J. Eur. Math. Soc. 19, 3549–3573
DOI 10.4171/JEMS/746

Abhishek Saha

On sup-norms of cusp forms of powerful level

Received January 7, 2015

Abstract. Let f be an L^2-normalized Hecke–Maass cuspidal newform of level N and Laplace eigenvalue λ. It is shown that $\|f\|_\infty \ll_{\lambda,\epsilon} N^{-1/12+\epsilon}$ for any $\epsilon > 0$. The exponent is further improved when N is not divisible by "small squares". Our work extends and generalizes previously known results in the special case of N squarefree.

Keywords. Maass form, sup-norm, Fourier coefficients, amplification

1. Introduction

The problem of bounding the sup-norms of L^2-normalized cuspidal automorphic forms has been much studied recently, beginning with the work of Iwaniec and Sarnak [IS95], who proved the first non-trivial bound in the eigenvalue aspect for Hecke–Maass cusp forms. Since then, this question has been considered in the eigenvalue/weight [Koy95, Van97, Don01, Rud05, Xia07, DS15, BT17, BP16, HRR14, BM15], volume/level [AU95, JK04, Lau10, Tem10, HT12, HT13, Tem14, Kir14] and hybrid [BH10, BM13, Tem15, BHM16] aspects for various types of automorphic forms. One reason why this problem is interesting is its connections with various other topics, such as the theory of quantum chaos, the subconvexity of L-functions, the combinatorics of Hecke algebras, and diophantine analysis.

Our interest in this paper is in the level aspect. We consider the sup-norm question for eigenfunctions on the arithmetic hyperbolic surface $\Gamma_0(N)\backslash\mathbb{H}$ equipped with the measure $\frac{dx\,dy}{y^2}$. It is natural to restrict to the case of newforms. Thus, we are interested in bounding the sup-norms of L^2-normalized Hecke–Maass newforms f of level N (and trivial character) in the N-aspect. The following upper bounds for $\|f\|_\infty$ in the N-aspect were known prior to this work:

- The "trivial bound" $\|f\|_\infty \ll_{\lambda,\epsilon} N^{\epsilon}$.
- $\|f\|_\infty \ll_{\lambda,\epsilon} N^{-25/914+\epsilon}$ for squarefree N, due to Blomer and Holowinsky [BH10], published in 2010.

A. Saha: Department of Mathematics, University of Bristol,
Bristol BS81TW, UK; e-mail: abhishek.saha@bris.ac.uk

Mathematics Subject Classification (2010): Primary 11F03

- $\|f\|_\infty \ll_{\lambda,\epsilon} N^{-1/22+\epsilon}$ for squarefree N, due to Templier [Tem10], published in 2010.
- $\|f\|_\infty \ll_{\lambda,\epsilon} N^{-1/20+\epsilon}$ for squarefree N, due to Helfgott–Ricotta (unpublished).
- $\|f\|_\infty \ll_{\lambda,\epsilon} N^{-1/12+\epsilon}$ for squarefree N, due to Harcos and Templier [HT12], published in 2012.
- $\|f\|_\infty \ll_{\lambda,\epsilon} N^{-1/6+\epsilon}$ for squarefree N due to Harcos and Templier [HT13], published in 2013.[1]

As the above makes clear, there has been fairly rapid progress in the squarefree case, yet no improvement has been obtained beyond the trivial bound when N is not squarefree. Indeed, all the above papers rely crucially on using Atkin–Lehner operators to move any point of $\mathbb{H}$ to a point of imaginary part $\geq 1/N$ (which is essentially equivalent to using a suitable Atkin–Lehner operator to move any cusp to infinity). This only works if N is squarefree.

