k+1)(4\pi)^{-k}}{V_{1,d_B N}} \sum_{i=1}^2 \sum_{\ell \mid d_B N} \frac{1}{\ell} \sum_j \frac{1}{T_j^2} \left(1 + \frac{\ell^2}{d_B N} \frac{T_j^2}{k+1}\right)^{-A} \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell;g_i) \cap \Omega^\star(1,T_j) \\ \det(\gamma_1)=\det(\gamma_2)}} 1 \end{aligned}$$

for any $A \geq 0$. Let us recall from Equation (4.1) that $V_{d_B,N}, V_{1,d_B N} = (d_B N)^{1+o(1)}$ and that for $\varphi \in \mathcal{F}^-_{\leq L}$ we have $\|\varphi^{\mathrm{JL}}\| \gg_L 1$ (see Equation (5.7)). In order to conclude the first part of the proposition, we note that the range of T may be limited from above to $\preccurlyeq (d_B N)^{\frac{1}{2}} \ell^{-1}$ by the superpolynomial decay and *any* polynomial bound on the second moment matrix count, which was noted in §3.4 and is the subject of the remaining sections §6 through §10. The second part of the proposition follows along the same lines, but we need to use the bound $\|\varphi^{\mathrm{JL}}\|^2 \gg \Gamma(k)(4\pi)^{-k}$ for $\varphi \in \mathcal{F}^{-,\mathrm{hol}}$ instead (see Equation (5.8)). □

Proof of Proposition 3.3. We follow the recipe of the previous proof, only this time for the family $\mathcal{F}^{-,\mathrm{hol}}$ to which the theta function $\Theta^{-,\mathrm{hol}}$ corresponds. As we shall see, the latter already possesses a finite L^2-norm. Hence, we need not consider a difference of theta functions. After the initial steps, we arrive at

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$$\sum_{\varphi\in\mathcal{F}^{-,\mathrm{hol}}}\frac{\|\varphi^{\mathrm{JL}}\|^2}{(V_{d_B,N})^2}|\varphi(g)|^4\ll\frac{1}{V_{1,d_BN}}\frac{k^2\Gamma(k-1)}{(4\pi)^k}$$
$$\times\sum_{\ell|d_BN}\frac{1}{\ell}\sum_{n\in\frac{1}{\ell}\mathbb{N}}\frac{1}{n}\left|\sum_{\substack{\gamma\in R(\ell;g)\\ \det(\gamma)=n}}\left(\frac{\det(\gamma)^{\frac12}}{\overline{X(\gamma)}}\right)^k\right|^2 Q(k-1,2\sqrt{3}\pi n\tfrac{\ell^2}{d_BN}). \quad (5.14)$$

We further simplify using the lower bound $\|\varphi^{\mathrm{JL}}\|^2\geqslant\Gamma(k)(4\pi)^{-k}$ (see Equation (5.8)), the approximations $V_{d_B,N},V_{1,d_BN}=(d_BN)^{1+o(1)}$ and the superpolynomial decay of normalized incomplete gamma function, as well as the identities

$$|X(\gamma)|^2=\det(\gamma)+u(\gamma)\quad\text{and}\quad 2u(\gamma)+\det(\gamma)=P(\gamma).$$

We obtain

$$\frac{1}{d_BNk}\sum_{\varphi\in\mathcal{F}^{-,\mathrm{hol}}}|\varphi(g)|^4\preccurlyeq_A\sum_{\ell|d_BN}\frac{1}{\ell}\sum_{n\in\frac{1}{\ell}\mathbb{N}}\frac{1}{n}\left(1+\frac{\ell^2}{d_BN}\frac{n}{k}\right)^{-A}\left|\sum_{\substack{\gamma\in R(\ell;g)\\ \det(\gamma)=n}}\left(1+\frac{u(\gamma)}{n}\right)^{-\frac{k}{2}}\right|^2, \quad (5.15)$$

for any $A\geq 0$. By the triangle inequality, we reduce to estimating similar expressions but with the sum over γ restricted by one of the following conditions:

(i) $u(\gamma)\leq k^{-1+\varepsilon}\det(\gamma)$,
(ii) $k^{-1+\varepsilon}\det(\gamma)\leq u(\gamma)\leq\det(\gamma)$ or
(iii) $\det(\gamma)\leq u(\gamma)$.

In Case (i), we bound $(1+u(\gamma)/n)^{-\frac{k}{2}}\leq 1$. Furthermore, we have $\det(\gamma)\asymp P(\gamma)$ and $u(\gamma)\ll k^{-1+\epsilon}P(\gamma)$. Hence, after dyadically partitioning the range of $P(\gamma)^{\frac12}$, we arrive at

$$\preccurlyeq_A\sum_{\ell|d_BN}\frac{1}{\ell}\sum_j\frac{1}{T_j^2}\left(1+\frac{\ell^2}{d_BN}\frac{T_j^2}{k}\right)^{-A}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Omega^\star(k^{-1+\varepsilon},T_j)\\ \det(\gamma_1)=\det(\gamma_2)>0}}1.$$

In Case (ii), we use that $(1+u(\gamma)/n)^{-\frac{k}{2}}\leq(1+k^{\varepsilon-1})^{-\frac{k}{2}}$ has superpolynomial decay in k (and hence also in (d_BN) as $k\gg_\eta(d_BN)^\eta$ by assumption). As in Case (i), we have $\det(\gamma)\asymp P(\gamma)$, but this time only $u(\gamma)\leq P(\gamma)$. We arrive at a contribution of

$$\preccurlyeq_{A,\eta}(kd_BN)^{-A}\sum_{\ell|d_BN}\frac{1}{\ell}\sum_j\frac{1}{T_j^2}\left(1+\frac{\ell^2}{d_BN}\frac{T_j^2}{k}\right)^{-A}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Omega^\star(1,T_j)\\ \det(\gamma_1)=\det(\gamma_2)>0}}1.$$

In Case (iii), we bound

$$\left(1+\frac{u(\gamma)}{n}\right)^{-\frac{k}{2}}\leq 2^{-\frac{k}{4}}\left(1+\frac{u(\gamma)}{n}\right)^{-\frac{k}{4}}.$$

The factor $2^{-\frac{k}{4}}$ we use for superpolynomial decay in (kd_BN) as before. The other factor we use as follows

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$$\frac{1}{n}\left(1+\frac{u(\gamma)}{n}\right)^{-1}\left(1+\frac{\ell^2}{d_BN}\frac{n}{k}\right)^{-\frac{k}{4}+1}\left(1+\frac{u(\gamma)}{n}\right)^{-\frac{k}{4}+1}\leq\frac{1}{u(\gamma)}\left(1+\frac{\ell^2}{d_BN}\frac{u(\gamma)}{k}\right)^{-\frac{k}{4}+1}.$$

Hence, Equation (5.15) is bounded by

$$\begin{aligned}&\ll_{A,\eta}(kd_BN)^{-A}\sum_{\ell\mid d_BN}\frac{1}{\ell}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\\ \det(\gamma_1)=\det(\gamma_2)}}\frac{1}{u(\gamma_1)+u(\gamma_2)}\left(1+\frac{\ell^2}{d_BN}\frac{u(\gamma_1)+u(\gamma_2)}{k}\right)^{-\frac{k}{4}+1}\\&\ll_{A,\eta}(kd_BN)^{-A}\sum_{\ell\mid d_BN}\frac{1}{\ell}\sum_j\frac{1}{T_j^2}\left(1+\frac{\ell^2}{d_BN}\frac{T_j^2}{k}\right)^{-\frac{k}{4}+1}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Omega^\star(1,T_j)\\ \det(\gamma_1)=\det(\gamma_2)>0}}1,\end{aligned}$$

where we have dyadically partitioned $\max_i\{P(\gamma_i)^{\frac{1}{2}}\}\asymp\sqrt{u(\gamma_1)+u(\gamma_2)}$. The proof of the proposition is now concluded as in the previous case. □

Proof of Proposition 3.4. We treat the definite spherical case in the same spirit as the indefinite spherical case. We readily arrive at the estimate

$$\begin{aligned}\sum_{\varphi\in\mathcal{F}_m^+}\frac{\|\varphi^{\mathrm{JL}}\|^2}{(V_{d_B,N})^2}\left(|\varphi(g_1)|^2-|\varphi(g_2)|^2\right)^2&\ll\frac{1}{V_{1,d_BN}}\frac{(2m+1)^2\Gamma(2m+1)}{(4\pi)^{2m+2}}\\&\times\sum_{i=1}^2\sum_{\ell\mid d_BN}\frac{1}{\ell}\sum_{n\in\frac{1}{\ell}\mathbb{N}}\frac{1}{n}\left|\sum_{\substack{\gamma\in R(\ell;g_i)\\ \det(\gamma)=n}}P_m\left(\frac{|X(\gamma)|^2-u(\gamma)}{n}\right)\right|^2Q(2m+1,2\sqrt{3}\pi n\tfrac{\ell^2}{d_BN}).\end{aligned}\tag{5.16}$$

We simplify the inequality by using the lower bound $\|\varphi^{\mathrm{JL}}\|^2\geqslant\Gamma(2m+2)(4\pi)^{-2m-2}$ (see Equation (5.8)), the approximations $V_{d_B,N},V_{1,d_BN}=(d_BN)^{1+o(1)}$ and the superpolynomial decay of the normalized incomplete Gamma function Q. We obtain

$$\begin{aligned}&\frac{1}{d_BN(m+1)}\sum_{\varphi\in\mathcal{F}_m^+}\left(|\varphi(g_1)|^2-|\varphi(g_2)|^2\right)^2\\&\ll_A\sum_{i=1}^2\sum_{\ell\mid d_BN}\frac{1}{\ell}\sum_{n\in\frac{1}{\ell}\mathbb{N}}\frac{1}{n}\left(1+\frac{\ell^2}{d_BN}\frac{n}{m+1}\right)^{-A}\left|\sum_{\substack{\gamma\in R(\ell;g_i)\\ \det(\gamma)=n}}P_m\left(\frac{|X(\gamma)|^2-u(\gamma)}{n}\right)\right|^2,\end{aligned}\tag{5.17}$$

for any $A\geq0$. We proceed further by appealing to the Bernstein inequality [Ber31] for the Legendre polynomials:

$$P_m(t)\leq\min\left\{1,\sqrt{\frac{2}{\pi m}}\frac{1}{(1-t^2)^{\frac{1}{4}}}\right\},\quad\text{for }|t|\leq1.\tag{5.18}$$

We recall that $\det(\gamma)=|X(\gamma)|^2+u(\gamma)$ so that, with γ and n as in the above sum,

$$t:=\frac{|X(\gamma)|^2-u(\gamma)}{n}=\frac{|X(\gamma)|^2-u(\gamma)}{|X(\gamma)|^2+u(\gamma)},\quad1-t^2=\frac{4|X(\gamma)|^2\cdot u(\gamma)}{n^2}\geq0.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Dyadically partitioning $P(\gamma)^{\frac{1}{2}} = \det(\gamma)^{\frac{1}{2}}$, we conclude that Equation (5.17) is bounded by

$$\ll_A \sum_{i=1}^{2} \sum_{\ell | d_B N} \frac{1}{\ell} \sum_j \frac{1}{T_j^2} \left(1 + \frac{\ell^2}{d_B N} \frac{T_j^2}{m+1}\right)^{-A} \times \sum_{T_j^2 \leq n < 4T_j^2} \left(\sum_{\substack{\gamma \in R(\ell; g_i) \\ \gamma \in \Omega^\star(1,2T_j) - \Omega^\star(1,T_j) \\ \det(\gamma) = n}} \min\left\{1, \frac{1}{(m+1)^{\frac{1}{2}}} \frac{T_j}{(|X(\gamma)|^2 \cdot u(\gamma))^{\frac{1}{4}}}\right\} \right)^2, \tag{5.19}$$

where $T_j = 2^j$ as before. The minimum in Equation (5.19) lies between $\asymp (m+1)^{-\frac{1}{2}}$ and 1. Let us consider the $\gamma \in \Omega^\star(1, 2T_j) - \Omega^\star(1, T_j)$ for which

$$\min\left\{1, \frac{1}{(m+1)^{\frac{1}{2}}} \frac{T_j}{(|X(\gamma)|^2 \cdot u(\gamma))^{\frac{1}{4}}}\right\} \asymp \frac{1}{(m+1)^{\frac{1}{2}}} \frac{1}{\delta^{\frac{1}{4}}}, \tag{5.20}$$

for some given δ with $1/(m+1)^2 \ll \delta \leq 1$. In particular, $|X(\gamma)|^2 \cdot u(\gamma) \ll \delta T_j^4$. Since $|X(\gamma)|^2 + u(\gamma) = P(\gamma) \asymp T_j^2$, both cannot be simultaneously small. Hence,

$$\min\{|X(\gamma)|^2, u(\gamma)\} = \frac{|X(\gamma)|^2 \cdot u(\gamma)}{\max\{|X(\gamma)|^2, u(\gamma)\}} \ll \delta T_j^2.$$

Thus, after replacing δ with its multiple by a scalar of the form $\asymp 1$ if needed (which has no affect on Equation (5.20)), we may assume that $\min\{|X(\gamma)|^2, u(\gamma)\} \leq \delta(2T_j)^2$, that is, that γ lies in either $\Omega^\star(\delta, 2T_j)$ or $\Psi^\star(\delta, 2T_j)$. We now consider dyadic scales δ_a of δ's between $\asymp 1/(m^2+1)$ and 1. The just mentioned arguments then allow us to bound second line in Equation (5.19) by

$$\sum_{T_j^2 \leq n < 4T_j^2} \left(\sum_a \sum_{\substack{\gamma \in R(\ell; g_i) \\ \gamma \in \Omega^\star(\delta_a, 2T_j) \cup \Psi^\star(\delta_a, 2T_j) \\ \det(\gamma) = n}} (m+1)^{-\frac{1}{2}} \delta_a^{-\frac{1}{4}} \right)^2.$$

There are at most $\ll 1$ dyadic scales in the range $1/(m^2+1) \ll \delta \leq 1$. Thus, after applying Cauchy–Schwarz in order to pull out the sum over δ_a, we bound Equation (5.19) by

$$\ll_A \sum_{i=1}^{2} \sum_{\ell | d_B N} \frac{1}{\ell} \sum_j \frac{1}{T_j^2} \left(1 + \frac{\ell^2}{d_B N} \frac{T_j^2}{m+1}\right)^{-A} \sum_a \sum_{\substack{\gamma_1, \gamma_2 \in R(\ell; g_i) \\ \gamma_1, \gamma_2 \in \Omega^\star(\delta_a, 2T_j) \cup \Psi^\star(\delta_a, 2T_j)) \\ \det(\gamma_1) = \det(\gamma_2)}} (m+1)^{-1} \delta_a^{-\frac{1}{2}}. \tag{5.21}$$

The proof is once more concluded as it was for the first proposition. □

Proof of Proposition 3.5. One final time we iterate the initial steps for the holomorphic family $\mathcal{F}^{\mathrm{hol}}$ in the definite case. We are again in the situation where the theta kernel $\Theta^{+,\mathrm{hol}}$ has finite L^2-norm, so we obtain, without having to take differences,

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$$\sum_{\varphi\in\mathcal{F}^{+,\mathrm{hol}}}\frac{\|\varphi^{\mathrm{JL}}\|^2}{(V_{d_B,N})^2}|\varphi(g)|^4 \ll \frac{1}{V_{1,d_BN}}\frac{(k+1)^2\Gamma(k+1)}{(4\pi)^{k+2}}$$
$$\times\sum_{\ell|d_BN}\frac{1}{\ell}\sum_{n\in\frac{1}{\ell}\mathbb{N}}\frac{1}{n}\left|\sum_{\substack{\gamma\in R(\ell;g)\\ \det(\gamma)=n}}\left(\frac{X(\gamma)^2}{n}\right)^{\frac{k}{2}}\right|^2 Q(k+1,2\sqrt{3}\pi n\tfrac{\ell^2}{d_BN}). \tag{5.22}$$

We simplify this estimate using the lower bound $\|\varphi^{\mathrm{JL}}\|^2 \geqslant \Gamma(k+2)(4\pi)^{-k-2}$ (see Equation (5.8)), the approximations $V_{d_B,N}, V_{1,d_BN} = (d_BN)^{1+o(1)}$, the superpolynomial decay of the normalized incomplete gamma function Q and the identity $|X(\gamma)|^2 = \det(\gamma) - u(\gamma)$. We thereby obtain

$$\frac{1}{d_BNk}\sum_{\varphi\in\mathcal{F}^{+,\mathrm{hol}}}|\varphi(g)|^4 \leqslant_A \sum_{\ell|d_BN}\frac{1}{\ell}\sum_{n\in\frac{1}{\ell}\mathbb{N}}\frac{1}{n}\left(1+\frac{\ell^2}{d_BN}\frac{n}{k}\right)^{-A}\left|\sum_{\substack{\gamma\in R(\ell;g)\\ \det(\gamma)=n}}\left(1-\frac{u(\gamma)}{n}\right)^{\frac{k}{2}}\right|^2, \tag{5.23}$$

for any $A \geq 0$. We dyadically partition $P(\gamma)^{\frac{1}{2}} = \det(\gamma)^{\frac{1}{2}}$ and distinguish the two cases:

(i) $u(\gamma) \leq k^{-1+\varepsilon}\det(\gamma)$, and
(ii) $k^{-1+\varepsilon}\det(\gamma) \leq u(\gamma)$.

