



Peter Humphries · Asbjørn Christian Nordentoft

Sparse equidistribution of geometric invariants of real quadratic fields

Received 15 November 2022; revised 30 March 2025

Abstract. Duke, Imamoğlu, and Tóth have recently constructed a new geometric invariant, a hyperbolic orbifold, associated to each narrow ideal class of a real quadratic field. Furthermore, they have shown that the projection of these hyperbolic orbifolds onto the modular surface $\Gamma \backslash \mathbb{H}$ equidistributes on average over a genus of the narrow class group as the fundamental discriminant D of the real quadratic field tends to infinity. We extend this construction of hyperbolic orbifolds to allow for a level structure, akin to Heegner points and closed geodesics of level q . Additionally, we refine this equidistribution result in several directions. First, we investigate sparse equidistribution in the level aspect, where we prove the equidistribution of level q hyperbolic orbifolds when restricted to a translate of $\Gamma \backslash \mathbb{H}$ in $\Gamma_0(q) \backslash \mathbb{H}$, which presents some new interesting features. Second, we explore sparse equidistribution in the subgroup aspect, namely equidistribution on average over small subgroups of the narrow class group. Third, we prove small scale equidistribution and give upper bounds for the discrepancy. Behind these refinements is a new interpretation of the Weyl sums arising in these equidistribution problems in terms of adèlic period integrals, which in turn are related to Rankin–Selberg L -functions via Waldspurger’s formula. The key remaining inputs are hybrid subconvex bounds for these L -functions and a certain homological version of the sup-norm problem.

Keywords: equidistribution, thin subgroup, hyperbolic orbifold, real quadratic field.

1. Introduction

1.1. Equidistribution of hyperbolic orbifolds

Let $E := \mathbb{Q}(\sqrt{D})$ be a real quadratic number field, where $D > 1$ is a positive fundamental discriminant. Associated to each narrow ideal class A of the narrow class group Cl_D^+ of E is a closed geodesic $\mathcal{C}_A$ on the modular surface $\Gamma \backslash \mathbb{H}$, where $\Gamma := \text{SL}_2(\mathbb{Z})$ denotes the

Peter Humphries: Department of Mathematics, University of Virginia, 141 Cabell Drive, Kerchof Hall, Charlottesville, VA 22904, USA; pclhumphries@gmail.com

Asbjørn Christian Nordentoft: Laboratoire de Mathématiques d’Orsay, Université Paris-Saclay, 307 Rue Michel Magat, 91405 Orsay Cedex, France; Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark; acnordentoft@outlook.com, nordentoft@math.ku.dk

Mathematics Subject Classification 2020: 11F12 (primary); 11E16, 11F67, 11R11 (secondary).

modular group. In [19], Duke, Imamoğlu, and Tóth introduced a new geometric invariant associated to each narrow ideal class A , a hyperbolic orbifold $\Gamma_A \backslash \mathcal{N}_A$ with boundary given by the closed geodesic $\mathcal{C}_A$. The group $\Gamma_A \subset \mathrm{PSL}_2(\mathbb{Z})$ is a Fuchsian group of the second kind whose construction is given in terms of certain invariants of A , while $\mathcal{N}_A \subset \mathbb{H}$ is the Nielsen region of Γ_A , namely the smallest nonempty Γ_A -invariant open convex subset of $\mathbb{H}$.

Duke, Imamoğlu, and Tóth additionally showed that these hyperbolic orbifolds equidistribute as D tends to infinity when projected onto the modular surface. For each positive fundamental discriminant D , one chooses a genus G_D in the group of genera $\mathrm{Gen}_D := \mathrm{Cl}_D^+ / (\mathrm{Cl}_D^+)^2$, so that G_D is a coset $C(\mathrm{Cl}_D^+)^2$ of narrow ideal classes for some $C \in \mathrm{Cl}_D^+$; then for every continuity set $B \subset \Gamma \backslash \mathbb{H}$,

$$\frac{\sum_{A \in G_D} \mathrm{vol}(\mathcal{F}_A \cap \Gamma B)}{\sum_{A \in G_D} \mathrm{vol}(\mathcal{F}_A)} = \frac{\mathrm{vol}(B)}{\mathrm{vol}(\Gamma \backslash \mathbb{H})} + o_B(1) \quad (1.1)$$

as D tends to infinity through fundamental discriminants [19, Theorem 2]. Here $\mathcal{F}_A$ denotes a canonical fundamental domain for $\Gamma_A \backslash \mathcal{N}_A$, while the volume measure on the upper half-plane $\mathbb{H} \ni z = x + iy$ is $d\mu(z) = y^{-2} dx dy$, so that $\mathrm{vol}(\Gamma \backslash \mathbb{H}) = \frac{\pi}{3}$. This equidistribution theorem can be viewed as an analogue of Duke's celebrated result on the equidistribution of closed geodesics and of Heegner points on the modular surface and additionally of lattice points on the sphere [17].

1.2. Sparse equidistribution in the level aspect

We generalise Duke, Imamoğlu, and Tóth's construction of hyperbolic orbifolds in the level aspect. Let q be an odd prime that splits in E . In Section 3, we construct hyperbolic orbifolds of level q , denoted by $\Gamma_A(q) \backslash \mathcal{N}_A(q)$, and a canonical fundamental domain $\mathcal{F}_A(q) \subset \mathbb{H}$. This extends the level 1 construction of the hyperbolic orbifolds $\Gamma_A \backslash \mathcal{N}_A$ and their canonical fundamental domains $\mathcal{F}_A$ introduced in [19]. These level q hyperbolic orbifolds are analogous to Heegner points of level q [29, p. 499] and closed geodesics of level q [29, p. 500] (cf. [14, Section 1]), but with several key new features. We refer to Figures 2, 3, and 4 in Section 3 for some illustrative examples.

The level q modular surface $\Gamma_0(q) \backslash \mathbb{H}$ may be written as

$$\Gamma_0(q) \backslash \mathbb{H} = \bigcup_{\omega_q \in \Gamma / \Gamma_0(q)} \omega_q^{-1} \Gamma \backslash \mathbb{H}.$$

This partitions $\Gamma_0(q) \backslash \mathbb{H}$ into $q + 1$ translates of $\Gamma \backslash \mathbb{H}$. We are interested in the *hybrid* problem concerning the asymptotic behaviour of the volume of $\mathcal{F}_A(q) \cap \Gamma_0(q) \omega_q^{-1} \Gamma \backslash \mathbb{H}$ on average over a genus G_D as *both* q and D tend to infinity; this is motivated by the work of Liu, Masri, and Young on the analogous problem for Heegner points [49, Theorem 1.4]. Since the volume of $\omega_q^{-1} \Gamma \backslash \mathbb{H}$ is equal to that of $\Gamma \backslash \mathbb{H}$, these translates are small in comparison to the total volume of $\Gamma_0(q) \backslash \mathbb{H}$, so that as q grows, we are studying the equidistribution of hyperbolic orbifolds in sets of shrinking volume.

Theorem 1.1. Fix $\delta \in [0, \frac{1}{36})$. For each positive squarefree fundamental discriminant D , choose a genus G_D in the group of genera Gen_D . For each odd prime q that splits in E with $q \leq D^\delta$, choose $\omega_q \in \Gamma/\Gamma_0(q)$. Then

$$\frac{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})}{\text{vol}(\Gamma\backslash\mathbb{H})} \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q) \cap \Gamma_0(q)\omega_q^{-1}\Gamma\backslash\mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} = 1 + o_\delta(1)$$

as qD tends to infinity. Assuming the generalised Lindelöf hypothesis, the same result holds for $q \leq D^\delta$ for some fixed $\delta \in [0, \frac{1}{12})$.

The second author [54] recently studied the related problem of determining the distribution of the homology classes of closed geodesics

$$[\mathcal{C}_A(q)] \in H_1(Y_0(q), \mathbb{Z}),$$

where $Y_0(q) := \Gamma_0(q)\backslash\mathbb{H}$ is the non-compact modular surface of level q and $\mathcal{C}_A(q)$ is the oriented closed geodesic associated to $A \in \text{Cl}_D^+$. As discussed in [49], these results can be viewed as an analogue of Linnik's theorem. Linnik's theorem, in its effective form, states that there exists an absolute constant $L > 1$ such that given a positive integer q and a reduced residue class a modulo q , there exists the expected number of primes, $(1 + o(1)) \frac{q^L}{\varphi(q)}$, that are less than q^L and are congruent to a modulo q . Theorem 1.1 can be viewed in a similar light: there exists an absolute constant $\delta > 0$ such that given a prime q and a translate $\omega_q^{-1}\Gamma\backslash\mathbb{H}$ of $\Gamma\backslash\mathbb{H}$ in $\Gamma_0(q)\backslash\mathbb{H}$, there exists the expected proportion of mass of projections of hyperbolic orbifolds of discriminant $D \leq q^{\frac{1}{\delta}}$ onto $\Gamma_0(q)\backslash\mathbb{H}$ that lie in the translate $\omega_q^{-1}\Gamma\backslash\mathbb{H}$.

1.3. Sparse equidistribution in the subgroup aspect

Our second refinement is to study the equidistribution of hyperbolic orbifolds averaged over *sparse* subsets of Cl_D^+ . Previously, we averaged over a genus G_D , which has cardinality $2^{1-\omega(D)}h_D^+$, where $h_D^+ := |\text{Cl}_D^+|$ denotes the narrow class number and $\omega(D)$ denotes the number of distinct prime divisors of D . We instead consider an *arbitrary* subgroup $H = H_D$ of the narrow class group Cl_D^+ in place of the subgroup $(\text{Cl}_D^+)^2$ and an *arbitrary* coset CH in place of a genus $G_D = C(\text{Cl}_D^+)^2$. Whereas a genus G_D satisfies $|G_D| \gg_\varepsilon D^{-\varepsilon}h_D^+$ for every $\varepsilon > 0$, we allow for the possibility that the cardinality of a coset CH may be significantly smaller than h_D^+ .

Theorem 1.2. Fix $\delta \in [0, \frac{1}{2826})$ and fix either $q = 1$ or q an odd prime. For each positive fundamental discriminant D for which q splits in E if $q > 1$, choose a coset CH with $C \in \text{Cl}_D^+$ and $H = H_D$ a subgroup of Cl_D^+ satisfying $|H| \gg D^{-\delta}h_D^+$. Then for each fixed continuity set $B \subset \Gamma_0(q)\backslash\mathbb{H}$,

$$\frac{\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q) \cap \Gamma_0(q)B)}{\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q))} = \frac{\text{vol}(B)}{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})} + o_{q,B,\delta}(1) \quad (1.2)$$

as D tends to infinity. Assuming the generalised Lindelöf hypothesis, the same result holds for $|H| \gg D^{-\delta}h_D^+$ for some fixed $\delta \in [0, \frac{1}{4})$.

Note that by taking H to be the trivial subgroup, Theorem 1.2 implies the equidistribution of *individual* hyperbolic orbifolds as D tends to infinity along fundamental discriminants for which $h_D^+ \ll D^\delta$ for some fixed $\delta < \frac{1}{2826} \approx 0.00035$ (cf. [57, Theorem 6.5.1]).

At the other extreme, we may take H to be Cl_D^+ , so that (1.2) gives the equidistribution of hyperbolic orbifolds averaged over the whole narrow class group. For $q = 1$, such a result is *trivial*, as observed by Duke, Imamoglu, and Tóth [19, Section 4]; as we discuss in Section 8, however, this result is no longer trivial for q an odd prime.

Theorem 1.2 is motivated by a conjecture of Michel and Venkatesh [53, Conjecture 1], where the analogous statement for Heegner points is conjectured to hold for any fixed $\delta \in [0, \frac{1}{2})$; it is noted that the generalised Lindelöf hypothesis implies such a conjecture in the range $\delta \in [0, \frac{1}{4})$. Harcos and Michel have proven this conjecture in the range $\delta \in [0, \frac{1}{2826})$ [32, Theorem 1.2] (see additionally [31, Corollary 1.4]), while Venkatesh has studied this problem in more general settings [61, Theorem 7.2].

More recently, a toy model of this problem was resolved by the first author [37, Theorem 1.5], namely the sparse equidistribution as q tends to infinity of the points

$$\left\{ \left(\frac{d}{q}, \frac{d'}{q} \right) \in \mathbb{T}^2 : d \in CH, d d' \equiv 1 \pmod{q} \right\}$$

in the torus $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ indexed by a coset CH of the group $(\mathbb{Z}/q\mathbb{Z})^\times$ with q a prime and $|H| \gg q^\delta$ for some fixed $\delta > 0$. In this setting, it is shown that this sparse equidistribution result is a simple consequence of a deep result of Bourgain on cancellation in certain exponential sums [11].

1.4. Small scale equidistribution and discrepancy bounds

The first author investigated a refinement of Duke, Imamoglu, and Tóth's equidistribution result in [36], namely small scale equidistribution, in which the continuity set $B \subset \Gamma \backslash \mathbb{H}$ in (1.1) is chosen to be a ball $B_R(w)$ whose radius R shrinks as D grows. One can think of the fastest rate at which this radius can shrink with respect to the growth of D for which equidistribution still holds as being the smallest *scale* of equidistribution. We prove the following unconditional result in this regard.

Theorem 1.3. *Fix $\delta \in [0, \frac{1}{2})$ and $w \in \Gamma \backslash \mathbb{H}$. For each positive squarefree fundamental discriminant D , choose a genus G_D in the group of genera Gen_D . Then for all $R \in [D^{-\delta}, 1]$, we have that*

$$\frac{\text{vol}(\Gamma \backslash \mathbb{H})}{\text{vol}(B_R(w))} \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} = 1 + o_\delta(1) \quad (1.3)$$

as D tends to infinity.

This improves upon [36, Theorem 1.24], where this result was proven under the assumption of the generalised Lindelöf hypothesis.¹ We use Theorem 1.3 to bound the

¹As stated, [36, Theorem 1.24] claims that formula (1.3) holds for $R \asymp D^{-\delta}$ for some fixed $\delta \in [0, \frac{1}{12})$ unconditionally and for some fixed $\delta \in [0, \frac{1}{4})$ under the assumption of the generalised

discrepancy associated to this equidistribution result, which may be thought of as a quantitative way of measuring *uniformly* the rate of equidistribution.

Theorem 1.4. *For each positive squarefree fundamental discriminant D , choose a genus G_D in the group of genera Gen_D . Then as D tends to infinity, we have that*

$$\sup_{B_R(w) \subset \Gamma \backslash \mathbb{H}} \left| \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} - \frac{\text{vol}(B_R(w))}{\text{vol}(\Gamma \backslash \mathbb{H})} \right| \ll_\varepsilon D^{-\frac{1}{12} + \varepsilon},$$

where the supremum is over all injective geodesic balls in $\Gamma \backslash \mathbb{H}$. Assuming the generalised Lindelöf hypothesis, the stronger bound $O_\varepsilon(D^{-\frac{1}{4} + \varepsilon})$ holds for the discrepancy.

Remark 1.5. Recently, the second named author and Ser Peow Tan [55] gave a geometric construction of the *partial coverings* studied in this paper and in [19] (i.e., the image of the projection $\mathcal{F}_A(q) \rightarrow \Gamma_0(q) \backslash \mathbb{H}$) applicable to any Fuchsian group of the first kind. This construction circumvents any reference to thin subgroups. Furthermore, [55, Theorem 1.1] shows that equidistribution of a sequence of collections of closed geodesics implies equidistribution of the associated partial coverings, where in both settings equidistribution is meant in the sense of smooth compactly supported functions. The methods are quite different to the present paper and in particular do not yield effective error terms (outside of the co-compact case, which is not treated in the present paper) and do not cover the case where the Fuchsian group is varying, as in Theorem 1.1.

1.5. A discussion on the proofs

1.5.1. Sparse equidistribution in the level aspect. Theorem 1.1 is proven via approximating the indicator function of $\omega_q^{-1} \Gamma \backslash \mathbb{H}$ by a smooth function and spectrally expanding this on $\Gamma_0(q) \backslash \mathbb{H}$. One then integrates this spectral expansion over the projection of $\mathcal{F}_A(q)$ onto $\Gamma_0(q) \backslash \mathbb{H}$ and sums over $A \in G_D$. The main contribution comes from the constant function. It remains to bound the contributions from the cuspidal and continuous spectrum, which involve Weyl sums. As is standard for equidistribution results of this form, these Weyl sums can be expressed in terms of the twisted L -functions $L(\frac{1}{2}, f \otimes \chi_{D_1})$, where $D_1 \mid D$ is a fundamental discriminant and χ_{D_1} denotes the primitive quadratic Dirichlet character modulo $|D_1|$. Unusually, there is additionally a new *topological contribution* involving

- (1) the twisted L -functions $L(\frac{1}{2}, h \otimes \chi_{D_1})$ for h a holomorphic Hecke cusp form of weight 2 and level q ;
- (2) line integrals of f along sides of a fundamental polygon $\mathcal{P}(q)$ of $\Gamma_0(q)$ (which are independent of D); and
- (3) certain cap product pairings between homology and cohomology of the compactification $X_0(q)$ of the modular surface $\Gamma_0(q) \backslash \mathbb{H}$ (which are also independent of D).

Lindelöf hypothesis, though these conditions should in fact be the stronger conditions $\delta \in [0, \frac{1}{6})$ and $\delta \in [0, \frac{1}{2})$, respectively.

A key step is to obtain strong bounds for certain moments of L -functions of the form

$$\sum_{D_1 D_2 = D} \sum_{\substack{f \in \mathcal{B}_0(\Gamma_0(q)) \\ T \leq t_f \leq 2T}} \frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)},$$

where $\mathcal{B}_0(\Gamma_0(q))$ denotes an orthonormal basis of $L^2(\Gamma_0(q) \backslash \mathbb{H})$ consisting of Hecke–Maaß cusp forms. Due the hybrid nature of Theorem 1.1, we require bounds for this moment that are uniform in q and D , and similarly hybrid bounds for related moments of L -functions associated to *holomorphic* Hecke cusp forms. Here we proceed via Hölder’s inequality together with hybrid bounds for the third moments of $L(\frac{1}{2}, f \otimes \chi_{D_1})$ and $L(\frac{1}{2}, f \otimes \chi_{D_2})$ due to Petrow and Young [56].

What remains is to bound the contribution from the topological term. By a detailed analysis of the structure of $\partial \mathcal{F}_A(q)$, the line integrals of f can be expressed in terms of Voronoï L -series (i.e., L -series of additive twists of f), which can in turn be bounded on average via the spectral large sieve. The bounding of the cap product pairing can be seen as a homological instance of the sup-norm problem and has been bounded recently by the second author [54] using a theta correspondence approach combined with techniques from geometric coding to treat the counting problem that arises.

1.5.2. Sparse equidistribution in the subgroup aspect. We reduce the proof of Theorem 1.2 to showing that for each fixed Hecke–Maaß cusp form $f \in \mathcal{B}_0(\Gamma_0(q))$, the quantity

$$\frac{|H|}{h_D^+} \frac{1}{\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q))} \sum_{\chi \in H^\perp} \bar{\chi}(C) W_{\chi, f}$$

tends to zero as D tends to infinity, as well as an analogous result for f replaced by an Eisenstein series. Here $H^\perp$ denotes the annihilator of H , namely the set of characters χ of Cl_D^+ that satisfy $\chi(A) = 1$ for all $A \in H$, while the Weyl sum $W_{\chi, f}$ is

$$W_{\chi, f} := \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A(q)} f(z) d\mu(z). \quad (1.4)$$

Following earlier work of Duke, Imamoğlu, and Tóth [19, Proposition 1], we prove a lower bound for the sums of volumes of the hyperbolic orbifolds of the form

$$\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q)) \gg_{q, \varepsilon} \frac{|H|}{h_D^+} D^{\frac{1}{2} - \varepsilon}.$$

Since $|H^\perp| = \frac{h_D^+}{|H|}$, equidistribution follows provided that we can show that there exists an absolute constant $\alpha > 0$ such that

$$W_{\chi, f} \ll_{q, f} \frac{|H|}{h_D^+} D^{\frac{1}{2} - \alpha}.$$

In this way, the proof of Theorem 1.2 is reduced to proving such bounds for Weyl sums.

To bound these Weyl sums, we make explicit a version of Waldspurger's formula [63] in order to show that the square of the absolute value of a Weyl sum is essentially equal to a special value of a Rankin–Selberg L -function. The proof is then completed upon invoking subconvex bounds for Rankin–Selberg L -functions due to Harcos and Michel [32].

1.5.3. Small scale equidistribution and discrepancy bounds. The proof of Theorem 1.3 concerning small scale equidistribution follows in a similar manner to previous work of Young [64, Theorem 2.1] on small scale equidistribution of Heegner points on $\Gamma \backslash \mathbb{H}$ and of the first author and Radziwiłł [39, Theorem 1.5] on small scale equidistribution of lattice points on the sphere. The chief idea is to approximate the indicator function of a ball via a smooth function and spectrally expand this on $\Gamma \backslash \mathbb{H}$. One then integrates this spectral expansion over the projection of $\Gamma_A \backslash \mathcal{N}_A$ onto $\Gamma \backslash \mathbb{H}$ and sums over $A \in G_D$. The main contribution comes from the constant function, and our goal becomes adequately bounding the ensuing sum over the cuspidal spectrum and integral over the continuous spectrum.

We approach this via a dyadic subdivision together with an application of pre-existing identities of Duke, Imamoğlu, and Tóth relating Weyl sums to L -functions [19, Theorems 3 and 4]. In this way, the problem is reduced to deducing strong bounds for the quantities

$$\sum_{D_1 D_2 = D} \sum_{\substack{f \in \mathcal{B}_0(\Gamma) \\ T \leq t_f \leq 2T}} \frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)},$$

$$\sum_{D_1 D_2 = D} \int_{T \leq |t| \leq 2T} \left| \frac{L(\frac{1}{2} + it, \chi_{D_1})^2 L(\frac{1}{2} + it, \chi_{D_2})^2}{\zeta(1 + 2it)} \right|^2 dt$$

that are uniform in both T and D . For this, we can appeal to recent work of the first author and Radziwiłł [39, Proposition 2.14]. The bounds for the discrepancy in Theorem 1.4 are then readily deduced from Theorem 1.3.

1.6. Complications

The above discussion of the proofs of Theorems 1.1, 1.2, 1.3, and 1.4 glosses over two key obstacles that we must overcome.

1.6.1. Hyperbolic orbifolds of level q . The first obstacle is defining the hyperbolic orbifolds of level q . For $q = 1$, Duke, Imamoğlu, and Tóth [19] associate to a narrow ideal class A of a real quadratic field a Fuchsian group of the second kind $\Gamma_A \leq \text{PSL}_2(\mathbb{Z})$ whose unique boundary component projects to the closed geodesic $\mathcal{C}_A(1) \subset X_0(1)$ associated to A . In the higher level setting, however, it is not possible to define a subgroup of $\Gamma_0(q)$ that is a Fuchsian group of the second kind whose boundary component projects to $\mathcal{C}_A(q) \subset X_0(q)$. This is caused by the possibility of a topological obstruction in the case of nontrivial genus. One can observe this intuitively by considering a loop around

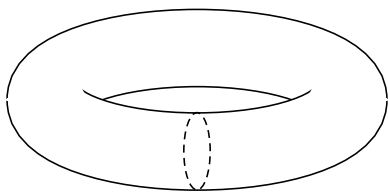


Fig. 1. A loop around one of the holes of the torus.

one of the holes of a torus, as in Figure 1. This loop is not the boundary of a submanifold of the torus, which is precisely due to the fact that the loop is nontrivial in the homology of the torus. In other words, we must compensate for the nontrivial homology of $\mathcal{C}_A(q)$ in a systematic way.

In Section 3, we define a Fuchsian group of the second kind $\Gamma_A(q) \leq \mathrm{PSL}_2(\mathbb{Z})$ (which in general is not contained in $\Gamma_0(q)$) as well as an associated canonical fundamental domain $\mathcal{F}_A(q) \subset \mathbb{H}$ of $\Gamma_A(q) \backslash \mathcal{N}_A(q)$, where $\mathcal{N}_A(q)$ is the associated Nielsen region. The equidistribution problem that we study concerns the image under the map $\mathcal{F}_A(q) \rightarrow X_0(q)$ given by restricting the projection $\mathbb{H} \rightarrow X_0(q)$. The starting point for this construction is a special fundamental polygon $\mathcal{P}(q)$ of $\Gamma_0(q)$ introduced by Kulkarni [46] and studied recently by Doan, Kim, Lang, and Tan [15]. Pictorially, the boundary of the fundamental polygon $\partial\mathcal{P}(q)$ defines a curve inside $X_0(q)$, and we compensate for the nontrivial homology of $\mathcal{C}_A(q)$ by subsegments of $\partial\mathcal{P}(q)$. The boundary $\partial\mathcal{P}(q)$ contributes to the topological terms in the Weyl sums alluded to above, which has not previously been observed before in the context of Duke's theorem. Notably, bounding these topological contributions requires several new ideas that go beyond the work of Duke, Imamoğlu, and Tóth [19], namely averaged bounds for Voronoï L -series and a homological version of the sup-norm problem.

1.6.2. Weyl sums. The second obstacle is relating the Weyl sum (1.4) to a special value of a Rankin–Selberg L -function. Let us first recall how this is done by Duke, Imamoğlu, and Tóth when $q = 1$ and $H = (\mathrm{Cl}_D^+)^2$. In this setting, a character $\chi \in H^\perp$ is a genus character associated to a pair of primitive quadratic Dirichlet characters χ_{D_1}, χ_{D_2} of conductors $|D_1|, |D_2|$, where $D_1, D_2 \in \mathbb{Z}$ are fundamental discriminants for which $D_1 D_2 = D$. In [19, Theorem 4], it is shown that the Weyl sum $W_{\chi, f}$ for this genus character χ is essentially equal to $b(D_1)b(D_2)$, where $b(n)$ denotes the n -th Fourier coefficient of the Shintani lift of the Hecke–Maaß form f of weight zero, namely the Maaß form of weight $\frac{1}{2}$ associated to f via the Shimura–Shintani correspondence, as elucidated by Katok and Sarnak [44]. One then uses Waldspurger's formula [62, Théorème 1], in an explicit form due to Baruch and Mao [5, Theorem 1.4], to relate $|b(D_1)|^2$ and $|b(D_2)|^2$ to the values at $s = \frac{1}{2}$ of the L -functions $L(s, f \otimes \chi_{D_1})$ and $L(s, f \otimes \chi_{D_2})$.

This method breaks down, however, when χ is not a genus character, for then there is no explicit relation between the Weyl sum and the Fourier coefficients of the Shintani lift. We show that nonetheless there is a relation between the square of the absolute value

of the Weyl sum and the special value of an L -function. The key observation is that [19, Lemma 1] identifies the Weyl sum with a weighted sum of cycle integrals, namely integrals of an automorphic form along a closed geodesic. With a careful classical-to-adèlic correspondence, we show that this weighted sum of cycle integrals can in turn be viewed as an adèlic period integral $\mathcal{P}_\Omega(\phi)$, where the integrand is the product of a distinguished automorphic form $\phi: \mathrm{GL}_2(\mathbb{A}_\mathbb{Q}) \rightarrow \mathbb{C}$ lying in the automorphic representation π_f associated to f and the Hecke character $\Omega: E^\times \backslash \mathbb{A}_E^1 \rightarrow \mathbb{C}^\times$ associated to χ .

It is known, going back to the work of Waldspurger [63], that $|\mathcal{P}_\Omega(\phi)|^2$ is essentially equal to the value at $s = \frac{1}{2}$ of the Rankin–Selberg L -function $L(s, \pi_f \otimes \pi_\Omega)$, where π_Ω denotes the automorphic induction of Ω to an automorphic representation of $\mathrm{GL}_2(\mathbb{A}_\mathbb{Q})$. When Ω is the idèlic lift of a narrow class character χ , this L -function may in turn be written as the Rankin–Selberg L -function $L(s, f \otimes \Theta_\chi)$, where Θ_χ denotes the theta series on $\Gamma_0(D) \backslash \mathbb{H}$ associated to χ .

It is instructive to consider the case when the Hecke character $\Omega: E^\times \backslash \mathbb{A}_E^1 \rightarrow \mathbb{C}^\times$ factors through the norm map from $\mathbb{A}_E^\times$ to $\mathbb{A}_\mathbb{Q}^\times$. When this occurs, the automorphic representation π_Ω is noncuspidal; it is the isobaric sum of a pair of Hecke characters $\omega_1, \omega_2: \mathbb{Q}^\times \backslash \mathbb{A}_\mathbb{Q}^1 \rightarrow \mathbb{C}^\times$ and the vector space of automorphic forms associated to π_Ω consists of Eisenstein series, so that the Rankin–Selberg L -function $L(s, \pi_f \otimes \pi_\Omega)$ factorises as $L(s, \pi_f \otimes \omega_1)L(s, \pi_f \otimes \omega_2)$. When Ω is additionally the adèlic lift of a narrow class character χ , then necessarily χ is a genus character associated to a pair of quadratic characters χ_{D_1}, χ_{D_2} whose idèlic lifts are ω_1, ω_2 , so that

$$L(s, \pi_f \otimes \omega_1) = L(s, f \otimes \chi_{D_1}) \quad \text{and} \quad L(s, \pi_f \otimes \omega_2) = L(s, f \otimes \chi_{D_2}).$$

In this way, our result encompasses that of Duke, Imamoğlu, and Tóth.

There is a caveat to our claim that $|\mathcal{P}_\Omega(\phi)|^2$ is essentially equal to $L(\frac{1}{2}, \pi_f \otimes \pi_\Omega)$; this is true only up to multiplication by a constant that depends sensitively on the automorphic form ϕ . Thankfully, this constant has been explicitly evaluated for certain choices of automorphic forms ϕ by Martin and Whitehouse [50], and while the cases of interest required in the proofs of our two main theorems do not quite fall under the umbrella of the cases considered in [50], only mild modifications are needed.

2. Oriented closed geodesics of level q

We begin by recording several details relating oriented ideals and narrow ideal classes of real quadratic fields, integral binary quadratic forms, embeddings of real quadratic fields into spaces of matrices, and closed geodesics on the level q modular surface $\Gamma_0(q) \backslash \mathbb{H}$. Useful references for this material include [29, Section 1], [14, Section 1], and [57, Section 6]. We work throughout with a positive fundamental discriminant $D > 1$ and a squarefree integer q for which every prime dividing q splits in $E := \mathbb{Q}(\sqrt{D})$. We also fix once and for all a residue class r modulo $2q$ for which $r^2 \equiv D \pmod{4q}$; the dependence on the choice of r is briefly discussed in Section 2.5.

2.1. Oriented ideals

Let α be a nonzero fractional ideal of E . Let $(\alpha_1, \alpha_2) \in E^2$ be generators over $\mathbb{Z}$ of α , so that $\alpha = \mathbb{Z}\alpha_1 + \mathbb{Z}\alpha_2$. The (absolute) norm of α is

$$N(\alpha) = \frac{|\alpha_1\sigma(\alpha_2) - \alpha_2\sigma(\alpha_1)|}{\sqrt{D}},$$

where σ denotes the nontrivial Galois automorphism of E . The ideal α is said to be oriented with respect to the ordered pair of generators $(\alpha_1, \alpha_2) \in E^2$ if $\alpha_1\sigma(\alpha_2) - \alpha_2\sigma(\alpha_1) > 0$ and to be of level q if

$$\frac{N(\alpha_1)}{N(\alpha)} \equiv 0 \pmod{q} \quad \text{and} \quad \frac{\text{Tr}(\alpha_1\sigma(\alpha_2))}{N(\alpha)} \equiv r \pmod{2q}.$$

We denote by $[\alpha; \alpha_1, \alpha_2]$ the oriented ideal α with respect to the generators (α_1, α_2) .

The congruence subgroup $\Gamma_0(q) \ni \gamma$ acts on the set of such triples $[\alpha; \alpha_1, \alpha_2]$ by acting trivially on the ideal α and mapping the ordered pair of generators $(\alpha_1, \alpha_2) \in E^2$, viewed as a row vector, to $(\alpha_1, \alpha_2)\gamma$. This action preserves oriented ideals of level q .

Let $\mathcal{O}_E$ denote the ring of integers of E . The set

$$\mathcal{P}_E^+ := \{(\alpha) = \alpha\mathcal{O}_E \subset E : \alpha \in E, \alpha, \sigma(\alpha) > 0\}$$

of totally positive principal fractional ideals – equivalently, the identity I in the narrow class group Cl_D^+ – acts on an oriented ideal $[\alpha; \alpha_1, \alpha_2]$ of level q via the map

$$(\alpha) \cdot [\alpha; \alpha_1, \alpha_2] := [(\alpha)\alpha; \alpha\alpha_1, \alpha\alpha_2],$$

and this action commutes with the action of $\Gamma_0(q)$.

In this way, each narrow ideal class of Cl_D^+ may be bijectively identified with equivalence classes of oriented ideals of level q modulo the action of $\Gamma_0(q)$ and $\mathcal{P}_E^+$, where we identify the identity narrow ideal class I with the equivalence class of oriented ideals containing $[\mathbb{Z}q + \mathbb{Z}\frac{r-\sqrt{D}}{2}; q, \frac{r-\sqrt{D}}{2}]$ for each $r \in \mathbb{Z}$ for which $r^2 \equiv D \pmod{4q}$. The narrow ideal class $J \in \text{Cl}_D^+$ containing the different $\mathfrak{d}$ then corresponds to the equivalence class of oriented ideals containing $[\mathbb{Z}q\sqrt{D} + \mathbb{Z}\frac{D-r\sqrt{D}}{2}; q\sqrt{D}, \frac{D-r\sqrt{D}}{2}]$.

2.2. Binary quadratic forms

For each trio of integers $(a, b, c) \in \mathbb{Z}^3$ having greatest common divisor equal to 1 and satisfying $b^2 - 4ac = D$, $a \equiv 0 \pmod{q}$, and $b \equiv r \pmod{2q}$, we define the integral binary quadratic form

$$Q(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + bxy + cy^2, \quad (2.1)$$

which is a primitive form of level q and discriminant $b^2 - 4ac = D$. We write $Q = [a, b, c]$ to denote this form and call such a form a Heegner form, following [14]; we denote the set of such forms by $\mathcal{Q}_D(q)$.

The congruence subgroup $\Gamma_0(q) \ni \gamma$ acts on Heegner forms via

$$(\gamma \cdot Q)(x, y) := Q((x, y)^t \gamma) = \begin{pmatrix} x & y \end{pmatrix}^t \gamma \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \gamma \begin{pmatrix} x \\ y \end{pmatrix},$$

where ${}^t\gamma$ denotes the transpose of γ and we view (x, y) as a row vector. Moreover, this action preserves $\mathcal{Q}_D(q)$.

To each equivalence class $\mathcal{P}_E^+ \cdot [\alpha; \alpha_1, \alpha_2]$ of oriented ideals of level q , we associate the Heegner form $Q = Q_{\mathcal{P}_E^+ \cdot [\alpha; \alpha_1, \alpha_2]} \in \mathcal{Q}_D(q)$ given by

$$Q(x, y) := \frac{N(\alpha_1 x + \alpha_2 y)}{N(\alpha)}.$$

Conversely, associated to each Heegner form $Q = [a, b, c] \in \mathcal{Q}_D(q)$ as in (2.1) is the equivalence class of oriented ideals of level q given by

$$\begin{cases} \mathcal{P}_E^+ \cdot \left[\mathbb{Z}a + \mathbb{Z}\frac{b - \sqrt{D}}{2}; a, \frac{b - \sqrt{D}}{2} \right] & \text{if } a > 0, \\ \mathcal{P}_E^+ \cdot \left[\mathbb{Z}(-a\sqrt{D}) + \mathbb{Z}\frac{D - b\sqrt{D}}{2}; -a\sqrt{D}, \frac{D - b\sqrt{D}}{2} \right] & \text{if } a < 0. \end{cases} \quad (2.2)$$

This map is a bijection between $\mathcal{Q}_D(q)$ and equivalence classes of oriented ideals of level q . This association descends to a bijection between narrow ideal classes of the narrow class group Cl_D^+ and equivalence classes of primitive integral binary quadratic forms of level q and discriminant D modulo the action of $\Gamma_0(q)$. In particular, for each $r \in \mathbb{Z}$ for which $r^2 \equiv D \pmod{4q}$, the equivalence class of elements of $\mathcal{Q}_D(q)$ modulo $\Gamma_0(q)$ that contains $[q, r, \frac{r^2 - D}{4q}]$ corresponds to the principal narrow ideal class $I \in \text{Cl}_D^+$, while the equivalence class containing $[-q, r, \frac{D - r^2}{4q}]$ corresponds to the narrow ideal class J containing the different $\mathfrak{d}$.

2.3. Oriented embeddings

Again, let $(a, b, c) \in \mathbb{Z}^3$ have greatest common divisor equal to 1 and satisfy $b^2 - 4ac = D$, $a \equiv 0 \pmod{q}$, and $b \equiv r \pmod{2q}$. We define an embedding $\Psi: E \hookrightarrow \text{Mat}_{2 \times 2}(\mathbb{Q})$ by

$$\Psi(x + \sqrt{D}y) := \begin{pmatrix} x + by & 2cy \\ -2ay & x - by \end{pmatrix} \quad (2.3)$$

for $x, y \in \mathbb{Q}$. This satisfies

$$\Psi(E) \cap \{g \in \text{Mat}_{2 \times 2}(\mathbb{Z}) : g_{2,1} \equiv 0 \pmod{q}\} = \Psi(\mathcal{O}_E);$$

that is, Ψ is an oriented optimal embedding of level q . Conversely, every oriented optimal embedding of level q arises from such a trio of integers $(a, b, c) \in \mathbb{Z}^3$.