In this paper, we introduce some new ideas and technical improvements which allows us to obtain a non-trivial result without any squarefree assumptions.[2]

Theorem 1. *Let f be an L^2-normalized Hecke–Maass cuspidal newform for the group $\Gamma_0(N)$ with Laplace eigenvalue $\lambda \leq T$.*

(1) *For any $\epsilon > 0$ we have the bound*

$$\|f\|_\infty \ll_{T,\epsilon} N^{-1/12+\epsilon}.$$

(2) *Suppose that there is no integer M in the range $1 < M < N^{1/6}$ such that M^2 divides N. Then we can improve the above bound to*

$$\|f\|_\infty \ll_{T,\epsilon} N^{\epsilon} \max(N^{-1/6}, N^{-1/4}N_0^{1/4})$$

where N_0 is the largest integer such that N_0^2 divides N. In particular, in this case we always have

$$\|f\|_\infty \ll_{T,\epsilon} N^{-1/8+\epsilon}.$$

Remark 1.1. Assertion (2) of the theorem can be regarded as dealing with the case when N is not divisible by "small squares". This includes, for instance, the squarefree case (in which case we recover the bound $\|f\|_\infty \ll_{T,\epsilon} N^{-1/6+\epsilon}$ due to Harcos–Templier), the case $N = p^2 N_2$ where N_2 is squarefree and p is a prime such that $p \geq N_2^{1/4}$, and the case $N = p^n$ where p is a prime and $1 \leq n \leq 6$.

Remark 1.2. All the results of this paper (and in particular the main result above) remain valid in the case of holomorphic newforms of fixed weight and varying level N.

[1] Templier [Tem15] has successfully combined this bound with the bound of Iwaniec–Sarnak in the eigenvalue aspect, to obtain a state-of-the-art hybrid estimate.

[2] A look at the wider sup-norm literature suggests that this is the first time that the squarefree barrier has been non-trivially broken for any kind of automorphic form on a domain that contains cusps.

Remark 1.3. In this paper we have restricted for simplicity to the case of trivial central character. We have also made no effort to obtain a hybrid bound, i.e., to quantify the dependence of our constants on the Laplace eigenvalue. However, we expect that the methods of this paper, with some modifications, will be able to deal with these cases. Further, we hope that this paper will shed some light on how to remove the squarefree restriction from sup-norm bounds for more general automorphic forms. We will come back to some of these questions in future work.

Remark 1.4 (Added in proof). Recently, we have succeeded in significantly improving the results of this paper, as well as obtained a hybrid bound. This is done in [Sah17], which uses a fairly different (and in our view, superior) adelic methodology.

Let us briefly explain the new ingredients in this paper compared to the paper by Harcos and Templier [HT13] (whose general strategy we broadly follow). Our key new idea is to look at the behavior of cusp forms around *cusps of width* 1. Recall that if N is squarefree, then the surface $\Gamma_0(N)\backslash\mathbb{H}$ has exactly one cusp of width 1, namely the cusp at infinity. However, if N is not squarefree, then there is always more than one cusp of width 1. Cusps of width 1 have several nice properties. First, any cusp can be conjugated to a cusp of width 1 by use of a suitable Atkin–Lehner operator. Secondly, this leads to a "gap principle", whereby any point of $\mathbb{H}$ can be moved by an Atkin–Lehner operator to another point which has high imaginary part and good diophantine properties when rewritten in the coordinates corresponding to a suitable cusp of width 1. Thirdly, if $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ is a matrix that takes the cusp at infinity to a cusp of width 1, then for any Hecke–Maass cuspidal newform f for $\Gamma_0(N)$, the function $f|\sigma$ is a Maass cusp form on the slightly smaller group $\Gamma_0(N; M) := \Gamma_0(N) \cap \Gamma_1(M)$ (where M^2 is a suitable divisor of N) and moreover $f|\sigma$ is an eigenfunction of the Hecke operators at all primes congruent to 1 mod M.

We exploit the above facts to reduce the sup-norm question from f to some suitable $f|\sigma$. However, several technical difficulties arise. First, the counting problem that lies at the heart of the amplification method becomes much more involved, especially for the *parabolic* matrices. Secondly, the bound via applying the Cauchy–Schwarz inequality on the Fourier expansion requires undertaking a deep study of the Fourier coefficients at the cusp σ. Thirdly, because the surface $\Gamma_0(N; M)\backslash\mathbb{H}$ has higher volume than $\Gamma_0(N)\backslash\mathbb{H}$ and because we can now amplify *only* over primes that are 1 mod M, we lose some sharpness in our bounds, and it is important to offset this in some way[3] so that this loss is not too prominent. These technical difficulties are, however, all successfully overcome, and in the end we get the theorem quoted above.

