



RESEARCH ARTICLE

# Concentration of closed geodesics in the homology of modular curves

Asbjørn Christian Nordentoft 

LAGA, Institut Galilée, 99 avenue Jean Baptiste Clément, 93430 Villetaneuse, France; E-mail: [acnordentoft@outlook.com](mailto:acnordentoft@outlook.com).

**Received:** 3 April 2023; **Revised:** 12 June 2023; **Accepted:** 5 July 2023

**2020 Mathematics Subject Classification:** *Primary* – 11F67

## Abstract

We prove that the homology classes of closed geodesics associated to subgroups of narrow class groups of real quadratic fields concentrate around the Eisenstein line. This fits into the framework of Duke’s Theorem and can be seen as a real quadratic analogue of results of Michel and Liu–Masri–Young on supersingular reduction of CM-elliptic curves. We also study the level aspect, as well as a homological version of the sup norm problem. Finally, we present applications to group theory and modular forms

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Statement of results	2
1.1.1	Varying the level	4
1.2	Applications	5
1.3	Geometric interpretation	6
<b>2</b>	<b>Supersingular reduction of CM elliptic curves</b>	<b>7</b>
2.1	Supersingular reduction of elliptic curves of varying level	9
<b>3</b>	<b>Arithmetic background</b>	<b>10</b>
3.1	Real quadratic fields and closed geodesics	10
3.1.1	Closed geodesics and class groups	11
3.2	(Co)homology of modular curves	11
3.2.1	Hecke operators	12
<b>4</b>	<b>Background on Fuchsian groups</b>	<b>15</b>
4.1	Fundamental polygons	15
4.1.1	Bounding coordinates of side pairing transformations	17
4.2	Zagier’s fundamental polygon	18
4.3	Special fundamental polygons	19
<b>5</b>	<b>Proof of Theorem 1.1</b>	<b>20</b>
5.1	The Eisenstein contribution	21
5.2	The cuspidal contribution	23
<b>6</b>	<b>A homological version of the sup norm problem</b>	<b>24</b>
6.1	The method of proof	25
6.2	A lower bound for modular symbols	26

6.3 Reduction to a counting problem . . . . . 29  
 6.4 The counting argument . . . . . 31  
 6.5 Proof of Theorem 6.1 . . . . . 33  
**7 The level aspect** . . . . . **33**  
**8 Applications** . . . . . **34**  
 8.1 A group theoretic application . . . . . 34  
 8.2 An application to modular forms . . . . . 36

**1. Introduction**

Let  $p$  be a prime and consider the modular curve  $Y_0(p) = \Gamma_0(p)\backslash\mathbb{H}$  of level  $p$  equipped with the hyperbolic line element  $|dz|/y$  and volume element  $dx dy/y^2$ , where  $\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\}$  is the upper half-plane and  $\Gamma_0(p) \leq \text{PSL}_2(\mathbb{Z})$  is the level  $p$  Hecke congruence group<sup>1</sup> acting on  $\mathbb{H}$  by linear fractional transformations. Given a real quadratic field  $K$  of discriminant  $d_K > 0$  such that  $p$  splits in  $K$ , one can associate to an element  $A \in \text{Cl}_K^+$  of the narrow class group of  $K$  an oriented closed geodesic  $\mathcal{C}_A(p)$  on  $Y_0(p)$  (see Section 3.1 for details). In a celebrated paper [Duk88], Duke proved that the geodesics  $\{\mathcal{C}_A(p) \subset Y_0(p) : A \in \text{Cl}_K^+\}$  equidistribute with respect to hyperbolic measure as  $d_K$  tends to infinity, meaning that

$$\frac{\sum_{A \in \text{Cl}_K^+} \int_{\mathcal{C}_A(p)} f(z) \frac{|dz|}{y}}{\sum_{A \in \text{Cl}_K^+} \int_{\mathcal{C}_A(p)} 1 \frac{|dz|}{y}} \rightarrow \frac{1}{\text{vol}(Y_0(p))} \int_{Y_0(p)} f(z) \frac{dx dy}{y^2}, \quad d_K \rightarrow \infty, \tag{1.1}$$

for  $f : Y_0(p) \rightarrow \mathbb{C}$  smooth of compact support (for further results, see [ELMV12], [MV06] and the reference therein). In this paper, we study the *homological behavior* of the oriented closed geodesics – that is, the map

$$\text{Cl}_K^+ \rightarrow H_1(Y_0(p), \mathbb{Z}), \quad A \mapsto [\mathcal{C}_A(p)] := \text{class of } \mathcal{C}_A(p), \tag{1.2}$$

from the narrow class group to the integral homology of  $Y_0(p)$  (which one can identify with the abelianization of  $\Gamma_0(p)$  modulo torsion) as  $d_K \rightarrow \infty$ . Our result can be stated as saying that the classes concentrate around the *Eisenstein line*. This is a real quadratic analogue of the equidistribution of supersingular reduction of CM elliptic curves as in [Mic04] (see Section 2). We also study the level aspect (analogue of [LMY15]) leading to a homological version of the *sup norm problem* (see Section 6) which might be of independent interest (for another application, consult [HN22]). We also present applications of our distribution results – one group theoretic and another concerning nonvanishing of cycle integrals of modular forms (see Section 1.2). Finally, we refer to Section 1.3 and Remark 1.9 below for geometric interpretations of our results.

**1.1. Statement of results**

Let  $p$  be prime and let  $g$  denote the genus of the (noncompact) Riemann surface  $Y_0(p)$  satisfying  $g = \frac{p}{12} + O(1)$ . The integral homology

$$H_1(Y_0(p), \mathbb{Z}) \cong \mathbb{Z}^{2g+1}$$

sits as a lattice inside the real homology

$$V_p := H_1(Y_0(p), \mathbb{R}) \cong \Gamma_0(p)^{\text{ab}} \otimes \mathbb{R} \cong \mathbb{R}^{2g+1}. \tag{1.3}$$

<sup>1</sup>Here and throughout we will (contrary to standard conventions) consider the Hecke congruence groups  $\Gamma_0(p)$  as subgroups of  $\text{PSL}_2(\mathbb{R})$ .

We will be interested in how elements of  $V_p$  distribute when projected to the  $2g$ -sphere, by which we mean the map

$$V_p - \{0\} \rightarrow \mathbf{S}(V_p) := (V_p - \{0\})/\mathbb{R}_{>0}, \quad v \mapsto \bar{v}, \tag{1.4}$$

where we endow  $\mathbf{S}(V_p)$  with the quotient topology of the Euclidean topology of  $V_p$ . We define the *Eisenstein class*

$$v_E(p) \in H_1(Y_0(p), \mathbb{Z}) \tag{1.5}$$

as the homology class of a simple loop going once around the cusp at  $\infty$  with positive orientation (which corresponds to the image of the matrix  $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  in  $\Gamma_0(p)^{\text{ab}}$  using the identification (1.3)). This is indeed a Hecke eigenclass with the same eigenvalues as the weight 2 Eisenstein series, as we will see in Section 3.2.1. Our first main result is the following.

**Theorem 1.1.** *Let  $p$  be a prime and consider a real quadratic field  $K$  of discriminant  $d_K$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$ . Consider a subgroup  $H \leq \text{Cl}_K^+$  such that  $\mathfrak{p}_1 \notin H$  and  $(\sqrt{d_K}) \notin H$ . Then as  $d_K \rightarrow \infty$ , the classes of the closed geodesics associated to  $H$  concentrate around the line generated by the Eisenstein element, meaning that*

$$\overline{\sum_{A \in H} [C_A(p)]} \rightarrow \overline{-v_E(p)}, \quad \text{as } d_K \rightarrow \infty \tag{1.6}$$

in the quotient topology of  $\mathbf{S}(V_p)$ .

This is the real quadratic analogue of a distribution result due to Michel [Mic04] (see also [EOY05], [Yan08], [Kan09], [LMY15] and [ALMW22]) concerning the map

$$\text{Cl}_K \rightarrow \mathcal{E}\ell^{s,s}(\mathbb{F}_{p^2}) \tag{1.7}$$

from the class group of an imaginary quadratic field  $K$  (in which  $p$  is inert) to the isomorphism classes of supersingular elliptic curves defined over  $\mathbb{F}_{p^2}$  (or equivalently over  $\overline{\mathbb{F}_p}$ ). We refer to Section 2 for a more elaborate explanation of the analogy between the two cases. We note also that our results can be phrased as a weak convergence statement (see Theorem 5.1) which resembles (1.1). Another useful analogue to have in mind is the distribution of lattice points on the unit sphere  $S^2 \subset \mathbb{R}^3$  [Duk88]:

$$\frac{1}{\sqrt{d}} \{(a, b, c) \in \mathbb{Z}^3 : a^2 + b^2 + c^2 = d\} \subset S^2,$$

as  $d \rightarrow \infty$  (which is the basic case of Linnik’s Problem; see [MV06]); in both cases, one rescales the points to get a convergence of measures to, respectively, the Haar measure and the point measure at (minus) the Eisenstein element. We have the following useful corollary. In Section 1.3 below, we will use this to explain the geometric content of our result.

**Corollary 1.2.** *Let  $K$  and  $H \leq \text{Cl}_K^+$  be as in Theorem 1.1 and consider a basis  $B$  of  $V_p$  containing  $v_E(p)$ . Then for  $d_K$  sufficiently large, the  $v_E(p)$ -coordinate of the vector*

$$\sum_{A \in H} [C_A(p)] \in V_p$$

in the basis  $B$  has strictly maximal absolute value among all coordinates (and in particular is nonzero).

**Remark 1.3.** The conditions on the level and discriminant in the theorems above are necessary. If either  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$  with  $\mathfrak{p}_1 \in (\text{Cl}_K^+)^2$  or  $(\sqrt{d_K}) \in H$ , then there is a basis for  $V_p$  (the Hecke basis) such that the  $v_E(p)$ -coordinate of

$$\sum_{A \in (\text{Cl}_K^+)^2} [\mathcal{C}_A(p)]$$

is zero. In Section 3.1, we construct for each  $p$  an infinite family of real quadratic fields  $K$  that satisfy the conditions in Theorem 1.1. It is unclear whether the statement should be true for any genus.

**Remark 1.4.** The restriction to prime level ensures that there are no old forms and also that there is a unique Eisenstein class in the (co)homology. The main steps in the proofs should, however, work for general (square-free) level, but the statements would have to be modified accordingly.

**1.1.1. Varying the level**

Our second result is concerned with the level aspect in the sense that we will obtain a distribution statement uniform in  $p$ . Notice, first of all, that for  $v_0, v_1, \dots \in V_p - \{0\}$ , the convergence

$$\overline{v_n} \longrightarrow \overline{v_0} \in \mathbf{S}(V_p) \text{ as } n \rightarrow \infty$$

is equivalent to

$$\left\| \frac{v_n}{\|v_n\|} - \frac{v_0}{\|v_0\|} \right\| \rightarrow 0 \text{ as } n \rightarrow \infty \tag{1.8}$$

for any (fixed) norm  $\| \cdot \|$  of  $V_p$ . A natural question is to ask for bounds for the left-hand side of (1.8) uniform in  $d_K$  and  $p$  for certain specific norms of  $V_p$ . A basis  $B$  for  $H_1(Y_0(p), \mathbb{R})$  defines an isomorphism of vector spaces  $V_p \cong \mathbb{R}^{2g+1}$  (by mapping  $B$  to the standard basis of  $\mathbb{R}^{2g+1}$ ), and by pulling back the  $L^r$ -norm for  $1 \leq r \leq \infty$ , we obtain a norm on  $V_p$  which we denote by  $\| \cdot \|_{B,r}$  (see (7.3) for details). We notice that one can choose a sequence of bases for each  $p$  such that the convergence in (1.8) (with say  $r = 1$ ) is arbitrarily slow in  $p$ . This is parallel to the case of the distribution of CM points on modular curves in the level aspect as considered in [LMY13]; here, one has to consider ‘compatible’ test functions as the level varies. Similarly, we will consider certain ‘compatible’ bases of  $H_1(Y_0(p), \mathbb{R})$ . To define these, we recall (see Section 4.2) that the following matrices generate  $\Gamma_0(p)$ ;

$$\mathcal{S}(p) := \left\{ \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} a & -(aa^*+1)/p \\ p & -a^* \end{pmatrix} : 0 < a < p, \right\}, \tag{1.9}$$

where  $0 < a^* < p$  is such that  $aa^* \equiv -1 \pmod p$ . We say that a basis  $B$  of  $H_1(Y_0(p), \mathbb{R})$  is a *basic basis of level  $p$*  if it consists of homology classes containing the oriented geodesic connecting  $i \in \mathbb{H}$  and  $\sigma i \in \mathbb{H}$  for some  $\sigma \in \mathcal{S}(p)$ . We think of these bases as analogues of the sets  $\mathcal{E}l\ell^{ss}(\mathbb{F}_{p^2})$  considered in the imaginary quadratic case (1.7) (see Section 2). Our second main result is the following.

**Theorem 1.5.** *Let  $p$  be prime and  $B$  a basic basis of level  $p$  with associated norm  $\| \cdot \|_{B,\infty}$  (note that  $\|v_E(p)\|_{B,\infty} = 1$ ). Let  $K$  be a real quadratic field of discriminant  $d_K$  with no unit of norm  $-1$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$  and  $\mathfrak{p}_1 \notin (\text{Cl}_K^+)^2$ . Then we have*

$$\left\| \frac{\sum_{A \in (\text{Cl}_K^+)^2} [\mathcal{C}_A(p)]}{\| \sum_{A \in (\text{Cl}_K^+)^2} [\mathcal{C}_A(p)] \|_{B,\infty}} + v_E(p) \right\|_{B,\infty} \ll_\varepsilon p^{2+\varepsilon} d_K^{-1/12+\varepsilon}. \tag{1.10}$$

As above, we get the following corollary.

**Corollary 1.6.** *Let  $p$  and  $K$  be as above and consider a basic basis  $B \subset V_p$  of level  $p$  containing  $v_E(p)$ . Then for  $d_K \gg_\varepsilon p^{24+\varepsilon}$ , the  $v_E(p)$ -coordinate of the vector*

$$\sum_{A \in (\text{Cl}_K^+)^2} [\mathcal{C}_A(p)] \in V_p$$

*in the basis  $B$  has strictly maximal absolute value among all coordinates (and in particular is nonzero).*

The above result has a very close analogue in the imaginary quadratic case as worked out by Liu–Masri–Young [LMY15], who studied a level  $p$  version of the equidistribution of the map (1.7). One can identify the finite set  $\mathcal{E}\ell\ell^{ss}(\mathbb{F}_{p^2}) = \{e_1, \dots, e_n\}$  with the connected components of a certain conic curve  $X^{p,\infty}$  defined from the quaternion algebra over  $\mathbb{Q}$  ramified at  $p$  and  $\infty$ . Thus,  $\{e_1, \dots, e_n\}$  defines a basis for the 0-th homology group  $H_0(X^{p,\infty}, \mathbb{Z})$ . In this language, the results of [LMY15] can be phrased exactly as (1.10) (see (2.5)). We will develop this analogy in greater detail in Section 2.

### 1.2. Applications

We will now present some applications of our results; one is group theoretic and the other has to do with nonvanishing of cycle integrals of modular forms.

Consider a prime  $p \equiv -1 \pmod{12}$  (for simplicity) and put  $n = \frac{p+1}{3}$ . Then  $\Gamma_0(p)$  considered as a subgroup of  $\mathrm{PSL}_2(\mathbb{Z})$  is torsion-free with  $\Gamma_0(p)^{\mathrm{ab}} \cong \mathbb{Z}^{n/2+1}$ , and thus we know by the Kurosh subgroup theorem that  $\Gamma_0(p)$  is a free group on  $n/2 + 1$  generators (being a subgroup of  $\mathrm{PSL}_2(\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z}$ ). Let

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

be a Farey symbol of level  $p$  in the terminology of Kulkarni [Kul91], meaning that  $a_i/b_i$  are reduced fractions such that  $a_i b_{i+1} - a_{i+1} b_i = 1$  for all  $1 \leq i < n$  and that there is a pairing  $i \leftrightarrow i^*$  on  $0, \dots, n-1$  satisfying

$$b_i b_{i^*} + b_{i+1} b_{i^*+1} \equiv 0 \pmod{p}.$$

Such a symbol always exists, even one that is symmetric around  $1/2$ , by [Kul91, Section 13]. It follows from [Kul91] and a classical result of Poincaré that  $\Gamma_0(p)$  is freely generated by (the images inside  $\mathrm{PSL}_2(\mathbb{Z})$  of)  $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  together with the  $n/2 = \frac{p+1}{6}$  matrices

$$\begin{pmatrix} a_{i^*+1} b_{i+1} + a_{i^*} b_i & -a_i a_{i^*} - a_{i+1} a_{i^*+1} \\ b_i b_{i^*} + b_{i+1} b_{i^*+1} & -a_{i+1} b_{i^*+1} - a_i b_{i^*} \end{pmatrix} \text{ for } i < i^* \text{ a pair.} \tag{1.11}$$

The following group theoretic application can be thought of as an analogue of Linnik’s Theorem on the smallest prime in arithmetic progressions (see also Theorem 2.3 below).

**Corollary 1.7.** *Let  $p \equiv -1 \pmod{12}$  be prime and consider a real quadratic field  $K$  of discriminant  $d_K$  and (wide) class number one such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1 \mathfrak{p}_2$  such that  $\mathfrak{p}_1$  does not have a generator of positive norm. Let  $(u, v)$  be the positive half-integer solution to  $u^2 - d_K v^2 = 1$  such that  $v$  is minimal among all such and let  $a, b, c \in \mathbb{Z}$  satisfy  $b^2 - 4ac = d_K$  and  $p|a$ . Then for  $d_K \gg_\varepsilon p^{24+\varepsilon}$ , the matrix  $\begin{pmatrix} u+bv & 2cv \\ -2av & u-bv \end{pmatrix} \in \Gamma_0(p)$  is not contained in the subgroup generated by the matrices (1.11).*

The conjugacy class in  $\Gamma_0(p)$  of the matrix  $\begin{pmatrix} u+bv & 2cv \\ -2av & u-bv \end{pmatrix}$  in the corollary above corresponds exactly to one of the two oriented closed geodesic associated to the class group of  $K$  (see Section 3.1 for details). We also obtain a related result when the wide class number of  $K$  is not one and for general prime levels  $p$ , which is, however, a bit more cumbersome to state. We will refer to Section 8 for details.

The next application is concerned with the nonvanishing of integrals of modular forms over closed geodesics. Let  $\sigma_1(n) = \sum_{d|n} d$  be the sum of divisors function. For a Hecke eigenform  $f \in \mathcal{M}_2(p)$  with Fourier coefficients  $a_f(n)$  (at  $\infty$ ), we have the trivial bound  $|a_f(n)| \ll \sigma_1(n)$ . This means that for any nonzero modular form  $f \in \mathcal{M}_2(p)$ , we can define

$$M_f := \inf\{c \geq 0 : |a_f(n)| \leq c\sigma_1(n), \forall n \geq 1\} \in (0, \infty),$$

where again,  $a_f(n)$  denotes the Fourier coefficients of  $f$  at  $\infty$ .

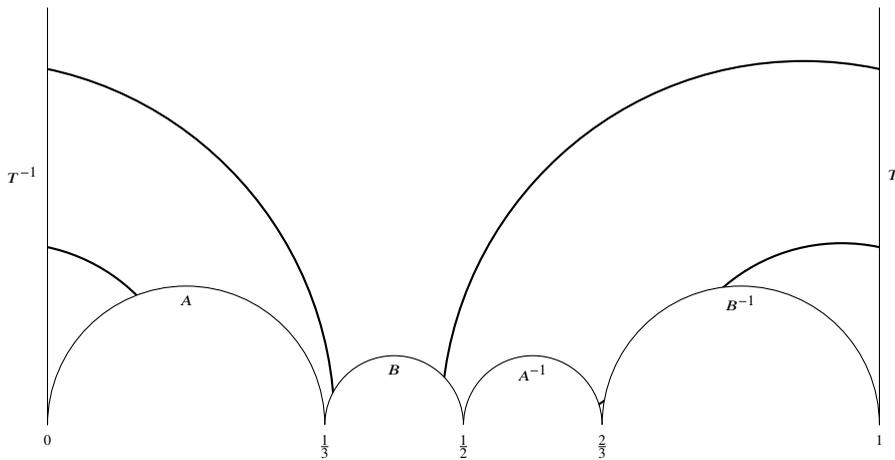


Figure 1. The closed geodesic on  $Y_0(11)$  associated to the principal class in  $\mathbb{Q}(\sqrt{23})$ .