The congruence subgroup $\Gamma_0(q)$ acts on the set of optimal oriented embeddings of level q by conjugation, namely

$$(\gamma \cdot \Psi)(x + \sqrt{D}y) := \gamma^{-1} \Psi(x + \sqrt{D}y) \gamma$$

for $\gamma \in \Gamma_0(q)$, and this action preserves optimal oriented embeddings of level q .

There is a natural bijection between oriented optimal embeddings Ψ of level q as in (2.3) and Heegner forms $Q = [a, b, c]$ as in (2.1), since these are both completed determined by $(a, b, c) \in \mathbb{Z}^3$; in turn, there is a bijection with equivalence classes of oriented ideals of level q as in (2.2). Again, this descends to a bijection between narrow ideal classes A of the narrow class group Cl_D^+ and equivalence classes of oriented optimal embeddings of level q modulo the action of $\Gamma_0(q)$. In particular, the equivalence classes of oriented ideals of level q corresponding to I and J contain the optimal embeddings of level q given, respectively, by

$$\Psi_I(x + \sqrt{D}y) := \begin{pmatrix} x + ry & \frac{r^2 - D}{2q}y \\ -2qy & x - ry \end{pmatrix}, \quad \Psi_J(x + \sqrt{D}y) := \begin{pmatrix} x + ry & \frac{D - r^2}{2q}y \\ 2qy & x - ry \end{pmatrix}.$$

2.4. Closed geodesics

Associated to a Heegner form $Q = [a, b, c] \in \mathcal{Q}_D(q)$ as in (2.1) is an oriented geodesic S_Q in the upper half-plane connecting the two points $\frac{-b - \sqrt{D}}{2a}$ and $\frac{-b + \sqrt{D}}{2a}$, namely the Euclidean semicircle

$$S_Q := \{z \in \mathbb{H} : a|z|^2 + b\text{Re}(z) + c = 0\} \quad (2.4)$$

oriented anticlockwise if $a > 0$ and clockwise if $a < 0$.

Let $\epsilon_D > 1$ be the least unit with positive norm in $\mathcal{O}_E$, so that $\epsilon_D = u + \sqrt{D}v$ with u, v positive half-integers that satisfy Pell's equation $u^2 - Dv^2 = 1$ and minimise v among all such positive half-integral solutions. For the oriented optimal embedding Ψ of level q as in (2.3) associated to Q , define the matrix

$$\gamma_Q := \Psi(\epsilon_D) = \begin{pmatrix} u + bv & 2cv \\ -2av & u - bv \end{pmatrix} \in \Gamma_0(q). \quad (2.5)$$

Together with $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, this generates the group of automorphs of Q ,

$$\Gamma_0(q)_Q := \{\gamma \in \Gamma_0(q) : \gamma \cdot Q = Q\}.$$

Note that γ_Q is a hyperbolic matrix, as $\text{Tr}(\gamma_Q) = 2u = 2\sqrt{1 + Dv^2} > 2$.

Let A be the narrow ideal class associated to the equivalence class of Heegner forms modulo $\Gamma_0(q)$ that contains Q . We let $\mathcal{C}_A(q) := \Gamma_0(q)_Q \backslash S_Q$ denote the oriented closed geodesic in $\Gamma_0(q) \backslash \mathbb{H}$ corresponding to A , which we may view explicitly as the reduction modulo $\Gamma_0(q)$ of the oriented geodesic segment from z_Q to $\gamma_Q z_Q$, where

$$z_Q := \begin{cases} \frac{-b + i\sqrt{D}}{2a} & \text{if } a > 0, \\ \frac{b + i\sqrt{D}}{-2a} & \text{if } a < 0, \end{cases} \quad (2.6)$$

$$\gamma_Q z_Q = \begin{cases} \frac{-b(u^2 + Dv^2) - 2Duv + i\sqrt{D}}{2a(u^2 + Dv^2)} & \text{if } a > 0, \\ \frac{b(u^2 + Dv^2) + 2Duv + i\sqrt{D}}{-2a(u^2 + Dv^2)} & \text{if } a < 0. \end{cases}$$

2.5. Atkin–Lehner operators

We return to our choice of r modulo $2q$ for which $r^2 \equiv D \pmod{4q}$. There are $2^{\omega(q)}$ such choices, and for each choice, there is an associated collection of h_D^+ distinct oriented closed geodesics in $\Gamma_0(q) \backslash \mathbb{H}$. Each of these collections of oriented closed geodesics is permuted by the $2^{\omega(q)}$ Atkin–Lehner operators on $\Gamma_0(q)$. More precisely, for each divisor q_1 of q , so that $q = q_1 q_2$, we let

$$W_{q_1} := \begin{pmatrix} a\sqrt{q_1} & \frac{b}{\sqrt{q_1}} \\ cq_2\sqrt{q_1} & d\sqrt{q_1} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \quad (2.7)$$

be an Atkin–Lehner operator on $\Gamma_0(q)$ associated to q_1 , where $a, b, c, d \in \mathbb{Z}$ are such that $adq_1 - bcq_2 = 1$. Any two such Atkin–Lehner operators associated to q_1 are equivalent modulo $\Gamma_0(q)$. These operators are elements of the normaliser of $\Gamma_0(q)$ and permute the cusps of $\Gamma_0(q) \backslash \mathbb{H}$. They act on oriented ideals, Heegner forms, oriented embeddings, and oriented closed geodesics; an Atkin–Lehner operator W_{q_1} , where $q_1 q_2 = q$, has the effect of replacing $r \in \mathbb{Z}/2q\mathbb{Z}$ with the element $r' \in \mathbb{Z}/2q\mathbb{Z}$ that satisfies $r' \equiv -r \pmod{2q_1}$ and $r' \equiv r \pmod{2q_2}$.

3. Hyperbolic orbifolds of level q

Let $D > 1$ be a positive fundamental discriminant and let q be an odd prime² that splits in $E = \mathbb{Q}(\sqrt{D})$; as in Section 2.4, we fix once and for all a residue class r modulo $2q$ for which $r^2 \equiv D \pmod{4q}$. Our goal for this section is to construct a hyperbolic orbifold whose boundary is the closed geodesic on $X_0(q)$ associated to an element of the narrow class group Cl_D^+ . Due to the possibility of nontrivial genus, these orbifolds have some new and interesting features that do not show up in the level 1 case.

3.1. Special fundamental domains

By the Kurosh subgroup theorem, any subgroup of

$$\mathrm{PSL}_2(\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z}$$

is isomorphic to a free product of a number of copies of $\mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/3\mathbb{Z}$, and $\mathbb{Z}$. In particular, if $\Gamma_0(q)$ is a torsion-free Hecke congruence subgroup, then it is a free group on $\mathrm{rank}_{\mathbb{Z}} \Gamma_0(q)^{\mathrm{ab}}$ generators. We now describe an explicit geometric way, following [46], for constructing an independent generating set using so-called *special fundamental polygons* of $\Gamma_0(q)$.

²Much of what we prove below generalises in a straightforward manner to squarefree level q (at least when there is no 3-torsion), but we stick to the case of q prime for simplicity.

3.1.1. Farey symbols of level q . Here and below, we denote by g the genus of $X_0(q)$ (suppressing q from the notation), and by $e_2 \in \{0, 2\}$ and $e_3 \in \{0, 2\}$ the number of conjugacy classes of subgroups in $\Gamma_0(q)$ of order 2 and of order 3, respectively (see [40, (2.12) and (2.13)]). We write

$$T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

for the standard generators of $\mathrm{PSL}_2(\mathbb{Z})$.

Definition 3.1 ([46]). A Farey symbol of level q is a sequence of reduced fractions,

$$\frac{0}{1} = \frac{a_0}{b_0} < \frac{a_1}{b_1} < \cdots < \frac{a_{n-1}}{b_{n-1}} < \frac{a_n}{b_n} = \frac{1}{1},$$

where $n := 4g + e_2 + e_3$, such that

- $a_{i+1}b_i - a_ib_{i+1} = 1$ for all $i \in \{0, \dots, n\}$,
- there are e_2 even indices i such that

$$b_i^2 + b_{i+1}^2 \equiv 0 \pmod{q},$$

- there are e_3 odd indices i such that

$$b_i^2 + b_ib_{i+1} + b_{i+1}^2 \equiv 0 \pmod{q},$$

- for the remaining $4g$ free indices, there is a pairing $i \leftrightarrow i^*$ satisfying

$$b_ib_{i^*} + b_{i+1}b_{i^*+1} \equiv 0 \pmod{q}.$$

Here we consider the indices i above modulo $n + 2$ by setting $a_{-1} = -1$, $b_{-1} = b_{n+1} = 0$, and $a_{n+1} = 1$.

Such a Farey symbol of level q always exists; moreover, one can even find one that is symmetric around $\frac{1}{2}$, so that $a_{n-i} = 1 - a_i$ and $b_{n-i} = b_i$ for each index i [46, Section 13]. Dooms, Jespers, and Kononov [16] have described an algorithm for determining Farey symbols of arbitrary level.

It is easily seen that any Farey symbol of level q satisfies the bound $b_i \ll e^{O(q)}$ for the denominators; on the other hand, we have the lower bound $\max_i |b_i| \gg \sqrt{q}$. It has recently been shown by Doan, Kim, Lang, and Tan that furthermore there exist Farey symbols of prime level that are minimal in the sense that this lower bound is essentially sharp.

Theorem 3.2 ([15, Theorem 1.1]). *Let q be an odd prime. There exists a Farey symbol of level q that is symmetric about $\frac{1}{2}$ with $b_i \leq \lfloor \sqrt{\frac{4q}{3}} \rfloor$ for each index $i \in \{0, \dots, n\}$, and such that*

$$\begin{aligned} b_i^2 + b_{i+1}^2 &= q && \text{for all even indices } i, \\ b_i^2 + b_ib_{i+1} + b_{i+1}^2 &= q && \text{for all odd indices } i, \text{ and} \\ b_ib_{i^*} + b_{i+1}b_{i^*+1} &= q && \text{for all free indices } i. \end{aligned}$$

From here onward, we let q be an odd prime and fix a Farey symbol of level q as in Theorem 3.2. Following [46, Section 2], we denote by $\mathcal{P}(q)$ the hyperbolic polygon in $\mathbb{H}$ with the following $e_2 + e_3 + n + 2$ vertices:

- ∞ ;
- the $n + 1$ fractions $\frac{a_i}{b_i}$ of the Farey symbol for each index $i \in \{0, \dots, n\}$;
- the e_2 $\mathrm{PGL}_2(\mathbb{Z})$ -translates of $\sqrt{-1}$ on the infinite geodesic connecting $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$ for each even index $i \in \{1, \dots, n-1\}$; and
- the e_3 $\mathrm{PGL}_2(\mathbb{Z})$ -translates of $\frac{1+i\sqrt{3}}{2}$ lying just below the infinite geodesic connecting $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$ for each odd index $i \in \{1, \dots, n-1\}$.

As in [46, Section 3], we define a side pairing on $\mathcal{P}(q)$ by pairing the following edges:

- the two vertical geodesics sharing the cusp ∞ ;
- the two geodesic segments connecting $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$ with the $\mathrm{PGL}_2(\mathbb{Z})$ -translate of $\sqrt{-1}$ lying between them for each of the e_2 even indices i ;
- the two geodesic segments connecting $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$ with the $\mathrm{PGL}_2(\mathbb{Z})$ -translate of $\frac{1+\sqrt{-3}}{2}$ lying between them for each of the e_3 odd indices i ; and
- the infinite geodesic between the cusps $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$ with the infinite geodesic between the cusps $\frac{a_{i^*+1}}{b_{i^*+1}}$ and $\frac{a_{i^*}}{b_{i^*}}$ for each of the $4g$ free indices i .

Definition 3.3. We define a *side* of $\mathcal{P}(q)$ to be a maximal geodesic segment on the boundary of $\mathcal{P}(q)$ (note that there is only one side for each even index). Given a side $\mathcal{L}$ of $\mathcal{P}(q)$ that is paired with a corresponding side $\mathcal{L}'$ (which might be $\mathcal{L}$ itself), we define the *label* of $\mathcal{L}$ as the element α of $\mathrm{PSL}_2(\mathbb{Z})$ that maps $\mathcal{L}'$ to $\mathcal{L}$ as defined by Kulkarni (see below). The set of all labels is called the set of *side pairing elements* of $\mathcal{P}(q)$.

By Poincaré's polygon theorem [51], the side pairing elements define a finite index subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ that has $\mathcal{P}(q)$ as a fundamental domain; by [46, Theorem 13.2], this subgroup is the congruence subgroup $\Gamma_0(q)$. Furthermore, since $\mathcal{P}(q)$ has a minimal number of sides (which follows from [40, Proposition 2.6]), the labels define an independent set of generators (upon forgetting inverses) of $\Gamma_0(q)$.

We denote by

$$\{iy : y > 0\} =: \mathcal{L}_1, \quad \mathcal{L}_2, \quad \dots, \quad \mathcal{L}_{N-1}, \quad \mathcal{L}_N := \{1 + iy : y > 0\},$$

the $N := 4g + 2 + e_2 + 2e_3$ oriented sides of $\mathcal{P}(q)$ taken in the positive orientation starting from the side containing 0 and ∞ . We refer to the vertex of $\mathcal{L}_j$ shared with $\mathcal{L}_{j-1}$ as the *left-most* or *first* vertex of $\mathcal{L}_j$ (which is ∞ for $j = 1$) and the vertex of $\mathcal{L}_j$ shared with $\mathcal{L}_{j+1}$ as the *right-most* or *second* vertex (which is ∞ for $j = N$).

We denote by α_j the label in $\Gamma_0(q)$ associated to the oriented side $\mathcal{L}_j$ of $\mathcal{P}(q)$, so that $\alpha_1 = T^{-1}$ and $\alpha_N = T$. From [46, Theorem 6.1], these labels are explicitly as follows:

- for a side $\mathcal{L}_j$ associated to an even index i ,

$$\alpha_j = \begin{pmatrix} a_{i+1}b_{i+1} + a_i b_i & -a_i^2 - a_{i+1}^2 \\ b_i^2 + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_i b_i \end{pmatrix};$$

- for a side $\mathcal{L}_j$ associated to an odd index i ,

$$\alpha_j = \begin{pmatrix} a_{i+1}b_{i+1} + a_i b_{i+1} + a_i b_i & -a_i^2 - a_i a_{i+1} - a_{i+1}^2 \\ b_i^2 + b_i b_{i+1} + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_{i+1}b_i - a_i b_i \end{pmatrix}^{\pm 1},$$

where $\pm = +$ if $\frac{a_i}{b_i}$ is to the left of the associated $\mathrm{PGL}_2(\mathbb{Z})$ -translate of $\frac{1+\sqrt{-3}}{2}$, while $\pm = -$ if $\frac{a_i}{b_i}$ is to the right;

- for a side $\mathcal{L}_j$ associated to a free index i ,

$$\alpha_j = \begin{pmatrix} a_{i^*+1}b_{i+1} + a_{i^*}b_i & -a_i a_{i^*} - a_{i+1}a_{i^*+1} \\ b_i b_{i^*} + b_{i+1}b_{i^*+1} & -a_{i+1}b_{i^*+1} - a_i b_{i^*} \end{pmatrix}.$$

We say that a side $\mathcal{L}_j$ is *hyperbolic*, *elliptic*, or *parabolic* if the associated label α_j is a hyperbolic, elliptic, or parabolic matrix, respectively; in particular, sides associated to even or odd indices are elliptic. We note that since $0 \leq a_i \leq b_i$ and $b_i \ll \sqrt{q}$, the Frobenius norm $\sqrt{\mathrm{Tr}(\alpha_j {}^t\alpha_j)}$ of any label α_j of $\mathcal{P}(q)$ is bounded by $O(q)$.

Here and below, we consider the indices j of $\mathcal{L}_j$ and α_j modulo N . By a slight abuse of notation, we denote by $*$: $\mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}$ the side pairing, so that for $j \in \mathbb{Z}/N\mathbb{Z}$,

$$\alpha_j \mathcal{L}_{j^*} = \mathcal{L}_j, \quad \alpha_{j^*} \mathcal{L}_j = \mathcal{L}_{j^*}.$$

3.1.2. Orbits of consecutive polygons. It is key for us to understand the orbits of the map $t: \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}$ given by

$$t(j) := j^* - 1.$$

To see why this map is relevant, let $\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2$ be three consecutive $\Gamma_0(q)$ -translates of $\mathcal{P}(q)$ that all contain some cusp $b \in \mathbb{P}^1(\mathbb{Q})$ (ordered in positive orientation). Let α_{j_1} be the label of the side of $\mathcal{P}_0$ shared with $\mathcal{P}_1$ and let α_{j_2} be the label of the side of $\mathcal{P}_1$ shared with $\mathcal{P}_2$. Then we have the equality

$$j_2 = j_1^* - 1 = t(j_1).$$

In other words, the map t encodes labels of consecutive translates of $\mathcal{P}(q)$.

Lemma 3.4. *Let $\mathcal{E} \subset \mathbb{Z}/N\mathbb{Z}$ denote the subset of indices j for which α_j is either elliptic of order 3 or parabolic and additionally $\mathcal{L}_j$ is to the left of $\mathcal{L}_{j^*}$. The map $t: \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}$ has $e_3 + 2$ orbits explicitly given by*

$$\{j\} \text{ for each } j \in \mathcal{E} \text{ and } (\mathbb{Z}/N\mathbb{Z}) \setminus \mathcal{E}.$$

Furthermore, for any t -orbit $\mathcal{S} \subset \mathbb{Z}/N\mathbb{Z}$ and for all $j \in \mathcal{S}$, the element of $\Gamma_0(q)$, given by

$$\alpha_j \alpha_{t(j)} \cdots \alpha_{t^{(|\mathcal{S}|-1)}(j)},$$

is either elliptic or parabolic.

Proof. First of all, for $j \in \mathcal{E}$ we have that $t(j) = j + 1 - 1 = j$. Next, consider the orbit $\mathcal{S}$ of 1 in $\mathbb{Z}/N\mathbb{Z}$. Observe that for $j \in \mathcal{S}$, the second vertex of the side $\mathcal{L}_j$ of $\mathcal{P}(q)$ is

cuspidal and not equal to ∞ . Pick an element $j \in \mathcal{S}$ and let $\mathfrak{b} \in \mathbb{P}^1(\mathbb{Q})$ be the second vertex of the side $\mathcal{L}_j$. Then the element

$$\alpha := \alpha_j \alpha_{t(j)} \alpha_{t(t(j))} \cdots \alpha_{t(|\mathcal{S}|-1)(j)} \in \Gamma_0(q)$$

lies in the stabiliser of $\mathfrak{b}$ as it maps the side $\mathcal{L}_j$ of $\mathcal{P}(q)$ containing $\mathfrak{b}$ to the side $\alpha \mathcal{L}_j$ of $\alpha \mathcal{P}(q)$ containing $\mathfrak{b}$ (using here that $t^{(|\mathcal{S}|)}(j) = j$). This implies that the width of the cusp between $\mathcal{L}_j$ and $\alpha \mathcal{L}_j$ is at least q (as this is the width of the cusp 0 in $\Gamma_0(q)$). On the other hand, all of the cuspidal zones near $\mathfrak{b}$ of

$$\alpha_j \mathcal{P}(q), \quad \alpha_j \alpha_{t(j)} \mathcal{P}(q), \quad \dots, \quad \alpha_j \alpha_{t(j)} \cdots \alpha_{t(|\mathcal{S}|-2)(j)} \mathcal{P}(q), \quad \alpha \mathcal{P}(q)$$

are $\Gamma_0(q)$ -inequivalent. Thus we conclude that α is in fact a generator of the stabiliser of $\mathfrak{b}$ and furthermore that $\mathcal{S}$ must contain all indices not in $\mathcal{E}$, as these are precisely the indices whose associated side has second vertex equal to a cusp in the $\Gamma_0(q)$ -orbit of 0. ■

We note that the size of the orbit containing $1 \in \mathbb{Z}/N\mathbb{Z}$ is equal to $4g + e_2 + e_3 + 1$. In particular, it follows from the above that

$$\alpha_1 \alpha_{t(1)} \cdots \alpha_{t(4g+e_2+e_3+1)(1)} = \begin{pmatrix} 1 & 0 \\ -q & 1 \end{pmatrix},$$

as the left-hand side is equal to the generator of the stabiliser in $\Gamma_0(q)$ of 0 and maps ∞ to a negative rational. Here we recall that $\alpha_{t(4g+e_2+e_3+1)(1)} = \alpha_1 = T^{-1}$.

Lemma 3.5. *For each $j \in \{0, \dots, 4g + e_2 + e_3\}$, the Frobenius norm of the matrix*

$$\alpha_1 \alpha_{t(1)} \cdots \alpha_{t(j)(1)} \in \Gamma_0(q),$$

is bounded by $O(q^{\frac{3}{2}})$.

Proof. Put

$$\gamma_j := \alpha_1 \alpha_{t(1)} \cdots \alpha_{t(j)(1)}. \tag{3.1}$$

Then the fundamental domain $\gamma_j \mathcal{P}(q)$ contains the cusp 0. This means that

$$\gamma_j \frac{a_i}{b_i} = 0 \tag{3.2}$$

for some vertex $\frac{a_i}{b_i}$ of $\mathcal{P}(q)$. Since all the cuspidal zones near 0 of

$$\alpha_1 \mathcal{P}(q), \quad \alpha_1 \alpha_{t(1)} \mathcal{P}(q), \quad \dots, \quad \gamma_j \mathcal{P}(q)$$

are $\Gamma_0(q)$ -inequivalent, we see that $\gamma_j \in \Gamma_0(q)$ is characterised as the element of $\Gamma_0(q)$ satisfying (3.2) for which $\gamma_j \infty$ is negative such that $|\gamma_j \infty|$ is maximised. This means explicitly that

$$\gamma_j = \begin{pmatrix} b_i & -a_i \\ -qx_0 & y_0 \end{pmatrix},$$

where $x_0 \in \mathbb{N}$ is the minimal positive solution to the congruence $a_i qx_0 \equiv -1 \pmod{b_i}$ and $y_0 := \frac{1+a_i qx_0}{b_i}$. This gives the desired result since $b_i \ll \sqrt{q}$. ■

3.2. Orbifolds associated to real quadratic fields

Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(q)$ be a primitive hyperbolic matrix with axis

$$S_\gamma := \{z \in \mathbb{H} : c|z|^2 + (d-a)\operatorname{Re}(z) - b = 0\} \quad (3.3)$$

oriented anticlockwise if $c < 0$ and clockwise if $c > 0$ and intersecting the fundamental polygon $\mathcal{P}(q)$ (which can always be arranged by conjugation). We denote by $\mathcal{C}_\gamma(q)$ the corresponding oriented closed geodesic in $\Gamma_0(q) \backslash \mathbb{H}$, namely $\mathcal{C}_\gamma(q) := \Gamma_0(q)_\gamma \backslash S_\gamma$ with $\Gamma_0(q)_\gamma := \{\gamma' \in \Gamma_0(q) : \gamma'^{-1}\gamma\gamma' = \gamma\}$ the associated group of automorphisms.

Let $\alpha_{i(0)}$ be the label of the side containing the second intersection between S_γ and $\mathcal{P}(q)$, $\alpha_{i(1)}$ the label of the side containing the second intersection between S_γ and $\alpha_{i(0)}\mathcal{P}(q)$, and so on. As the projection of S_γ onto $\mathcal{P}(q)$ is a *closed* geodesic, this defines a *periodic* sequence of side pairing elements (see, e.g., [43])

$$\alpha_{i(0)}, \alpha_{i(1)}, \alpha_{i(2)}, \dots$$

Let $m_\gamma + 1$ denote the period of this sequence. We then have the following representation of γ in terms of the labels of $\mathcal{P}(q)$, namely the *Morse code* of γ :

$$\gamma = \alpha_{i(0)}\alpha_{i(1)} \cdots \alpha_{i(m_\gamma)}. \quad (3.4)$$

The representation of γ in terms of the side pairing elements of $\mathcal{P}(q)$ is unique up to the relation $\alpha_j^2 = \alpha_{j^*}$ for the elliptic labels α_j of order 3. We consider the index as a map $i: \mathbb{Z}/(m_\gamma + 1)\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}$. For $i_1, i_2 \in \mathbb{Z}/N\mathbb{Z}$, we denote by $\operatorname{arc}(i_1, i_2)$ the indices that lie on the positively oriented open arc strictly between i_1 and i_2 when $\mathbb{Z}/N\mathbb{Z}$ is embedded into the unit circle S^1 via the homomorphism defined by

$$1 \pmod{N} \mapsto e^{\frac{2\pi i}{N}}.$$

Thus if we view $i_1, i_2 \in \mathbb{Z}/N\mathbb{Z}$ as positive integers in $\{1, \dots, N\}$, then

$$\operatorname{arc}(i_1, i_2) = \begin{cases} \{i_1 + 1, \dots, i_2 - 1\} & \text{if } i_1 < i_2, \\ \emptyset & \text{if } i_1 = i_2, \\ \{1, \dots, i_2 - 1, i_1 + 1, \dots, N\} & \text{if } i_1 > i_2. \end{cases}$$

Given an integer $k \in \mathbb{Z}/N\mathbb{Z}$, we define $\sigma_k \in \operatorname{PSL}_2(\mathbb{Z})$ by

$$\sigma_k := \begin{cases} \alpha_k & \text{if } \alpha_k \text{ is elliptic,} \\ \begin{pmatrix} a_{i+1} & a_i \\ b_{i+1} & b_i \end{pmatrix} S \begin{pmatrix} a_{i+1} & a_i \\ b_{i+1} & b_i \end{pmatrix}^{-1} & \text{otherwise,} \end{cases}$$

where the index i is such that the side $\mathcal{L}_k$ associated to α_k is contained in the strip between $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$. Note that if i is a free index, then

$$\sigma_k = \begin{pmatrix} a_{i+1}b_{i+1} + a_ib_i & -a_i^2 - a_{i+1}^2 \\ b_i^2 + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_ib_i \end{pmatrix},$$

which is not an element of $\Gamma_0(q)$ since by definition i is not an even index. Note also that if i is a free or even index, then α_k is a π -rotation around $\mathcal{L}_k$.

We now construct a subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ from the Morse code of γ . Special care has to be given to labels of order 3 since the geometry is different at the corresponding sides. We define

$$\delta_+(j) := \begin{cases} 1 & \text{if } \alpha_{i(j)}^3 = 1 \text{ and } i(j)^* \in \mathrm{arc}(i(j), i(j-1)^*), \\ 0 & \text{otherwise,} \end{cases}$$

$$\delta_-(j) := \begin{cases} 1 & \text{if } \alpha_{i(j-1)}^3 = 1 \text{ and } i(j-1) \in \mathrm{arc}(i(j), i(j-1)^*), \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{J}_3 \subset \{0, \dots, m_\gamma\}$ denote the set of j such that $\alpha_{i(j-1)}^3 = 1$, $i(j-1) \in \mathrm{arc}(i(j), i(j-1)^*)$ and $i(j-1) \neq i(j-2)$. For $j \notin \mathcal{J}_3$, we define

$$\Omega_{\gamma,j}(q) := \{\alpha_{i(0)} \cdots \alpha_{i(j-1)} \sigma_k \alpha_{i(j-1)}^{-1} \cdots \alpha_{i(0)}^{-1} : \\ k \in \mathrm{arc}(i(j) + \delta_+(j), i(j-1)^* - \delta_-(j))\}, \quad (3.5)$$

and for $j \in \mathcal{J}_3$ (in which case $\delta_-(j) = 1$), we define

$$\Omega_{\gamma,j}(q) := \{\alpha_{i(0)} \cdots \alpha_{i(j-1)} \sigma_k \alpha_{i(j-1)}^{-1} \cdots \alpha_{i(0)}^{-1} : \\ k \in \mathrm{arc}(i(j) + \delta_+(j), i(j-1)^* - \delta_-(j))\} \\ \cup \{\alpha_{i(0)} \cdots \alpha_{i(j-2)} \alpha_{i(j-1)}^* \sigma_k \alpha_{i(j-1)}^{-1} \cdots \alpha_{i(0)}^{-1} : \\ k \in \mathrm{arc}(i(j-1)^*, i(j-1))\}. \quad (3.6)$$

As we shall see in the next section, the additional set of matrices in the case $j \in \mathcal{J}_3$ corresponds geometrically to “adding an extra copy of $\mathcal{P}(q)$ ” necessary to ensuring that the fundamental domain is convex.

We put

$$\Omega_\gamma(q) := \bigcup_{j=0}^{m_\gamma} \Omega_{\gamma,j}(q)$$

and define the following subgroup of $\mathrm{PSL}_2(\mathbb{Z})$:

$$\Gamma_\gamma(q) := \langle \Omega_\gamma(q), \gamma \rangle.$$

We shall see in the next section that (outside of one exceptional case) $\Gamma_\gamma(q)$ is a *thin subgroup* of $\mathrm{PSL}_2(\mathbb{Z})$, meaning Zariski dense and of infinite index. Note in particular that $\Omega_\gamma(q)$ is nonempty since γ is hyperbolic. We denote by $\mathcal{N}_\gamma(q)$ the *Nielsen region* (or *convex core*) for $\Gamma_\gamma(q)$, namely the smallest nonempty $\Gamma_\gamma(q)$ -invariant open convex subset of $\mathbb{H}$. Notice that we have a well-defined map $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q) \rightarrow X_0(1)$ arising from the inclusion $\Gamma_\gamma(q) \subset \mathrm{PSL}_2(\mathbb{Z})$. We do not, however, have a canonical map from $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$ to $X_0(q)$. In order to define such a map, we construct a canonical fundamental domain for $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$ inside $\mathbb{H}$ and then consider the projection of this to $X_0(q)$.

3.2.1. A fundamental domain for $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$. We begin by constructing an explicit fundamental domain for $\Gamma_\gamma(q) \backslash \mathbb{H}$. This allows us to define a map $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q) \rightarrow X_0(q)$ and furthermore to study the geometry of $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$. There are geometrical obstacles if $\alpha_{i(m_\gamma)}^3 = 1$ and $i(m_\gamma) \in \text{arc}(i(0), i(m_\gamma)^*)$. By conjugation, we can always arrange so that this is *not* the case unless there is 3-torsion and $\gamma \in \{\alpha_{k_1} \alpha_{k_2}, \alpha_{k_2} \alpha_{k_1}\}$ with $1 < k_1 < k_2 < N$ and $(\mathcal{L}_{k_1}, \mathcal{L}_{k_1+1}), (\mathcal{L}_{k_2}, \mathcal{L}_{k_2+1})$ the two pairs of sides of $\mathcal{P}(q)$ with order 3 labels (this follows by a quick case study recalling that $e_3 \in \{0, 2\}$). If the discriminant $D > 0$ of $\alpha_{k_1} \alpha_{k_2}$ is fundamental, we can exclude it from our considerations by changing the sign of the orientation r modulo $2q$ as in Section 2 (i.e., $r^2 \equiv D \pmod{4q}$). Thus we will henceforth assume that $\gamma \in \Gamma_0(q)$ is primitive hyperbolic and *non-exceptional*, meaning that either $\alpha_{i(m_\gamma)}^3 \neq 1$ or $i(m_\gamma) \notin \text{arc}(i(0), i(m_\gamma)^*)$.

With this in mind, let $\tilde{\mathcal{F}}_\gamma(q)$ be the hyperbolic polygon bounded by the following geodesic segments:

- the complete geodesics containing

$$\mathcal{L}_{i(m_\gamma)^*}, \quad \gamma \mathcal{L}_{i(m_\gamma)^*},$$

- for each $j \in \{0, \dots, m_\gamma\}$ and each $k \in \text{arc}(i(j), i(j-1)^*)$, the geodesic segment

$$\alpha_{i(0)} \cdots \alpha_{i(j-1)} \mathcal{L}_k,$$

- for each $j \in \mathcal{I}_3$, as defined above equation (3.5), and each $k \in \text{arc}(i(j-1)^*, i(j-1))$, the geodesic segment

$$\alpha_{i(0)} \cdots \alpha_{i(j-2)} \alpha_{i(j-1)^*} \mathcal{L}_k.$$

Note that all of the geodesic segments are disjoint by the assumption that γ is non-exceptional. Moreover, $\tilde{\mathcal{F}}_\gamma(q)$ is convex by construction. If α_k is hyperbolic, parabolic, or elliptic of order 2, the geodesic segment $\alpha_{i(0)} \cdots \alpha_{i(j-1)} \mathcal{L}_k$ is mapped to itself by the corresponding element of $\Omega_{\gamma,j}(q)$, whereas if α_k is of order 3, the corresponding element of $\Omega_{\gamma,j}(q)$ maps this side to

$$\alpha_{i(0)} \cdots \alpha_{i(j-1)} \mathcal{L}_{k \pm 1},$$

where $\pm$ depends on whether $\mathcal{L}_k$ is the left-most or right-most side in the elliptic pair of sides.

From the above, we define the hyperbolic polygon

$$\mathcal{F}_\gamma(q) := \tilde{\mathcal{F}}_\gamma(q) \cap \mathcal{N}_\gamma(q) \subset \mathbb{H}, \quad (3.7)$$

which we shall soon see is a fundamental domain for $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$, and in fact that $\mathcal{F}_\gamma(q) = \tilde{\mathcal{F}}_\gamma(q) \setminus \text{int}(S_\gamma)$, where $\text{int}(S_\gamma) \subset \mathbb{H}$ denotes the interior (relative to the orientation) of the axis S_γ of γ (see Figures 2 and 3).

Example 3.6. It is instructive to consider the case of $q = 11$. The genus of $X_0(11)$ is $g = 1$, while there are no conjugacy classes in $\Gamma_0(11)$ of order 2 or 3, so that $e_2 = e_3 = 0$, and consequently

$$n = 4g + e_2 + e_3 = 4 \quad \text{and} \quad N = n + e_3 + 2 = 6.$$

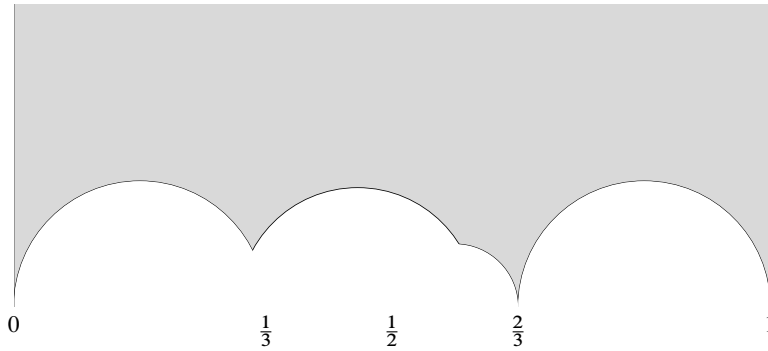


Fig. 2. The fundamental polygon $\mathcal{F}_\gamma(11)$ associated to the hyperbolic matrix $\gamma = \begin{pmatrix} 7 & -2 \\ 11 & -3 \end{pmatrix}$. The boundary consists of 4 semicircles as well as two incomplete arcs connected by the closed geodesic (thick) associated to γ .

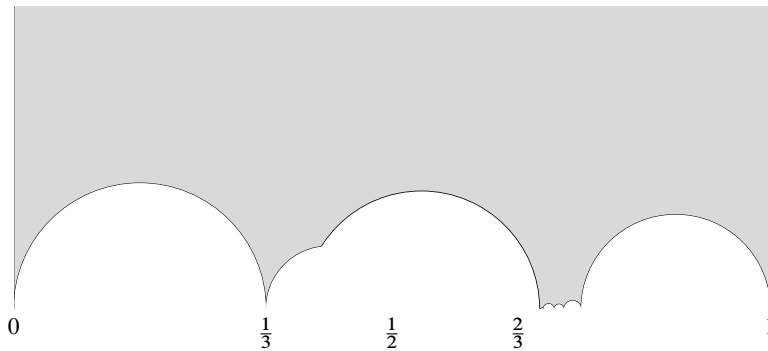


Fig. 3. The fundamental polygon $\mathcal{F}_\gamma(11)$ associated to the hyperbolic matrix $\gamma = \begin{pmatrix} 107 & -41 \\ 154 & -59 \end{pmatrix}$. The boundary consists of 9 semicircles (two of which are very small) as well as two incomplete geodesic segments connected by the closed geodesic (thick) associated to γ .

Figure 4 shows a special fundamental domain $\mathcal{P}(11)$ given by the hyperbolic polygon with vertices $\infty, 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$ with corresponding labels

$$\begin{aligned} \alpha_1 = T^{-1} &= \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, & \alpha_2 &= \begin{pmatrix} -3 & 2 \\ -11 & 7 \end{pmatrix}, & \alpha_3 &= \begin{pmatrix} 4 & -3 \\ 11 & -8 \end{pmatrix}, \\ \alpha_4 = \alpha_2^{-1} &= \begin{pmatrix} 7 & -2 \\ 11 & -3 \end{pmatrix}, & \alpha_5 = \alpha_3^{-1} &= \begin{pmatrix} -8 & 3 \\ -11 & 4 \end{pmatrix}, & \alpha_6 = T &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Figure 2 gives an example of a surface $F_\gamma(11)$ with γ of reduced word length 1 in the free generators $\{T, \alpha_2, \alpha_3\}$ (cf. [19, Figures 2 and 4]). In this case, $\gamma = \alpha_2^{-1} = \begin{pmatrix} 7 & -2 \\ 11 & -3 \end{pmatrix}$ is equal to the matrix γ_Q as in (2.5) associated to the Heegner form $Q = [-11, 10, -2] \in \mathcal{Q}_{12}(11)$. With $r = 10$, Q corresponds to the narrow ideal class $J \in \text{Cl}_{12}^+$.

