RANDOM MODULAR SYMBOLS

AVNER ASH AND DAN YASAKI

Abstract. Let Γ denote the subgroup $\Gamma_0^{\pm}(N)$ of $\mathrm{GL}_2(\mathbb{Z})$, N prime. Let V be the space of holomorphic modular forms for Γ. Let $V_\alpha \subset V$ denote the various Hecke eigenspaces, with the last V_α denoting the Eisenstein subspace. If $M \in V$ is a modular symbol, define the type of M to be $(t_1, \dots, t_k, t_E)$ where $t_\alpha = 1$ if the projection of M to V_α is nonzero, and $t_\alpha = 0$ otherwise.

For each $N \leq 100$, we compute the types of the modular symbols in an increasing series of concentric boxes. We prove one obstruction for a given type to occur, related to the existence of "Eisenstein primes." For any given type that survives this obstruction, we give computational evidence that the proportion of its occurrence in a box stabilizes as the boxes grow larger. We interpret the limit of this ratio (assuming it exists) as the box size goes to infinity as the probability that a random modular symbol will have this type.

Contrary to our original expectation, it does not appear to be the case that with probability 1 a random symbol will project nontrivially to each V_i. Whether the limit referred to in the previous paragraph actually exists, and why the limits have the various value that appear in our computations, are open questions.

1. General framework

Define

$$\Gamma_0^{\pm}(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}_2(\mathbb{Z}) \;\middle|\; c \equiv 0 \bmod N \right\} \subset \mathrm{GL}_2(\mathbb{Z}).$$

We work with this group, rather than the more usual $\Gamma_0(N) = \Gamma_0^{\pm}(N) \cap \mathrm{SL}_2(\mathbb{Z})$ because the phenomena we are investigating are already seen using the larger group, and the computations run twice as fast.

Fix a prime N, and let $\Gamma = \Gamma_0^{\pm}(N)$. Let $\mathcal{H}$ denote the tame Hecke algebra, generated over $\mathbb{Z}$ by T_ℓ for $\ell \nmid N$. Let V denote the cohomology $V = H^1(\Gamma, \mathbb{Q})$. It is an $\mathcal{H}$-module. Decompose V into a direct sum of $\mathbb{Q}\mathcal{H}$-irreducible subspaces:

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k \oplus E,$$

where E is the line spanned by the Eisenstein series. (Because N is prime, E is one-dimensional.) It is well-known that every package of Hecke eigenvalues belonging to a cuspform of weight 2 and level N appears in one of the V_i, and also that E appears, as written above.

For $i = 1, 2, \dots, k$, let $\pi_i \colon V \to V_i$ denote the projection onto V_i. [1] Let $\pi_E \colon V \to E$ denote projection onto the Eisenstein space E.

Date: November 10, 2023 18:03.

[1]Note that if we decompose V_i further into 1-dimensional $\overline{\mathbb{Q}}\mathcal{H}$ eigenspaces, and if $v \in V$, then the projection of v to any of these eigenlines is nonzero if and only if $\pi_i(v) \neq 0$. This is seen using the $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ action on $V \otimes_{\mathbb{Q}} \overline{\mathbb{Q}}$ and the fact that v is rational.

Definition 1.1. For $v \in V$, define the *type of* v to be

$$t(v) = (t_1, t_2, \ldots, t_k, t_E),$$

where

$$t_\alpha = \begin{cases} 1 & \text{if } \pi_\alpha(v) \neq 0, \\ 0 & \text{if } \pi_\alpha(v) = 0. \end{cases}$$

We say that a type is *cuspidal* if and only if $t_E = 0$.

The main task of this paper is to explore which types occur and with what probabilities. This question seems not to be treated in the literature. We will see that there is an obstruction to certain types occurring coming from Eisenstein primes. If a type escapes this obstruction, we can ask if it occurs, and if so, how often on average. The type $(1, 1, \ldots, 1)$ always survives the Eisenstein obstruction. A first guess might be that this type should occur with probability 1, since one might think that a random modular symbol is in "general position" in the cohomology. Our computations show that this is not likely to be the case. However, we have no explanation for the detailed probabilities we encounter in our computations.

Given two cusps $x, y \in \mathbb{P}^1(\mathbb{Q})$, by definition the modular symbol $[x, y]$ is the fundamental class of the geodesic in the bordified upper half plane $\overline{H}$ from x to y in the relative homology $H_1(\overline{H}, \partial\overline{H}, \mathbb{Q})$. Its image modulo Γ in $H_1(\overline{H}/\Gamma, \partial\overline{H}/\Gamma, \mathbb{Q})$ is denoted $[x, y]_\Gamma$. As recalled in [AY21], $H_1(\overline{H}/\Gamma, \partial\overline{H}/\Gamma, \mathbb{Q})$ can be canonically identified with $H_0(\Gamma, \mathrm{St} \otimes_{\mathbb{Z}} \mathbb{Q})$ where St is the Steinberg representation of $\mathrm{GL}_2(\mathbb{Q})$.

Let $\Gamma_0 = \Gamma \cap \mathrm{SL}_2(\mathbb{Z})$, and let $M_2(\Gamma_0)$ denote the complex vector space of holomorphic modular forms of weight 2 for Γ_0, and $M_2^{\mathrm{cusp}}(\Gamma_0)$ the subspace of cuspforms. Borel-Serre duality gives a canonical isomorphism

$$H_0(\Gamma_0, \mathrm{St}) \xrightarrow{\sim} H^1(\Gamma_0, \mathbb{Q}),$$

and the latter cohomology group tensored with $\mathbb{C}$ is isomorphic to $M_2(\Gamma_0)$. We define $H^1_{\mathrm{cusp}}(\Gamma_0, \mathbb{Q})$ to be $H^1(\Gamma_0, \mathbb{Q}) \cap M_2^{\mathrm{cusp}}(\Gamma_0)$.

In [AY21] we erroneously asserted that Borel-Serre duality holds if Γ_0 is replaced by Γ. This is not true, because Γ contains the element $J = \mathrm{diag}(1, -1)$ that reverses orientation on H. However, we may identify $H^1(\Gamma, \mathbb{Q})$ with the J-invariants in $H^1(\Gamma_0, \mathbb{Q})$. What we actually compute in this paper is $[x, y]_\Gamma$ in the space of J-invariants in $H_1(\overline{H}/\Gamma_0, \partial\overline{H}/\Gamma_0, \mathbb{Q})$ (where J acts on H by $z \mapsto -\bar{z}$). In this way, we will view $[x, y]_\Gamma$ as an element of $H^1(\Gamma, \mathbb{Q})$, after identifying $H^1(\Gamma_0, \mathbb{Q})$ with a $\mathbb{Q}$-subspace of $M_2(\Gamma_0)$ and using Borel-Serre duality for Γ_0.