**Corollary 1.8.** Let  $p$  be prime and let  $f \in \mathcal{M}_2(p)$  be a holomorphic modular form of weight 2 and level  $p$  with constant Fourier coefficient equal to 1. Consider a real quadratic field  $K$  of discriminant  $d_K$  and (wide) class number one such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$  such that  $\mathfrak{p}_1$  does not have a generator of positive norm. Let  $C$  denote the geodesic associated to the class group of  $K$ . Then we have for  $d_K \gg_\varepsilon (M_f)^{12+\varepsilon} p^{48+\varepsilon}$  that

$$\int_C f(z)dz \neq 0.$$

Notice that the above does not depend on the choice of orientation of  $C$ .

### 1.3. Geometric interpretation

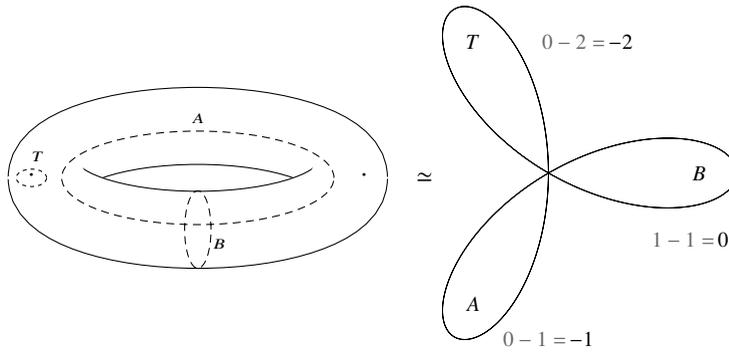
We will now explain the geometric content of Theorem 1.1. Recall that topologically,  $Y_0(p)$  is a genus  $g$  curve with two punctures where the genus satisfies  $g = \frac{p}{12} + O(1)$ . We are interested in understanding the sum of the homology classes of the oriented closed geodesics of the principal genus. The associated geodesics will travel around  $Y_0(p)$  in a complicated way but our results can be interpreted as saying that

‘closed geodesics from the principal genus wind around the cusp at infinity a lot’.

To illustrate this, let us consider the simplest nontrivial case  $p = 11$  where the genus is one. A fundamental polygon for  $Y_0(11)$  is given by the hyperbolic polygon with vertices  $\infty, 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1$  as illustrated in Figure 1. The associated side pairing transformations (forgetting inverses) are

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, A = \begin{pmatrix} 3 & -2 \\ 11 & -7 \end{pmatrix}, B = \begin{pmatrix} 4 & -3 \\ 11 & -8 \end{pmatrix},$$

which define a set of free generators for  $\Gamma_0(11)$  (this follows from Poincaré’s Theorem as explained in Section 4). As illustrated in Figure 2, when viewing  $Y_0(11)$  as a double punctured torus, the matrix  $T$  corresponds to a simple loop around the puncture at  $\infty$ , and  $A, B$  correspond to the loops going around the two ‘holes’ of the torus. The content of Theorem 1.1 is concerned with the coordinates of the closed geodesics in the basis  $\{T, A, B\}$  in the homology (or equivalently, when the corresponding



**Figure 2.** The  $\{T, A, B\}$ -coordinates in the homology of  $Y_0(11)$  of the closed geodesic associated to the principal class of  $\mathbb{Q}(\sqrt{23})$ .

hyperbolic conjugacy classes are projected to the abelianization of  $\Gamma_0(11)$ . Now consider the quadratic field  $K = \mathbb{Q}(\sqrt{23})$  which has narrow class number two and wide class number one. In this case, we have associated to the principal class  $I \in \text{Cl}_K^+$  the conjugacy class of the following matrix:

$$\gamma_I := \begin{pmatrix} 26 & -35 \\ 55 & -74 \end{pmatrix},$$

and one can check that we have

$$\gamma_I = BA^{-1}T^{-1}B^{-1}T^{-1}$$

in our basis (see Figure 1). Observe that  $Y_0(11)$  is homotopic to a wedge of three circles, and we are counting the (oriented) number of times the geodesic goes around each of the three circles as illustrated in Figure 2. We already see in this numerically very small example a tendency towards large  $T$ -coordinate.

**Remark 1.9.** Our results can be interpreted in the context of the classical problem of understanding the distribution of the projection

$$\pi_1(M) \twoheadrightarrow \text{Conj}(\pi_1(M)) \twoheadrightarrow \pi_1(M)^{\text{ab}} \cong H_1(M, \mathbb{Z}) \tag{1.12}$$

for a manifold  $M$ . Note that if  $\Gamma_0(p)$  is torsion free, then we have  $\pi_1(Y_0(p)) \cong \Gamma_0(p)$  and  $H_1(Y_0(p), \mathbb{Z}) \cong \Gamma_0(p)^{\text{ab}}$ . For  $M$  a compact Riemann surface, there is a 1-to-1 correspondence between conjugacy classes in  $\pi_1(M)$  and oriented closed geodesics. Phillips and Sarnak [PS87] obtained an asymptotic expansion for the number of primitive geodesics of length  $\leq X$  with specified image under the map (1.12). In the same setup, Petridis and Risager [PR08] obtained an equidistribution statement for subsets  $A \subset H_1(M, \mathbb{Z})$  with asymptotic density. Also, Petridis and Risager [PR05] showed that given a splitting  $H_1(M, \mathbb{Z}) = \mathbb{Z}v \oplus V$  with  $v \in H_1(M, \mathbb{Z})$ , the  $v$ -coordinate of closed geodesics become normally distributed (when properly normalized) again ordered by the length of the geodesics. See also [NC20], [Bv22]. From an arithmetic point of view, the ordering by geodesic length is not very natural as it gives very large weight to discriminants with a large class group (compared to the discriminant), whereas the ordering by discriminant does not seem to admit any nice geometric description.

## 2. Supersingular reduction of CM elliptic curves

The result of Duke [Duk88] mentioned in the introduction regarding equidistribution of closed geodesics has an imaginary quadratic analogue, which amounts to the equidistribution of elliptic curves with complex multiplication inside the moduli space of elliptic curves over  $\mathbb{C}$  (i.e., CM points on the modular

curve; see again [Duk88]). Similarly, our results can be seen as a real quadratic analogue of the distribution of the supersingular reduction of elliptic curves with complex multiplication as investigated by many authors [Mic04], [EOY05], [Yan08], [Kan09], [LMY15], [ALMW22]. This analogy between oriented closed geodesics in homology and supersingular reduction of CM elliptic curves is very natural and appears, for example, in the recent work of Darmon–Harris–Rotger–Venkatesh [DHRV21].

Let  $\mathcal{E}ll^{ss}(\mathbb{F}_{p^2}) = \{e_1, \dots, e_n\}$  denote the set of isomorphism classes of supersingular elliptic curves defined over  $\mathbb{F}_{p^2}$ . It is known that  $|\mathcal{E}ll^{ss}(\mathbb{F}_{p^2})| = \frac{p}{12} + O(1)$ . We endow  $\mathcal{E}ll^{ss}(\mathbb{F}_{p^2})$  with the measure defined by

$$\mu_p(\{e_i\}) = \frac{w_i^{-1}}{\sum_{j=1}^n w_j^{-1}}, \tag{2.1}$$

where  $w_i$  denotes the size of the endomorphism group of the elliptic curves corresponding to  $e_i$ . Let  $\mathcal{E}ll_K$  denote the set of isomorphism classes of elliptic curve (defined over  $\overline{\mathbb{Q}}$ ) with complex multiplication by the ring of integers of the imaginary quadratic field  $K$  with discriminant  $d_K < 0$  and class group  $Cl_K$ . The set  $\mathcal{E}ll_K$  carries a natural  $Cl_K$ -action which is free and transitive. If  $p$  is inert in  $K$ , we have a map

$$r_{\mathfrak{p}} : \mathcal{E}ll_K \rightarrow \mathcal{E}ll^{ss}(\mathbb{F}_{p^2})$$

given by taking the mod  $\mathfrak{p}$  reduction where  $\mathfrak{p}$  is a prime ideal of the Hilbert class field of  $K$  lying over  $p$ . The following is [Mic04, Theorem 3].

**Theorem 2.1** (Michel). *Consider a CM elliptic curve  $E \in \mathcal{E}ll_K$  and a subgroup  $H \leq Cl_K$  of index  $\leq |d_K|^{1/2015}$ . Then the orbits  $H.E = \{r_{\mathfrak{p}}(A.E) \in \mathcal{E}ll^{ss}(\mathbb{F}_{p^2}) : A \in H\}$  become equidistributed as  $d_K \rightarrow -\infty$  with respect to the measure  $\mu_p$  given by (2.1).*

The proof goes through an identification of  $\mathcal{E}ll^{ss}(\mathbb{F}_{p^2})$  with the set of connected components of a certain conic curve

$$X^{p,\infty} := \mathbf{PB}^\times(\mathbb{Q}) \backslash \mathbf{PB}^\times(\mathbb{A}_{\mathbb{Q}}) / \mathbf{PB}^\times(\widehat{\mathbb{Z}})K_\infty, \tag{2.2}$$

where  $\mathbf{B}$  denotes the unique quaternion algebra over  $\mathbb{Q}$  ramified at  $p$  and  $\infty$ ,  $\mathbf{PB}^\times$  is its group of projective units,  $K_\infty$  denotes a maximal compact torus of  $\mathbf{PB}^\times(\mathbb{R})$  and  $\mathbf{PB}^\times(\widehat{\mathbb{Z}})$  denotes the projective units of  $\widehat{\mathbb{Z}} \otimes \mathcal{O}$  where  $\mathcal{O} \subset \mathbf{B}$  is a maximal order. Using the Jacquet–Langlands correspondence and a formula of Gross, this reduces the distribution problem to subconvexity bounds of certain Rankin–Selberg  $L$ -functions which is resolved (see [Mic04, Section 5] for details).

In order to set up the analogy with the real quadratic case, we will slightly reformulate the statement in Theorem 2.1 above. Denote by  $H_0(X^{p,\infty}, \mathbb{Z})$  the 0-th (singular) homology group of  $X^{p,\infty}$  with integral coefficients (one should picture a copy of  $\mathbb{Z}$  at each connected component of  $X^{p,\infty}$ ), which is a lattice inside the real homology group  $H_0(X^{p,\infty}, \mathbb{R})$ . Note that both of these abelian groups carry a natural action of the Hecke algebra coming from the description in terms of quaternion algebras (see, for example, [BD96, Section 1.5]), and as such is isomorphic to the space of modular forms  $\mathcal{M}_2(p)$  of level  $p$  and weight 2. We have a natural basis of  $H_0(X^{p,\infty}, \mathbb{Z})$  of geometric nature corresponding to the classes  $e_1, \dots, e_n$  (using suggestive notation) associated to each connected component of  $X^{p,\infty}$ . In this basis, the Eisenstein element is the following:

$$e_0 := \sum_{i=1}^n w_i^{-1} e_i,$$

meaning that  $T_\ell e_0 = (\ell + 1)e_0$  for  $\ell \neq p$  prime and  $T_\ell$  the  $\ell$ -th Hecke operator. Using the above identifications with supersingular elliptic curves and after fixing an elliptic curve  $E \in \mathcal{E}ll_K$ , we get a

map

$$r_p : \text{Cl}_K \rightarrow \{e_1, \dots, e_n\} \subset H_0(X^{p,\infty}, \mathbb{Z}) \subset H_0(X^{p,\infty}, \mathbb{R}), \quad A \mapsto r_p(A.E), \quad (2.3)$$

which will serve as an imaginary quadratic analogue of the map (1.2). We will now consider Theorem 2.1 as a statement about convergence (with respect to the standard topology) on the  $(n - 1)$ -sphere which we identify with  $\mathbf{S}(V_{p,\infty}) := (V_{p,\infty} - \{0\})/\mathbb{R}_{>0}$  equipped with the quotient topology where  $V_{p,\infty} := H_0(X^{p,\infty}, \mathbb{R}) \cong \mathbb{R}^n$  (equipped with the Euclidean topology). As above for  $v \in V_{p,\infty} - \{0\}$ , we denote by  $\bar{v} \in \mathbf{S}(V_{p,\infty})$  the image under the natural projection  $V_{p,\infty} - \{0\} \rightarrow \mathbf{S}(V_{p,\infty})$ . We can then recast the equidistribution statement of Michel as follows.

**Theorem 2.2** (‘Vector space’-version of Theorem 2.1). *Let  $p > 2$  be prime and let  $K$  be an imaginary quadratic field of discriminant  $d_K < 0$  such that  $p$  is inert in  $K$ . Consider a subgroup  $H \leq \text{Cl}_K$  of index  $\leq |d_K|^{1/2015}$  and a coset  $CH \subset \text{Cl}_K$ . Then we have as  $d_K \rightarrow -\infty$  that*

$$\overline{\sum_{A \in CH} r_p(A)} \longrightarrow \bar{e}_0 \quad (2.4)$$

in the standard topology of  $\mathbf{S}(V_{p,\infty})$ .

We will now show that this is equivalent to Theorem 2.1. Let  $B = \{e_1, \dots, e_n\}$  be the standard basis for  $V_{p,\infty}$  which defines an isomorphism  $V_{p,\infty} \cong \mathbb{R}^n$ . As explained above, we get a norm  $\|\cdot\|_{B,1}$  on  $V_{p,\infty}$  by pulling back the  $L^1$ -norm in the standard basis of  $\mathbb{R}^n$ . Now recall that the convergence on the  $(n - 1)$ -sphere  $\mathbf{S}(V_{p,\infty})$  is equivalent to the convergence statement (1.8) using, for example, the norm  $\|\cdot\|_{B,1}$ . Notice that we have

$$\left\| \sum_{A \in CH} r_p(A) \right\|_{B,1} = |H|, \quad \|e_0\|_{B,1} = \sum_{i=1}^n w_i^{-1},$$

recalling that the  $e_i$ -coordinate of  $e_0$  is equal to  $w_i^{-1}$ . Now the convergence in (1.8) implies that the  $e_i$ -coordinates of  $\sum_{A \in CH} r_p(A)$  converge to those of  $e_0$  (both normalized). This recovers the statement of Michel as in Theorem 2.1 up to the fact that Theorem 2.2 does not see that  $r_p(A)$  is always equal to one of the vectors  $e_1, \dots, e_n$ .

### 2.1. Supersingular reduction of elliptic curves of varying level

It is natural to ask what happens if we let  $p$  vary with  $d_K$  as has been considered by Liu–Masri–Young [LMY15]. In the above terminology, they consider the convergence with respect to the basis corresponding to the connected components  $\{e_1, \dots, e_n\}$  of  $X^{p,\infty}$ . Then [LMY15, Theorem 1.1] amounts to the following:

$$\left\| \frac{\sum_{A \in \text{Cl}_K} r_p(A)}{\|\sum_{A \in \text{Cl}_K} r_p(A)\|_{B,1}} - \frac{e_0}{\|e_0\|_{B,1}} \right\|_{B,\infty} \ll_\varepsilon p^{1/8+\varepsilon} |d_K|^{-1/16+\varepsilon}, \quad (2.5)$$

where again,  $\|\cdot\|_{B,1}$  and  $\|\cdot\|_{B,\infty}$  denote the pullback of, respectively, the  $L^1$ -norm and sup norm under the isomorphism  $V_{p,\infty} \cong \mathbb{R}^n$  defined by  $B = \{e_1, \dots, e_n\}$ . We note that the statement (2.5) is exactly parallel to Theorem 1.5. Translating back to the language of elliptic curves, (2.5) implies the following analogue of Linnik’s Theorem on the smallest prime in arithmetic progressions.

**Theorem 2.3** (Liu–Masri–Young). *The reduction map  $r_q : \mathcal{E}\ell_K \rightarrow \mathcal{E}\ell_{p^2}^{SS}$  is surjective for  $|d_K| \gg_\varepsilon p^{18+\varepsilon}$ .*

We think of Corollary 1.7 (and more generally Corollary 8.1 below) as a real quadratic analogue of the above.

### 3. Arithmetic background

In this section, we will introduce some basic facts about, respectively, oriented closed geodesics associated to class groups of real quadratic fields and (co)homology of modular curves.

#### 3.1. Real quadratic fields and closed geodesics

We will refer to [Pop06] and [Dar94] for an in-depth account of the following material. Let  $K$  be a real quadratic extension of  $\mathbb{Q}$  of discriminant  $d_K > 0$ . Let  $p$  be a prime which splits in  $K$  and fix throughout a residue  $r \pmod{2p}$  such that  $r^2 \equiv d_K \pmod{4p}$ . Then we have the following equality of ideals:

$$p\mathcal{O}_K = \left[ p, \frac{r - \sqrt{d_K}}{2} \right] \left[ p, \frac{r + \sqrt{d_K}}{2} \right],$$

where we use the following notation for  $\alpha, \beta \in K$ :

$$[\alpha, \beta] := \mathbb{Z}\alpha + \mathbb{Z}\beta \subset K. \tag{3.1}$$

In other words,  $[p, \frac{r \pm \sqrt{d_K}}{2}]$  are the two prime ideals of  $\mathcal{O}_K$  lying over  $p$ .

We let  $\text{Cl}_K^+$  denote the narrow class group of  $K$  (i.e., fractional ideals modulo principal ideals generated by elements of positive norms). Below, when the discriminant  $d_K$  is clear from context, we let  $I \in \text{Cl}_K^+$  denote the class containing the principal fractional ideal  $(1) = \mathcal{O}_K \subset K$  and  $J \in \text{Cl}_K^+$  denote the class containing the different  $(\sqrt{d_K}) = \sqrt{d_K}\mathcal{O}_K \subset K$ . Observe that for fundamental discriminants divisible by a prime  $q \equiv 3 \pmod{4}$ , there exists no unit with norm  $-1$  (as  $-1$  is not a quadratic residue modulo  $q$ ) which implies that  $J \neq I$ . In this case, a principal ideal belongs to the class  $J$  exactly if it has a generator with negative norm.

We will now show the following result, which is closely related to the conditions appearing in our main results as we will see below.

**Proposition 3.1.** *Let  $p$  be an odd prime. Then there exists infinitely many positive fundamental discriminants  $d$  such that  $J \neq I$  and  $[p, \frac{r - \sqrt{d}}{2}] \in J$  inside  $\text{Cl}_K^+$  with  $K = \mathbb{Q}(\sqrt{d})$ .*

*Proof.* Pick an odd prime  $q \equiv 3 \pmod{4}$  such that  $-p$  is a quadratic residue mod  $q$ . Then by simple considerations about quadratic residues, there is a residue  $t \pmod{8q}$  such that

$$q|t^2 + p, \quad q^2 \nmid t^2 + p \quad \text{and} \quad (t^2 + p \pmod{16}) \in \{1, 5, 8, 9, 12, 13\},$$

meaning in particular that  $t^2 + p \equiv 0$  or  $1 \pmod{4}$ . Now for each  $n \equiv t \pmod{8q}$ , there is a unique positive fundamental discriminant  $d > 0$  and integer  $m > 0$  such that  $dm^2 = n^2 + p$ . By construction, we have  $q|d$ , and thus  $J \neq I$  inside  $\text{Cl}_K^+$  where  $K = \mathbb{Q}(\sqrt{d})$ . Furthermore, we have the factorization

$$-p = (n - \sqrt{dm})(n + \sqrt{dm}),$$

which means that  $[p, \frac{r - \sqrt{d}}{2}]$  is a principal ideal generated by  $n \pm \sqrt{dm}$  (for some choice of sign  $\pm$ ). By construction, the norm of  $n \pm \sqrt{dm}$  is negative (and equal to  $-p$ ). By the above, this implies  $[p, \frac{r - \sqrt{d}}{2}] \in J$  in  $\text{Cl}_K^+$  as wanted.  $\square$

Now it follows that if  $J \neq I$ , then  $J \notin (\text{Cl}_K^+)^2$ . Thus, for  $p, d$  as in Proposition 3.1 and each subgroup  $H \leq (\text{Cl}_K^+)^2$ , we have  $[p, \frac{r - \sqrt{d}}{2}] \notin H$ . This gives plenty of examples for which Theorems 1.1 and 1.5 apply, recalling that  $\{\mathfrak{p}_1, \mathfrak{p}_2\} = \{[p, \frac{r \pm \sqrt{d}}{2}]\}$ .