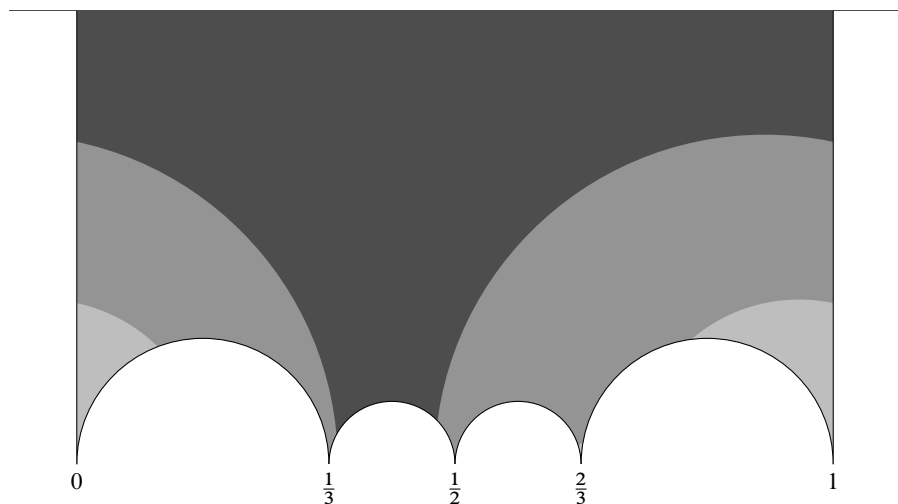


Fig. 4. The projection onto $X_0(11)$ of the fundamental polygon $\mathcal{F}_\gamma(11)$ associated to the hyperbolic matrix $\gamma = \begin{pmatrix} 19 & 10 \\ 55 & 29 \end{pmatrix}$, which corresponds to the narrow ideal class $J \in \text{Cl}_{92}^+$.

Figure 3 gives an example of a surface $\mathcal{F}_\gamma(11)$ with γ of reduced word length 3 in the free generators $\{T, \alpha_2, \alpha_3\}$. In this case, $\gamma = \alpha_3^{-1}\alpha_2^{-1}\alpha_3^{-1} = \begin{pmatrix} 107 & -41 \\ 154 & -59 \end{pmatrix}$. Note that the discriminant $D = 2300 = 5^2 \cdot 92$ of this hyperbolic matrix is not a fundamental discriminant. The boundary consists of 9 semicircles (two of which are very small) as well as two incomplete geodesic segments connected by the closed geodesic associated to γ .

We are interested in the projections of surfaces $\mathcal{F}_\gamma(11)$ to $X_0(11)$, where γ corresponds to an ideal class in a real quadratic number field E such that 11 splits in E . Consider $E = \mathbb{Q}(\sqrt{92})$, which has narrow class number 2 and wide class number 1. Figure 4 shows the projection of $\mathcal{F}_\gamma(11)$ to $\mathcal{P}(11)$, where $\gamma = \alpha_3\alpha_6\alpha_2\alpha_3^{-1}\alpha_6 = \begin{pmatrix} 19 & 10 \\ 55 & 29 \end{pmatrix}$ is equal to the matrix γ_Q as in (2.5) associated to the Heegner form $Q = [-11, -2, 2] \in \mathcal{Q}_{92}(11)$. With $r = -2$, this corresponds to the narrow ideal class $J \in \text{Cl}_{92}^+$ (cf. [19, Figure 6]).

We now show how to use these explicit fundamental domains to extract geometric information about the hyperbolic orbifold $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$.

Proposition 3.7. *Let $\gamma \in \Gamma_0(q)$ be a (non-exceptional) primitive hyperbolic matrix such that the axis S_γ given by (3.3) intersects the fundamental polygon $\mathcal{P}(q)$.*

- (1) *The polygon $\tilde{\mathcal{F}}_\gamma(q)$ is a fundamental domain for $\Gamma_\gamma(q) \backslash \mathbb{H}$.*
- (2) *The subgroup $\Gamma_\gamma(q) \leq \text{PSL}_2(\mathbb{Z})$ is thin.*
- (3) *The orbifold $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$ has genus 0, a unique cusp, and a unique boundary component, which projects to the oriented closed geodesic $\mathcal{C}_\gamma(1) \subset X_0(1)$. Furthermore, the corresponding geodesic segment on the boundary of $\mathcal{F}_\gamma(q)$ projects to $\mathcal{C}_\gamma(q) \subset X_0(q)$.*

(4) We have that

$$\text{vol}(\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)) = \pi |\Omega_\gamma(q)| - \frac{2\pi}{3} e_{3,\gamma}, \quad (3.8)$$

where $e_{3,\gamma}$ denotes the number of conjugacy classes of order 3 subgroups in $\Gamma_\gamma(q)$.

By comparing formula (3.8) to [19, (3.3)], we see that the cardinality of $\Omega_\gamma(q)$ plays the role of the length ℓ_A of the minus continuous fraction in the level 1 case in [19].

Proof of Proposition 3.7. It is a simple consequence of Poincaré's polygon theorem [51] that $\tilde{\mathcal{F}}_\gamma(q)$ is a fundamental domain for $\Gamma_\gamma(q) \backslash \mathbb{H}$ with side pairing elements $\Omega_\gamma(q) \cup \{\gamma, \gamma^{-1}\}$. Since the volume of $\tilde{\mathcal{F}}_\gamma(q)$ is infinite and $\Gamma_\gamma(q)$ contains two hyperbolic elements with different axes (e.g., γ and $\sigma\gamma\sigma^{-1}$ for any $\sigma \in \Omega_\gamma(q)$), we conclude that it is a thin subgroup of $\text{PSL}_2(\mathbb{Z})$.

It follows from the explicit side pairing on $\tilde{\mathcal{F}}_\gamma(q)$ that there is a unique cusp and that the genus is 0. Furthermore, $\Gamma_\gamma(q) \backslash \mathbb{H}$ has exactly one boundary circle, namely corresponding on $\tilde{\mathcal{F}}_\gamma(q)$ to the interval of $\mathbb{P}^1(\mathbb{R})$ between the right-most vertex v of (the complete geodesic containing) $\mathcal{L}_{i(m_\gamma)^*}$ and γv (which is the left-most vertex of the complete geodesic containing $\gamma \mathcal{L}_{i(m_\gamma)^*}$). Notice here that $v \neq \gamma v$ since γ is hyperbolic. Consider the oriented geodesic $\mathcal{C}$ connecting the intersections between the axis $S_\gamma \subset \mathbb{H}$ of γ with $\mathcal{L}_{i(m_\gamma)^*}$ and $\gamma \mathcal{L}_{i(m_\gamma)^*}$, respectively. By construction, this is freely homotopic to the boundary circle of $\Gamma_\gamma(q) \backslash \mathbb{H}$. Since there is a unique closed geodesic with this property, we conclude that (the projection of) $\mathcal{C}$ is exactly the boundary component of $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$. In particular, the boundary component of the fundamental domain $\mathcal{F}_\gamma(q)$ equals the geodesic $\mathcal{C}$ which projects to $\mathcal{C}_\gamma(q) \subset X_0(q)$ by construction. From this, it also follows that the boundary component of $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$ projects to $\mathcal{C}_\gamma(1) \subset X_0(1)$.

As above, denote by $e_{2,\gamma}$ the number of elliptic points of order 2 of $\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)$ and by $e_{3,\gamma}$ the number of elliptic points of order 3. Then clearly we have that

$$|\Omega_\gamma(q)| = e_{2,\gamma} + 2e_{3,\gamma}.$$

Thus the Gauß–Bonnet theorem [19, (2.8)] gives that

$$\text{vol}(\Gamma_\gamma(q) \backslash \mathcal{N}_\gamma(q)) = 2\pi \left(\frac{1}{2} e_{2,\gamma} + \frac{2}{3} e_{3,\gamma} \right) = \pi |\Omega_\gamma(q)| - \frac{2\pi}{3} e_{3,\gamma},$$

as desired. ■

For $\gamma = \gamma_Q$ associated to a Heegner form $Q \in \mathcal{Q}_D(q)$, we write

$$\Gamma_Q(q) := \Gamma_{\gamma_Q}(q), \quad \mathcal{F}_Q(q) := \mathcal{F}_{\gamma_Q}(q), \quad \Omega_{Q,j}(q) := \Omega_{\gamma_Q,j}(q), \quad \Omega_Q(q) := \Omega_{\gamma_Q}(q).$$

Notice that different choices of $\Gamma_0(q)$ -equivalent Heegner forms $Q \in \mathcal{Q}_D(q)$ (all corresponding to the same narrow ideal class $A \in \text{Cl}_D^+$) give rise to subgroups $\Gamma_Q(q)$ of $\text{PSL}_2(\mathbb{Z})$ that are $\Gamma_0(q)$ -conjugates; furthermore, the projection of $\mathcal{F}_Q(q)$ to $X_0(q)$ is independent of this choice. For $A \in \text{Cl}_D^+$, we denote by $\Gamma_A(q) \backslash \mathcal{N}_A(q)$ any such orbifold $\Gamma_Q(q) \backslash \mathcal{N}_Q(q)$ and by $\Omega_A(q)$ and $\mathcal{F}_A(q)$ the corresponding set of matrices and fundamental domain.

As we are interested in the projection of $\mathcal{F}_A(q)$ to $X_0(q)$, we may use this abuse of notation with impunity.

Remark 3.8. Recall that $Q = [a, b, c] \in \mathcal{Q}_D(q)$ is such that $(a, b, c) = 1$, $a \equiv 0 \pmod{q}$, and $b \equiv r \pmod{2q}$, where r is a fixed choice of residue class modulo $2q$ for which $r^2 \equiv D \pmod{4q}$. As q is prime, there are two such possible choices of $r \pmod{2q}$. Thus there are two closed geodesics, and hence two hyperbolic orbifolds, associated to each narrow ideal class $A \in \text{Cl}_D^+$: one to each choice of r . By fixing r , we are thereby fixing a choice of one of these two closed geodesics and one of these two hyperbolic orbifolds associated to A . Since we regard r as being fixed, we suppress from the notation the dependence of $\Gamma_A(q) \backslash \mathcal{N}_A(q)$ and of $\mathcal{F}_A(q)$ on r .

3.2.2. A lower bound for the volume. In this section, we use formula (3.8) to obtain a lower bound for the hyperbolic volume of $\mathcal{F}_A(q)$. In the level 1 setting, this was obtained in [19] by relating the volume to minus continued fractions and using some explicit relations to the regulator $\log \epsilon_D$, while in [55, Section 5.1], such a lower bound was obtained using geometric considerations relying on equidistribution of the closed geodesics. In the level q setting, on the other hand, such routes do not seem to be readily available. We instead use a group-theoretic approach.

For $\gamma \in \Gamma_0(q)$, let $\text{wl}(\gamma)$ denote the word length of γ in the independent generators defined from $\mathcal{P}(q)$, namely the labels $\alpha_1, \dots, \alpha_N$ (forgetting inverses). Given a Heegner form $Q \in \mathcal{Q}_D(q)$ with associated hyperbolic matrix γ_Q , we first observe that the entries of γ_Q are bounded by $q^{O(\text{wl}(\gamma_Q))}$ since the Frobenius norms of the labels α_j are $O(q)$. In particular, we have that

$$\text{Tr}(\gamma_Q) \ll q^{O(\text{wl}(\gamma_Q))},$$

which implies that

$$\text{wl}(\gamma_Q) \gg \frac{\log \text{Tr}(\gamma_Q)}{\log q} \gg \frac{\log \epsilon_D}{\log q}.$$

This is close to what we want. Alas, a lower bound for the word length does not quite give a lower bound for $|\Omega_Q(q)|$ in general. For a *generic* hyperbolic matrix $\gamma \in \Gamma_0(q)$ of word length tending to infinity, it is not hard to see that in fact $|\Omega_\gamma(q)| \asymp q \text{wl}(\gamma)$. Thus one might hope that for subgroups $H \subset \text{Cl}_D^+$ of sufficiently small index (say, $[\text{Cl}_D^+ : H] \ll D^\delta$ for some small $\delta > 0$), one has that

$$\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q)) \gg q|H| \log \epsilon_D$$

for any coset $CH \subset \text{Cl}_D^+$ (cf. [19, Proposition 1]). We make progress towards by showing the following.

Proposition 3.9. *For $A \in \text{Cl}_D^+$, we have that*

$$\text{vol}(\mathcal{F}_A(q)) \gg \frac{\log \epsilon_D}{\log qD}. \quad (3.9)$$

Proof. Let $Q \in \mathcal{Q}_D(q)$ be a Heegner form corresponding to the narrow ideal class $A \in \text{Cl}_D^+$ with corresponding hyperbolic matrix $\gamma_Q \in \Gamma_0(q)$ such that its axis $S_Q \subset \mathbb{H}$ (i.e., the infinite geodesic fixed by γ_Q) intersects $\mathcal{P}(q)$. We start by bounding the word length $\text{wl}(\gamma_Q)$ in terms of D . Recall from [46] that the special fundamental polygon $\mathcal{P}(q)$ is tiled by

$$\left\{ z \in \mathbb{H} : |z| \geq 1, 0 \leq \text{Re}(z) < \frac{1}{2} \right\}. \quad (3.10)$$

The word length $\text{wl}(\gamma_Q)$ of γ_Q with respect to the independent generators of $\Gamma_0(q)$ defined from $\mathcal{P}(q)$ (forgetting inverses) is precisely the number of intersections between the geodesic $\mathcal{C}_A(q)$ from z_Q to $\gamma_Q z_Q$ and the $\Gamma_0(q)$ -translates of $\mathcal{P}(q)$ and thus is bounded by the number of intersections with (3.10). We shall now relate this intersection number to continued fractions.

Consider the tessellation of $\mathbb{H}$ by the *Farey triangle* Δ_{Far} with vertices $\{0, 1, \infty\}$. Note that Δ_{Far} is tiled by six copies of triangle (3.10). As explained in [60, Theorem A and Corollary 3.3.5], the *left-right cutting sequence* of the axis S_Q of γ_Q with respect to this tessellation is periodic with period $a_1 + \cdots + a_{\ell'}$, where $[b_0, \dots, b_r, \overline{a_1, \dots, a_{\ell'}}]$ is the continued fraction expansion of the attracting endpoint of S_Q , which is eventually periodic with period ℓ' . Consider the fundamental domain $\mathcal{F}_0$ for $\text{PSL}_2(\mathbb{Z})$ with vertices $\{0, \frac{1+\sqrt{-3}}{2}, \infty\}$ and side pairing elements $\{S, TS, ST^{-1}\}$. Note that Δ_{Far} is tiled by three copies of $\mathcal{F}_0$ such that all the sides of Δ_{Far} correspond to the side of $\mathcal{F}_0$ containing 0 and ∞ with label S . Furthermore, a “left” (resp. “right”) in the cutting sequence corresponds, in terms of $\mathcal{F}_0$, to the label S followed by either the label TS or twice its inverse (resp. ST^{-1} or twice its inverse). Since $(ST^{-1})^2 = TS$, we conclude that the cutting sequence uniquely determines the representation of $\gamma_Q \in \text{PSL}_2(\mathbb{Z})$ in terms of the independent generators S, TS of $\text{PSL}_2(\mathbb{Z})$ of order two and three, respectively. In particular, the minimal period of the cutting sequence of S_Q is at most a third of the length of the Morse code of γ_Q with respect to $\mathcal{F}_0$. Since $\mathcal{F}_0$ is tiled by two copies of triangle (3.10), we conclude that

$$\text{wl}(\gamma_Q) \leq 6(a_1 + \cdots + a_{\ell'}) \ll D^{\frac{3}{2}} \quad (3.11)$$

using the elementary bounds $a_i \ll \sqrt{D}$ and $\ell' \ll D$, which follow from the classical algorithm for continued fractions as described in [34].

Let $0 < j(1) < \cdots < j(\ell) < m_\gamma + 1$ be the nonzero indices in the Morse code (3.4) of γ_Q for which $\Omega_{Q,j(k)}(q) = \emptyset$ with $\Omega_{Q,j}(q) = \Omega_{\gamma_Q,j}(q)$ as defined in (3.5) and (3.6). Put $j(0) = 0$ and $j(\ell + 1) = m_\gamma + 1$. We now argue that for all $0 \leq k \leq \ell$, the product

$$\beta_k := \alpha_{i(j(k))} \cdots \alpha_{i(j(k+1)-1)}$$

can be written as $\alpha_h \alpha_{t(h)} \cdots \alpha_{t(m)(h)}$ for some $h \in \mathbb{Z}/N\mathbb{Z}$ and $m \geq 0$. Assume first that $e_3 = 0$. Then $\Omega_{Q,j}(q) = \emptyset$ means exactly that $i(j) = i(j-1)^* - 1 = t(i(j-1))$ and the claim follows. In the presence of 3-torsion, $\Omega_{Q,j}(q) = \emptyset$ implies that

$$\text{arc}(i(j), i(j-1)^*) \subset \{i(j-1), i(j)^*\}.$$

We need to consider the four possible geometric configurations corresponding to the nonempty subsets of $\{i(j-1), i(j)^*\}$. If $\text{arc}(i(j), i(j-1)^*) = \emptyset$, we get as above that $i(j) = t(i(j-1))$. Next, assume that $\text{arc}(i(j), i(j-1)^*) = \{i(j-1)\}$. Then since $\Omega_{Q,j}(q) = \emptyset$, we conclude that $\alpha_{i(j-1)}^3 = 1$, $i(j-1) < i(j-1)^*$ and $i(j) = i(j-1) - 1$. Furthermore, since $j \notin \mathcal{J}_3$, we must that $i(j-2) = i(j-1)$, implying that $i(j-1) = t(i(j-2))$. Since we have $\alpha_{i(j-2)}\alpha_{i(j-1)} = \alpha_{i(j-1)^*}$, $i(j-2)^* \in \text{arc}(i(j-2), i(j-3)^*)$, and $i(j) = (i(j-1)^*)^* - 1 = t(i(j-1)^*)$, we see that we can alter the Morse code to get the desired shape. The two final cases can be treated similarly, which yields the claim.

We conclude by Lemma 3.4 that we may write

$$\gamma_Q = \beta_0 \beta_1 \cdots \beta_\ell,$$

where for all $0 \leq k \leq \ell$ either

- (1) $\beta_k = \alpha_j$ is a label of $\mathcal{P}(q)$, in which case the Frobenius norm is bounded by $O(q)$;
- (2) $\beta_k = T^n$ for some $n \in \mathbb{Z}$, in which case the Frobenius norm is bounded by $O(|n|)$;
or
- (3) we have that

$$\beta_k = \gamma_{j_1}^{-1} \begin{pmatrix} 1 & 0 \\ q & 1 \end{pmatrix}^n \gamma_{j_2},$$

for some $j_1, j_2 \in \{1, \dots, 4g + e_2 + e_3 + 1\}$ (with γ_{j_i} defined as in (3.1)) and $n \in \mathbb{Z}$, in which case the Frobenius norm is bounded by $O(q^{\frac{3}{2}} \cdot q|n| \cdot q^{\frac{3}{2}})$ via Lemma 3.5.

Now clearly the power n of the parabolic elements appearing in β_k satisfies

$$|n| \leq \text{wl}(\beta_k) \leq \text{wl}(\gamma_Q) \ll D^{\frac{3}{2}},$$

using bound (3.11). Thus the entries of γ_Q are bounded by

$$q^{O(\ell)} D^{O(\ell)}.$$

By computing the trace, we see that

$$\ell \gg \frac{\log \epsilon_D}{\log qD}.$$

Since trivially $|\Omega_Q(q)| \geq \ell$ and $e_{3,\gamma_Q} \leq \frac{|\Omega_Q(q)|}{2}$, Proposition 3.7 gives the desired result. ■

Remark 3.10. If we consider an *arbitrary* special polygon $\mathcal{P}(q)$ instead of the one constructed based on Theorem 3.2, then in place of the bound $b_i \ll \sqrt{q}$ for the denominators in the Farey sequence of level q , we would only have that $b_i \ll e^{O(q)}$. The above argument yields in this case the weaker lower bound

$$\text{vol}(\Gamma_Q(q) \backslash \mathcal{N}_Q(q)) \gg \frac{\log \epsilon_D}{q + \log D}.$$

Remark 3.11. To prove a polynomial bound for $\text{wl}(\gamma_Q)$, one might instead use the bound $\text{wl}(\gamma_Q) \leq m_A + \ell_A$, where m_A is as in [19, (2.1)]. Via the identity $m_A = 2\ell_A + \ell_{AJ}$ [20, (46) and (49)], this reduces the problem to bounding ℓ_A in terms of D . It is claimed in [19, p. 968] that one can apply a general argument of Eichler [21] to show that $\ell_A \ll \log \epsilon_D$ uniformly for $A \in \text{Cl}_D^+$, which would yield the desired result. Unfortunately, however, Eichler's argument does not apply directly: Eichler's result gives a method for determining upper bounds for the word length of the *geometric code* of a closed geodesic (in the sense of [43]), whereas ℓ_A is the word length of the *arithmetic code* of a closed geodesic (in the sense of [43]), and in general these need not coincide [43, Theorem 1].

4. Adèlisation of Maaß cusp forms

We review some standard notions about Maaß cusp forms of weight k , level q , and principal nebentypus, with an emphasis on forms of weight 0 and the action of raising and lowering operators on such forms. We then describe the relation between such classical automorphic forms and adèlic automorphic forms, highlighting the correspondence between Whittaker functions of representations of $\text{GL}_2(\mathbb{R})$ and $\text{GL}_2(\mathbb{Q}_p)$, the Whittaker expansion of an adèlic automorphic form, and the Fourier expansion of a classical automorphic form. This explicit correspondence is invaluable in Section 6, where we prove an identity between integrals of Maaß forms over hyperbolic orbifolds and period integrals of adèlic automorphic forms. Useful references for this material include [18, Section 4], [27, Chapters 3 and 4], [58, 59].

4.1. Maaß cusp forms

Let k be an integer, q be a positive integer, and denote by $\mathcal{C}_k(\Gamma_0(q))$ the vector subspace of $L^2(\Gamma_0(q) \backslash \mathbb{H})$ spanned by Maaß cusp forms of weight k , level q , and principal nebentypus, in the sense of [18, Section 4]. Such a Maaß cusp form is a real analytic function $f: \mathbb{H} \rightarrow \mathbb{C}$ for which

- f is an eigenfunction of the weight k Laplacian

$$\Delta_k = -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + iky \frac{\partial}{\partial x},$$

so that $\Delta_k f(z) = \lambda_f f(z)$ for some $\lambda_f \in \mathbb{C}$ (and necessarily $\lambda_f \in [\frac{1}{4} - (\frac{7}{64})^2, \infty)$),

- f is automorphic, so that

$$j_\gamma(z)^{-k} f(\gamma z) = f(z)$$

for all $z \in \mathbb{H}$ and $\gamma \in \Gamma_0(q)$, where for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2^+(\mathbb{R})$, the space of 2×2 matrices with real entries and positive determinant,

$$gz := \frac{az + b}{cz + d}, \quad j_g(z) := \frac{cz + d}{|cz + d|}, \quad (4.1)$$

- f is of moderate growth, and
- f is cuspidal, so that for each cusp $\mathfrak{b}$ of $\Gamma_0(q)\backslash\mathbb{H}$,

$$\int_0^1 j_{\sigma_{\mathfrak{b}}}(z)^{-k} f(\sigma_{\mathfrak{b}}z) dx = 0$$

for all $y > 0$, where $\sigma_{\mathfrak{b}} \in \mathrm{SL}_2(\mathbb{R})$ is a scaling matrix for $\mathfrak{b}$.

4.1.1. The Fourier expansion and Hecke eigenvalues of a Maaß cusp form. The Fourier expansion at the cusp at infinity of a weight 0 Maaß cusp form $f \in \mathcal{C}_0(\Gamma_0(q))$ is

$$f(z) = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \rho_f(n) W_{0, it_f}(4\pi|n|y) e(nx), \quad (4.2)$$

where $W_{\alpha, \beta}$ denotes the classical Whittaker function and $t_f \in \mathbb{R} \cup i[-\frac{7}{64}, \frac{7}{64}]$ is the spectral parameter of f , so that $\lambda_f = \frac{1}{4} + t_f^2$. If f is additionally a Hecke–Maaß newform, namely an eigenfunction of the n -th Hecke operator T_n for all $n \in \mathbb{N}$ as well as the reflection operator $X: \mathcal{C}_0(\Gamma_0(q)) \rightarrow \mathcal{C}_0(\Gamma_0(q))$ given by $(Xf)(z) := f(-\bar{z})$, the Fourier coefficients $\rho_f(n)$ and Hecke eigenvalues $\lambda_f(n)$ of f satisfy

- $\rho_f(1)\lambda_f(n) = \sqrt{n}\rho_f(n)$ for $n \in \mathbb{N}$,
- $\rho_f(n) = \epsilon_f \rho_f(-n)$ for $n \in \mathbb{Z}$, where $\epsilon_f \in \{1, -1\}$ is the parity of f , so that $Xf = \epsilon_f f$,
- for all $m, n \in \mathbb{N}$, the Hecke eigenvalues satisfy the multiplicativity relations

$$\lambda_f(m)\lambda_f(n) = \sum_{\substack{d|(m,n) \\ (d,q)=1}} \lambda_f\left(\frac{mn}{d^2}\right), \quad \lambda_f(mn) = \sum_{\substack{d|(m,n) \\ (d,q)=1}} \mu(d)\lambda_f\left(\frac{m}{d}\right)\lambda_f\left(\frac{n}{d}\right),$$

- for each $p \nmid q$, there exists $\alpha_f(q) \in \mathbb{C}$ satisfying $p^{-\frac{7}{64}} \leq |\alpha_f(q)| \leq p^{\frac{7}{64}}$ such that for all $r \geq 1$,

$$\lambda_f(p^r) = \sum_{m=0}^r \alpha_f(q)^m \alpha_f^{-1}(q)^{r-m}, \quad (4.3)$$

- for each $p \parallel q$, there exists $\alpha_f(q) \in \{1, -1\}$ such that for all $r \geq 1$,

$$\lambda_f(p^r) = \frac{\alpha_f(q)^r}{p^{\frac{r}{2}}}, \quad (4.4)$$

- for each prime p for which $p^2 \mid q$, we have that $\lambda_f(p^r) = 0$ for all $r \geq 1$.

Furthermore, if $f \in \mathcal{C}_0(\Gamma_0(q))$ is a Hecke–Maaß newform and q is squarefree, then the L^2 -norm of f and the first Fourier coefficient $\rho_f(1)$ of f satisfy the relation [38, Lemma 4.6]

$$\int_{\Gamma_0(q)\backslash\mathbb{H}} |f(z)|^2 d\mu(z) = \frac{2q|\rho_f(1)|^2 L(1, \mathrm{ad} f)}{\cosh \pi t_f}. \quad (4.5)$$

Here we note that the measure $d\mu(z) = y^{-2} dx dy$ is such that

$$\mathrm{vol}(\Gamma_0(q)\backslash\mathbb{H}) = \frac{\pi}{3} v(q), \quad v(q) := [\Gamma : \Gamma_0(q)] = q \prod_{p|q} \left(1 + \frac{1}{p}\right). \quad (4.6)$$

4.1.2. *Raising and lowering operators.* The weight k raising operator

$$R_k := \frac{k}{2} + (z - \bar{z}) \frac{\partial}{\partial z} = \frac{k}{2} + iy \left(\frac{\partial}{\partial x} - i \frac{\partial}{\partial y} \right)$$

acts on $\mathcal{C}_k(\Gamma_0(q))$ and raises the weight by 2; that is, its image lies in $\mathcal{C}_{k+2}(\Gamma_0(q))$. Similarly, the weight k lowering operator

$$L_k := -\frac{k}{2} - (z - \bar{z}) \frac{\partial}{\partial \bar{z}} = -\frac{k}{2} - iy \left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right)$$

maps $\mathcal{C}_k(\Gamma_0(q))$ to $\mathcal{C}_{k-2}(\Gamma_0(q))$. From [27, Proposition 3.9.13], if $f \in \mathcal{C}_0(\Gamma_0(q))$ is a Hecke–Maaß newform of weight 0 and level q with Fourier expansion (4.2), then the Fourier expansion of the weight k Maaß cusp form $F_k \in \mathcal{C}_k(\Gamma_0(q))$ defined by

$$F_k := \begin{cases} R_{k-2} \cdots R_0 f & \text{if } k \in 2\mathbb{N}, \\ f & \text{if } k = 0, \\ L_{k+2} \cdots L_0 f & \text{if } k \in -2\mathbb{N}, \end{cases} \quad (4.7)$$

is given by

$$\begin{aligned} F_k(z) &= \frac{\Gamma(\frac{k+1}{2} + it_f) \Gamma(\frac{k+1}{2} - it_f)}{\Gamma(\frac{1}{2} + it_f) \Gamma(\frac{1}{2} - it_f)} \epsilon_f \rho_f(1) \sum_{n=-\infty}^{-1} \frac{\lambda_f(|n|)}{\sqrt{|n|}} W_{-\frac{k}{2}, it_f}(4\pi|n|y) e(nx) \\ &\quad + (-1)^{\frac{k}{2}} \rho_f(1) \sum_{n=1}^{\infty} \frac{\lambda_f(n)}{\sqrt{n}} W_{\frac{k}{2}, it_f}(4\pi ny) e(nx) \end{aligned} \quad (4.8)$$

if $k \in 2\mathbb{N} \cup \{0\}$, while if $k \in -2\mathbb{N}$, the Fourier expansion is

$$\begin{aligned} F_k(z) &= (-1)^{\frac{k}{2}} \epsilon_f \rho_f(1) \sum_{n=-\infty}^{-1} \frac{\lambda_f(|n|)}{\sqrt{|n|}} W_{-\frac{k}{2}, it_f}(4\pi|n|y) e(nx) \\ &\quad + \frac{\Gamma(\frac{1-k}{2} + it_f) \Gamma(\frac{1-k}{2} - it_f)}{\Gamma(\frac{1}{2} + it_f) \Gamma(\frac{1}{2} - it_f)} \rho_f(1) \sum_{n=1}^{\infty} \frac{\lambda_f(n)}{\sqrt{n}} W_{\frac{k}{2}, it_f}(4\pi ny) e(nx). \end{aligned} \quad (4.9)$$

4.1.3. *Atkin–Lehner operators.* Let q be squarefree. For each divisor q_1 of q , such that $q = q_1 q_2$, we define an Atkin–Lehner operator $W_{q_1} \in \mathrm{SL}_2(\mathbb{R})$ as in (2.7), namely

$$W_{q_1} := \begin{pmatrix} a\sqrt{q_1} & \frac{b}{\sqrt{q_1}} \\ cq_2\sqrt{q_1} & d\sqrt{q_1} \end{pmatrix},$$

where $a, b, c, d \in \mathbb{Z}$ are such that $adq_1 - bcq_2 = 1$. If $f \in \mathcal{C}_k(\Gamma_0(q))$ is a Hecke–Maaß newform, then it is an *eigenfunction* of each Atkin–Lehner operator, in the sense that there exists some constant $\eta_f(q_1) \in \{1, -1\}$, dependent on f and q_1 but not on a, b, c, d , such that

$$j_{W_{q_1}}(z)^{-k} f(W_{q_1} z) = \eta_f(q_1) f(z)$$

for all $z \in \mathbb{H}$, where $j_g(z)$ is as in (4.1).

4.2. Eisenstein series

We recall that the Eisenstein series $E(z, s)$ for $\Gamma \backslash \mathbb{H}$ is given for $\operatorname{Re}(s) > 1$ by the absolutely convergent series

$$E(z, s) := \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \operatorname{Im}(\gamma z)^s,$$

where $\Gamma_\infty := \{\pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{Z}\}$ is the stabiliser of the cusp at infinity. The Eisenstein series extends meromorphically to $\mathbb{C}$ with a simple pole at $s = 1$ with residue $\frac{1}{\operatorname{vol}(\Gamma \backslash \mathbb{H})}$, independently of z . It is an eigenfunction of the weight 0 Laplacian with eigenvalue $s(1-s)$ and is automorphic and of moderate growth, but it is *not* cuspidal.

For $s = \frac{1}{2} + it$, the Laplacian eigenvalue of $E(z, \frac{1}{2} + it)$ is $\frac{1}{4} + t^2$, the parity is $\epsilon = 1$, while the Fourier coefficients $\rho(n, t)$ and Hecke eigenvalues $\lambda(n, t)$ satisfy

- $\rho(1, t)\lambda(n, t) = \sqrt{n}\rho(n, t)$ for $n \in \mathbb{N}$,
- $\rho(n, t) = \rho(-n, t)$ for $n \in \mathbb{Z}$,
- for all $m, n \in \mathbb{N}$, the Hecke eigenvalues satisfy the multiplicativity relations

$$\lambda(m, t)\lambda(n, t) = \sum_{d|(m, n)} \lambda\left(\frac{mn}{d^2}, t\right), \quad \lambda(mn, t) = \sum_{d|(m, n)} \mu(d)\lambda\left(\frac{m}{d}, t\right)\lambda\left(\frac{n}{d}, t\right),$$

- for all $n \in \mathbb{N}$, the Hecke eigenvalues are given explicitly by

$$\lambda(n, t) = \sum_{ab=n} a^{it} b^{-it},$$

- the first Fourier coefficient is given explicitly by

$$\rho(1, t) = \frac{1}{\xi(1 + 2it)},$$

where $\xi(s) := \pi^{-\frac{s}{2}} \Gamma(\frac{s}{2}) \zeta(s)$.

The weight 0 raising and lowering operators act on $E(\cdot, \frac{1}{2} + it)$ and raise and lower the weight by 2. The Fourier expansion of $E(z, \frac{1}{2} + it)$ reads

$$\begin{aligned} E\left(z, \frac{1}{2} + it\right) &= y^{\frac{1}{2}+it} + \frac{\xi(1-2it)}{\xi(1+2it)} y^{\frac{1}{2}-it} + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \rho(n, t) W_{0, it_f}(4\pi|n|y) e(nx) \\ &= y^{\frac{1}{2}+it} + \frac{\xi(1-2it)}{\xi(1+2it)} y^{\frac{1}{2}-it} \\ &\quad + \frac{1}{\xi(1+2it)} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{\lambda(|n|, t)}{\sqrt{|n|}} W_{0, it_f}(4\pi|n|y) e(nx), \end{aligned} \quad (4.10)$$

while the Fourier expansion of the weight 2 Eisenstein series $(R_0 E)(z, \frac{1}{2} + it)$ is given by

$$(R_0 E)\left(z, \frac{1}{2} + it\right) = \left(\frac{1}{2} + it\right) y^{\frac{1}{2}+it} + \left(\frac{1}{2} - it\right) \frac{\xi(1-2it)}{\xi(1+2it)} y^{\frac{1}{2}-it}$$

$$\begin{aligned}
& + \frac{(\frac{1}{4} + t^2)}{\xi(1 + 2it)} \sum_{n=-\infty}^{-1} \frac{\lambda(|n|, t)}{\sqrt{|n|}} W_{-1, it}(4\pi|n|y)e(nx) \\
& - \frac{1}{\xi(1 + 2it)} \sum_{n=1}^{\infty} \frac{\lambda(n, t)}{\sqrt{n}} W_{1, it_f}(4\pi ny)e(nx). \quad (4.11)
\end{aligned}$$

We additionally define

$$\begin{aligned}
(\widetilde{R_0 E})\left(z, \frac{1}{2} + it\right) &:= (R_0 E)\left(z, \frac{1}{2} + it\right) - \left(\frac{1}{2} + it\right) \operatorname{Im}(z)^{\frac{1}{2} + it} \\
&\quad - \left(\frac{1}{2} - it\right) \frac{\xi(1 - 2it)}{\xi(1 + 2it)} \operatorname{Im}(z)^{\frac{1}{2} - it}. \quad (4.12)
\end{aligned}$$

This is square-integrable but no longer automorphic.

Finally, we record here the following useful result concerning the behaviour of an Eisenstein series at a cusp.