This work was supported by a grant from the Simons Foundation (848154, DY).

2. An Eisenstein obstruction

Let $U \subset V$ be the set

$$U = \{[x, y]_\Gamma \mid x, y \in \mathbb{P}^1(\mathbb{Q})\},$$

and let $U' \subset U$ be the "cuspidal subset",

$$U' = \{[x, \gamma x]_\Gamma \mid x \in \mathbb{P}^1(\mathbb{Q}), \gamma \in \Gamma\}.$$

It is well-known that U generates $H^1(\Gamma, \mathbb{Q})$ over $\mathbb{Q}$. In [AY21] we show that the elements of U' are cuspidal cohomology classes.

Let $e = 1/0$, and let $f = 0/1$ in $\mathbb{P}^1(\mathbb{Q})$. Then $\pi_E([e, f]_\Gamma) \neq 0$, and $\pi_E([x, y]_\Gamma) = 0$ for all $[x, y]_\Gamma \in U'$. (Note that $[e, f]$ is fixed under J, so the Eisenstein class for Γ_0 is indeed in the cohomology of Γ.) It follows easily that U' generates $H^1_{\mathrm{cusp}}(\Gamma, \mathbb{Q})$ over $\mathbb{Q}$.

Proposition 2.1. *Let $\alpha = a/b \in \mathbb{P}^1(\mathbb{Q})$ be written in reduced form, and let N be prime. Then α is Γ-equivalent to $e = 1/0$ if N divides b and is Γ-equivalent to $f = 0/1$ otherwise.*

Proof. This is standard. See for example page 238 in [MS76]. □

Definition 2.2. Let $\Lambda \subset V$ denote the $\mathbb{Z}$-lattice generated by U. We call this the *modular symbol lattice*. Let $\Lambda' \subset \Lambda$ denote the sublattice generated by U', which we call the *cuspidal modular symbol lattice*.

Remark 2.3. Although Λ and Λ' are defined from an infinite amount of data (they are lattices spanned by an infinite number of modular symbols), we can reduce to a finite computation by identifying $\Gamma \backslash \mathrm{GL}_2(\mathbb{Z})$ with $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$.

It is clear that $\Lambda \otimes \mathbb{Q} = V$ and $\Lambda' \otimes \mathbb{Q} = V_1 \oplus \cdots \oplus V_k$. Our first goal is to study the subsets U and U' of these lattices.

Theorem 2.4.

(1) *The image of the cuspidal modular symbols in V is the cuspidal modular symbol lattice,*
$$U' = \Lambda'.$$
(2) *The image of the modular symbols in V is the union of three cosets in Λ/Λ',*
$$U = \Lambda' \cup ([e, f]_\Gamma + \Lambda') \cup (-[e, f]_\Gamma + \Lambda').$$

Proof. First we show (1). From [AY21, Theorem 5.3 (ii)], we have
$$U' = \{[f, \gamma f]_\Gamma \mid \gamma \in \Gamma\}.$$
It suffices to show that U' is closed under negation and addition.

Let $\gamma \in \Gamma$, and consider $[f, \gamma f]_\Gamma$. Then
$$-[f, \gamma f]_\Gamma = [\gamma f, f]_\Gamma = [f, \gamma^{-1} f]_\Gamma$$
which is in U', as desired.

Let $\gamma, \tau \in \Gamma$, and consider $[f, \gamma f]_\Gamma$ and $[f, \tau f]_\Gamma$ in U'. Then
$$\begin{aligned} [f, \gamma f]_\Gamma + [f, \tau f]_\Gamma &= [\gamma^{-1} f, f]_\Gamma + [f, \tau f]_\Gamma \\ &= [\gamma^{-1} f, \tau f]_\Gamma \\ &= [f, \gamma\tau f]_\Gamma, \end{aligned}$$
which is in U', as desired.

Next we show (2). Suppose $[x, y]_\Gamma \in U \setminus U'$ is not cuspidal. Then x is not equivalent to y. From Proposition 2.1, either

(1) x is Γ-equivalent to f and y is Γ-equivalent to e; or
(2) x is Γ-equivalent to e and y is Γ-equivalent to f.

Suppose $x = \gamma f$ and $y = \tau e$ for some $\gamma, \tau \in \Gamma$. Then
$$\begin{aligned} [x, y]_\Gamma + [e, f]_\Gamma &= [\gamma f, \tau e]_\Gamma + [e, f]_\Gamma \\ &= [\tau^{-1}\gamma f, e]_\Gamma + [e, f]_\Gamma \\ &= [\tau^{-1}\gamma f, f]_\Gamma \\ &= [f, \gamma^{-1}\tau f]_\Gamma. \end{aligned}$$

It follows that
$$[x,y]_\Gamma = -[e,f]_\Gamma + [f,\sigma f]_\Gamma, \tag{1}$$
for some $\sigma \in \Gamma$. In particular, $[x,y]_\Gamma \in -[e,f]_\Gamma + \Lambda'$.

If instead x is Γ-equivalent to e and y is Γ-equivalent to f, then a similar argument shows $[x,y]_\Gamma \in [e,f]_\Gamma + \Lambda'$. The result follows. $\square$

Next we look more closely at how the image of U sits inside the unimodular lattice Λ. For $i = 1, 2, \ldots, k$, let $\Lambda_i \subset V_i$ be the lattice
$$\Lambda_i = \operatorname{span}_{\mathbb{Z}}\{\pi_i([x,y]_\Gamma) \mid [x,y]_\Gamma \in U\}.$$
Let $\Lambda_i' \subseteq \Lambda_i$ denote the submodule generated by $\pi_i([x,y]_\Gamma)$ with $[x,y]_\Gamma \in U'$. It follows from Theorem 2.4 that the index $[\Lambda_i \colon \Lambda_i']$ is 1 if and only if $\pi_i([e,f]_\Gamma) \in \Lambda_i'$.

Definition 2.5. We call a type "cupsidal" if $t_E = 0$ and "noncuspidal" if $t_E = 1$.