**3.1.1. Closed geodesics and class groups**

Let  $p$  be prime and  $K$  a real quadratic field of discriminant  $d_K$  such that  $p$  splits in  $K$ . Let  $r \pmod{2p}$  as above be such that  $r^2 \equiv d_K \pmod{4p}$ . Denote by  $\mathcal{Q}_{K,p}$  (suppressing  $r$  in the notation) the set of all integral binary quadratic form

$$Q(x, y) = ax^2 + bxy + cy^2$$

of discriminant  $b^2 - 4ac = d_K$  and level  $p$ , meaning that  $a \equiv 0 \pmod{p}$  and  $b \equiv r \pmod{2p}$ . The group  $\Gamma_0(p)$  acts naturally on  $\mathcal{Q}_{K,p}$  by coordinate transformation. It is a classical fact, essentially due to Gauß, that there is a natural bijection (depending on the choice of  $r \pmod{p}$ )

$$\text{Cl}_K^+ \xrightarrow{\sim} \Gamma_0(p) \backslash \mathcal{Q}_{K,p}. \tag{3.2}$$

When the level is trivial, the above bijection is induced by mapping  $ax^2 + bxy + cy^2$  to the narrow ideal class of the fractional ideal  $[1, \frac{-\text{sign}(a)b + \sqrt{d_K}}{2|a|}]$  (using the notation (3.1)).

Given an integral binary quadratic form  $Q(x, y) = ax^2 + bxy + cy^2$  of discriminant  $d_K$  and level  $p$ , we associate the following matrix:

$$\gamma_Q := \begin{pmatrix} u + bv & 2cv \\ -2av & u - bv \end{pmatrix} \in \Gamma_0(p), \tag{3.3}$$

where  $u, v$  are positive half-integers satisfying Pell’s equation  $u^2 - d_K v^2 = 1$  and such that  $v$  is minimal among all such solutions (i.e., the fundamental positive unit of  $K$  is  $\epsilon_K = u + v\sqrt{d_K}$ ).

For  $A \in \text{Cl}_K^+$ , we denote by  $\mathcal{C}_A(p)$  the oriented closed geodesic on  $Y_0(p)$  obtained by projecting the oriented geodesic connecting  $z_Q$  and  $\gamma_Q z_Q$ , where

$$z_Q := \frac{-\text{sign}(a)b + i\sqrt{d_K}}{2|a|}$$

and  $Q \in \mathcal{Q}_{K,p}$  corresponds to  $A$  under the isomorphism (3.2) (the image in  $Y_0(p)$  is independent of the choice of representative  $Q$ ).

**3.2. (Co)homology of modular curves**

Here and throughout, we will consider matrices in  $\text{SL}_2(\mathbb{R})$  as elements of  $\text{PSL}_2(\mathbb{R})$  without further mentioning. Let  $p$  be prime and consider the Hecke congruence group (or more precisely, its projection to  $\text{PSL}_2(\mathbb{R})$ )

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

Let

$$Y_0(p) := \Gamma_0(p) \backslash \mathbb{H}, \quad X_0(p) := \overline{Y_0(p)} = Y_0(p) \cup \Gamma_0(p) \backslash \mathbb{P}^1(\mathbb{Q})$$

be resp., the modular curve of level  $p$  and its compactification which is a compact Riemann surface of genus  $g = \frac{p}{12} + O(1)$  (see, for example, [Shi94, Proposition 1.40]). We can consider the integral singular homology group [Hat02, Chapter 2]

$$H_1(Y_0(p), \mathbb{Z}) \cong \mathbb{Z}^{2g+1},$$

which sits as a lattice inside the real homology

$$H_1(Y_0(p), \mathbb{R}) \cong \mathbb{R}^{2g+1}.$$

We will be interested in the distribution of oriented closed geodesics inside the lattice  $H_1(Y_0(p), \mathbb{Z})$ .

We have the cap product pairing

$$\langle \cdot, \cdot \rangle : H_1(Y_0(p), \mathbb{R}) \times H^1(Y_0(p), \mathbb{R}) \rightarrow \mathbb{R} \tag{3.4}$$

between real homology and cohomology which identifies  $H^1(Y_0(p), \mathbb{R})$  with the linear dual  $H_1(Y_0(p), \mathbb{R})^*$ . Given a basis  $B$  of  $H_1(Y_0(p), \mathbb{R})$ , we denote by  $B^* \subset H^1(Y_0(p), \mathbb{R})$  the dual basis of  $B$  with respect to the cap product pairing as in (3.4).

Recall that the de Rham isomorphism gives a description of  $H^1(Y_0(p), \mathbb{R})$  in terms of real valued harmonic 1-forms on  $Y_0(p)$ . Any such 1-form is a linear combination of forms of the type

$$\omega_f = 2\pi i f(z) dz, \quad \overline{\omega_f} = \overline{2\pi i f(z) dz},$$

where  $f \in \mathcal{M}_2(p)$  is a weight 2 and level  $p$  holomorphic form (not necessarily cuspidal). We also have a surjective map

$$\Gamma_0(p) \twoheadrightarrow H_1(Y_0(p), \mathbb{Z}), \quad \gamma \mapsto \{z, \gamma z\}, \tag{3.5}$$

where  $z \in \mathbb{H}$  (the class does not depend on the choice of  $z$ ), and we are using the following notation for  $z_1, z_2 \in \mathbb{H}$ :

$$\{z_1, z_2\} := [\text{class of the oriented geodesic connecting } z_1 \text{ and } z_2] \in H_1(Y_0(p), \mathbb{R}), \tag{3.6}$$

which defines an element of the real homology  $H_1(Y_0(p), \mathbb{R})$  via integration against 1-forms. The map (3.5) induces an isomorphism

$$\Gamma_0(p)^{\text{ab}} / (\Gamma_0(p)^{\text{ab}})_{\text{tor}} \cong H_1(Y_0(p), \mathbb{Z}). \tag{3.7}$$

We note that for a closed oriented geodesic  $\mathcal{C}_A(p)$  as in the previous section, the homology class  $[\mathcal{C}_A(p)] \in H_1(Y_0(p), \mathbb{Z})$  corresponds exactly to the image of (any)  $\gamma_Q$  as in (3.3) under the map (3.5). Using the above identifications, the cap product pairing is induced by the map

$$\Gamma_0(p) \times \mathcal{M}_2(p) \ni (\gamma, f) \mapsto \int_z^{\gamma z} \omega_f,$$

for any  $z \in \mathbb{H}$ .

The natural pullback map induced by inclusion fits into a short exact sequence of  $\mathbb{R}$ -vector spaces

$$0 \rightarrow H^1(X_0(p), \mathbb{R}) \rightarrow H^1(Y_0(p), \mathbb{R}) \rightarrow \mathbb{R} \rightarrow 0, \tag{3.8}$$

using here that we have two cusps (which, as we will see, implies that the Eisenstein space of weight 2 is one-dimensional). This identifies  $H^1(X_0(p), \mathbb{R})$  with the parabolic classes in  $H^1(Y_0(p), \mathbb{R})$  (i.e., classes which vanish on all parabolic elements of  $\Gamma_0(p)$  using the above pairing). Under the de Rham isomorphism, the parabolic classes correspond to 1-forms obtained from holomorphic cusp(!) forms of weight 2 and level  $p$  (see, for example, [Shi94, Section 5]). Similarly, we have the pushforward map  $H_1(Y_0(p), \mathbb{R}) \rightarrow H_1(X_0(p), \mathbb{R})$  whose kernel is exactly given by the image of the parabolic elements of  $\Gamma_0(p) \otimes \mathbb{R}$  inside  $H_1(Y_0(p), \mathbb{R})$  under the map (3.5).

### 3.2.1. Hecke operators

Due to the arithmetic nature of  $\Gamma_0(p)$ , we have a family of commuting linear operators acting on all of the above-mentioned (co)homology groups, namely the *Hecke operators*. The Hecke action is induced by the following: on the space of holomorphic forms  $\mathcal{M}_2(p)$  of weight 2 and level  $p$ , the  $n$ -th Hecke

operator  $T_n$  acts by (see, for example, [IK04, (14.46)])

$$T_n f(z) := \frac{1}{n} \sum_{\substack{ad=n, \\ (a,p)=1}} a^2 \sum_{0 \leq b < d} f\left(\frac{az+b}{d}\right), \tag{3.9}$$

the Fricke involution  $W_p$  acts as

$$W_p f(z) := p^{-1} z^{-2} f(-1/(pz)) \tag{3.10}$$

and we also have the involution  $\iota$  given by

$$\iota f(z) := f(-\bar{z}), \tag{3.11}$$

defined for  $f \in \mathcal{M}_2(p) \oplus \overline{\mathcal{M}_2(p)}$ . Notice that, with this normalization, the Ramanujan conjecture amounts to the bound  $\leq d(n)n^{1/2}$  for the Hecke eigenvalues. See also [Shi94, Section 5] for an intrinsic definition in terms of group cohomology using double cosets. Similarly, we can define an action on the homology groups by (using the notation (3.6))

$$T_n \{z_1, z_2\} := \sum_{\substack{ad=n, \\ (a,p)=1}} \sum_{0 \leq b < d} \left\{ \frac{az_1+b}{d}, \frac{az_2+b}{d} \right\}, \tag{3.12}$$

$$W_p \{z_1, z_2\} := \{-1/(pz_1), -1/(pz_2)\} \tag{3.13}$$

and

$$\iota \{z_1, z_2\} := \{-\bar{z}_1, -\bar{z}_2\}. \tag{3.14}$$

One can now check that all of these operators are pairwise adjoint with respect to the cap product pairing as above. Furthermore, it can be shown that all of these linear operators commute, and thus we can find a common eigen-basis. Explicitly, such a Hecke eigen-basis for  $H^1(Y_0(p), \mathbb{R})$  is given by

$$B_{\text{Hecke}}(p) := \{\omega_f^\epsilon : f \in \mathcal{B}_p, \epsilon \in \{\pm\}\} \cup \{\omega_E(p)\}, \tag{3.15}$$

where

$$\omega_f^\pm := \frac{2\pi i f(z) dz \pm 2\pi i \overline{f(z)} dz}{1+i \pm (1-i)} \in H^1(Y_0(p), \mathbb{R}),$$

with  $\mathcal{B}_p := \{f_1, \dots, f_g\} \subset \mathcal{S}_2(p)$  a basis of Hecke normalized (i.e., the first Fourier coefficient is 1) holomorphic cuspidal eigen forms of weight 2 and level  $p$ . And

$$\omega_E(p) := E_{2,p}(z) dz \in H^1(Y_0(p), \mathbb{R}) \tag{3.16}$$

is the normalized *Eisenstein class* defined from the weight 2 Eisenstein series of level  $p$ :

$$E_{2,p}(z) := \frac{pE_2(pz) - E_2(z)}{p-1} = \frac{pE_2^*(pz) - E_2^*(z)}{p-1} = 1 + \frac{24}{p-1} e^{2\pi iz} + \dots,$$

where  $E_2$  denotes the the weight 2 Eisenstein series of level 1 and  $E_2^*$  the modified Eisenstein series given by

$$E_2(z) := 1 - 24 \sum_{n \geq 1} \sigma_1(n) e^{2\pi i n z} = E_2^*(z) + \frac{3}{\pi y}. \tag{3.17}$$

Here,  $\sigma_1(n) = \sum_{d|n} d$  is the sum of divisors function. One obtains a Hecke eigen basis for the homology as the dual basis of  $B_{\text{Hecke}}(p)$  with respect to the cap product pairing (3.4) which we denote by

$$B_{\text{Hecke}}(p)^* := \{v_f^\pm : f \in \mathcal{B}_p, \pm\} \cup \{v_E(p)\} \subset H_1(Y_0(p), \mathbb{R}), \tag{3.18}$$

where  $\langle v_f^\pm, \omega_f^\pm \rangle = 1$  for  $f \in \mathcal{B}_p$  and  $\langle v_E(p), \omega_E(p) \rangle = 1$ . Since the constant Fourier coefficient of  $E_{2,p}$  is 1, one sees that, in fact,

$$v_E(p) = \{i, i + 1\} \in H_1(Y_0(p), \mathbb{Z}), \tag{3.19}$$

with notation as in (3.6). In other words,  $v_E(p)$  is the (normalized) Eisenstein class in homology appearing in Theorem 1.1 which geometrically is the class of a simple loop going around the cusp at  $\infty$ .

We have the following classical formula for  $\gamma \in \text{PSL}_2(\mathbb{Z})$  (see, for example, [DIT18, (55)]):

$$\int_z^{\gamma z} E_2^*(z) dz = \Psi(\gamma),$$

where

$$\Psi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) := \begin{cases} \frac{a+d}{c} - 12 \operatorname{sign}(c)s(a, c) - 3 \operatorname{sign}(c(a+d)), & c \neq 0 \\ \frac{b}{d}, & c = 0 \end{cases}$$

is the Rademacher symbol with

$$s(a, c) := \sum_{n=1}^c \left(\left(\frac{n}{c}\right)\right) \left(\left(\frac{na}{c}\right)\right),$$

the classical Dedekind sum with

$$\left(\left(x\right)\right) = \begin{cases} x - [x] - 1/2, & x \notin \mathbb{Z} \\ 0, & x \in \mathbb{Z} \end{cases}$$

the sawtooth function. Notice that  $\Psi(\gamma)$  does not depend on the representative of  $\gamma \in \text{PSL}_2(\mathbb{Z})$  as should be the case. This implies the following key formula for  $\gamma \in \Gamma_0(p)$ :

$$\langle \{z, \gamma z\}, \omega_E(p) \rangle = \frac{\Psi(\gamma') - \Psi(\gamma)}{p - 1}, \tag{3.20}$$

where

$$\gamma' = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \gamma \begin{pmatrix} 1/p & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & bp \\ c/p & d \end{pmatrix} \in \text{PSL}_2(\mathbb{Z}),$$

with  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ . Using the trivial bound  $\Psi(\gamma) \ll \frac{|a+d|}{c} + c$  (for  $c > 0$ ), we get the following useful estimate:

$$\langle \{z, \gamma z\}, \omega_E(p) \rangle \ll \frac{p|a+d| + c^2}{pc}, \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(p), c > 0. \tag{3.21}$$

Recall the short exact sequence (3.8) above. A Hecke-equivariant splitting is given by mapping  $1 \in \mathbb{R}$  to the Eisenstein class  $\omega_E(p) = E_{2,p}(z)dz$  (this fits into a general framework due to Franke [Fra98]). This defines an isomorphism

$$H^1(Y_0(p), \mathbb{R}) \cong H^1(X_0(p), \mathbb{R}) \oplus \mathbb{R}\omega_E(p), \tag{3.22}$$

and we will denote the projection to the ‘cuspidal subspace’ (which we will identify with  $H^1(X_0(p), \mathbb{R})$ ) with respect to this splitting by

$$\mathbb{P}_{\text{cusp}} : H^1(Y_0(p), \mathbb{R}) \rightarrow H^1(X_0(p), \mathbb{R}) \subset H^1(Y_0(p), \mathbb{R}), \tag{3.23}$$

which is explicitly given by

$$\mathbb{P}_{\text{cusp}}\omega = \omega - \langle v_E(p), \omega \rangle \omega_E(p),$$

where  $v_E(p), \omega_E(p)$  are the Eisenstein classes defined above.

### 4. Background on Fuchsian groups

In this section, we will review some useful facts regarding the geometry of fundamental polygons of Fuchsian groups. We will refer to [Bea83, Section 9] for an in-depth treatment.

#### 4.1. Fundamental polygons

Let  $\Gamma \subset \text{PSL}_2(\mathbb{R})$  be a Fuchsian subgroup of the first kind (i.e., a discrete and cofinite subgroup of  $\text{PSL}_2(\mathbb{R})$ ). A *fundamental domain* for  $\Gamma$  is a locally finite domain  $\mathcal{F} \subset \mathbb{H}$  such that

1. For any  $z \in \mathbb{H}$ , we have  $\gamma z \in \overline{\mathcal{F}}$  for some  $\gamma \in \Gamma$ .
2. If  $z_1, z_2 \in \overline{\mathcal{F}}$  are  $\Gamma$ -equivalent, then  $z_1 = z_2$  or  $z_1, z_2 \in \partial\mathcal{F}$ .

We say that  $\mathcal{F}$  is a *fundamental polygon* for  $\Gamma$  if  $\mathcal{F}$  is furthermore (hyperbolically) convex with a piecewise geodesic boundary. We define a *side of  $\mathcal{F}$*  as a nonempty subset of the shape  $\gamma\overline{\mathcal{F}} \cap \overline{\mathcal{F}}$  with  $\text{Id} \neq \gamma \in \Gamma$ . We define a *vertex of  $\mathcal{F}$*  as a non-empty subset of the shape  $\gamma_1\overline{\mathcal{F}} \cap \gamma_2\overline{\mathcal{F}} \cap \overline{\mathcal{F}}$  with  $\text{Id}, \gamma_1, \gamma_2 \in \Gamma$  pairwise distinct. It can be shown that a fundamental polygon  $\mathcal{F}$  has an even number of sides which are pairwise  $\Gamma$ -equivalent, and  $\Gamma\mathcal{F}$  gives a tessellation of  $\mathbb{H}$ . The set of elements identifying sides of  $\mathcal{F}$  are called the *side pairing transformations* associated to  $\mathcal{F}$  which we denote by  $\mathcal{S}(\mathcal{F}) \subset \Gamma$  (which formally is a multiset if there are order two elements in  $\Gamma$ ). Given a side  $L$  of  $\mathcal{F}$  we refer to *the side pairing transformation associated to the side  $L$*  as the element  $\sigma \in \mathcal{S}(\mathcal{F})$  such that  $\sigma^{-1}L$  is also a side of  $\mathcal{F}$ . Similarly we say that  $\sigma$  is the side pairing transformation associated to the side  $\gamma L$  of the translate  $\gamma\mathcal{F}$  for each  $\gamma \in \Gamma$ .

It is known that  $\mathcal{S}(\mathcal{F})$  generates  $\Gamma$  for any fundamental polygon  $\mathcal{F}$ , and it is a fundamental fact that one can understand the relation between the elements of  $\mathcal{S}(\mathcal{F})$  from the geometry of a fundamental polygon and vice versa (see Lemma 4.3). A simple but key incarnation is the following.

**Proposition 4.1.** *Let  $\mathcal{F}$  be a fundamental polygon for a discrete and co-finite subgroup  $\Gamma \subset \text{PSL}_2(\mathbb{R})$ . Consider a sequence of consecutive  $\Gamma$ -translates of  $\mathcal{F}$ :*

$$\mathcal{F}, \gamma_1\mathcal{F}, \dots, \gamma_n\mathcal{F}.$$

Then

$$\gamma_n = \sigma_1\sigma_2 \cdots \sigma_n,$$

where  $\sigma_i \in \mathcal{S}(\mathcal{F})$  denotes the side pairing transformation associated to the side shared between  $\gamma_{i-1}\mathcal{F}$  and  $\gamma_i\mathcal{F}$  (here, we put  $\gamma_0 = \text{Id}$ ).