We separate them by the triangle inequality. Using the inequality $(1-u(\gamma)/n) \leq 1$, we see that the contribution of the first case is bounded by

$$\leqslant_A \sum_{\ell|d_BN}\frac{1}{\ell}\sum_j\frac{1}{T_j^2}\left(1+\frac{\ell^2}{d_BN}\frac{T_j^2}{k}\right)^{-A}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Omega^\star(k^{-1+\varepsilon},T_j)\\ \det(\gamma_1)=\det(\gamma_2)}}1,$$

where $T_j = 2^j$, $j\in\mathbb{Z}$. In the second case, we see that $(1-k^{\varepsilon-1})^{\frac{k}{2}}$ enjoys superpolynomial decay in k. The contribution of the second case is thus bounded by

$$\leqslant_A k^{-A}\sum_{\ell|d_BN}\frac{1}{\ell}\sum_j\frac{1}{T_j^2}\left(1+\frac{\ell^2}{d_BN}\frac{T_j^2}{k}\right)^{-A}\sum_{\substack{\gamma_1,\gamma_2\in R(\ell;g)\cap\Omega^\star(1,T_j)\\ \det(\gamma_1)=\det(\gamma_2)}}1.$$

The proof is now concluded as it was for the previous propositions. □

6. Preliminaries on the geometry of numbers

6.1. Bounds on successive minima

Definition 6.1. Let V be an n-dimensional real vector space. Let $L \subseteq V$ be a lattice (i.e., a cocompact discrete subgroup). Given a compact convex 0-symmetric subset $\mathcal{K}$ of V with nonempty interior, we define a function $N : V \to \mathbb{R}_{\geq 0}$ by $N(v) := \inf\{t > 0 : v \in t\mathcal{K}\}$. Given a positive-definite quadratic form Q on V, we define such a function by $N(v) := Q(v)^{1/2}$, or equivalently, by applying the previous definition with $\mathcal{K}$ the unit ball for Q. In either case, we define the *successive minima* $\lambda_1 \leq \cdots \leq \lambda_n$ of $\mathcal{K}$ on L (or of Q on L) as: λ_k is the smallest positive real for which there is a linearly independent subset $\{v_1,\ldots,v_k\}$ of L for which $N(v_j) \leq \lambda_k$ for each $1 \leq j \leq k$.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Lemma 6.2. *Let $z \in \mathbb{H}$ with maximal imaginary part under the orbit of the Atkin–Lehner operators $A_0(N)$ of $\Gamma_0(N)$ with N squarefree. Then, we have*

$$\Im(z) \geq \frac{\sqrt{3}}{2N} \quad \text{and} \quad |cz+d|^2 \geq \frac{(c,N)}{N}$$

for any $(c,d) \in \mathbb{Z}^2$ distinct from $(0,0)$.

Proof. This is essentially [HT12, Lemma 1]. That reference gives the slightly weaker bound obtained by omitting the factor (c,N), but the stronger bound that we have stated follows from their proof, keeping track of (c,N) at each step rather than bounding it from below by 1. □

6.2. Lattice counting

Lemma 6.3. *Let $f_{\mathcal{K}}$ be the distance function of a closed convex 0-symmetric set $\mathcal{K} \subseteq \mathbb{R}^n$ of positive volume. Let $\Lambda \subset \mathbb{R}^n$ be a lattice, and let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the successive minima (see Definition 6.1) of $\mathcal{K}$ on Λ. Then, there is a basis $v_1, \ldots, v_n$ of Λ such that $f_{\mathcal{K}}(v_i) \asymp_n \lambda_i$.*

Proof. This is [GL87, Thm. 2, p. 66]. □

Lemma 6.4. *Let $\mathcal{K} \subseteq \mathbb{R}^n$ be a closed convex 0-symmetric set of positive volume. Let $\Lambda \subset \mathbb{R}^n$ be a lattice, and let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ denote the successive minima of $\mathcal{K}$ on Λ. Then*

$$|\mathcal{K} \cap \Lambda| \asymp_n \prod_{i=1}^{n} \left(1 + \frac{1}{\lambda_i}\right).$$

Proof. The lower bound follows from van der Corput's generalization of Minkowski's first theorem [vdC36]. It states that for $\mathcal{K}' \subset \mathbb{R}^d$ a closed convex 0-symmetric set and $\Lambda \subset \mathbb{R}^d$ a lattice, one has

$$|\mathcal{K}' \cap \Lambda'| + 1 \geq |\operatorname{int} \mathcal{K}' \cap \Lambda'| + 1 \geq 2^{1-d} \frac{\operatorname{vol}(\mathcal{K}')}{\operatorname{vol}(\mathbb{R}^d/\Lambda')}. \tag{6.1}$$

Let d be the largest integer such that $\lambda_d \leq 1$. Let $v_i \in \Lambda$, for $i = 1, \ldots, d$, be a set of linearly independent vectors such that $\lambda_i^{-1} v_i \in \mathcal{K}$. Let $\mathcal{K}'$ be the convex hull of the vectors $\pm\lambda_i^{-1} v_i$ and Λ' the span of the vectors v_i. In particular, $\mathcal{K}'$ is nonempty, hence $0 \in \mathcal{K}' \cap \Lambda'$, and so

$$2|\mathcal{K}' \cap \Lambda'| \geq |\mathcal{K}' \cap \Lambda'| + 1.$$

Using Equation (6.1), it follows now that

$$|\mathcal{K} \cap \Lambda| \geq |\mathcal{K}' \cap \Lambda'| \geq 2^{-d} \frac{\operatorname{vol}(\mathcal{K}')}{\operatorname{vol}(\mathbb{R}^d/\Lambda')} = \frac{1}{d!} \prod_{i=1}^{d} \frac{1}{\lambda_i}.$$

For the upper bound, we refer to [BHW93, Prop. 2.1]. □

Lemma 6.5. *Let $\Lambda \subset \mathbb{R}^2$ be a lattice of rank 2 and $B \subseteq \mathbb{R}^2$ a ball of radius R (not necessarily centred at 0). If $\lambda_1 \leq \lambda_2$ are the successive minima of Λ, then*

$$|B \cap \Lambda| \ll 1 + \frac{R}{\lambda_1} + \frac{R^2}{\lambda_1 \lambda_2}.$$

Proof. See [HT13, Lemma 2.1]. □

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

7. Local preliminaries on orders

7.1. Quadratic preliminaries

Let F be a non-Archimedean local field of characteristic $\neq 2$. Let E/F be a separable quadratic extension, thus E is either the split quadratic extension $F \oplus F$ or a quadratic field extension. We write $\mathfrak{o}$ (resp. $\mathfrak{o}_E$) for the ring of integers in F (resp. E), $x \mapsto \bar{x}$ for the canonical involution on E and

$$\mathrm{nr}(x) = x\bar{x}, \quad \mathrm{tr}(x) = x + \bar{x}$$

for the norm and trace. Recall that the different ideal $\mathfrak{d}$ for this extension is the smallest $\mathfrak{o}_E$-ideal for which $\mathrm{tr}(\mathfrak{d}^{-1}) \subseteq \mathfrak{o}$, and in fact $\mathfrak{d}^{-1} = \{x \in E : \mathrm{tr}(x\mathfrak{o}_E) \subseteq \mathfrak{o}\}$. If E/F is split or unramified, then $\mathfrak{d} = \mathfrak{o}_E$.

We may regard E as a two-dimensional vector space over F.

Let $q : E \to F$ be a nondegenerate binary quadratic form with the property that for all $e, x \in E$, we have $q(ex) = \mathrm{nr}(e)q(x)$. In other words, q is an F-multiple of nr, specifically $q = q(1)\,\mathrm{nr}$.

For $x, y \in E$, we set $\langle x, y\rangle := q(x+y) - q(x) - q(y) = q(1)\,\mathrm{tr}(x\bar{y})$ so that $q(x) = \langle x, x\rangle/2$.

Let $\mathfrak{a} \subset E$ be a fractional $\mathfrak{o}_E$-ideal. Write $\mathfrak{a}^\vee$ for the dual of $\mathfrak{a}$ with respect to the quadratic form q, that is, $\mathfrak{a}^\vee := \{x \in E : \langle x, \mathfrak{a}\rangle \subseteq \mathfrak{o}\}$.

Let $\mathfrak{n}$ denote the fractional $\mathfrak{o}$-ideal generated by $q(\mathfrak{a})$.

Lemma 7.1. *We have* $\mathfrak{a} = \mathfrak{d}\mathfrak{n}\mathfrak{a}^\vee$.

Proof. Let α be a generator of $\mathfrak{a}$. Then $\mathfrak{n} = q(1)\,\mathrm{nr}(\mathfrak{a}) = \mathfrak{o}q(1)\alpha\bar{\alpha}$, $\mathfrak{a}^\vee = \{q(1)^{-1}x : x \in E, \mathrm{tr}(x\bar{\mathfrak{a}}) \subseteq \mathfrak{o}\} = q(1)^{-1}\bar{\alpha}^{-1}\mathfrak{d}^{-1}$. Multiplying through, the conclusion follows. □

Corollary 7.2. *Suppose that E/F is unramified and that q is integral on $\mathfrak{a}$ so that $\mathfrak{a} \subseteq \mathfrak{a}^\vee$. Then the elementary divisors for the $\mathfrak{o}$-module inclusion $\mathfrak{a} \hookrightarrow \mathfrak{a}^\vee$ are $(\mathfrak{n}, \mathfrak{n})$.*

Proof. Our hypotheses imply that $\mathfrak{d} = \mathfrak{o}$ and that $\mathfrak{n}$ is an integral ideal. The lemma implies that there is an isomorphism (first of $\mathfrak{o}_E$-modules, then of $\mathfrak{o}$-modules) $\mathfrak{a}^\vee/\mathfrak{a} \cong \mathfrak{o}_E/\mathfrak{n}\mathfrak{o}_E \cong (\mathfrak{o}/\mathfrak{n})^2$, whence the conclusion. □

Remark 7.3. Under the hypotheses of the corollary, the discriminant ideal of the binary quadratic form $(q, \mathfrak{a})$ is $\mathfrak{n}^2$. More generally, under the hypotheses of the lemma, the discriminant ideal is $\mathfrak{D}\mathfrak{n}^2$, with $\mathfrak{D} = \mathrm{nr}(\mathfrak{d})$. Conversely, given the discriminant ideal, we may compute $\mathfrak{n}$ as its square root.

7.2. Quaternionic preliminaries: general case

Let F be a non-Archimedean local field, let B be a quaternion F-algebra and let E be a separable quadratic F-subalgebra of B. We equip B with the quadratic form $q : B \to F$ given by the reduced norm, whose bilinearization $\langle\,,\rangle$ as above is described by the reduced trace and the main involution on B via the formula $\langle x, y\rangle = \mathrm{tr}(x\bar{y})$. We have a canonical decomposition $B = E \oplus E^\perp$, where $E^\perp = \{x \in B : \langle x, y\rangle = 0 \text{ for all } y \in E\}$.

Let $\mathfrak{o}$ and $\mathfrak{o}_E$ denote the respective maximal orders of F and E. We write $\mathfrak{d}$ for the different ideal, as before.

Let us say that an order R in B is *E-adapted* if it is of the form $R = \mathfrak{o}_E \oplus \mathfrak{a}$ for some $\mathfrak{o}_E$-submodule $\mathfrak{a}$ of $E^\perp$ (for the action by either left or right multiplication – it doesn't matter which because they are conjugates of each other).

Consider such an order R. Its traceless submodule is given by

$$R^0 = \mathfrak{o}_E^0 \oplus \mathfrak{a}.$$

We aim to compute the dual lattice $(R^0)^\vee$ with respect to q. To that end, it suffices to dualize each summand in the above decomposition because $(R^0)^\vee = (\mathfrak{o}_E^0)^\vee \oplus \mathfrak{a}^\vee$. We generally have

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$(\mathfrak{d}^{-1})^0 \subseteq (\mathfrak{o}_E^0)^\vee \subseteq \frac{1}{2}(\mathfrak{d}^{-1})^0$. If E is unramified or split, then $(\mathfrak{o}_E^0)^\vee = \frac{1}{2}\mathfrak{o}_E^0 = \frac{1}{2}(\mathfrak{d}^{-1})^0$. On the other hand, we can compute $\mathfrak{a}^\vee$ using the results of the previous section. Indeed, the choice of any invertible element $j \in E^\perp$ defines an isomorphism $E \to E^\perp$, $x \mapsto xj$. Transporting q and $\mathfrak{a}$ via the inverse of this isomorphism gives us a fractional ideal in E and a quadratic form on E that satisfy the hypotheses of that section. We obtain

$$\mathfrak{a}^\vee = \mathfrak{d}^{-1}\mathfrak{n}^{-1}\mathfrak{a},$$

where $\mathfrak{n}$ is the integral $\mathfrak{o}$-ideal characterized by either of the following properties:

- $\mathfrak{n}$ is generated by $q(\mathfrak{a})$.
- $\mathfrak{D}\mathfrak{n}^2$ is the discriminant ideal of $(q, \mathfrak{a})$.

Let $\mathcal{D}$ denote the discriminant ideal of R. The discriminant ideal of the summand $(q, \mathfrak{o}_E)$ is $\mathfrak{D}$. Since the discriminant ideal is multiplicative with respect to direct sums, we obtain

$$\mathcal{D} = \mathfrak{D}^2\mathfrak{n}^2. \tag{7.1}$$

We may regard this last identity as a formula for $\mathfrak{n}$.

7.3. *Quaternionic preliminaries: unramified case*

Let us restrict henceforth to the case that E/F *is unramified*. (The ramified case would be relevant for studying, for example, the 'minimal vectors' considered in [HNS19, HN18, Sah20].)

The above formula then simplifies to

$$\mathfrak{n}^2 = \mathcal{D},$$

and we obtain

$$\mathfrak{a}^\vee/\mathfrak{a} \cong (\mathfrak{o}/\mathfrak{n})^2.$$

There are three possibilities for $F \subset E \subset B$, up to isomorphism:

(i) B is split and $E \cong F \oplus F$. We may then find an isomorphism $B \cong \mathrm{Mat}_{2\times 2}(F)$ under which E identifies with the diagonal subalgebra.
(ii) B is split and E/F is the unique unramified quadratic field extension.
(iii) B is nonsplit and E/F is the unique unramified quadratic field extension.

In Case (i), the E-adapted orders R are just the Eichler orders. An Eichler order of level $\mathfrak{q}$ has discriminant $\mathfrak{q}^2$ (cf. [Voi18, §23.4]), hence the ideal $\mathfrak{n}$ defined above is $\mathfrak{q}$.

Case (ii) corresponds to another type of 'minimal vectors', which we do not consider in this paper.

In Case (iii), the maximal order R is E-adapted and has discriminant ideal $\mathfrak{p}^2$ (cf. [Voi18, §15.2.11, §23.4]), hence $\mathfrak{n} = \mathfrak{p}$.

7.4. *Bounds for commutators of elements of* R^0

Lemma 7.4. *Let E/F unramified and R be an E-adapted order. Let $[\gamma_1, \gamma_2] = \gamma_1\gamma_2 - \gamma_2\gamma_1$ denote the commutator for two elements γ_1, γ_2 in B. Then, we have*

(i) $q([\gamma_1, \gamma_2]) \in \mathfrak{n}$ *for all* $\gamma_1, \gamma_2 \in R$,
(ii) $[\gamma_1, \gamma_2] \in \mathfrak{n}^{-1}R$ *for all* $\gamma_1, \gamma_2 \in R^\vee$,
(iii) $q([\gamma_1, \gamma_2]) \in \mathfrak{n}^{-2}$ *for all* $\gamma_1, \gamma_2 \in R^\vee$.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Proof. By the previous discussion, we may write $\gamma_i = \alpha_i + \beta_i$ with $\alpha_i \in \mathfrak{o}_E$ and $\beta_i \in \mathfrak{a}$ if $\gamma_i \in R$, respectively $\mathfrak{n}^{-1}\mathfrak{a}$ if $\gamma_i \in R^\vee$. We have $[\alpha_1, \alpha_2] = 0$ and $[\alpha_1, \beta_2] = 2\alpha_1\beta_2 \in \mathfrak{a}$, respectively $\in \mathfrak{n}^{-1}\mathfrak{a}$. Lastly, we have $\mathfrak{a}^2 = \mathfrak{n}\mathfrak{a}^\vee\mathfrak{a} = \mathfrak{n}$ and $(\mathfrak{a}^\vee)^2 = \mathfrak{n}^{-1}\mathfrak{a}\mathfrak{a}^\vee = \mathfrak{n}^{-1}$. Hence,

$$[\beta_1, \beta_2] = \beta_1\beta_2 - \beta_2\beta_1 \in \mathfrak{n}, \text{ respectively } \mathfrak{n}^{-1}.$$

By the bilinearity of the commutator, we obtain Claim (ii) and subsequently Claim (iii). Similarly, we have $q([\beta_1, \beta_2]) \in \mathfrak{n}^2$ for $\beta_i \in \mathfrak{a}$. Thus, by orthogonality, we have

$$q([\gamma_1, \gamma_2]) = q(2\alpha_1\beta_2 - 2\alpha_2\beta_1 + [\beta_1, \beta_2]) = q(2\alpha_1\beta_2 - 2\alpha_2\beta_1) + q([\beta_1, \beta_2]) \in \mathfrak{n},$$

for $\gamma_i \in R$, which gives Claim (i). □

8. Invariants of rational quadratic forms

Let V be an n-dimensional $\mathbb{Q}$-vector space and $q : V \to \mathbb{Q}$ a nondegenerate quadratic form. We normalize the polarization $\langle\,,\rangle$ of q such that $\langle x, x\rangle = 2q(x)$.

Given a lattice $L \subseteq V$ and a positive-definite quadratic form Q on $V \otimes_{\mathbb{Q}} \mathbb{R}$, one can define the successive minima of the pair (L, Q). Our aim in this section is to provide certain estimates for those successive minima in terms of other invariants of (L, Q). Our results may be understood as a generalization of those of Blomer–Michel [BM13, §3, §4], who treated the special case that q is definite and $Q = q$. We mention also the work of Saha [Sah20, §2].

8.1. Non-Archimedean invariants

Let $L \subseteq V$ be a lattice, that is, a $\mathbb{Z}$-submodule whose rank is the dimension of V. Define

- the *content* C of L to be the greatest common divisor of $q(L)$,
- the *level* N of L to be the reciprocal of the content of the dual lattice $L^\vee$ and
- the (unsigned) *discriminant* Δ of L to be the absolute value of the determinant of the Gram matrix of q on L. (The discriminant of a quadratic form is traditionally defined without taking absolute values, but the sign will not matter for us.)

Remark 8.1. In general, the content of L does not agree with the content of the Gram matrix of q on L, that is, the greatest common divisor of the entries. However, if q is integral on L, then the level of L agrees with the level of the Gram matrix of q.

Remark 8.2. For our purposes, the content, respectively level, may be replaced by the first, respectively last, elementary divisor of the Gram matrix of q as these quantities differ by a bounded power of 2.