We end this introduction with a few speculative remarks regarding the true order of magnitude for $\|f\|_\infty$. The trivial *lower bound* for $\|f\|_\infty$ in the N-aspect is $\|f\|_\infty \gg_{T,\epsilon} N^{-1/2-\epsilon}$, and this bound is also valid for L^2-normalized Hecke–Maass newforms with non-trivial character. However, if the conductor of the character is large relative to N,

[3] This is achieved by a twofold process. First, our gap principle contains a factor of M^2, which makes the bounds obtained via the Fourier expansion extremely strong when M is relatively large. Secondly, our counting arguments are refined to mostly account for the presence of M.

local effects (coming from the behavior of local Whittaker newforms for ramified principal series representations) lead to stronger lower bounds. For example, if f is an L^2-normalized Hecke–Maass newform of level N with N a perfect square, and the conductor of the character attached to f is also equal to N, then Templier [Tem14] showed that $\|f\|_\infty \gg_{T,\epsilon} N^{-1/4-\epsilon}$. In recent work by the author [Sah17, Sah16], the results of this paper, as well as Templier's example, are generalized to a wide variety of cases with nontrivial character. Moreover, we will precisely measure the local effects coming from the ramified Whittaker newforms, and thus are able to make a conjecture about the true size of $\|f\|_\infty$. In the case of trivial central character as in this paper, or more generally if the exponent of each prime dividing the conductor of the character is at most half the exponent of the prime dividing the square-full part of N, we will (optimistically) conjecture that $N^{-1/2-\epsilon} \ll_{T,\epsilon} \|f\|_\infty \ll_{T,\epsilon} N^{-1/2+\epsilon}$.

Some basic notations and definitions

- The symbols $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, S^1, $\mathbb{Z}_p$ and $\mathbb{Q}_p$ have the usual meanings. $\mathbb{A}$ denotes the ring of adeles of $\mathbb{Q}$.
- For any two complex numbers α, z, we let $K_\alpha(z)$ denote the modified Bessel function of the second kind. We write $e(z) := e^{2\pi i z}$. For each positive integer n, we let $\phi(n)$ denote the Euler phi function $\phi(n) = \#(\mathbb{Z}/n)^\times = \#\{a \in \mathbb{Z} : 1 \leq a \leq n,\ (a,n) = 1\}$.
- Given two integers a and b, we use $a \mid b$ to denote that a divides b, and we write $a \mid b^\infty$ when $a \mid b^n$ for some positive integer n. We use (a,b) for the greatest common divisor of a and b, which by our convention is always positive. We let (a, b^∞) denote the limit $\lim_{n\to\infty}(a, b^n)$, which always exists. We write $a^n \mid\mid b$ to mean that $a^n \mid b$ and a^{n+1} does not divide b. For any real number α, we let $\lfloor \alpha \rfloor$ denote the greatest integer less than or equal to α, and $\lceil \alpha \rceil$ is the smallest integer greater than or equal to α.
- For any commutative ring R and positive integer n, $M_n(R)$ denotes the ring of n by n matrices with entries in R, and $\mathrm{GL}_n(R)$ is the group of invertible matrices in $M_n(R)$. We use $R^\times$ to denote $\mathrm{GL}_1(R)$.
- The groups SL_2, PSL_2 and $\Gamma_0(N)$ have their usual meanings. We let $\mathrm{GL}_2^+(\mathbb{R})$ denote the subgroup of $\mathrm{GL}_2(\mathbb{R})$ consisting of matrices with positive determinant.
- We let $\mathbb{H} = \{x + iy : x, y \in \mathbb{R},\ y > 0\}$ denote the upper half-plane. For any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $\mathrm{GL}_2^+(\mathbb{R})$, and any $z \in \mathbb{H}$, we define $\gamma(z)$ or γz to equal $\frac{az+b}{cz+d}$. This action of $\mathrm{GL}_2^+(\mathbb{R})$ on $\mathbb{H}$ extends naturally to the boundary of $\mathbb{H}$. For any $g \in \mathrm{GL}_2^+(\mathbb{R})$ and any function f on $\mathbb{H}$, we let $f|\gamma$ denote the function on $\mathbb{H}$ defined by $f|\gamma(z) = f(\gamma z)$.
- For any congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$, and any bounded function $f : \mathbb{H} \to \mathbb{C}$ satisfying $f(\gamma z) = f(z)$ for all $\gamma \in \Gamma$, we define $\langle f, f\rangle_\Gamma = \int_{\Gamma\backslash\mathbb{H}} |f(z)|^2 \frac{dx\,dy}{y^2}$, and $\|f\|_\infty = \sup_{z\in\mathbb{H}} f(z)$. We say that such an f is a *Maass cusp form* for/on Γ if f is an eigenfunction of the hyperbolic Laplacian $\Delta := y^{-2}(\partial_x^2 + \partial_y^2)$ on $\mathbb{H}$ and decays rapidly at the cusps of Γ. The *Laplace eigenvalue* of such a Maass cusp form f is the real number λ satisfying $(\Delta + \lambda)f = 0$. We can write $\lambda = 1/4 + r^2$ where $r \in \mathbb{R} \cup i[0, 1/2]$; this follows from the non-negativity of Δ. We say that f is *L^2-normalized* if $\langle f, f\rangle_\Gamma = 1$.