*Proof.* Note that the side of  $\mathcal{F}$  shared with  $\gamma_1\mathcal{F}$  will exactly have associated  $\sigma_1 = \gamma_1 \in \mathcal{S}(\mathcal{F})$ . Now if the side shared between  $\gamma_1\mathcal{F}$  and  $\gamma_2\mathcal{F}$  has associated  $\sigma_2 \in \mathcal{S}(\mathcal{F})$ , then we have

$$\gamma_2\mathcal{F} = (\sigma_1\sigma_2\sigma_1^{-1})(\sigma_1\mathcal{F}) = \sigma_1\sigma_2\mathcal{F}.$$

Continuing like this, we get

$$\gamma_n\mathcal{F} = \sigma_1 \cdots \sigma_n\mathcal{F},$$

and since  $\gamma\mathcal{F} = \gamma'\mathcal{F}$  implies  $\gamma = \gamma' \in \Gamma$  by condition (2), we conclude the wanted equality. □

A first application of Proposition 4.1 is the following slight reformulation.

**Corollary 4.2.** *Let  $c : [0, 1] \rightarrow \mathbb{H}$  be a continuous, injective curve with  $c(0) = z \in \mathcal{F}$  and  $c(1) = \gamma z$  for some  $\gamma \in \Gamma$ . Assume that  $c([0, 1])$  does not intersect the set of  $\Gamma$ -translates of the vertices of  $\mathcal{F}$ . Then one can code the element  $\gamma$  in the following way; let  $\sigma_1, \dots, \sigma_n \in \mathcal{S}(\mathcal{F})$  be the (ordered) sequence of side pairing transformations associated to the intersection between the curve  $c$  and  $\Gamma$ -translates of the sides of  $\mathcal{F}$ . Then one has*

$$\gamma = \sigma_1 \cdots \sigma_n.$$

When  $c([0, 1])$  is a closed geodesic (when projected to  $\Gamma \backslash \mathbb{H}$ ), this is known as *geometric coding* of geodesics (see [Kat96] for a nice treatment).

As a second consequence, we note that if the sequence of translates ‘loops around’, meaning that  $\gamma_n = 1$ , then one obtains a relation between the side pairing transformations  $\mathcal{S}(\Gamma)$ . This yields immediately that if two sides of  $\mathcal{F}$  are paired, then the associated elements in  $\mathcal{S}(\mathcal{F})$  are inverses. We call these the *inverse relations* of  $\mathcal{F}$ . We also get relations by looping around the vertices of  $\mathcal{F}$  which we will now make precise. Observe that the embedding  $\mathcal{F} \subset \mathbb{H}$  defines an orientation on the boundary  $\partial\mathcal{F}$ . Let  $L_1$  be a side of  $\mathcal{F}$  with leftmost (wrt. the orientation) vertex  $v_1$  and let  $\sigma_1 \in \mathcal{S}(\mathcal{F})$  be the associated side pairing transformation. Let  $L_2$  be the side of  $\mathcal{F}$  different from  $\sigma_1^{-1}L_1$  containing the vertex  $v_2 = \sigma_1^{-1}v_1$  and let  $\sigma_2 \in \mathcal{S}(\mathcal{F})$  be the associated side pairing transformation. Continuing like this yields a periodic sequence of pairs

$$(L_1, v_1), (L_2, v_2), \dots,$$

with minimal period  $m \geq 1$ , say, which we call a *cycle of  $\mathcal{F}$* . It is now clear that  $\sigma_1\sigma_2 \dots \sigma_m$  fixes  $v_1$ . Thus, if we put  $\nu = |\Gamma_{v_1}|$  (i.e., the size of the stabilizer of  $v_1$  inside  $\Gamma$ ), then we get the following relation on the side pairing transformations:

$$(\sigma_1\sigma_2 \dots \sigma_m)^\nu = 1,$$

where if  $\nu = \infty$  (i.e.,  $v_1$  is a boundary vertex), this is understood as the empty relation. Notice that the relation does not depend on the choice of starting point  $(L_1, v_1)$ . We call these the *cycle relations of  $\mathcal{F}$* , and we have the following key theorem of Poincaré; see [Mas71] for a proof.

**Lemma 4.3** (Poincaré’s Theorem). *The side pairing transformations  $\mathcal{S}(\mathcal{F})$  generate  $\Gamma$ , and the inverse and cycle relations give a complete set of relations for  $\mathcal{S}(\mathcal{F})$ .*

Secondly, it is clear that each side appears in exactly one cycle. Thus, we obtain the following useful fact.

**Lemma 4.4** (cf. Lemma 5.3 of [Voi09]). *Each side pairing transformation appears exactly once in a cycle relation.*

**4.1.1. Bounding coordinates of side pairing transformations**

Now we consider the image of the inverse and cycle relations when mapped to  $V_p$  under the composition of maps:

$$\Gamma_0(p) \twoheadrightarrow \Gamma_0(p)^{ab} \twoheadrightarrow H_1(Y_0(p), \mathbb{Z}) \hookrightarrow V_p = H_1(Y_0(p), \mathbb{R}) \cong \mathbb{R}^{2g+1}.$$

Notice that since  $V_p$  is torsion-free, we can divide the relations by  $v$  so that all coefficients in the relations are contained in  $\{-1, 0, 1\}$ . By killing the inverse relations, this yields a system of linear relations with a variable for each pair  $\{\sigma, \sigma^{-1}\} \subset \mathcal{S}(\mathcal{F})$  of side pairing transformations so that each such variable appears once with positive and once with negative sign (which follows from Lemma 4.4).

Let  $B = \{v_0, \dots, v_{2g}\} \subset V_p$  be a basis of the real homology corresponding to side pairing transformations  $\sigma \in \mathcal{S}(\mathcal{F})$  under the surjection (3.5). Combining the cycle relations with the choice of  $B$  gives rise to a system of linear equations in  $V_p$ :

$$\mathcal{L}_{\mathcal{F},B}(x_0, \dots, x_n) = \left\{ \sum_{0 \leq i \leq n} a_{ik}x_i + \sum_{0 \leq j \leq 2g} b_{jk}v_j = 0 \right\}_{0 \leq k \leq K},$$

where  $x_0, \dots, x_n$  are variables, one for each pair  $\{\sigma, \sigma^{-1}\} \subset \mathcal{S}(\mathcal{F})$  not corresponding to an element of  $B$ . Lemmata 4.3 and 4.4 translate to two key properties:

1.  $\mathcal{L}_{\mathcal{F},B}$  has exactly one solution  $(x_0, \dots, x_n) \in (V_p)^{n+1}$ .
2. For any subset  $A \subset \{0, \dots, K\}$ , and any indices  $0 \leq i \leq n, 0 \leq j \leq 2g$ , we have

$$\sum_{k \in A} a_{ik}, \sum_{k \in A} b_{jk} \in \{-1, 0, 1\}.$$

We have the following general result about such systems of linear equations.

**Lemma 4.5.** *Let*

$$\mathcal{L}(x_0, \dots, x_n) = \left\{ \sum_{0 \leq i \leq n} a_{ik}x_i + \sum_{0 \leq j \leq 2g} b_{jk}v_j = 0 \right\}_{0 \leq k \leq K}$$

*be a system of linear equations in  $V_p$  satisfying (1) and (2).*

*Then the unique solution  $(x_0, \dots, x_n) \in (V_p)^{n+1}$  satisfies*

$$x_i = \sum_{0 \leq j \leq 2g} c_{ij}v_j, \quad c_{ij} \in \{-1, 0, 1\}, \tag{4.1}$$

*for all  $0 \leq i \leq n$ .*

*Proof.* We proceed by induction on  $K + 1 \geq 0$  (i.e., the number of equations). If  $K + 1 = 0$ , then there is nothing to prove. Now assume the claim is known for systems of  $K' < K + 1$  equations. We start by making the following reductions. We may assume that all equations contain a variable (i.e.,  $\forall k \exists i : a_{ik} \neq 0$ ), since otherwise we can remove such an equation. Furthermore, we may assume that there are some variables appearing exactly once (i.e.,  $\exists i : |\{k : a_{ik} \neq 0\}| = 1$ ). If this is not the case, then by (1) and (2) every variable appears exactly twice with coefficients  $\pm 1$ , respectively. This means that the equation obtained by adding all of the equations in  $\mathcal{L}(x_0, \dots, x_n)$  has to be trivial. Thus, by removing any equation, we obtain an equivalent system with  $K$  equations, and we are done by the induction hypothesis.

The key observation is that there is always an equation with exactly one variable appearing (i.e.,  $\exists k : |\{i : a_{ik} \neq 0\}| = 1$ ). If not, then since every variable appears at most twice and there is one variable appearing once (by the above reductions) we have a system of linear equations with more variables than

equations. Thus, the number of solutions  $(x_0, \dots, x_n) \in (V_p)^{n+1}$  is either 0 or  $\infty$ , which contradicts (1). Let  $x_i$  be the only variable appearing in some equation. Then this equation determines  $x_i$ , which by (2) satisfies (4.1). Now if  $-x_i$  does not appear, we are done by the induction hypothesis. Otherwise, we add the equation containing  $x_i$  to the one containing  $-x_i$  which gives a system of  $K$  linear equations again satisfying (1) and (2). Thus, the claim follows from the induction hypothesis.  $\square$

**Corollary 4.6.** *Let  $\Gamma$  be a Fuchsian group with a fundamental polygon  $\mathcal{F}$ , and let  $B$  be a basis for  $H_1(\Gamma \backslash \mathbb{H}, \mathbb{R})$  consisting of classes of the form  $\{z, \sigma z\}$  with  $\sigma \in \mathcal{S}(\mathcal{F})$  and  $z \in \mathbb{H}$ .*

*Then we have for any  $\sigma \in \mathcal{S}(\mathcal{F})$ ,  $z \in \mathbb{H}$  and  $\omega \in B^*$  that*

$$|\langle \{z, \sigma z\}, \omega \rangle| \leq 1.$$

*Proof.* Observe that if the variable  $x_i$  corresponds to  $\{\sigma, \sigma^{-1}\} \in \mathcal{S}(\mathcal{F})$ , then the numbers

$$(\langle \{z, \sigma z\}, \omega \rangle)_{\omega \in B^*}$$

are exactly the coordinates of either  $x_i$  or  $-x_i$  in the basis  $B$  where  $(x_0, \dots, x_n) \in (V_p)^{n+1}$  is the unique solution to  $\mathcal{L}_{\mathcal{F}, B}$ . Now the result follows directly from Lemma 4.5.  $\square$

**4.2. Zagier’s fundamental polygon**

We will now consider a fundamental polygon for  $\Gamma_0(p)$  introduced by Zagier [Zag85, Section 3] which will give rise to a set of natural bases for the homology. The following is a fundamental polygon for  $\Gamma_0(p)$ :

$$\tilde{\mathcal{F}}_{\text{Zag}}(p) = \cup_{i=0}^p \sigma_i \mathcal{F}_{\text{std}},$$

where  $\sigma_i = \begin{pmatrix} 0 & -1 \\ 1 & i \end{pmatrix}$  for  $0 \leq i < p$  and  $\sigma_p = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ , and

$$\mathcal{F}_{\text{std}} := \{z \in \mathbb{H} : |\operatorname{Re} z| < 1/2, |z| > 1\} \tag{4.2}$$

is the standard fundamental polygon for  $\operatorname{PSL}_2(\mathbb{Z})$ . This gives rise to the side pairing transformation set

$$\mathcal{S}(\tilde{\mathcal{F}}_{\text{Zag}}(p)) = \left\{ \begin{pmatrix} 1 & 0 \\ \pm p & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} -a^* & -1 \\ (aa^*+1) & j \end{pmatrix} : 0 < a < p \right\},$$

where  $0 < a^* < p$  is such that  $aa^* \equiv -1 \pmod p$ . Recall that  $\Gamma_0(p)$  is normalized by the matrix  $W_p = \begin{pmatrix} 0 & -1/\sqrt{p} \\ \sqrt{p} & 0 \end{pmatrix}$ , which implies that also  $\mathcal{F}_{\text{Zag}}(p) := W_p \tilde{\mathcal{F}}_{\text{Zag}}(p)$  is a fundamental polygon for  $\Gamma_0(p)$  with side pairing transformations given by

$$\begin{aligned} \mathcal{S}(\mathcal{F}_{\text{Zag}}(p)) &= W_p \left( \left\{ \begin{pmatrix} 1 & 0 \\ \pm p & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} -a^* & -1 \\ (aa^*+1) & a \end{pmatrix} : 0 < a < p \right\} \right) W_p^{-1} \\ &= \left\{ \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 1 & 0 \\ \pm p & 1 \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} a & -(aa^*+1)/p \\ p & -a^* \end{pmatrix} : 0 < a < p \right\}, \end{aligned}$$

which, by Lemma 4.3, generate  $\Gamma_0(p)$  (as claimed in the introduction). Both of these fundamental polygons have the nice property that all cuspidal sides (i.e., sides containing a cusp) are paired by a parabolic element of  $\Gamma_0(p)$  (which is not the case for all fundamental polygons). The elements of  $\mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$  are minimal in the sense that the archimedean sizes of the entries are as small as one can hope for ( $\leq p$ ). Explicitly, the fundamental polygon  $\mathcal{F}_{\text{Zag}}(p)$  has  $p + 3$  vertices: the two cusps 0 and  $\infty$  as well as  $\frac{2j-1+i\sqrt{3}}{2p}$  for  $0 \leq j \leq p$ . We see that the matrix  $W_p$  takes the hyperbolic triangle with vertices  $\{\frac{-1+i\sqrt{3}}{2p}, 0, \frac{1+i\sqrt{3}}{2p}\}$  to the hyperbolic triangle with vertices  $\{\frac{-1+i\sqrt{3}}{2}, \infty, \frac{1+i\sqrt{3}}{2}\}$ . In particular, the

subgroup  $\langle W_p, \Gamma_0(p) \rangle \leq \text{PSL}_2(\mathbb{R})$  has a fundamental domain contained in

$$\{z \in \mathbb{H} : |\text{Re } z| \leq 1/2, \text{Im } z \geq \sqrt{3}/(2p)\}, \tag{4.3}$$

which will be useful later on.

We define the following compatible family of bases of the homology groups.

**Definition 4.7.** A basic basis of level  $p$  is a basis  $B \subset H_1(Y_0(p), \mathbb{R})$  consisting of elements  $\{z, \sigma z\}$  with  $z \in \mathbb{H}$  and  $\sigma \in \mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$ .

### 4.3. Special fundamental polygons

By the Kurosh subgroup theorem, we know that any subgroup of  $\text{PSL}_2(\mathbb{Z}) \cong \mathbb{Z}/2\mathbb{Z} * \mathbb{Z}/3\mathbb{Z}$  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, a torsion-free Hecke congruence subgroup  $\Gamma_0(p)$  is a free group on  $k = \text{rank}_{\mathbb{Z}} \Gamma_0(p)^{\text{ab}}$  generators. We will now describe an explicit geometric way due to Kulkarni [Kul91] for constructing a set of independent generators of  $\Gamma_0(p)$ . The starting points are so-called *special fundamental polygons* of  $\Gamma_0(p)$ .

Let  $g$  be the genus of  $Y_0(p)$ , and  $e_2, e_3$  the number of conjugacy classes of subgroups in  $\Gamma_0(p)$  of order, respectively, 2 and 3. Following [Kul91], we define a *Farey symbol of level  $p$*  as a sequence of reduced fractions

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

with  $n = 4g + e_2 + e_3$  such that  $a_{i+1}b_i - a_ib_{i+1} = 1$  for all  $1 \leq i < n$ . Furthermore (considering below the indices modulo  $n + 2$ ),

- there are  $e_2$  even indices  $i$  such that

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

- there are  $e_3$  odd indices  $i$  such that

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

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

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

Such a symbol always exists, and one can even find one which is symmetric around  $1/2$  [Kul91, Section 13]. Dooms–Jesper–Konolalov [DJK10] have described an algorithm for determining Farey symbols of a general level.

Consider the polygon  $\mathcal{P}(p)$  with vertices at  $\infty$ , at the fractions of the Farey symbol, at the midpoint of the geodesic circle connecting  $a_i/b_i$  and  $a_{i+1}/b_{i+1}$  for  $i$  an even index, and for an odd index  $i$  at the  $\text{PGL}_2(\mathbb{Z})$ -translate of  $\frac{1+i\sqrt{3}}{2}$  lying between  $a_i/b_i$  and  $a_{i+1}/b_{i+1}$  (for details see [Kul91, Section 2]). Note that  $\mathcal{P}(p)$  consists of  $\text{PGL}_2(\mathbb{Z})$ -translates of

$$\mathcal{F}^+ := \{z \in \mathbb{H} : 0 < \text{Re } z < 1/2, |z| > 1\}. \tag{4.4}$$

The map  $*$  defines a side pairing transformation on this polygon by identifying the half circle connecting  $\frac{a_i}{b_i}, \frac{a_{i+1}}{b_{i+1}}$  and the one connecting  $\frac{a_{i^*+1}}{b_{i^*+1}}, \frac{a_{i^*}}{b_{i^*}}$ , as well as identifying the vertical sides of  $\mathcal{P}(p)$  and the elliptic sides. By Poincaré’s theorem,  $\mathcal{P}(p)$  together with this pairing defines a subgroup of  $\text{PSL}_2(\mathbb{R})$  which can be shown to be equal to  $\Gamma_0(p)$ . Furthermore, since  $\mathcal{P}(p)$  has a minimal number of sides, it follows that an

independent set of generators of  $\Gamma_0(p)$  is given by the matrices which map between the sides identified by the pairing induced from  $*$ . These matrices are explicitly given by (see [Kul91, Theorem 6.1])

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{4.5}$$

and the  $e_2$  matrices of order 2 and  $e_3$  matrices of order 3

$$\begin{pmatrix} a_{i+1}b_{i+1} + a_i b_i & -a_i^2 - a_{i+1}^2 \\ b_i^2 + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_i b_i \end{pmatrix}, \tag{4.6}$$

$$\begin{pmatrix} a_{i+1}b_{i+1} + a_i b_{i+1} + a_i b_i & -a_i^2 - a_i a_{i+1} - a_{i+1}^2 \\ b_i^2 + b_i b_{i+1} + b_{i+1}^2 & -a_{i+1}b_{i+1} - a_{i+1}b_i - a_i b_i \end{pmatrix}, \tag{4.7}$$

together with the  $2g$  hyperbolic matrices

$$\begin{pmatrix} a_{i^*+1}b_{i+1} + a_i^* b_i & -a_i a_{i^*} - a_{i+1} a_{i^*+1} \\ b_i b_{i^*} + b_{i+1} b_{i^*+1} & -a_{i+1} b_{i^*+1} - a_i b_{i^*} \end{pmatrix} \text{ for } i < i^* \text{ a pair.} \tag{4.8}$$

Observe that the  $2g$  hyperbolic matrices above, together with  $T$ , define a basis for  $H_1(Y_0(p), \mathbb{R})$  under the map (3.5). Recently, Doan–Kim–Lang–Tan [DKLP22] have shown that one can find *minimal special fundamental polygons*  $\mathcal{P}_{\min}(p)$ , meaning that we have

$$b_i b_{i^*} + b_{i+1} b_{i^*+1} = p, \quad 0 \leq i \leq n,$$

which implies that, in fact,  $\mathcal{S}(\mathcal{P}_{\min}(p)) \subset \mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$ .

### 5. Proof of Theorem 1.1

In this section, we will prove the following statement which implies our first main result.

**Theorem 5.1.** *Fix a prime  $p$  and  $\delta \in (0, \frac{1}{2|14})$ . Consider a real quadratic field  $K$  of discriminant  $d_K$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$ . Consider a subgroup  $H \leq \text{Cl}_K^+$  with  $\mathfrak{p}_1 \notin H$  and  $J = (\sqrt{d_K}) \notin H$ .*

*Then for any  $\omega \in H^1(Y_0(p), \mathbb{R})$ , we have as  $d_K \rightarrow \infty$*

$$\frac{\sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega \rangle}{|\sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle|} = -\langle v_E(p), \omega \rangle + O_\omega(d_K^{-\delta}), \tag{5.1}$$

where  $\langle \cdot, \cdot \rangle$  denotes the cap product pairing, and  $v_E(p), \omega_E(p)$  are the Eisenstein classes in (co)homology defined in (3.16) and (3.19).

First of all, let us see how Theorem 1.1 follows from this.