Lemma 4.1. *Let q be squarefree. For a cusp $\mathfrak{b}$ of $\Gamma_0(q) \backslash \mathbb{H}$ with scaling matrix $\sigma_{\mathfrak{b}} \in \Gamma_0(q)$ and for $\ell \mid q$, we have that*

$$\left. \frac{\partial}{\partial z} \right|_{z=x+iY} E\left(\sigma_{\mathfrak{b}}(\ell z), \frac{1}{2} + it\right) \ll_{q,t} \frac{1}{\sqrt{Y}}.$$

Proof. We begin by noting that $E(\ell z, \frac{1}{2} + it)$ may be written as a finite linear combination of Eisenstein series $E_{\alpha}(z, \frac{1}{2} + it)$ associated to cusps α of $\Gamma_0(q) \backslash \mathbb{H}$ [65, Theorem 7.1]. From [40, (6.18)], the Fourier expansion of $E_{\alpha}(\sigma_{\mathfrak{b}} z, \frac{1}{2} + it)$ is of the form

$$\begin{aligned}
E_{\alpha}\left(\sigma_{\mathfrak{b}} z, \frac{1}{2} + it\right) &= \delta_{\alpha, \mathfrak{b}} y^{\frac{1}{2} + it} + \varphi_{\alpha, \mathfrak{b}}\left(\frac{1}{2} + it\right) y^{\frac{1}{2} - it} \\
&\quad + \sum_{\substack{n=-\infty \\ n \neq 0}} \rho_{\alpha, \mathfrak{b}}(n, t) W_{0, it}(4\pi|n|y)e(nx), \quad (4.13)
\end{aligned}$$

where $\varphi_{\alpha, \mathfrak{b}}(s)$ is an entry of the scattering matrix, while the Fourier coefficients $\rho_{\alpha, \mathfrak{b}}(n, t)$ grow at most polynomially in $|n|$. The desired result now follows by differentiating term by term. ■

4.3. Adèlic automorphic forms

4.3.1. The adèlic lift of a Maaß cusp form. Following [27, Sections 4.11 and 4.12], we describe the adèlic lift of a Maaß cusp form $f \in \mathcal{C}_k(\Gamma_0(q))$. We first lift $f \in \mathcal{C}_k(\Gamma_0(q))$ to a function $\tilde{f}: \mathrm{GL}_2^+(\mathbb{R}) \rightarrow \mathbb{C}$ defined via

$$\tilde{f}(g) := j_g(i)^{-k} f(gi). \quad (4.14)$$

For all $g \in \mathrm{GL}_2^+(\mathbb{R})$ and $\theta \in [0, 2\pi)$, this satisfies

$$\tilde{f}\left(g \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}\right) = e^{ik\theta} \tilde{f}(g).$$

Next, we lift $\tilde{f}$ to an adèlic automorphic form $\phi = \phi_f$ on $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$, where $\mathbb{A}_{\mathbb{Q}}$ denotes the ring of adèles of $\mathbb{Q}$. To describe this lift, we first let $K_0(q) \ni k = (k_{\infty}, k_2, k_3, k_5, \dots)$ denote the congruence subgroup of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ of the form

$$K_0(q) := \left\{ k \in \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}) : k_{\infty} = 1_2, k_p = \begin{pmatrix} a_p & b_p \\ c_p & d_p \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}_p) \right. \\ \left. \text{with } c_p \in p^r \mathbb{Z}_p \text{ if } p^r \parallel q \right\}. \quad (4.15)$$

Here 1_2 denotes the 2×2 identity matrix. We view $\mathrm{GL}_2^+(\mathbb{R}) \ni g_{\infty}$ as a subgroup of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ via the embedding $g_{\infty} \mapsto (g_{\infty}, 1_2, 1_2, \dots)$. Finally, we view $\mathrm{GL}_2(\mathbb{Q}) \ni \gamma$ as a subgroup of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ via the diagonal embedding $\gamma \mapsto (\gamma, \gamma, \gamma, \dots)$. Then via the strong approximation theorem,

$$\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}) = \mathrm{GL}_2(\mathbb{Q}) \mathrm{GL}_2^+(\mathbb{R}) K_0(q), \quad (4.16)$$

so that every $g \in \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ can be written (non-uniquely) as $g = \gamma g_{\infty} k$ for some $\gamma \in \mathrm{GL}_2(\mathbb{Q})$, $g_{\infty} \in \mathrm{GL}_2^+(\mathbb{R})$, and $k \in K_0(q)$. The adèlic lift $\phi = \phi_f$ of a Maaß cusp form $f \in \mathcal{C}_k(\Gamma_0(q))$ is then given by

$$\phi(g) = \phi(\gamma g_{\infty} k) := \tilde{f}(g_{\infty}). \quad (4.17)$$

This is well defined even though the decomposition $g = \gamma g_{\infty} k$ is not unique. In particular,

$$\phi(g) = f(x + iy) \quad (4.18)$$

for $g = g_{\infty} = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R}) \subset \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ with $y > 0$ and $x \in \mathbb{R}$.

4.3.2. The Whittaker expansion of an adèlic automorphic form. Let $\psi: \mathbb{Q} \backslash \mathbb{A}_{\mathbb{Q}} \rightarrow \mathbb{C}$ be the standard adèlic additive character defined as in [27, Definition 1.7.1], such that $\psi(u) = \psi_{\infty}(u_{\infty}) \prod_p \psi_p(u_p)$ for $u = (u_{\infty}, u_2, u_3, \dots) \in \mathbb{A}_{\mathbb{Q}}$ with $\psi_{\infty}: \mathbb{R} \rightarrow \mathbb{C}$ the additive character $\psi_{\infty}(u_{\infty}) := e(u_{\infty}) := e^{2\pi i u_{\infty}}$ and $\psi_p: \mathbb{Q}_p \rightarrow \mathbb{C}$ the standard unramified additive character defined in [27, Definition 1.6.3]. The Whittaker function $W_{\phi}: \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}}) \rightarrow \mathbb{C}$ of a cuspidal adèlic automorphic form ϕ is

$$W_{\phi}(g) := \int_{\mathbb{Q} \backslash \mathbb{A}_{\mathbb{Q}}} \phi \left(\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} g \right) \bar{\psi}(u) du,$$

which satisfies

$$W_{\phi} \left(\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} g \right) = \psi(u) W_{\phi}(g) \quad (4.19)$$

for all $u \in \mathbb{A}_{\mathbb{Q}}$ and $g \in \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$. The automorphic form ϕ has the Whittaker expansion

$$\phi(g) = \sum_{\alpha \in \mathbb{Q}^{\times}} W_{\phi} \left(\begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} g \right). \quad (4.20)$$

4.3.3. The Whittaker expansion of an adèlic lift. Let $\phi = \phi_{F_k}$ be the adèlic lift of a Maaß cusp form $F_k \in \mathcal{C}_k(\Gamma_0(q))$ of weight k associated via raising or lowering operators to a Hecke–Maaß newform $f \in \mathcal{C}_0(\Gamma_0(q))$ of weight 0 as in (4.7). Then ϕ is a pure tensor lying in the vector space of a cuspidal automorphic representation $\pi = \pi_f = \pi_\infty \otimes \bigotimes_p \pi_p$ of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$, where each π_p is a generic irreducible admissible unitary representation of $\mathrm{GL}_2(\mathbb{Q}_p)$ and π_∞ is a generic irreducible unitary Casselman–Wallach representation of $\mathrm{GL}_2(\mathbb{R})$. In particular, for $g = (g_\infty, g_2, g_3, \dots) \in \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$, we have the factorisation

$$W_\phi(g) = c_\phi W_\infty(g_\infty) \prod_p W_p(g_p), \quad (4.21)$$

where c_ϕ is a constant independent of g , each W_p is a Whittaker function in the Whittaker model $\mathcal{W}(\pi_p, \psi_p)$ of π_p , and similarly $W_\infty \in \mathcal{W}(\pi_\infty, \psi_\infty)$.

We show in Lemma 4.2 that the Whittaker functions W_p are such that for $\alpha \in \mathbb{Q}^\times$,

$$\prod_p W_p \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} = \begin{cases} \frac{\lambda_f(|n|)}{\sqrt{|n|}} & \text{if } \alpha = n \in \mathbb{Z} \setminus \{0\}, \\ 0 & \text{otherwise.} \end{cases}$$

In Lemma 4.3, we show that for $n \in \mathbb{Z} \setminus \{0\}$ and $y > 0$,

$$W_\infty \begin{pmatrix} ny & 0 \\ 0 & 1 \end{pmatrix} = \begin{cases} (-1)^{\frac{k}{2}} W_{\frac{k}{2}, it}(4\pi ny), & n \in \mathbb{N}, k \in 2\mathbb{N} \cup \{0\}, \\ \epsilon_f \frac{\Gamma(\frac{k+1}{2} + it) \Gamma(\frac{k+1}{2} - it)}{\Gamma(\frac{1}{2} + it) \Gamma(\frac{1}{2} - it)} W_{-\frac{k}{2}, it}(4\pi |n|y), & n \in -\mathbb{N}, k \in 2\mathbb{N} \cup \{0\}, \\ \frac{\Gamma(\frac{1-k}{2} + it) \Gamma(\frac{1-k}{2} - it)}{\Gamma(\frac{1}{2} + it) \Gamma(\frac{1}{2} - it)} W_{\frac{k}{2}, it}(4\pi ny), & n \in \mathbb{N}, k \in -2\mathbb{N}, \\ \epsilon_f (-1)^{\frac{k}{2}} W_{-\frac{k}{2}, it}(4\pi |n|y), & n \in -\mathbb{N}, k \in -2\mathbb{N}. \end{cases}$$

From this, we see that the constant c_ϕ in (4.21) is equal to $\rho_f(1)$ by taking $g = g_\infty = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R}) \subset \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ with $y > 0$ and $x \in \mathbb{R}$ in (4.20), so that by (4.18) and (4.19),

$$F_k(x + iy) = \phi(g) = c_\phi \sum_{\alpha \in \mathbb{Q}^\times} W_\infty \begin{pmatrix} \alpha y & 0 \\ 0 & 1 \end{pmatrix} \prod_p W_p \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} e(\alpha x),$$

and comparing this adèlic Whittaker expansion to the classical Fourier expansion at the cusp at infinity (4.8) and (4.9).

4.3.4. Nonarchimedean Whittaker functions. Let $\phi = \phi_{F_k}$ be the adèlic lift of a Maaß cusp form $F_k \in \mathcal{C}_k(\Gamma_0(q))$ associated to a Hecke–Maaß newform $f \in \mathcal{C}_0(\Gamma_0(q))$ as in definition (4.7) with q squarefree. For each prime p , the local Whittaker functions $W_p \in \mathcal{W}(\pi_p, \psi_p)$ are of a distinguished form. One can explicitly describe the values of the Whittaker function $W_p(g_p)$ for $g_p = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$ with $a \in \mathbb{Q}_p^\times$; see [59, Section 2.4].

Lemma 4.2. (1) For $p \nmid q$, the representation π_p is a spherical principal series representation $\omega_p \boxplus \omega_p^{-1}$, where ω_p is an unramified character of $\mathbb{Q}_p^\times$ satisfying $p^{-\frac{7}{64}} \leq |\omega_p(q)|_p \leq p^{\frac{7}{64}}$ and $|\cdot|_p$ denotes the p -adic absolute value normalised such that $|p|_p = p^{-1}$. This character is such that $\omega_p(q)$ is equal to $\alpha_f(q)$ as in (4.3). For $a \in \mathbb{Q}_p^\times$, let $v(a) \in \mathbb{Z}$ be such that $|a|_p = p^{-v(a)}$. There is a distinguished Whittaker function, the spherical Whittaker function, that is right $\mathrm{GL}_2(\mathbb{Z}_p)$ -invariant and satisfies

$$W_p \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \begin{cases} \sum_{m=0}^{v(a)} \omega_p(q)^m \omega_p^{-1}(q)^{v(a)-m} |a|_p^{\frac{1}{2}} & \text{if } 0 < |a|_p \leq 1, \text{ so that } v(a) \geq 0, \\ 0 & \text{if } |a|_p \geq p, \text{ so that } v(a) \leq -1. \end{cases}$$

(2) For $p \mid q$, the representation π_p is a special representation $\omega_p \mathrm{St}_p$, where ω_p is an unramified unitary character of $\mathbb{Q}_p^\times$, so that $\omega_p(q) \in \{1, -1\}$. This character is such that $\omega_p(q)$ is equal to $\alpha_f(q)$ as in (4.4). There is a distinguished Whittaker function, the Whittaker newform, that is right-invariant under the congruence subgroup of $\mathrm{GL}_2(\mathbb{Z}_p)$ consisting of elements $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ for which $c \in p\mathbb{Z}_p$ and satisfies

$$W_p \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \begin{cases} \omega_p(a) |a|_p & \text{if } 0 < |a|_p \leq 1, \text{ so that } v(a) \geq 0, \\ 0 & \text{if } |a|_p \geq p, \text{ so that } v(a) \leq -1. \end{cases}$$

4.3.5. Archimedean Whittaker functions. Let $\phi = \phi_{F_k}$ be the adèlic lift of a Maaß cusp form $F_k \in \mathcal{C}_k(\Gamma_0(q))$ associated to a Hecke–Maaß newform $f \in \mathcal{C}_0(\Gamma_0(q))$ as in (4.7). The local Whittaker function

$$W_\infty = W_\infty^k \in \mathcal{W}(\pi_\infty, \psi_\infty)$$

is again of a distinguished form. Since f has weight 0, the representation π_∞ is a principal series representation of the form $\mathrm{sgn}^\kappa |\cdot|_\infty^{it} \boxplus \mathrm{sgn}^\kappa |\cdot|_\infty^{-it}$ with

$$\kappa = \kappa_f \in \{0, 1\} \quad \text{and} \quad t = t_f \in \mathbb{R} \cup i \left[-\frac{7}{64}, \frac{7}{64} \right]$$

such that $(-1)^{\kappa_f} = \epsilon_f$ is the parity of f and t_f is the spectral parameter of f , so that $\frac{1}{4} + t_f^2 = \lambda_f$ is the Laplacian eigenvalue of f . Here $|\cdot|_\infty = |\cdot|$ is the usual archimedean absolute value on $\mathbb{R}$.

The following claims are essentially implicit (albeit with some typographical errors) in the seminal work of Jacquet and Langlands [42, Section 2.5], as further detailed by Godement [26, Sections 2.3–2.6]; see also [58]. For the sake of completeness, we give explicit proofs.

Lemma 4.3. (1) For each $k \in 2\mathbb{Z}$, there exists a distinguished Whittaker function $W_\infty^k \in \mathcal{W}(\pi_\infty, \psi_\infty)$, where $\pi_\infty = \mathrm{sgn}^\kappa |\cdot|_\infty^{it} \boxplus \mathrm{sgn}^\kappa |\cdot|_\infty^{-it}$, that is of weight k , so that for all $g_\infty \in \mathrm{GL}_2(\mathbb{R})$ and $\theta \in [0, 2\pi)$, it satisfies

$$W_\infty^k \left(g_\infty \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \right) = e^{ik\theta} W_\infty^k(g_\infty).$$

(2) For all $a \in \mathbb{R}^\times$, this distinguished Whittaker function satisfies

$$W_\infty^k \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \begin{cases} (-1)^{\frac{k}{2}} W_{\frac{k}{2}, it}(4\pi a), & a > 0, k \in 2\mathbb{N} \cup \{0\}, \\ (-1)^\kappa \frac{\Gamma(\frac{k+1}{2} + it) \Gamma(\frac{k+1}{2} - it)}{\Gamma(\frac{1}{2} + it) \Gamma(\frac{1}{2} - it)} \\ \quad \times W_{-\frac{k}{2}, it}(4\pi |a|), & a < 0, k \in 2\mathbb{N} \cup \{0\}, \\ \frac{\Gamma(\frac{1-k}{2} + it) \Gamma(\frac{1-k}{2} - it)}{\Gamma(\frac{1}{2} + it) \Gamma(\frac{1}{2} - it)} \\ \quad \times W_{\frac{k}{2}, it}(4\pi a), & a > 0, k \in -2\mathbb{N}, \\ (-1)^{\kappa + \frac{k}{2}} W_{-\frac{k}{2}, it}(4\pi |a|), & a < 0, k \in -2\mathbb{N}. \end{cases} \quad (4.22)$$

(3) For $\kappa' \in \{0, 1\}$ and $\operatorname{Re}(s) \geq \frac{1}{2}$, we have that

$$\begin{aligned} & \int_{\mathbb{R}^\times} W_\infty^2 \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \operatorname{sgn}^{\kappa'}(a) |a|^{s-\frac{1}{2}} d^\times a \\ &= \begin{cases} -2\pi^{-s} \Gamma\left(\frac{s+1+it}{2}\right) \Gamma\left(\frac{s+1-it}{2}\right) & \text{if } \kappa \equiv \kappa' + 1 \pmod{2}, \\ \left(\frac{1}{2} - s\right) \pi^{-s} \Gamma\left(\frac{s+it}{2}\right) \Gamma\left(\frac{s-it}{2}\right) & \text{if } \kappa \equiv \kappa' \pmod{2}, \end{cases} \end{aligned} \quad (4.23)$$

where $d^\times a := |a|^{-1} da$ denotes the multiplicative Haar measure on $\mathbb{R}^\times$.

(4) We have that

$$\int_{\mathbb{R}^\times} \left| W_\infty^2 \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \right|^2 d^\times a = \left(\frac{1}{4} + t^2\right) \Gamma\left(\frac{1}{2} + it\right) \Gamma\left(\frac{1}{2} - it\right). \quad (4.24)$$

Proof. (1) Let $\pi_\infty = \operatorname{sgn}^\kappa | \cdot |_0^{t_1} \boxplus \operatorname{sgn}^\kappa | \cdot |_0^{t_2}$ be a principal series representation with $\kappa \in \{0, 1\}$ and $t_1, t_2 \in \mathbb{C}$. We initially assume that $\operatorname{Re}(t_1) > \operatorname{Re}(t_2)$. For each $k \in 2\mathbb{Z}$ and $(x_1, x_2) \in \mathbb{R}^2$, let

$$\begin{aligned} \Phi^k(x_1, x_2) &:= (x_2 - \operatorname{sgn}(k)ix_1)^{|k|} e^{-\pi(x_1^2 + x_2^2)} \\ &= \begin{cases} (x_2 - ix_1)^k e^{-\pi(x_1^2 + x_2^2)} & \text{if } k \in 2\mathbb{N} \cup \{0\}, \\ (x_2 + ix_1)^{-k} e^{-\pi(x_1^2 + x_2^2)} & \text{if } k \in -2\mathbb{N}. \end{cases} \end{aligned}$$

Define the Godement section $\varphi_\infty^k: \operatorname{GL}_2(\mathbb{R}) \rightarrow \mathbb{C}$ by

$$\varphi_\infty^k(g_\infty) := \pi^{\frac{|k|}{2}} \operatorname{sgn}^\kappa(\det g_\infty) |\det g_\infty|^{t_1 + \frac{1}{2}} \int_{\mathbb{R}^\times} |y|^{t_2 - t_1 - 1} \Phi^k((0, y^{-1})g_\infty) d^\times y.$$

This Godement section converges absolutely and defines an element of the induced model of π_∞ of weight k ; that is,

$$\varphi_\infty^k \left(\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} g_\infty \right) = \operatorname{sgn}^\kappa(a) |a|^{t_1 + \frac{1}{2}} \operatorname{sgn}^\kappa(d) |d|^{t_2 - \frac{1}{2}} \varphi_\infty^k(g_\infty), \quad (4.25)$$

and

$$\varphi_\infty^k \left(g_\infty \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \right) = e^{ik\theta} \varphi_\infty^k(g_\infty) \quad (4.26)$$

for all $a, d \in \mathbb{R}^\times$, $b \in \mathbb{R}$, $g_\infty \in \mathrm{GL}_2(\mathbb{R})$, and $\theta \in [0, 2\pi)$. Taking g_∞ to be the identity, we have the normalisation

$$\varphi_\infty^k \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \pi^{-\frac{1+t_1-t_2}{2}} \Gamma\left(\frac{1+|k|+t_1-t_2}{2}\right). \quad (4.27)$$

The Jacquet integral

$$W_\infty^k(g_\infty) := \int_{\mathbb{R}} \varphi_\infty^k \left(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} g_\infty \right) e(-u) du \quad (4.28)$$

$$= \pi^{\frac{|k|}{2}} \mathrm{sgn}^k(\det g_\infty) |\det g_\infty|^{t_1+\frac{1}{2}} \int_{\mathbb{R}^\times} |y|^{t_2-t_1} \\ \times \int_{\mathbb{R}} \Phi^k((y^{-1}, u)g_\infty) e^{-2\pi i y u} du d^\times y \quad (4.29)$$

converges absolutely and defines an element of the Whittaker model $\mathcal{W}(\pi_\infty, \psi_\infty)$ of π_∞ of weight k . By the Iwasawa decomposition, every $g_\infty \in \mathrm{GL}_2(\mathbb{R})$ can be written in the form

$$g_\infty = \begin{pmatrix} z & 0 \\ 0 & z \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

with $x \in \mathbb{R}$, $a, z \in \mathbb{R}^\times$, and $\theta \in [0, 2\pi)$. From (4.25), (4.26), and the change of variables $u \mapsto u - x$, we see that for this value of g_∞ ,

$$W_\infty^k(g_\infty) = |z|^{t_1+t_2} e(x) e^{ik\theta} W_\infty^k \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}.$$

The Whittaker function W_∞^k extends holomorphically as a function of the complex variables $t_1, t_2 \in \mathbb{C}$ to $(t_1, t_2) = (it, -it)$ with $t \in \mathbb{R} \cup i[-\frac{7}{64}, \frac{7}{64}]$. This holomorphic extension defines a weight k element of the Whittaker model $\mathcal{W}(\pi_\infty, \psi_\infty)$ of $\pi_\infty = \mathrm{sgn}^k | \cdot |_\infty^{t_1} \boxplus \mathrm{sgn}^k | \cdot |_\infty^{-it}$.

(2) We again let $\pi_\infty = \mathrm{sgn}^k | \cdot |_\infty^{t_1} \boxplus \mathrm{sgn}^k | \cdot |_\infty^{t_2}$ and initially assume that $\mathrm{Re}(t_1) > \mathrm{Re}(t_2)$. Since

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1+\frac{u^2}{a^2}}} & -\frac{u}{a\sqrt{1+\frac{u^2}{a^2}}} \\ 0 & a\sqrt{1+\frac{u^2}{a^2}} \end{pmatrix} \begin{pmatrix} \frac{u}{a\sqrt{1+\frac{u^2}{a^2}}} & -\frac{1}{\sqrt{1+\frac{u^2}{a^2}}} \\ \frac{1}{\sqrt{1+\frac{u^2}{a^2}}} & \frac{u}{a\sqrt{1+\frac{u^2}{a^2}}} \end{pmatrix},$$

we have by (4.25)–(4.28), and the change of variables $u \mapsto au$ that

$$W_\infty^k \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = (-1)^{\frac{k}{2}} \pi^{-\frac{1+t_1-t_2}{2}} \Gamma\left(\frac{1+|k|+t_1-t_2}{2}\right) \mathrm{sgn}(a)^k |a|^{\frac{1}{2}+t_2} \\ \times \int_{\mathbb{R}} (1+iu)^{-\frac{1-k+t_1-t_2}{2}} (1-iu)^{-\frac{1+k+t_1-t_2}{2}} e(-au) du.$$

From [28, 3.384.9], this integral is equal to

$$\frac{\pi^{\frac{1+t_1-t_2}{2}} |a|^{-\frac{1-t_1+t_2}{2}}}{\Gamma(\frac{1+\operatorname{sgn}(a)k+t_1-t_2}{2})} W_{\operatorname{sgn}(a)\frac{k}{2}, \frac{t_1-t_2}{2}}(4\pi|a|).$$

Analytically continuing to $t_1 = it$ and $t_2 = -it$, we obtain (4.22).

(3) For $a \in \mathbb{R}^\times$, we have that

$$W_\infty^k \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \pi^{\frac{|k|}{2}} \operatorname{sgn}^\kappa(a) |a|^{it+\frac{1}{2}} \int_{\mathbb{R}^\times} |y|^{-2it} \int_{\mathbb{R}} \Phi^k(y^{-1}a, u) e^{-2\pi i y u} du d^\times y \quad (4.30)$$

from (4.29). We insert identity (4.30) for W_∞^2 into the left-hand side of (4.23), make the change of variables $a \mapsto ya$ and $u \mapsto u + ia$, then shift the contour of integration back to the line $\operatorname{Im}(u) = 0$, yielding

$$\pi \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(a) |a|^{s+it} \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(y) |y|^{s-it} e^{2\pi i ya} \int_{\mathbb{R}} u^2 e^{-\pi u^2} e^{-2\pi i(y+a)u} du d^\times y d^\times a.$$

The innermost integral may be evaluated via integration by parts, leading to

$$\begin{aligned} & \frac{1}{2} \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(a) |a|^{s+it} e^{-\pi a^2} d^\times a \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(y) |y|^{s-it} e^{-\pi y^2} d^\times y \\ & - \pi \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(a) |a|^{s+it} e^{-\pi a^2} d^\times a \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(y) |y|^{s+2-it} e^{-\pi y^2} d^\times y \\ & - \pi \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(a) |a|^{s+2+it} e^{-\pi a^2} d^\times a \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'}(y) |y|^{s-it} e^{-\pi y^2} d^\times y \\ & - 2\pi \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'+1}(a) |a|^{s+1+it} e^{-\pi a^2} d^\times a \int_{\mathbb{R}^\times} \operatorname{sgn}^{\kappa+\kappa'+1}(y) |y|^{s+1-it} e^{-\pi y^2} d^\times y. \end{aligned}$$

The first three expressions vanish if $\kappa \equiv \kappa' + 1 \pmod{2}$, while the last vanishes if $\kappa \equiv \kappa' \pmod{2}$. The result then follows via the recurrence relation $\Gamma(s+1) = s\Gamma(s)$.

(4) The left-hand side of (4.24) is

$$\int_0^\infty \left(W_{1,it}(4\pi a)^2 + \left(\frac{1}{4} + t^2 \right)^2 W_{-1,it}(4\pi a)^2 \right) \frac{da}{a}$$

from (4.22). The desired identity then follows from the change of variables $a \mapsto \frac{a}{4\pi}$ together with [28, 7.611.4, 8.365.1, and 8.365.8]. ■

5. The boundary of $\mathcal{F}_A(q)$

We turn our attention to the boundary of the canonical fundamental domain $\mathcal{F}_A(q)$ of the hyperbolic orbifold $\Gamma_A(q) \backslash \mathcal{N}_A(q)$. We give an explicit description of the boundary in terms of the homology of $X_0(q)$, which relies crucially on the fact that $\mathcal{P}(q)$ has a minimal number of sides. For related arguments in the case of a general fundamental

polygon, see [55, Section 3]. Furthermore, we introduce the notion of regularised integrals of Eisenstein series along geodesics. These two concepts are key in understanding the *topological terms* in the Weyl sums for our equidistribution problem.

Recall that the boundary of $\mathcal{F}_A(q)$ consists of two parts: the closed geodesic $\mathcal{C}_A(q)$ and certain $\Gamma_0(q)$ -translates of the sides of $\mathcal{P}(q)$. When considering the boundary of $\mathcal{F}_A(q)$ modulo $\Gamma_0(q)$, the sides of $\mathcal{P}(q)$ that are paired together are indistinguishable. With this in mind, we let

$$\mathcal{C}_1, \dots, \mathcal{C}_{2g+1+e_2+e_3} \subset \partial\mathcal{P}(q) \quad (5.1)$$

be a sequence of edges of $\mathcal{P}(q)$ such that for each pair of $\Gamma_0(q)$ -equivalent edges $\mathcal{L}, \mathcal{L}'$ of $\mathcal{P}(q)$ with $\mathcal{L}$ to the left of $\mathcal{L}'$, we have $\mathcal{C}_i = \mathcal{L}$ for some $1 \leq i \leq 2g + 1 + e_2 + e_3$ and such that $\mathcal{C}_1, \dots, \mathcal{C}_{2g}$ are hyperbolic sides. Note that $\mathcal{C}_i$ for $1 \leq i \leq 2g$ define closed curves on $X_0(q)$ since the endpoints are all $\Gamma_0(q)$ -equivalent to the cusp 0.

5.1. Homology of modular curves

Let q be an odd prime. The closed modular surface $X_0(q)$ of level q is a compact Riemann surface of genus $g = \frac{q}{12} + O(1)$. The associated integral singular homology and cohomology groups $H_1(X_0(q), \mathbb{Z})$ and $H^1(X_0(q), \mathbb{Z})$ are free abelian groups of rank $2g$ that sit as lattices inside $H_1(X_0(q), \mathbb{R})$ and $H^1(X_0(q), \mathbb{R})$, respectively. By general principles, we have the cap product pairing between homology and cohomology

$$\langle \cdot, \cdot \rangle_{\text{cap}}: H_1(X_0(q), \mathbb{R}) \times H^1(X_0(q), \mathbb{R}) \rightarrow \mathbb{R}, \quad (5.2)$$

which is a perfect pairing [33, Section 3.3]. The cap product pairing identifies the homology group with the (linear) dual of the cohomology group and vice versa. Furthermore, the Hecke operators as well as the Atkin–Lehner operator W_q are self-adjoint with respect to the cap product pairing (see, e.g., [54, Section 3.2.1] for explicit formulæ for the action of Hecke and Atkin–Lehner operators on homology and cohomology).

Given a closed curve $\mathcal{C}$ inside $X_0(q)$, we denote by

$$[\mathcal{C}] \in H_1(X_0(q), \mathbb{Z}) \subset H_1(X_0(q), \mathbb{R})$$

the associated homology class. Recall that $H_1(X_0(q), \mathbb{Z})$ is generated as an abelian group by such classes [33, Chapter 2]. If h is a holomorphic cusp form of weight 2 and level q , then $h(z) dz$ and $\overline{h(z)} d\overline{z}$ define complex-valued harmonic 1-forms on $X_0(q)$ and such 1-forms span the entire space of harmonic 1-forms. Given any harmonic 1-form ω , we get an associated (complex-valued) cohomology class:

$$[\mathcal{C}] \mapsto \int_{\mathcal{C}} \omega,$$

for all closed curves $\mathcal{C}$. By theorems of de Rham and Hodge, this association yields an isomorphism between the real cohomology group $H^1(X_0(q), \mathbb{R})$ and the space of real-valued harmonic 1-forms.

We consider two bases for the homology and cohomology groups.

5.1.1. The Hecke basis. Let $\mathcal{B}_2^{\text{hol}}(\Gamma_0(q))$ be the orthogonal basis of weight 2 and level q holomorphic cusp forms of level q consisting of Hecke eigenforms normalised so that the first Fourier coefficient of each cusp form is 1; note that these are all newforms since there are no cusp forms of weight 2 and level 1. For each $h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))$,

$$\frac{1}{2}i^{\frac{1-\tau}{2}}(h(z)dz \pm \overline{h(z)}d\bar{z}) \quad (5.3)$$

defines a real-valued closed 1-form. We denote by $\omega_h^\pm \in H^1(X_0(q), \mathbb{R})$ the associated cohomology class. It follows by dimensional considerations that

$$\{\omega_h^\epsilon : h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q)), \epsilon \in \{+, -\}\} \quad (5.4)$$

is a basis for $H^1(X_0(q), \mathbb{R})$. We denote by

$$\{v_h^\epsilon : h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q)), \epsilon \in \{+, -\}\}$$

the basis of $H_1(X_0(q), \mathbb{R})$ that is dual to (5.4) with respect to the cap product pairing (5.2); by this, we mean that for $h_1, h_2 \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))$ and $\epsilon_1, \epsilon_2 \in \{+, -\}$,

$$\langle v_{h_1}^{\epsilon_1}, \omega_{h_2}^{\epsilon_2} \rangle_{\text{cap}} = \begin{cases} 1 & \text{if } h_1 = h_2 \text{ and } \epsilon_1 = \epsilon_2, \\ 0 & \text{otherwise.} \end{cases}$$

5.1.2. The special basis. By general principles (since $\mathbb{H}$ is contractible), we have a sequence of surjective maps

$$\Gamma_0(q) \twoheadrightarrow \text{Conj}(\Gamma_0(q)) \twoheadrightarrow \Gamma_0(q)^{\text{ab}} \twoheadrightarrow H_1(Y_0(q), \mathbb{Z}) \twoheadrightarrow H_1(X_0(q), \mathbb{Z}), \quad (5.5)$$

where the composition $\Gamma_0(q) \twoheadrightarrow H_1(X_0(q), \mathbb{Z})$ is given by

$$\gamma \mapsto [\mathcal{C}_\gamma],$$

with $\mathcal{C}_\gamma$ any piecewise geodesic curve connecting z with γz , where $z \in \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$ (as the homology class is independent of these choices). The surjectivity of the first two maps is evident and the composition of all of the maps can be thought of as the map from the orbifold fundamental group $\pi_1^{\text{orb}}(Y_0(p)) \cong \Gamma_0(p)$ (using that $\mathbb{H}$ is the universal orbifold covering space of the orbifold $Y_0(p)$) to the orbifold homology group $H_1^{\text{orb}}(Y_0(p), \mathbb{Z}) \cong \Gamma_0(p)^{\text{ab}}$ composed with the surjective map from $H_1^{\text{orb}}(Y_0(p), \mathbb{Z})$ to the homology of the underlying topological space of $X_0(p)$ (see, e.g., [13, Section 2.2]). Now one sees directly that the kernel of the last map contains images of the parabolic and elliptic conjugacy classes in $\Gamma_0(q)$ and furthermore by dimension considerations that these classes even generate the kernel. Thus we conclude that the images of the hyperbolic labels α_j of $\mathcal{P}(q)$ under (5.5) generate the lattice $H_1(X_0(q), \mathbb{Z})$ and so, in particular, span $H_1(X_0(q), \mathbb{R})$.

Lemma 5.1. *The cohomology classes $[\mathcal{C}_1], \dots, [\mathcal{C}_{2g}]$ define an integral basis for the lattice $H_1(X_0(q), \mathbb{Z})$.*

Proof. We start by noticing that the homology class associated to a hyperbolic label α_j with $j > j^*$ can be written as

$$\sum_{j^* < k < j} [\mathcal{L}_k], \quad (5.6)$$

where $[\mathcal{L}_k]$ denotes the homology class of the side $\mathcal{L}_k$ of $\mathcal{P}(q)$. To see this, observe that the left-most vertex of the side $\mathcal{L}_j$ is equal to $\alpha_j v$, where v is the right-most vertex of $\mathcal{L}_{j^*}$. Thus if $j^* < j$, we see that the concatenation of the sides

$$\mathcal{L}_{j^*+1}, \mathcal{L}_{j^*+2}, \dots, \mathcal{L}_{j-1}$$

defines a curve connecting v and $\alpha_j v$, which gives the claimed expression (5.6).

This shows that the homology class associated to a hyperbolic label α_j lies in the $\mathbb{Z}$ -span of $\{[\mathcal{C}_i] : 1 \leq i \leq 2g + 1 + e_2 + e_3\}$. Since the classes associated to the hyperbolic labels via map (5.5) generate the lattice $H_1(X_0(q), \mathbb{Z})$ and the classes of parabolic and elliptic sides vanish, we get the desired conclusion by dimension considerations. ■

We denote by

$$\omega_1, \dots, \omega_{2g} \in H^1(X_0(q), \mathbb{Z}) \subset H^1(X_0(q), \mathbb{R}) \quad (5.7)$$

the dual basis of $[\mathcal{C}_1], \dots, [\mathcal{C}_{2g}]$ with respect to the cap product pairing. We refer to these bases as the *special basis* of homology and cohomology, respectively.

5.2. A homological description of the boundary

In studying the projection of $\mathcal{F}_A(q)$ to the closed modular surface $X_0(q)$, we need to understand the boundary of this projection. Unlike the level 1 case, this is not simply the geodesic $\mathcal{C}_A(q)$, for $\mathcal{C}_A(q)$ may be nontrivial in homology (which corresponds to the fact that $\Gamma_A(q)$ need not be contained in $\Gamma_0(q)$).

First of all, denote by $\partial\mathcal{F}_A(q)$ the image of the oriented boundary of $\mathcal{F}_A(q)$ under the projection to $X_0(q)$. Formally, this is nothing but a singular 1-chain (as in the theory of singular homology [33, Chapter 2]), which is to say an element of

$$\mathbb{Z}[\{\varphi: [0, 1] \rightarrow X_0(q) : \text{continuous}\}].$$

We write

$$\partial\mathcal{F}_A(q) = \mathcal{C}_A(q) \cup \bigcup_{i=1}^{2g+1+e_2+e_3} m_i(A, q) \mathcal{C}_i,$$

where $m_i(A, q) \in \mathbb{Z}$ is the multiplicity of the edge $\mathcal{C}_i$ in $\mathcal{F}_A(q)$, namely the signed number of times the boundary of $\mathcal{F}_A(q)$ contains $(\Gamma_0(q)$ -translates of) the two oriented edges of $\mathcal{P}(q)$ corresponding to $\mathcal{C}_i$. Explicitly, this means that for a compactly supported 1-form ω on $X_0(q)$,

$$\int_{\partial\mathcal{F}_A(q)} \omega = \int_{\mathcal{C}_A(q)} \omega + \sum_{i=1}^{2g+1+e_2+e_3} m_i(A, q) \int_{\mathcal{C}_i} \omega.$$

Lemma 5.2 (Cf. [55, Lemma 3.1]). *Let $\mathcal{C}_i$ be an elliptic or parabolic edge of $\mathcal{P}(q)$. Then we have that*

$$m_i(A, q) = 0.$$

Proof. Consider the boundary of the special fundamental polygon $\mathcal{P}(q)$ as a (cyclic) graph with $4g + 2 + 2e_2 + 2e_3$ vertices with its natural action of $\Gamma_0(q)$, and denote by

$$\mathcal{G}(q) := \partial\mathcal{P}(q)/\Gamma_0(q)$$

the quotient graph. The graph $\mathcal{G}(q)$ has $2 + e_2 + e_3$ vertices and $2g + 1 + e_2 + e_3$ edges. The vertices corresponding to the cusp ∞ and the elliptic points all have a single edge connected to the vertex corresponding to the cusp 0, and the vertex 0 has $2g$ self-edges (i.e., the degree of the vertex 0 is $4g + 1 + e_2 + e_3$).

The oriented boundary of $\mathcal{F}_A(q)$ can be described as a weighted version of $\mathcal{G}(q)$; for an edge E of $\mathcal{G}(q)$ corresponding to $\mathcal{C}_i$, we associated the weight $m_i(A, q) \in \mathbb{Z}$. The key observation is now that by construction these weights are all obtained by taking a cycle in the signed graph and recording the signed number of times the cycle crosses each edge. The claim now follows since the vertices of $\mathcal{G}(q)$ corresponding to ∞ and elliptic points have degree 1 (i.e., there is only one way to go to the elliptic points and to ∞). ■

Lemma 5.2 shows that the oriented boundary of $\mathcal{F}_A(q)$, when projected to $X_0(q)$, consists only of hyperbolic sides of $\mathcal{P}(q)$. This means that we can rewrite the boundary solely in terms of the hyperbolic sides $\mathcal{C}_1, \dots, \mathcal{C}_{2g}$ defined above:

$$\partial\mathcal{F}_A(q) = \mathcal{C}_A(q) \cup \bigcup_{i=1}^{2g} m_i(A, q) \mathcal{C}_i.$$

Lemma 5.3. *For $i \in \{1, \dots, 2g\}$, we have that*

$$m_i(A, q) = -\langle [\mathcal{C}_A(q)], \omega_i \rangle_{\text{cap}},$$

where $\omega_1, \dots, \omega_{2g} \in H^1(X_0(q), \mathbb{Z})$ is the special basis of cohomology defined in (5.7).