- We say that f is a *cuspidal Hecke–Maass newform for* $\Gamma_0(N)$ (also referred to as a cuspidal Hecke–Maass newform of level N and trivial character) if it is a Maass cusp form for $\Gamma_0(N)$ and is a newform in the sense of Atkin–Lehner (i.e., it is orthogonal to all oldforms, and is an eigenfunction of all the Hecke and Atkin–Lehner operators). A cuspidal Hecke–Maass newform f is always either even or odd, i.e., there exists $\epsilon_f \in \{\pm 1\}$ such that $f(-\overline{z}) = \epsilon_f f(z)$.
- We use the notation $A \ll_{x,y,z} B$ to signify that there exists a positive constant C, depending at most upon x, y, z, such that $|A| \leq C|B|$.
- The symbol ϵ will denote a small positive quantity, whose value may change from line to line, and the value of the constant implicit in $\ll_{\epsilon,\ldots}$ may also change from line to line. An assertion such as "Let $1 \leq L \leq N^{O(1)}$. Then $f(\epsilon, L, N, \ldots) \ll_{\epsilon,\ldots} N^{O(\epsilon)} g(L, N, \ldots)$" means "For every $C > 0$ and $1 \leq L \leq N^C$, there is a constant $D > 0$ depending only on C such that $f(\epsilon, L, N, \ldots) \ll_{C,\epsilon,\ldots} N^{D\epsilon} g(L, N, \ldots)$."

2. Cusps of width 1 and Atkin–Lehner operators

Let $N = \prod_p p^{n_p}$ be a positive integer. Let $\mathbb{P}^1(\mathbb{Q})$ denote the set of all boundary points of the upper half-plane $\mathbb{H}$ that are stabilized by a non-trivial element of $\mathrm{PSL}_2(\mathbb{Z})$; precisely, $\mathbb{P}^1(\mathbb{Q})$ is the union of ∞ and the rational points on the real line. The set $\mathcal{C}(\Gamma_0(N)) = \Gamma_0(N)\backslash\mathbb{P}^1(\mathbb{Q})$ is the set of *cusps* of $\Gamma_0(N)$. For any ring R, let $N(R) = \left\{\begin{pmatrix} 1 & n \\ & 1 \end{pmatrix} : n \in R\right\}$ and $Z(R) = \left\{\begin{pmatrix} z & \\ & z \end{pmatrix} : z \in R^\times\right\}$. Via the correspondence $z \leftrightarrow \gamma(\infty)$, the set $\mathcal{C}(\Gamma_0(N))$ can be identified with the double coset space $\Gamma_0(N)\backslash \mathrm{SL}_2(\mathbb{Z})/N(\mathbb{Z})$. Given any $\tau \in \mathrm{SL}_2(\mathbb{Z})$, we can therefore speak of (some property of) the cusp (corresponding to) τ.