*Proof of Theorem 1.1 assuming Theorem 5.1.* Let  $B$  be any basis of  $V_p = H_1(Y_0(p), \mathbb{R})$  containing  $v_E(p)$  and let  $B^*$  denote the dual basis of  $H^1(Y_0(p), \mathbb{R})$  with respect to the cap product pairing. Consider the isomorphism  $V_p \cong \mathbb{R}^{2g+1}$  defined by sending  $B$  to the standard basis of  $\mathbb{R}^{2g+1}$  and denote by  $\|\cdot\| = \|\cdot\|_{B, \infty}$  the norm on  $V_p$  obtained by pulling back the sup norm with respect to the standard basis of  $\mathbb{R}^{2g+1}$ . We will prove the convergence (1.8) which implies Theorem 1.1. By Theorem 5.1, we have for  $d_K$  large enough that

$$\left\| \sum_{A \in H} [\mathcal{C}_A(p)] \right\| = \max_{\omega \in B^*} \left| \sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega \rangle \right| = \left| \sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle \right| \max_{\omega \in B^*} \left( |\langle v_E(p), \omega \rangle| + O_\omega(d_K^{-\delta}) \right) \tag{5.2}$$

$$= \left| \sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle \right| (1 + O_B(d_K^{-\delta})). \tag{5.3}$$

By the triangle inequality, we conclude that

$$\left\| \frac{\sum_{A \in H} [\mathcal{C}_A(p)]}{\|\sum_{A \in H} [\mathcal{C}_A(p)]\|} + v_E(p) \right\| \tag{5.4}$$

$$\leq \left\| \frac{\sum_{A \in H} [\mathcal{C}_A(p)]}{|\langle \sum_{A \in H} [\mathcal{C}_A(p)], \omega_E(p) \rangle|} + v_E(p) \right\| + \left| \frac{\|\sum_{A \in H} [\mathcal{C}_A(p)]\|}{|\langle \sum_{A \in H} [\mathcal{C}_A(p)], \omega_E(p) \rangle|} - 1 \right|. \tag{5.5}$$

Finally by (5.3) and Theorem 5.1, the above is bounded by  $O_B(d_K^{-\delta})$ , which yields the wanted expression.  $\square$

The rest of this section is occupied with the proof of Theorem 5.1. The idea is to do a change of coordinates to the Hecke basis (3.15) of the real cohomology. Then using formulas due to Hecke and Waldspurger (more precisely, an explicit extension due in this case to Popa [Pop08]), we reduce the problem to a question about certain special values of  $L$ -functions. Now the result follows upon applying subconvexity bounds (as well as lower bounds for  $L$ -functions on the critical line). This is exactly the same proof structure as is used in the automorphic approach to Duke’s Theorem (see, for example, [MV06] and the references therein).

Given  $\omega \in H^1(Y_0(p), \mathbb{R})$ , we obtain by expanding in the Hecke basis for homology defined in (3.18) that

$$\langle [\mathcal{C}_A(p)], \omega \rangle = \langle v_E(p), \omega \rangle \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle + \sum_{f \in \mathcal{B}_{p, \pm}} \langle v_f^\pm, \omega \rangle \langle [\mathcal{C}_A(p)], \omega_f^\pm \rangle.$$

Now we want to average over cosets  $CH \subset \text{Cl}_K^+$  for subgroups  $H \leq \text{Cl}_K^+$  of the narrow class group. Using a standard trick from Fourier analysis, we can write

$$\begin{aligned} \sum_{A \in CH} \langle [\mathcal{C}_A(p)], \omega \rangle &= \frac{|H|}{|\text{Cl}_K^+|} \sum_{\chi \in \overline{\text{Cl}_K^+} : \chi|_H=1} \chi(C) \sum_{A \in \text{Cl}_K^+} \langle [\mathcal{C}_A(p)], \omega \rangle \bar{\chi}(A) \\ &= \langle v_E(p), \omega \rangle \sum_{A \in CH} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle \\ &\quad + \sum_{f \in \mathcal{B}_{p, \pm}} \langle v_f^\pm, \omega \rangle \frac{|H|}{|\text{Cl}_K^+|} \sum_{\chi \in \overline{\text{Cl}_K^+} : \chi|_H=1} \chi(C) \sum_{A \in \text{Cl}_K^+} \langle [\mathcal{C}_A(p)], \omega_f^\pm \rangle \bar{\chi}(A). \end{aligned} \tag{5.6}$$

The inner sums in the cuspidal contribution are twisted Weyl sums, which we will have to estimate. These sums fits into the framework of *toric periods* (see, for example, [ELMV11, Section 4]) and they turn out to be related to special values of automorphic  $L$ -functions as first proved by Waldspurger [Wal85]. This is exactly the reason why we have changed to the Hecke coordinates.

### 5.1. The Eisenstein contribution

In order to estimate the Eisenstein contribution, we will rely on the following classical formula of Hecke (see [DIT18, (68)]):

$$\sum_{A \in \text{Cl}_K^+} \int_{\mathcal{C}_A(p)} E_2^*(z) dz \chi(A) = 6(1 - \chi(J))L(\chi, 0),$$

where  $E_2^*(z) = E_2(z) - \frac{3}{\pi \text{Im} z}$  (with  $E_2(z)$  defined in (3.17)),  $J = (\sqrt{D}) \in \text{Cl}_K^+$  is the different of  $K$  and  $L(\chi, s)$  denotes the (finite part) of the Hecke  $L$ -function associated to the narrow class group character

$\chi$ . Thus, by a change of variable, we get

$$\sum_{A \in \text{Cl}_K^+} \int_{\mathcal{C}_A(p)} p E_2^*(pz) dz \chi(A) = \sum_{A \in \text{Cl}_K^+} \int_{p \mathcal{C}_A(p)} E_2^*(z) dz \chi(A).$$

We notice by direct computation that  $p \mathcal{C}_A(p)$  (i.e., the dilation of  $\mathcal{C}_A(p)$  by the factor  $p$ ) is again a closed geodesic as follows. If  $\mathcal{C}_A(p)$  corresponds to the quadratic form  $ax^2 + bxy + cy^2$  of discriminant  $d_K$  and level  $p$ , then  $p \mathcal{C}_A(p)$  corresponds to  $\frac{a}{p}x^2 + bxy + cpy^2$  (which now might not be of level  $p$ ). Recall that in Section 3.1 we fixed a residue  $r \pmod{2p}$  such that  $r^2 \equiv d_K \pmod{4p}$ . One can now check by direct computation on ideals that

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

if  $a > 0$ , and similarly,

$$\left( \mathbb{Z} \left( -\frac{a}{p} \sqrt{d_K} \right) + \mathbb{Z} \frac{d - b\sqrt{d_K}}{2} \right) \left( \mathbb{Z} p + \mathbb{Z} \frac{r - \sqrt{d_K}}{2} \right) = \mathbb{Z}(-a\sqrt{d_K}) + \mathbb{Z} \frac{d - b\sqrt{d_K}}{2}$$

if  $a < 0$ . This implies that when projected to the modular curve  $\text{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}$  of level 1, we have

$$p \mathcal{C}_A(p) = \mathcal{C}_{AA_p}(1),$$

where  $A_p = [p, \frac{r - \sqrt{d_K}}{2}] \in \text{Cl}_K^+$  (using the notation (3.1)). This implies that

$$\begin{aligned} \sum_{A \in \text{Cl}_K^+} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle \chi(A) &= \sum_{A \in \text{Cl}_K^+} \int_{\mathcal{C}_A(p)} E_{2,p}(z) dz \chi(A) \\ &= \frac{6}{p-1} (1 - \chi(J)) (\overline{\chi(A_p)} - 1) L(\chi, 0), \end{aligned} \tag{5.7}$$

using that  $(p-1)E_{2,p}(z) = pE_2^*(pz) - E_2^*(z)$ . Note that the above vanishes when  $\chi(J) = 1$  or  $\chi(A_p) = 1$ .

By the functional equation for odd class group  $L$ -functions [DIT18, p. 13] and the fact that class group characters are self-dual (i.e.,  $\chi \circ c = \overline{\chi}$  where  $c$  denotes complex conjugation), we conclude for  $\chi \in \widehat{\text{Cl}_K^+}$  with  $\chi(J) = -1$  that

$$\overline{L(\chi, 0)} = L(\overline{\chi}, 0) = L(\chi \circ c, 0) = L(\chi, 0) = \pi^{-2} d_K^{1/2} L(\chi, 1) > 0. \tag{5.8}$$

By standard estimates for  $L$ -functions on the critical line for all  $\varepsilon > 0$ , there is some uniform (but ineffective) constant  $c_\varepsilon > 0$  such that

$$L(\chi, 0) \geq c_\varepsilon d_K^{1/2-\varepsilon}. \tag{5.9}$$

**Proposition 5.2.** *Let  $H \leq \text{Cl}_K^+$  be a subgroup such that  $A_p \notin H$  and  $J \notin H$ . Then for each  $\varepsilon > 0$ , there exists a constant  $c_\varepsilon > 0$  such that*

$$\sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle \leq -c_\varepsilon \frac{d_K^{1/2-\varepsilon}}{p}. \tag{5.10}$$

*Proof.* By orthogonality, the Hecke formula (5.7), and equations (5.8) and (5.9), we conclude

$$\sum_{A \in H} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle = \frac{|H|}{|\text{Cl}_K^+|} \sum_{\chi \in \widehat{\text{Cl}}_K^+ : \chi|_H=1} \sum_{A \in \text{Cl}_K^+} \langle [\mathcal{C}_A(p)], \omega_E(p) \rangle \chi(A) \tag{5.11}$$

$$= -\frac{12|H|}{(p-1)|\text{Cl}_K^+|} \sum_{\substack{\chi \in \widehat{\text{Cl}}_K^+ : \\ \chi|_H=1, \chi(J)=-1}} \sum_{A \in \text{Cl}_K^+} (1 - \text{Re } \chi(A_p)) L(\chi, 0) \tag{5.12}$$

$$\leq -c_\varepsilon \frac{d_K^{1/2-\varepsilon}}{p} \frac{|H|}{|\text{Cl}_K^+|} \sum_{\substack{\chi \in \widehat{\text{Cl}}_K^+ : \\ \chi|_H=1, \chi(J)=-1}} (1 - \text{Re } \chi(A_p)), \tag{5.13}$$

where we are using that  $(1 - \text{Re } \chi(A_p)) \geq 0$ . Now we observe that if  $H' := \langle H, J \rangle \leq \text{Cl}_K^+$  denotes the group generated by  $H \leq \text{Cl}_K^+$  and  $J = (\sqrt{d_K})$ , then we can write

$$\sum_{\substack{\chi \in \widehat{\text{Cl}}_K^+ : \\ \chi|_H=1, \chi(J)=-1}} (1 - \text{Re } \chi(A_p)) = \frac{|\text{Cl}_K^+|}{|H'|} - \text{Re} \left( \sum_{\chi \in \widehat{\text{Cl}}_K^+ / H'} \chi(A_p H) - \sum_{\chi \in \widehat{\text{Cl}}_K^+ / H'} \chi(A_p H') \right) \tag{5.14}$$

$$= \begin{cases} \frac{|\text{Cl}_K^+|}{2|H|}, & A_p \notin H' \\ \frac{|\text{Cl}_K^+|}{|H|}, & A_p \in H' \end{cases} \tag{5.15}$$

Here, we are using that  $J \notin H$ , which implies that  $|H'| = 2|H|$ , and  $A_p \notin H$ . Inserting this into the above yields the wanted estimate. □

### 5.2. The cuspidal contribution

In the case of cuspidal Weyl sums, we will use the following formula due to Popa [Pop06, Theorem 6.3.1] (see also the first formula in the introduction of [Pop06]):

$$\left| \sum_{A \in \text{Cl}_K^+} \langle [\mathcal{C}_A(p)], \omega_f \rangle \bar{\chi}(A) \right|^2 = d_K^{1/2} L(\pi_f \otimes \pi_\chi, 1/2), \tag{5.16}$$

where  $\pi_f$  denotes the  $\text{GL}_2$ -automorphic representation associated to  $f$ ,  $\pi_\chi$  denotes the  $\text{GL}_2$ -automorphic representation associated to  $\chi$  via automorphic induction and  $L(\pi_f \otimes \pi_\chi, 1/2)$  denotes the central value of the (finite part of the) Rankin–Selberg  $L$ -function of  $\pi_f$  and  $\pi_\chi$ . This implies

$$\sum_{A \in \text{Cl}_K^+} \langle [\mathcal{C}_A(p)], \omega_f^\pm \rangle \bar{\chi}(A) = \frac{d_K^{1/4}}{2} \left( |\epsilon_{f, \chi}| L(\pi_f \otimes \pi_\chi, 1/2)^{1/2} \pm |\epsilon_{f, \bar{\chi}}| L(\pi_f \otimes \pi_{\bar{\chi}}, 1/2)^{1/2} \right), \tag{5.17}$$

with  $|\epsilon_{f, \chi}| = |\epsilon_{f, \bar{\chi}}| = 1$ . By the subconvexity bound

$$L(\pi_f \otimes \pi_\chi, 1/2) \ll_f d_K^{1/2-1/1057}$$

due to Michel [Mic04, Theorem 2], we conclude the following bound on the twisted Weyl sums:

$$\sum_{A \in \text{Cl}_K^+} \langle [C_A(p)], \omega_f^\pm \rangle \bar{\chi}(A) \ll_f d_K^{1/2-1/2114}. \tag{5.18}$$

Combining this with the lower bound for the Eisenstein contribution yields the proof of our main result.

*Proof of Theorem 5.1.* Let  $H \leq \text{Cl}_K^+$  be a subgroup as in the statement. Then by Proposition 5.2, we have for all  $\varepsilon > 0$  that there exists an absolute constant  $c_\varepsilon > 0$  such that

$$\left| \sum_{A \in H} \langle [C_A(p)], \omega_E(p) \rangle \right| = - \sum_{A \in H} \langle [C_A(p)], \omega_E(p) \rangle \geq c_\varepsilon d_K^{1/2-\varepsilon} p^{-1}. \tag{5.19}$$

Now the result follows from (5.6) (with  $C = 1$ ) by dividing through by (5.19) and bounding the cuspidal contribution using (5.18). Here, we are using that the number of narrow class group characters  $\chi$  such that  $\chi|_H = 1$  is exactly  $\frac{|\text{Cl}_K^+|}{|H|}$ , and the remaining terms in (5.6) do not depend on  $d_K$ .  $\square$

### 6. A homological version of the sup norm problem

In this section, we will deal with the problem of obtaining a version of Theorem 5.1 where the level  $p$  is allowed to vary. By (5.6) and (5.17), this requires, first of all, bounds for  $L(\pi_f \otimes \pi_\chi, 1/2)$  in terms of both  $d_K$  and  $p$  which in the case of genus characters has been studied by Petrow and Young [PY19]. Secondly, we will need estimates for the cap product pairings  $\langle v_f^\pm, \omega \rangle$  (with  $f \in \mathcal{B}_p$ ) in terms of  $p$ . This can be thought of as an analogue of the *sup norm problem* from *arithmetic quantum chaos* (see, for example, [BH10]) as we will explain below. We will have to restrict to certain compatible families of  $\omega \in H^1(Y_0(p), \mathbb{R})$ , which in our case will be dual bases of basic bases of level  $p$  as in Definition 4.7.

**Theorem 6.1.** *Let  $p$  be prime and let  $B \subset H_1(Y_0(p), \mathbb{Z})$  be a basic basis of level  $p$ . Then for  $\omega \in B^*$  (the dual basis of  $B$  with respect to (3.4)), we have*

$$\sum_{\varepsilon \in \{\pm\}} \sum_{f \in \mathcal{B}_p} |\langle v_f^\varepsilon, \omega \rangle|^2 \ll_\varepsilon p^{1+\varepsilon}. \tag{6.1}$$

See Remark 6.2 below for thoughts on the optimal bound that one can expect. Notice that since the newforms  $f \in \mathcal{B}_p$  are Hecke normalized, the dual vectors  $v_f^\pm$  are very subtle quantities as they are related to the *minimal periods*  $c_f^\pm$  of  $f$  (characterized when  $f$  has rational coefficients by  $(c_f^+)^{-1} \text{Re } f(z) dz \in H^1(X_0(p), \mathbb{Z})$  being primitive, and similarly for  $c_f^-$ ). Upper bounding  $c_f^\pm$  in terms of the level is extremely hard as any polynomial bound implies (a weak form of) the ABC-conjecture (see [Go02]).

To put Theorem 6.1 into perspective, it is again useful to compare to the imaginary quadratic analogue of supersingular reduction of CM elliptic curves. Recall that in Section 2 we defined an  $n$ -dimensional vector space  $H_0(X^{p,\infty}, \mathbb{R})$  having a canonical basis  $e_1, \dots, e_n$  corresponding to the connected components of the conic curve  $X^{p,\infty}$  as defined in (2.2), which, in turn, can be identified with isomorphism classes of supersingular elliptic curves defined over  $\mathbb{F}_{p^2}$ . There is a canonical bilinear form  $\langle \cdot, \cdot \rangle_{\text{ss}}$  on  $H_0(X^{p,\infty}, \mathbb{R})$  given by

$$\langle e_i, e_j \rangle_{\text{ss}} := \delta_{i,j} w_i, \tag{6.2}$$

where  $w_i$  is the size of the endomorphism ring of the elliptic curves corresponding to  $e_i$ . The homology group  $H_0(X^{p,\infty}, \mathbb{R})$  carries a natural action of the Hecke algebra (of level  $p$ ) defined via correspondences, and these linear operators are self-adjoint with respect to the bilinear form  $\langle \cdot, \cdot \rangle_{\text{ss}}$ . The homology group  $H_0(X^{p,\infty}, \mathbb{R})$  is (by the Jacquet–Langlands correspondence) isomorphic as a Hecke algebra to the space

$\mathcal{M}_2(p)$  of holomorphic modular forms of level  $p$  and weight 2. In particular, one can associate to each Hecke eigenform  $f$  a unique element  $e_f \in H_0(X^{p,\infty}, \mathbb{R})$  such that  $\langle e_f, e_f \rangle_{ss} = 1$  and  $T_\ell e_f = \lambda_f(\ell)e_f$ , where  $\lambda_f(\ell)$  denotes the  $\ell$ -th Hecke eigenvalues of  $f$  with  $\ell \neq p$  prime. In this setting, we are interested in upper bounds for the ‘ $L^r$ -norms’:

$$\|e_f\|_r := \left( \sum_{i=1}^n |\langle e_f, e_i \rangle_{ss}|^r \right)^{1/r}, \quad 1 \leq r < \infty, \quad \|e_f\|_\infty := \max_{1 \leq i \leq n} |\langle e_f, e_i \rangle_{ss}|.$$

Putting  $r = 2$  and using Parseval, one obtains  $\|e_f\|_2 = 1$  which implies the trivial (or convexity) bound  $\|e_f\|_r \leq 1$  for all  $r \geq 2$ . In the case  $r = \infty$ , Blomer and Michel [BM11] were the first to go beyond this by proving  $\|e_f\|_\infty \ll p^{-1/6+\varepsilon}$ . This was recently improved by Khayutin–Nelson–Steiner [KNS22, Corollary 2.3] who obtained  $\|e_f\|_\infty \ll p^{-1/4+\varepsilon}$ . Liu–Masri–Young [LMY15, Proposition 1.12] obtained the  $L^4$ -bound  $\|e_f\|_4 \ll p^{-1/8+\varepsilon}$ .

The fact that the Hecke operators are self-adjoint with respect to the bilinear form (6.2) on  $H_0(X^{p,\infty}, \mathbb{R})$  implies that the dual basis

$$e_1^*, \dots, e_n^* \in H_0(X^{p,\infty}, \mathbb{R})^*$$

of  $e_1, \dots, e_n$  (which we can identify with  $w_1^{-1}\langle \cdot, e_1 \rangle_{ss}, \dots, w_n^{-1}\langle \cdot, e_n \rangle_{ss}$ ) have the same behavior under the action of the Hecke algebra. This is why in the proof of the level aspect version (2.5) by Liu–Masri–Young one does not need any nontrivial input to bound the factors  $\langle e_f, e_i \rangle_{ss}$  appearing when spectrally expanding; one can simply employ the trivial Parseval bound  $\sum_{f \in \mathcal{B}_p} |\langle e_f, e_i \rangle_{ss}|^2 \leq 1$ . This is not necessarily the case for the basic bases  $B \subset H_1(Y_0(p), \mathbb{R})$  in our setting. This means, in particular, that we do not know whether

$$\langle v, \omega_f^\pm \rangle \langle v_f^\pm, v^* \rangle \tag{6.3}$$

is positive or not for  $v \in B$  and  $v^* \in B^*$  such that  $\langle v, v^* \rangle = 1$ . This lack of positivity makes it hard to obtain any bound at all.