Proof. We observe that in homology, we have that

$$[\mathcal{C}_A(q)] + \sum_{i=1}^{2g} m_i(A, q) [\mathcal{C}_i] = [\partial\mathcal{F}_A(q)] = 0 \in H_1(X_0(q), \mathbb{Z}),$$

since $\mathcal{F}_A(q) \subset \mathbb{H}$ and $\mathbb{H}$ is contractible. Now the claim follows since the coefficient of the basis element $[\mathcal{C}_i]$ in the expansion of $[\mathcal{C}_A(q)]$ with respect to the special basis from Lemma 5.1 is exactly

$$\langle [\mathcal{C}_A(q)], \omega_i \rangle_{\text{cap}},$$

by the definition of the dual basis. ■

5.3. Regularised integrals

For a cusp $\mathfrak{b} \in \mathbb{P}^1(\mathbb{Q})$ and $Y > 0$, we define the cuspidal zone

$$\mathcal{F}_{\mathfrak{b}}(Y) := \{z \in \mathbb{H} : 0 < \operatorname{Re}(\sigma_{\mathfrak{b}}^{-1}z) < 1, \operatorname{Im}(\sigma_{\mathfrak{b}}^{-1}z) \geq Y\}, \quad (5.8)$$

where $\sigma_{\mathfrak{b}} \in \Gamma$ is a scaling matrix for $\mathfrak{b}$, so that $\sigma_{\mathfrak{b}}\infty = \mathfrak{b}$; note that $\mathcal{F}_{\mathfrak{b}}(Y)$ is independent of the choice of such a scaling matrix. One can readily check that for $\gamma \in \Gamma$, this satisfies

$$\gamma \mathcal{F}_{\mathfrak{b}}(Y) = \mathcal{F}_{\gamma\mathfrak{b}}(Y). \quad (5.9)$$

Furthermore, for $\ell \in \mathbb{N}$ with corresponding matrix

$$g_{\ell} := \begin{pmatrix} \sqrt{\ell} & 0 \\ 0 & \frac{1}{\sqrt{\ell}} \end{pmatrix} \in \operatorname{SL}_2(\mathbb{R}),$$

we have that

$$g_{\ell} \mathcal{F}_{\mathfrak{b}}(Y) = \mathcal{F}_{g_{\ell}\mathfrak{b}}\left(\frac{(\ell, \operatorname{den}(\mathfrak{b}))^2}{\ell} Y\right), \quad (5.10)$$

where $\operatorname{den}(\mathfrak{b})$ denotes the denominator of $\mathfrak{b} \in \mathbb{P}^1(\mathbb{Q})$ written in reduced form, with the convention that

$$\operatorname{den}(0) := 1 \quad \text{and} \quad \operatorname{den}(\infty) := 0.$$

Lemma 5.4 (Cf. [8, Appendix]). *Let $\mathcal{C} \subset \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ be a geodesic with endpoints $\mathfrak{b}_1, \mathfrak{b}_2 \in \mathbb{P}^1(\mathbb{Q})$ with $\mathfrak{b}_1 \neq \mathfrak{b}_2$. Then the limit*

$$\begin{aligned} & \int_{\mathcal{C}}^* (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)} \\ &:= \lim_{Y \rightarrow \infty} \int_{\mathcal{C} \setminus (\mathcal{F}_{\mathfrak{b}_1}(Y) \cup \mathcal{F}_{\mathfrak{b}_2}(Y))} (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)} \end{aligned} \quad (5.11)$$

exists.

We may think of limit (5.11) as a regularised integral, akin to Zagier's regularisation of integrals on $\Gamma \backslash \mathbb{H}$ [66].

Proof of Lemma 5.4. We let $\sigma_{\mathfrak{b}_1} \in \Gamma$ be a scaling matrix for $\mathfrak{b}_1$. Upon making the change of variables $z \mapsto \sigma_{\mathfrak{b}_1}^{-1}z$ and recalling (5.9), we may reduce to the case that $\mathfrak{b}_1 = \infty$ and $\mathfrak{b}_2 = \frac{a}{b}$ with $a \in \mathbb{Z}$, $b \in \mathbb{N}$, and $(a, b) = 1$. For sufficiently large $Y > 0$, we may write

$$\begin{aligned} & \int_{\mathcal{C} \setminus (\mathcal{F}_{\infty}(Y) \cup \mathcal{F}_{\frac{a}{b}}(Y))} (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)} \\ &= \int_{\frac{a}{b} + \frac{i}{b}}^{\frac{a}{b} + iY} (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)} + \int_{\frac{a}{b} + \frac{i}{b^2 Y}}^{\frac{a}{b} + \frac{i}{b}} (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)}. \end{aligned}$$

Let $\sigma_{\mathfrak{b}_2} \in \Gamma$ be a scaling matrix for $\mathfrak{b}_2$; in particular, we may choose $\sigma_{\mathfrak{b}_2} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ if $a = 0$, while if $a \neq 1$, then there exists some $\bar{a} \in \mathbb{Z}$ such that $a\bar{a} \equiv 1 \pmod{b}$, in which case we may take

$$\sigma_{\mathfrak{b}_2} = \begin{pmatrix} a & \frac{a\bar{a}-1}{b} \\ b & \bar{a} \end{pmatrix}.$$

Making the change of variables $z \mapsto \sigma_{\mathfrak{b}_2}^{-1}z$ in the second integral above, we see that the right-hand side is equal to

$$\int_{\frac{a}{b} + \frac{i}{b}}^{\frac{a}{b} + iY} (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)} - \int_{-\frac{\bar{a}}{b} + \frac{i}{b}}^{-\frac{\bar{a}}{b} + iY} (R_0 E) \left(z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)}$$

if $a \neq 0$, while the same holds for $a = 0$ with $\bar{a}$ replaced by 0. Now we insert the Fourier expansion (4.11). Since the constant terms cancel, we see that the limit exists as Y tends to infinity due the rapid decay of the Whittaker function, namely $W_{\pm 1, it}(4\pi y) \ll ye^{-2\pi y}$ as y tends to infinity. ■

Corollary 5.5. *Let $\ell \in \mathbb{N}$ be squarefree. Let $\mathcal{C}$ be the geodesic connecting $\mathfrak{b}_1$ and $\mathfrak{b}_2$, where $\mathfrak{b}_1, \mathfrak{b}_2 \in \mathbb{P}^1(\mathbb{Q})$ are $\Gamma_0(\ell)$ -equivalent. Then the limit*

$$\lim_{Y \rightarrow \infty} \int_{\mathcal{C} \setminus (\mathcal{F}_{\mathfrak{b}_1}(Y) \cup \mathcal{F}_{\mathfrak{b}_2}(Y))} (R_0 E) \left(\ell z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)} \quad (5.12)$$

exists and is equal to

$$\int_{\mathcal{C}}^* (R_0 E) \left(\ell z, \frac{1}{2} + it \right) \frac{dz}{\operatorname{Im}(z)}.$$

Proof. This follows directly from Lemma 5.4 combined with (5.10) since $\frac{a_1}{b_1}, \frac{a_2}{b_2} \in \mathbb{P}^1(\mathbb{Q})$ are $\Gamma_0(\ell)$ -equivalent precisely when $(b_1, \ell) = (b_2, \ell)$. ■

6. Weyl sums

6.1. Weyl sums for newforms and L -functions

Let q be a positive squarefree integer for which every prime dividing q splits in E , let $f \in \mathcal{C}_0(\Gamma_0(q))$ be a Hecke–Maaß newform, and let χ be a narrow class character of E . Our goal is to relate the Weyl sum

$$W_{\chi, f} := \sum_{A \in \operatorname{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A(q)} f(z) d\mu(z) \quad (6.1)$$

to a special value of the Rankin–Selberg L -function $L(s, f \otimes \Theta_\chi)$, where Θ_χ denotes the theta series associated to χ , as in [38, Appendix A.1], which is a newform of weight 0, level D , nebentypus χ_D , Laplacian eigenvalue $\lambda_{\Theta_\chi} = \frac{1}{4}$, and parity $\epsilon_{\Theta_\chi} = \chi(J) \in \{1, -1\}$. The automorphic form Θ_χ is a cusp form if and only if χ is complex; otherwise Θ_χ is an Eisenstein series and χ is a genus character.

Proposition 6.1. *Let q be either 1 or a prime that splits in E , let $f \in \mathcal{C}_0(\Gamma_0(q))$ be a Hecke–Maaß newform normalised such that*

$$\int_{\Gamma_0(q) \backslash \mathbb{H}} |f(z)|^2 d\mu(z) = 1,$$

and let χ be a narrow class character of E . Then there exist constants $c_{\chi,f} \in \mathbb{C}$ and $c_{\chi,h}^{\pm} \in \mathbb{C}$ for each $h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))$ such that

$$W_{\chi,f} = c_{\chi,f} \frac{L(\frac{1}{2}, f \otimes \Theta_{\chi})^{\frac{1}{2}}}{L(1, \text{ad } f)^{\frac{1}{2}}} + \sum_{j=1}^{2g} \int_{\mathcal{C}_j} (R_0 f)(z) \frac{dz}{\text{Im}(z)} \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} \sum_{\pm} \langle v_h^{\pm}, \omega_j \rangle_{\text{cap}} \\ \times \left(c_{\chi,h}^+ L\left(\frac{1}{2}, h \otimes \Theta_{\bar{\chi}}\right)^{\frac{1}{2}} \pm c_{\chi,h}^- L\left(\frac{1}{2}, h \otimes \Theta_{\chi}\right)^{\frac{1}{2}} \right).$$

Here the sum over j is empty if $q = 1$ (so that $g = 0$), $\mathcal{C}_1, \dots, \mathcal{C}_{2g}$ are the hyperbolic sides of $\mathcal{P}(q)$ as in (5.1), and $\omega_1, \dots, \omega_{2g}$ is the special basis of cohomology as in (5.7). Moreover, the constants $c_{\chi,f}$ and $c_{\chi,h}^{\pm}$ satisfy

$$|c_{\chi,f}|^2 = \frac{(1 - \epsilon_f \chi(J)) \sqrt{D} \Gamma(\frac{3}{4} + \frac{it_f}{2})^2 \Gamma(\frac{3}{4} - \frac{it_f}{2})^2}{q(\frac{1}{4} + t_f^2)^2 \Gamma(\frac{1}{2} + it_f) \Gamma(\frac{1}{2} - it_f)}, \\ |c_{\chi,h}^{\pm}|^2 = \frac{\sqrt{D}}{16\pi^2(\frac{1}{4} + t_f^2)^2},$$

where $t_f \in \mathbb{R} \cup i[-\frac{7}{64}, \frac{7}{64}]$ denotes the spectral parameter and $\epsilon_f \in \{1, -1\}$ denotes the parity of f .

Remark 6.2. In terms of the completed L -functions

$$\Lambda(s, f \otimes \Theta_{\chi}) := \pi^{-2(s+|\kappa_f-\kappa_{\chi}|)} \Gamma\left(\frac{s+|\kappa_f-\kappa_{\chi}|+it_f}{2}\right)^2 \Gamma\left(\frac{s+|\kappa_f-\kappa_{\chi}|-it_f}{2}\right)^2 \\ \times L(s, f \otimes \Theta_{\chi}), \\ \Lambda(s, \text{ad } f) := \pi^{-\frac{3s}{2}} \Gamma\left(\frac{s}{2} + it_f\right) \Gamma\left(\frac{s}{2}\right) \Gamma\left(\frac{s}{2} - it_f\right) L(s, \text{ad } f),$$

where $\kappa_f, \kappa_{\chi} \in \{0, 1\}$ are such that $(-1)^{\kappa_f} = \epsilon_f$ and $(-1)^{\kappa_{\chi}} = \epsilon_{\Theta_{\chi}} = \chi(J)$, we have that

$$|c_{\chi,f}|^2 \frac{L(\frac{1}{2}, f \otimes \Theta_{\chi})}{L(1, \text{ad } f)} = \frac{(1 - \epsilon_f \chi(J)) \pi^2 \sqrt{D}}{q(\frac{1}{4} + t_f^2)^2} \frac{\Lambda(\frac{1}{2}, f \otimes \Theta_{\chi})}{\Lambda(1, \text{ad } f)}.$$

Remark 6.3. Stirling’s formula implies that

$$c_{\chi,f} \ll D^{\frac{1}{4}} q^{-\frac{1}{2}} (1 + |t_f|)^{-\frac{3}{2}}, \quad c_{\chi,h}^{\pm} \ll D^{\frac{1}{2}} (1 + |t_f|)^{-2}. \quad (6.2)$$

Remark 6.4. It is instructive to consider the case of $q = 1$ and χ a genus character associated to the pair of primitive quadratic Dirichlet characters χ_{D_1} and χ_{D_2} modulo $|D_1|$ and $|D_2|$, respectively, where D_1 and D_2 are fundamental discriminants for

which $D_1 D_2 = D$. Then Θ_χ is the Eisenstein newform associated to χ_{D_1} and χ_{D_2} , as described in [65], and so $L(s, f \otimes \Theta_\chi) = L(s, f \otimes \chi_{D_1}) L(s, f \otimes \chi_{D_2})$. Since $D > 1$, either $D_1, D_2 > 0$ or $D_1, D_2 < 0$; in the former case, we have that $\chi(J) = 1$, while $\chi(J) = -1$ in the latter case. Proposition 6.1 then gives the identity

$$|W_{\chi, f}|^2 = \frac{(1 - \epsilon_f \chi(J)) \sqrt{D}}{q(\frac{1}{4} + t_f^2)^2} \frac{\Gamma(\frac{3}{4} + \frac{it_f}{2})^2 \Gamma(\frac{3}{4} - \frac{it_f}{2})^2}{\Gamma(\frac{1}{2} + it_f) \Gamma(\frac{1}{2} - it_f)} \times \frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)}. \quad (6.3)$$

We see that the Weyl sum $W_{\chi, f}$ vanishes if f is even and $D_1, D_2 > 0$ or if f is odd and $D_1, D_2 < 0$; additionally, $W_{\chi, f}$ vanishes if f is odd and $D_1, D_2 > 0$, for then the root numbers of $f \otimes \chi_{D_1}$ and $f \otimes \chi_{D_2}$ are both equal to -1 [38, Lemma A.2], and hence $L(s, f \otimes \chi_{D_1})$ and $L(s, f \otimes \chi_{D_2})$ both vanish at $s = \frac{1}{2}$. These vanishing results and identity (6.3) when f is even and $D_1, D_2 < 0$ are in exact accordance with the work of Duke, Imamoğlu, and Tóth [19, Theorem 4 and (5.17)].

Similarly, letting $E(z, s)$ denote the Eisenstein series on $\Gamma \backslash \mathbb{H}$, which has parity 1, we relate the Weyl sum

$$W_{\chi, t} := \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A} E\left(z, \frac{1}{2} + it\right) d\mu(z) \quad (6.4)$$

to a special value of $L(s, \Theta_\chi) = L(s, \chi)$.

Proposition 6.5. *Let χ be a narrow class character of E . For $t \in \mathbb{R}$, we have that*

$$W_{\chi, t} = \frac{(1 - \chi(J)) D^{\frac{1}{4} + \frac{it}{2}}}{\frac{1}{4} + t^2} \frac{\Gamma(\frac{3}{4} + \frac{it}{2})^2}{\Gamma(\frac{1}{2} + it)} \frac{L(\frac{1}{2} + it, \Theta_\chi)}{\zeta(1 + 2it)}.$$

Remark 6.6. This is

$$\frac{(1 - \chi(J)) \pi D^{\frac{1}{4} + \frac{it}{2}}}{\frac{1}{4} + t^2} \frac{\Lambda(\frac{1}{2} + it, \Theta_\chi)}{\xi(1 + 2it)}$$

in terms of the completed L -functions

$$\Lambda(s, \Theta_\chi) := \pi^{-s-1} \Gamma\left(\frac{s+1}{2}\right)^2 L(s, \Theta_\chi),$$

$$\xi(s) := \pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) \zeta(s).$$

The proofs of Propositions 6.1 and 6.5 are given in Section 6.6. Our method is to first prove identities relating certain adèlic period integrals to L -functions, then show that these adèlic period integrals are equal to integrals over closed geodesics, and finally relate these integrals over closed geodesics to integrals over hyperbolic orbifolds.

In Section 6.7, we extend Propositions 6.1 and 6.5 to oldforms; note that if $q > 1$ is squarefree, every Eisenstein series is an oldform [65].

6.2. Adèlic period integrals and choices of test vectors

The first step towards proving Proposition 6.1 is to apply a formula due to Martin and Whitehouse [50], extending work of Waldspurger [63], relating certain adèlic period integrals to a ratio of special values of L -functions. We now describe in some detail how the results of Martin and Whitehouse apply in our specific case.

6.2.1. Waldspurger-type formulae. Following the pioneering work of Waldspurger [63], there has been considerable work in obtaining explicit formulae relating adèlic period integrals and central values of Rankin–Selberg L -functions, as in work of Gross [30], Zhang [67], Jacquet and Chen [41], Martin and Whitehouse [50], and File, Martin, and Pitale [24], among others. The setting is as follows: let π be an automorphic representation of $\mathrm{GL}_2(\mathbb{A}_F)$ for some number field F , let E be a quadratic extension of F embedded in a quaternion algebra D defined over F , let $\Omega: E^\times \backslash \mathbb{A}_E^\times \rightarrow \mathbb{C}^\times$ be a unitary Hecke character for which $\Omega|_{\mathbb{A}_F^\times}$ is equal to the central character of π , and let ϕ be a test vector in the automorphic representation π^D of $D^\times(\mathbb{A}_F)$ corresponding to π via the Jacquet–Langlands correspondence. We then define the adèlic period integral

$$\mathcal{P}_\Omega^D(\phi) := \int_{\mathbb{A}_F^\times E^\times \backslash \mathbb{A}_E^\times} \phi(x) \Omega^{-1}(x) dx.$$

Note that implicitly this depends on a choice of embedding $\mathbb{A}_E^\times \hookrightarrow D^\times(\mathbb{A}_F)$, which is suppressed in the notation, as well as a choice of normalisation of the measure dx .

A remarkable result of Waldspurger [63] is the formula

$$\frac{|\mathcal{P}_\Omega^D(\phi)|^2}{\langle \phi, \phi \rangle} = c_{\Omega, \phi} \frac{\Lambda(\frac{1}{2}, \pi_E \otimes \Omega)}{\Lambda(1, \mathrm{ad} \pi)}, \quad (6.5)$$

where ϕ is *any* nonzero test vector in π^D and $c_{\Omega, \phi}$ is a finite product of local factors. Here π_E denotes the base change of π to an automorphic representation of $\mathrm{GL}_2(\mathbb{A}_E)$; alternatively, we may write

$$\Lambda(s, \pi_E \otimes \Omega) = \Lambda(s, \pi \otimes \pi_\Omega),$$

where π_Ω denotes the automorphic induction of the Hecke character Ω to an automorphic representation of $\mathrm{GL}_2(\mathbb{A}_F)$.

6.2.2. An explicit formula. For applications in analytic number theory, it is essential that we have at our disposal a completely explicit form of Waldspurger’s formula (6.5). Building on the work of Jacquet and Chen [41], Martin and Whitehouse [50, Theorem 4.1] provide such a formula for a *specific* choice of test vector $\phi \in \pi^D$ (under some local assumptions, which were slightly relaxed by File, Martin, and Pitale [24]). For this specific choice of test vector, the local factors (whose product we denoted by $c_{\Omega, \phi}$ in (6.5)) are described in [50, Section 4.2].

For our application, we can restrict to the case where $F = \mathbb{Q}$, $E = \mathbb{Q}(\sqrt{D})$ with D a positive fundamental discriminant, $\pi = \pi_f = \pi_\infty \otimes \bigotimes_p \pi_p$ is a cuspidal automorphic

representation of $\mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ associated to a Hecke–Maaß newform $f \in \mathcal{C}_0(\Gamma_0(q))$ of weight 0, principal nebentypus, and squarefree level q for which every prime dividing q splits in E , and Ω is the idèlic lift of a narrow class character χ ; later, we specialise to q either equal to 1 or equal to an odd prime. With this choice of data, the quaternion algebra D is simply the matrix algebra $\mathrm{Mat}_{2 \times 2}$, so that $\pi^D = \pi$. We shorten notation and write $\mathcal{P}_{\Omega}(\phi) := \mathcal{P}_{\Omega}^D(\phi)$ in this case.

The choice of test vector $\phi \in \pi$ used in [50] is characterised by some local compatibilities with the Hecke character Ω and thus implicitly depends on the choice of embedding $\Psi_{\mathbb{A}_{\mathbb{Q}}} : \mathbb{A}_E \hookrightarrow D(\mathbb{A}_{\mathbb{Q}})$. The properties that characterise the local test vectors ϕ_p and ϕ_{∞} are described in [50, p. 172] and are as follows.

- At a nonarchimedean place p , Martin and Whitehouse pick $\phi_p \in \pi_p$ to be nonzero and invariant under the units $R^{\times}$ of a certain order R in the local quaternion algebra (which determines ϕ_p up to scaling). In our setting, R is simply the Eichler order in $\mathrm{GL}_2(\mathbb{Q}_p)$ of reduced discriminant $p^{c(\pi_p)}$ such that

$$R \cap \Psi_p(E_p) = \Psi_p(\mathcal{O}_{E_p}),$$

where $c(\pi_p)$ denotes the conductor exponent of π_p and $E_p := E \otimes \mathbb{Q}_p$.

- At the archimedean place, we let $K_{\infty} \cong \mathrm{O}(2)$ be a maximal compact subgroup of $\mathrm{GL}_2(\mathbb{R})$ such that $K_{\infty} \cap \Psi_{\infty}(E_{\infty}^{\times}) \cong (\mathbb{Z}/2\mathbb{Z})^2$ is a maximal compact subgroup of $\Psi_{\infty}(E_{\infty}^{\times}) \cong (\mathbb{R}^{\times})^2$, where $E_{\infty} := E \otimes \mathbb{R} \cong \mathbb{R}^2$. Martin and Whitehouse pick ϕ_{∞} such that $K_{\infty} \cap \Psi_{\infty}(E_{\infty}^{\times})$ acts (via π) on ϕ_{∞} in the same way as $\Omega_{\infty} : E_{\infty}^{\times} \rightarrow \mathbb{C}^{\times}$ and ϕ_{∞} lies in the *minimal* such K_{∞} -type in the sense of Popa [58, Theorem 1] (which also uniquely determines ϕ_{∞} up to scaling).

In our application, we slightly modify the choice of test vector ϕ_{∞} .

In order to obtain an explicit formula, we must now specify an embedding $\Psi_{\mathbb{A}_{\mathbb{Q}}} : \mathbb{A}_E \hookrightarrow \mathrm{Mat}_{2 \times 2}(\mathbb{A}_{\mathbb{Q}})$ and then determine which choice of local test vectors ϕ_p the above described conditions imply.

6.2.3. A specific test vector. We construct an embedding $\Psi_{\mathbb{A}_{\mathbb{Q}}}$ using an oriented optimal embedding $\Psi : E \hookrightarrow \mathrm{Mat}_{2 \times 2}(\mathbb{Q})$ of level q as described in (2.3) associated to a Heegner form $Q = [a, b, c] \in \mathcal{Q}_D(q)$ as in (2.1). By tensoring with $\mathbb{A}_{\mathbb{Q}}$, we get an embedding

$$\Psi_{\mathbb{A}_{\mathbb{Q}}} = (\Psi_{\infty}, \Psi_2, \Psi_3, \Psi_5, \dots) : \mathbb{A}_E \hookrightarrow \mathrm{Mat}_{2 \times 2}(\mathbb{A}_{\mathbb{Q}}).$$

Since Ψ is an *optimal* embedding, the Eichler order R is exactly the standard order of level $p^{c(\pi_p)}$ for each prime p , so that

$$\prod_p R^{\times} = K_0(q),$$

where $K_0(q) \subset \mathrm{GL}_2(\mathbb{A}_{\mathbb{Q}})$ is the congruence subgroup of level q as in (4.15). This means that we can choose the local component of ϕ at each prime to be the same as those of the adèlisation ϕ_{F_2} of our Maaß cusp form $F_2 = R_0 f$.

At the archimedean place, there is the slight complication that $\Psi_\infty: E_\infty \hookrightarrow \text{Mat}_{2 \times 2}(\mathbb{R})$ is *not* the diagonal embedding. If, however, we conjugate Ψ_∞ by the matrix $\gamma_\infty \in \text{GL}_2^+(\mathbb{R})$ given by

$$\gamma_\infty := \begin{cases} \begin{pmatrix} -b - \sqrt{D} & b - \sqrt{D} \\ 2a & -2a \end{pmatrix} & \text{if } a > 0, \\ \begin{pmatrix} b + \sqrt{D} & b - \sqrt{D} \\ -2a & -2a \end{pmatrix} & \text{if } a < 0, \end{cases}$$

where $(a, b, c) \in \mathbb{Z}^3$ are associated to Ψ as in Section 2.3, then we obtain the diagonal embedding, namely

$$\begin{aligned} (\gamma_\infty \cdot \Psi_\infty)(x + \sqrt{D}y, x - \sqrt{D}y) &:= \gamma_\infty^{-1} \Psi_\infty(x + \sqrt{D}y, x - \sqrt{D}y) \gamma_\infty \\ &= \begin{pmatrix} x + \sqrt{D}y & 0 \\ 0 & x - \sqrt{D}y \end{pmatrix}. \end{aligned} \quad (6.6)$$

Note that $\gamma_\infty i = z_Q$ and $\Psi_\infty(\epsilon_D, \epsilon_D^{-1}) \gamma_\infty i = \gamma_Q z_Q$, where z_Q and $\gamma_Q z_Q$ are as in (2.6).

With this in mind, Martin and Whitehouse choose the local component of ϕ at the archimedean place to be

$$\begin{cases} \pi_\infty(\gamma_\infty) \phi_{F_0, \infty} & \text{if } \epsilon_f = \chi(J), \\ \pi_\infty(\gamma_\infty) \phi_{F_2, \infty} - \pi_\infty(\gamma_\infty) \phi_{F_{-2}, \infty} & \text{if } \epsilon_f = -\chi(J), \end{cases}$$

where $\phi_{F_0, \infty}$, $\phi_{F_2, \infty}$, and $\phi_{F_{-2}, \infty}$ are the local components of the adèlisations ϕ_{F_0} , ϕ_{F_2} , and $\phi_{F_{-2}}$ of $F_0 = f$, $F_2 = R_0 f$, and $F_{-2} = \mathcal{L}_0 f$, respectively. We instead merely take the local component of ϕ at the archimedean place to be $\pi_\infty(\gamma_\infty) \phi_{F_2, \infty}$.

Altogether, the above discussion implies that when using the embedding $\Psi_{\mathbb{A}_\mathbb{Q}}: \mathbb{A}_E \hookrightarrow \text{Mat}_{2 \times 2}(\mathbb{A}_\mathbb{Q})$, our test vector is

$$\phi = \pi(\gamma_\infty) \phi_{F_2},$$

where we view $\gamma_\infty \in \text{GL}_2^+(\mathbb{R})$ as an element of $\text{GL}_2(\mathbb{A}_\mathbb{Q})$.

6.3. A formula for certain adèlic period integrals

Let $\phi_{F_2}: \text{GL}_2(\mathbb{A}_\mathbb{Q}) \rightarrow \mathbb{C}$ denote the adèlic lift of $F_2 := R_0 f \in \mathcal{C}_2(\Gamma_0(q))$, which is an element of the cuspidal automorphic representation $\pi = \pi_f$ of $\text{GL}_2(\mathbb{A}_\mathbb{Q})$ associated to f . Let $\Omega \in E^\times \backslash \mathbb{A}_E^1$ be the idèlic lift of χ , so that Ω is a unitary Hecke character that is unramified at every nonarchimedean place of E and has local components at the two archimedean places of E of the form $(\text{sgn}^{\kappa_\chi}, \text{sgn}^{\kappa_\chi})$ with $\kappa_\chi \in \{0, 1\}$ such that $(-1)^{\kappa_\chi} = \chi(J)$. We study the adèlic period integral

$$\mathcal{P}_\Omega(\pi(\gamma_\infty) \phi_{F_2}) := \int_{\mathbb{A}_\mathbb{Q}^\times E^\times \backslash \mathbb{A}_E^\times} \phi_{F_2}(\Psi_{\mathbb{A}_\mathbb{Q}}(x) \gamma_\infty) \Omega^{-1}(x) dx. \quad (6.7)$$

The measure dx is normalised such that $\mathbb{A}_\mathbb{Q}^\times E^\times \backslash \mathbb{A}_E^\times$ has volume $2\Lambda(1, \chi_D) = 2L(1, \chi_D)$, where

$$\Lambda(s, \chi_D) := \pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) L(s, \chi_D).$$

Lemma 6.7. *We have that*

$$|\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2})|^2 = \frac{1 - \epsilon_f \chi(J)}{q\sqrt{D}} \frac{\Gamma(\frac{3}{4} + \frac{it_f}{2})^2 \Gamma(\frac{3}{4} - \frac{it_f}{2})^2}{\Gamma(\frac{1}{2} + it_f) \Gamma(\frac{1}{2} - it_f)} \frac{L(\frac{1}{2}, f \otimes \Theta_\chi)}{L(1, \text{ad } f)}.$$

Proof. We apply [50, Theorem 4.1] with $F = \mathbb{Q}$, $E = \mathbb{Q}(\sqrt{D})$, $\varphi = \pi(\gamma_\infty)\phi_{F_2}$ (so that $\pi = \pi_f$), and Ω as above. With this choice of data, we have that $S'(\pi) = S(\Omega) = \emptyset$, $\text{Ram}(\pi) = \{p : p \mid q\}$, $\Delta_F = 1$, $\Delta_E = D$, $c(\Omega) = 1$, and $\Sigma_F^\infty = \{\infty\}$ in the notation of [50, Theorem 4.1].

There is a slight caveat; [50, Theorem 4.1] does not quite apply since although the automorphic form ϕ_{F_2} has the same local Whittaker functions $W_p \in \mathcal{W}(\pi_p, \psi_p)$ at every nonarchimedean place to that appearing in [50, Theorem 4.1] (compare Lemma 4.2 to [50, Section 2]), the local Whittaker function $W_\infty^2 \in \mathcal{W}(\pi_\infty, \psi_\infty)$ at the archimedean place, as described in Lemma 4.3, has a slightly different form than that appearing in [50, Theorem 4.1]. This issue is readily circumvented: we replace the term $C_\infty(E, \pi, \Omega)$ appearing in [50, Theorem 4.1] with its definition in [50, Section 4.2.2] in terms of local archimedean L -functions and $\tilde{J}_{\pi_\infty}(f_\infty)$, where now

$$\tilde{J}_{\pi_\infty}(f_\infty) = \frac{\left| \int_{\mathbb{R}^\times} W_\infty^2 \left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \right) \text{sgn}^{\kappa_\chi}(a) d^\times a \right|^2}{\int_{\mathbb{R}^\times} \left| W_\infty^2 \left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \right) \right|^2 d^\times a}.$$

This local distribution is just as in [50, Section 3.3] except that we have projected onto the local Whittaker function $W_\infty^2 \in \mathcal{W}(\pi_\infty, \psi_\infty)$ associated to ϕ_{F_2} instead of the local Whittaker function $W_\infty \in \mathcal{W}(\pi_\infty, \psi_\infty)$ for which the numerator is equal to the local archimedean L -function and additionally satisfying

$$W_\infty \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = (-1)^{\kappa_\Omega} W_\infty \begin{pmatrix} -a & 0 \\ 0 & 1 \end{pmatrix}$$

for all $a \in \mathbb{R}^\times$.

With this minor modification at the archimedean place, we deduce from [50, Theorem 4.1] that

$$\begin{aligned} |\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2})|^2 &= \frac{\pi}{2\sqrt{D}} \prod_{p \mid q} \frac{1}{1 - p^{-1}} \tilde{J}_{\pi_\infty}(f_\infty) \frac{L(\frac{1}{2}, \pi_f \otimes \pi_\Omega)}{L(1, \text{ad } \pi_f)} \\ &\quad \times \int_{Z(\mathbb{A}_\mathbb{Q}) \text{GL}_2(\mathbb{Q}) \backslash \text{GL}_2(\mathbb{A}_\mathbb{Q})} |\phi_{F_2}(g)|^2 dg, \end{aligned} \quad (6.8)$$

where $Z(\mathbb{A}_\mathbb{Q})$ denotes the centre of $\text{GL}_2(\mathbb{A}_\mathbb{Q})$ and the measure dg is normalised to be the Tamagawa measure multiplied by $\xi^q(2)$, where

$$\xi^q(s) := \pi^{-\frac{s}{2}} \Gamma\left(\frac{s}{2}\right) \zeta(s) \prod_{p \mid q} \left(1 - \frac{1}{p^s}\right),$$

so that

$$\text{vol}(Z(\mathbb{A}_\mathbb{Q}) \text{GL}_2(\mathbb{Q}) \backslash \text{GL}_2(\mathbb{A}_\mathbb{Q})) = 2\xi^q(2) = \frac{\pi}{3} \prod_{p \mid q} \left(1 - \frac{1}{p^2}\right), \quad (6.9)$$

since the Tamagawa number of PGL_2 is 2. From (4.23) and (4.24) with $\kappa = \kappa_f$, $\kappa' = \kappa_\chi$, $t = t_f$, and $s = \frac{1}{2}$,

$$\tilde{J}_{\pi_\infty}(f_\infty) = \begin{cases} \frac{4}{\pi(\frac{1}{4} + t_f^2)} \frac{\Gamma(\frac{3}{4} + \frac{it_f}{2})^2 \Gamma(\frac{3}{4} - \frac{it_f}{2})^2}{\Gamma(\frac{1}{2} + it_f) \Gamma(\frac{1}{2} - it_f)} & \text{if } \kappa_f \equiv \kappa_\chi + 1 \pmod{2}, \\ 0 & \text{if } \kappa_f \equiv \kappa_\chi \pmod{2}. \end{cases} \quad (6.10)$$

Furthermore,

$$\begin{aligned} & \int_{Z(\mathbb{A}_\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{Q}) \backslash \mathrm{GL}_2(\mathbb{A}_\mathbb{Q})} |\phi_{F_2}(g)|^2 dg \\ &= \frac{1}{q} \prod_{p|q} \left(1 - \frac{1}{p}\right) \int_{\Gamma_0(q) \backslash \mathbb{H}} |(R_0 f)(z)|^2 d\mu(z) \\ &= \frac{1}{q} \prod_{p|q} \left(1 - \frac{1}{p}\right) \left(\frac{1}{4} + t_f^2\right) \int_{\Gamma_0(q) \backslash \mathbb{H}} |f(z)|^2 d\mu(z), \end{aligned} \quad (6.11)$$

where the first equality holds via the strong approximation theorem, (4.16), while the second equality follows from [18, (4.38)]; to check that the normalisation of measures in the first equality is correct, we replace ϕ_{F_2} with the constant function 1 and recall (6.9) and (4.6). We obtain the result upon combining (6.8), (6.10), and (6.11) and noting that $L(s, \pi_f \otimes \pi_\Omega) = L(s, f \otimes \Theta_\chi)$ and $L(s, \mathrm{ad} \pi_f) = L(s, \mathrm{ad} f)$, and finally that

$$1 - \epsilon_f \chi(J) = \begin{cases} 2 & \text{if } \kappa_f \equiv \kappa_\chi + 1 \pmod{2}, \\ 0 & \text{if } \kappa_f \equiv \kappa_\chi \pmod{2}. \end{cases} \quad \blacksquare$$

6.4. From adèlic period integrals to cycle integrals

We relate the adèlic period integral (6.7) to a certain sum of cycle integrals over oriented geodesics in $\Gamma_0(q) \backslash \mathbb{H}$ indexed by narrow ideal classes. We first define these cycle integrals and show that they are well defined. Given $f \in \mathcal{C}_0(\Gamma_0(q))$ and a Heegner form $Q \in \mathcal{Q}_D(q)$, we consider the cycle integral

$$\int_{z_Q}^{\gamma_Q z_Q} (R_0 f)(z) \frac{dz}{\mathrm{Im}(z)}, \quad (6.12)$$

where z_Q and $\gamma_Q z_Q$ are as in (2.6) and the contour of integration is the geodesic segment between these two points.

Lemma 6.8. *For all $\gamma \in \Gamma_0(q)$, the cycle integral (6.12) is invariant under replacing Q by $\gamma \cdot Q$.*

For this reason, we may write (6.12) as $\int_{\mathcal{C}_A(q)} (R_0 f)(z) \frac{dz}{\mathrm{Im}(z)}$ without ambiguity, where $\mathcal{C}_A(q)$ denotes the oriented geodesic in $\Gamma_0(q) \backslash \mathbb{H}$ associated to a narrow ideal class A corresponding to Q as in Section 2.4, since this cycle integral is independent of the choice of Heegner form Q associated to A .

Proof of Lemma 6.8. Suppose that $Q' = \gamma \cdot Q$ for some $\gamma \in \Gamma_0(q)$. We make the change of variables $z \mapsto \gamma^{-1}z$ in (6.12). The integrand remains unchanged since

$$(R_0 f)(\gamma z) = j_\gamma(z)^2 (R_0 f)(z), \quad \frac{d}{dz}(\gamma z) = \frac{\operatorname{Im}(\gamma z)}{\operatorname{Im}(z)} j_\gamma(z)^{-2}.$$

It is easily checked that $\gamma^{-1}\gamma_Q\gamma = \gamma_{Q'}$, and so the new contour of integration is the geodesic segment from γz_Q to $\gamma_{Q'}\gamma z_Q$ on semicircle (2.4) associated to Q' . Further changes of variables by powers of $\gamma_{Q'}$ rotate the contour of integration along this semicircle while leaving the integrand intact, and so there is an appropriate power of $\gamma_{Q'}$ for which the resulting geodesic segment intersects nontrivially with the geodesic segment from $z_{Q'}$ to $\gamma_{Q'}z_{Q'}$. We then break up the integral into two parts, and for the part that does not intersect this geodesic segment, we make one last change of variables by either $\gamma_{Q'}$ or $\gamma_{Q'}^{-1}$ as appropriate; recombining, we obtain (6.12) with Q' in place of Q . ■

With this in hand, we now write the adèlic period integral $\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2})$ defined in (6.7) as a geodesic cycle integral.