From Theorem 2.4, we have that the image in V of the cuspidal modular symbols is the full cuspidal modular symbol lattice, $U' = \Lambda'$. Because the decomposition $V = \oplus_\alpha V_\alpha$ is rational, it follows that all 2^k cuspidal types will occur infinitely often in U'.

An obstruction to a non-cuspidal type occurring in U can be measured by a certain index:

Proposition 2.6. *Suppose $[\Lambda_i \colon \Lambda_i'] \neq 1$, and let $[x,y]_\Gamma \in V$ have type*
$$t([x,y]_\Gamma) = (t_1, t_2, \ldots, t_k, t_E).$$
Then $t_E = 1$ implies that $t_i = 1$.

Proof. Suppose $t_E = 1$. Then $[x,y]_\Gamma$ is not cuspidal. Without loss of generality, from (1), we have
$$\pi_i([x,y]_\Gamma) = \pi_i([f,\gamma f]_\Gamma) - \pi_i([e,f]_\Gamma),$$
for some $\gamma \in \Gamma$. Since $[f,\gamma f]_\Gamma$ is cuspidal, we have $\pi_i([f,\gamma f]_\Gamma) \in \Lambda_i'$. On the other hand, if $[\Lambda_i \colon \Lambda_i'] \neq 1$, then $\pi_i([e,f]) \notin \Lambda_i'$. Then
$$\pi_i([x,y]_\Gamma) = \pi_i([f,\gamma f]_\Gamma) - \pi_i([e,f]_\Gamma) \neq 0,$$
so $t_i = 1$, as desired, $\square$

Here is a theorem that was first suggested by our data:

Theorem 2.7. *Let $N \geq 11$, $i \geq 1$, and p a prime. Suppose the index $[\Lambda_i \colon \Lambda_i']$ is divisible by p. Then there is a newform of level N and weight 2 whose Hecke eigenvalue a_ℓ is congruent modulo p to $\ell + 1$ for all ℓ not dividing N. If $p \neq 2$, then p divides $N - 1$.*

Proof. Tensor everything with $\mathbb{C}$ so that we may identify modular symbols with holomorphic modular forms of weight 2. Let $\varepsilon \in E \cap \Lambda$ be primitive in Λ. Since from Theorem 2.4 we know that $\Lambda = \Lambda' + \mathbb{Z}[e,f]_\Gamma$, write
$$\varepsilon = [x,\gamma x]_\Gamma + b[e,f]_\Gamma$$
for some cusp x, some $\gamma \in \Gamma$, and some $b \in \mathbb{Z}$. Then
$$0 = \pi_i(\varepsilon) = \pi_i([x,\gamma x]_\Gamma) + b\pi_i([e,f]_\Gamma).$$
Since $\pi_i([x,\gamma x]_\Gamma) \in \Lambda_i'$, we find that $b\pi_i([e,f]_\Gamma) \in \Lambda_i'$,

We have proven that $\pi_i([e,f]_\Gamma)$ is a generator of Λ_i/Λ_i'. Therefore the index $[\Lambda_i \colon \Lambda_i']$ divides b and so p divides b. It follows that
$$\varepsilon \equiv [x,\gamma x]_\Gamma \mod p\Lambda.$$

Since ε was chosen to be primitive in Λ, we have $\varepsilon \notin p\Lambda$. So the mod p reduction $\overline{\varepsilon} \in \Lambda/p\Lambda$ is an $\mathcal{H}$-eigenvector, where the eigenvalue of T_ℓ is $\ell + 1$ for all $\ell \nmid N$. But $[x, \gamma x]_\Gamma$ is cuspidal. Therefore this package of $\mathcal{H}$-eigenvalues must be equal modulo a prime over p to the package of eigenvalues of a newform in Λ'_i.

In other words p is an Eisenstein prime, a prime in the support of the Eisenstein ideal. It follows that if p is odd, p divides $N - 1$. See [WWE21, Section 1.1.]. □

There are 1229 primes less than 10,000. With the exception of prime levels $N \in \{2, 3, 5, 7, 13\}$, each prime level in this range has a nontrivial cuspidal space. For these prime levels $N < 10{,}000$, we observe from our data that

- the product of the indices divides $N - 1$, i.e.,

$$\prod_{i=1}^{k} [\Lambda_i \colon \Lambda'_i] \mid (N - 1);$$

- the quotient

$$Q = \frac{(N-1)}{\prod_i [\Lambda_i \colon \Lambda'_i]}$$

  is a positive power of 2 times a nonnegative power of 3;
- for all N in the range mentioned above, the index $[\Lambda_i \colon \Lambda'_i]$ is 1 for all i except for exactly one or two i's. There are two exceptional i's exactly when

$$N = 71, 211, 307, 397, 487, 577, 673, 1871, 1999, 3001, 4621, 9931.$$

  For these primes, see Table 1.

We do not have an explanation for these observations.

Table 1. Exceptional prime levels $N < 10{,}000$ with two Hecke irreducible cuspidal spaces V_i with indices $[\Lambda_i \colon \Lambda'_i] > 1$.

N	$\dim(V_1)$	$\dim(V_2)$	$[\Lambda_1 \colon \Lambda'_1]$	$[\Lambda_2 \colon \Lambda'_2]$
71	3	3	7	5
211	2	9	5	7
307	2	9	3	17
397	5	10	11	3
487	2	16	3	27
577	2	18	3	8
673	4	24	7	4
1871	2	98	5	187
1999	2	94	3	111
3001	2	132	5	25
4621	2	196	5	77
9931	2	434	5	331

3. Prevalence

Next, we wish to learn what a "random" modular symbol looks like. From Theorem 2.4, we see that the image of the modular symbols in V is the union of three cosets in Λ/Λ', namely the cuspidal lattice Λ' and its translates by $\pm[e, f]_\Gamma$. If a "random" modular symbol is chosen, in which coset do we expect it land? [2] Can we say more about the type of a random modular symbol? For example, if we compute the images of modular symbols in some box, what are the relative proportions of the types that occur? In this section, we examine these questions.

Proposition 2.1 allows us to predict the likelihood that a given symbol $[x, y]_\Gamma$ is cuspidal, since cuspidal symbols must have x and y Γ-equivalent to each other.

Lemma 3.1. *Let N be prime, let $\Gamma = \Gamma_0^\pm(N)$, and let r be a positive integer. There are exactly*

$$N^{2r}\left(\frac{2}{1+N}\right) + 2N^r - \frac{2}{1+N}$$

pairs $(a, b) \in \mathbb{Z}^2 - \{(0, 0)\}$ with $-N^r \le a \le N^r$ and $0 \le b \le N^r$ such that $a/b \in \mathbb{P}^1(\mathbb{Q})$ is Γ-equivalent to $e = 1/0$.