### 6.1. The method of proof

The first natural approach to bounding the left-hand side of (6.1) would be to use a version of the amplified pre-trace formula approach to the sup norm problem in arithmetic quantum chaos (see, for example, [BH10]). This is, however, not very effective due to the possible non-positivity of (6.3) (i.e., we are working in a Banach space rather than a Hilbert space). However, it has become apparent from the work of Steiner [Ste20], Khayutin–Steiner [KS20] and Khayutin–Nelson–Steiner [KNS22] that one in many cases can obtain very strong sup norm bounds by using the theta correspondence. We will employ a version of this argument in our setting. The basic fact is that for  $v_0 \in H_1(X_0(p), \mathbb{R})$  and  $\omega \in H^1(X_0(p), \mathbb{R})$ ,

$$\sum_{n \geq 1} \langle T_n v_0, \omega \rangle e^{2\pi i n z} \tag{6.4}$$

defines a cusp form of level  $p$ . This follows easily by the fact that the  $\pm 1$  isotypic components of  $\iota$  acting on  $H_1(X_0(p), \mathbb{R})$  are isomorphic as Hecke modules to the space of holomorphic cusp forms  $\mathcal{S}_2(p)$  of weight 2 and level  $p$  (which is an incarnation of the Eichler–Shimura isomorphism [Shi94, Section 8.2]). This can be seen as an instance of the theta correspondence (see [DHRV21, Section 5.3]). We will give a simple proof below of the exact statement that we need. Now the key fact is that if  $\iota v_0 = \pm v_0$ , then the  $L^2$ -norm of (6.4) is (by spectrally expanding) equal to

$$\sum_{f \in \mathcal{B}_p} |\langle v_0, \omega_f^\pm \rangle|^2 |\langle v_f^\pm, \omega \rangle|^2.$$

Note that now we have obtained positivity for free! Thus, we are reduced to, on the one hand, a lower bound for  $|\langle v_0, \omega_f^\pm \rangle|^2$  which we resolve in Corollary 6.5, and on the other hand, a bound for the  $L^2$ -norm of (6.4) which is essentially equal to

$$\sum_{1 \leq n \leq p} \frac{|\langle T_n v_0, \mathbb{P}_{\text{cusp}} \omega \rangle|^2}{n}. \tag{6.5}$$

The above is an analogue of the second moment matrix count that appears in, for example, [KS20, Section 9]. We will bound (6.5) using geometric coding of geodesics.

**Remark 6.2.** Let  $B$  be a basic basis of level  $p$  and let  $v \in B$  and  $v^* \in B^*$  such that  $\langle v, v^* \rangle = 1$ . Then we get by expanding in the Hecke basis

$$1 = \sum_{f \in \mathcal{B}_{p, \pm}} \langle v_f^\pm, v^* \rangle \langle v, \omega_f^\pm \rangle.$$

As we will see in Corollary 6.5, we have on average that  $\langle v, \omega_f^\pm \rangle \asymp 1$ . Thus, if the classes  $v_f^\pm$  are perfectly distributed (relative to  $B$ ), then we would have  $\langle v_f^\pm, v^* \rangle \asymp 1/p$ . Thus, the strongest sup norm conjecture one can hope for is

$$\langle v_f^\pm, \omega \rangle \ll_\varepsilon p^{-1+\varepsilon},$$

for  $\omega \in B^*$ . This might, however, be too much to hope for.

**Remark 6.3.** The dual sup norm problem corresponds to obtaining bounds for the modular symbols

$$\sup_{v \in B} |\langle v, \omega_f \rangle|,$$

for  $v \in B$  as  $p \rightarrow \infty$ , where  $\omega_f = f(z)dz$  with  $f \in \mathcal{B}_p$  a holomorphic Hecke eigenform of weight 2 and level  $p$ . Let  $v = \{\gamma z, z\} \in H_1(Y_0(p), \mathbb{Z})$  with  $\gamma \in \mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$ , which we may assume is hyperbolic. Picking  $z$  on the fixed circle of  $\gamma$  and using Hölder’s inequality, we see that

$$|\langle v, \omega_f \rangle| \leq 2\pi \int_{\mathcal{C}_\gamma} |\text{Im}(z)f(z)| \frac{dz}{\text{Im}(z)} \ll_\varepsilon p^{1/4+\varepsilon},$$

using the sup norm estimate coming from [KNS22, Theorem 1.6]. Here,  $\mathcal{C}_\gamma$  denotes the closed geodesic on  $X_0(p)$  associated to  $\gamma$  with hyperbolic length  $\ell(\mathcal{C}_\gamma) \ll \log p$ . Notice that we also have the trivial bound

$$\left( \sum_{f \in \mathcal{B}_p} |\langle v, \omega_f \rangle|^2 \right)^{1/2} \leq 2\pi \int_{\mathcal{C}_\gamma} \left( \sum_{f \in \mathcal{B}_p} |\text{Im}(z)f(z)|^2 \right)^{1/2} \frac{dz}{\text{Im}(z)} \ll_\varepsilon p^{1/2+\varepsilon}$$

by Minkowski’s integral inequality and the pre-trace formula (thanks to the referee for pointing this out).

### 6.2. A lower bound for modular symbols

In this section, we will construct the homology class  $v_0 \in H_1(X_0(p), \mathbb{R})$  mentioned above such that  $\langle v_0, \omega_f^\pm \rangle$  is ‘not too small’ in terms of the level. We will be constructing  $v_0$  in terms of classes of the shape  $\{\frac{x}{p}, \infty\}$  for  $1 \leq x < p$  (using the notation (3.6)). This reduces the problem to a lower bound for the additive twist  $L$ -functions of  $f$ . One could naively try to calculate the average of all additive twists with  $1 \leq x < p$ . But this yields very poor results as there is a lot of cancellation in this sum. Instead, we use that the second moment coincides with the second moment of Dirichlet twists of the  $L$ -function of  $f$

(by the Birch–Stevens formula). By Cauchy–Schwarz, it suffices to calculate the first moment of these  $L$ -functions instead. Such a calculation is quite standard using an approximate functional equation. The exact case we need, however, does not seem to have been considered before as we are in the case of joint ramification. For the sake of completeness, we have provided a detailed argument below.

To be more precise, let  $f \in \mathcal{B}_p$  be a Hecke normalized eigenform and let  $\chi$  be a primitive Dirichlet character modulo  $q$  with  $p|q$ . Then we define the twisted  $L$ -function as the analytic continuation of

$$L(f, \chi, s) := \sum_{n \geq 1} \frac{\lambda_f(n)\chi(n)}{n^{s+1/2}}$$

(here the Ramanujan conjecture is  $|\lambda_f(n)| \leq d(n)n^{1/2}$ ). By the Birch–Stevens formula (see, for example, [Nor21, Proposition 6.1]), we have for a primitive  $\chi$  with  $\chi(-1) = \pm 1$

$$i^{(\chi(-1)-1)/2} \tau(\bar{\chi}) L(f, \chi, s) = \sum_{a \bmod q}^* L^\pm(f, a/q, s) \overline{\chi(a)}, \tag{6.6}$$

where

$$L^\pm(f, a/q, s) := \frac{L(f, a/q, s) \pm L(f, -a/q, s)}{1 + i \pm (1 - i)}, \quad L(f, a/q, s) := \sum_{n \geq 1} \frac{\lambda_f(n)e(na/q)}{n^{s+1/2}}, \quad \text{Re } s > 1$$

are the additive twist  $L$ -series satisfying analytic continuation and the functional equation

$$\gamma_f(s) q^s L(f, a/q, s) = -\gamma_f(1-s) q^{1-s} L(f, -\bar{a}/q, 1-s),$$

where  $a\bar{a} \equiv 1 \pmod q$  and  $\gamma_f(s) = \Gamma(s + 1/2)(2\pi)^{-s}$  (see e.g. [Nor21, Proposition 3.3]). This implies the functional equation

$$\gamma_f(s) q^s L(f, \chi, s) = \varepsilon_\chi \gamma_f(1-s) q^{1-s} L(f, \bar{\chi}, 1-s), \tag{6.7}$$

where  $\varepsilon_\chi = \frac{-\tau(\chi)^2}{q}$ . Note that by inserting the Fourier expansion of  $f$  and interchanging sum and integral, we arrive at the identity

$$\langle \{ \frac{a}{q}, \infty \}, \omega_f^\pm \rangle = -L^\pm(f, \frac{a}{q}, 1/2). \tag{6.8}$$

**Lemma 6.4.** *Let  $f \in \mathcal{B}_p$  and let  $q$  be an integer such that  $p|q$ . Then we have (uniformly in  $p$  and  $f$ ) that*

$$\frac{1}{\varphi_\pm^*(q)} \sum_{\chi \bmod q}^{*, \pm} L(f, \chi, 1/2) = 1 + O_\varepsilon(q^{-1/4+\varepsilon}),$$

where the sum is restricted to primitive characters  $\chi$  modulo  $q$  with  $\chi(-1) = \pm 1$ , and  $\varphi_\pm^*(q)$  denotes the total number of such characters.

*Proof.* By the approximate functional equation [IK04, Theorem 5.3] using (6.7), we can write the moment in question as

$$\frac{2}{\varphi^*(q)} \sum_{\chi \bmod q}^{*, \pm} \sum_{n \geq 1} \frac{\lambda_f(n)}{n} \left( \chi(n) V\left(\frac{n}{q^\lambda}\right) + \varepsilon_\chi \bar{\chi}(n) V\left(\frac{n}{q^{2-\lambda}}\right) \right), \tag{6.9}$$

for some  $0 < \lambda < 2$  to be chosen and  $V : \mathbb{R}_{>0} \rightarrow \mathbb{R}$  a smooth, rapidly decaying function satisfying  $V(y) = 1 + O(y^{1/3})$  as  $y \rightarrow 0$  (see [IK04, Proposition 5.4]). Now by a simple application of Möbius

inversion, we get that for  $(n, q) = 1$ ,

$$\sum_{\chi \bmod q}^{*,\pm} \chi(n) = \sum_{d|q} \mu(q/d)\varphi(d)(\delta_{n\equiv 1(d)} \pm \delta_{n\equiv -1(d)})/2, \tag{6.10}$$

$$\sum_{\chi \bmod q}^{*,\pm} \varepsilon_\chi \bar{\chi}(n) = -\frac{1}{q} \sum_{d|q} \mu(q/d)\varphi(d) \frac{K_2(n; d) \pm K_2(-n; d)}{2}, \tag{6.11}$$

where  $K_2(n; d) = \sum_{x+y\equiv n(d)} e((x+y)/d)$  denotes the usual 2-dimensional Kloosterman sum. By Weil’s bound and standard estimates, we conclude that for  $(n, q) = 1$ ,

$$\sum_{\chi \bmod q}^{*,\pm} \varepsilon_\chi \bar{\chi}(n) \ll_\varepsilon q^{1/2+\varepsilon}. \tag{6.12}$$

Now for the primary part of the sum (6.9) we obtain using the analytic properties of  $V$  mentioned above as well as (6.10), we get

$$\frac{1}{\varphi_\pm^*(q)} \sum_{\chi \bmod q}^{*,\pm} \sum_{n \geq 1} \frac{\lambda_f(n)}{n} \chi(n) V\left(\frac{n}{q^\lambda}\right) \tag{6.13}$$

$$= 1 + O_\varepsilon \left( \frac{q^{1-\lambda/3}}{\varphi_\pm^*(q)} + \frac{1}{\varphi_\pm^*(q)} \sum_{d|q} \varphi(d) \sum_{\substack{n \equiv \pm 1(d) \\ 1 < n \ll q^{\lambda+\varepsilon}}} \frac{|\lambda_f(n)|}{n} \right), \tag{6.14}$$

with the main term corresponding to  $n = 1$ . Now using the Ramanujan bound  $|\lambda_f(n)| \leq d(n)n^{1/2} \ll_\varepsilon n^{1/2+\varepsilon}$ , we get

$$\sum_{\substack{n \equiv \pm 1(d) \\ 1 < n \ll q^{\lambda+\varepsilon}}} \frac{|\lambda_f(n)|}{n} \ll_\varepsilon \sum_{1 \leq m \ll q^{\lambda+\varepsilon}/d} \sum_{\pm} (md \pm 1)^{-1/2+\varepsilon} \ll q^{\lambda/2+\varepsilon} d^{-1}, \tag{6.15}$$

which bounds the error-term in (6.13) by  $O_\varepsilon((q^{1-\lambda/3} + q^{\lambda/2+\varepsilon})/\varphi^*(q))$ . Similarly, by using (6.11), we can bound the dual sum by  $O_\varepsilon(q^{3/2-\lambda/2+\varepsilon}/\varphi_\pm^*(q))$ . Choosing  $\lambda = 3/2$  and recalling that  $\varphi_\pm^*(q) \gg_\varepsilon q^{1-\varepsilon}$ , we get the wanted asymptotic formula.  $\square$

**Corollary 6.5.** *There exists an absolute constant  $c_0 > 0$  such that for any prime  $p$ ,  $f \in \mathcal{B}_p$  and  $\varepsilon \in \{\pm\}$ , we have*

$$\frac{1}{p-1} \sum_{0 < a < p} |\langle \{\frac{a}{p}, \infty\}, \omega_f^\varepsilon \rangle|^2 \geq c_0. \tag{6.16}$$

*Proof.* First of all, recall that the following matrices generate  $\Gamma_0(p)$ :

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} a & -(aa^*+1)/p \\ p & -a^* \end{pmatrix}: \quad 0 < a < p,$$

where  $0 < a^* < p$  are such that  $aa^* \equiv -1 \pmod p$  (as is proved in Section 4.2). Since  $\omega_f^\pm$  are nonzero cohomology classes vanishing on parabolic elements, we conclude that the left-hand side of (6.16) is always nonzero.

We now define the Fourier transform of

$$(\mathbb{Z}/p\mathbb{Z})^\times \ni a \mapsto \langle \{\frac{a}{p}, \infty\}, \omega_f^\pm \rangle$$

as follows for a Dirichlet character  $\chi \pmod p$ :

$$\widehat{L}_f^\pm(\chi) := \sum_{a \pmod p}^* \langle \{\frac{a}{p}, \infty\}, \omega_f^\pm \rangle \overline{\chi(a)}.$$

For  $\chi$  primitive with  $\chi(-1) = \pm 1$ , we have by the equality (6.8) and the Birch–Stevens formula (6.6) that

$$\widehat{L}_f^\pm(\chi) = -i^{(\chi(-1)-1)/2} \tau(\overline{\chi}) L(f, \chi, 1/2).$$

Now by Parseval, we have

$$\sum_{0 < a < p} |\langle \{\frac{a}{p}, \infty\}, \omega_f^\pm \rangle|^2 = \frac{1}{p-1} \sum_{\chi \pmod p} |\widehat{L}_f^\pm(\chi)|^2 \geq \frac{p}{p-1} \sum_{\chi \pmod p}^{*,\pm} |L(f, \chi, 1/2)|^2.$$

Using the previous lemma, we conclude by Cauchy–Schwarz that

$$\frac{1}{\varphi_\pm^*(p)} \sum_{\chi \pmod p}^{*,\pm} |L(f, \chi, 1/2)|^2 \geq \left( \frac{1}{\varphi_\pm^*(p)} \sum_{\chi \pmod p}^{*,\pm} L(f, \chi, 1/2) \right)^2 \geq 1/2,$$

for  $p$  large enough. This yields the wanted lower bound since  $\varphi_\pm^*(p) \asymp p - 1$ . □

**Remark 6.6.** Using a subconvexity bound for  $L(f \otimes \chi, 1/2)$  as in [MV10], the above implies that there is some constant  $\delta > 0$  such that  $L(f \otimes \chi, 1/2)$  is nonvanishing for at least  $\gg p^\delta$  of both odd and even characters. This improves on Merel [Mer09, Corollaire 2], who has shown using modular symbol techniques that for  $f \in \mathcal{B}_p$ , there exists at least one odd and one even character  $\chi$  of conductor  $p$  such that  $L(f \otimes \chi, 1/2) \neq 0$ . One can probably get much better results using mollification.

### 6.3. Reduction to a counting problem

We will now use the theta correspondence in a simple form to reduce the sup norm problem to a certain counting problem.

**Lemma 6.7.** *Let  $\omega \in H^1(X_0(p), \mathbb{R})$  be a cuspidal cohomology class. Then we have*

$$\sum_{\epsilon \in \{\pm\}} \sum_{f \in \mathcal{B}_p} |\langle v_f^\epsilon, \omega \rangle|^2 \ll_\epsilon p^{-1+\epsilon} \sum_{n \geq 1} \frac{\frac{1}{p} \sum_{0 < x < p} (|\langle T_n \{\frac{x}{p}, \infty\}, \omega \rangle|^2 + |\langle T_n \{\frac{1}{x}, 0\}, \omega \rangle|^2)}{n} e^{-8n/p}$$

uniformly in  $p$ , where  $\langle \cdot, \cdot \rangle$  denotes the cap product pairing (3.4) between homology and cohomology and  $T_n$  denotes the  $n$ -th Hecke operator.