Lemma 6.9. *We have that*

$$\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2}) = -\frac{i\bar{\chi}(A_\Psi)}{\sqrt{D}} \sum_{A \in \operatorname{Cl}_D^+} \bar{\chi}(A) \int_{\mathcal{C}_A(q)} (R_0 f)(z) \frac{dz}{\operatorname{Im}(z)}, \quad (6.13)$$

where $A_\Psi \in \operatorname{Cl}_D^+$ is the element of the narrow class group associated to the oriented optimal embedding Ψ . In particular, $\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2})$ is independent of the choice of oriented optimal embedding Ψ of level q within an equivalence class of embeddings modulo the action of $\Gamma_0(q)$.

A related result is proven by Popa in [57, Section 6], namely the identity

$$\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi) = \frac{i^k \bar{\chi}(A_\Psi)}{D^{\frac{k-1}{2}}} \sum_{A \in \operatorname{Cl}_D^+} \bar{\chi}(A) \int_{\mathcal{C}_A(q)} g(z) Q(z, 1)^{k-1} dz.$$

Here g is a holomorphic Hecke newform of weight k , level q , and trivial nebentypus and ϕ is the adèlic lift of g . Popa additionally proves an identity relating $|\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi)|^2$ to $L(\frac{1}{2}, g \otimes \Theta_\chi)$ akin to Lemma 6.7 via the theta correspondence [57, Theorem 5.4.1]. The proof of Lemma 6.9 given below closely follows that of Popa in [57, Section 6].

Proof of Lemma 6.9. Since Ω is the idèlic lift of a narrow class character, it is trivial on both $\hat{\mathcal{O}}_E^\times$ and $\mathbb{A}_\mathbb{Q}^\times$. Furthermore, we have the inclusion $\Psi_{\mathbb{A}_\mathbb{Q}}(\hat{\mathcal{O}}_E^\times) \subset K_0(q)$ since Ψ is an optimal embedding. As the newform f has level q , it follows that both $\pi(\gamma_\infty)\phi_{F_2}$ and Ω are well defined on the double quotient $\mathbb{A}_\mathbb{Q}^\times E^\times \backslash \mathbb{A}_E^\times / \hat{\mathcal{O}}_E^\times$. Since $\hat{\mathcal{O}}_E^\times$ has measure 1, we deduce that

$$\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2}) = \int_{\mathbb{A}_\mathbb{Q}^\times E^\times \backslash \mathbb{A}_E^\times / \hat{\mathcal{O}}_E^\times} \phi_{F_2}(\Psi_{\mathbb{A}_\mathbb{Q}}(x)\gamma_\infty)\Omega^{-1}(x) dx.$$

Via the strong approximation theorem, we have the decomposition

$$\mathbb{A}_{\mathbb{Q}}^{\times} E^{\times} \backslash \mathbb{A}_E^{\times} / \widehat{\mathcal{O}}_E^{\times} \cong \bigsqcup_{A \in \text{Cl}_D^+} A \cdot \epsilon_D^{\mathbb{Z}} \backslash E_{\infty}^1,$$

where $A = (A_v) \in \mathbb{A}_E^{\times}$ runs through a set of finite idèle representatives of the narrow class group Cl_D^+ (so that $A_v = 1$ if v is an archimedean place of E), which we freely identify with elements of the narrow class group, while $E_{\infty}^1 := \{(t, t^{-1}) \in E_{\infty} : t > 0\} \cong \mathbb{R}_{+}^{\times}$, which we view as a subgroup of $\mathbb{A}_E^{\times}$ via the embedding $(t, t^{-1}) \mapsto (t, t^{-1}, 1, 1, \dots)$, and $\epsilon_D^{\mathbb{Z}} := \{(\epsilon_D^m, \epsilon_D^{-m}) \in E_{\infty} : m \in \mathbb{Z}\}$. Thus every $x \in \mathbb{A}_{\mathbb{Q}}^{\times} E^{\times} \backslash \mathbb{A}_E^{\times} / \widehat{\mathcal{O}}_E^{\times}$ can be written as $x = A(t, t^{-1}, 1, 1, \dots)$ with $t \in [1, \epsilon_D)$. From this, we may write

$$\mathcal{P}_{\Omega}(\pi(\gamma_{\infty})\phi_{F_2}) = \frac{2}{\sqrt{D}} \sum_{A \in \text{Cl}_D^+} \bar{\chi}(A) \int_1^{\epsilon_D} \phi_{F_2}(\Psi_{\mathbb{A}_{\mathbb{Q}}}(A) \Psi_{\infty}(t, t^{-1}) \gamma_{\infty}) d^{\times} t.$$

We can check that the normalisation of measures here is correct by replacing $\pi(\gamma_{\infty})\phi_{F_2}$ with the constant function 1 and taking χ to be the trivial character, noting that $\text{vol}(\mathbb{A}_{\mathbb{Q}}^{\times} E^{\times} \backslash \mathbb{A}_E^{\times})$ is $2L(1, \chi_D)$, whereas $h_D^+ \log \epsilon_D = \sqrt{D}L(1, \chi_D)$ by the narrow class number formula.

Let g_A denote the inverse of the $\text{GL}_2^+(\mathbb{R})$ component g_{∞} of the representation $\gamma g_{\infty} k$ of the idèle $\Psi_{\mathbb{A}_{\mathbb{Q}}}(A) \in \mathbb{A}_{\mathbb{Q}}^{\times}$, as in (4.16). By definition (4.17) and (4.14) of the adèlic lift together with the fact that $\det \gamma_{\infty} > 0$, we have that

$$\begin{aligned} \mathcal{P}_{\Omega}(\pi(\gamma_{\infty})\phi_{F_2}) &= \frac{2}{\sqrt{D}} \sum_{A \in \text{Cl}_D^+} \bar{\chi}(A) \int_1^{\epsilon_D} j_{g_A^{-1} \Psi_{\infty}(t, t^{-1}) \gamma_{\infty}}(i)^{-2} \\ &\quad \times (R_0 f)(g_A^{-1} \Psi_{\infty}(t, t^{-1}) \gamma_{\infty} i) d^{\times} t. \end{aligned} \quad (6.14)$$

We make the change of variables $z = g_A^{-1} \Psi_{\infty}(t, t^{-1}) \gamma_{\infty} i = g_A^{-1} \gamma_{\infty} (it^2)$; the contour of integration in (6.14) then becomes the oriented geodesic segment from $g_A^{-1} \gamma_{\infty} i = g_A^{-1} z_Q$ to $g_A^{-1} \gamma_Q z_Q$ parametrised by $g_A^{-1} \Psi_{\infty}(t, t^{-1}) z_Q$, where z_Q and $\gamma_Q z_Q$ are as in (2.6) with $Q = Q_{\Psi} = [a, b, c]$ the Heegner form associated to the optimal embedding Ψ . We have that

$$d^{\times} t = -\frac{i}{2} j_{\gamma_{\infty}^{-1} g_A}(z)^{-2} \frac{dz}{\text{Im}(z)},$$

since $t^2 = -i \gamma_{\infty}^{-1} g_A z$ and

$$\frac{d}{dz}(gz) = \frac{\text{Im}(gz)}{\text{Im}(z)} j_g(z)^{-2}$$

for any $g \in \text{GL}_2(\mathbb{R})$, while the cocycle relation $j_{g_1 g_2}(z) = j_{g_2}(z) j_{g_1}(g_2 z)$ implies that

$$j_{g_A^{-1} \Psi_{\infty}(t, t^{-1}) \gamma_{\infty}}(i) j_{\gamma_{\infty}^{-1} g_A}(z) = j_{\gamma_{\infty}^{-1} \Psi_{\infty}(t, t^{-1}) \gamma_{\infty}}(i) = 1,$$

where the last equality follows upon recalling (4.1) and (6.6). We deduce that (6.14) is equal to

$$-\frac{i}{\sqrt{D}} \sum_{A \in \text{Cl}_D^+} \bar{\chi}(A) \int_{g_A^{-1} z_Q}^{g_A^{-1} \gamma_Q z_Q} (R_0 f)(z) \frac{dz}{\text{Im}(z)}.$$

It is shown in [57, Theorem 6.2.2 (i)] that as A runs through the narrow class group Cl_D^+ , $g_A^{-1}\Psi g_A$ runs through a set of representatives of equivalence classes of oriented optimal embeddings of level q modulo the action of $\Gamma_0(q)$. Furthermore, letting Q' denote the Heegner form of level q associated to the oriented optimal embedding $g_A^{-1}\Psi g_A$, we have that the contour of integration from $g_A^{-1}z_Q$ and $g_A^{-1}\gamma_Q z_Q = \gamma_{Q'} g_A^{-1}z_Q$ is a geodesic segment of length $2 \log \epsilon_D$ of semicircle (2.4) associated to Q' , and hence

$$\int_{g_A^{-1}z_Q}^{g_A^{-1}\gamma_Q z_Q} (R_0 f)(z) \frac{dz}{\text{Im}(z)} = \int_{\mathcal{C}_{A'}(q)} (R_0 f)(z) \frac{dz}{\text{Im}(z)}$$

by the proof of Lemma 6.8, where $A' \in \text{Cl}_D^+$ is the element of the narrow class group associated to Q' . It remains to note that by [57, Theorem 6.2.2 (ii)], the oriented optimal embedding $g_A^{-1}\Psi g_A$ is associated to $AA_\Psi \in \text{Cl}_D^+$, where A_Ψ is the element of Cl_D^+ corresponding to Ψ . ■

6.5. From cycle integrals to Weyl sums for newforms

Having shown in Lemma 6.9 that the adèlic period integral $\mathcal{P}_\Omega(\pi(\gamma_\infty)\phi_{F_2})$ defined in (6.7) may be reexpressed as a geodesic cycle integral defined in (6.13), we now show that this integral of a Maaß cusp form over a closed geodesic $\mathcal{C}_A(q)$ is related to an integral over the hyperbolic orbifold $\Gamma_A(q) \backslash \mathcal{N}_A(q)$ plus a certain *topological contribution*. Furthermore, we show that a similar relation holds for Eisenstein series, with a much simpler proof. It is at this point that we specialise to the setting of q being either equal to 1 or equal to an odd prime.

Lemma 6.10 (Cf. [19, Lemmata 1 and 2]). *Let $f \in \mathcal{C}_0(\Gamma_0(q))$ be a Maaß cusp form of weight 0, level q , where either $q = 1$ or q is an odd prime that splits in $E = \mathbb{Q}(\sqrt{D})$, and Laplacian eigenvalue $\lambda_f = \frac{1}{4} + t_f^2$. Then for each $A \in \text{Cl}_D^+$, we have that*

$$\begin{aligned} \int_{\mathcal{F}_A(q)} f(z) d\mu(z) &= \frac{1}{\frac{1}{4} + t_f^2} \left(\int_{\mathcal{C}_A(q)} (R_0 f)(z) \frac{dz}{\text{Im}(z)} \right. \\ &\quad \left. + \sum_{j=1}^{2g} \int_{\mathcal{C}_j} (R_0 f)(z) \frac{dz}{\text{Im}(z)} \langle [\mathcal{C}_A(q)], \omega_j \rangle_{\text{cap}} \right). \end{aligned}$$

Similarly, let $E(z, \frac{1}{2} + it)$ be the Eisenstein series on $\Gamma \backslash \mathbb{H}$ with Laplacian eigenvalue $\frac{1}{4} + t^2$, where $t \in \mathbb{R}$. Then for $\ell \mid q$, we have that

$$\begin{aligned} \int_{\mathcal{F}_A(q)} E\left(\ell z, \frac{1}{2} + it\right) d\mu(z) &= \frac{1}{\frac{1}{4} + t^2} \left(\int_{\mathcal{C}_A(q)} (R_0 E)\left(\ell z, \frac{1}{2} + it\right) \frac{dz}{\text{Im}(z)} \right. \\ &\quad \left. + \sum_{j=1}^{2g} \int_{\mathcal{C}_j}^* (R_0 E)\left(\ell z, \frac{1}{2} + it\right) \frac{dz}{\text{Im}(z)} \langle [\mathcal{C}_A(q)], \omega_j \rangle_{\text{cap}} \right), \end{aligned}$$

where $\int_{\mathcal{C}_j}^*$ denotes the regularised integral defined in (5.12).

Proof. We show this for q prime; the case of $q = 1$ follows in the same way except that $g = 0$, so that the sum over j is empty. Let $\mathcal{F}_A(q) \subset \mathbb{H}$ be the fundamental domain for $\Gamma_A(q) \backslash \mathcal{N}_A(q)$ defined in (3.7). For each vertex v of $\mathcal{F}_A(q)$ and for $Y > 1$, let $\mathcal{F}_v(Y)$ denote the cuspidal zone defined in (5.8). Let ϕ be either a Hecke–Maaß cusp form or an Eisenstein series as above and denote by λ_ϕ the associated Laplace eigenvalue. By Stokes’ theorem, we have that for all Y sufficiently large,

$$\begin{aligned} & \int_{\mathcal{F}_A(q) \setminus \bigcup_v \mathcal{F}_v(Y)} \frac{\partial^2}{\partial \bar{z} \partial z} \phi(z) dz d\bar{z} \\ &= - \int_{\partial \mathcal{F}_A(q) \setminus \bigcup_v \mathcal{F}_v(Y)} \frac{\partial}{\partial \bar{z}} \phi(z) dz \\ &\quad - \sum_{i=1}^{n+2} n_i(A, q) \int_{I_i(q)} \frac{\partial}{\partial z} \Big|_{z=x+iY} \phi \left(\begin{pmatrix} a_i & a_{i-1} \\ b_i & b_{i-1} \end{pmatrix} z \right) dx. \end{aligned} \quad (6.15)$$

Here $n_i(A, q)$ is the multiplicity in $\mathcal{F}_A(q)$ of the cuspidal zone $\mathcal{F}_{\frac{a_i}{b_i}}(Y)$ around the vertex $\frac{a_i}{b_i}$ (which is some number independent of Y), while $I_i(q) \subset \mathbb{R}$ is the horocycle interval at height Y around the vertex $\frac{a_i}{b_i}$ of $\mathcal{P}(q)$ defined by

$$\begin{pmatrix} a_i & a_{i-1} \\ b_i & b_{i-1} \end{pmatrix}^{-1} \mathcal{P}(q) \cap \{x + iY : x \in \mathbb{R}\} = \{x + iY : x \in I_i(q)\}.$$

Since

$$\frac{\partial^2}{\partial \bar{z} \partial z} = \frac{1}{4} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right), \quad dz d\bar{z} = -2iy^2 d\mu(z),$$

the left-hand side of (6.15) is

$$\frac{i\lambda_\phi}{2} \int_{\mathcal{F}_A(q) \setminus \bigcup_v \mathcal{F}_v(Y)} \phi(z) d\mu(z).$$

As Y tends to infinity, this converges to

$$\frac{i\lambda_\phi}{2} \int_{\mathcal{F}_A(q)} \phi(z) d\mu(z).$$

The first term on the right-hand side of (6.15) is equal to

$$\begin{aligned} & - \int_{\mathcal{C}_A(q)} \frac{\partial}{\partial \bar{z}} \phi(z) dz - \sum_{j=1}^{\frac{N}{2}} m_j(A, q) \int_{\mathcal{C}_j \setminus \bigcup_v \mathcal{F}_v(Y)} \frac{\partial}{\partial \bar{z}} \phi(z) dz \\ &= \frac{i}{2} \int_{\mathcal{C}_A(q)} (R_0 \phi)(z) \frac{dz}{\operatorname{Im}(z)} + \frac{i}{2} \sum_{j=1}^{\frac{N}{2}} m_j(A, q) \int_{\mathcal{C}_j \setminus \bigcup_v \mathcal{F}_v(Y)} (R_0 \phi)(z) \frac{dz}{\operatorname{Im}(z)}, \end{aligned}$$

where $m_j(A, q)$ denotes the multiplicity of the side $\mathcal{C}_j$ of $\mathcal{P}(q)$ in $\Gamma_A(q) \backslash \mathcal{N}_A(q)$ as defined in Section 5.2. The second term on the right-hand side of (6.15) converges to zero as Y tends to infinity; for ϕ cuspidal, this follows from the rapid decay at each cusp $\mathfrak{b}$, while for Eisenstein series, this follows from Lemma 4.1. Thus upon taking the limit as Y tends to infinity, we obtain the desired identity by Lemmata 5.2 and 5.3. ■

6.6. Proofs of Propositions 6.1 and 6.5

We are finally in a position to prove Proposition 6.1.

Proof of Proposition 6.1. Lemmata 6.9 and 6.10 imply that the Weyl sum $W_{\chi,f}$ defined in (6.1) satisfies the identity

$$W_{\chi,f} = -\frac{i\bar{\chi}(A\Psi)\sqrt{D}}{\frac{1}{4} + t_f^2} \mathcal{P}_{\Omega}(\pi(\gamma_{\infty})\phi_{F_2}) \\ + \frac{1}{\frac{1}{4} + t_f^2} \sum_{j=1}^{2g} \int_{\mathcal{C}_j} (R_0 f)(z) \frac{dz}{\text{Im}(z)} \sum_{A \in \text{Cl}_D^+} \chi(A) \langle [\mathcal{C}_A(q)], \omega_j \rangle_{\text{cap}}.$$

Expressing the class $[\mathcal{C}_A(q)] \in H_1(X_0(q), \mathbb{R})$ in terms of the Hecke basis defined in Section 5.1.1 and recalling definition (5.3) gives

$$\sum_{A \in \text{Cl}_D^+} \chi(A) \langle [\mathcal{C}_A(q)], \omega_j \rangle_{\text{cap}} \\ = \frac{1}{2} \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} \sum_{\pm} i^{\frac{1\mp 1}{2}} \langle v_h^{\pm}, \omega_j \rangle_{\text{cap}} \sum_{A \in \text{Cl}_D^+} \chi(A) \left(\int_{\mathcal{C}_A(q)} h(z) dz \pm \int_{\mathcal{C}_A(q)} \overline{h(z)} d\bar{z} \right).$$

Combining this, the result follows from Lemma 6.7 as well as the explicit form of Waldspurger's formula for $h \in \mathcal{B}_2^{\text{hol}}(q)$ due to Popa [58, Theorem 6.3.1], namely

$$\left| \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{C}_A(q)} h(z) dz \right|^2 = \frac{\sqrt{D}}{4\pi^2} L\left(\frac{1}{2}, h \otimes \Theta_{\bar{\chi}}\right). \quad \blacksquare$$

The proof of Proposition 6.5 is a little simpler, since we can circumvent the adèlic formulation of this Weyl sum.

Proof of Proposition 6.5. From [19, (7.3)], we have that

$$\sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{C}_A} (R_0 E)(z, s) \frac{dz}{\text{Im}(z)} = (1 - \chi(J)) D^{\frac{s}{2}} \frac{\Gamma(\frac{s+1}{2})^2}{\Gamma(s)} \frac{L(s, \chi)}{\zeta(2s)}$$

for $\text{Re}(s) > 1$. Via analytic continuation, this identity extends to $s = \frac{1}{2} + it$. The result then follows via Lemma 6.10. $\blacksquare$

6.7. Weyl sums for oldforms

Finally, we show that Weyl sums for oldforms are essentially equal to Weyl sums for the associated newform of lower level. We define for a Hecke–Maaß newform f of weight 0 and level 1 and a narrow class character χ ,

$$W_{\chi,f}^{\ell,q} := \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A(q)} f(\ell z) d\mu(z) \quad (6.16)$$

and similarly for the Eisenstein series $E(\cdot, \frac{1}{2} + it)$ for $\Gamma \backslash \mathbb{H}$,

$$W_{\chi, t}^{\ell, q} := \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A(q)} E\left(\ell z, \frac{1}{2} + it\right) d\mu(z). \quad (6.17)$$

We also write $W_{\chi, f}^{1, q} = W_{\chi, f}$ for f a Hecke–Maaß newform of level q .

Lemma 6.11. *Let q be an odd prime and let g denote the genus of $X_0(q)$. Let $\ell \mid q$ and let $A_{\mathfrak{l}} \in \text{Cl}_D^+$ denote the narrow ideal class containing the oriented ideal*

$$\left[\mathfrak{l}; \ell, \frac{r - \sqrt{D}}{2}\right]. \quad (6.18)$$

Then for a Hecke–Maaß newform f of weight 0 and level 1 and a narrow class character χ ,

$$\begin{aligned} W_{\chi, f}^{\ell, q} &= \chi(A_{\mathfrak{l}}) \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A} f(z) d\mu(z) \\ &\quad + \delta_{\ell, q} \sum_{j=1}^{2g} \int_{\mathcal{C}_j} (R_0 f)(\ell z) \frac{dz}{\text{Im}(z)} \sum_{A \in \text{Cl}_D^+} \chi(A) \langle [\mathcal{C}_A(q)], \omega_j \rangle_{\text{cap}}. \end{aligned} \quad (6.19)$$

Similarly, for the Eisenstein series $E(z, \frac{1}{2} + it)$ for $\Gamma \backslash \mathbb{H}$,

$$\begin{aligned} W_{\chi, t}^{\ell, q} &= \chi(A_{\mathfrak{l}}) \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A} E\left(z, \frac{1}{2} + it\right) d\mu(z) \\ &\quad + \delta_{\ell, q} \sum_{j=1}^{2g} \int_{\mathcal{C}_j}^* (R_0 E)\left(\ell z, \frac{1}{2} + it\right) \frac{dz}{\text{Im}(z)} \sum_{A \in \text{Cl}_D^+} \chi(A) \langle [C_A(q)], \omega_j \rangle_{\text{cap}}. \end{aligned} \quad (6.20)$$

Note that although r is only defined modulo $2q$, the oriented ideal (6.18) is well defined since it contains ℓ and hence also q .

Proof of Lemma 6.11. We give the proof for (6.19); the same method yields (6.20). Let $Q = [a, b, c] \in \mathcal{Q}_D(q)$ be a Heegner form associated to the oriented closed geodesic $\mathcal{C}_A(q)$. Since $\mathcal{C}_A(q)$ is the reduction modulo $\Gamma_0(q)$ of the geodesic segment from z_Q to $\gamma_Q z_Q$, as in (2.6), Lemma 6.10 together with the change of variables

$$z \mapsto g_{\ell}^{-1} z, \quad \text{where } g_{\ell} := \begin{pmatrix} \sqrt{\ell} & 0 \\ 0 & \frac{1}{\sqrt{\ell}} \end{pmatrix} \in \text{SL}_2(\mathbb{R}),$$

implies that

$$\begin{aligned} \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{\mathcal{F}_A(q)} f(\ell z) d\mu(z) &= \frac{1}{\frac{1}{4} + t_f^2} \sum_{A \in \text{Cl}_D^+} \chi(A) \int_{g_{\ell} z_Q}^{g_{\ell} \gamma_Q z_Q} (R_0 f)(z) \frac{dz}{\text{Im}(z)} \\ &\quad + \sum_{j=1}^{2g} \int_{\mathcal{C}_j} (R_0 f)(\ell z) \frac{dz}{\text{Im}(z)} \sum_{A \in \text{Cl}_D^+} \chi(A) \langle [\mathcal{C}_A(q)], \omega_j \rangle_{\text{cap}}. \end{aligned}$$

The Heegner form associated to the geodesic segment from $g_\ell z_Q$ to $g_\ell \gamma_Q z_Q$, as in Section 2.4, is $Q_\ell := [\frac{a}{\ell}, b, c\ell] \in \mathcal{Q}_D(\frac{q}{\ell})$. We claim that the narrow ideal class associated to Q_ℓ is $A_\Gamma^{-1}A$, from which (6.19) follows. From (2.2), this is implied by the fact that

$$\left(\mathbb{Z}\frac{a}{\ell} + \mathbb{Z}\frac{b - \sqrt{D}}{2}\right)\left(\mathbb{Z}\ell + \mathbb{Z}\frac{r - \sqrt{D}}{2}\right) = \mathbb{Z}a + \mathbb{Z}\frac{b - \sqrt{D}}{2}$$

if $a > 0$, and similarly

$$\left(\mathbb{Z}\left(-\frac{a}{\ell}\sqrt{D}\right) + \mathbb{Z}\frac{D - b\sqrt{D}}{2}\right)\left(\mathbb{Z}\ell + \mathbb{Z}\frac{r - \sqrt{D}}{2}\right) = \mathbb{Z}(-a\sqrt{D}) + \mathbb{Z}\frac{D - b\sqrt{D}}{2}$$

if $a < 0$. Finally, if $\ell = 1$, then the fact that $\mathcal{C}_j$ is fixed by an order 2 matrix in Γ implies that

$$\int_{\mathcal{C}_j} (R_0 f)(z) \frac{dz}{\text{Im}(z)} = 0. \quad \blacksquare$$

7. Sparse equidistribution in the level aspect

7.1. Explicit orthonormal bases

In order to prove Theorem 1.1, we require the spectral decomposition of $L^2(\Gamma_0(q)\backslash\mathbb{H})$. This decomposition involves three parts: the residual spectrum, consisting of the constant function; the cuspidal spectrum, spanned by Hecke–Maaß cusp forms; and the continuous spectrum, spanned by incomplete Eisenstein series.

For the cuspidal spectrum, we may choose an orthonormal basis of this subspace consisting of certain linear combinations of oldforms and newforms.

Lemma 7.1 ([38, Lemma 3.1]). *Let q be squarefree. An orthonormal basis of $\mathcal{C}_0(\Gamma_0(q))$ with respect to the inner product*

$$\langle f_1, f_2 \rangle_q := \int_{\Gamma_0(q)\backslash\mathbb{H}} f_1(z) \overline{f_2(z)} d\mu(z)$$

is given by

$$\{f_\ell \in \mathcal{C}_0(\Gamma_0(q)) : f \in \mathcal{B}_0^*(\Gamma_0(q_1)), q_1 q_2 = q, \ell \mid q_2\},$$

where $\mathcal{B}_0^*(\Gamma_0(q_1))$ denotes an orthonormal basis of Hecke–Maaß newforms f of weight 0 and level q_1 with respect to the inner product $\langle \cdot, \cdot \rangle_{q_1}$, while f_ℓ is associated to $f \in \mathcal{B}_0^*(\Gamma_0(q_1))$ by

$$f_\ell(z) := \left(L_\ell(1, \text{ad } f) \frac{\varphi(\ell)}{\ell v(q_2)}\right)^{\frac{1}{2}} \sum_{v|w=\ell} \frac{v(v)}{v} \frac{\mu(w) \lambda_f(w)}{\sqrt{w}} f(vz).$$

Here for $\ell \nmid q_1$,

$$L_\ell(s, \text{ad } f) := \prod_{p|\ell} \frac{1}{1 - \lambda_f(p^2)p^{-s} + \lambda_f(p^2)p^{-2s} - p^{-3s}}.$$

A similar construction is valid for the continuous spectrum by choosing an orthonormal set of Eisenstein series.

Lemma 7.2 ([65, Section 8.4]). *Let q be squarefree. An orthonormal basis of the vector space of Eisenstein series of weight 0 and level q is given by*

$$\left\{ E_\ell\left(\cdot, \frac{1}{2} + it\right) : \ell \mid q \right\},$$

where

$$E_\ell\left(z, \frac{1}{2} + it\right) := \left(\frac{\zeta_\ell(1 + 2it)\zeta_\ell(1 - 2it)}{v(q)} \right)^{\frac{1}{2}} \sum_{vw=\ell} \frac{v(v)}{v} \frac{\mu(w)\lambda(w, t)}{\sqrt{w}} E\left(vz, \frac{1}{2} + it\right).$$

Here

$$\zeta_\ell(s) := \prod_{p \mid \ell} \frac{1}{1 - p^{-s}}.$$

7.2. Averaged bounds for regularised integrals

We require bounds in terms of q for the integrals of both Maaß cusp forms and Eisenstein series over the sides of $\mathcal{P}(q)$. In order to bound the regularised integrals that show up in the Weyl sums, we start by showing a useful lemma: all the hyperbolic sides of $\mathcal{P}(q)$ can be brought to a standard form under the action of the Atkin–Lehner operator

$$W_q = \begin{pmatrix} 0 & -\frac{1}{\sqrt{q}} \\ \sqrt{q} & 0 \end{pmatrix},$$

as in (2.7), and $\Gamma_0(q)$.

Lemma 7.3. *Let $\mathcal{L}$ be a hyperbolic side of $\mathcal{P}(q)$ with endpoints $\frac{a_i}{b_i}$ and $\frac{a_{i+1}}{b_{i+1}}$. There exists a unique element $\gamma_{\mathcal{L}} \in \Gamma_0(q)$ such that*

$$\gamma_{\mathcal{L}} W_q \mathcal{L} = \left\{ z \in \mathbb{H} : \operatorname{Re}(z) = \frac{v}{q} \right\}$$

for $v \in \{1, \dots, q-1\}$ satisfying $vb_{i+1} \equiv b_i \pmod{q}$. Moreover, v uniquely determines $\mathcal{L}$.

Proof. Let $v \in \{1, \dots, q-1\}$ be such that $vb_{i+1} \equiv b_i \pmod{q}$ (note that $(b_{i+1}, q) = 1$) and define

$$\gamma_{\mathcal{L}} := \begin{pmatrix} a_{i+1}v - a_i & \frac{b_{i+1}v - b_i}{q} \\ a_{i+1}q & b_{i+1} \end{pmatrix} \in \Gamma_0(q).$$

Then the fact that $a_{i+1}b_i - a_ib_{i+1} = 1$ implies that

$$\gamma_{\mathcal{L}} W_q \frac{a_{i+1}}{b_{i+1}} = \infty, \quad \gamma_{\mathcal{L}} W_q \frac{a_i}{b_i} = \frac{v}{q},$$

as desired.

To show uniqueness, we suppose that $\gamma'_{\mathcal{L}} W_q$ also maps $\frac{a_{i+1}}{b_{i+1}}$ to ∞ and $\frac{a_i}{b_i}$ to $\frac{v}{q}$. Then $\gamma'_{\mathcal{L}} \gamma_{\mathcal{L}}^{-1} \infty = \infty$, so that $\gamma'_{\mathcal{L}} \gamma_{\mathcal{L}}^{-1} \in \Gamma_{\infty}$, and hence there exists some $n \in \mathbb{Z}$ such that $\gamma'_{\mathcal{L}} \gamma_{\mathcal{L}}^{-1} = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$. Since $\gamma'_{\mathcal{L}} \gamma_{\mathcal{L}}^{-1} \frac{v}{q} \in (0, 1)$, we must have that $n = 0$, so that $\gamma'_{\mathcal{L}} = \gamma_{\mathcal{L}}$.

Finally, if we have two hyperbolic sides $\mathcal{L}$ and $\mathcal{L}'$ such that

$$\gamma_{\mathcal{L}} W_q \mathcal{L} = \gamma'_{\mathcal{L}} W_q \mathcal{L}',$$

then by applying W_q^{-1} and using that W_q normalises $\Gamma_0(q)$, we conclude that $\mathcal{L}$ and $\mathcal{L}'$ are $\Gamma_0(q)$ -equivalent. This implies that $b_i b_{i^*+1} \equiv b_{i+1} b_i^* \pmod{q}$. But we also have $b_i b_i^* \equiv -b_{i+1} b_{i^*+1} \pmod{q}$, which implies $b_i^2 + b_{i+1}^2 \equiv 0 \pmod{q}$, and thus $\mathcal{L} = \mathcal{L}'$ is an even side, contrary to the assumption. ■

We start by bounding the cuspidal case.

Lemma 7.4. *Let q be an odd prime and let $\mathcal{L}$ be a hyperbolic side of $\mathcal{P}(q)$. Then for $q_1 q_2 = q$, we have that*

$$\frac{1}{q_2} \sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ |t_f| \leq T}} \left| \int_{\mathcal{L}} (R_0 f)(q_2 z) \frac{dz}{\text{Im}(z)} \right|^2 \ll_{\varepsilon} q^{\varepsilon} T^{3+\varepsilon}. \quad (7.1)$$

Proof. We make the change of variables $z \mapsto W_q^{-1} \gamma_{\mathcal{L}}^{-1} z$, where $\gamma_{\mathcal{L}} \in \Gamma_0(q)$ is as in Lemma 7.3. As $g_{q_2} W_q^{-1} W_{q_1} \in \Gamma_0(q_1)$, we have via Lemma 7.3 that

$$\int_{\mathcal{L}} (R_0 f)(q_2 z) \frac{dz}{\text{Im}(z)} = \eta_f(q_1) \int_{\frac{v}{q}}^{i\infty} (R_0 f)(z) \frac{dz}{\text{Im}(z)},$$

where $\eta_f(q_1) \in \{1, -1\}$ is the Atkin–Lehner eigenvalue of f under the action of W_{q_1} . So upon defining

$$\Lambda\left(s, f, \frac{v}{q}\right) := q^{s-\frac{1}{2}} \int_{\frac{v}{q}}^{i\infty} (R_0 f)(z) \text{Im}(z)^{s-\frac{1}{2}} \frac{dz}{\text{Im}(z)}$$

for $s \in \mathbb{C}$ and recalling Lemma 7.1, we deduce that the left-hand side of (7.1) is equal to

$$\frac{1}{q_2} \sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ |t_f| \leq T}} \left| \Lambda\left(\frac{1}{2}, f, \frac{v}{q}\right) \right|^2. \quad (7.2)$$

Let $\bar{v} \in \{1, \dots, q-1\}$ be such that $v\bar{v} \equiv 1 \pmod{q}$. Then

$$(R_0 f)\left(\frac{v}{q} + \frac{iy}{q}\right) = -(R_0 f)\left(-\frac{\bar{v}}{q} + \frac{i}{qy}\right)$$

via the automorphy of $R_0 f$, namely $(R_0 f)(\gamma z) = j_{\gamma}(z)^2 (R_0 f)(z)$ with

$$\gamma = \begin{pmatrix} v & \frac{v\bar{v}-1}{q} \\ q & \bar{v} \end{pmatrix} \in \Gamma_0(q) \quad \text{and} \quad z = -\frac{\bar{v}}{q} + \frac{i}{qy}.$$

This yields the functional equation

$$\begin{aligned}\Lambda\left(s, f, \frac{v}{q}\right) &= i \int_1^\infty \left((R_0 f) \left(\frac{v}{q} + \frac{iy}{q} \right) y^{s-\frac{1}{2}} - (R_0 f) \left(-\frac{\bar{v}}{q} + \frac{iy}{q} \right) y^{\frac{1}{2}-s} \right) \frac{dy}{y} \\ &= -\Lambda\left(1-s, f, -\frac{\bar{v}}{q}\right),\end{aligned}\quad (7.3)$$

noting that the integral represents an entire function. Moreover, for $\operatorname{Re}(s) > 1$, we may insert the Fourier expansion (4.8) for $R_0 f$ in order to see that

$$\begin{aligned}\Lambda\left(s, f, \frac{v}{q}\right) &= -i\rho_f(1)q^{s-\frac{1}{2}} \sum_{n=1}^\infty \frac{\lambda_f(n)}{n^s} e\left(\frac{nv}{q}\right) \int_0^\infty W_{1, it_f}(4\pi y) y^{s-\frac{1}{2}} \frac{dy}{y} \\ &\quad + i\epsilon_f \rho_f(1)q^{s-\frac{1}{2}} \sum_{n=1}^\infty \frac{\lambda_f(n)}{n^s} e\left(-\frac{nv}{q}\right) \\ &\quad \times \int_0^\infty \left(\frac{1}{4} + t_f^2\right) W_{-1, it_f}(4\pi y) y^{s-\frac{1}{2}} \frac{dy}{y}.\end{aligned}\quad (7.4)$$

From (4.22) and (4.23), the first integral above is

$$\pi^{-s} \Gamma\left(\frac{s+1+it_f}{2}\right) \Gamma\left(\frac{s+1-it_f}{2}\right) + \frac{1}{2}\left(s - \frac{1}{2}\right) \pi^{-s} \Gamma\left(\frac{s+it_f}{2}\right) \Gamma\left(\frac{s-it_f}{2}\right), \quad (7.5)$$

while the second is

$$\pi^{-s} \Gamma\left(\frac{s+1+it_f}{2}\right) \Gamma\left(\frac{s+1-it_f}{2}\right) - \frac{1}{2}\left(s - \frac{1}{2}\right) \pi^{-s} \Gamma\left(\frac{s+it_f}{2}\right) \Gamma\left(\frac{s-it_f}{2}\right). \quad (7.6)$$

With this in hand, we derive the approximate functional equation for $\Lambda(\frac{1}{2}, f, \frac{v}{q})$, namely the identity

$$\Lambda\left(\frac{1}{2}, f, \frac{v}{q}\right) = \frac{i\rho_f(1)}{\sqrt{\pi}} \sum_{\pm} \mp \epsilon_f^{\frac{1\mp 1}{2}} \sum_{n=1}^\infty \frac{\lambda_f(n)}{\sqrt{n}} \left(e\left(\pm \frac{nv}{q}\right) + e\left(\mp \frac{n\bar{v}}{q}\right) \right) V_{\pm}\left(\frac{n}{q}, t_f\right), \quad (7.7)$$

where for $x > 0$, $t \in \mathbb{R} \cup i[-\frac{7}{64}, \frac{7}{64}]$, and $\sigma > 0$,

$$\begin{aligned}V_{\pm}(x, t) &:= \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} x^{-s} e^{s^2} \left(\Gamma\left(\frac{s}{2} + \frac{3}{4} + \frac{it}{2}\right) \Gamma\left(\frac{s}{2} + \frac{3}{4} - \frac{it}{2}\right) \right. \\ &\quad \left. \pm \frac{s}{2} \Gamma\left(\frac{s}{2} + \frac{1}{4} + \frac{it}{2}\right) \Gamma\left(\frac{s}{2} + \frac{1}{4} - \frac{it}{2}\right) \right) \frac{ds}{s}.\end{aligned}\quad (7.8)$$

To see this, consider

$$\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \Lambda\left(\frac{1}{2} + s, f, \frac{v}{q}\right) e^{s^2} \frac{ds}{s} - \frac{1}{2\pi i} \int_{-\sigma-i\infty}^{-\sigma+i\infty} \Lambda\left(\frac{1}{2} + s, f, \frac{v}{q}\right) e^{s^2} \frac{ds}{s} \quad (7.9)$$

with $\sigma > \frac{1}{2}$. This is equal to the left-hand side of (7.7) simply by shifting the first contour of integration from $\operatorname{Re}(s) = \sigma$ to $\operatorname{Re}(s) = -\sigma$, which picks up a residue at $s = 0$ equal to $\Lambda(\frac{1}{2}, f, \frac{v}{q})$. On the other hand, this is equal to the right-hand side of (7.7) by first using

the functional equation (7.3) for the second integral and making the change of variables $s \mapsto -s$, and then inserting expression (7.4) for $\Lambda(\frac{1}{2} + s, f, \frac{v}{q})$ and $\Lambda(\frac{1}{2} + s, f, -\frac{\bar{v}}{q})$, interchanging the order of summation and integration, and recalling (7.5) and (7.6).