Let $\tau \in \mathrm{SL}_2(\mathbb{Z})$. The cusp $\tau(\infty)$ contains a representative of the form a/c, where $a, c \in \mathbb{Z}$, $c \mid N$, $c > 0$, $(a, c) = 1$. The integer c is uniquely determined. We will denote $C(\tau) = c$ and refer to $C(\tau)$ as the *denominator* of the cusp corresponding to τ. It can be easily checked that if $\tau = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $C(\tau) = (c, N)$. If $\Gamma_0(N)\tau_1 N(\mathbb{Z}) = \Gamma_0(N)\tau_2 N(\mathbb{Z})$, then $C(\tau_1) = C(\tau_2)$.

We let $W(\tau)$ denote the width of the cusp corresponding to τ; precisely, $W(\tau) = [N(\mathbb{Z}) : N(\mathbb{Z}) \cap \tau^{-1}\Gamma_0(N)\tau]$. Since the group $N(\mathbb{Z}) \cap \tau^{-1}\Gamma_0(N)\tau$ always contains $\left\{\begin{pmatrix} 1 & Nn \\ & 1 \end{pmatrix} : n \in \mathbb{Z}\right\}$ which has index N in $N(\mathbb{Z})$, it follows that $W(\tau)$ divides N.

For the convenience of the reader, we note down a few standard facts, proofs of which can be found for example in [NPS14, Sec. 3.4.1].

- For each $c \mid N$, the number of cusps with denominator c equals $\phi((c, N/c))$. Thus, the total number of cusps equals $\sum_{c|N} \phi((c, N/c))$. Moreover, there exists only one cusp of denominator N, namely the cusp ∞ $(= 1/N)$.
- For each $\tau \in \mathrm{SL}_2(\mathbb{Z})$ we have $W(\tau) = N/(C(\tau)^2, N)$. In particular, $W(\tau) = 1$ if and only if $N \mid C(\tau)^2$.
- If N is squarefree, then there is exactly one cusp of width 1, namely ∞. However, if N is not squarefree, then there is always more than one such cusp.

Remark 2.1. From the above facts, it is clear that an element $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ satisfies $W(\sigma) = 1$ if and only if $C(\sigma) = N/M$ for some positive integer M such that $M^2 \mid N$.

For each prime p, let

$$K_0(p^{n_p}) = \mathrm{GL}_2(\mathbb{Z}_p) \cap \begin{pmatrix} \mathbb{Z}_p & \mathbb{Z}_p \\ p^{n_p}\mathbb{Z}_p & \mathbb{Z}_p \end{pmatrix}.$$

For any divisor M of N, we define the congruence subgroup $\Gamma_0(N; M)$ as follows:

$$\Gamma_0(N; M) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : a \equiv d \equiv 1 \bmod M,\ c \equiv 0 \bmod N \right\}.$$

Note that $\Gamma_0(N; 1) = \Gamma_0(N)$ and $\Gamma_0(N; N) = \Gamma_1(N)$. We have the following proposition.

Proposition 2.2. *Suppose that M is a positive integer such that M^2 divides N. Let $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ be such that $C(\sigma) = N/M$. Then $\sigma\Gamma_0(N; M)\sigma^{-1} \subseteq \Gamma_0(N; M)$.*

Proof. Recall that $C(\sigma) = N/M$ iff the lower left entry of σ is a multiple of N/M. Now the result follows from the equation

$$\begin{aligned} &\begin{pmatrix} a & b \\ Nc/M & d \end{pmatrix}^{-1} \begin{pmatrix} 1+Mp & q \\ Nr & 1+Ms \end{pmatrix} \begin{pmatrix} a & b \\ Nc/M & d \end{pmatrix} \\ &= \begin{pmatrix} 1+Madp+N(bdr-bcs)-(N/M)acq & a^2q+bM(as-ap-br(N/M)) \\ N(dcp+d^2r-cds-c^2q(N/M^2)) & 1+Mads+N(-bdr-bcp)+(N/M)acq \end{pmatrix}. \end{aligned} \tag{1}$$

□

Let $\mathcal{P}_N$ denote the set of primes dividing N. For each subset $S \subseteq \mathcal{P}_N$, we define $N_S = \prod_{p\in S} p^{n_p}$ where we understand $N_\emptyset = 1$. We set

$$\begin{aligned} \mathcal{W}(S) = \Big\{ W \in M_2(\mathbb{Z}) : W &\equiv \begin{pmatrix} 0 & * \\ 0 & 0 \end{pmatrix} \bmod N_S, \\ W &\equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \bmod N, \det(W) = N_S \Big\}. \end{aligned}$$