Proof. Let $(a, b) \in \mathbb{Z}^2$ with $-N^r \le a \le N^r$ and $0 \le b \le N^r$. Let $\alpha = \frac{a}{b} \in \mathbb{P}^1(\mathbb{Q})$. From Proposition 2.1, α is Γ-equivalent to $e = 1/0$ if and only if N divides the denominator of α written in reduced form. Equivalently, $\nu_N(a) < \nu_N(b)$ and $a \neq 0$, where $\nu_N(x)$ is the largest e such that N^e divides x, except that $\nu_N(0) = \infty$.

Let S be the set

$$S = \{(a, b) \in \mathbb{Z}^2 \mid -N^r \le a \le N^r, 0 \le b \le N^r, a \neq 0, \nu_N(a) < \nu_N(b)\}.$$

There are $2N^r$ pairs of the form $(a, 0)$ in S and $2N^r - 2$ pairs of the form (a, N^r) in S. For $k = 1, 2, \ldots, r-1$, there are $N^{r-k} - N^{r-k-1}$ choices for b with $\nu_N(b) = k$ and $2N^r - 2N^{r-k}$ choices for a with $\nu_N(a) < k$ for $(a, b) \in S$. Thus

$$\#S = 4N^r - 2 + 2\sum_{k=1}^{r-1}(N^r - N^{r-k})(N^{r-k} - N^{r-k-1}).$$

We expand the summand and break up the sum into two pieces. The first piece is a telescoping sum and the second piece is geometric, namely the alternating sum of the powers of N from N to $-N^{2r-2}$, so

$$\begin{aligned}
\sum_{k=1}^{r-1}(N^r - N^{r-k})(N^{r-k} - N^{r-k-1}) &= \sum_{k=1}^{r-1}(N^{2r-k} - N^{2r-k-1}) + \sum_{k=1}^{r-1}(N^{2r-2k-1} - N^{2r-2k}) \\
&= (N^{2r-1} - N^r) + \left(\frac{N - N^{2r-1}}{1+N}\right) \\
&= -N^r + \frac{N^{2r-1} + N^{2r} + N - N^{2r-1}}{1+N} \\
&= -N^r + \frac{N^{2r} + N}{1+N}.
\end{aligned}$$

[2]Since $[x, y]_\Gamma = -[y, x]_\Gamma$, the two nontrivial cosets will occur an equal number of times, and the real question in this regard is simply whether the chosen modular symbol is cuspidal or not.

We compute

$$\begin{aligned}
\#S &= 4N^r - 2 + 2\left(-N^r + \frac{N^{2r}+N}{1+N}\right)\\
&= 2N^r - 2 + 2\left(\frac{N^{2r}}{1+N} + \frac{N}{1+N}\right)\\
&= 2N^r + N^{2r}\left(\frac{2}{1+N}\right) - 2 + \frac{2N}{1+N}\\
&= N^{2r}\left(\frac{2}{1+N}\right) + 2N^r - \frac{2}{1+N},
\end{aligned}$$

as desired. □

Theorem 3.2. *Let N be prime, let $\Gamma = \Gamma_0^{\pm}(N)$, and let r be a positive integer. Let M_r denote the set of matrices*

$$M_r = \left\{ \begin{bmatrix} a & p \\ b & q \end{bmatrix} \;\middle|\; \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} p \\ q \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}, -N^r \leq a, p \leq N^r, 0 \leq b, q \leq N^r \right\}.$$

Let $C_r \subset M_r$ denote the subset of matrices $\begin{bmatrix} a & p \\ b & q \end{bmatrix}$ that give rise to cuspidal modular symbols $[a/b, p/q]_\Gamma \in U'$. The the limiting proportion of cuspidal symbols is

$$\lim_{r\to\infty} \frac{\#C_r}{\#M_r} = \frac{1+N^2}{(1+N)^2}.$$

Proof. First we compute $\#M_r$. For $\begin{bmatrix} a & p \\ b & q \end{bmatrix} \in M_r$, there are

$$(2N^r+1)(N^r+1) - 1 = 2N^{2r} + 4N^r$$

choices for (a, b) and the same number for (p, q). Thus

$$\#M_r = (2N^{2r} + 4N^r)^2.$$

From Lemma 3.1, we have $N^{2r}\left(\frac{2}{1+N}\right) + 2N^r - \frac{2}{1+N}$ pairs (a, b) that give cusps that are Γ-equivalent to $e = 1/0$. The remaining $N^{2r}\left(\frac{2N}{1+N}\right) + 2N^r + \frac{2}{1+N}$ pairs are Γ-equivalent to $f = 0/1$.

It follows that

$$\#C_r = \left(N^{2r}\left(\frac{2}{1+N}\right) + 2N^r - \frac{2}{1+N}\right)^2 + \left(N^{2r}\left(\frac{2N}{1+N}\right) + 2N^r + \frac{2}{1+N}\right)^2.$$

Let $B = N^r$. Then as $r \to \infty$, we have $B \to \infty$ and only the highest order term affects the limit. Thus

$$\begin{aligned}
\lim_{r\to\infty} \frac{\#C_r}{\#M_r} &= \lim_{B\to\infty} \frac{B^4\left(\left(\frac{2}{1+N}\right)^2 + \left(\frac{2N}{1+N}\right)^2\right)}{4B^4}\\
&= \frac{1}{4}\left(\left(\frac{2}{1+N}\right)^2 + \left(\frac{2N}{1+N}\right)^2\right)\\
&= \frac{1+N^2}{(1+N)^2}.
\end{aligned}$$