Next, we note that by (4.5) and Stirling's formula, we have that for any $A > 0$,

$$\rho_f(1)V_{\pm}(x, t_f) \ll_{\varepsilon} \begin{cases} q_1^{-\frac{1}{2}+\varepsilon} (1 + |t_f|)^{\frac{1}{2}+\varepsilon} & \text{for } x \leq 1 + |t_f|, \\ q_1^{-\frac{1}{2}+\varepsilon} (1 + |t_f|)^{A+\frac{1}{2}+\varepsilon} x^{-A} & \text{for } x \geq 1 + |t_f|. \end{cases}$$

Here we have used the fact that

$$L(1, \text{ad } f) \gg_{\varepsilon} (q_1(1 + |t_f|))^{-\varepsilon};$$

additionally, for $x \leq 1 + |t_f|$, we have shifted the contour to $\text{Re}(s) = -\frac{25}{64} + \varepsilon$, while we have shifted the contour to $\text{Re}(s) = A$ for $x \geq 1 + |t_f|$. This bound shows that the portion of the sum over $n \in \mathbb{N}$ in (7.7) for which $n \geq (q(1 + |t_f|))^{1+\varepsilon}$ is negligibly small.

The result now follows by inserting the approximate functional equation (7.7) into (7.2), dividing the sum over $n \in \mathbb{N}$ in (7.7) into dyadic ranges, and applying the spectral large sieve, namely the well-known bound

$$\sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ |t_f| \leq T}} \left| \sum_{n \leq N} a_n \lambda_f(n) \right|^2 \ll_{\varepsilon} (q_1 T N)^{\varepsilon} (q_1 T^2 + N) \sum_{n \leq N} |a_n|^2. \quad \blacksquare$$

An analogous result holds for Eisenstein series.

Lemma 7.5. *Let q be an odd prime and let $\mathcal{L}$ be a hyperbolic side of $\mathcal{P}(q)$. Then we have that*

$$\frac{1}{q} \int_{-T}^T \left| \int_{\mathcal{L}}^* (R_0 E) \left(qz, \frac{1}{2} + it \right) \frac{dz}{\text{Im}(z)} \right|^2 dt \ll_{\varepsilon} q^{\varepsilon} T^{3+\varepsilon}.$$

Proof. For $\text{Re}(s) > 1$, we define

$$\Lambda \left(s, E_t, \frac{v}{q} \right) := q^{s-\frac{1}{2}} \int_{\frac{v}{q}}^{i\infty} (\widetilde{R_0 E}) \left(z, \frac{1}{2} + it \right) \text{Im}(z)^{s-\frac{1}{2}} \frac{dz}{\text{Im}(z)},$$

where $(\widetilde{R_0 E})(z, \frac{1}{2} + it)$ is as in (4.12). The automorphy of $(R_0 E)(z, \frac{1}{2} + it)$ implies the functional equation

$$\begin{aligned} \Lambda \left(s, E_t, \frac{v}{q} \right) &= i \int_1^{\infty} \left((\widetilde{R_0 E}) \left(\frac{v}{q} + \frac{iy}{q}, \frac{1}{2} + it \right) y^{s-\frac{1}{2}} \right. \\ &\quad \left. - (\widetilde{R_0 E}) \left(-\frac{\bar{v}}{q} + \frac{iy}{q}, \frac{1}{2} + it \right) y^{\frac{1}{2}-s} \right) \frac{dy}{y} \\ &\quad + i \sum_{\pm} \left(\frac{1}{2} \pm it \right) \frac{\xi(1 \pm 2it)}{\xi(1 + 2it)} q^{-\frac{1}{2} \mp it} \left(\frac{1}{s \pm it} - \frac{1}{1 - s \pm it} \right) \\ &= -\Lambda \left(1 - s, E_t, -\frac{\bar{v}}{q} \right). \end{aligned}$$

Note that the first integral is entire, while the second term, which arises from the constant terms in the Fourier expansion of $(R_0 E)(z, \frac{1}{2} + it)$, is meromorphic with simple poles at $s = 1 \pm it$ and $s = \mp it$ and a zero at $s = \frac{1}{2}$. Furthermore, for $\operatorname{Re}(s) > 1$, we have that

$$\begin{aligned} \Lambda\left(s, E_t, \frac{v}{q}\right) &= -i \frac{q^{s-\frac{1}{2}}}{\xi(1+2it)} \sum_{n=1}^{\infty} \frac{\lambda(n, t)}{n^s} e\left(\frac{nv}{q}\right) \int_0^{\infty} W_{1, it}(4\pi y) y^{s-\frac{1}{2}} \frac{dy}{y} \\ &\quad + i \frac{q^{s-\frac{1}{2}}}{\xi(1+2it)} \sum_{n=1}^{\infty} \frac{\lambda(n, t)}{n^s} e\left(-\frac{nv}{q}\right) \\ &\quad \times \int_0^{\infty} \left(\frac{1}{4} + t^2\right) W_{-1, it}(4\pi y) y^{s-\frac{1}{2}} \frac{dy}{y}; \end{aligned}$$

these integrals are explicitly calculated in (7.5) and (7.6). Finally, the approximate functional equation in this setting reads

$$\Lambda\left(\frac{1}{2}, E_t, \frac{v}{q}\right) = \frac{i}{\sqrt{\pi}\xi(1+2it)} \sum_{\pm} \mp \sum_{n=1}^{\infty} \frac{\lambda(n, t)}{\sqrt{n}} \left(e\left(\pm \frac{nv}{q}\right) + e\left(\mp \frac{n\bar{v}}{q}\right) \right) V_{\pm}\left(\frac{n}{q}, t\right),$$

where $V_{\pm}(x, t)$ is as in (7.8). This follows by the same method as for Hecke–Maaß newforms, namely considering integral (7.9) with f replaced by E_t ; although there are additional poles of the integrand at $s = \frac{1}{2} \pm it$ and $s = -\frac{1}{2} \mp it$, these ultimately cancel out.

The same approach as for the Hecke–Maaß form case shows that the left-hand side of (7.1) is equal to

$$\frac{1}{q} \int_{-T}^T \left| \Lambda\left(\frac{1}{2}, E_t, \frac{v}{q}\right) \right|^2 dt.$$

The result once more follows from the spectral large sieve, which states that

$$\int_{-T}^T \left| \sum_{n \leq N} a_n \lambda(n, t) \right|^2 dt \ll_{\varepsilon} (TN)^{\varepsilon} (T^2 + N) \sum_{n \leq N} |a_n|^2. \quad \blacksquare$$

7.3. Proof of Theorem 1.1

Proof of Theorem 1.1. Let $u: \mathbb{H} \rightarrow \mathbb{R}$ be a smooth compactly supported function such that no two points of the support of u are Γ -equivalent. This gives rise to a smooth compactly supported function $U(z): \Gamma \backslash \mathbb{H} \rightarrow \mathbb{R}$ given by

$$U(z) := \sum_{\gamma \in \Gamma} u(\gamma z).$$

For each odd prime q , choose $\omega_q \in \Gamma/\Gamma_0(q)$ and define

$$U_q(z) := \sum_{\gamma \in \Gamma_0(q)} u(\omega_q \gamma z) = \begin{cases} U(z) & \text{if } z \in \omega_q^{-1} \Gamma \backslash \mathbb{H}, \\ 0 & \text{otherwise.} \end{cases}$$

Via Lemmata 7.1 and 7.2, the spectral expansion of $U_q(z)$ reads

$$U_q(z) = \frac{1}{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})} \int_{\Gamma\backslash\mathbb{H}} U(z) d\mu(z) + \sum_{q_1 q_2 = q} \sum_{\ell|q_2} \sum_{f \in \mathcal{B}_0^*(\Gamma_0(q_1))} \langle U_q, f_\ell \rangle_q f_\ell(z) \\ + \frac{1}{4\pi} \sum_{\ell|q} \int_{-\infty}^{\infty} \left\langle U_q, E_\ell\left(\cdot, \frac{1}{2} + it\right) \right\rangle_q E_\ell\left(z, \frac{1}{2} + it\right) dt.$$

Then by character orthogonality and Lemma 6.11, the quantity

$$\frac{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \sum_{A \in G_D} \int_{\mathcal{F}_A(q)} U_q(z) d\mu(z),$$

where $G_D = C(\text{Cl}_D^+)^2 \in \text{Gen}_D$ is a genus with $C \in \text{Cl}_D^+$, is equal to the sum of the contribution from the residual spectrum,

$$\int_{\Gamma\backslash\mathbb{H}} U(z) d\mu(z),$$

the contribution from the cuspidal spectrum,

$$\frac{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \frac{1}{2^{\omega(D)-1}} \sum_{\chi \in \widehat{\text{Gen}_D}} \chi(C) \sum_{q_1 q_2 = q} \sum_{\ell_1 \ell_2 | q_2} \frac{\sqrt{\varphi(\ell_1 \ell_2)} v(\ell_1) \mu(\ell_2) \lambda_f(\ell_2)}{\ell_1^{\frac{3}{2}} \ell_2 \sqrt{v(q_2)}} \\ \times \sum_{f \in \mathcal{B}_0^*(\Gamma_0(q_1))} \sqrt{L_{\ell_1 \ell_2}(1, \text{ad } f)} \langle U_q, f_{\ell_1 \ell_2} \rangle_q W_{\chi, f}^{\ell_1, q},$$

and the contribution from the continuous spectrum,

$$\frac{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \frac{1}{2^{\omega(D)-1}} \sum_{\chi \in \widehat{\text{Gen}_D}} \chi(C) \sum_{\ell_1 \ell_2 | q} \frac{v(\ell_1) \mu(\ell_2) \lambda(\ell_2, t)}{\ell_1 \sqrt{\ell_2 v(q)}} \\ \times \frac{1}{4\pi} \int_{-\infty}^{\infty} |\zeta_{\ell_1 \ell_2}(1 + 2it)| \left\langle U_q, E_{\ell_1 \ell_2}\left(\cdot, \frac{1}{2} + it\right) \right\rangle_q W_{\chi, t}^{\ell_1, q} dt.$$

Here $\widehat{\text{Gen}_D}$ is the group of genus characters, namely characters that are trivial on $(\text{Cl}_D^+)^2$, while the Weyl sums $W_{\chi, f}^{\ell, q}$ and $W_{\chi, t}^{\ell, q}$ are as in (6.16) and (6.17).

We fix $\varepsilon > 0$ and break up the contributions from the cuspidal and continuous spectra into two parts: the contributions for which $|t_f|, |t| \leq D^\varepsilon$ and their complements. These complementary contributions are negligibly small, since for any nonnegative integer A , we have that

$$\langle U_q, f_\ell \rangle_q \ll_{U, A} (1 + |t_f|)^{-A}, \quad \left\langle U_q, E_\ell\left(\cdot, \frac{1}{2} + it\right) \right\rangle_q \ll_{U, A} (1 + |t|)^{-A}, \quad (7.10)$$

which follows by repeated integration by parts (see, for example [49, (6.1)]). Thus from the Cauchy–Schwarz inequality,

$$\left| \frac{\text{vol}(\Gamma_0(q)\backslash\mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \sum_{A \in G_D} \int_{\mathcal{F}_A(q)} U_q(z) d\mu(z) - \int_{\Gamma\backslash\mathbb{H}} U(z) d\mu(z) \right|^2$$

is bounded by the sum of two terms plus a negligibly small term. These two terms are those for which $|t_f| \leq D^\varepsilon$ and $|t| \leq D^\varepsilon$. By Bessel's inequality, these two terms are

$$\ll_\varepsilon D^\varepsilon q^\varepsilon |\langle U, U \rangle_1|^2 \left(\frac{\text{vol}(\Gamma_0(q) \backslash \mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \right)^2 \sum_{\chi \in \widehat{\text{Gen}_D}} \sum_{q_1 \ell | q} \sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ |t_f| \leq D^\varepsilon}} \frac{1}{\ell} |W_{\chi, f}^{\ell, q}|^2 \quad (7.11)$$

and

$$\ll_\varepsilon D^\varepsilon q^\varepsilon |\langle U, U \rangle_1|^2 \left(\frac{\text{vol}(\Gamma_0(q) \backslash \mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \right)^2 \sum_{\chi \in \widehat{\text{Gen}_D}} \sum_{\ell | q} \frac{1}{\ell} \int_{-D^\varepsilon}^{D^\varepsilon} |W_{\chi, t}^{\ell, q}|^2 dt, \quad (7.12)$$

respectively.

We recall that the set of genus characters $\chi \in \widehat{\text{Gen}_D}$ is indexed by ordered pairs of fundamental discriminants (D_1, D_2) such that $D_1 D_2 = D$; each genus character χ is then associated to the pair of primitive quadratic Dirichlet characters χ_{D_1} and χ_{D_2} modulo $|D_1|$ and $|D_2|$, respectively. Note that Θ_χ is an Eisenstein series, so that $L(s, f \otimes \Theta_\chi) = L(s, f \otimes \chi_{D_1}) L(s, f \otimes \chi_{D_2})$ and $L(s, \Theta_\chi) = L(s, \chi_{D_1}) L(s, \chi_{D_2})$.

For $\ell q_1 | q$ and $f \in \mathcal{B}_0^*(\Gamma_0(q_1))$, we deduce from Proposition 6.1 and Stirling's formula (6.2) applied to the geodesic term, as well as the Cauchy–Schwarz inequality applied to the topological terms, that

$$\begin{aligned} |W_{\chi, f}^{\ell, q}|^2 &\ll_\varepsilon \frac{\sqrt{D}}{q_1 (1 + |t_f|)^3} \frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)} \\ &\quad + \delta_{q_1 \ell, q} \frac{\sqrt{D}}{(1 + |t_f|)^4} \sum_{j=1}^{2g} \left| \int_{\mathcal{C}_j} (R_0 f)(\ell z) \frac{dz}{\text{Im}(z)} \right|^2 \\ &\quad \times \sum_{j=1}^{2g} \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} \sum_{\pm} |\langle v_h^\pm, \omega_j \rangle_{\text{cap}}|^2 \\ &\quad \times \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} L\left(\frac{1}{2}, h \otimes \chi_{D_1}\right) L\left(\frac{1}{2}, h \otimes \chi_{D_2}\right), \end{aligned} \quad (7.13)$$

and similarly

$$\begin{aligned} |W_{\chi, t}^{\ell, q}|^2 &\ll_\varepsilon \frac{\sqrt{D}}{(1 + |t|)^3} \left| \frac{L(\frac{1}{2} + it, \chi_{D_1}) L(\frac{1}{2} + it, \chi_{D_2})}{\zeta(1 + 2it)} \right|^2 \\ &\quad + \delta_{\ell, q} \sqrt{D} \sum_{j=1}^{2g} \left| \int_{\mathcal{C}_j} (R_0 E)\left(\ell z, \frac{1}{2} + it\right) \frac{dz}{\text{Im}(z)} \right|^2 \\ &\quad \times \sum_{j=1}^{2g} \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} \sum_{\pm} |\langle v_h^\pm, \omega_j \rangle_{\text{cap}}|^2 \\ &\quad \times \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} L\left(\frac{1}{2}, h \otimes \chi_{D_1}\right) L\left(\frac{1}{2}, h \otimes \chi_{D_2}\right). \end{aligned} \quad (7.14)$$

We first bound the portions of (7.11) and (7.12) coming from the first terms on the right-hand sides of (7.13) and (7.14), namely the geodesic terms. We use Hölder's inequality with exponents $(3, 3, 3)$ together with the Weyl law upper bound

$$\sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ |t_f| \leq T}} 1 \ll T^2 q_1$$

and with the third moment bounds

$$\sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ T-1 \leq |t_f| \leq T}} \frac{L(\frac{1}{2}, f \otimes \chi_D)^3}{L(1, \text{ad } f)} \ll_\varepsilon D^{1+\varepsilon} q_1^{1+\varepsilon} T^{1+\varepsilon}, \quad (7.15)$$

$$\int_{T-1 \leq |t| \leq T} \left| \frac{L(\frac{1}{2} + it, \chi_D)^3}{\zeta(1 + 2it)} \right|^2 dt \ll_\varepsilon D^{1+\varepsilon} T^{1+\varepsilon}. \quad (7.16)$$

The sixth moment bound (7.16) is [64, Theorem 1]. The third moment bound (7.15) with $T^{1+\varepsilon}$ replaced by T^A for some unspecified large constant A was proven for *holomorphic* cusp forms by Petrow and Young [56, Theorem 1]; it was then improved to form (7.15) in [4, Theorem 4.1] and in a more general form in [25, Theorem 11.1], which both also recover bound (7.16). Combining this with the fact that $|\widehat{\text{Gen}}_D| = 2^{\omega(D)-1} \ll_\varepsilon D^\varepsilon$, and that the lower bound for the volume $\text{vol}(\mathcal{F}_A(q))$ coming from (3.9), we obtain the bounds $O_\varepsilon(D^{-\frac{1}{6}+\varepsilon} q^{2+\varepsilon})$ for these terms.

We next bound the portions of (7.11) and (7.12) coming from the second terms on the right-hand sides of (7.13) and (7.14), namely the topological terms. For the innermost sum over $h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))$, we use Hölder's inequality with exponents $(3, 3, 3)$ together with the dimension upper bound $|\mathcal{B}_2^{\text{hol}}(\Gamma_0(q))| \ll q$ and the third moment bound

$$\sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} L\left(\frac{1}{2}, h \otimes \chi_D\right)^3 \ll_\varepsilon D^{1+\varepsilon} q^{1+\varepsilon}$$

due to Petrow and Young [56, Theorem 1] (see also [25, Theorem 11.1]). Next, we want to apply the L^2 -bound for the cap product pairing derived by the second author [54, Theorem 6.1] which applies to (dual bases of) *basic bases* of homology in the terminology of [54]. These are bases of the vector space $H_1(Y_0(q), \mathbb{R})$ consisting of classes containing the geodesic connecting z and γz , where $\gamma \in \Gamma_0(q)$ and $\gamma\infty = \frac{v}{q}$ with $0 < v < q$. Lemma 7.3 implies that indeed

$$W_q[\mathcal{C}_1], \dots, W_q[\mathcal{C}_{2g}]$$

is the projection to the homology $H_1(X_0(q), \mathbb{R})$ of a basic basis, where W_q denotes the Atkin–Lehner involution acting on homology. Since W_q is self-adjoint with respect to the cap product pairing and $W_q h(z) = \eta_h(q) h(z)$, where $\eta_h(q) \in \{1, -1\}$ is the Atkin–Lehner eigenvalue of $h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))$, we get

$$|\langle v_h^\pm, \omega_j \rangle_{\text{cap}}|^2 = |\langle v_h^\pm, W_q \omega_j \rangle_{\text{cap}}|^2.$$

Since W_q commutes with taking dual bases (using that W_q is self-adjoint), we conclude by [54, Theorem 6.1] the bound

$$\sum_{j=1}^{2g} \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} \sum_{\pm} |\langle v_h^{\pm}, \omega_j \rangle_{\text{cap}}|^2 \ll_{\varepsilon} q^{2+\varepsilon}.$$

As discussed in [54, Section 6], this should be viewed as a homological version of the sup-norm problem.

Finally, we apply Lemmata 7.4 and 7.5 to bound the remaining second moment of integrals along the hyperbolic sides $\mathcal{C}_j$. Combined, this yields the bounds $O_{\varepsilon}(D^{-\frac{1}{6}+\varepsilon}q^{6+\varepsilon})$ for these terms.

We deduce that

$$\begin{aligned} & \frac{\text{vol}(\Gamma_0(q) \backslash \mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \sum_{A \in G_D} \int_{\mathcal{F}_A(q)} U_q(z) d\mu(z) \\ &= \int_{\Gamma \backslash \mathbb{H}} U(z) d\mu(z) + O_{U,\varepsilon}(D^{-\frac{1}{12}+\varepsilon}q^{3+\varepsilon}). \end{aligned} \quad (7.17)$$

We improve the error term in (7.17) to $O_{\varepsilon}(D^{-\frac{1}{4}+\varepsilon}q^{3+\varepsilon})$ under the assumption of the generalised Lindelöf hypothesis by invoking the conditional bounds

$$\begin{aligned} & \frac{L(\frac{1}{2}, f \otimes \chi_{D_1})L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)} \ll_{\varepsilon} (Dq_1(1 + |t_f|))^{\varepsilon}, \\ & L\left(\frac{1}{2}, h \otimes \chi_{D_1}\right)L\left(\frac{1}{2}, h \otimes \chi_{D_2}\right) \ll_{\varepsilon} (Dq)^{\varepsilon}, \\ & \left| \frac{L(\frac{1}{2} + it, \chi_{D_1})L(\frac{1}{2} + it, \chi_{D_2})}{\zeta(1 + 2it)} \right|^2 \ll_{\varepsilon} (D(1 + |t|))^{\varepsilon}, \end{aligned}$$

and using the Weyl law instead of applying Hölder's inequality.

We are almost able to deduce Theorem 1.1 upon taking $u(z)$ to closely approximate a fundamental domain of $\Gamma \backslash \mathbb{H}$. There is one remaining obstacle, namely that we are assuming that u is *compactly supported*, and so we have not precluded the possibility of the escape of mass in this equidistribution problem.

We circumvent this by performing a dyadic partition of unity and choosing u to be a function of the form $u(z) := \delta_{\text{Re}(z) \in [0,1]} \Psi(\frac{\text{Im}(z)}{Y})$, with $Y > 1$ a large parameter and $\Psi: (0, \infty) \rightarrow \mathbb{R}$ a smooth bump function equal to 1 on $[1, 2]$ and vanishing outside $[\frac{1}{2}, \frac{5}{2}]$. We use the same framework as above except that bounds (7.10) are replaced by

$$\langle U_q, f_{\ell} \rangle_q = 0, \quad \left\langle U_q, E_{\ell}\left(\cdot, \frac{1}{2} + it\right) \right\rangle_q \ll_{\Psi} Y^{-\frac{1}{2}}(1 + |t|)^{-A}$$

for any $A > 0$. These follow by inserting the Fourier expansions (4.2) and (4.10): the cusp form term vanishes since there is no constant term in the Fourier expansion, while for the Eisenstein case, we make the change of variables $y \mapsto Yy$ and then repeatedly integrate by parts. Thus there is no contribution from the cuspidal spectrum, while we bound the

contribution from the continuous spectrum using the same bounds as above. In this way, we find that

$$\begin{aligned} & \frac{\text{vol}(\Gamma_0(q) \backslash \mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A(q))} \sum_{A \in G_D} \int_{\mathcal{F}_A(q)} U_q(z) d\mu(z) \\ &= \int_{\Gamma \backslash \mathbb{H}} U(z) d\mu(z) + O_{U,\varepsilon}(Y^{-1} D^{-\frac{1}{12}+\varepsilon} q^{\frac{3}{2}+\varepsilon}). \end{aligned}$$

This yields the desired result. ■

Remark 7.6. Instead of proceeding via Hölder's inequality together with bounds for the third moments of L -functions, it is tempting to try to directly bound the moments

$$\begin{aligned} & \sum_{\substack{f \in \mathcal{B}_0^*(\Gamma_0(q_1)) \\ |t_f| \leq D^\varepsilon}} \frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)}, \\ & \sum_{h \in \mathcal{B}_2^{\text{hol}}(\Gamma_0(q))} L\left(\frac{1}{2}, h \otimes \chi_{D_1}\right) L\left(\frac{1}{2}, h \otimes \chi_{D_2}\right), \\ & \int_{-D^\varepsilon}^{D^\varepsilon} \left| \frac{L(\frac{1}{2} + it, \chi_{D_1}) L(\frac{1}{2} + it, \chi_{D_2})}{\zeta(1 + 2it)} \right|^2 dt, \end{aligned}$$

via approximate functional equations, the Kuznetsov and Petersson formulæ, and the Voronoi summation formula (or equivalently two applications of the Poisson summation formula). Work of Holowinsky and Templier [35, Theorem 1] suggests that it should be possible via this approach to bound each of these by $O_\varepsilon(q^{1+\varepsilon} + D^{\frac{1}{2}+\varepsilon})$ (cf. [38, Remark 6.2]). Unfortunately, this falls just shy of being sufficient for our purposes; we instead require a bound of the form $O(D^{\frac{1}{2}-\delta})$ for some $\delta > 0$. This is a level-aspect analogue of the obstacle discussed in [39, Section 8] connecting small-scale equidistribution to sub-Weyl subconvexity.

Remark 7.7. Liu, Masri, and Young [49, Theorem 1.4] have proven the equidistribution of level q Heegner points in translates $\omega_q^{-1} \Gamma \backslash \mathbb{H}$ for odd negative fundamental discriminants D and for primes q that split in $\mathbb{Q}(\sqrt{D})$ as $q(-D)$ tends to infinity under the proviso that $q \leq (-D)^\delta$ for some fixed $\delta < \frac{1}{20}$. The method of proof of Theorem 1.1 may be used to strengthen this result to relax the condition on q to be squarefree, D to be any negative fundamental discriminant (not necessarily odd), and $q \leq (-D)^\delta$ for some fixed $\delta < \frac{1}{12}$. Moreover, an analogous result can also be shown to hold for the equidistribution of level q closed geodesics. The input for this improvement is the usage of the Weyl-strength third moment bounds (7.15) and (7.16) in place of Burgess-strength pointwise bounds

$$\begin{aligned} L\left(\frac{1}{2}, f \otimes \chi_D\right) &\ll_{t_f, \varepsilon} D^{\frac{3}{8}+\varepsilon} q^{\frac{1}{2}+\varepsilon}, \\ L\left(\frac{1}{2} + it, \chi_D\right) &\ll_{t, \varepsilon} D^{\frac{3}{16}+\varepsilon}. \end{aligned}$$

8. Sparse equidistribution in the subgroup aspect

Proof of Theorem 1.2. Via the Weyl equidistribution criterion together with the lower bound (3.9) for $\text{vol}(\mathcal{F}_A(q))$ and the orthonormal bases for the cuspidal and continuous spectra given in Lemmata 7.1 and 7.2, it suffices to show that for each $q_1 \mid q$, $\ell \mid \frac{q}{q_1}$, and $f \in \mathcal{B}_0^*(\Gamma_0(q_1))$, and for each $t \in \mathbb{R}$,

$$\sum_{A \in CH} \int_{\mathcal{F}_A(q)} f(\ell z) d\mu(z) = o_{q,t,f} \left(\frac{|H|}{h_D^+} \frac{\sqrt{D} L(1, \chi_D)}{\log D} \right),$$

$$\sum_{A \in CH} \int_{\mathcal{F}_A(q)} E\left(\ell z, \frac{1}{2} + it\right) d\mu(z) = o_{q,t} \left(\frac{|H|}{h_D^+} \frac{\sqrt{D} L(1, \chi_D)}{\log D} \right),$$

where we have used the narrow class number formula $h_D^+ \log \epsilon_D = \sqrt{D} L(1, \chi_D)$. Via character orthogonality, these expressions are respectively equal to

$$\frac{|H|}{h_D^+} \sum_{\chi \in H^\perp} \bar{\chi}(C) W_{\chi,f}^{\ell,q}, \quad \frac{|H|}{h_D^+} \sum_{\chi \in H^\perp} \bar{\chi}(C) W_{\chi,t}^{\ell,q},$$

where the Weyl sums $W_{\chi,f}^{\ell,q}$ and $W_{\chi,t}^{\ell,q}$ are as in (6.16) and (6.17). From Lemma 6.11 and Propositions 6.1 and 6.5, we have that

$$|W_{\chi,f}^{\ell,q}|^2 \ll_{q,t,f} \sqrt{D} \left(L\left(\frac{1}{2}, f \otimes \Theta_\chi\right) + \sum_{h \in \mathcal{B}_2^{\text{hol}}(q)} \left(L\left(\frac{1}{2}, h \otimes \Theta_\chi\right) + L\left(\frac{1}{2}, h \otimes \Theta_{\bar{\chi}}\right) \right) \right),$$

$$|W_{\chi,t}^{\ell,q}|^2 \ll_{q,t} \sqrt{D} \left(\left| L\left(\frac{1}{2} + it, \Theta_\chi\right) \right|^2 + \sum_{h \in \mathcal{B}_2^{\text{hol}}(q)} \left(L\left(\frac{1}{2}, h \otimes \Theta_\chi\right) + L\left(\frac{1}{2}, h \otimes \Theta_{\bar{\chi}}\right) \right) \right).$$

Theorem 1.2 thereby holds unconditionally for $\delta < \frac{1}{2826}$ due to the subconvexity estimates

$$L\left(\frac{1}{2}, f \otimes \Theta_\chi\right) \ll_{q_1,t,f,\varepsilon} D^{\frac{1}{2} - \frac{1}{1413} + \varepsilon}, \quad (8.1)$$

$$L\left(\frac{1}{2}, h \otimes \Theta_\chi\right) \ll_q D^{\frac{1}{2} - \frac{1}{1057}}, \quad (8.2)$$

$$L\left(\frac{1}{2} + it, \Theta_\chi\right) \ll_t D^{\frac{1}{4} - \frac{1}{1889}}, \quad (8.3)$$

combined with the fact that $|H^\perp| = \frac{h_D^+}{|H|}$, and the (ineffective) Siegel bound $L(1, \chi_D) \gg_\varepsilon D^{-\varepsilon}$. When χ is not a genus character, so that Θ_χ is a cusp form, the first bound above is due to Harcos and Michel [32, Theorem 1] (see additionally [31, Theorem 1.3]), the second is due to Michel [52, Theorem 2], while the third is due to Blomer, Harcos, and Michel [9, Theorem 2]. When χ is a genus character, then stronger bounds are known via the work of Petrow and Young [56, Theorem 1].

We obtain Theorem 1.2 for $\delta < \frac{1}{4}$ under the assumption of the generalised Lindelöf hypothesis, since this implies the stronger bounds

$$|W_{\chi,f}^{\ell,q}|^2 \ll_{q,t,f,\varepsilon} D^{\frac{1}{2} + \varepsilon} \quad \text{and} \quad |W_{\chi,t}^{\ell,q}|^2 \ll_{q,t,\varepsilon} D^{\frac{1}{2} + \varepsilon}. \quad \blacksquare$$

Remark 8.1. Improvements of bound (8.3) exist in the literature (see, for example, [10, Theorem 1] for D prime); the obstacles in unconditionally enlarging the range of δ in Theorem 1.2 are improvements of bounds (8.1) and (8.2).

In [19, Section 4], it is observed that Theorem 1.2 is *trivial* when $q = 1$ and $H = \text{Cl}_D^+$ is the whole narrow class group, for then for every fixed continuity set $B \subset \Gamma_0(q) \backslash \mathbb{H}$, we have the *equality*

$$\frac{\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q) \cap \Gamma_0(q)B)}{\sum_{A \in CH} \text{vol}(\mathcal{F}_A(q))} = \frac{\text{vol}(B)}{\text{vol}(\Gamma_0(q) \backslash \mathbb{H})} \quad (8.4)$$

with *no* error term. Moreover, if

$$q = 1 \quad \text{and} \quad H = (\text{Cl}_D^+)^2,$$

so that $CH \in \text{Gen}_D$ is a genus, then we also have equality (8.4) if J lies in the principal genus $(\text{Cl}_D^+)^2 \in \text{Gen}_D$, where $J \in \text{Cl}_D^+$ is the narrow ideal class containing the different $\mathfrak{d} := (\sqrt{D})$, so that J^2 is the principal narrow ideal class I ; this occurs if and only if D is not divisible by a prime congruent to 3 modulo 4.

The reason for the trivial equality (8.4) is simply due to the fact that the pair $\Gamma_A \backslash \mathcal{N}_A$ and $\Gamma_{JA^{-1}}(1) \backslash \mathcal{N}_{JA^{-1}}(1)$ are *complementary*, in the sense that their union covers $\Gamma \backslash \mathbb{H}$ evenly and the images of their boundary geodesics are the same as sets but with opposite orientations. As a consequence, we have the following result.

Corollary 8.2. *Let $q = 1$ and let D be a positive fundamental discriminant. Let CH be a coset of Cl_D^+ with H a subgroup of Cl_D^+ and $C \in \text{Cl}_D^+$. Then we have equality (8.4) for every fixed continuity set $B \subset \Gamma \backslash \mathbb{H}$ if $C^2 J \in H$.*

Proof. The oriented closed geodesics $\mathcal{C}_A(1)$ and $\mathcal{C}_{JA^{-1}}(1)$ are the same curve with opposite orientations, which means that $\Gamma_A(1) \backslash \mathcal{N}_A(1)$ and $\Gamma_{JA^{-1}}(1) \backslash \mathcal{N}_{JA^{-1}}(1)$ cover $\Gamma_0(1) \backslash \mathbb{H}$ evenly. Thus (8.4) holds if $JA^{-1} \in CH$ for every $A \in CH$. This condition is met precisely when $C^2 J \in H$. ■

For $q \neq 1$, on the other hand, it is no longer the case that $\Gamma_A(q) \backslash \mathcal{N}_A(q)$ and $\Gamma_{JA^{-1}}(q) \backslash \mathcal{N}_{JA^{-1}}(q)$ are complementary. Indeed, if a narrow ideal class A is associated to a Heegner form

$$Q = [a, b, c] \in \mathcal{Q}_D(q), \quad \text{where } q > 1 \text{ and } b \equiv r \pmod{2q}$$

for some fixed residue class r modulo $2q$ such that $r^2 \equiv D \pmod{4q}$, then there does *not* exist a Heegner form

$$Q' = [a', b', c'] \in \mathcal{Q}_D(q) \quad \text{with } b' \equiv r \pmod{2q}$$

such that $\Gamma_Q(q) \backslash \mathcal{N}_Q(q)$ and $\Gamma_{Q'}(q) \backslash \mathcal{N}_{Q'}(q)$ are complementary. Instead, the complementary Heegner form is $[-a, -b, -c] \in \mathcal{Q}_D(q)$, which is such that $-b \equiv -r \pmod{2q}$, and hence does not correspond to a narrow ideal class $A \in CH$ appearing in the sums in (8.4). As a notable consequence, Theorem 1.2 is nontrivial if $q > 1$ and $H = \text{Cl}_D^+$.

9. Small scale equidistribution and discrepancy bounds

9.1. Automorphic kernels and Selberg–Harish-Chandra transforms

For $z, w \in \mathbb{H}$, set

$$\rho(z, w) := \log \frac{|z - \bar{w}| + |z - w|}{|z - \bar{w}| - |z - w|}, \quad u(z, w) := \frac{|z - w|^2}{4 \operatorname{Im}(z) \operatorname{Im}(w)} = \sinh^2 \frac{\rho(z, w)}{2}.$$

The function $u: \mathbb{H} \times \mathbb{H} \rightarrow [0, \infty)$ is a point-pair invariant for the symmetric space $\mathbb{H} \cong \operatorname{SL}_2(\mathbb{R})/\operatorname{SO}(2)$; that is, $u(gz, gw) = u(z, w)$ for all $g \in \operatorname{SL}_2(\mathbb{R})$ and $z, w \in \mathbb{H}$. From this, a function $k: [0, \infty) \rightarrow \mathbb{C}$ gives rise to a point-pair invariant $k(u(z, w))$ on $\mathbb{H}$.

We take $k(u(z, w)) = k_R(u(z, w))$ to be equal to the indicator function of a ball of radius R centred at a point w ,

$$B_R(w) := \{z \in \mathbb{H} : \rho(z, w) \leq R\} = \left\{z \in \mathbb{H} : u(z, w) \leq \sinh^2 \frac{R}{2}\right\},$$

normalised by the volume of this ball,

$$\operatorname{vol}(B_R) = 4\pi \sinh^2 \frac{R}{2},$$

namely

$$k_R(u(z, w)) := \begin{cases} \frac{1}{\operatorname{vol}(B_R)} & \text{if } u(z, w) \leq \sinh^2 \frac{R}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

We additionally consider the convolution

$$k * k'(u(z, w)) = \int_{\mathbb{H}} k(u(z, \zeta)) k'(u(\zeta, w)) d\mu(\zeta)$$

of two point-pair invariants k, k' on $\mathbb{H}$.