The elements of $\mathcal{W}(S)$ (considered as operators on $\mathbb{H}$) are called the *Atkin–Lehner operators*. It is well-known [AL70] that all elements W in $\mathcal{W}(S)$ satisfy $W^2 \in Z(\mathbb{Q})\Gamma_0(N)$. The main other property of an Atkin–Lehner operator W we need is that

$$W \in \begin{cases} K_0(p^{n_p}) \begin{pmatrix} 0 & -1 \\ p^{n_p} & 0 \end{pmatrix} & \text{if } p \in S, \\ K_0(p^{n_p}) & \text{if } p \notin S, \end{cases}$$

which follows directly from the definitions.

Proposition 2.3. *Let $\tau \in \mathrm{SL}_2(\mathbb{Z})$. Then there exists a subset S of $\mathcal{P}_N$, an Atkin–Lehner operator $W \in \mathcal{W}(S)$, positive integers M_1, M such that $M^2 \mid N$, $M_1 = (M, N_S)$, $M_1^2 \mid N_S$, and an element $n \in N(\mathbb{Q})$ such that the matrix σ defined by*

$$\sigma = W\tau n \begin{pmatrix} 1/M_1 & 0 \\ 0 & M_1/N_S \end{pmatrix}$$

has the following properties:

$$\sigma \in \mathrm{SL}_2(\mathbb{Z}), \qquad C(\sigma) = N/M.$$

The proof will be essentially local in nature. For any $\tau = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{Z}_p)$, we define $c_p(\tau) = \min(v_p(c), n_p)$ and $w_p(\tau) = \max(n_p - 2c_p(\tau), 0)$. Note that the integers $c_p(\tau)$ and $w_p(\tau)$ both range between 0 and n_p.

Lemma 2.4. *The integers $c_p(\tau)$ and $w_p(\tau)$ depend only on the double coset $K_0(p^{n_p})\tau N(\mathbb{Z}_p)$. Moreover, this double coset contains the matrix $\left(\begin{smallmatrix} 1 & 0 \\ p^{c_p(\tau)} & v \end{smallmatrix}\right)$ for some $v \in \mathbb{Z}_p^\times$.*

Proof. The first assertion is immediate by looking at the matrix products in $K_0(p^{n_p})\tau N(\mathbb{Z}_p)$ modulo p^{n_p}. For the second assertion, we consider three cases for $\tau = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. The first case is when $v_p(c) = 0$. Then c is a unit and the result follows from the equation

$$\begin{pmatrix} 1 & (1-a)/c \\ 0 & 1/c \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & (ad-bc-d)/c \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & (ad-bc)/c \end{pmatrix}.$$

The second case is $0 < k = v_p(c) < n_p$. Then a, d and $c_1 = c/p^k$ are all units. The result follows from the equation

$$\begin{pmatrix} 1/a & 0 \\ 0 & 1/c_1 \end{pmatrix} \begin{pmatrix} a & b \\ p^k c_1 & d \end{pmatrix} \begin{pmatrix} 1 & -b/a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ p^k & d/c_1 + p^k b/a \end{pmatrix}.$$

The third case is $0 < n_p \le k = v_p(c)$. Then a is a unit and the result follows from the equation

$$\begin{pmatrix} 1/a & 0 \\ (p^{n_p} - c)/a & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & (1 - c/p^{n_p})/a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ p^{n_p} & (ad-bc)/a \end{pmatrix}. \qquad \square$$

Lemma 2.5. *Let $\tau \in \mathrm{SL}_2(\mathbb{Z})$. Then $W(\tau) = \prod_{p|N} p^{w_p(\tau)}$ and $C(\tau) = \prod_{p|N} p^{c_p(\tau)}$.*

Proof. The equation $C(\tau) = \prod_{p|N} p^{c_p(\tau)}$ follows immediately from the relevant definitions. The relation $W(\tau) = \prod_{p|N} p^{w_p(\tau)}$ follows from the formulas $w_p(\tau) = \min(n_p - 2c_p(\tau), 0)$ and $W(\tau) = N/(C(\tau)^2, N)$. $\square$