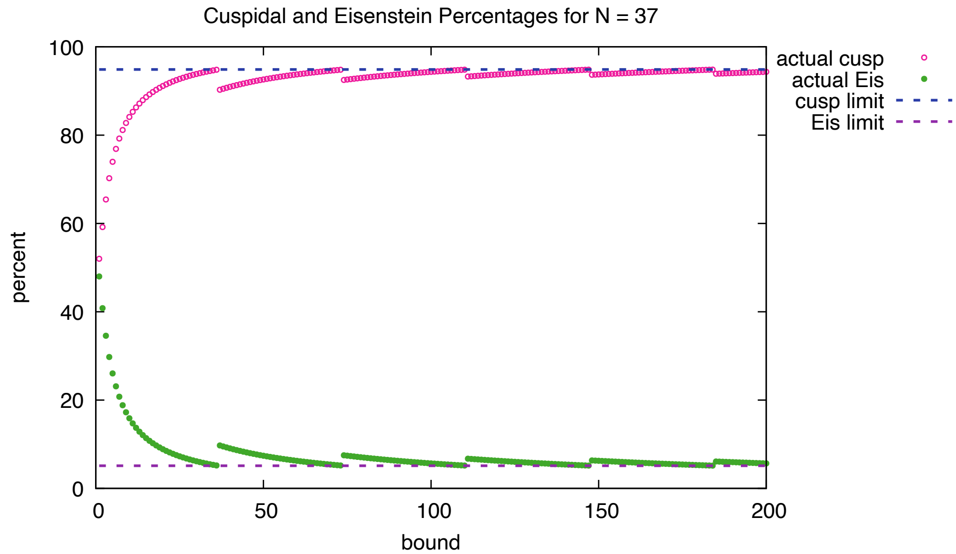

FIGURE 1. The percentage of cuspidal and noncuspidal (Eis) modular symbols observed for level 37 as a function of box size, verifying the limiting value of approximately 94.875% cuspidal symbols and 5.125% Eisenstein symbols as the box size grows. Note that $(1+37^2)/(1+37)^2 \approx 0.94875$.

$\square$

Corollary 3.3. *Let* $a, b, p, q \in \mathbb{Z}$ *with* $\begin{bmatrix} a \\ b \end{bmatrix}$ *and* $\begin{bmatrix} p \\ q \end{bmatrix}$ *not equal to* $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. *Let* $[a/b, p/q]_\Gamma \in V$ *be chosen randomly from a rectangular box as in the preceding theorem and suppose it has type*

$$t([a/b, p/q]_\Gamma) = (t_1, t_2, \ldots, t_k, t_E).$$

Then as the box grows to infinity, the probability that $t_E = 0$ *is* $\frac{1+N^2}{(1+N)^2}$, *and the probability that* $t_E = 1$ *is* $\frac{2N}{(1+N)^2}$.

Remark 3.4. This result shows that a random modular symbol is likely cuspidal, with the likelihood getting higher as the level increases. Note, however, that neither the result nor the proof, gives any indication of the proportion of zero symbols, i.e. those of type $(0, 0, \ldots, 0)$. Nevertheless, the experimental results in Section 4 suggest that most random modular symbols are nontrivial and cuspidal.

We verify Corollary 3.3 in our computations. The predicted probabilities in Corollary 3.3 are indeed observed. The convergence to the limiting proportions is fairly fast, as may be seen for example in Figure 1.

In the next section we provide a view of the data we collected.

4. Experimental results on prevalence

In this section, we examine in detail the types and prevalence of each type that occur for the prime levels $N < 100$. All calculations are done in Magma V2.25-8 [BCP97].

We collect data from a box of size B for each level N as follows. Let

$$M(B) = \left\{ \begin{bmatrix} a & p \\ b & q \end{bmatrix} \;\middle|\; \begin{bmatrix} a \\ b \end{bmatrix}, \begin{bmatrix} p \\ q \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}, -B \leq a, p \leq B, 0 \leq b, q \leq B \right\}.$$

For each type τ, let $T_N(B,\tau) \subset M(B)$ denote the subset of matrices $\begin{bmatrix} a & p \\ b & q \end{bmatrix}$ that give rise to modular symbols $[a/b, p/q]_\Gamma$ of type τ. Let $p_N(B,\tau)$ denote the proportion of $T_N(B,\tau)$ in $M(B)$,

$$p_N(B,\tau) = \frac{\#T_N(B,\tau)}{\#M(B)}.$$

Here are some observations on the prevalence of various types in our data.

- It appears that, given N, the probability that a certain type occurs does stabilize asymptotically as the box grows, i.e., for each type τ, it appears that as $B \to \infty$, the proportion $p_N(B,\tau)$ converges.
- Let τ_0 denote the trivial type $(0,0,\ldots,0)$. Let

  $$z_N = \lim_{B\to\infty} p_N(B,\tau_0).$$

  Then from our data it appears that $z_N > 0$, and that $z_N \to 0$ as $N \to \infty$.
- The previous observation together with Corollary 3.3 suggests most random modular symbols are nontrivial and cuspidal.
- For most N there seems to be a dominant type. However, for $N = 89$, it seems like two types may have the same asymptotic probability. Could there be some kind of symmetry that would account for this (if it is indeed the case)? It is possible that a similar behavior would occur for $N = 79$ if we took B large enough.
- The case of $N = 61$ is interesting because the graphs of two types cross around $B = 120$, a behavior that doesn't appear in the other graphs.
- When $N = 89$ there occurs a phenomenon unique among our data. Namely, the prevalence of one of the types, although nonzero, is extremely small.

It would be interesting to have even a heuristic explanation of the phenomena we have discovered.

The graphs in Figures 2, 3, and 4 are constructed as follows. The different kinds of dots correspond to the different types, as shown. The horizontal axis is the B-axis. The vertical axis measures the proportion of modular symbols in $M(B)$ which have the given type. Thus the dots measure the cumulative proportions as the boxes grow. There are discontinuities in the graphs when B is a multiple of N. This is because when B is a multiple of N, there are suddenly many more fractions that have denominators divisible by N and therefore cusps equivalent to e. This causes a momentary increase in the number the modular symbols in the $\pm[e,f]_\Gamma$ coset, i.e., non-cuspidal. We compute dots for all $B \leq 200$ in increments of 1. We analyzed $\#M(200) = 6{,}496{,}360{,}000$ symbols for each prime level $N < 100$. These plots and more are available at `https://mathstats.uncg.edu/yasaki/random-modular-symbols/`.

We also provide a table listing N, then for each N, the types which survive the Eisenstein obstruction, and for each type τ, the value of $p_N(B,\tau)$ for $B = 200$. See Tables 3, 6, and 8.

4.1. **One Hecke irreducible subspace.** For $N \in \{2,3,5,7,13\}$, we have that the cuspidal space is trivial so $V = E$, the 1-dimensional Eisenstein space. Thus the only cuspidal type is (0), and the only Eisenstein type is (1). Thus the prevalence of Eisenstein and cuspidal symbols given in Corollary 3.3 coincides exactly with the values of the prevalence of the two types.