Lemma 9.1. *For $0 < \rho < R$, the convolution $k_R * k_\rho(u(z, w))$ is nonnegative, bounded by $\frac{1}{\operatorname{vol}(B_R)}$, and satisfies*

$$k_R * k_\rho(u(z, w)) = \begin{cases} \frac{1}{\operatorname{vol}(B_R)} & \text{if } u(z, w) \leq \sinh^2 \frac{R-\rho}{2}, \\ 0 & \text{if } u(z, w) \geq \sinh^2 \frac{R+\rho}{2}, \end{cases}$$

so that

$$\frac{\operatorname{vol}(B_{R-\rho})}{\operatorname{vol}(B_R)} k_{R-\rho} * k_\rho(u(z, w)) \leq k_R(u(z, w)) \leq \frac{\operatorname{vol}(B_{R+\rho})}{\operatorname{vol}(B_R)} k_{R+\rho} * k_\rho(u(z, w))$$

for all $z, w \in \mathbb{H}$.

Proof. This follows from the triangle inequality for the distance function $\rho(z, w)$. ■

Given $k: [0, \infty) \rightarrow \mathbb{C}$, we define the automorphic kernel $K: \Gamma \backslash \mathbb{H} \times \Gamma \backslash \mathbb{H} \rightarrow \mathbb{C}$ by

$$K(z, w) := \sum_{\gamma \in \Gamma} k(u(\gamma z, w)).$$

We write

$$K_R(z, w) := \sum_{\gamma \in \Gamma} k_R(u(\gamma z, w)).$$

This is the indicator function of an injective geodesic ball in $\Gamma \backslash \mathbb{H}$ provided that

$$k_R(u(\gamma_1 z, w)) = k_R(u(\gamma_2 z, w)) = 1$$

if and only if $\gamma_1 = \gamma_2$.

A necessary condition for this is the requirement that $2 \operatorname{ht}(w) \sinh R \leq 1$, where $\operatorname{ht}(w) := \max_{\gamma \in \Gamma} \operatorname{Im}(\gamma w)$. For $0 < \rho < R$, we consider the convolution kernel

$$K_R * K_\rho(z, w) := \sum_{\gamma \in \Gamma} k_R * k_\rho(u(\gamma z, w)).$$

Corollary 9.2. *For $0 < \rho < R$, the convolution kernel $K_R * K_\rho(z, w)$ is nonnegative and satisfies*

$$\frac{\operatorname{vol}(B_{R-\rho})}{\operatorname{vol}(B_R)} K_{R-\rho} * K_\rho(z, w) \leq K_R(z, w) \leq \frac{\operatorname{vol}(B_{R+\rho})}{\operatorname{vol}(B_R)} K_{R+\rho} * K_\rho(z, w) \quad (9.1)$$

for all $z, w \in \Gamma \backslash \mathbb{H}$ satisfying $2 \operatorname{ht}(w) \sinh R \leq 1$.

The spectral expansion for the convolution kernel $K_{R \pm \rho} * K_\rho$ involves a sum over an orthonormal basis $\mathcal{B}_0(\Gamma)$ of the space of Maaß cusp forms (which we may choose to consist of Hecke–Maaß eigenforms), where the inner product is

$$\langle f, g \rangle := \int_{\Gamma \backslash \mathbb{H}} f(z) \overline{g(z)} d\mu(z),$$

and an integral over $t \in \mathbb{R}$ indexing the Eisenstein series $E(z, \frac{1}{2} + it)$. It also involves the Selberg–Harish-Chandra transforms $h_{R \pm \rho}$ of $k_{R \pm \rho}$ and h_ρ of k_ρ . The Selberg–Harish-Chandra transform takes sufficiently well-behaved functions $k: [0, \infty) \rightarrow \mathbb{C}$ to functions $h: \mathbb{R} \cup i[-\frac{1}{2}, \frac{1}{2}] \rightarrow \mathbb{C}$ via

$$h(t) := 2\pi \int_0^\infty P_{-\frac{1}{2}+it}(\cosh \rho) k\left(\sinh^2 \frac{\rho}{2}\right) \sinh \rho d\rho,$$

where $P_\lambda^\mu(z)$ denotes the associated Legendre function. In particular,

$$h_R(t) = \frac{2\pi}{\operatorname{vol}(B_R)} \int_0^R P_{-\frac{1}{2}+it}(\cosh \rho) \sinh \rho d\rho.$$

The size of this transform is given by the following well-known bounds.

Lemma 9.3 ([39, Lemma 2.33]). *Suppose that $0 < R \leq \frac{1}{\varepsilon}$ for some fixed $\varepsilon > 0$. For $t \in \mathbb{R}$,*

$$h_R(t) \ll \begin{cases} 1 & \text{for } |t| \leq \frac{1}{R}, \\ R^{-\frac{3}{2}} |t|^{-\frac{3}{2}} & \text{for } |t| \geq \frac{1}{R}. \end{cases} \quad (9.2)$$

The advantage of convolving is that it smooths the point-pair invariant and improves the decay of the Selberg–Harish-Chandra transform, since the transform of the convolution $k_1 * k_2$ is the product $h_1(t)h_2(t)$ of the individual Selberg–Harish-Chandra transforms. This ensures that the convolution kernel has a spectral expansion on $L^2(\Gamma \backslash \mathbb{H})$ that not only converges in L^2 but uniformly.

Lemma 9.4 ([40, Theorems 1.14 and 7.4]). *The automorphic kernel K_R satisfies*

$$\begin{aligned} \int_{\Gamma \backslash \mathbb{H}} K_R(z, w) d\mu(z) &= h_R\left(\frac{i}{2}\right) = 1, \\ \int_{\Gamma \backslash \mathbb{H}} f(z) K_R(z, w) d\mu(z) &= h_R(t_f) f(w), \\ \int_{\Gamma \backslash \mathbb{H}} E\left(z, \frac{1}{2} + it\right) K_R(z, w) d\mu(z) &= h_R(t) E\left(w, \frac{1}{2} + it\right) \end{aligned}$$

for every $f \in \mathcal{B}_0(\Gamma)$, $t \in \mathbb{R}$, and $w \in \Gamma \backslash \mathbb{H}$. Moreover, the convolved kernel $K_{R \pm \rho} * K_\rho$ has the spectral expansion

$$\begin{aligned} K_{R \pm \rho} * K_\rho(z, w) &= \frac{1}{\text{vol}(\Gamma \backslash \mathbb{H})} + \sum_{f \in \mathcal{B}_0(\Gamma)} h_{R \pm \rho}(t_f) h_\rho(t_f) f(z) \overline{f(w)} \\ &\quad + \frac{1}{4\pi} \int_{-\infty}^{\infty} h_{R \pm \rho}(t) h_\rho(t) E\left(z, \frac{1}{2} + it\right) \overline{E\left(w, \frac{1}{2} + it\right)} dt, \end{aligned} \quad (9.3)$$

which converges absolutely and uniformly.

9.2. Bounds for fractional moments of L -functions

The proofs of Theorems 1.3 and 1.4 require the following bounds for fractional moments of certain L -functions.

Lemma 9.5. *Let D_1 and D_2 be odd squarefree fundamental discriminants and let χ_{D_1} and χ_{D_2} be the primitive quadratic Dirichlet characters modulo $|D_1|$ and $|D_2|$. Then for $T \geq 1$ and $w \in \Gamma \backslash \mathbb{H}$, we have that*

$$\left\{ \sum_{\substack{f \in \mathcal{B}_0(\Gamma) \\ T \leq t_f \leq 2T}} \sqrt{\frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)}} |f(w)| \right. \\ \left. \int_{T \leq |t| \leq 2T} \left| \frac{L(\frac{1}{2} + it, \chi_{D_1}) L(\frac{1}{2} + it, \chi_{D_2})}{\zeta(1 + 2it)} \right| \left| E\left(w, \frac{1}{2} + it\right) \right| dt \right\}$$

$$\ll_{\varepsilon} \begin{cases} T^{\frac{3}{2}}(T^{\frac{1}{2}} + \text{ht}(w)^{\frac{1}{2}})D^{\frac{1}{6}+\varepsilon} & \text{if } T \leq D^{\frac{1}{12}}, \\ T^{\frac{1}{2}}(T^{\frac{1}{2}} + \text{ht}(w)^{\frac{1}{2}})D^{\frac{1}{4}+\varepsilon} & \text{if } D^{\frac{1}{12}} \leq T \leq D^{\frac{1}{4}}, \\ T^{\frac{3}{2}+\varepsilon}(T^{\frac{1}{2}} + \text{ht}(w)^{\frac{1}{2}}) & \text{if } T \geq D^{\frac{1}{4}}. \end{cases} \quad (9.4)$$

Assuming the generalised Lindelöf hypothesis, we have the improved bounds

$$\left\{ \sum_{\substack{f \in \mathcal{B}_0(\Gamma) \\ T \leq t_f \leq 2T}} \sqrt{\frac{L(\frac{1}{2}, f \otimes \chi_{D_1})L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)}} |f(w)| \right. \\ \left. \int_{T \leq |t| \leq 2T} \left| \frac{L(\frac{1}{2} + it, \chi_{D_1})L(\frac{1}{2} + it, \chi_{D_2})}{\zeta(1 + 2it)} \right| \left| E\left(w, \frac{1}{2} + it\right) \right| dt \right\} \\ \ll_{\varepsilon} T^{\frac{3}{2}+\varepsilon}(T^{\frac{1}{2}} + \text{ht}(w)^{\frac{1}{2}})D^{\varepsilon} \quad (9.5)$$

for all $T \geq 1$.

Proof. We apply the Cauchy–Schwarz inequality and use the local Weyl law [40, Proposition 7.2]

$$\left\{ \sum_{t_f \leq T} |f(w)|^2 \int_{-T}^T \left| E\left(w, \frac{1}{2} + it\right) \right|^2 dt \right\} \ll T^2 + T \text{ht}(w)$$

and the bound

$$\left\{ \sum_{\substack{f \in \mathcal{B}_0(\Gamma) \\ T \leq t_f \leq 2T}} \frac{L(\frac{1}{2}, f \otimes \chi_{D_1})L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)} \right. \\ \left. \int T \leq |t| \leq 2T \left| \frac{L(\frac{1}{2} + it, \chi_{D_1})^2 L(\frac{1}{2} + it, \chi_{D_2})^2}{\zeta(1 + 2it)} \right|^2 dt \right\} \\ \ll_{\varepsilon} \begin{cases} T^2 D^{\frac{1}{3}+\varepsilon} & \text{if } T \leq D^{\frac{1}{12}}, \\ D^{\frac{1}{2}+\varepsilon} & \text{if } D^{\frac{1}{12}} \leq T \leq D^{\frac{1}{4}}, \\ T^{2+\varepsilon} & \text{if } T \geq D^{\frac{1}{4}}. \end{cases} \quad (9.6)$$

This is proven for $D_1 = 1$ and $D_2 = D$ in [39, Proposition 2.14]. The same method of proof yields the above bound, with the only notable difference being a slightly different application of the Voronoï summation formula for

$$\sum_{m=1}^{\infty} \frac{\lambda_{\chi_{D_1}, \chi_{D_2}}(m, 0) e(\frac{md}{c})}{m^s},$$

where $\lambda_{\chi_{D_1}, \chi_{D_2}}(m, 0) := \sum_{ab=m} \chi_{D_1}(a) \chi_{D_2}(b)$, which is given in [47, Appendix A]. Assuming the generalised Lindelöf hypothesis, the Weyl law implies the improved bounds $O_{\varepsilon}(T^{2+\varepsilon} D^{\varepsilon})$ for all $T \geq 1$ for (9.6). ■

9.3. Proofs of Theorems 1.3 and 1.4

Proof of Theorem 1.3. For $w \in \Gamma \backslash \mathbb{H}$, the difference

$$\frac{\text{vol}(\Gamma \backslash \mathbb{H})}{\text{vol}(B_R)} \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} - 1 \quad (9.7)$$

is equal to

$$\frac{\text{vol}(\Gamma \backslash \mathbb{H})}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} \int_{\mathcal{F}_A} K_R(z, w) d\mu(z) - 1.$$

By the upper bound (9.1), the spectral expansion (9.3), and character orthogonality, we see that for any $0 < \rho < R$, (9.7) is bounded from above by the sum of the three terms

$$\frac{\text{vol}(B_{R+\rho})}{\text{vol}(B_R)} - 1, \quad (9.8)$$

$$\frac{\text{vol}(B_{R+\rho}) \text{vol}(\Gamma \backslash \mathbb{H})}{\text{vol}(B_R) \sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} \frac{1}{2^{\omega(D)-1}} \times \sum_{\chi \in \widehat{\text{Gen}}_D} \chi(G_D) \sum_{f \in \mathcal{B}_0(\Gamma)} h_{R+\rho}(t_f) h_\rho(t_f) \overline{f(w)} W_{\chi, f}, \quad (9.9)$$

$$\frac{\text{vol}(B_{R+\rho}) \text{vol}(\Gamma \backslash \mathbb{H})}{\text{vol}(B_R) \sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} \frac{1}{2^{\omega(D)-1}} \times \sum_{\chi \in \widehat{\text{Gen}}_D} \chi(G_D) \frac{1}{4\pi} \int_{-\infty}^{\infty} h_{R+\rho}(t) h_\rho(t) \overline{E\left(w, \frac{1}{2} + it\right)} W_{\chi, t} dt. \quad (9.10)$$

Here the Weyl sums $W_{\chi, f}$ and $W_{\chi, t}$ are as in (6.1) and (6.4). Similarly, (9.7) is bounded from below by the sum of the same three terms except with $R + \rho$ replaced by $R - \rho$ at each instance. Thus to prove Theorem 1.3, it suffices to show that for $R > D^{-\delta}$ with $\delta < \frac{1}{2}$, there exists some choice of $\rho \in (0, R)$ for which each of the three terms (9.8), (9.9), and (9.10) is $o(1)$.

Term (9.8) is $O(\rho R^{-1})$ since $\text{vol}(B_R) \asymp R^2$. For terms (9.9) and (9.10), we begin by recalling that the set of genus characters $\chi \in \widehat{\text{Gen}}_D$ is indexed by ordered pairs of fundamental discriminants (D_1, D_2) such that $D_1 D_2 = D$; each genus character χ is then associated to the pair of primitive quadratic Dirichlet characters χ_{D_1} and χ_{D_2} modulo $|D_1|$ and $|D_2|$, respectively. From Proposition 6.1 and Stirling's formula, we deduce that

$$|W_{\chi, f}|^2 \ll \frac{\sqrt{D}}{(1 + |t_f|)^3} \frac{L(\frac{1}{2}, f \otimes \chi_{D_1}) L(\frac{1}{2}, f \otimes \chi_{D_2})}{L(1, \text{ad } f)},$$

$$|W_{\chi, t}|^2 \ll \frac{\sqrt{D}}{(1 + |t|)^3} \frac{L(\frac{1}{2} + it, \chi_{D_1}) L(\frac{1}{2} - it, \chi_{D_1}) L(\frac{1}{2} + it, \chi_{D_2}) L(\frac{1}{2} - it, \chi_{D_2})}{\zeta(1 + 2it) \zeta(1 - 2it)}.$$

We also use the bound $\sum_{A \in G_D} \text{vol}(\mathcal{F}_A) \gg_\varepsilon D^{\frac{1}{2}-\varepsilon}$ [19, Proposition 1]. We then dyadically divide up the sum over $f \in \mathcal{B}_0(\Gamma)$ and the integral over $t \in \mathbb{R}$ based on the size

of t_f , $|t| \in [T, 2T]$ and employ bounds (9.2) for $h_{R \pm \rho}$ and h_ρ and bounds (9.4) for the ensuing fractional moment of L -functions. It remains to choose

$$\rho = \begin{cases} R^{\frac{1}{2}} D^{-\frac{1}{12}} & \text{if } R \geq D^{-\frac{1}{12}}, \\ RD^{-\frac{1}{24}} & \text{if } D^{-\frac{5}{12}} \leq R \leq D^{-\frac{1}{12}}, \\ \frac{1}{2} R^{\frac{1}{2}} D^{-\frac{1}{4}} & \text{if } D^{-\frac{1}{2}} \leq R \leq D^{-\frac{5}{12}}, \\ \frac{1}{2} R & \text{if } 0 < R \leq D^{-\frac{1}{2}}. \end{cases}$$

In this way, we find that

$$\begin{aligned} & \left| \frac{\text{vol}(\Gamma \backslash \mathbb{H})}{\text{vol}(B_R)} \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} - 1 \right| \\ & \ll_\varepsilon \begin{cases} R^{-\frac{1}{2}} D^{-\frac{1}{12} + \varepsilon} & \text{if } D^{-\frac{1}{12}} \leq R \leq 1, \\ D^{-\frac{1}{24} + \varepsilon} & \text{if } D^{-\frac{5}{12}} \leq R \leq D^{-\frac{1}{12}}, \\ R^{-\frac{1}{2}} D^{-\frac{1}{4} + \varepsilon} & \text{if } R \leq D^{-\frac{5}{12}}. \end{cases} \end{aligned} \quad (9.11)$$

This is $o(1)$ provided that $R \geq D^{-\delta}$ for some $\delta < \frac{1}{2}$. ■

Proof of Theorem 1.4. From (9.11) and the fact that $\text{vol}(B_r) \asymp R^2$ for $R \leq 1$, we have that for any $w \in \Gamma \backslash \mathbb{H}$,

$$\left| \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} - \frac{\text{vol}(B_R)}{\text{vol}(\Gamma \backslash \mathbb{H})} \right| \ll_\varepsilon \begin{cases} R^{\frac{3}{2}} D^{-\frac{1}{12} + \varepsilon} & \text{if } D^{-\frac{1}{12}} \leq R \leq 1, \\ R^2 D^{-\frac{1}{24} + \varepsilon} & \text{if } D^{-\frac{5}{12}} \leq R \leq D^{-\frac{1}{12}}, \\ R^{\frac{3}{2}} D^{-\frac{1}{4} + \varepsilon} & \text{if } R \leq D^{-\frac{5}{12}}. \end{cases}$$

In particular,

$$\sup_{B_R(w) \subset \Gamma \backslash \mathbb{H}} \left| \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} - \frac{\text{vol}(B_R)}{\text{vol}(\Gamma \backslash \mathbb{H})} \right| \ll_\varepsilon D^{-\frac{1}{12} + \varepsilon}.$$

This result may be strengthened under the assumption of the generalised Lindelöf hypothesis by conditionally improving bounds (9.11). Instead of using the unconditional bounds (9.4), we use the conditional bounds (9.5). Choosing

$$\rho = \begin{cases} \frac{1}{2} R^{\frac{1}{2}} D^{-\frac{1}{4}} & \text{if } D^{-\frac{1}{2}} \leq R \leq 1, \\ \frac{1}{2} R & \text{if } R \leq D^{-\frac{1}{2}}, \end{cases}$$

we find that

$$\left| \frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \cap \Gamma B_R(w))}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A)} - \frac{\text{vol}(B_R)}{\text{vol}(\Gamma \backslash \mathbb{H})} \right| \ll_\varepsilon R^{\frac{3}{2}} D^{-\frac{1}{4} + \varepsilon}$$

for all $R \leq 1$, which yields the conditional bound $O_\varepsilon(D^{-\frac{1}{4} + \varepsilon})$ for the discrepancy. ■

10. Questions

We end by raising some natural follow-up questions beyond the results contained in this paper.

Question 10.1. Can one prove Theorem 1.2 for $\delta = 0$ using techniques purely from arithmetic ergodic theory?

This question was raised by Duke, Imamoğlu, and Tóth [19, Section 4]. Indeed, Duke's theorems on the equidistribution of closed geodesics and of Heegner points on the modular surface and additionally of lattice points on the sphere [17] were first proved under the additional hypothesis of a splitting condition by Linnik [48] using techniques from ergodic theory. Einsiedler, Lindenstrauss, Michel, and Venkatesh [23] showed that in the case of closed geodesics, Linnik's ergodic method is valid even without the splitting condition hypothesis.

Question 10.2. (1) For any fixed $\delta > 0$, does there exist a sequence of positive fundamental discriminants D and subsets H of Cl_D^+ that are not necessarily cosets with $\frac{|H|}{h_D^+} \ll D^{-\delta}$ for which the equidistribution result (1.2) in Theorem 1.2 fails to hold?

(2) Does the equidistribution result (1.2) in Theorem 1.2 hold for arbitrary subsets H of Cl_D^+ that are not necessarily cosets provided that $\frac{|H|}{h_D^+} \log D$ tends to infinity with D ?

The condition that CH be a coset of Cl_D^+ in Theorem 1.2 may be thought of as imposing the requirement that we restrict to a subset of Cl_D^+ with an algebraic structure. By comparing to related results on the sparse equidistribution of closed geodesics, namely [1, Theorem 4.1] and [12, Theorem 1.8], we expect that such an algebraic condition is necessary in order for equidistribution to hold if the cardinality of this subset is $O(D^{-\delta} h_D^+)$ for some $\delta > 0$. On the other hand, analogous results for closed geodesics, namely [1, Theorem 1.2], lead us to expect that Theorem 1.2 holds for subsets H of Cl_D^+ without any algebraic structure provided that $\frac{|H|}{h_D^+} \log D$ tends to infinity with D .

Question 10.3. Does the joint equidistribution of hyperbolic orbifolds hold?

Here, by joint equidistribution, we mean the following. For each positive fundamental discriminant D , choose a genus G_D in the group of genera Gen_D and choose $C \in \text{Cl}_D^+$ such that the minimal norm of any integral ideal representing C tends to infinity as D tends to infinity. Joint equidistribution of hyperbolic orbifolds is the statement that for each fixed continuity set $B \subset \Gamma \backslash \mathbb{H} \times \Gamma \backslash \mathbb{H}$,

$$\frac{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \times \mathcal{F}_{AC} \cap (\Gamma \times \Gamma)B)}{\sum_{A \in G_D} \text{vol}(\mathcal{F}_A \times \mathcal{F}_{AC})} = \frac{\text{vol}(B)}{\text{vol}(\Gamma \backslash \mathbb{H} \times \Gamma \backslash \mathbb{H})} + o_B(1)$$

as D tends to infinity. This is a natural analogue of the mixing conjecture of Michel and Venkatesh [53] for Heegner points, which has been resolved by Khayutin [45, Theorem 1.3] under the assumptions of a splitting hypothesis and of the nonexistence of Landau–Siegel zeroes. The first author resolved a toy model of this problem in the much simpler setting of modular inverses on the torus [37, Theorem 1.6].

Finally, we briefly mention that one can pose a closely related variant of joint equidistribution, namely *simultaneous equidistribution*. Associated to each narrow ideal class of a quadratic field $\mathbb{Q}(\sqrt{D})$ are geometric invariants for *each* quaternion algebra that is unramified at each place that splits in E . Michel and Venkatesh [53] pose the question of whether such geometric invariants equidistribute *simultaneously* in *multiple* different quaternion algebras. For progress on this problem and variants thereof, we direct the reader to [2, 3, 6, 7, 22].

Funding. The first author was supported by the National Science Foundation grant DMS-2302079 and the Simons Foundation (award 965056). The second author was supported by a grant from the Independent Research Fund Denmark DFF (1025-00020B) and by a public grant from Fondation Mathématique Jacques Hadamard.

References

- [1] Aka, M., Einsiedler, M.: [Duke's theorem for subcollections](#). *Ergodic Theory Dynam. Systems* **36**, 335–342 (2016) Zbl [1355.37004](#) MR [3503026](#)
- [2] Aka, M., Einsiedler, M., Shapira, U.: [Integer points on spheres and their orthogonal lattices](#). *Invent. Math.* **206**, 379–396 (2016) Zbl [1410.11117](#) MR [3570295](#)
- [3] Aka, M., Luethi, M., Michel, P., Wieser, A.: [Simultaneous supersingular reductions of CM elliptic curves](#). *J. Reine Angew. Math.* **786**, 1–43 (2022) Zbl [1536.11083](#) MR [4434745](#)
- [4] Andersen, N., Wu, H.: [Hybrid subconvexity and the partition function](#). *J. Number Theory* **242**, 154–180 (2023) Zbl [1506.11072](#) MR [4474855](#)
- [5] Baruch, E. M., Mao, Z.: [A generalized Kohnen–Zagier formula for Maass forms](#). *J. Lond. Math. Soc. (2)* **82**, 1–16 (2010) Zbl [1217.11047](#) MR [2669637](#)
- [6] Blomer, V., Brumley, F.: [Simultaneous equidistribution of toric periods and fractional moments of \$L\$ -functions](#). *J. Eur. Math. Soc. (JEMS)* **26**, 2745–2796 (2024) Zbl [1552.11046](#) MR [4756945](#)
- [7] Blomer, V., Brumley, F., Khayutin, I.: The mixing conjecture under GRH. [arXiv:2212.06280v1](#) (2022)
- [8] Blomer, V., Harcos, G.: [A hybrid asymptotic formula for the second moment of Rankin–Selberg \$L\$ -functions](#). *Proc. Lond. Math. Soc. (3)* **105**, 473–505 (2012) Zbl [1294.11063](#) MR [2974197](#)
- [9] Blomer, V., Harcos, G., Michel, P.: [Bounds for modular \$L\$ -functions in the level aspect](#). *Ann. Sci. École Norm. Sup. (4)* **40**, 697–740 (2007) Zbl [1185.11034](#) MR [2382859](#)
- [10] Blomer, V., Khan, R.: [Uniform subconvexity and symmetry breaking reciprocity](#). *J. Funct. Anal.* **276**, 2315–2358 (2019) Zbl [1492.11098](#) MR [3912807](#)
- [11] Bourgain, J.: [Mordell's exponential sum estimate revisited](#). *J. Amer. Math. Soc.* **18**, 477–499 (2005) Zbl [1072.11063](#) MR [2137982](#)
- [12] Bourgain, J., Kontorovich, A.: [Beyond expansion II: low-lying fundamental geodesics](#). *J. Eur. Math. Soc. (JEMS)* **19**, 1331–1359 (2017) Zbl [1403.11054](#) MR [3635355](#)
- [13] Caramello Jr., F. C.: Introduction to orbifolds. [arXiv:1909.08699v6](#) (2022)
- [14] Darmon, H.: [Heegner points, Heegner cycles, and congruences](#). In: *Elliptic curves and related topics*, CRM Proc. Lecture Notes 4, American Mathematical Society, Providence, RI, 45–59 (1994) Zbl [0813.11036](#) MR [1260954](#)
- [15] Doan, N. M., Kim, S.-h., Lang, M. L., Tan, S. P.: [Optimal special polygons for the congruence subgroups \$\Gamma_0\(p\)\$ and \$\Gamma_0\(pq\)\$](#) . *J. Geom. Anal.* **35**, article no. 216 (2025) MR [4917150](#)

- [16] Dooms, A., Jespers, E., Kononov, A.: [From Farey symbols to generators for subgroups of finite index in integral group rings of finite groups](#). *J. K-Theory* **6**, 263–283 (2010) Zbl 1209.16030 MR 2735087
- [17] Duke, W.: [Hyperbolic distribution problems and half-integral weight Maass forms](#). *Invent. Math.* **92**, 73–90 (1988) Zbl 0628.10029 MR 931205
- [18] Duke, W., Friedlander, J. B., Iwaniec, H.: [The subconvexity problem for Artin \$L\$ -functions](#). *Invent. Math.* **149**, 489–577 (2002) Zbl 1056.11072 MR 1923476
- [19] Duke, W., Imamoğlu, Ö., Tóth, Á.: [Geometric invariants for real quadratic fields](#). *Ann. of Math. (2)* **184**, 949–990 (2016) Zbl 1372.11056 MR 3549627
- [20] Duke, W., Imamoğlu, Ö., Tóth, Á.: [Kronecker’s first limit formula, revisited](#). *Res. Math. Sci.* **5**, article no. 20 (2018) Zbl 1441.11085 MR 3782448
- [21] Eichler, M.: [Grenzkreisgruppen und kettenbruchartige Algorithmen](#). *Acta Arith.* **11**, 169–180 (1965) Zbl 0148.32503 MR 228436
- [22] Einsiedler, M., Lindenstrauss, E.: [Joinings of higher rank torus actions on homogeneous spaces](#). *Publ. Math. Inst. Hautes Études Sci.* **129**, 83–127 (2019) Zbl 1423.22014 MR 3949028
- [23] Einsiedler, M., Lindenstrauss, E., Michel, P., Venkatesh, A.: [The distribution of closed geodesics on the modular surface, and Duke’s theorem](#). *Enseign. Math. (2)* **58**, 249–313 (2012) Zbl 1312.37032 MR 3058601
- [24] File, D., Martin, K., Pitale, A.: [Test vectors and central \$L\$ -values for \$GL\(2\)\$](#) . *Algebra Number Theory* **11**, 253–318 (2017) Zbl 1425.11104 MR 3641876
- [25] Ganguly, S., Humphries, P., Lin, Y., Nunes, R.: [Strong hybrid subconvexity for twisted selfdual \$GL_3\$ \$L\$ -functions](#). *arXiv:2408.00596v1* (2024)
- [26] Godement, R.: [Notes on Jacquet–Langlands’ theory \(with commentaries by Robert Langlands and Herve Jacquet\)](#). *CTM. Class. Top. Math.* **8**, Higher Education Press, Beijing (2018) Zbl 1416.11002 MR 3889963
- [27] Goldfeld, D., Hundley, J.: [Automorphic representations and \$L\$ -functions for the general linear group. Volume I](#). *Cambridge Stud. Adv. Math.* **129**, Cambridge University Press, Cambridge (2011) Zbl 1247.11070 MR 2807433
- [28] Gradshteyn, I. S., Ryzhik, I. M.: [Table of integrals, series, and products](#). 8th ed., Elsevier, Amsterdam (2015) Zbl 1300.65001 MR 3307944
- [29] Gross, B., Kohnen, W., Zagier, D.: [Heegner points and derivatives of \$L\$ -series. II](#). *Math. Ann.* **278**, 497–562 (1987) Zbl 0641.14013 MR 909238
- [30] Gross, B. H.: [Local orders, root numbers, and modular curves](#). *Amer. J. Math.* **110**, 1153–1182 (1988) Zbl 0675.12011 MR 970123
- [31] Harcos, G.: [Subconvex bounds for automorphic \$L\$ -functions and applications](#). D. Sc. Thesis, Hungarian Academy of Sciences, <https://www.renyi.hu/~gharcos/ertekezes.pdf>, visited on 21 August 2025 (2011)
- [32] Harcos, G., Michel, P.: [The subconvexity problem for Rankin–Selberg \$L\$ -functions and equidistribution of Heegner points. II](#). *Invent. Math.* **163**, 581–655 (2006) Zbl 1111.11027 MR 2207235
- [33] Hatcher, A.: [Algebraic topology](#). Cambridge University Press, Cambridge (2002) Zbl 1044.55001 MR 1867354
- [34] Hickerson, D. R.: [Length of period simple continued fraction expansion of \$\sqrt{d}\$](#) . *Pacific J. Math.* **46**, 429–432 (1973) Zbl 0236.40003 MR 321881
- [35] Holowinsky, R., Templier, N.: [First moment of Rankin–Selberg central \$L\$ -values and subconvexity in the level aspect](#). *Ramanujan J.* **33**, 131–155 (2014) Zbl 1290.11085 MR 3142436
- [36] Humphries, P.: [Equidistribution in shrinking sets and \$L^4\$ -norm bounds for automorphic forms](#). *Math. Ann.* **371**, 1497–1543 (2018) Zbl 1448.11084 MR 3831279
- [37] Humphries, P.: [Distributing points on the torus via modular inverses](#). *Q. J. Math.* **73**, 1–16 (2022) Zbl 1503.11106 MR 4395069

- [38] Humphries, P., Khan, R.: [On the random wave conjecture for dihedral Maaß forms](#). *Geom. Funct. Anal.* **30**, 34–125 (2020) Zbl [1462.11041](#) MR [4081056](#)
- [39] Humphries, P., Radziwiłł, M.: [Optimal small scale equidistribution of lattice points on the sphere, Heegner points, and closed geodesics](#). *Comm. Pure Appl. Math.* **75**, 1936–1996 (2022) Zbl [1504.11084](#) MR [4465906](#)
- [40] Iwaniec, H.: [Spectral methods of automorphic forms](#). 2nd ed., Grad. Stud. Math. 53, American Mathematical Society, Providence, RI (2002) Zbl [1006.11024](#) MR [1942691](#)
- [41] Jacquet, H., Chen, N.: [Positivity of quadratic base change \$L\$ -functions](#). *Bull. Soc. Math. France* **129**, 33–90 (2001) Zbl [1069.11017](#) MR [1871978](#)
- [42] Jacquet, H., Langlands, R. P.: [Automorphic forms on \$GL\(2\)\$](#) . *Lecture Notes in Math.* 114, Springer, Berlin (1970) Zbl [0236.12010](#) MR [401654](#)
- [43] Katok, S.: [Coding of closed geodesics after Gauss and Morse](#). *Geom. Dedicata* **63**, 123–145 (1996) Zbl [0884.20030](#) MR [1413625](#)
- [44] Katok, S., Sarnak, P.: [Heegner points, cycles and Maass forms](#). *Israel J. Math.* **84**, 193–227 (1993) Zbl [0787.11016](#) MR [1244668](#)
- [45] Khayutin, I.: [Joint equidistribution of CM points](#). *Ann. of Math. (2)* **189**, 145–276 (2019) Zbl [1431.11080](#) MR [3898709](#)
- [46] Kulkarni, R. S.: [An arithmetic-geometric method in the study of the subgroups of the modular group](#). *Amer. J. Math.* **113**, 1053–1133 (1991) Zbl [0758.11024](#) MR [1137534](#)
- [47] Lau, Y.-K., Tsang, K.-M.: [A mean square formula for central values of twisted automorphic \$L\$ -functions](#). *Acta Arith.* **118**, 231–262 (2005) Zbl [1081.11036](#) MR [2168765](#)
- [48] Linnik, Yu. V.: [Ergodic properties of algebraic fields](#). *Ergeb. Math. Grenzgeb.* 45, Springer, New York (1968) Zbl [0162.06801](#) MR [238801](#)
- [49] Liu, S.-C., Masri, R., Young, M. P.: [Subconvexity and equidistribution of Heegner points in the level aspect](#). *Compos. Math.* **149**, 1150–1174 (2013) Zbl [1329.11046](#) MR [3078642](#)
- [50] Martin, K., Whitehouse, D.: [Central \$L\$ -values and toric periods for \$GL\(2\)\$](#) . *Int. Math. Res. Not. IMRN* **2009**, 141–191 (2009) Zbl [1193.11046](#) MR [2471298](#)
- [51] Maskit, B.: [On Poincaré’s theorem for fundamental polygons](#). *Adv. Math.* **7**, 219–230 (1971) Zbl [0223.30008](#) MR [297997](#)
- [52] Michel, P.: [The subconvexity problem for Rankin–Selberg \$L\$ -functions and equidistribution of Heegner points](#). *Ann. of Math. (2)* **160**, 185–236 (2004) Zbl [1068.11033](#) MR [2119720](#)
- [53] Michel, P., Venkatesh, A.: [Equidistribution, \$L\$ -functions and ergodic theory: on some problems of Yu. Linnik](#). In: *International Congress of Mathematicians. Vol. II*, European Mathematical Society, Zürich, 421–457 (2006) Zbl [1157.11019](#) MR [2275604](#)
- [54] Nordentoft, A. C.: [Concentration of closed geodesics in the homology of modular curves](#). *Forum Math. Sigma* **11**, article no. e91 (2023) Zbl [1540.11051](#) MR [4651628](#)
- [55] Nordentoft, A. C., Tan, S. P.: [Equidistribution of partial coverings defined from closed geodesics](#). *arXiv:2502.08472v1* (2025)
- [56] Petrow, I., Young, M. P.: [A generalized cubic moment and the Petersson formula for newforms](#). *Math. Ann.* **373**, 287–353 (2019) Zbl [1446.11069](#) MR [3968874](#)
- [57] Popa, A. A.: [Central values of Rankin \$L\$ -series over real quadratic fields](#). *Compos. Math.* **142**, 811–866 (2006) Zbl [1144.11041](#) MR [2249532](#)
- [58] Popa, A. A.: [Whittaker newforms for archimedean representations of \$GL\(2\)\$](#) . *J. Number Theory* **128**, 1637–1645 (2008) Zbl [1146.11030](#) MR [2419183](#)
- [59] Schmidt, R.: [Some remarks on local newforms for \$GL\(2\)\$](#) . *J. Ramanujan Math. Soc.* **17**, 115–147 (2002) Zbl [0997.11040](#) MR [1913897](#)
- [60] Series, C.: [The modular surface and continued fractions](#). *J. Lond. Math. Soc. (2)* **31**, 69–80 (1985) Zbl [0545.30001](#) MR [810563](#)
- [61] Venkatesh, A.: [Sparse equidistribution problems, period bounds and subconvexity](#). *Ann. of Math. (2)* **172**, 989–1094 (2010) Zbl [1214.11051](#) MR [2680486](#)

- [62] Waldspurger, J.-L.: Sur les coefficients de Fourier des formes modulaires de poids demi-entier. *J. Math. Pures Appl.* (9) **60**, 375–484 (1981) Zbl [0431.10015](#) MR [646366](#)
- [63] Waldspurger, J.-L.: Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie. *Compos. Math.* **54**, 173–242 (1985) Zbl [0567.10021](#) MR [783511](#)
- [64] Young, M. P.: [Weyl-type hybrid subconvexity bounds for twisted \$L\$ -functions and Heegner points on shrinking sets](#). *J. Eur. Math. Soc. (JEMS)* **19**, 1545–1576 (2017) Zbl [1430.11067](#) MR [3635360](#)
- [65] Young, M. P.: [Explicit calculations with Eisenstein series](#). *J. Number Theory* **199**, 1–48 (2019) Zbl [1454.11083](#) MR [3926186](#)
- [66] Zagier, D.: The Rankin–Selberg method for automorphic functions which are not of rapid decay. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **28**, 415–437 (1981) Zbl [0505.10011](#) MR [656029](#)
- [67] Zhang, S.-W.: [Gross–Zagier formula for \$GL_2\$](#) . *Asian J. Math.* **5**, 183–290 (2001) Zbl [1111.11030](#) MR [1868935](#)