To get acquainted with these quantities, consider for instance the case that L admits a basis $e_1, \dots, e_n$ with respect to which q is given by the diagonal quadratic form $q(\sum x_i e_i) = \frac{1}{2}\sum a_i x_i^2$ for some nonzero rational numbers $a_1, \dots, a_n$. Then

$$C = \tfrac{1}{2}\gcd(a_1, \dots, a_n),$$

$$N = 2/\gcd(1/a_1, \dots, 1/a_n) = 2 \cdot \operatorname{lcm}(a_1, \dots, a_n),$$

$$\Delta = |a_1 \cdots a_n|.$$

We may relate the invariants attached to homothetic lattices: the effect of the substitution $L \mapsto mL$ for a nonzero rational scalar m is

$$C \mapsto m^2 C, \qquad N \mapsto m^2 N, \qquad \Delta \mapsto m^{2n} N.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

8.2. Archimedean invariants

Next, write $V_{\mathbb{R}} := V \otimes_{\mathbb{Q}} \mathbb{R}$ and let $Q : V_{\mathbb{R}} \to \mathbb{R}$ be a positive-definite quadratic form. We may find a basis $e_1, \dots, e_n$ of $V_{\mathbb{R}}$ so that, writing $x = \sum x_i e_i$, we have $Q(x) = \frac{1}{2}\sum x_i^2$ and $q(x) = \frac{1}{2}\sum a_i x_i^2$ for some nonzero real numbers $a_1, \dots, a_n$. The *dual* $Q^\vee$ of Q with respect to q may be defined by $Q^\vee(x) = \frac{1}{2}\sum a_i^2 x_i^2$. We note that in the scaled coordinates $y = \sum a_i^{-1} y_i e_i$, we have $Q^\vee(y) = \frac{1}{2}\sum y_i^2$ and $q(y) = \frac{1}{2}\sum a_i^{-1} y_i^2$. The *Gram matrix relative to q* of Q is defined to be the diagonal matrix with entries $(a_1, \dots, a_n)$.

Define

- the *content* C of Q to be the infimum of the ratio $Q/|q|$ over the set where $q \neq 0$,
- the *level* N of Q to be the reciprocal of the content of the dual $Q^\vee$ of Q and
- the *discriminant* Δ of Q to be the absolute value of the reciprocal of the determinant of the Gram matrix relative to q of Q.

The invariants of Q may be described in terms of coordinates as above by

$$C = 1/\max(|a_1|, \dots, |a_n|) = \min(1/|a_1|, \dots, 1/|a_n|),$$

$$N = \max(1/|a_1|, \dots, 1/|a_n|) = 1/\min(|a_1|, \dots, |a_n|).$$

$$\Delta = 1/|a_1 \cdots a_n|.$$

These again behave predictably under homotheties: If we substitute $Q \mapsto Q_m := [x \mapsto Q(m^{-1}x)]$ for some nonzero real scalar m (which has the effect of multiplying the Q-unit ball by m), then the coefficients transform like $a_i \mapsto m^2 a_i$ and hence the invariants like

$$C \mapsto m^{-2}C, \qquad N \mapsto m^{-2}N, \qquad \Delta \mapsto m^{-2n}\Delta.$$

For later reference, it will be convenient to explicate the definition of 'level' in terms of matrices. To that end, we note first that for any basis $e_1, \dots, e_n$ of V, we may find symmetric matrices S and P that represent q and Q in the sense that, for example, $q(\sum x_i e_i) = \frac{1}{2}\sum\sum x_i S_{ij} x_j$. By singular value decomposition, we may find nonsingular matrices A and D, with D diagonal, so that

$$P = A^t A \quad \text{and} \quad S = A^t D A. \tag{8.1}$$

The level of Q is then the operator norm of the matrix D^{-1}. For instance, if we choose our basis so that $Q(\sum x_i e_i) = \frac{1}{2}\sum x_i^2$ and $q(\sum x_i e_i) = \frac{1}{2}\sum d_i x_i^2$, then the level of Q is $\max |d_i|^{-1}$.

Remark 8.3. We may relate the above definition of level to more standard notions. We recall that the form Q is a *majorant* of q if $PS^{-1}P = S$, or equivalently, if $D^2 = 1$. Suppose that $|q| \leq Q$. Then the level of Q is always at least 1, and it is equal to 1 if and only if Q is a majorant of q. Indeed, the assumption $|q| \leq Q$ implies that $|d_i| \leq 1$ for each i, with equality precisely when $D^2 = 1$.

8.3. Duality

We note that replacing L (resp. Q) with its dual $L^\vee$ (resp. $Q^\vee$) has the following effect on the invariants:

$$(C, N, \Delta) \mapsto (1/N, 1/C, 1/\Delta). \tag{8.2}$$

In what follows, this relation allows us to reduce slightly the number of computations required. For instance, we can read off the invariants of the dual of an Eichler order (or its traceless submodule) from those of the Eichler order itself.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

8.4. Adelic invariants

Let (Q, L) be a pair consisting of a positive-definite quadratic form Q and a lattice L as above. We define the *content* (resp. *level*, *discriminant*) of the pair to be the product of the corresponding invariants of Q and L.

We note that the invariants C, N, Δ assigned to the pair (Q, L) are invariant by rational homotheties, that is, replacing L by mL and Q by Q_m for the same nonzero rational scalar m, and also under automorphisms of V that preserve q.

Furthermore, the discriminant of the pair (Q, L) is the same as the determinant of the Gram matrix of Q with respect to a $\mathbb{Z}$-basis of L, as the discriminant of Q is nothing but the inverse of the determinant of the matrix D in the singular value decomposition (8.1).

8.5. Statement of result

Proposition 8.4. *Let V be an n-dimensional $\mathbb{Q}$-vector space. Let $q : V \to \mathbb{Q}$ be an anisotropic quadratic form. Let $L \subset V$ be a lattice. Let $Q : V_{\mathbb{R}} \to \mathbb{R}$ be a positive-definite quadratic form. Let $\lambda_1 \le \cdots \le \lambda_n$ denote the successive minima of Q on L (see Definition 6.1). Let C, N, Δ denote the content, level and discriminant of the pair (Q, L). Then,*

(i) $\lambda_1 \ge C^{1/2}$,
(ii) $\lambda_1 \cdots \lambda_n \asymp \Delta^{1/2}$ *and*
(iii) $\lambda_1 \cdots \lambda_{n-1} \gg (\Delta/N)^{1/2}$.

In particular, for $n = 3$, we have for all $X > 0$ that

$$|\{v \in L : Q(v) \le X^2\}| \ll 1 + \frac{X}{\sqrt{C}} + \frac{X^2}{\sqrt{\Delta/N}} + \frac{X^3}{\sqrt{\Delta}}.$$

Remark 8.5. This last estimate is scale-invariant in the sense that replacing (Q, L) with (Q_m, mL) for a positive rational scalar m has no effect on the right-hand side. This feature is not surprising in view of the multiplication-by-m bijection $\{v \in L : Q(v) \le X^2\} \cong \{v \in mL : Q_m(v) \le X^2\}$. The estimate is likewise invariant under replacing (Q, X) by $(Q_m, X/m^2)$ for some nonzero real number m, as one might expect for similar reasons.

Proof. We follow the basic strategy of Blomer–Michel [BM13, §3, §4], who established the corresponding result for q positive-definite and $Q = q$.

Let v be a nonzero element of L. Since L is anisotropic, we have $q(v) \ne 0$. Let C_Q, C_L denote the content of Q, respectively L. By the definition of the content, we have $C_L|q(v)$ and $Q \ge C_Q|q|$. Thus, $Q(v) \ge C_Q|q(v)| \ge C_Q C_L = C$, giving Claim (i).

By Lemma 6.3, we may find a basis $e_1, \ldots, e_n$ of L so that the submodules $L_m := \sum_{j \le m} \mathbb{Z} e_j$ have covolume $\asymp \prod_{j \le m} \lambda_j$ in their real span, with volume defined using the restriction of Q. On the other hand, that covolume is the square root of the Gram determinant of Q on L_m. Write $\det(Q, L_m)$ and $\det(q, L_m)$ for the respective Gram determinants of Q and q. Then

$$\prod_{j \le m} \lambda_j \asymp \det(Q, L_m)^{\frac{1}{2}}. \tag{8.3}$$

Since $\det(Q, L_n) = \Delta$ by the remark in §8.4, the case $m = n$ of this estimate gives Claim (ii).

For the proof of Claim (iii), observe first that in view of Claim (ii), it is equivalent to check that $\lambda_n \ll (N_L N_Q)^{1/2}$, where N_L, respectively N_Q, is the level of L, respectively Q. To that end, write $P = A^t A$ and $S = A^t D A$ for the matrices of Q and q, as in §8.2, and consider the final matrix entry $(P^{-1})_{nn}$ of the inverse of P. Cramer's rule expresses $(P^{-1})_{nn} = \det(Q, L_{n-1})/\det(Q, L_n)$, so by the

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

cases $m = n-1, n$ of Equation (8.3), we have $(P^{-1})_{nn} \asymp 1/\lambda_n^2$. On the other hand, since N_Q bounds the operator norm of D^{-1}, we have

$$|\,(S^{-1})_{nn}\,| = |\,\langle A^{-t}e_n, D^{-1}A^{-t}e_n\rangle\,| \le N_Q\langle A^{-t}e_n, A^{-t}e_n\rangle = N_Q(P^{-1})_{nn}.$$

Cramer's rule likewise expresses $(S^{-1})_{nn} = \det(q, L_{n-1})/\det(q, L_n)$ as a ratio of Gram determinants. Since q is anisotropic, both determinants are nonzero. Since $2N_L(S^{-1})_{nn} \in \mathbb{Z}$, it follows that $1/(2N_L) \le |(S^{-1})_{nn}|$. Thus, $1/\lambda_n^2 \asymp (P^{-1})_{nn} \ge 1/(2N_LN_Q)$, giving the required estimate. □

9. Type I estimates

The local computations of Section §7, together with the behavior of invariants under duality recorded in §8.2, imply that the elementary divisors of the Gram matrix of the reduced trace form on $g^{-1}R(\ell)^0g$ are given by

$$\begin{cases} \left(\frac{1}{\ell}, \frac{d_BN}{\ell^2}, \frac{2d_BN}{\ell}\right), & 2 \nmid d_BN, \\ \left(\frac{2}{\ell}, \frac{d_BN}{\ell^2}, \frac{d_BN}{\ell}\right), & 2|d_BN, 2 \nmid \ell, \\ \left(\frac{1}{\ell}, \frac{d_BN}{\ell^2}, \frac{d_BN}{2\ell}\right), & 2|d_BN, 2|\ell. \end{cases} \tag{9.1}$$

Hence, the content, level and discriminant of $g^{-1}R(\ell)^0g$ with respect to the reduced norm are comparable to $1/\ell$, d_BN/ℓ and $(d_BN)^2/\ell^4$ respectively. Here, 'comparable to' means the ratios are bounded from above and below by positive constants. Suppose that the reduced norm on R is anisotropic. In this case, we wish to apply Proposition 8.4 to the lattice $g^{-1}R(\ell)^0g$ with q given by the reduced norm. Recall the notation P, u and X from Equation (5.1). As a first choice, we let $Q = P + \delta^{-1}u$, whose content, level and discriminant are comparable to $1, \delta^{-1}$ and δ^{-2}, respectively. This yields that the first successive minima of $g^{-1}R(\ell)^0g$ with respect to $P + \delta^{-1}u$ is $\gg \ell^{-\frac{1}{2}}$, and hence also with respect to $\Omega(\delta, 1) \cap B_\infty^0$. Furthermore,

$$|g^{-1}R(\ell)^0g \cap \Omega(\delta, T)| \prec 1 + \ell^{\frac{1}{2}}T + \frac{\ell^{\frac{3}{2}}\delta^{\frac{1}{2}}}{(d_BN)^{\frac{1}{2}}}T^2 + \frac{\ell^2\delta}{d_BN}T^3.$$

This proves the first half of Theorem 2.5. Similarly, we have for the choice $Q = P + \delta^{-1}|X|^2$ that the content, level and discriminant Q are comparable to $1, \delta^{-1}$ and δ^{-1}, respectively. Thus, the first successive minima of $g^{-1}R(\ell)^0g$ with respect to $\Psi(\delta, 1) \cap B_\infty^0$ is $\gg \ell^{-\frac{1}{2}}$ and

$$|g^{-1}R(\ell)^0g \cap \Psi(\delta, T)| \prec 1 + \ell^{\frac{1}{2}}T + \frac{\ell^{\frac{3}{2}}}{(d_BN)^{\frac{1}{2}}}T^2 + \frac{\ell^2\delta^{\frac{1}{2}}}{d_BN}T^3,$$

which is the second half of Theorem 2.5.

We now turn to the case that the quaternion algebra B is split. Here, we proceed in a more ad hoc manner. First, we note that $R(\ell)^0$ is normalized by the Atkin–Lehner operators. Thus, we need only consider $g \in G(\mathbb{R})$ such that $g \cdot i = x + iy$ has maximal imaginary part under the action of the Atkin–Lehner operators. In particular, we have $H(g) = y$. Let $\lambda_1 \le \lambda_2 \le \lambda_3$ be the successive minima (Definition 6.1) of the closed convex 0-symmetric set $\Omega(\delta, 1) \cap B_\infty^0$ with respect to the lattice $g^{-1}R(\ell)^0g$. Since $\Omega(\delta, 1)$ is both left and right K_∞-invariant, we may further assume that $g = \left(\begin{smallmatrix} 1 & x \\ & 1 \end{smallmatrix}\right)\mathrm{diag}(y^{\frac{1}{2}}, y^{-\frac{1}{2}})$. By Lemma 6.4, we have

$$|g^{-1}R(\ell)^0g \cap \Omega(\delta, T)| = |g^{-1}R(\ell)^0g \cap T\Omega(\delta, 1)| \asymp 1 + \frac{T}{\lambda_1} + \frac{T^2}{\lambda_1\lambda_2} + \frac{T^3}{\lambda_1\lambda_2\lambda_3}.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Let $\beta_0 = \left(\begin{smallmatrix} a & b \\ c & -a \end{smallmatrix}\right) \in R(\ell)^0$, thus $a \in \mathbb{Z}, b \in \frac{1}{\ell}\mathbb{Z}, c \in \frac{N}{\ell}\mathbb{Z}$. Let

$$\alpha_0 = g^{-1}\beta_0 g = \begin{pmatrix} a - cx & \frac{1}{y}(2ax + b - cx^2) \\ cy & cx - a \end{pmatrix}.$$

Suppose $\beta_0 \neq 0$. If $(a, c) \neq (0, 0)$, then by Lemma 6.2 we have $P(\alpha_0) \geq \frac{1}{2}|cz - a|^2 \geq \frac{1}{2\ell}$. Otherwise, we have $u(\alpha_0) = (\frac{b}{y})^2 \geq \frac{1}{2(\ell y)^2}$. Hence, we have $\lambda_1 \gg \min\{\ell^{-\frac{1}{2}}, \ell^{-1}y^{-1}\delta^{-\frac{1}{2}}\}$. In order to get a lower bound on $\lambda_1\lambda_2$ and $\lambda_1\lambda_2\lambda_3$, we shall give an upper bound on $|g^{-1}R(\ell)^0 g \cap \Omega(\delta, T)|$ along the lines of Harcos–Templier [HT13]. First, we bound the number of choices of c by $\ll 1 + \frac{\ell T}{Ny}$ as $|cy| \leq P(\alpha_0)^{\frac{1}{2}} \leq T$. For each such choice of c, the equation

$$\frac{1}{y^2}\left|-cz^2 + 2az + b\right|^2 = u(\alpha_0) \leq \delta T^2$$

defines a circle of radius $\delta^{\frac{1}{2}}Ty$ and center cz^2 in which we need to count lattice points of the lattice generated by $2z$ and $\frac{1}{\ell}$. This lattice has covolume $2y/\ell$ and first successive minima $\geq (\ell N)^{-\frac{1}{2}}$ by Lemma 6.2. We may thus apply Lemma 6.5 to bound the number of (a, b) by $\ll 1 + \ell^{\frac{1}{2}}N^{\frac{1}{2}}\delta^{\frac{1}{2}}Ty + \ell\delta T^2 y$. We obtain

$$\frac{T^2}{\lambda_1\lambda_2} + \frac{T^3}{\lambda_1\lambda_2\lambda_3} \ll \left(1 + \frac{\ell T}{Ny}\right)\left(1 + \ell^{\frac{1}{2}}N^{\frac{1}{2}}\delta^{\frac{1}{2}}Ty + \ell\delta T^2 y\right).$$

By letting T tend to ∞, it follows that $\lambda_1\lambda_2\lambda_3 \gg N\ell^{-2}\delta^{-1}$. By taking $T = N^{\frac{1}{2}}/(\ell\delta)^{\frac{1}{2}}$, we obtain

$$\frac{1}{\lambda_1\lambda_2} \ll \frac{\ell\delta}{N} + \frac{\ell^{\frac{3}{2}}\delta^{\frac{1}{2}}}{N^{\frac{3}{2}}y} + \ell\delta y + \frac{\ell^{\frac{3}{2}}\delta^{\frac{1}{2}}}{N^{\frac{1}{2}}} \ll \frac{\ell^{\frac{3}{2}}\delta^{\frac{1}{2}}}{N^{\frac{1}{2}}} + \ell\delta y.$$

We thereby conclude the proof of the final case of Theorem 2.5.

10. Type II estimates

10.1. Bounds for representation numbers of binary quadratic forms

Lemma 10.1. *Let M be a free $\mathbb{Z}$-module of rank 2. Let $q : M \to \mathbb{Z}$ be a nondegenerate integral binary quadratic form. Let Q be a positive-definite quadratic form on $M \otimes \mathbb{R}$ such that $|q| \leq Q$. Let n be a nonzero integer and let $X \geq 1$. Set*

$$S := \{\beta \in M : q(\beta) = n, Q(\beta) \leq X^2\}.$$

Then,

$$|S| \ll_\varepsilon (X|n|)^\varepsilon.$$

Proof. We begin with a preliminary reduction. Suppose we can find a lattice M' in $M \otimes \mathbb{Q}$ that contains M and on which the $\mathbb{Q}$-bilinear extension of q is integral. It suffices then to verify the lemma after replacing M with M'. Indeed, doing so enlarges the set S. In particular, we may assume that the quadratic form M is *primitive*.