Lemma 2.6. *Let $p \mid N$ and suppose that $\tau \in \mathrm{GL}_2(\mathbb{Z}_p)$ satisfies $w_p(\tau) > 0$. Then there exists $n \in N(\mathbb{Q}_p)$ such that for all $y_1, y_2 \in \mathbb{Z}_p^\times$ and $\gamma \in K_0(p^{n_p})$, the element $\sigma \in \mathrm{GL}_2(\mathbb{Q}_p)$ defined via*

$$\sigma = \gamma \begin{pmatrix} 0 & -1 \\ p^{n_p} & 0 \end{pmatrix} \tau n \begin{pmatrix} y_1 p^{-c_p(\tau)} & 0 \\ 0 & y_2 p^{c_p(\tau)-n_p} \end{pmatrix}$$

has the following properties:

$$\sigma \in \mathrm{GL}_2(\mathbb{Z}_p), \qquad c_p(\sigma) = n_p - c_p(\tau).$$

Proof. Note that if σ has the required properties, then so do all elements in $K_0(p^{n_p})\sigma$. Hence we may assume that $\gamma = 1$. Moreover, since $\left(\begin{smallmatrix} 0 & -1 \\ p^{n_p} & 0 \end{smallmatrix}\right)$ normalizes $K_0(p^{n_p})$, it follows that if the proposition is true for some τ, it is true for all $\tau \in K_0(p^{n_p})\tau N(\mathbb{Z}_p)$. Hence, using Lemma 2.4 we can assume without loss of generality that $\tau = \left(\begin{smallmatrix} 1 & 0 \\ p^{c_p(\tau)} & v \end{smallmatrix}\right)$ for some $v \in \mathbb{Z}_p^\times$. Define $n = \left(\begin{smallmatrix} 1 & -vp^{-c_p(\tau)} \\ & 1 \end{smallmatrix}\right)$. The condition $w_p(\tau) > 0$ is equivalent to $2c_p(\tau) < n_p$. Then

$$\sigma = \begin{pmatrix} 0 & -1 \\ p^{n_p} & 0 \end{pmatrix} \tau n \begin{pmatrix} y_1 p^{-c_p(\tau)} & 0 \\ 0 & y_2 p^{c_p(\tau)-n_p} \end{pmatrix} = \begin{pmatrix} -y_1 & 0 \\ p^{n_p - c_p(\tau)} y_1 & -vy_2 \end{pmatrix}.$$

So by inspection, we see that $\sigma \in \mathrm{GL}_2(\mathbb{Z}_p)$ and $c_p(\sigma) = n_p - c_p(\tau)$. □

Lemma 2.7. *Let $p \mid N$ and suppose that $\tau \in \mathrm{GL}_2(\mathbb{Z}_p)$ satisfies $w_p(\tau) = 0$. Then for all $y_1, y_2 \in \mathbb{Z}_p^\times$, $\gamma \in K_0(p^{n_p})$, $n \in N(\mathbb{Z}_p)$, the element $\sigma \in \mathrm{GL}_2(\mathbb{Z}_p)$ defined via*

$$\sigma = \gamma \tau n \begin{pmatrix} y_1 & 0 \\ 0 & y_2 \end{pmatrix}$$

satisfies $c_p(\sigma) = c_p(\tau)$.

Proof. This is immediate from the definitions. □

Proof of Proposition 2.3. Let S be the set of primes p for which $w_p(\tau) > 0$ (i.e., $c_p(\tau) < n_p/2$). Define $M_1 = \prod_{p \in S} p^{c_p(\tau)}$; note that $M_1^2 \mid N_S$. For each $p \mid N$, we set $m_p = c_p(\tau)$ if $p \in S$ and $m_p = n_p - c_p(\tau)$ if $p \notin S$. Define

$$M = M_1 \prod_{p \mid N,\, p \notin S} p^{n_p - c_p(\tau)} = \prod_{p \mid N} p^{m_p}.$$