*Proof.* For  $0 < x < p$  and  $\epsilon = (\epsilon_1, \epsilon_2)$  with  $\epsilon_i \in \{\pm 1\}$ , we consider the following class in the compactly supported homology

$$v_x^\epsilon := (1 + \epsilon_1 \iota)(1 + \epsilon_2 W_p) \left\{ \frac{x}{p}, \infty \right\} \tag{6.17}$$

$$= \left\{ \frac{x}{p}, \infty \right\} + \epsilon_1 \left\{ -\frac{x}{p}, \infty \right\} + \epsilon_2 \left\{ -\frac{1}{x}, 0 \right\} + \epsilon_1 \epsilon_2 \left\{ \frac{1}{x}, 0 \right\} \in H_1(X_0(p), \mathbb{Z}). \tag{6.18}$$

By construction,  $v_x^\epsilon$  is contained in the  $\epsilon_1$  eigenspace of the involution  $\iota$  as in (3.14), and in the  $\epsilon_2$  eigenspace of the Fricke involution  $W_p$  as in (3.13). Associated to  $0 < x < p$ ,  $\epsilon \in \{(\pm 1, \pm 1)\}$  and  $\omega \in H^1(X_0(p), \mathbb{R})$  (suppressed in the notation), we define  $g : \mathbb{H} \rightarrow \mathbb{C}$  by

$$g(z) := \sum_{n \geq 1} \langle T_n v_x^\epsilon, \omega \rangle e^{2\pi i n z}. \tag{6.19}$$

By expanding in the Hecke basis of cohomology, we obtain

$$\langle T_n v_x^\epsilon, \omega \rangle = \sum_{\substack{f \in \mathcal{B}_p: \\ W_p f = \epsilon_2 f}} \lambda_f(n) \langle v_x^\epsilon, \omega_f^{\epsilon_1} \rangle \langle v_f^{\epsilon_1}, \omega \rangle.$$

By newform theory, we know that for  $f \in \mathcal{B}_p$ , we have  $\lambda_f(n) = a_f(n)$ , where  $a_f(n)$  denotes the Fourier coefficients (at  $\infty$ ) of  $f$ . Thus, we conclude that

$$g(z) = \sum_{\substack{f \in \mathcal{B}_p: \\ W_p f = \epsilon_2 f}} \langle v_x^\epsilon, \omega_f^{\epsilon_1} \rangle \langle v_f^{\epsilon_1}, \omega \rangle f(z),$$

and, in particular,  $g \in \mathcal{S}_2(p)$  is a holomorphic cusp form of weight 2 for  $\Gamma_0(p)$  contained in the  $\epsilon_2$  eigenspace of  $W_p$ . This can be seen as an instance of the theta correspondence as explained above. By orthogonality of Hecke eigenforms, this implies the key identity

$$\langle g, g \rangle_{\text{Pet}} = \sum_{\substack{f \in \mathcal{B}_p: \\ W_p f = \epsilon_2 f}} \langle f, f \rangle_{\text{Pet}} |\langle v_x^\epsilon, \omega_f^{\epsilon_1} \rangle \langle v_f^{\epsilon_1}, \omega \rangle|^2, \tag{6.20}$$

where  $\langle f, g \rangle_{\text{Pet}} = \int_{Y_0(p)} f(z) \overline{g(z)} dx dy$  denotes the Petersson inner-product on  $\mathcal{S}_2(p)$ . Recall from (4.3) that the subgroup

$$\Gamma_0^*(p) := \langle W_p, \Gamma_0(p) \rangle \subset \text{PSL}_2(\mathbb{R})$$

has  $\Gamma_0(p)$  as an index two subgroup and has a fundamental domain contained in

$$\{z \in \mathbb{H} : |\text{Re } z| \leq 1/2, \text{Im } z \geq \sqrt{3}/(2p)\}.$$

By unfolding and using that  $|g|^2$  is invariant under the Fricke involution  $W_p$ , by construction we arrive at the following bound:

$$\begin{aligned} \sum_{\substack{f \in \mathcal{B}_p: \\ W_p f = \epsilon_2 f}} \langle f, f \rangle_{\text{Pet}} |\langle v_x^\epsilon, \omega_f^{\epsilon_1} \rangle \langle v_f^{\epsilon_1}, \omega \rangle|^2 &= 2 \int_{\Gamma_0^*(p) \backslash \mathbb{H}} |g(z)|^2 dx dy \ll \int_{\sqrt{3}/(2p)}^\infty \int_{-1/2}^{1/2} |g(z)|^2 dx dy \\ &\ll \sum_{n \geq 1} |\langle T_n v_x^\epsilon, \omega \rangle|^2 \int_{\sqrt{3}/(2p)}^\infty e^{-4\pi n y} dy \\ &\ll \sum_{n \geq 1} \frac{|\langle T_n v_x^\epsilon, \omega \rangle|^2}{n} e^{-2\sqrt{3}\pi n/p}. \end{aligned} \tag{6.21}$$

Notice that for  $f \in \mathcal{B}_p$  with  $W_p f = \epsilon_2 f$ , we have by self-adjointness of  $\iota$  and  $W_p$  with respect to the cap product pairing that

$$\langle v_x^\epsilon, \omega_f^{\epsilon_1} \rangle = 4 \langle \{ \frac{x}{p}, \infty \}, \omega_f^{\epsilon_1} \rangle.$$

Now we sum over  $0 < x < p$ ,  $\epsilon \in \{(\pm 1, \pm 1)\}$  and apply Corollary 6.5 to the left-hand side of (6.21). Finally, by expressing the Petersson inner product in terms of a special value of an adjoint  $L$ -function (see, for example, [PR18, (8.6)]) and using the lower bounds of Hoffstein–Lockhart [HL94], we arrive at

$$\langle f, f \rangle_{\text{Pet}} = \frac{pL(\text{sym}^2 f, 1)}{8\pi^3} \gg_{\epsilon} p^{1-\epsilon}.$$

Finally, we have by linearity of  $T_n$  that

$$|\langle T_n v_x^{\epsilon}, \omega \rangle|^2 \ll |\langle T_n \{ \frac{x}{p}, \infty \}, \omega \rangle|^2 + |\langle T_n \{ -\frac{x}{p}, \infty \}, \omega \rangle|^2 + |\langle T_n \{ \frac{1}{x}, \infty \}, \omega \rangle|^2 + |\langle T_n \{ -\frac{1}{x}, \infty \}, \omega \rangle|^2,$$

from which we conclude the wanted inequality since  $2\sqrt{3}\pi > 8$ . □

### 6.4. The counting argument

By Lemma 6.7, we are reduced to a certain second moment count. We think of this as an analogue of the matrix counts that show up in most approaches to the (arithmetic) sup norm problem (see, for example, [Tem15]). Recall the Ramanujan bound  $|\lambda_f(n)| \leq n^{1/2}d(n)$  for Hecke eigenforms  $f \in \mathcal{B}_p$ . This implies for any class  $v \in H_1(X_0(p), \mathbb{R})$  and  $\omega \in H^1(X_0(p), \mathbb{R})$ , we have

$$\langle T_n v, \omega \rangle \ll_{v, \omega} n^{1/2}d(n), \quad \text{as } n \rightarrow \infty.$$

This is, however, not useful as for us the main point is exactly the dependence on  $v$  and  $\omega$ . Our approach is to use the explicit description of the Hecke operators and then lift the counting from  $\Gamma_0(p)^{\text{ab}}$  to  $\Gamma_0(p)$ . As a first step, we use geometric coding as in Proposition 4.1 to obtain the following.

**Lemma 6.8.** *Let  $\omega \in H^1(X_0(p), \mathbb{R})$  be a cuspidal cohomology class and  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(p)$  with  $c > 0$ . Then we have*

$$|\langle \{\gamma\infty, \infty\}, \omega \rangle| \ll \left( \max_{0 < x < p} |\langle \{ \frac{x}{p}, \infty \}, \omega \rangle| \right) \log c, \tag{6.22}$$

uniformly in  $p$ .

*Proof.* Observe that  $\{\gamma\infty, \infty\}$  only depends on  $a/c \pmod 1$ . This means that without changing the homology class, we may assume that  $a, d$  are integers satisfying  $2c \leq |a + d| \leq 4c$  as well as  $-c < a - d < c$ . This implies that all entries of  $\gamma$  are  $O(c)$ , and that the half circle  $S_{\gamma}$  fixed by  $\gamma$  intersects the standard fundamental domain  $\mathcal{F}_{\text{std}}$  for  $\text{PSL}_2(\mathbb{Z})$ . Recall Zagier’s fundamental polygon  $\mathcal{F}'_{\text{Zag}}(p)$  for  $\Gamma_0(p)$  defined in Section 4.2 which consists of  $\text{PSL}_2(\mathbb{Z})$ -translates of  $\mathcal{F}_{\text{std}}$ , as well as the fundamental polygon  $\mathcal{F}_{\text{Zag}}(p) := W_p \mathcal{F}'_{\text{Zag}}(p)$ . This latter polygon has the advantage that all entries of the associated side pairing transformations are of size  $\leq p$ , whereas for  $\mathcal{F}'_{\text{Zag}}(p)$ , the lower left entry can be of magnitude  $p^2$ . Notice also that the assumptions on  $\gamma$  above ensure that the half circle  $S_{\gamma}$  fixed by  $\gamma$  intersects the Siegel domain  $\{z \in \mathbb{H} : |\text{Re } z| \leq 1/2, \text{Im } z \geq 1\}$  which is contained in  $\mathcal{F}_{\text{Zag}}(p)$ .

We now consider the geometric coding of  $\gamma$  with respect to  $\mathcal{F}_{\text{Zag}}(p)$  as in Corollary 4.2. This yields an expression

$$\gamma = \sigma_1 \cdots \sigma_N,$$

where  $\sigma_i \in \mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$  are side pairing transformations. When considering the class  $\{\gamma\infty, \infty\} \in H_1(X_0(p), \mathbb{Z})$  in the compact homology, we can ignore the parabolic elements among the  $\sigma_i$ ’s. More precisely, if  $\sigma'_1, \dots, \sigma'_\ell$  is the subsequence of  $\sigma_1, \dots, \sigma_N$  consisting of non-parabolic elements, then we have

$$\{\gamma^\infty, \infty\} = \sum_{i=1}^{\ell} \{\sigma'_i \infty, \infty\} \in H_1(X_0(p), \mathbb{Z}).$$

Now  $\ell$  is exactly the number of intersections between the geodesic from  $w$  to  $\gamma w$  (for any  $w \in \mathcal{F}_{\text{Zag}}(p) \cap S_\gamma$ ) and  $\Gamma_0(p)$ -translates of the sides of  $\mathcal{F}_{\text{Zag}}(p)$  being paired by non-parabolic elements. This is equal to the number of intersections between the geodesic from  $w$  to  $\gamma'w$  (where  $\gamma' = W_p \gamma W_p^{-1}$  and  $w \in \mathcal{F}'_{\text{Zag}}(p) \cap S_{\gamma'}$ ) and  $\Gamma_0(p)$ -translates of the non-parabolic sides of  $\mathcal{F}'_{\text{Zag}}(p)$  (this follows by conjugating everything which preserves parabolicity). Recall that for  $\mathcal{F}'_{\text{Zag}}(p)$ , all sides containing a cusp are paired by parabolic elements. Since  $\mathcal{F}'_{\text{Zag}}(p)$  consists of  $\text{PSL}_2(\mathbb{Z})$ -translates of  $\mathcal{F}_{\text{std}}$ , we can bound  $\ell$  by the number of intersections between the geodesic from  $w$  to  $\gamma'w$  (for  $w \in \mathcal{F}_{\text{std}} \cap S_{\gamma'}$ ) and  $\text{PSL}_2(\mathbb{Z})$ -translates of the non-parabolic side of  $\mathcal{F}_{\text{std}}$  – that is, the arc

$$\{z \in \mathbb{H} : |z| = 1, -1/2 < \text{Re } z < 1/2\}.$$

It now follows from a result of Eichler [Eic65, Satz 1] that

$$\ell \ll \log((a')^2 + (b')^2 + (c')^2 + (d')^2) \ll \log c,$$

where  $a', b', c', d'$  are the entries of  $\gamma'$ , which by the assumptions on the entries of  $\gamma$  are all  $O(pc) = O(c^2)$ . By definition of  $\mathcal{F}_{\text{Zag}}(p)$ , we have for all  $1 \leq i \leq \ell$  that  $\sigma'_i \infty$  is of the form  $\frac{x}{p}$  with  $0 < x < p$ . Thus, we conclude

$$|\langle \{\gamma^\infty, \infty\}, \omega \rangle| \ll \ell \left( \max_{0 < x < p} |\langle \{\frac{x}{p}, \infty\}, \omega \rangle| \right) \ll \log c \left( \max_{0 < x < p} |\langle \{\frac{x}{p}, \infty\}, \omega \rangle| \right),$$

as wanted. □

This implies the following result for basic bases. Recall the definition of the cuspidal projection operator  $\mathbb{P}_{\text{cusp}}$  in (3.23).

**Corollary 6.9.** *Let  $B$  be a basic basis of level  $p$ . Then for  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(p)$  with  $c > 0$  and  $\omega \in B^*$ , we have*

$$|\langle \{\gamma^\infty, \infty\}, \mathbb{P}_{\text{cusp}} \omega \rangle| \ll \log c,$$

uniformly in  $p$ .

*Proof.* By Proposition 6.8, we are reduced to proving that

$$\max_{0 < x < p} |\langle \{\frac{x}{p}, \infty\}, \mathbb{P}_{\text{cusp}} \omega \rangle| \ll 1$$

for  $\omega \in B^*$ . Recall that we have

$$\langle \{\frac{x}{p}, \infty\}, \mathbb{P}_{\text{cusp}} \omega \rangle = \langle \{\gamma_x z, z\}, \omega \rangle - \langle v_E(p), \omega \rangle \langle \{\gamma_x z, z\}, \omega_E(p) \rangle,$$

where  $\gamma_x := \begin{pmatrix} x & -(xx^*+1)/p \\ p & -x^* \end{pmatrix}$  with  $0 < x^* < p$  such that  $xx^* \equiv -1 \pmod p$  and  $z \in \mathbb{H}$  is arbitrary. Now by Corollary 4.6, we have

$$\langle v_E(p), \omega \rangle \ll 1, \quad \langle \{\gamma_x z, z\}, \omega \rangle \ll 1,$$

and by (3.21), we have

$$\langle \{\gamma_x z, z\}, \omega_E(p) \rangle \ll \frac{p|x - x^*| + p^2}{p^2} \ll 1,$$

which yields the wanted bound. □

### 6.5. Proof of Theorem 6.1

Combining all of the above, we are now ready to prove our sup norm bounds.

*Proof of Theorem 6.1.* Let  $B$  be a basic basis of level  $p$ . By the definition of the Hecke operators acting on homology (3.12), we have for  $\omega \in B^*$  by Corollary 6.9

$$\begin{aligned}
 |\langle T_n\{\frac{x}{p}, \infty\}, \mathbb{P}_{\text{cusp}} \omega \rangle| &\leq \sum_{\substack{ad=n, \\ (a,p)=1}} \sum_{0 \leq b < d} \left| \left\langle \left\{ \frac{ax+bp}{pd}, \infty \right\}, \mathbb{P}_{\text{cusp}} \omega \right\rangle \right| \\
 &\ll \sum_{\substack{ad=n, \\ (a,p)=1}} \sum_{0 \leq b < d} \log pd \ll_{\varepsilon} p^{\varepsilon} n^{1+\varepsilon}.
 \end{aligned} \tag{6.23}$$

Here, we are using that  $(ax + bp, p) = 1$ , meaning that  $\frac{ax+bp}{pd}$  is of the shape  $\gamma_{\infty}$  for  $\gamma \in \Gamma_0(p)$  with left lower entry of size  $O(pd)$ . By changing contours, we have the following equality of compactly supported homology classes:

$$\left\{ \frac{1}{x}, 0 \right\} = \left\{ \left( \frac{\bar{x}}{x\bar{x}-1} \ 1 \right) 0, 0 \right\} = \left\{ \left( \frac{\bar{x}}{x\bar{x}-1} \ 1 \right) \infty, \infty \right\} = \left\{ \frac{\bar{x}}{x\bar{x}-1}, \infty \right\}, \quad 0 < \bar{x} < p : x\bar{x} \equiv 1 \pmod{p}.$$

A similar argument as above, again using Corollary 6.9, yields

$$|\langle T_n\{\frac{1}{x}, \infty\}, \mathbb{P}_{\text{cusp}} \omega \rangle| \ll_{\varepsilon} p^{\varepsilon} n^{1+\varepsilon}, \quad \omega \in B^*.$$

Thus, by Lemma 6.7, we conclude that

$$\sum_{f \in \mathcal{B}_p} |\langle v_f^{\pm}, \omega \rangle|^2 \ll_{\varepsilon} p^{-1+\varepsilon} \sum_{n \geq 1} \frac{n^{2+\varepsilon}}{n} e^{-8n/p} \ll_{\varepsilon} p^{1+\varepsilon}, \quad \omega \in B^*,$$

as wanted. □

## 7. The level aspect

Using the second moment bound from Theorem 6.1, we are ready to prove our main result.

**Theorem 7.1.** *Let  $p$  be prime and let  $K$  be a real quadratic field of discriminant  $d_K$  with no unit of norm  $-1$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$  and  $\mathfrak{p}_1 \notin H$  where  $H = (Cl_K^+)^2$ .*

*For  $B$  a basic basis of level  $p$  and  $\omega \in B^*$ , we have*

$$\frac{\sum_{A \in H} \langle [C_A(p)], \omega \rangle}{|\sum_{A \in H} \langle [C_A(p)], \omega_E(p) \rangle|} = -\langle v_E(p), \omega \rangle + O_{\varepsilon}(d_K^{-1/12+\varepsilon} p^{2+\varepsilon}), \tag{7.1}$$

where  $v_E(p), \omega_E(p)$  denotes the Eisenstein classes in homology and cohomology as defined in (3.19) and (3.16), respectively.

*Proof.* Starting from (5.6), we want to bound the cuspidal contribution on the right-hand side, which by (5.17) can be expressed in terms of  $L$ -functions. Recall that class group characters  $\chi$  such that  $\chi_{|(Cl_K^+)^2}$  (i.e., genus characters) correspond to factorizations  $d_1 d_2 = d_K$  in terms of fundamental discriminants, and we have

$$L(\pi_f \otimes \pi_{\chi}, s) = L(f \otimes \chi_{d_1}, s) L(f \otimes \chi_{d_2}, s),$$

where  $\chi_{d_i} = \left( \frac{d_i}{\cdot} \right)$  are quadratic characters. Now we apply Cauchy–Schwarz followed by Hölder’s inequality with exponents  $(3, 3, 3)$  combined with Theorem 6.1 as well as the third moment bound of Petrow and Young [PY19, Theorem 1]:

$$\begin{aligned}
 & \left| \sum_{A \in H} \langle [\mathcal{C}_A(p)], \mathbb{P}_{\text{cusp}} \omega \rangle \right|^2 \\
 & \ll d_K^{1/2+\varepsilon} \left( \sum_{f \in \mathcal{B}_{p,\pm}} |\langle v_f^\pm, \omega \rangle|^2 \right) \left( \sum_{d_1 d_2 = d_K} \sum_{f \in \mathcal{B}_p} L(f \otimes \chi_{d_1}, 1/2) L(f \otimes \chi_{d_2}, 1/2) \right) \\
 & \ll d_K^{1/2+\varepsilon} \left( \sum_{f \in \mathcal{B}_{p,\pm}} |\langle v_f^\pm, \omega \rangle|^2 \right) \\
 & \quad \times \sum_{d_1 d_2 = d_K} p^{1/3} \left( \sum_{f \in \mathcal{B}_p} L(f \otimes \chi_{d_1}, 1/2)^3 \right)^{1/3} \left( \sum_{f \in \mathcal{B}_p} L(f \otimes \chi_{d_2}, 1/2)^3 \right)^{1/3} \tag{7.2} \\
 & \ll_\varepsilon p^{2+\varepsilon} d_K^{5/6+\varepsilon},
 \end{aligned}$$

for all  $\omega \in B^*$  (using here also that  $L(f \otimes \chi_{d_i}, 1/2) \geq 0$ ). It follows from Proposition 5.2 that the Eisenstein contribution on the right-hand side of (5.6) is  $\gg_\varepsilon d_K^{1/2-\varepsilon} p^{-1}$ . Inserting all of this into (5.6) yields as wanted.  $\square$

Recall that given a basis  $B \subset V_p = H_1(Y_0(p), \mathbb{R})$ , we get an associated isomorphism  $V_p \cong \mathbb{R}^{2g+1}$  by mapping  $B$  to the standard basis of  $\mathbb{R}^{2g+1}$ . By pulling back the  $L^r$ -norm with respect to the standard basis of  $\mathbb{R}^{2g+1}$ , we get the following norm on  $V_p$ :

$$\|v\|_{B,r} := \left( \sum_{\omega \in B^*} |\langle v, \omega \rangle|^r \right)^{1/r} \quad \text{for } 1 \leq r < \infty, \quad \|v\|_{B,\infty} := \sup_{\omega \in B^*} |\langle v, \omega \rangle|, \tag{7.3}$$

with  $B^* \subset H^1(Y_0(p), \mathbb{R})$  the dual basis of  $B$ . Combining Theorem 7.1 and equations (5.2) and (5.4), we arrive at the following (we will skip the details).

**Theorem 7.2.** *Let  $p$  be prime and let  $K$  be a real quadratic field of discriminant  $d_K$  with no unit of norm  $-1$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$  and  $\mathfrak{p}_1 \notin H$  where  $H = (\text{Cl}_K^+)^2$ .*

*For  $B$  a basic basis of level  $p$ , we have*

$$\left\| \frac{\sum_{A \in H} \langle [\mathcal{C}_A(p)] \rangle}{\| \sum_{A \in H} \langle [\mathcal{C}_A(p)] \rangle \|_{B,\infty}} + v_E(p) \right\|_{B,\infty} \ll_\varepsilon d_K^{-1/12+\varepsilon} p^{2+\varepsilon}. \tag{7.4}$$

### 8. Applications

In this section, we will present certain applications of Theorem 7.2.