Suppose now that M is primitive and anisotropic. Without loss of generality, M is either positive-definite or indefinite. By the form-ideal correspondence, we may assume then that M is an invertible

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

ideal of an order $\mathfrak{o}$ in a quadratic field, with q given by the element norm ν divided by the ideal norm $\nu(M)$ of M:

$$q(\beta) = \nu(\beta)/\nu(M).$$

We will establish the estimate

$$|S| \ll \log(2 + X^2/|n|)\tau(n), \tag{10.1}$$

which suffices in view of the divisor bound. Let $\mathfrak{o}_{\max}$ denote the maximal order in the quadratic field containing $\mathfrak{o}$. For each $\beta \in S$, the $\mathfrak{o}$-ideal $M^{-1}\beta$ has norm $|n|$, as does the $\mathfrak{o}_{\max}$-ideal $\mathfrak{o}_{\max}M^{-1}\beta$. The number of $\mathfrak{o}_{\max}$-ideals of norm $|n|$ is at most $\tau(n)$. Suppose two elements $\beta_0, \beta \in S$ give rise to the same $\mathfrak{o}_{\max}$-ideal. Then, β/β_0 is a norm one unit in $\mathfrak{o}_{\max}^\times$. The required estimate follows in the imaginary quadratic case (without the logarithmic factor) because $|\mathfrak{o}_{\max}^\times| \leq 6$. In the real quadratic case, we fix a positive generator η for the group $\cong \mathbb{Z}$ of non-root-of-unity norm one units in $\mathfrak{o}_{\max}$ and write $\beta = \pm\beta_0\eta^\ell$ for some $\ell \in \mathbb{Z}$. It will suffice then to verify that $\ell \ll \log(2+X^2/|n|)$. To that end, we estimate $q(\beta_0+\beta)$ in two ways. On the one hand, the triangle inequality for the Euclidean norm defined by Q gives the upper bound $|q(\beta_0+\beta)| \leq Q(\beta_0+\beta) \ll Q(\beta_0) + Q(\beta) \ll X^2$. On the other hand, the multiplicativity of ν gives the identity $q(\beta_0+\beta) = n\nu(1\pm\eta^\ell)$. The lower bound $\nu(1\pm\eta^\ell) \geq \frac{1}{4}\cdot 1.618^\ell$ for fundamental units now yields the required estimate for ℓ.

It remains to consider the case that M is isotropic and q nondegenerate. In that case, after applying our preliminary reduction to enlarge M if necessary, we may assume that $M = \mathbb{Z}^2$ and $q(x,y) = xy$. Indeed, we may choose a basis e_1, e_2 for M with e_1 isotropic. Then, q is given with respect to the coordinates $xe_1 + ye_2$ by $q(x,y) = axy + by^2$ for some $a, b \in \mathbb{Z}$, with $a \neq 0$. Then $q(\frac{x-by}{a}, y) = xy$ and $M \subseteq M' := \{\frac{x-by}{a}e_1 + ye_2 : x, y \in \mathbb{Z}\}$, so (M', q) gives the required enlargement. Now, since $n \neq 0$, the divisor bound gives $|S| = 2\tau(n) \ll_\varepsilon |n|^\varepsilon$. □

10.2. Local quaternionic preliminaries

Let B be a quaternion algebra over the rationals. We write d_B for its reduced discriminant and q for its reduced norm.

10.2.1. Non-Archimedean preliminaries

Let $R \subset B$ be an Eichler order of level N, with N coprime to d_B.

Lemma 10.2. *For $x, y \in R(\ell)^0$, we have $[x,y] \in \frac{1}{\ell}R^0$ and $q([x,y]) \in \frac{d_B N}{\ell^3}\mathbb{Z}$.*

Proof. This follows from the local computations in Lemma 7.4 together with the fact that the trace of a commutator is zero. □

10.2.2. Archimedean preliminaries

Recall the notation 'Ω' from §2.4.4 and 'P' from Equation (5.1).

Lemma 10.3. *For $0 < \delta \leq 1$ and $x, y \in \Omega(\delta, T)$, we have*

$$q([x,y]) \ll \delta T^4$$

and

$$P([x,y]) \ll \delta T^4.$$

Proof. For $\delta = 1$, the required estimates reduce via homogeneity to the compactness of the unit ball and the continuity of multiplication. We turn to the case $0 < \delta < 1$. We embed $B_\infty \hookrightarrow \mathrm{Mat}_{2\times 2}(\mathbb{C})$ by $\mathbf{i} \mapsto \left(\begin{smallmatrix} i & 0 \\ 0 & -i \end{smallmatrix}\right)$, $\mathbf{j} \mapsto \left(\begin{smallmatrix} 0 & 1 \\ \pm 1 & 0 \end{smallmatrix}\right)$. Then, P is asymptotic to the restriction of the squared Euclidean norm on the

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

matrix entries, while q is the restriction of the determinant. We may assume that $x = [a_1, b_1, c_1]$ and $y = [a_2, b_2, c_2]$ are nonzero. We may assume (by the known $\delta = 1$ case) that δ is sufficiently small. The required conclusion then reduces via homogeneity to the following assertion: The commutator of any two matrices of the form

$$\begin{pmatrix} a_1 i & O(\delta^{1/2}) \\ O(\delta^{1/2}) & -a_1 i \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a_2 i & O(\delta^{1/2}) \\ O(\delta^{1/2}) & -a_2 i \end{pmatrix}$$

is of the form $\begin{pmatrix} O(\delta) & O(\delta^{1/2}) \\ O(\delta^{1/2}) & O(\delta) \end{pmatrix}$, and in particular has Euclidean norm $O(\delta^{1/2})$ and determinant $O(\delta)$. Indeed, the product of any two matrices of the indicated form is readily computed to be of the form $\begin{pmatrix} -a_1 a_2 + O(\delta) & O(\delta^{1/2}) \\ O(\delta^{1/2}) & -a_1 a_2 + O(\delta) \end{pmatrix}$, hence the commutator of two such matrices, being a difference of such products, has the required form. □

10.3. The nonsplit case

We retain the above setting and further assume that B is *nonsplit* and that the level N of the Eichler order R is *squarefree*.

The following estimate is, in some sense, the most intricate one in the paper. It requires us to bound certain matrix counts in the critical range (see Remark 10.5 below) by essentially $O(1)$, uniformly in the discriminant and level. To achieve such uniformity seems to require the delicate argument involving commutators recorded below.

Proposition 10.4. *Let n be a nonzero integer, let $0 < \delta \leq 1$ and let $T \geq \ell^{-\frac{1}{2}}$. Then the set $S := R(\ell)^0 \cap \Omega(\delta, T) \cap q^{-1}(\{n\})$ has cardinality*

$$|S| \ll_\varepsilon (\ell T)^\varepsilon \tau(d_B N) \left(1 + \frac{\ell^2}{d_B N} \min\left\{\delta^{\frac{1}{2}} T^2, \frac{\delta T^4}{|n|}\right\}\right).$$

Remark 10.5. The critical range is when $n \asymp T^2 \asymp d_B N(1+k)/\ell^2$ and $\delta \asymp (1+k)^{-1}$; in that range, we obtain $|S| \ll_\varepsilon (d_B N(1+k))^\varepsilon$.

Proof. Suppose S is not empty, and let $\gamma_1, \gamma_2 \in S$. Our strategy will be to bound for each γ_1 the number of possibilities for γ_2.

Set $\beta := [\gamma_1, \gamma_2] \in \frac{1}{\ell} R^0$ and $a := \operatorname{tr}(\gamma_1 \gamma_2) \in \frac{1}{\ell}\mathbb{Z}$. Then, $2\gamma_1\gamma_2 = a + \beta$, $4n^2 = a^2 + q(\beta)$. In particular, $\gamma_2 = (a+\beta)/2\gamma_1$, so it suffices to bound the number of possibilities for a and β.

Lemmas 10.2 and 10.3 give $q(\beta) \ll \delta T^4$ and $q(\beta) \in \frac{d_B N}{\ell^3}\mathbb{Z}$, that is,

$$a^2 = 4n^2 + O(\delta T^4),$$

$$a^2 \equiv 4n^2 \quad (\tfrac{d_B N}{\ell^3}),$$

thus

$$a = \pm 2n + O\left(\min\left\{\delta^{\frac{1}{2}} T^2, \frac{\delta T^4}{|n|}\right\}\right),$$

$$a \equiv a_0 \quad (\tfrac{d_B N}{\ell^2})$$

for some sign $\pm$ and some residue class a_0 modulo $d_B N/\ell^2$ with $a_0^2 \equiv 4n^2$ $(\frac{d_B N}{\ell^2})$. Since $d_B N$ is squarefree, there are at most $\tau(d_B N)$ such classes. For each a_0, the number of possibilities for a is $O(1 + (\ell^2/d_B N) \cdot \min\{\delta^{\frac{1}{2}} T^2, \delta T^4/|n|\})$.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

We now bound for each a the number of possibilities for β. Let M denote the orthogonal complement in $\frac{1}{\ell}R^0$ of γ_1, thus $M = \{\gamma \in \frac{1}{\ell}R^0 : \mathrm{tr}(\gamma\gamma_1) = 0\}$. By restricting q to M, we obtain an integral binary quadratic form. Since B is nonsplit, M is anisotropic. Since $\mathrm{tr}(\gamma_1\gamma_2\gamma_1) = \mathrm{tr}(\gamma_2\gamma_1\gamma_1)$, we have $\beta \in M$. From Lemma 10.3, we obtain $P(\beta) \ll T^4$. Thus, β satisfies the system

$$\beta \in M, \quad q(\beta) = 4n^2 - a^2, \quad P(\beta) \ll T^4.$$

Since $q(M) \subseteq \frac{1}{\ell^2}\mathbb{Z}$ and $4n^2 - a^2 \ll \delta T^4$, we see by Lemma 10.1 that the number of possibilities for β is $\ll_\varepsilon (\ell T)^\varepsilon$.

By multiplying together the number of possibilities for $\pm, a_0, a$ and β, we achieve the required bound. □

10.4. Extension to the split case

Recall from §2.2, that we may assume our Eichler order of level N in the split quaternion algebra $B = \mathrm{Mat}_{2\times 2}(\mathbb{Q})$ to be of the shape $g^{-1}Rg$, where $R = \left(\begin{smallmatrix} \mathbb{Z} & \mathbb{Z} \\ N\mathbb{Z} & \mathbb{Z} \end{smallmatrix}\right)$ and $g \in G(\mathbb{R})$. Our aim is to bound the cardinality of the set

$$S := g^{-1}R(\ell)^0 g \cap \Omega(\delta, T) \cap \det^{-1}(\{n\}).$$

In the nonsplit case, we had verified that

$$|S| \ll_\varepsilon C,$$

where

$$C := 1 + (\ell T)^\varepsilon \tau(d_B N)\left(1 + \frac{\ell^2}{d_B N}\min\left\{\delta^{\frac{1}{2}}T^2, \frac{\delta T^4}{|n|}\right\}\right).$$

We extend this to the split case as follows.

Proposition 10.6. *Let n be an integer. If $-n$ is not a square, then $|S| \ll_\varepsilon C$. If $-n$ is a square, then $|S| \ll_\varepsilon C + \delta^{1/2}T\ell H(g)$, where H denotes the normalized height function defined in §2.2.*

Proof. We proceed as in the original argument, aiming to bound for fixed $\gamma_1 \in gSg^{-1}$ the number of possible $\gamma_2 \in gSg^{-1}$. As before, we write

$$\alpha = \mathrm{tr}(\gamma_1\gamma_2) = \gamma_1\gamma_2 + \gamma_2\gamma_1 \in \tfrac{1}{\ell}\mathbb{Z},$$

$$\beta = [\gamma_1, \gamma_2] = \gamma_1\gamma_2 - \gamma_2\gamma_1 \in \tfrac{1}{\ell}R^0 \tag{10.2}$$

so that

$$2\gamma_1\gamma_2 = \alpha + \beta,$$

$$4n^2 = q(\gamma_1\gamma_2) = \alpha^2 + q(\beta).$$

Since $\gamma_1 = (\alpha+\beta)/2\gamma_2$, it suffices to count the number of possible pairs (α, β). The pairs with $q(\beta) \neq 0$ may be counted as before after noting that the restriction of q to the orthogonal complement of γ_1 is nondegenerate. There are at most two pairs with $\beta = 0$, since then $\alpha = \pm 2n$. We thereby reduce to counting the number of pairs for which

$$q(\beta) = 0, \quad \beta \neq 0.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Recall the notation of §4.3.1. We note that for each cusp $\mathfrak{a}$, we may and shall choose $\sigma_{\mathfrak{a}} \in \Gamma_0(N)\tau_t$ for some $t|N$. With this choice, we have

$$\sigma_{\mathfrak{a}}^{-1} R \sigma_{\mathfrak{a}} = \begin{pmatrix} \mathbb{Z} & w_{\mathfrak{a}}\mathbb{Z} \\ \frac{N}{w_{\mathfrak{a}}}\mathbb{Z} & \mathbb{Z} \end{pmatrix}, \tag{10.3}$$

where $w_{\mathfrak{a}}$ is the cusp width of the cusp $\mathfrak{a}$. This is easily verified locally. We further introduce the following notation: For $\mathfrak{a} \in \mathbb{P}^1(\mathbb{Z})$ and $\kappa \in B^0$, we set $\kappa^{\mathfrak{a}} := \sigma_{\mathfrak{a}}^{-1}\kappa\sigma_{\mathfrak{a}}$.

We observe, by Equation (10.2), that $\ell\beta$ is a nonzero element of $R^0 \subseteq \mathrm{Mat}_{2\times 2}(\mathbb{Z})^0$. There is thus a unique (up to sign) primitive element β_0 of $\mathrm{Mat}_{2\times 2}(\mathbb{Z})^0$ that generates $\mathbb{Q}\ell\beta \cap \mathrm{Mat}_{2\times 2}(\mathbb{Z})^0$. We then have $\beta \in \frac{1}{\ell}\mathbb{Z}\beta_0$.

We may and shall choose $\mathfrak{a}$ so that $\beta_0^{\mathfrak{a}} = \left(\begin{smallmatrix} 0 & \pm 1 \\ 0 & 0 \end{smallmatrix}\right)$. Note that β (equivalently, β_0) is orthogonal not only to γ_1 (as was used in the original argument), but also to γ_2. From this, we deduce that

$$\gamma_2^{\mathfrak{a}} = \begin{pmatrix} a & b \\ 0 & -a \end{pmatrix}$$

for some $a \in \mathbb{Z}$ and $b \in \frac{1}{\ell}\mathbb{Z}$. We have $n = q(\gamma_2) = -a^2$, which shows that $-n$ must be a square and also that there are at most two possibilities for a. It remains to verify that the number of possibilities for b is $O(1 + \delta^{1/2}T\ell H(g))$. We will show in fact that the number of b is

$$O(1 + \delta^{1/2}T\ell y_{\mathfrak{a}}/w_{\mathfrak{a}}),$$

where $\sigma_{\mathfrak{a}}^{-1} z = z_{\mathfrak{a}} = x_{\mathfrak{a}} + iy_{\mathfrak{a}}$. To see this, observe first that the condition $\gamma_2 \in R(\ell)^0 \subseteq \frac{1}{\ell}R^0 \Leftrightarrow \gamma_2^{\mathfrak{a}} \in \sigma_{\mathfrak{a}}^{-1}R(\ell)^0\sigma_{\mathfrak{a}} \subseteq \frac{1}{\ell}\sigma_{\mathfrak{a}}^{-1}R^0\sigma_{\mathfrak{a}}$ yields the congruence $b \equiv 0 \ (\frac{w_{\mathfrak{a}}}{\ell})$; see Equation (10.3). The condition $\gamma_2 \in g\Omega(\delta, T)g^{-1}$ may be restated as

$$\gamma_2^{\mathfrak{a}} \in \sigma_{\mathfrak{a}}^{-1} g \Omega(\delta, T)(\sigma_{\mathfrak{a}}^{-1} g)^{-1}.$$

Let g' be an upper-triangular element of $G(\mathbb{R})$ for which $g' \cdot i = \sigma_{\mathfrak{a}}^{-1} g i = \sigma_{\mathfrak{a}}^{-1} z = z_{\mathfrak{a}}$. Then, by the K_∞ invariance of $\Omega(\delta, T)$ on the left and right, we have $\gamma_2^{\mathfrak{a}} \in g'\Omega(\delta, T)(g')^{-1}$. We compute

$$(g')^{-1}\gamma_2^{\mathfrak{a}} g' = \begin{pmatrix} a & y_{\mathfrak{a}}^{-1}(b + 2ax_{\mathfrak{a}}) \\ 0 & -a \end{pmatrix} \in \Omega(\delta, T).$$

This last condition forces b to lie in an interval of length $O(\delta^{1/2}Ty_{\mathfrak{a}})$. We thereby obtain the required bound for the number of possible b's. □

10.5. Proof of Theorem 2.6

We may split the set $\Omega(\delta, T)$ into $\Omega(1/16, 4\delta^{\frac{1}{2}}T)$ and the dyadic sets which are comprised of the elements $[a, b, c] + d \in B_\infty$ for which

$$\tfrac{1}{2}\delta_j T^2 \le a^2 + b^2 + c^2 + d^2 \le \delta_j T^2, \quad b^2 + c^2 \le \delta T^2,$$

for some $16\delta \le \delta_j \le 1$. We note that these are contained in $\Omega(\delta\delta_j^{-1}, \delta_j^{\frac{1}{2}}T)$. In order for the dyadic sets to contain an element of trace 0 and norm n, one must have $|n| \asymp a^2 \asymp \delta_j T^2$. Hence, if we apply Propositions 10.4 and 10.6, we get

$$|g^{-1}R(\ell)^0 g \cap \Omega(\delta\delta_j^{-1}, \delta_j^{\frac{1}{2}}T) \cap \det^{-1}(\{n\})| \prec 1 + \ell\delta^{\frac{1}{2}}H(g)T + \frac{\ell^2}{d_B N}\delta T^2,$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

where we used $|n| \asymp \delta_J T^2$, and

$$|g^{-1}R(\ell)^0 g \cap \Omega(\tfrac{1}{16}, 4\delta^{\frac{1}{2}}T) \cap \det^{-1}(\{n\})| \prec 1 + \ell\delta^{\frac{1}{2}}H(g)T + \frac{\ell^2}{d_B N}\delta T^2.$$

If δ happens to be very small, say, if $\delta \leq (16 d_B N T^2)^{-1}$, then it suffices to consider only $\delta_J \geq (d_B N T^2)^{-1}$ and the final set $\Omega(\delta d_B N T^2, (d_B N)^{-\frac{1}{2}})$. This avoids a factor $\delta^{-o(1)}$ for a too small δ. We conclude Theorem 2.6.