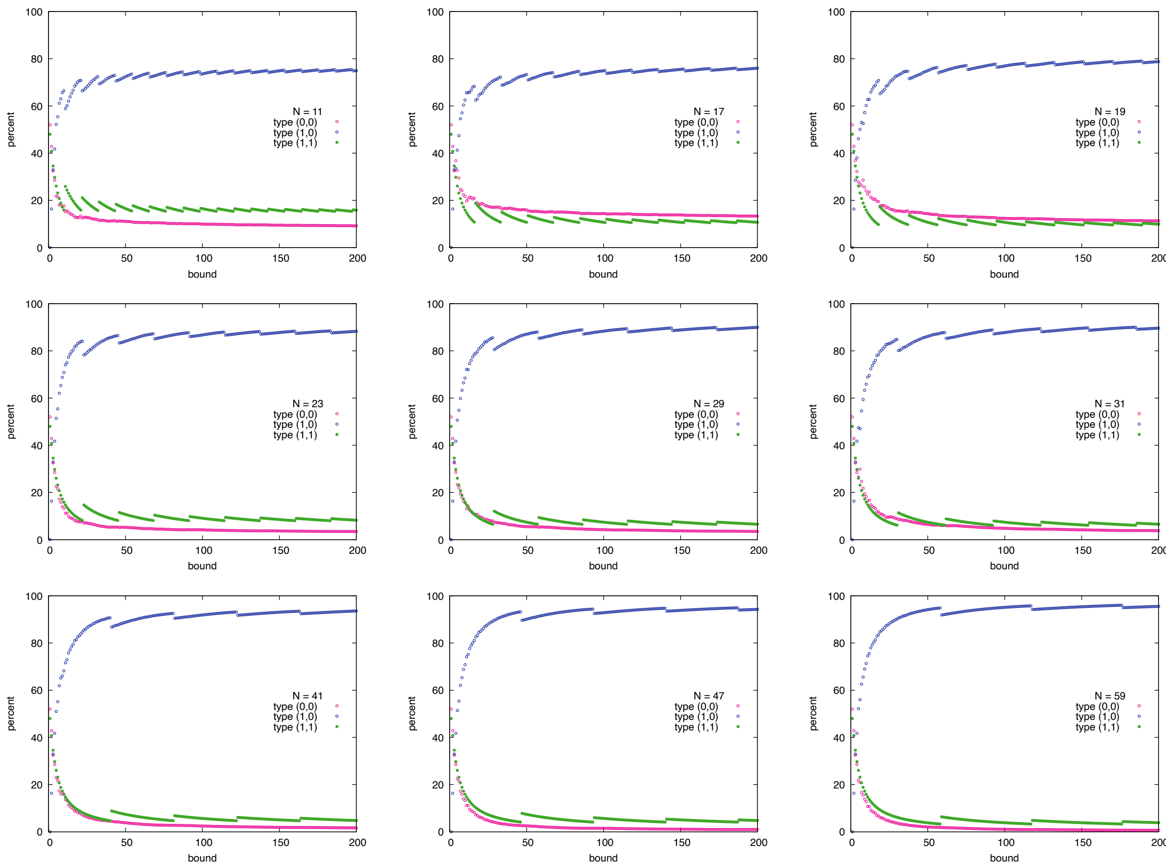

FIGURE 2. Prevalence plots for levels $N \in \{11, 17, 19, 23, 29, 31, 41, 47, 59\}$.

4.2. **Two Hecke irreducible subspaces.** For $N \in \{11, 17, 19, 23, 29, 31, 41, 47, 59\}$, we have that V decomposes as a direct sum of Hecke irreducible subspaces

$$V = V_1 \oplus E.$$

The dimension $\dim(V_1)$, the index $[\Lambda_1 \colon \Lambda_1']$, and the quotient $Q = (N-1)/[\Lambda_1 \colon \Lambda_1']$ are given in Table 2. The type of $[e, f]_\Gamma$ is $(1, 1)$, while the fact that $[\Lambda_1 \colon \Lambda_1'] \neq 1$ combined with Proposition 2.6 shows that type $(0, 1)$ does not arise for these levels. Thus, the types that arise for $N \in \{11, 17, 19, 23, 29, 31, 41, 47, 59\}$ are exactly $\{(0, 0), (1, 0), (1, 1)\}$.

For N in $\{11, 17, 19, 23, 29, 31, 41, 47, 59\}$, the space V_1 is the span of the newforms 11.2.a.a, 17.2.a.a, 19.2.a.a, 23.2.a.a, 29.2.a.a, 31.2.a.a, 41.2.a.a, 47.2.a.a, and 59.2.a.a, respectively, in the notation of the L-functions and modular forms database (LMFDB) [LMF23].

4.3. **Three Hecke irreducible subspaces.** For $N \in \{37, 43, 53, 61, 71, 79, 83, 97\}$, we have that V decomposes as a direct sum of Hecke irreducible subspaces

$$V = V_1 \oplus V_2 \oplus E.$$

The dimensions $\dim(V_i)$ and the indices $[\Lambda_i \colon \Lambda_i']$ for $i = 1, 2$ are given in Table 4. We also give the quotient $Q = (N-1)/\Pi_{i=1}^{2}[\Lambda_i \colon \Lambda_i']$.

For $N \in \{37, 43, 53, 61, 79, 83, 97\}$, we have $t([e, f]_\Gamma) = (0, 1, 1)$ and $[\Lambda_2 \colon \Lambda_2'] \neq 1$. Thus any type of the form $(*, 0, 1)$ does not arise. In particular, for $N \in \{37, 43, 53, 61, 79, 83, 97\}$, the

TABLE 2. Lattice indices and the quotient $Q = (N-1)/[\Lambda_1 \colon \Lambda_1']$ of modular symbols for level $N \in \{11, 17, 19, 23, 29, 31, 41, 47, 59\}$.

N	$\dim(V_1)$	$[\Lambda_1 \colon \Lambda_1']$	Q
11	1	5	2
17	1	2	2^3
19	1	3	$2 \cdot 3$
23	2	11	2
29	2	7	2^2
31	2	5	$2 \cdot 3$
41	3	5	2^3
47	4	23	2
59	5	29	2

TABLE 3. Observed asymptotic prevalence for each type for level $N \in \{11, 17, 19, 23, 29, 31, 41, 47, 59\}$.

type	11	17	19	23	29	31	41	47	59
$(0,0)$	9.203	13.352	11.315	3.456	3.482	3.809	1.659	0.972	0.628
$(1,0)$	74.964	75.977	78.806	88.299	89.978	89.651	93.577	94.264	95.523
$(1,1)$	15.832	10.671	9.879	8.244	6.540	6.540	4.764	4.764	3.849

TABLE 4. Lattice indices, and the quotient $Q = (N-1)/\Pi_{i=1}^{2}[\Lambda_i \colon \Lambda_i']$ of modular symbols for level $N \in \{37, 43, 53, 61, 71, 79, 83, 97\}$.