#### 8.1. A group theoretic application

Recall the independent generators of  $\Gamma_0(p)$  described in Section 4.3 coming from special fundamental polygons. Given a real quadratic field  $K$  of discriminant  $d_K$  such that  $p$  splits in  $K$ , the oriented closed geodesics  $\mathcal{C}_A(p)$  associated to  $A \in \text{Cl}_K^+$  corresponds to a conjugacy class of matrices inside  $\Gamma_0(p)$  (given by the matrices  $\gamma_Q$  defined in (3.3) where  $Q \in \mathcal{Q}_{K,p}$  runs through integral binary quadratic form of discriminant  $d_K$  and level  $p$  corresponding to  $A$  under the identification (3.2)). We will apply our results to understand the representation of the matrices  $\gamma_Q$  in terms of the independent generators (4.5) (4.6), (4.7) and (4.8) coming from a special fundamental domain of  $\Gamma_0(p)$ .

**Corollary 8.1.** *Let  $p$  be prime and let  $\mathcal{P}(p)$  be a special fundamental polygon for  $\Gamma_0(p)$ . Consider a real quadratic field  $K$  of discriminant  $d_K$  with no unit of norm  $-1$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$*

and such that  $\mathfrak{p}_1 \notin (\text{Cl}_K^+)^2$ . Then for  $d_K \gg_\varepsilon p^{24+\varepsilon}$ , there is some class  $A \in (\text{Cl}_K^+)^2$  such that none of the matrices  $\gamma_Q \in \Gamma_0(p)$  with  $Q \in \mathcal{Q}_{K,p}$  corresponding to  $A$  are contained in the subgroup generated by the matrices in (4.6), (4.7) and (4.8).

*Proof.* Let  $B_{\text{sp}} \subset H_1(Y_0(p), \mathbb{R})$  be a basis obtained from the side pairing transformations of  $\mathcal{P}(p)$ . Note that  $v_E(p) \in B_{\text{sp}}$  and let  $\omega_0 \in B_{\text{sp}}^*$  be characterized by  $\langle v_E(p), \omega_0 \rangle = 1$ . Let  $B$  be a basic basis of level  $p$  containing  $v_E(p)$ . Then by expanding in this basis, we get by Theorem 7.2 that

$$\begin{aligned} \left\langle \frac{\sum_{A \in H} [C_A(p)]}{\|\sum_{A \in H} [C_A(p)]\|}, \omega_0 \right\rangle &= - \left\langle \frac{v_E(p)}{\|v_E(p)\|}, \omega_0 \right\rangle + \left\langle \frac{\sum_{A \in H} [C_A(p)]}{\|\sum_{A \in H} [C_A(p)]\|} + \frac{v_E(p)}{\|v_E(p)\|}, \omega_0 \right\rangle \\ &= -1 + \sum_{v \in B} \langle v, \omega_0 \rangle \left\langle \frac{\sum_{A \in H} [C_A(p)]}{\|\sum_{A \in H} [C_A(p)]\|} + \frac{v_E(p)}{\|v_E(p)\|}, v^* \right\rangle \\ &= -1 + O_\varepsilon \left( d_K^{-1/12+\varepsilon} p^{2+\varepsilon} \sum_{v \in B} |\langle v, \omega_0 \rangle| \right), \end{aligned} \tag{8.1}$$

where  $\|\cdot\| = \|\cdot\|_{B,\infty}$  and  $v^* \in B^*$  is characterized by  $\langle v, v^* \rangle = 1$ . For  $v \in B - \{v_E(p)\}$ , let  $\gamma_v \in \mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$  be such that

$$v = \{z, \gamma_v z\} \in H_1(Y_0(p), \mathbb{R}),$$

using the notation (3.6) (note that the lower left entry of  $\gamma_v$  is equal to  $p$ ). Now for  $Y > 0$ , consider a curve  $c_Y : [0, 1] \rightarrow \mathbb{H}$  connecting the three points

$$\gamma_v^{-1}\infty + iY, \quad \gamma_v\infty + iY, \quad \gamma_v(\gamma_v^{-1}\infty + iY) = \gamma_v\infty + ip^{-2}Y^{-1}$$

by a (Euclidean) straight line. We assume that  $Y$  is large enough so that

$$\{x + iY : \min(\gamma_v\infty, \gamma_v^{-1}\infty) < x < \max(\gamma_v\infty, \gamma_v^{-1}\infty)\} \subset \mathcal{P}(p).$$

Observe that by Corollary 4.2, the quantity  $|\langle v, \omega_0 \rangle|$  is bounded by the number of intersections between  $c_Y$  and  $\Gamma_0(p)$ -translates of  $\{iy : y > 0\}$  (which is a side of the special fundamental polygon  $\mathcal{P}(p)$ ). Since  $0 < \gamma_v\infty, \gamma_v^{-1}\infty < 1$ , there is no such intersection for the horizontal segment of  $c_Y$ . Intersections with the vertical segment correspond to integers  $a, b, c, d$  such that

$$p|c, \quad ad - bc = 1, \quad \text{and} \quad \frac{a}{c} < \gamma_v\infty < \frac{b}{d} \text{ or } \frac{b}{d} < \gamma_v\infty < \frac{a}{c}. \tag{8.2}$$

In particular, we have  $\gamma_v \neq \frac{a}{c}$ , which implies

$$\left| \gamma_v\infty - \frac{a}{c} \right| \geq \frac{1}{|c|}, \quad \left| \frac{b}{d} - \frac{a}{c} \right| = \frac{1}{|cd|},$$

but this contradicts (8.2). This implies that  $\langle v, \omega_0 \rangle = 0$  for  $v \in B - \{v_E(p)\}$  which by (8.1) yields the wanted error-term. □

From the above, Corollary 1.7 follows in the case where  $K$  has wide class number one and narrow class number two by writing out explicitly the matrices  $\gamma_Q$  associated to  $Q \in \mathcal{Q}_{K,p}$  as in (3.3) and recalling that since  $K$  does not have a unit of norm  $-1$ , the condition  $\mathfrak{p}_1 \notin (\text{Cl}_K^+)^2$  is equivalent to  $\mathfrak{p}_1$  not having a generator of positive norm.

### 8.2. An application to modular forms

Recall the following definition mentioned in the introduction for a modular form  $f \in \mathcal{M}_2(p)$ :

$$M_f := \inf\{c \geq 0 : |a_f(n)| \leq c\sigma_1(n), n \geq 1\} < \infty,$$

where  $a_f(n)$  denotes the Fourier coefficients of  $f$  (at  $\infty$ ) and  $\sigma_1(n) = \sum_{d|n} d$ . We have the following nonvanishing result for cycle integrals of modular forms.

**Corollary 8.2.** *Let  $p$  be prime and let  $f \in \mathcal{M}_2(p)$  be a holomorphic modular form of weight 2 and level  $p$  with constant Fourier coefficient equal to  $a_f(0) = 1$ . Consider a real quadratic field  $K$  of discriminant  $d_K$  with no unit of norm  $-1$  such that  $p$  splits in  $K$  with  $p\mathcal{O}_K = \mathfrak{p}_1\mathfrak{p}_2$  and  $\mathfrak{p}_1 \notin (\mathcal{O}_K^\times)^2$ . Then for  $d_K \gg_\varepsilon (M_f)^{12+\varepsilon} p^{48+\varepsilon}$  there is some  $A \in (\mathcal{O}_K^\times)^2$  such that*

$$\langle [C_A(p)], f(z)dz \rangle = \int_{C_A(p)} f(z)dz \neq 0.$$

*Proof.* Let

$$f(z)dz = \omega_+ + i\omega_- \in H^1(Y_0(p), \mathbb{R}) \oplus iH^1(X_0(p), \mathbb{R})$$

be the cohomology class (with complex coefficients) associated to  $f$ . By expanding in a basic basis  $B \subset H_1(Y_0(p), \mathbb{R})$  as in (8.1) and using Cauchy–Schwarz followed up by (7.4), we conclude that

$$\left\langle \frac{\sum_{A \in H} [C_A(p)]}{\|\sum_{A \in H} [C_A(p)]\|_{B, \infty}} - v_E(p), \omega_\pm \right\rangle \ll_\varepsilon p^{3+\varepsilon} d_K^{-1/12+\varepsilon} \sup_{v \in B} |\langle v, \omega_\pm \rangle|.$$

Now by the assumptions on the Fourier expansion of  $f$ , we get

$$f(x + iy) \ll 1 + M_f \sum_{n \geq 1} \sigma_1(n) e^{-2\pi yn} \ll_\varepsilon 1 + M_f \int_1^\infty t^{1+\varepsilon} e^{-2\pi ty} \frac{dt}{t} \ll_\varepsilon \max(1, M_f y^{-1-\varepsilon}).$$

In order to bound the quantities  $\langle v, \omega_\pm \rangle$  for  $v \in B$ , we recall that we can write  $v = \{\gamma z, z\}$  for  $z \in \mathbb{H}$  (which we may choose freely) and some matrix  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{S}(\mathcal{F}_{\text{Zag}}(p))$  as in Section 4.2. In particular, we observe that by construction, we have  $a, b, c, d \ll p$  and  $c \in \{0, p\}$ . If  $c = 0$ , then clearly  $\langle v, \omega_\pm \rangle \in \{0, 1\}$ . If  $c = p$ , we make the following convenient choice  $z = \frac{-d}{p} + \frac{i}{p}$  satisfying  $\gamma z = \frac{a}{p} + \frac{i}{p}$ . This gives

$$\langle v, \omega_\pm \rangle = \int_{\frac{a}{p}}^{-d/p} \frac{f(x + ip^{-1}) \pm \overline{f(x + ip^{-1})}}{2i^{(1 \mp 1)/2}} dx \ll_\varepsilon M_f \frac{|a + d|}{p} p^{1+\varepsilon} \ll_\varepsilon M_f p^{1+\varepsilon},$$

which we plug into the above. Since  $\langle v_E(p), \omega_+ \rangle = 1$  and  $\langle v_E(p), \omega_- \rangle = 0$ , we conclude that  $\langle \sum_{A \in H} [C_A(p)], f(z)dz \rangle$  is indeed nonvanishing for  $d_K \gg_\varepsilon (M_f)^{12+\varepsilon} p^{48+\varepsilon}$ , as wanted.  $\square$

Now Corollary 1.8 follows from the above in the case of wide class number one and narrow class number two.

**Acknowledgements.** The author would like to thank Peter Humphries, Raphael Steiner, Daniel Kriz and Farrell Brumley for useful discussions. The author would also like to thank the referee for a careful and insightful reading with many valuable suggestions, as well as pointing out a mistake in a key step of an earlier version of the proof of the main result.

**Competing interest.** The author have no competing interest to declare.

**Funding statement.** This research was supported by a grant from the Independent Research Fund Denmark DFF(1025-00020B).

## References

- [ALMW22] M. Aka, M. Luethi, P. Michel and A. Wieser, ‘Simultaneous supersingular reductions of cm elliptic curves’, *Journal für die reine und angewandte Mathematik (Crelles Journal)* 2022(786) (2022), 1–43.
- [BD96] M. Bertolini and H. Darmon, ‘Heegner points on Mumford-Tate curves’, *Invent. Math.* **126**(3) (1996), 413–456.
- [Bea83] A. F. Beardon, *The Geometry of Discrete Groups* vol. 91 (Springer, Cham, 1983).
- [BH10] V. Blomer and R. Holowinsky, ‘Bounding sup-norms of cusp forms of large level’, *Invent. Math.* **179**(3) (2010), 645–681.
- [BM11] V. Blomer and P. Michel, ‘Sup-norms of eigenfunctions on arithmetic ellipsoids’, *Int. Math. Res. Not.* **2011**(21) (2011), 4934–4966.
- [Bv22] C. Burrin and F. von Essen, ‘Windings of prime geodesics’, Preprint, 2022, [arXiv:2209.06233](https://arxiv.org/abs/2209.06233).
- [Dar94] H. Darmon, ‘Heegner points, Heegner cycles, and congruences’, in *Elliptic Curves and Related Topics* vol. 4 of *CRM Proc. Lecture Notes* (Amer. Math. Soc., Providence, RI, 1994), 45–59.
- [DHRV21] H. Darmon, M. Harris, V. Rotger and A. Venkatesh, The derived Hecke algebra for dihedral weight one forms. February 2021.
- [DIT18] W. Duke, Ö. Imamoğlu and Á. Tóth, ‘Kronecker’s first limit formula, revisited’, *Res. Math. Sci.* **5**(2) (2018), 21.
- [DJK10] A. Dooms, E. Jespers and A. Konovalov, ‘From Farey symbols to generators for subgroups of finite index in integral group rings of finite groups’, *J. K-Theory* **6**(2) (2010), 263–283.
- [DKLP22] N. M. Doan, S.-H. Kim, M. L. Lang and S. P. Tan, ‘Optimal independent generating system for the congruence subgroups  $\Gamma_0(p)$ ’, Preprint, 2022, [arXiv:2209.13937](https://arxiv.org/abs/2209.13937).
- [Duk88] W. Duke, ‘Hyperbolic distribution problems and half-integral weight Maass forms’, *Invent. Math.* **92**(1) (1988), 73–90.
- [Eic65] M. Eichler, ‘Grenzkreisgruppen und kettenbruchartige Algorithmen’, *Acta Arith.* **11** (1965), 169–180.
- [ELMV11] M. Einsiedler, E. Lindenstrauss, P. Michel and A. Venkatesh, ‘Distribution of periodic torus orbits and Duke’s theorem for cubic fields’, *Ann. of Math. (2)* **173**(2) (2011), 815–885.
- [ELMV12] M. Einsiedler, E. Lindenstrauss, P. Michel and A. Venkatesh, ‘The distribution of closed geodesics on the modular surface, and Duke’s theorem’, *Enseign. Math. (2)* **58**(3–4) (2012), 249–313.
- [EOY05] N. Elkies, K. Ono and T. Yang, ‘Reduction of CM elliptic curves and modular function congruences’, *Int. Math. Res. Not.* **2005**(44) (2005), 2695–2707.
- [Fra98] J. Franke, ‘Harmonic analysis in weighted  $L_2$ -spaces’, *Ann. Sci. Éc. Norm. Supér. (4)* **31**(2) (1998), 181–279.
- [Gol02] D. Goldfeld, ‘Modular forms, elliptic curves and the abc-conjecture’, in *A Panorama in Number Theory or The View from Baker’s Garden* (Cambridge University Press, Cambridge, 2002), 128–147.
- [Hat02] A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, 2002).
- [HL94] J. Hoffstein and P. Lockhart, ‘Coefficients of Maass forms and the Siegel zero. Appendix: An effective zero-free region, by Dorian Goldfeld, Jeffrey Hoffstein and Daniel Lieman’, *Ann. Math. (2)* **140**(1) (1994), 161–181.
- [HN22] P. Humphries and A. C. Nordentoft, ‘Sparse equidistribution of geometric invariants of real quadratic fields’, Preprint, 2022, [arXiv:2211.05890](https://arxiv.org/abs/2211.05890).
- [IK04] H. Iwaniec and E. Kowalski, *Analytic Number Theory* vol. 53 of *American Mathematical Society Colloquium Publications* (American Mathematical Society, Providence, RI, 2004).
- [Kan09] B. Kane, ‘CM liftings of supersingular elliptic curves’, *J. Théor. Nombres Bordx.* **21**(3) (2009), 635–663.
- [Kat96] S. Katok, ‘Coding of closed geodesics after Gauss and Morse’, *Geom. Dedicata* **63**(2) (1996), 123–145.
- [KNS22] I. Khayutin, P. D. Nelson and R. S. Steiner, ‘Theta functions, fourth moments of eigenforms, and the sup-norm problem II’, Preprint, 2022, [arXiv:2207.12351](https://arxiv.org/abs/2207.12351).
- [KS20] I. Khayutin and R. S. Steiner, ‘Theta functions, fourth moments of eigenforms, and the sup-norm problem I’, Preprint, 2020, [arXiv:2009.07194](https://arxiv.org/abs/2009.07194).
- [Kul91] R. S. Kulkarni, ‘An arithmetic-geometric method in the study of the subgroups of the modular group’, *Am. J. Math.* **113**(6) (1991), 1053–1133.
- [LMY13] S.-C. Liu, R. Masri and M. P. Young, ‘Subconvexity and equidistribution of Heegner points in the level aspect’, *Compos. Math.* **149**(7) (2013), 1150–1174.
- [LMY15] S.-C. Liu, R. Masri and M. P. Young, ‘Rankin-Selberg  $L$ -functions and the reduction of CM elliptic curves’, *Res. Math. Sci.* **2** (2015), 23.
- [Mas71] B. Maskit, ‘On Poincaré’s theorem for fundamental polygons’, *Adv. Math.* **7** (1971), 219–230.
- [Mer09] L. Merel, ‘Symboles de Manin et valeurs de fonctions  $L$ ’, in *Algebra, Arithmetic, and Geometry: in Honor of Yu. I. Manin. Vol. II* vol. 270 (Birkhäuser Boston, Inc., Boston, MA, 2009), 283–309.
- [Mic04] P. Michel, ‘The subconvexity problem for Rankin-Selberg  $L$ -functions and equidistribution of Heegner points’, *Ann. Of Math. (2)* **160**(1) (2004), 185–236.
- [MV06] P. Michel and A. Venkatesh, ‘Equidistribution,  $L$ -functions and ergodic theory: on some problems of Yu. V. Linnik’, in *International Congress of Mathematicians* vol. II (Eur. Math. Soc., Zürich, 2006), 421–457.
- [MV10] P. Michel and A. Venkatesh, ‘The subconvexity problem for  $GL_2$ ’, *Publ. Math. Inst. Hautes Études Sci.* **111** (2010), 171–271.

- [NC20] A. C. Nordentoft and P. Constantinescu, 'Residual equidistribution of modular symbols and cohomology classes for quotients of hyperbolic  $n$ -space', *Trans. Amer. Math. Soc.* **375** (2022), 7001–7034.
- [Nor21] A. C. Nordentoft, 'Central values of additive twists of cuspidal  $L$ -functions', *Journal für die reine und angewandte Mathematik (Crelles Journal)* **2021**(776) (2021), 255–293.
- [Pop06] A. A. Popa, 'Central values of Rankin  $L$ -series over real quadratic fields', *Compos. Math.* **142**(4) (2006), 811–866.
- [Pop08] A. A. Popa, 'Whittaker newforms for Archimedean representations', *J. Number Theory* **128**(6) (2008), 1637–1645.
- [PR05] Y. N. Petridis and M. S. Risager, 'The distribution of values of the Poincaré pairing for hyperbolic Riemann surfaces', *J. Reine Angew. Math.* **579** (2005), 159–173.
- [PR08] Y. N. Petridis and M. S. Risager, 'Equidistribution of geodesics on homology classes and analogues for free groups', *Forum Math.* **20**(5) (2008), 783–815.
- [PR18] Y. N. Petridis and M. S. Risager, 'Arithmetic statistics of modular symbols', *Invent. Math.* **212**(3) (2018), 997–1053.
- [PS87] R. Phillips and P. Sarnak, 'Geodesics in homology classes', *Duke Math. J.* **55** (1987), 287–297.
- [PY19] I. Petrow and M. P. Young, 'A generalized cubic moment and the Petersson formula for newforms', *Math. Ann.* **373**(1–2) (2019), 287–353.
- [Shi94] G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions* vol. 11 (Princeton University Press, Princeton, NJ, 1994).
- [Ste20] R. S. Steiner, 'Sup-norm of Hecke-Laplace eigenforms on  $S^3$ ', *Math. Ann.* **377**(1–2) (2020), 543–553.
- [Tem15] N. Templier, 'Hybrid sup-norm bounds for Hecke-Maass cusp forms', *J. Eur. Math. Soc. (JEMS)* **17**(8) (2015), 2069–2082.
- [Voi09] J. Voight, 'Computing fundamental domains for Fuchsian groups', *J. Théor. Nombres Bordx.* **21**(2) (2009), 469–491.
- [Wal85] J.-L. Waldspurger, 'Sur les valeurs de certaines fonctions  $L$  automorphes en leur centre de symétrie', *Compositio Math.* **54**(2) (1985), 173–242.
- [Yan08] T. Yang, 'Minimal CM liftings of supersingular elliptic curves', *Pure Appl. Math. Q.* **4**(4) (2008), 1317–1326.
- [Zag85] D. Zagier, 'Modular parametrizations of elliptic curves', *Canad. Math. Bull.* **28** (1985), 372–384.