A. The theta lift

In this section, we use the theta correspondence for the reductive dual pair $(\mathrm{O}_{\mathrm{det}}, \mathrm{SL}_2)$ to derive the necessary properties of the theta kernels in use. The group $\mathrm{O}_{\mathrm{det}}$ is the affine algebraic group over $\mathbb{Q}$ representing the orthogonal group of $(B, \det)$. Recall that G is the linear algebraic group defined over $\mathbb{Q}$ satisfying $G(L) = L^\times \backslash (B \otimes L)^\times$ for any $\mathbb{Q}$-algebra L. Denote by M the algebraic group representing the functor

$$M(L) = {}_{\Delta(L^\times)}\backslash \left\{(g_1, g_2) \in (B \otimes L)^\times \times (B \otimes L)^\times : \ \det g_1 = \det g_2\right\},$$

for any $\mathbb{Q}$-algebra L. Then, M is defined over $\mathbb{Q}$ and it is isomorphic to the special orthogonal group $\mathrm{SO}_{\mathrm{det}}$ via the action $(g_1, g_2).x = g_1 x g_2^{-1}$. We also define the algebraic group G' over $\mathbb{Q}$ to be the simply connected form of G, that is, $G'(L) = \mathrm{SL}_1(B \otimes L)$ for any $\mathbb{Q}$-algebra L. The natural map $G' \times G' \to M$ is an isogeny. The left-hand side is the simply connected form, that is, the Spin group, and the right-hand side the adjoint one.

The determinant map provides two exact sequences

$$1 \to \prod_{v \leq \infty} \langle (I, -I) \rangle \to M(\mathbb{A}) \xrightarrow{\iota^{(1)}} (G \times G)(\mathbb{A}) \xrightarrow{\det(\bullet)} {}_{\mathbb{A}^{\times 2}}\backslash \mathbb{A}^\times \to 1,$$

$$1 \to \prod_{v \leq \infty} \langle (-I, -I) \rangle \to (G' \times G')(\mathbb{A}) \xrightarrow{\iota'} M(\mathbb{A}) \xrightarrow{\det} {}_{\mathbb{A}^{\times 2}}\backslash \mathbb{A}^\times \to 1.$$

A.1. Restriction of automorphic representations

We would like to understand the behavior of irreducible cuspidal representation under pull-back by $\iota^{(1)}$ and ι'. We proceed to discuss some generalities that apply to these isogenies. Let H be a semisimple algebraic group defined over $\mathbb{Q}$, and fix a maximal compact open subgroup $K_f = \prod_{v<\infty} K_v < H(\mathbb{A}_f)$. Let $K_\infty < H(\mathbb{R})$ a maximal compact real subgroup and set $K = K_\infty K_f$. For an automorphic representation Π of $H(\mathbb{A})$, we denote by $\Pi^\infty \subset \Pi$ the dense subset of K-finite vectors. That is, every $v \in \Pi^\infty$ is invariant under a finite-index subgroup of K_f and its K_∞-orbit spans a finite-dimensional subspace. Then, Π^∞ is an admissible $H(\mathbb{A})$ representation.

Assume $\jmath : H' \to H$ is a homomorphism of algebraic groups satisfying the following conditions:

(i) $\ker \jmath$ is a finite central subgroup of H',
(ii) $\Im \jmath$ is a normal subgroup of H and $\operatorname{coker} \jmath$ is a finite abelian group.
(iii) $H(\mathbb{Q}_p) = K_p \jmath(H'(\mathbb{Q}_p))$ for almost all primes p.

In particular, the Lie algebras of H and H' are isomorphic, hence H' is semisimple. These assumptions are satisfied when $\jmath$ is an isogeny of semisimple algebraic groups and for the inclusion map $\mathrm{SO}_{\mathrm{det}} \to \mathrm{O}_{\mathrm{det}}$. The group $H_{\mathrm{char}} = H(\mathbb{Q})\backslash H(\mathbb{A})/\jmath(H'(\mathbb{A}))$ is a compact abelian group, that is often infinite. In particular, $\jmath(H'(\mathbb{A}))$-orbits on $[H]$ can have measure zero and the operation of restricting a function in $L^2([H])$ to an orbit of $\jmath(H'(\mathbb{A}))$ is ill-defined. Nevertheless, we have the following.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Lemma A.1. *The pullback* $\jmath^*\colon \mathrm{Res}^{H(\mathbb{A})}_{H'(\mathbb{A})} L^2([H])^\infty \to L^2([H'])^\infty$ *is a well-defined operator that restricts to an intertwining operator of the cuspidal spectrum* $\jmath^*\colon \mathrm{Res}^{H(\mathbb{A})}_{H'(\mathbb{A})} L^2_{\mathrm{cusp}}([H])^\infty \to L^2_{\mathrm{cusp}}([H'])^\infty$. *Moreover,* $f \in L^2([H])^\infty$ *is cuspidal if and only if for any class* $[h] \in H_{\mathrm{char}}$, *the vector* $\jmath^* R_h f$ *is cuspidal for some representative* $h \in [h] \subset H(\mathbb{A})$.

Remark. Notice that $\jmath^*$ does not preserve inner products.

Proof. Restricting to a $\jmath(H'(\mathbb{A}))$-orbit is a well-defined operation on $L^2([H])^\infty$ because every vector $v_0 \in L^2([H])^\infty$ is invariant under some compact-open subgoup $K_0 < H(\mathbb{A}_f)$ and H_{char}/K_0 is finite. We now show that the L^2-norm of $\jmath^* v_0$ is finite. The push-forward of the probability Haar measure on $[H]$ to H_{char} is invariant under the action of $H(\mathbb{A})$, hence it is the probability Haar measure on H_{char}. If we disintegrate the Haar measure on $[H]$ under the factor map $[H] \to H_{\mathrm{char}}$, then the atoms are exactly the $\jmath(H'(\mathbb{A}))$-orbits and the conditional measure on a.e. atom is $\jmath(H'(\mathbb{A}))$-invariant, hence it is the push-forward of the probability Haar measure on $[H']$ to the atom. We can now deduce that

$$\|v_0\|_2^2 = |H_{\mathrm{char}}/K_0|^{-1} \sum_{h \in H_{\mathrm{char}}/K_0} \|\jmath^*(R_h v_0)\|^2_{2,[H']}.$$

Hence, $\|\jmath^*(v_0)\|_{2,[H']} \le \sqrt{|H_{\mathrm{char}}/K_0|}\|v_0\|_2 < \infty$.

We show next that the image of a cuspidal vector is cuspidal. Fix $v_0 \in L^2([H])^\infty$. Let $P < H'$ be a parabolic subgroup defined over $\mathbb{Q}$, and let N_P be its unipotent radical. The kernel $\ker \jmath$ is a central subgroup, hence it is diagonalizable and its intersection with N_P is trivial. Then, $\jmath \restriction_{N_P}$ is an isomorphism onto its image $N_{\tilde{P}}$, which is the unipotent radical of a parabolic $\tilde{P} < H$. Specifically, $\tilde{P}$ is the parabolic associated to the same root data as P. For every $g \in H'(\mathbb{A})$, we have, writing c_P and $c_{\tilde{P}}$ for the maps assigning to a function its corresponding constant term,

$$c_P \jmath^* v_0(g) = \int_{[N_P]} \jmath^*(v_0)(ng)\,\mathrm{d}n = \int_{[N_{\tilde{P}}]} v_0(n\jmath(g))\,\mathrm{d}n = c_{\tilde{P}} v_0(\jmath(g)).$$

Hence, the constant term of the push-forward of a cuspidal vector vanishes. This formula also establishes the last claim. □

Next, we describe the transformation of the Haar measure. If $K < H(\mathbb{A})$ is a compact subgroup, we denote by $[H]_K$ the double quotient $H(\mathbb{Q})\backslash H(\mathbb{A})/K$.

Lemma A.2. *Fix compact open subgroups* $K'_f < H'(\mathbb{A}_f)$ *and* $K_f < H(\mathbb{A}_f)$ *satisfying* $\jmath^{-1}(K_f) = K'_f$. *Assume the following conditions:*

(i) $H_{\mathrm{char}}/K_f = 1$.
(ii) *The preimage of* $K_f \bmod \jmath(H'(\mathbb{A}))$ *under the map* $\mathrm{coker}\,\jmath(\mathbb{Q}) \to \mathrm{coker}\,\jmath(\mathbb{A})$ *is trivial.*
(iii) $\ker \jmath \restriction_{H'(\mathbb{A})} < Z_{H'}(\mathbb{Q}) \cdot K'_f$.

Then, the induced map $\jmath\colon [H']_{K'_f} \to [H]_{K_f}$ *is a homeomorphism and an isomorphism of Borel measure spaces when each space is endowed with the respective probability Haar measure.*

Proof. To show the map $[H']_{K'_f} \to [H]_{K_f}$ is surjective, we need to find for every $x \in H(\mathbb{A})$ elements $\gamma \in H(\mathbb{Q})$, $k \in K_f$ and $x' \in H'(\mathbb{A})$ such that $x = \gamma\jmath(x')k$. Equivalently, we need the class of $[\gamma^{-1}xk^{-1}] = [xk^{-1}]$ in H_{char} to be trivial and this follows from the assumption that $H_{\mathrm{char}}/K_f = 1$.

To verify the map is injective, we consider $x'_1, x'_2 \in H'(\mathbb{A})$ satisfying $\jmath(x'_1) = \gamma\jmath(x'_2)k$, with $\gamma \in H(\mathbb{Q})$ and $k \in K_f$. We first demonstrate that $\gamma \in \jmath(H'(\mathbb{Q}))$. Because $\gamma = \jmath(x'_{1,\infty})k^{-1}\jmath({x'_{2,\infty}}^{-1})$ and $\mathrm{coker}\,\jmath$ is abelian, we see that the class of γ in $\mathrm{coker}\,\jmath(\mathbb{A}) = \jmath(H'(\mathbb{A}))\backslash H(\mathbb{A})$ is the same as the class of k. The second assumption then implies that the class of γ in $\mathrm{coker}\,\jmath(\mathbb{Q}) = \jmath(H'(\mathbb{Q}))\backslash H(\mathbb{Q})$ is trivial as claimed. Hence, $\gamma = \jmath(\gamma_0)$ for some $\gamma_0 \in H'(\mathbb{Q})$. We can now write $k = \jmath({x'_2}^{-1}\gamma_0^{-1}x'_1) \in \jmath(H'(\mathbb{A})) \cap K_f$. Because we assumed $\jmath^{-1}(K_f) = K'_f$, we can write $k = \jmath(k')$ for $k' \in K'_f$ and $\jmath({x'_1}^{-1}\gamma_0 x'_2 k') = 1$.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Thus, $x_1'^{-1}\gamma_0 x_2' k' = z k_0'$ for some z in the center of $H'(\mathbb{Q})$ and $k_0' \in K_f'$. Because the center of $H'(\mathbb{Q})$ is contained in the center of $H'(\mathbb{A})$, we deduce $x_1' = z^{-1}\gamma_0 x_2' k' k_0'^{-1} \in H'(\mathbb{Q}) x_2' K_f'$ as required.

We have established that $\jmath\colon [H']_{K_f'} \to [H]_{K_f}$ is a continuous bijection. It is a homeomorphism because it is also a smooth function between two real manifolds with everywhere nonvanishing differential. The probability Haar measure on $[H]_{K_f}$ is the unique $H(\mathbb{R})$-invariant Borel probability measure that gives equal mass to each of the finitely many $H(\mathbb{R})$-orbits. The same holds for $[H']_{K_f'}$ and $H'(\mathbb{R})$. Using this characterization, it is easy to check that the push-forward of the probability Haar measure from $[H]_{K_f}$ to $[H']_{K_f'}$ under $\jmath^{-1}$ is the probability Haar measure. □

A.2. The theta transfer

We now revert to our setting of interest, as described at the start of §A. Recall that $R \subset B$ is an Eichler order. For a finite rational place v, we define $R_v = R \otimes_{\mathbb{Z}} \mathbb{Z}_v$ and $\widetilde{K}_{R_v} = R_v^\times$. Define K_{R_v} to be the image of $\widetilde{K}_{R_v}$ under the map $B_v^\times \to G(\mathbb{Q}_v)$. Finally, set $\widetilde{K}_R = \prod_{v<\infty} \widetilde{K}_{R_v}$, $K_R = \prod_{v<\infty} K_{R_v}$, and let K_M denote the preimage of $K_R \times K_R$ under $\iota^{(1)}$. Then, K_R is a compact and open subgroup of $G(\mathbb{A}_f)$. We also assume that R is of squarefree level N (see §2.1).

We verify now that the hypotheses of Lemma A.2 hold for both of the maps

$$G' \times G' \xrightarrow{\iota'} M \xrightarrow{\iota^{(1)}} G \times G.$$

Indeed:

(i) We have $\det\left(\begin{smallmatrix}\bullet & \\ & \bullet\end{smallmatrix}\right)(K_R \times K_R) = \widehat{\mathbb{Z}}^\times$ and $\det(K_M) = \widehat{\mathbb{Z}}^\times$. Because $\mathbb{Q}^\times \mathbb{A}^{\times 2} \backslash \mathbb{A}^\times / \widehat{\mathbb{Z}}^\times \cong 1$ the equality $(G \times G)_{\mathrm{char}}/(K_R \times K_R) = M_{\mathrm{char}}/K_M = 1$ holds.

(ii) This condition is easy to verify by applying the maps $\det\left(\begin{smallmatrix}\bullet & \\ & \bullet\end{smallmatrix}\right)$ and det, and the fact that a rational number is a square if and only if it is positive and has even valuation at each finite place.

(iii) Consider the case $\jmath = \iota^{(1)}$. The last condition can be checked locally at each finite place to see that $(I, -I)_v \in (K_{R_v} \times K_{R_v})^{(1)}$ for all $v < \infty$. At the Archimedean place, we use the diagonal embedding of $(I, -I)$ in $M(\mathbb{A})$ to arrive at $(I, -I)_\infty \in Z_M(\mathbb{Q}) \cdot K_M$. The argument for $\jmath = \iota'$ follows mutatis mutandis.

Lemma A.2 now implies that the following maps are measure preserving homeomorphisms.

$$[G' \times G']_{K_R^1 \times K_R^1} \xrightarrow{\iota'} [M]_{K_M} \xrightarrow{\iota^{(1)}} [G \times G]_{K_R \times K_R}. \tag{A.1}$$

We get isomorphisms of Hilbert spaces

$$L^2([G' \times G'])^{K_R^1 \times K_R^1} \xleftarrow[\sim]{\iota'^*} L^2([M])^{K_M} \xleftarrow[\sim]{\iota^{(1)^*}} L^2([G \times G])^{K_R \times K_R}. \tag{A.2}$$

By Lemma A.1, these restrict to isomorphisms of the respective spaces of cusp forms.

Set $\tilde{U}_0(p^n) = \left(\begin{smallmatrix}\mathbb{Z}_p & \mathbb{Z}_p \\ p^n\mathbb{Z}_p & \mathbb{Z}_p\end{smallmatrix}\right) \cap \mathrm{GL}_2(\mathbb{Z}_p)$ – a compact and open subgroup of $\mathrm{GL}_2(\mathbb{Q}_p)$. Let $\tilde{U}_R = \prod_p \tilde{U}_p < \mathrm{GL}_2(\mathbb{A}_f)$ to be defined by $\tilde{U}_p = \tilde{U}_0(1) = \mathrm{GL}_2(\mathbb{Z}_p)$ for all primes p where G is unramified and R_p is maximal. If G ramifies at p or R_p is not maximal, then define $\tilde{U}_p = \tilde{U}_0(p)$. Note that we assume that R has squarefree level. Finally, set $U_R^1 = \tilde{U}_R \cap \mathrm{SL}_2(\mathbb{A}_f)$ and let U_R be the projection of $\tilde{U}_R$ to $\mathrm{PGL}_2(\mathbb{A}_f)$. Similarly to the previous discussion, the natural map $\iota_0\colon [\mathrm{SL}_2]_{U_R^1} \to [\mathrm{PGL}_2]_{U_R}$ is a homeomorphism that sends the probability Haar measure on the left-hand side to the probability Haar measure on the right-hand side. This induces an isomorphism of Hilbert spaces

$$L^2([\mathrm{SL}_2])^{U_R^1} \xleftarrow[\sim]{\iota_0^*} L^2([\mathrm{PGL}_2])^{U_R}, \tag{A.3}$$

that descends to an isomorphism of the cuspidal subspaces.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Recall that $B_p = B \otimes \mathbb{Q}_p$. We also denote $B_\infty := B \otimes \mathbb{R}$. We denote by ρ the Weil representation of the reductive dual pair $(\mathrm{O}_{\det}, \mathrm{SL}_2)$ associated to the quadratic space $(B, \det)$. We refrain at the moment from specifying the exact space of test functions on $B_{\mathbb{A}} := B_\infty \times \prod_p' B_p$ on which we let ρ act. If $\Phi\colon B_{\mathbb{A}} \to \mathbb{C}$ is a test function, then the group $M(\mathbb{A}) \cong \mathrm{SO}_{\det}$ acts by determinant preserving transformations, $\rho(l, r; e).\Phi(x) = \Phi(l^{-1}xr)$ and the action of the group $\mathrm{SL}_2(\mathbb{A})$ is described in [Wei64, Shi72]. Specifically, the definition of the $\mathrm{SL}_2(\mathbb{A})$-action depends on a global character $\psi\colon \mathbb{Q}\backslash\mathbb{A} \to \mathbb{C}$. We fix $\psi = \prod_v \psi_v$ with ψ_v everywhere unramified and $\psi_\infty(x) = \exp(2\pi i x)$.

Let $\Phi = \prod_v \Phi_v\colon B_{\mathbb{A}} \to \mathbb{C}$ be a test function with $\Phi_\infty\colon B_\infty \to \mathbb{C}$ the Bergman test function from [KS20, §6] or a Schwartz function. Assume for $v < \infty$ that Φ_v is Schwartz–Bruhat and that $\Phi_v = \mathbb{1}_{R_v}$ for almost all v. If Φ_∞ is the Bergman test function, we let ρ act on the space of functions defined in [KS20, §3], otherwise we let ρ act on the space of Schwartz–Bruhat functions as usual. The theta kernel associated to Φ is the function $\Theta_\Phi\colon M(\mathbb{A}) \times \mathrm{SL}_2(\mathbb{A}) \to \mathbb{C}$ defined by

$$\Theta_\Phi(l, r; s) = \sum_{\xi \in B} (\rho(l, r; s).\Phi)(\xi). \tag{A.4}$$

The series defining $\Theta_\Phi(l, r; s)$ is absolutely convergent, [KS20, §3.6], and is of moderate growth on $M(\mathbb{A}) \times \mathrm{SL}_2(\mathbb{A})$, [RS75]. Moreover, it is $M(\mathbb{Q}) \times \mathrm{SL}_2(\mathbb{Q})$ invariant on the left, cf. [Wei64], [Shi72, Proposition 1], [KS20, §3.6].