N	$\dim(V_1)$	$\dim(V_2)$	$[\Lambda_1 \colon \Lambda_1']$	$[\Lambda_2 \colon \Lambda_2']$	Q
37	1	1	1	3	$2^2 \cdot 3$
43	1	2	1	7	$2 \cdot 3$
53	1	3	1	13	2^2
61	1	3	1	5	$2^2 \cdot 3$
71	3	3	7	5	2
79	1	5	1	13	$2 \cdot 3$
83	1	6	1	41	2
97	3	4	1	4	$2^3 \cdot 3$

types that arise are exactly

$$\{(0,0,0), (0,1,0), (0,1,1), (1,0,0), (1,1,0), (1,1,1)\}.$$

For $N = 71$, we have $t([e,f]_\Gamma) = (1,1,1)$ and $[\Lambda_i \colon \Lambda_i'] \neq 1$ for $i = 1, 2$. Thus any type of the form $(*,0,1)$ or $(0,*,1)$ does not arise. In particular, for $N = 71$, the types that arise are exactly

$$\{(0,0,0), (0,1,0), (1,0,0), (1,1,0), (1,1,1)\}.$$

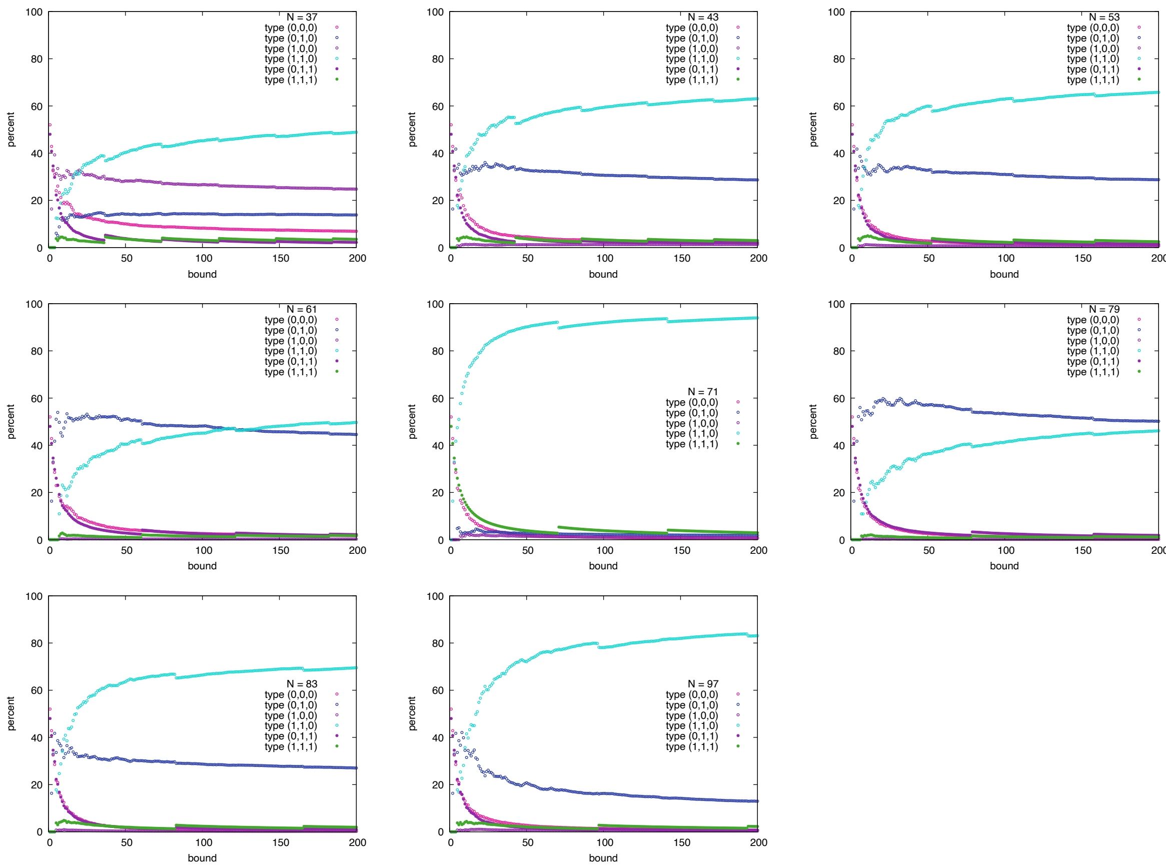

FIGURE 3. Prevalence plots for levels $N \in \{37, 43, 53, 61, 71, 79, 83, 97\}$.

TABLE 5. Cuspforms for level $N \in \{37, 43, 53, 61, 71, 79, 83, 97\}$.

N	V_1	V_2
37	37.2.a.a	37.2.a.b
43	43.2.a.a	43.2.a.b
53	53.2.a.a	53.2.a.b
61	61.2.a.a	61.2.a.b
71	71.2.a.a	71.2.a.b
79	79.2.a.a	79.2.a.b
83	83.2.a.a	83.2.a.b
97	97.2.a.a	97.2.a.b

4.4. **Four Hecke irreducible subspaces.** For $N \in \{67, 73, 89\}$, we have that V decomposes as a direct sum of Hecke irreducible subspaces

$$V = V_1 \oplus V_2 \oplus V_3 \oplus E.$$

The dimensions $\dim(V_i)$ and the indices $[\Lambda_i : \Lambda_i']$ for $i = 1, 2, 3$ are given in Table 9. We also give the quotient $Q = (N-1)/\Pi_{i=1}^{3}[\Lambda_i : \Lambda_i']$.