Definition A.3. Let $\varphi, \varphi' \in L^2_{\mathrm{cusp}}([G])^\infty$ and $\varphi^* \in L^2_{\mathrm{cusp}}([\mathrm{SL}_2])^\infty$. Fix a test function Φ as above. Then, the theta transfer of $\varphi \otimes \varphi'$ and φ^* relative to Φ is defined by

$$\begin{aligned}(\varphi \otimes \varphi')_\Phi(s) &= \int_{[M]} \Theta_\Phi(l, r; s)\varphi(l)\varphi'(r)\,\mathrm{d}(l, r)\\ &= \int_{[M]} \Theta_\Phi(l, r; s)\iota^{(1)^*}(\varphi \otimes \varphi')(l, r)\,\mathrm{d}(l, r),\\ \varphi^{*\Phi}(l, r) &= \int_{[\mathrm{SL}_2]} \Theta_\Phi(l, r; s)\varphi^*(s)\,\mathrm{d}s.\end{aligned}$$

The former is a complex-valued function on $\mathrm{SL}_2(\mathbb{A})$, and the latter is a function on $M(\mathbb{A})$. Both integrals converge absolutely because Θ_Φ is of moderate growth and φ,φ', φ^* are of rapid decay. By abuse of notation, we will also denote

$$\varphi_\Phi = (\varphi \otimes \overline{\varphi})_\Phi.$$

Note also that the modularity of the theta kernel Θ_Φ implies that $(\varphi \otimes \varphi')_\Phi$ is left $\mathrm{SL}_2(\mathbb{Q})$-invariant, and $\varphi^{*\Phi}$ is left $M(\mathbb{Q})$-invariant.

If, moreover, φ is K_R-invariant and Φ is both left and right $\widetilde{K}_R$-invariant then, because the maps in Equation (A.1) are isomorphisms of measure spaces, we have

$$\varphi_\Phi(s) = \int_{[G']} \int_{[G']} \Theta_\Phi(l, r; s)\varphi(l)\overline{\varphi(r)}\,\mathrm{d}l\,\mathrm{d}r. \tag{A.5}$$

We will need the following lemma. It is mostly a corollary of [Ral84, Mœg97, KR94].

Lemma A.4. *Let $\varphi, \varphi' \in L^2_{\mathrm{cusp}}([G])^\infty$ and $\varphi^* \in L^2_{\mathrm{cusp}}([\mathrm{SL}_2])^\infty$. Assume that φ, φ' are K_R-invariant and that Φ is both left and right $\widetilde{K}_R$-invariant. Then, $(\varphi \otimes \varphi')_\Phi \in L^2_{\mathrm{cusp}}([\mathrm{SL}_2])$ and $\varphi^{*\Phi} \in L^2_{\mathrm{cusp}}([M])$.*

Proof. The fact that $\varphi^{*\Phi}$ is square-integrable and cuspidal is trivial whenever G is anisotropic. If $G \cong \mathrm{PGL}_2$ is split, then this follows from Rallis' tower property [Ral84, Mœg97] and the fact that the

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

theta transfer of any cuspidal automorphic representation of $\mathrm{SL}_2 \cong \mathrm{Sp}_2$ to the orthogonal group of the hyperbolic plane $\mathrm{O}(1,1)$ vanishes.[5]

That the lift of $\varphi \otimes \varphi'$ is cuspidal follows similarly, except that we need to use the theta transfer from $\mathrm{O}_{\mathrm{det}}$ to SL_2, that is, we need first to lift ${\iota^{(1)}}^* \varphi \otimes \varphi'$ to $\mathrm{O}_{\mathrm{det}}$. For that purpose, we use the homomorphism $\iota\colon M \to \mathrm{O}_{\mathrm{det}}$, which is the composition of the isomorphism $M \cong \mathrm{SO}_{\mathrm{det}}$ with the embedding $\mathrm{SO}_{\mathrm{det}} \hookrightarrow \mathrm{O}_{\mathrm{det}}$. This map satisfies the assumptions of §A.1. For every finite place v, let $\mathrm{O}_{\mathrm{det}}(R_v)$ to be the group of orthogonal transformations of B_v that send R_v to itself and define $\mathrm{O}_{\mathrm{det}}(\hat{R}) = \prod_{v<\infty} \mathrm{O}_{\mathrm{det}}(R_v)$. Then, the conditions of Lemma A.2 are easily verified and we deduce that the pull-back ι^* induces an isomorphism of $L^2([\mathrm{O}_{\mathrm{det}}])^{\mathrm{O}_{\mathrm{det}}(\hat{R})}$ and $L^2([M])^{K_M}$. We deduce from Lemma A.1 that $(\iota^*)^{-1}{\iota^{(1)}}^* \varphi \otimes \varphi'$ is cuspidal and

$$(\varphi \otimes \varphi')_\Phi(s) = \int_{[\mathrm{O}_{\mathrm{det}}]} \Theta_\Phi(\bullet; s)(\iota^*)^{-1}{\iota^{(1)}}^* \varphi \otimes \varphi' \, \mathrm{d}m_{\mathrm{O}_{\mathrm{det}}}. \tag{A.6}$$

Here, we have extended the definition of Θ_Φ in Equation (A.4) to $\mathrm{O}_{\mathrm{det}}(\mathbb{A}) \times \mathrm{SL}_2(\mathbb{A})$ in the obvious way. The integral in Equation (A.6) is a theta lift of a cuspidal function in $L^2([\mathrm{O}_{\mathrm{det}}])^\infty$ to $L^2([\mathrm{SL}_2])^\infty$. In this case, [Ral84] verifies that the theta lift of a cuspidal function to SL_2 is cuspidal. □

Lemma A.5. *Let $\varphi^* \in L^2_{\mathrm{cusp}}([\mathrm{SL}_2])^\infty$ and $\varphi, \varphi' \in L^2_{\mathrm{cusp}}([G])^\infty$. Assume that φ, φ' are K_R-invariant and that Φ is both left and right $\widetilde{K}_R$-invariant. Then*

$$\left\langle \varphi^{*\Phi}, {\iota^{(1)}}^* (\varphi \otimes \varphi') \right\rangle = \left\langle \varphi^*, (\varphi \otimes \varphi')_{\overline{\Phi}} \right\rangle.$$

Proof. This follows from Fubini and the fact that cusp forms are of rapid decay. □

Proposition A.6. *Assume $\varphi, \varphi' \in L^2_{\mathrm{cusp}}([G])^\infty$ are K_R-invariant and that Φ is both left and right $\widetilde{K}_R$-invariant. For $s \in \mathrm{GL}_2(\mathbb{A})$, define $s_1 = \left(\begin{smallmatrix} (\det s)^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right) s$. The Whittaker function of the theta lift satisfies*

$$W_{\iota_0^{*-1}(\varphi \otimes \varphi')_\Phi}(s) = |\det s|_{\mathbb{A}} \langle T_s^\Phi \varphi, \overline{\varphi'}(\bullet \alpha_{\det s}) \rangle_{[G']},$$

where

$$\left(T_s^\Phi \varphi\right)(r) = \begin{cases} \int_{G'(\mathbb{A})} \varphi(rl^{-1}) \left(\rho\left(s_1\right).\Phi\right)(l\alpha_{\det s}) \, \mathrm{d}l & \det s \in \det B_{\mathbb{A}} \\ 0 & \det s \notin \det B_{\mathbb{A}} \end{cases}$$

and $\alpha_{\det s} \in B_{\mathbb{A}}$ is any element satisfying $\det \alpha_{\det s} = \det s$. Moreover, we can replace the inner product in $L^2([G'])$ in the formula above by an inner product in $L^2([G])$.

Proof. First, observe

$$\iota_0^{*-1}(\varphi \otimes \varphi')_\Phi(s) = |\det s|_{\mathbb{A}} \int_{[G']} \int_{[G']} \sum_{\xi \in B} (\rho(s_1).\Phi)(l^{-1}\xi r \alpha_{\det s}) \varphi(l) \varphi'(r\alpha_{\det s}) \, \mathrm{d}l \, \mathrm{d}r.$$

Consider both sides of the first equality as functions on $\mathrm{GL}_2(\mathbb{A})^\dagger = \{x \in \mathrm{GL}_2(\mathbb{A})\colon \det x \in \det B_{\mathbb{A}}\}$. Set $\mathrm{GL}_2(\mathbb{Q})^\dagger = \{x \in \mathrm{GL}_2(\mathbb{Q})\colon \det x \in \det B\}$ and note that $\mathrm{GL}_2(\mathbb{Q})^\dagger \backslash \mathrm{GL}_2(\mathbb{A})^\dagger \cong \mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A})$ because $\mathrm{GL}_2(\mathbb{Q})\mathrm{GL}_2(\mathbb{A})^\dagger = \mathrm{GL}_2(\mathbb{A})$. The first equality then follows from Equation (A.5) by noticing that both sides are $\mathrm{GL}_2(\mathbb{Q})^\dagger$-invariant on the left, $Z_{\mathrm{GL}_2}(\mathbb{A})$-invariant, U_R-invariant on the right and coincide on $\mathrm{SL}_2(\mathbb{A})$. A standard unfolding argument in the l variable (see [Shi72, KS20]) applied to the last expression shows for $\det s \in \det B_{\mathbb{A}}$

[5] This is simple to deduce from the fact that a theta series arising from the two-dimensional isotropic quadratic form is a pseudo-Eisenstein series; see, for example, [Nel21].

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$$W_{\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi}(s) = |\det s|_{\mathbb{A}} \int_{[G']} \int_{G'(\mathbb{A})} (\rho(s_1).\Phi)(l^{-1}r\alpha_{\det s})\varphi(l)\varphi'(r\alpha_{\det s})\,\mathrm{d}l\,\mathrm{d}r,$$

and $W_{\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi}(s) = 0$ if $\det s \notin \det B_{\mathbb{A}}$. Using the change of variables $l^{-1}r \mapsto l$, we can write

$$T_s^\Phi\varphi(r) = \int_{G'(\mathbb{A})} (\rho(s_1).\Phi)(l^{-1}r\alpha_{\det s})\varphi(l)\,\mathrm{d}l = \int_{G'(\mathbb{A})} \varphi(rl^{-1})(\rho(s_1).\Phi)(l\alpha_{\det s})\,\mathrm{d}l.$$

This establishes the first formula.

The last formula extends naturally to any $r \in G(\mathbb{A})$, and the result is left $G(\mathbb{Q})$-invariant. If $\det s \notin \det G(\mathbb{A})$, then we extend $T_s^\Phi\varphi$ by zero to $G(\mathbb{A})$. For any $k \in K_R$, using the invariance properties of Φ and φ, we can apply the change of variables $kl\alpha_{\det s}^{-1}k^{-1}\alpha_{\det s} \mapsto l$ to see that $T_s^\Phi\varphi$ is right $\alpha_{\det s}^{-1}K_R\alpha_{\det s}$-invariant. The same holds for $\varphi'(\bullet\alpha_{\det s})$. Because the groups $\alpha_{\det s}K_R\alpha_{\det s}^{-1}$, $\alpha_{\det s}K_R^1\alpha_{\det s}^{-1}$ and the isogeny $G' \to G$ satisfy the assumptions of Lemma A.2, we see that we can replace the inner product in $[G']$ by an inner product in $[G]$. □

Corollary A.7. *Assume Φ is both left and right $\widetilde{K}_R$-invariant. Let $\varphi, \varphi' \in L^2_{\mathrm{cusp}}([G])^\infty$ be K_R-invariant, and denote by π and π' the cuspidal automorphic representations generated by φ and φ', respectively. If π is disjoint from $\pi'^\vee$, then $(\varphi\otimes\varphi')_\Phi = 0$.*

Proof. In this case, we see that $\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi$ is cuspidal with a vanishing Whittaker function. □

Corollary A.8. *Assume Φ is invariant under the conjugation action of K_R. Let $\pi \subset L^2_{\mathrm{cusp}}([G])^\infty$ be an irreducible representation. Assume $\varphi, \overline{\varphi'} \in \pi$ are K_R-invariant decomposable vectors, that is, $\varphi, \overline{\varphi'} \mapsto \otimes\varphi_v, \otimes\overline{\varphi'_v}$ in $\pi \cong \bigotimes' \pi_v$. Then*

$$W_{\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi}(s) = V^{-1}|\det s|_{\mathbb{A}} \prod_v \left\langle \varphi_v \star_{G'(\mathbb{Q}_v)} (\rho(s_{v,1}).\Phi_v)(\bullet\alpha_{\det s,v}), \pi_v(\alpha_{\det s,v}).\overline{\varphi'_v} \right\rangle,$$

where V is the volume of the (possibly disconnected) real manifold $G'(\mathbb{Q})\backslash G'(\mathbb{A})/K_f^1$ with respect to the Haar measure of $G'(\mathbb{R})$ (see §4.2), and we normalize the Haar measure on $G'(\mathbb{Q}_p)$ so that R_p^1 has unit volume.

Proof. This follows directly from Proposition A.6. The constant V^{-1} arises as a measure normalization constant. Specifically, denote by $m_{G'(\mathbb{Q}_p)}$ the Haar measure on $G'(\mathbb{Q}_p)$ satisfying $m_{G'(\mathbb{Q})_p}(R_p^1) = 1$. The Haar measure on $G'(\mathbb{A})$ satisfies $m_{G'} = c\bigotimes_v m_{G'(\mathbb{Q}_v)}$ with some measure normalization constant $c > 0$. Specifically, this equality holds for linear combinations of standard test functions $\prod_v f_v$ with $f_p = \mathbb{1}_{R_p^1}$ for a.e. p. To compute c, we write $G'(\mathbb{Q})\backslash G'(\mathbb{A})/G'(\mathbb{R})K_f^1 = \{\delta_1, \ldots, \delta_h\}$ and denote by $\mathcal{F}_i \subset G'(\mathbb{R})$ a fundamental domain for the right action of $\Gamma_i = G(\mathbb{Q}) \cap \delta_i K_R^1 \delta_i^{-1}$ on $G'(\mathbb{R})$. Then $\bigsqcup_{i=1}^h \delta_i\mathcal{F}_iK_R^1 \subset G(\mathbb{A})$ is a fundamental domain for the left action of $G(\mathbb{Q})$ on $G(\mathbb{A})$, and we deduce

$$1 = m_G\left(\bigsqcup_{i=1}^h \delta_i\mathcal{F}_iK_R^1\right) = c\sum_{i=1}^h m_{G'(\mathbb{R})}(\mathcal{F}_i) = cV.$$

□

Lemma A.9. *Fix $\Phi_v = \mathbb{1}_{R_v}$ for all finite places $v < \infty$, and let Φ_∞ be a Schwartz function or the Bergman test function from [KS20]. Fix $s = (\iota_0(s_\infty), u_f)$ with $s_\infty \in \mathrm{SL}_2(\mathbb{R})$ and $u_f \in U_R$. Assume that $\varphi \in L^2_{\mathrm{cusp}}([G])^\infty$ has weight m and is a K_R-invariant newvector in an irreducible cuspidal automorphic representation π. If $\rho(\bullet,\bullet;s)\Phi_\infty$ is $K_\infty \times K_\infty$-isotypical of weight $(-m, m)$, then*

$$\begin{aligned} W_{\iota_0^{*-1}\varphi_\Phi}(s) &= \frac{\|\varphi\|_2^2}{V}\mathrm{Tr}\left(\mathrm{Res}^{G(\mathbb{R})}_{G'(\mathbb{R})}\pi_\infty\right)(\rho(s_\infty).\Phi_\infty \restriction_{G'(\mathbb{R})}) \\ &= \frac{\|\varphi\|_2^2}{V}\overline{\left\langle f_{\varphi_\infty,\varphi_\infty}, (\rho(s_\infty).\Phi_\infty)\right\rangle}_{G'(\mathbb{R})}, \end{aligned}$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

where $f_{\varphi_\infty,\varphi_\infty}(g) = \langle \pi(g).\varphi_\infty, \varphi_\infty\rangle$ *is the matrix coefficient attached to the Archimedean component of* φ *in* $\bigotimes_v' \pi_v$*, normalized so that* $\|\varphi_\infty\|_2 = 1$.

Proof. It is sufficient to establish the claim when $\|\varphi\|_2 = 1$. By [KS20, §4], the theta transfer φ_Φ is U_R^1-invariant, thus $\iota_0^{*-1}\varphi_\Phi$ is U_R-invariant and we can assume without loss of generality that $u_f = e$. Then $\det s = 1$ and we take $\alpha_{\det s} = e$.

The newvector φ decomposes as $\varphi \mapsto \otimes\varphi_v$ in $\pi \cong \bigotimes' \pi_v$. We normalize φ_∞ to have norm 1, then $\prod_p \|\varphi_p\|_2 = 1$ as well. We also normalize the measure on $G'(\mathbb{Q}_p)$ so that $R_p^1 = K_p^1$ has unit volume. Corollary A.8 now implies $W_{\iota_0^{*-1}\varphi_\Phi}(s) = V^{-1}\prod_v \left\langle \varphi_v \star_{G'(\mathbb{Q}_v)} (\rho(s_v).\Phi_v), \varphi_v \right\rangle$. For a finite place p, we have $s_p = e$ and φ_p is K_p-invariant, hence $\varphi_p \star \mathbb{1}_{K_p^1} = \varphi_p$. We conclude

$$W_{\iota_0^{*-1}\varphi_\Phi}(s) = V^{-1}\langle \varphi_\infty \star (\rho(s_\infty).\Phi_\infty) \restriction_{G'(\mathbb{R})}, \varphi_\infty\rangle.$$

This expression coincides with the trace if the convolution operator $\star_{G'(\mathbb{R})}(\rho(s_\infty).\Phi_\infty)$ annihilates the orthogonal complement to φ_∞ in π_∞. This follows from the facts that every K_∞-isotypical component of the admissible unitary representation π_∞ is at most one-dimensional and the assumption that $\rho(\bullet,\bullet;s)\Phi_\infty$ is $K_\infty \times K_\infty$-isotypical.

To show the formula in terms of a matrix coefficient, we use Fubini to deduce

$$\begin{aligned}
\langle \varphi_\infty \star (\rho(s_\infty).\Phi_\infty) \restriction_{G'(\mathbb{R})}, \varphi_\infty\rangle &= \int_{G'(\mathbb{R})}\int_{G(\mathbb{R})} \varphi_\infty(gh^{-1})(\rho(s_\infty).\Phi_\infty)(h)\overline{\varphi_\infty(g)}\,\mathrm{d}h\,\mathrm{d}g \\
&= \int_{G'(\mathbb{R})}\int_{G(\mathbb{R})} \varphi_\infty(g)\overline{\varphi_\infty(gh)}(\rho(s_\infty).\Phi_\infty)(h)\,\mathrm{d}g\,\mathrm{d}h \\
&= \left\langle \overline{f_{\varphi_\infty,\varphi_\infty}}, \overline{\rho(s_\infty).\Phi_\infty} \right\rangle_{G'(\mathbb{R})}.
\end{aligned}$$

The conditions of Fubini's theorem are satisfied because the test function $\Phi_\infty \restriction_{G'(\mathbb{R})}$ is in $L^q(G'(\mathbb{R}))$ for all $q \geq 1$ and $f_{\varphi_\infty,\varphi_\infty} \in L^p(G'(\mathbb{R}))$ for some $p \geq 2$. □

Proposition A.10. *Fix* $\Phi_v = \mathbb{1}_{R_v}$ *for all finite places* $v < \infty$*. Let* $\pi \subset L^2_{\mathrm{cusp}}([G])^\infty$ *be an irreducible cuspidal automorphic representation, and denote by* π^{JL} *its Jacquet–Langlands transfer to* $L^2_{\mathrm{cusp}}([\mathrm{PGL}_2])^\infty$*. In case G is split, we define* $\pi^{\mathrm{JL}} = \pi$*. Assume* $\varphi \in \pi$*,* $\varphi' \in \pi^\vee$ *are nonvanishing* K_R*-invariant vectors, then* $\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi \in \pi^{\mathrm{JL}}$.

Moreover, if Φ_∞ *is* $\rho(K_\infty, K_\infty; \mathrm{SO}_2(\mathbb{R}))$*-isotypical with weight* $(-m, m, \kappa)$*,* π *has conductor* K_R *and* φ *is a newvector of weight m, then either* φ_Φ *vanishes or* $\iota_0^{*-1}\varphi_\Phi$ *is a newvector of weight* κ *in the Jacquet–Langlands transfer* π^{JL}.

Proof. Any smooth vector in π^{K_R} is a linear combination of K_R-invariant factorizable vectors in the representation $\pi \cong \bigotimes_v' \pi_v$. Thus, we assume without loss of generality that φ and φ' are factorizable in π and $\pi^\vee$, respectively.

The function $\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi$ is cuspidal by Lemma A.4. Because $\varphi \mapsto \otimes_v\varphi_v$ and $\varphi' \mapsto \otimes_v\varphi_v'$ are factorizable, Corollary A.8 implies that the Whittaker function of $\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi$ decomposes into a product.

Assume $(\varphi\otimes\varphi')_\Phi$ does not vanish. Let S be a finite set of rational places containing the Archimedean place, all places where G ramifies, all places where π ramifies and all places where either φ_v or φ_v' is not spherical. Shimizu [Shi72] computes the local Whittaker function $|\det s|_v\langle \varphi_v \star_{G'(\mathbb{Q}_v)} \Phi_v(\bullet\alpha_{\det s,v}), \pi_v(\alpha_{\det s,v}).\overline{\varphi_v'}\rangle_{\pi_v}$ for every place $v \notin S$ and it coincides with the Whittaker function of a spherical newvector in π_v^{JL}. Hence, every irreducible component $\sigma \cong \bigotimes_v' \sigma_v$ of the representation generated by $\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi$ satisfies $\sigma_v \cong \pi_v^{\mathrm{JL}}$ for all $v \notin S$. Using the strong multiplicity one property for PGL_2, we deduce that $\sigma = \pi^{\mathrm{JL}}$ for every irreducible component σ as above. Hence, the representation generated by $\iota_0^{*-1}(\varphi\otimes\varphi')_\Phi$ is π^{JL}.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Assume next that π has conductor K_R and that φ is a newvector of weight m. Then, the assumption that Φ_∞ has weight $(-m, m, \kappa)$ implies that φ_Φ has weight κ. By [KS20, §4], the theta transfer φ_Φ is U_R^1-invariant. Because the conductor of the Jacquet–Langlands transfer is exactly U_R and $\iota_0^{*-1}\varphi_\Phi = \iota_0^{*-1}(\varphi \otimes \overline{\varphi})_\Phi \in \pi^{\mathrm{JL}}$, we deduce that $\iota_0^{*-1}\varphi_\Phi$ is a newvector as claimed. □

Corollary A.11. *Assume $\varphi, \varphi' \in L^2_{\mathrm{cusp}}([G])^\infty$ are K_R-invariant and generate disjoint automorphic cuspidal representations, then $\langle \varphi_\Phi, \varphi'_\Phi \rangle = 0$.*

Proof. The Jacquet–Langlands transfers of disjoint automorphic representations are disjoint. Hence, Proposition A.10 above implies that $\iota_0^{*-1}\varphi_\Phi$, $\iota_0^{*-1}\varphi'_\Phi$ generate mutually orthogonal subrepresentations of $L^2_{\mathrm{cusp}}([\mathrm{PGL}_2])^\infty$. □

Proposition A.12. *Fix $\Phi_v = \mathbb{1}_{R_v}$ for all finite places $v < \infty$, and assume that Φ_∞ is $\rho(K_\infty, K_\infty; \mathbf{SO}_2(\mathbb{R}))$-isotypical with weight $(-m, m, \kappa)$. Let $\pi \subset L^2_{\mathrm{cusp}}([G])^\infty$ be an irreducible cuspidal automorphic representation with conductor K_R. Assume $0 \neq \varphi \in \pi$ is a newvector of weight m. Then*

$$(\varphi_\Phi)^{\overline{\Phi}} = \left(\frac{\|\varphi_\Phi\|_2}{\|\varphi\|_2^2}\right)^2 \iota^{(1)^*}(\varphi \otimes \overline{\varphi}). \tag{A.7}$$

Proof. Assume that

$$(\varphi_\Phi)^{\overline{\Phi}} = \alpha \iota^{(1)^*}(\varphi \otimes \overline{\varphi}), \tag{A.8}$$

for some $\alpha \in \mathbb{C}$. Then, $\alpha\|\varphi\|_2^4 = \langle (\varphi_\Phi)^{\overline{\Phi}}, \iota^{(1)^*}(\varphi \otimes \overline{\varphi}) \rangle$ and Lemma A.5 implies that $\alpha = \|\varphi_\Phi\|_2^2/\|\varphi\|_2^4$.

Because $(\varphi_\Phi)^{\overline{\Phi}}$ is continuous and cuspidal by Lemma A.4, to establish Equation (A.8) it is enough to show that $\iota^{(1)^{*-1}}(\varphi_\Phi)^{\overline{\Phi}}$ is orthogonal to the orthogonal complement of $\mathbb{C}(\varphi \otimes \overline{\varphi})$ in $L^2_{\mathrm{cusp}}([G \times G])^{K_R \times K_R}$. Both $(\varphi \otimes \overline{\varphi})$ and $\iota^{(1)^{*-1}}(\varphi_\Phi)^{\overline{\Phi}}$ transform with weight $(m, -m)$ under $K_\infty \times K_\infty$. Hence, it is enough to check orthogonality in the $(m, -m)$ isotypical subspace

$$V_m = L^2_{\mathrm{cusp}}([G \times G])^{((K_\infty, m)\cdot K_R) \times ((K_\infty, -m)\cdot K_R)}.$$

Denote by V_m^0 the orthogonal compliment of $\mathbb{C}(\varphi \otimes \overline{\varphi})$ in V_m. We can choose an orthonormal basis for V_m^0 consisting of vectors $\psi \otimes \psi'$ with $\psi, \overline{\psi'} \in L^2_{\mathrm{cusp}}([G])^{(K_\infty, m)\cdot K_R}$ and ψ, ψ' generate irreducible cuspidal automorphic representations of $G(\mathbb{A})$. Because π has conductor K_R, either the representation generated by ψ is disjoint from π or the representation generated by ψ' is disjoint from $\pi^\vee$. Fix ψ, ψ' as above. We need to show $\left\langle \iota^{(1)^{*-1}}(\varphi_\Phi)^\Phi, \psi \otimes \psi' \right\rangle = 0$. Denote by σ, σ' the irreducible automorphic representations generated by ψ, ψ', respectively. We apply Lemma A.5 to deduce $\left\langle \iota^{(1)^{*-1}}(\varphi_\Phi)^{\overline{\Phi}}, \psi \otimes \psi' \right\rangle = \langle \varphi_\Phi, (\psi \otimes \psi')_\Phi \rangle$. If $\sigma' \neq \sigma^\vee$, then $(\psi \otimes \psi')_\Phi = 0$ by Corollary A.7. If $\sigma' = \sigma^\vee$, then σ is disjoint from π, and $(\iota_0^*)^{-1}(\psi \otimes \psi')_\Phi \in \sigma^{\mathrm{JL}}$ by Proposition A.10. The Jacquet–Langlands transfers of disjoint representations are disjoint. Hence, $\pi^{\mathrm{JL}} \perp \sigma^{\mathrm{JL}}$ and $\langle \varphi_\Phi, (\psi \otimes \psi')_\Phi \rangle = 0$ as claimed. □

A.3. Explicit theta kernels

Definition A.13. We now define the Archimedean test functions on B_∞ that give rise to the theta series from §5.1.1.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

$$\Phi_\infty^{-,k}(g) = X(g)^k e^{-2\pi P(g)},$$
$$\Phi_\infty^{-,\mathrm{hol}}(g) = \frac{k-1}{4\pi}\begin{cases}(\det g)^{k-1}\overline{X(g)}^{(-k)} e^{-2\pi \det g} & \det g > 0\\ 0 & \det g \le 0\end{cases},$$
$$\Phi_\infty^{+,m}(g) = (2m+1)(\det g)^m P_m\left(\frac{|X(g)|^2 - u(g)}{\det g}\right) e^{-2\pi \det g},$$
$$\Phi_\infty^{+,\mathrm{hol}}(g) = (k+1)X(g)^k e^{-2\pi \det g}.$$

The first two test functions are defined when $G(\mathbb{R})$ is split, and the last two are defined when $G(\mathbb{R})$ is ramified.

Lemma A.14. *Let Φ_∞ be one of the kernels in Definition A.13 above, and set $\kappa = k, k, 2m+2, k+2$ for the different kernels respectively. Then, $\rho(k_\theta).\Phi_\infty = e^{i\kappa\theta}\Phi_\infty$ for all $k_\theta \in \mathrm{SO}_2(\mathbb{R})$.*

Proof. This is verified by Vignéras' method [Vig77]. In all cases under consideration except $\Phi_\infty^{-,\mathrm{hol}}$, the test function is Schwartz, hence it is enough to check that Φ_∞ satisfies the partial differential equation in [KS20, §3.3] and then use Lemma 3.4, *op. cit.*. In case $\Phi_\infty = \Phi_\infty^{-,\mathrm{hol}}$, the test function is not Schwartz and a technical argument is required to circumvent this issue. This case is treated [KS20, §6]. We proceed to verify the three other cases.

Recall the notation $x = [a,b,c] + d \in B_\infty$ from §2.4.3. The Laplace operator with Fourier multiplier $-4\pi^2 \det(x)$ is then given by $\Delta = \frac{1}{4}(\frac{\partial^2}{\partial a^2} \mp (\frac{\partial^2}{\partial b^2} + \frac{\partial^2}{\partial c^2}) + \frac{\partial^2}{\partial d^2})$, where the sign is $-$ if B is indefinite and $+$ otherwise. The differential equation in §3.3 of [KS20] for the test function Φ_∞ is equivalent to

$$-\Delta\Phi(x) + (2\pi)^2 \det(x)\Phi(x) = 2\pi\kappa\Phi(x). \tag{A.9}$$

We note that for each of the remaining test functions, we may write $\Phi_\infty(x) = Q(x)e^{-2\pi P(x)}$, where $P(x) = a^2+b^2+c^2+d^2$ and Q a harmonic polynomial of homogeneous degree. For the first and last test function, this may be seen by a well-known criteria (c.f. [Iwa97, Thm 9.1]) noting that $[i,0,0]+1 \in B_\infty \otimes \mathbb{C}$ is an isotropic vector. For the third test function, this follows from [LPS87, p. 405]. With this in mind, we have $\Delta Q = 0$ and $(a\frac{\partial}{\partial a} + b\frac{\partial}{\partial b} + c\frac{\partial}{\partial c} + d\frac{\partial}{\partial d})Q = \deg(Q)Q$, which allows one to easily verify that Φ_∞ satisfies Equation (A.9) in the remaining cases. □

Proposition A.15. *Let $\Phi = \Phi_\infty \cdot \prod_{v<\infty} \mathbb{1}_{R_v}$, where Φ_∞ is any one of the test functions in Definition A.13 above. Set $\kappa = k, k, 2m+2, k+2$ for the different kernels, respectively. Denote by θ_g the matching classical theta function from §5.1.1. For $\tau \in \mathrm{SL}_2(\mathbb{Q})$, we denote by $(\tau)_\infty$ the image of τ in the Archimedean coordinate of $\mathrm{SL}_2(\mathbb{A})$. Then, for every $l \mid d_B N$ and $g \in G(\mathbb{A})$*

$$\Theta_\Phi(g,g;(\tau_\ell)_\infty s_\infty U_R^1) = \frac{\mu(\gcd(\ell, d_B))}{\ell}\theta_{g,\ell}(z)e^{i\kappa\theta},$$

where μ is the Möbius function, $s_\infty = \left(\begin{smallmatrix} y^{1/2} & xy^{-1/2} \\ 0 & y^{-1/2}\end{smallmatrix}\right)\left(\begin{smallmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{smallmatrix}\right)$ and $z = x+iy$, that is, $s_\infty.i = z$. Moreover, $\theta_{g,\ell}(z)$ is a $\Gamma_0(d_B N)$-invariant function on $\mathbb{H}$ of moderate growth at the cusps.

Proof. We already know Θ_Φ has weight κ in the Archimedean symplectic variable s_∞. Moreover, in [KS20, §3.5] it is shown that Θ_Φ is U_R^1 invariant. Denote by $(\tau_\ell)_f$ the diagonal image of τ_ℓ in $\mathrm{SL}_2(\mathbb{A}_f)$. The left $\mathrm{SL}_2(\mathbb{Q})$-invariance of the theta kernel implies $\Theta_\Phi(g,g;(\tau_\ell)_\infty s_\infty) = \Theta_\Phi(g,g;s_\infty(\tau_\ell)_f^{-1})$. For every prime $p \nmid d_B N$, we have $\tau_\ell \in \mathrm{SL}_2(\mathbb{Z}) \subset \mathrm{SL}(\mathbb{Z}_p) = U_p^1$. If $p \mid l$, then $\tau_\ell \equiv w \bmod p$, where $w = \left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$, hence $\tau_\ell \in wU_p^1$. If $p \mid \frac{d_B N}{l}$, then $\tau_\ell \equiv e \bmod p$, hence $\tau_\ell \in U_p^1$. Because $\rho(U_p^1).\mathbb{1}_{R_p} = \mathbb{1}_{R_p}$, we can write

$$\rho\left(g,g;s_\infty(\tau_\ell)_f^{-1}\right).\Phi = \rho(s_\infty).\Phi_\infty(g_\infty^{-1}\bullet g_\infty)\prod_{p\nmid l}\mathbb{1}_{g_p R_p g_p^{-1}}\prod_{p\mid l}\rho(w).\mathbb{1}_{g_p R_p g_p^{-1}}.$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

The Weil action of w is by the Fourier transform for $p \nmid d_B$ and it is by the negative of the Fourier transform for $p \mid d_B$. Specifically, it is shown in [KS20, Section §4] that for $p \mid d_B N$

$$\rho(w).\mathbb{1}_{R_p} = \gamma_p p^{-1} \mathbb{1}_{R_p^\vee},$$

where $\gamma_p = 1$ if B_p is split and $\gamma_p = -1$ if B_p is ramified. We conclude that

$$\rho\left(g, g, s_\infty (\tau_\ell)_f^{-1}\right).\Phi = \frac{\mu(\gcd(l, d_B))}{\ell} \rho(s_\infty).\Phi_\infty(g_\infty^{-1} \bullet g_\infty) \prod_{p \nmid \ell} \mathbb{1}_{g_p R_p g_p^{-1}} \prod_{p \mid \ell} \mathbb{1}_{g_p R_p^\vee g_p^{-1}}.$$

Because $\bigcap_{p\nmid\ell} g_p R_p g_p^{-1} \bigcap_{p\mid\ell} g_p R_p^\vee g_p^{-1} = R(\ell; g_f)$, we have for $\xi \in B$ that

$$\left(\prod_{p\nmid\ell} \mathbb{1}_{g_p R_p g_p^{-1}} \prod_{p\mid\ell} \mathbb{1}_{g_p R_p^\vee g_p^{-1}}\right)(\xi) = \mathbb{1}_{R(\ell;g_f)}(\xi),$$

and we can write

$$\begin{aligned}\Theta_\Phi(g, g; (\tau_\ell)_\infty s_\infty U_R^1) &= \frac{\mu(\gcd(l, d_B))}{\ell} \sum_{\xi \in R(\ell; g_f)} (\rho(s_\infty).\Phi_\infty)\,(g_\infty^{-1} \xi g_\infty) \\ &= \frac{\mu(\gcd(l, d_B))}{\ell} \sum_{x \in R(\ell; g)} (\rho(s_\infty).\Phi_\infty)\,(x).\end{aligned}$$

The last equality holds because $g_\infty^{-1} R(\ell; g_f) g_\infty = R(\ell; g)$. The claim now follows from Lemma A.14 above and the formulæ for the Weil action of the diagonal and unipotent subgroups. The moderate growth of $\theta_{g,\ell}$ now follows from the moderate growth of Θ_Φ in the symplectic variable s. The $\Gamma_0(d_B N)$-modularity of $\theta_{g,\ell}$ follows from the left $\mathrm{SL}_2(\mathbb{Q})$-invariance and right U_R^1-invariance of Θ_Φ in the symplectic variable, and the fact that τ_ℓ normalizes $\Gamma_0(d_B N)$. □

Proposition A.16. *Let $\Phi = \Phi_\infty \cdot \prod_{v<\infty} \mathbb{1}_{R_v}$, with Φ_∞ given by any of test functions listed in Definition A.13. Let $\mathcal{G}$ be any of the families of automorphic forms corresponding to Θ_Φ according to Table 1. Then, for any $\varphi \in \mathcal{G} \subseteq L^2([G])^\infty$, a K_R-invariant Hecke eigenform, we have $\varphi_\Phi = V^{-1}\varphi^{\mathrm{JL}}$, where φ^{JL} is the arithmetically normalized Jacquet–Langlands lift of φ, as defined in §5.1.2.*

Proof. Let κ be the entry of Table 1 corresponding to $\mathcal{G}$ and Θ_Φ. Lemma A.14 shows that φ_Φ is of weight κ and Proposition A.10 that $\iota_0^{*-1}\varphi_\Phi$ is newvector (or zero) of level U_R of the Jacquet–Langlands transfer π^{JL} of the representation π generated by φ. The subspace of vectors in π^{JL} satisfying these two properties is one-dimensional. This implies that φ_Φ is proportional to φ^{JL}. In order to find the constant of proportionality ρ_1, we compute and compare the Whittaker functions at the identity. The Whittaker function of φ^{JL} is recorded in §5.1.2 and those of φ_Φ we shall compute with the aid of Lemma A.9.