For $N \in \{67, 73, 89\}$, we have $t([e, f]_\Gamma) = (1, 0, 1, 1)$ and $[\Lambda_3 : \Lambda_3'] \neq 1$. Thus any type of the form $(*, *, 0, 1)$ does not arise. In particular, for $N \in \{67, 73, 89\}$, the types that arise

TABLE 6. Observed asymptotic prevalence for each type for level $N \in \{37, 43, 53, 61, 71, 79, 83, 97\}$.

type	37	43	53	61	71	79	83	97
$(0,0,0)$	6.902	2.254	1.025	1.682	0.511	0.691	0.373	0.729
$(0,1,0)$	24.774	28.668	28.753	44.563	1.811	50.177	27.027	12.971
$(0,1,1)$	2.191	1.805	1.389	2.151		1.801	0.982	0.671
$(1,0,0)$	13.810	1.254	0.601	0.234	0.792	0.103	0.172	0.319
$(1,1,0)$	48.853	63.060	65.773	49.673	93.972	46.115	69.514	83.067
$(1,1,1)$	3.470	2.959	2.459	1.697	2.914	1.113	1.932	2.243

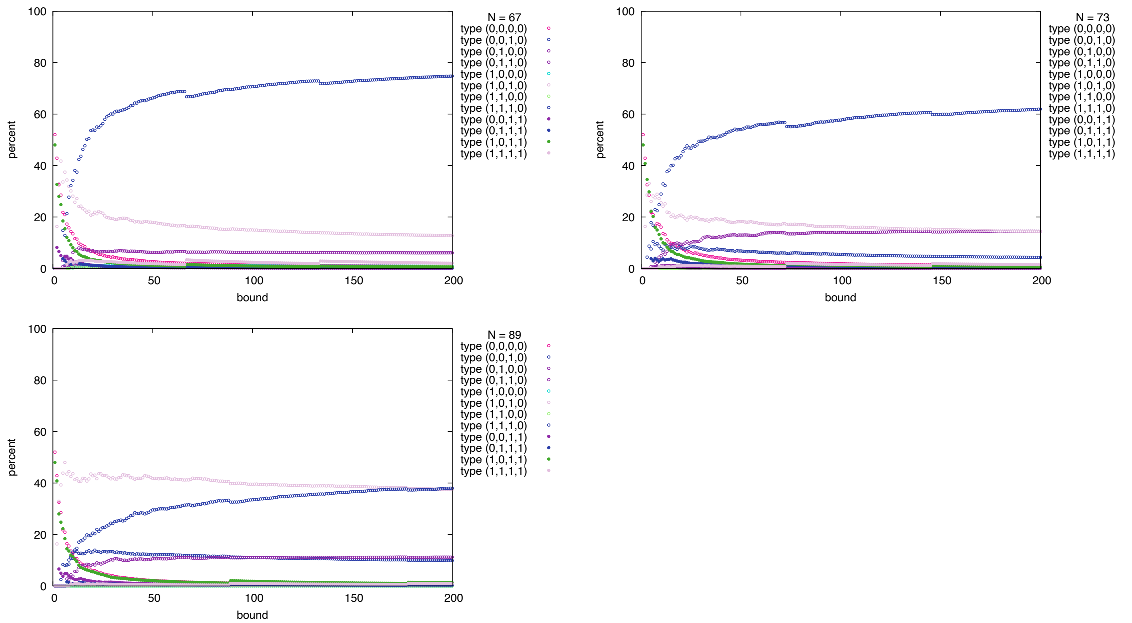

FIGURE 4. Prevalence plots for levels $N \in \{67, 73, 89\}$.

TABLE 7. Lattice indices, and the quotient $Q = (N-1)/\Pi_{i=1}^{3}[\Lambda_i \colon \Lambda_i']$ of modular symbols for level $N \in \{67, 73, 89\}$.

N	$\dim(V_1)$	$\dim(V_2)$	$\dim(V_3)$	$[\Lambda_1 \colon \Lambda_1']$	$[\Lambda_2 \colon \Lambda_2']$	$[\Lambda_3 \colon \Lambda_3']$	Q
67	1	2	2	1	1	11	$2 \cdot 3$
73	1	2	2	1	1	3	$2^3 \cdot 3$
89	1	1	5	1	1	11	2^3

are exactly

$$\{(0,0,0,0), (0,0,1,0), (0,1,0,0), (0,1,1,0), (0,1,1,1), (1,1,1,1), \\ (1,0,0,0), (1,0,1,0), (1,1,0,0), (1,0,1,1), (1,1,1,0), (1,1,1,1)\}.$$

Table 8. Observed asymptotic prevalence for each type for level $N \in \{67, 73, 89\}$.

type	67	73	89
$(0,0,0,0)$	0.808	1.102	0.506
$(0,0,1,0)$	0.564	4.277	9.862
$(0,0,1,1)$	0.078	0.102	0.416
$(0,1,0,0)$	0.530	0.521	0.056
$(0,1,1,0)$	6.079	14.456	11.178
$(0,1,1,1)$	0.225	0.618	0.338
$(1,0,0,0)$	0.861	0.272	0.003
$(1,0,1,0)$	12.761	14.423	37.509
$(1,0,1,1)$	0.585	0.748	1.273
$(1,1,0,0)$	0.733	0.119	9.41×10^{-5}
$(1,1,1,0)$	74.750	61.916	37.973
$(1,1,1,1)$	2.025	1.447	0.887

Table 9. Cuspforms for level $N \in \{67, 73, 89\}$.

N	V_1	V_2	V_3
67	67.2.a.a	67.2.a.b	67.2.a.c
73	73.2.a.a	73.2.a.b	73.2.a.c
89	89.2.a.b	89.2.a.a	89.2.a.c

References

[AY21] Avner Ash and Dan Yasaki, *Steinberg homology, modular forms, and real quadratic fields*, J. Number Theory **224** (2021), 323–367. MR 4244157

[BCP97] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265, Computational algebra and number theory (London, 1993). MR MR1484478

[LMF23] The LMFDB Collaboration, *The L-functions and modular forms database*, `http://www.lmfdb.org`, 2023, [Online; accessed 1 November 2023].

[MS76] Barry Mazur and Jean-Pierre Serre, *Points rationnels des courbes modulaires $X_0(N)$ (d'après A. Ogg)*, Séminaire Bourbaki (1974/1975), Exp. No. 469, Lecture Notes in Math, vol. Vol. 514, Springer, Berlin, 1976, pp. pp 238–255. MR 485882

[WWE21] Preston Wake and Carl Wang-Erickson, *The Eisenstein ideal with squarefree level*, Adv. Math. **380** (2021), Paper No. 107543, 62. MR 4200464

Boston College, Chestnut Hill, MA 02467
Email address: `Avner.Ash@bc.edu`

UNCG, Greensboro, NC 27412
Email address: `d_yasaki@uncg.edu`