The case $\mathcal{G} = \mathcal{F}^-, \Phi_\infty = \Phi_\infty^{-,0}, \kappa = 0$:

Suppose that $\varphi \in \mathcal{F}^-_{\frac{1}{4}+t^2} \subseteq \mathcal{F}^-$. Then, the representation π_∞ is a principal series representation obtained by normalized induction of the character $\left(\begin{smallmatrix}\lambda & * \\ 0 & \mu\end{smallmatrix}\right) \mapsto \mathrm{sgn}(\lambda/\mu)^\alpha |\lambda/\mu|^{it}$ for some $\alpha \in \{0, 1\}$. The equality of Whittaker functions yields the following equation for the constant of proportionality ρ_1:

$$2\rho_1 K_{it}(2\pi) = V^{-1} \operatorname{Tr}\left(\mathrm{Res}^{\mathrm{PGL}_2(\mathbb{R})}_{\mathrm{SL}_2(\mathbb{R})} \pi_\infty\right)(\Phi_\infty^{-,0} \restriction_{G'(\mathbb{R})}).$$

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Because $\Phi_\infty^{-,0}\restriction_{G'(\mathbb{R})}$ is bi-K_∞-invariant, the trace is the Fourier transform of the Abel–Satake transform of $\Phi_\infty^{-,0}\restriction_{G'(\mathbb{R})}$. Compute first the Abel–Satake transform

$$\begin{aligned}\mathcal{S}\Phi_\infty^{-,0}\restriction_{G'(\mathbb{R})}(\tau) &= e^{\tau/2}\int_{-\infty}^{\infty}\Phi_\infty^{-,0}\left(\begin{pmatrix}e^{\tau/2} & 0\\ 0 & e^{-\tau/2}\end{pmatrix}\begin{pmatrix}1 & n\\ 0 & 1\end{pmatrix}\right)\mathrm{d}n\\ &= e^{\tau/2}\int_{-\infty}^{\infty}e^{-\pi(2\cosh\tau+e^{\tau}n^2)}\,\mathrm{d}n = e^{-2\pi\cosh\tau}.\end{aligned}$$

The trace is proportional to the Fourier transform of $\mathcal{S}\Phi_\infty^{-,0}\restriction_{G'(\mathbb{R})}$. Using our measure normalization this becomes

$$\mathrm{Tr}\left(\mathrm{Res}^{\mathrm{PGL}_2(\mathbb{R})}_{\mathrm{SL}_2(\mathbb{R})}\pi_\infty\right)(\Phi_\infty^{-,0}\restriction_{G'(\mathbb{R})}) = \int_{-\infty}^{\infty}e^{-2\pi\cosh\tau}e^{it\tau}\,\mathrm{d}\tau = 2K_{it}(2\pi).$$

Hence, $\rho_1 = V^{-1}$.

The case $\mathcal{G} = \mathcal{F}^{-,\mathrm{hol}}$, $\Phi_\infty = \Phi_\infty^{-,k}$, $\kappa = k$:

The equality of Whittaker functions, yields the following equation for the constant of proportionality ρ_1:

$$\rho_1 e^{-2\pi} = V^{-1}\overline{\left\langle f_{\varphi_\infty,\varphi_\infty},\Phi_\infty^{-,k}\right\rangle}_{G'(\mathbb{R})}.$$

The representation π_∞ is the discrete series representation with parameter k, and $\varphi_\infty\in\pi_\infty$ is the L^2-normalized minimal weight vector. The matrix coefficient in this case is exactly[6] $f_{\varphi_\infty,\varphi_\infty}(g) = \overline{X(g)}^{(-k)}$ and we compute

$$\begin{aligned}\rho_1 e^{-2\pi} &= V^{-1}\int_{G'(\mathbb{R})}e^{-2\pi P(g)}\,\mathrm{d}g = V^{-1}2\pi\int_0^{\infty}e^{-2\pi\cosh\tau}\sinh\tau\,\mathrm{d}\tau\\ &= V^{-1}2\pi\int_1^{\infty}e^{-2\pi\xi}\,\mathrm{d}\xi = V^{-1}e^{-2\pi}.\end{aligned}$$

The remaining cases:

Here, we verify[7] that the matrix coefficient satisfies $f_{\varphi_\infty,\varphi_\infty} = e^{2\pi}d_{\pi_\infty}^{-1}\Phi_\infty$, where $\varphi_\infty\in\pi_\infty$ is the Archimedean component of φ, L^2-normalized and d_{π_∞} the (formal) degree of π_∞. The equality of Whittaker functions, then yields the following equation for the constant of proportionality ρ_1:

$$\begin{aligned}\rho_1 e^{-2\pi} &= V^{-1}\overline{\left\langle f_{\varphi_\infty,\varphi_\infty},\Phi_\infty\right\rangle}_{G'(\mathbb{R})} = V^{-1}e^{-2\pi}d_{\pi_\infty}\overline{\left\langle f_{\varphi_\infty,\varphi_\infty},f_{\varphi_\infty,\varphi_\infty}\right\rangle}_{G'(\mathbb{R})}\\ &= V^{-1}e^{-2\pi}\|\varphi_\infty\|_2^4 = V^{-1}e^{-2\pi},\end{aligned}$$

where we have used Schur–Weyl orthogonality for matrix coefficients. □

Acknowledgements. We would like to thank E. Assing, V. Blomer, F. Brumley, G. Harcos, Y. Hu, S. Marshall, A. Saha, W. Sawin and R. Toma for their helpful feedback on an earlier draft. We would also like to thank P. Sarnak for fruitful discussions on this and surrounding topics as well as his continued encouragement and support. We also thank the referee for their careful read-through of the manuscript and their valuable comments.

[6]This can be computed succinctly using the model of the discrete series as a subrepresentation of $L^2(G(\mathbb{R}))$.

[7]For $\varphi\in\mathcal{F}_m^+$, this can be computed easily from the model of the representation spanned by spherical harmonics, using the identity $P_m(\langle \mathrm{v},\mathrm{v}'\rangle) = \frac{4\pi}{2m+1}\sum_{n=-m}^{m}Y_{mn}(\mathrm{v})\overline{Y_{mn}(\mathrm{v}')}$, and the orthogonality of spherical harmonics. For $\varphi\in\mathcal{F}^{+,\mathrm{hol}}$, this can be computed easily from the model of the representation on the space of homogeneous binary complex polynomials of degree k.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

Funding statement. I.K. is deeply grateful for support of the AMS Centennial fellowship and the Sloan Research Fellowship.

This paper was completed while P.N. was at the Institute for Advanced Study during the academic year 2021–2022, where he was supported by the National Science Foundation under Grant No. DMS-1926686. Some work on this project also occurred during a short-term visit of P.N. to the Institute for Advanced Study in February 2020. Whilst the paper was revised for publication in April 2024, P.N. was supported by the research grant (VIL54509) from VILLUM FONDEN .

R.S. wishes to extend his gratitude to 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, where this research was conducted. During revision for publication in April 2024, R.S. was employed at the Huawei Research Center in Zurich, whom he also thanks.

Competing interest. The authors have no competing interest to declare.

Data availability statement. Not applicable.

Ethical standards. No ethical standards were required in the pursuit of this research.

Author contributions. All authors contributed equally.

References

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

[Ass21] E. Assing, 'The sup-norm problem beyond the newform', Preprint, 2021, arXiv:2111.01893.

[Ass24] E. Assing, 'On sup-norm bounds part I: Ramified Maaß newforms over number fields', *J. Eur. Math. Soc. (JEMS)* 2024. Published online first.

[Ber31] S. Bernstein, 'Sur les polynomes orthogonaux relatifs à un segment fini (seconde partie)', *Journal de Mathématiques Pures et Appliquées* **10** (1931), 219–286.

[Bér77] P. H. Bérard, 'On the wave equation on a compact Riemannian manifold without conjugate points', *Math. Z.* **155**(3) (1977), 249–276.

[BH10] V. Blomer and R. Holowinsky, 'Bounding sup-norms of cusp forms of large level', *Invent. Math.* **179**(3) (2010), 645–681.

[BHM16] V. Blomer, G. Harcos and D. Milićević, 'Bounds for eigenforms on arithmetic hyperbolic 3-manifolds', *Duke Math. J.* **165**(4) (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**(1) (2020), 1–53.

[BHMM22] V. Blomer, G. Harcos, P. Maga and D. Milićević, 'Beyond the spherical sup-norm problem', *J. Math. Pures Appl. (9)* **168** (2022), 1–64.

[BHW93] U. Betke, M. Henk and J. M. Wills, 'Successive-minima-type inequalities', *Discrete Comput. Geom.* **9**(2) (1993), 165–175.

[BK15] Jack Buttcane and Rizwanur Khan. L^4-norms of Hecke newforms of large level. *Math. Ann.* **362**(3-4): 699–715, 2015.

[Blo13] V. Blomer, 'On the 4-norm of an automorphic form', *J. Eur. Math. Soc. (JEMS)* **15**(5) (2013), 1825–1852.

[BM11] V. Blomer and P. Michel, 'Sup-norms of eigenfunctions on arithmetic ellipsoids', *Int. Math. Res. Not. IMRN* (**21**) (2011), 4934–4966.

[BM13] V. Blomer and P. Michel, 'Hybrid bounds for automorphic forms on ellipsoids over number fields', *J. Inst. Math. Jussieu* **12**(4) (2013), 727–758.

[Bor63] A. Borel, 'Some finiteness properties of adele groups over number fields', *Inst. Hautes Études Sci. Publ. Math.* (**16**) (1963), 5–30.

[Bum97] D. Bump. *Automorphic Forms and Representations*, *Cambridge Studies in Advanced Mathematics*, vol. 55 (Cambridge University Press, Cambridge, 1997).

[Cas73] W. Casselman, 'On some results of Atkin and Lehner', *Math. Ann.* **201** (1973), 301–314.

[Com21] F. Comtat, 'Optimal sup norm bounds for newforms on GL_2 with maximally ramified central character', *Forum Math*. **33**(1) (2021), 1–16.

[Del71] P. Deligne, 'Formes modulaires et représentations ℓ-adiques', *Séminaire Bourbaki , Vol. 1968/1969, Exp. 347-363* **179** (1971), 139–172.

[Del74] P. Deligne, 'La conjecture de Weil. I', *Inst. Hautes Études Sci. Publ. Math.* **43** (1974), 273–307.

[GL87] P. M. Gruber and C. G. Lekkerkerker, *Geometry of Numbers*, second edn., North-Holland Mathematical Library, vol. 37 (North-Holland Publishing Co., Amsterdam, 1987).

[Har03] G. Harcos, 'An additive problem in the Fourier coefficients of cusp forms', *Math. Ann.* **326**(2) (2003), 347–365.

[HC19] F. Hou and B. Chen, 'Level aspect subconvexity for twisted L-functions', *J. Number Theory* **203** (2019), 12–31.

[HL94] J. Hoffstein and P. Lockhart, 'Coefficients of Maass forms and the Siegel zero', *Ann. of Math. (2)* **140**(1) (1994), 161–181. With an appendix by Dorian Goldfeld, Hoffstein and D. Lieman.

[HM06] G. Harcos and P. Michel, 'The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points. II', *Invent. Math.* **163**(3) (2006), 581–655.
[HN18] Y. Hu and P. D. Nelson, 'New test vector for Waldspurger's period integral, relative trace formula, and hybrid subconvexity bounds', Preprint, 2018, arXiv:1810.11564.
[HNS19] Y. Hu, P. D. Nelson and A. Saha, 'Some analytic aspects of automorphic forms on $GL(2)$ of minimal type', *Comment. Math. Helv.* **94**(4) (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**(11) (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* (**20**) (2012), 4764–4774.
[HT13] G. Harcos and N. Templier, 'On the sup-norm of Maass cusp forms of large level. III', *Math. Ann.* **356**(1) (2013), 209–216.
[IS95] H. Iwaniec and P. Sarnak, 'L^∞ norms of eigenfunctions of arithmetic surfaces', *Ann. of Math. (2)* **141**(2) (1995), 301–320.
[Iwa90] H. Iwaniec, 'Small eigenvalues of Laplacian for $\Gamma_0(N)$', *Acta Arith.* **56**(1) (1990), 65–82.
[Iwa97] H. Iwaniec, *Topics in Classical Automorphic*, Forms Graduate Studies in Mathematics, vol. 17 (American Mathematical Society, Providence, RI, 1997).
[JL70] H. Jacquet and R. P. Langlands, *Automorphic forms on* (2), Lecture Notes in Mathematics, vol. 114 (Springer-Verlag, Berlin, 1970).
[KR94] S. S. Kudla and S. Rallis, 'A regularized Siegel–Weil formula: The first term identity', *Ann. of Math. (2)* **140**(1) (1994), 1–80.
[KS20] I. Khayutin and R. S. Steiner, 'Theta functions, fourth moments of eigenforms, and the sup-norm problem I', Preprint, 2020, arXiv:2009.07194.
[LPS87] A. Lubotzky, R. Phillips and P. Sarnak, 'Hecke operators and distributing points on S^2', II. *Comm. Pure Appl. Math.* **40**(4) (1987), 401–420.
[Mar16] S. Marshall, 'Local bounds for L^p norms of Maass forms in the level aspect', *Algebra Number Theory* **10**(4) (2016), 803–812.
[Mœg97] C. Mœglin, 'Quelques propriétés de base des séries théta', *J. Lie Theory* **7**(2) (1997), 231–238.
[Nel15] P. D. Nelson, 'Evaluating modular forms on Shimura curves', *Math. Comp.* **84**(295) (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), Paper No. e44, 21.
[Nel21] P. D. Nelson, 'The spectral decomposition of $|\theta|^2$', *Math. Z.* **298**(3-4) (2021), 1425–1447.
[Nor21] A. C. Nordentoft, 'Hybrid subconvexity for class group L-functions and uniform sup norm bounds of Eisenstein series', *Forum Math.* **33**(1) (2021), 39–57.
[PY19] I. Petrow and M. P. Young, 'The fourth moment of Dirichlet L-functions along a coset and the Weyl bound', *arXiv e-prints*, 2019, arXiv:1908.10346.
[Ral84] S. Rallis, 'On the Howe duality conjecture', *Compositio Math.* **51**(3) (1984), 333–399.
[RS75] S. Rallis and G. Schiffmann, 'Distributions invariantes par le groupe orthogonal', in *Analyse harmonique sur les groupes de Lie (Sém., Nancy-Strasbourg, 1973–75)*, Lecture Notes in Math., vol. 497 (1975), 494–642.
[Sah17] A. Saha, 'On sup-norms of cusp forms of powerful level', *J. Eur. Math. Soc. (JEMS)* **19**(11) (2017), 3549–3573.
[Sah20] A. Saha, 'Sup-norms of eigenfunctions in the level aspect for compact arithmetic surfaces', *Math. Ann.* **376**(1-2) (2020), 609–644.
[Saw21] W. Sawin, 'A geometric approach to the sup-norm problem for automorphic forms: the case of newforms on $GL_2\left(F_q(T)\right)$ with squarefree level', *Proc. Lond. Math. Soc. (3)* **123**(1) (2021), 1–56.
[Shi72] H. Shimizu, 'Theta series and automorphic forms on GL_2', *J. Math. Soc. Japan* **24** (1972), 638–683.
[Ste20] R. S. Steiner, 'Sup-norm of Hecke-Laplace eigenforms on S^3', *Math. Ann.* **377**(1-2) (2020), 543–553.
[Ste23] R. S. Steiner, 'Small diameters and generators for arithmetic lattices in $SL_2(\mathbb{R})$ and certain Ramanujan graphs', *Ramanujan J.* **62**(4) (2023), 953–966.
[Tem10] N. Templier, 'On the sup-norm of Maass cusp forms of large level', *Selecta Math. (N.S.)* **16**(3) (2010), 501–531.
[Tem15] N. Templier, 'Hybrid sup-norm bounds for Hecke-Maass cusp forms', *J. Eur. Math. Soc. (JEMS)* **17**(8) (2015), 2069–2082.
[Tom23] R. Toma, 'Hybrid bounds for the sup-norm of automorphic forms in higher rank', *Trans. Amer. Math. Soc.* **376**(8) (2023), 5573–5600.
[vdC36] J. G. van der Corput, 'Verallgemeinerung einer Mordellschen Beweismethode in der Geometrie der Zahlen, Zweite Mitteilung', *Acta Arithmetica* **2**(1) (1936), 145–146.
[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.
[Voi18] J. Voight, 'Quaternion algebras', 2018, https://math.dartmouth.edu/jvoight/quat.html.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press

[Wal85] J.-L. Waldspurger, 'Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie', *Compositio Math.* **54**(2) (1985), 173–242.
[Wat08] T. C. Watson, 'Rankin triple products and quantum chaos', Preprint, 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**(2) (2007), 325–327.
[You17] M. P. Young, 'Weyl-type hybrid subconvexity bounds for twisted L-functions and Heegner points on shrinking sets', *J. Eur. Math. Soc. (JEMS)* **19**(5) (2017), 1545–1576.

https://doi.org/10.1017/fmp.2024.9 Published online by Cambridge University Press