Note that $M^2 \mid N$. Pick any $W \in \mathcal{W}(S)$. Note that W considered as an element of $\mathrm{GL}_2(\mathbb{Q}_p)$ lies in $K_0(p^{n_p})\left(\begin{smallmatrix} 0 & -1 \\ p^{n_p} & 0 \end{smallmatrix}\right)$ for each p in S and lies in $K_0(p^{n_p})$ for each prime outside S. For each $p \in S$, Lemma 2.6 provides an $x_p \in N(\mathbb{Q}_p)$ such that

$$\sigma_p = W \tau x_p \begin{pmatrix} 1/M_1 & 0 \\ 0 & M_1/N_S \end{pmatrix}$$

has the following properties:

$$\sigma_p \in \mathrm{GL}_2(\mathbb{Z}_p), \qquad c_p(\sigma_p) = n_p - c_p(\tau),$$

By strong approximation, we can pick $n \in N(\mathbb{Q})$ such that

$$n \equiv x_p \bmod p^{n_p} \quad \text{for all } p \in S, \qquad n \in N(\mathbb{Z}_p) \quad \text{for all } p \notin S.$$

We claim that this choice of n has the required properties. Indeed, let

$$\sigma = W\tau n \begin{pmatrix} 1/M_1 & 0 \\ 0 & M_1/N_S \end{pmatrix}.$$

Then our choice implies that $\det(\sigma) = 1$, and moreover Lemmas 2.6 and 2.7 ensure that for all primes p, we have $\sigma \in \mathrm{GL}_2(\mathbb{Z}_p)$, $c_p(\sigma) = n_p - m_p$. It follows that $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ and $C(\sigma) = N/M$. □

3. A gap principle

Our goal in this section is to prove the following proposition.

Proposition 3.1. *Let $z \in \mathbb{H}$. Then there exists a subset S of $\mathcal{P}_N$, an Atkin–Lehner operator $W \in \mathcal{W}(S)$, an integer M such that $M^2 \mid N$, and an element $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ such that*

$$C(\sigma) = N/M, \qquad \mathrm{Im}(\sigma^{-1}Wz) \geq \sqrt{3}\, M^2/(2N),$$

and for any $(0, 0) \neq (c, d) \in \mathbb{Z}^2$, we have

$$|c(\sigma^{-1}Wz) + d|^2 \geq \frac{3M^2(c, N/M^2)}{4N}.$$

Remark 3.2. In the above proposition, we can always choose σ to lie in some fixed set of representatives for $\Gamma_0(N)\backslash\mathrm{SL}_2(\mathbb{Z})/N(\mathbb{Z})$. This is because both $\mathrm{Im}(\sigma^{-1}Wz)$ and the set of possible values for $(c\sigma^{-1}Wz + d)$ depend only on the class of σ in $\mathrm{SL}_2(\mathbb{Z})/N(\mathbb{Z})$, while any product of $\Gamma_0(N)$ on the left of σ can be absorbed into the W. In particular, if N is squarefree, then σ can be taken to equal the identity (in which case $M = 1$ and our result essentially reduces to [HT12, Lemma 1]).

We begin with an elementary lemma.

Lemma 3.3. *Let $z_0 \in \mathbb{H}$ be such that $\mathrm{Im}(z_0) \geq \sqrt{3}/2$. Then for all $n \in N(\mathbb{R})$ and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we have $\mathrm{Im}(z_0) \geq \frac{3}{4}\,\mathrm{Im}(\gamma n z_0)$.*

Proof. By replacing z_0 by nz_0 if necessary, we may assume that $n = 1$. Also, by translating z_0 horizontally by an integer, we may assume that $\mathrm{Re}(z_0)$ lies between $-1/2$ and $1/2$. Now, the conclusion is immediate from the standard tiling of the upper half-plane by $\mathrm{SL}_2(\mathbb{Z})$-translates of the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$. □

Proof of Proposition 3.1. By the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$, there exists $\tau \in \mathrm{SL}_2(\mathbb{Z})$ such that $z = \tau z_0$ where $\mathrm{Im}(z_0) \geq \sqrt{3}/2$. Now, let σ, W, N_S, M, M_1 be as in Proposition 2.3. Then $C(\sigma) = N/M$. Note that $M_1^2/N_S \geq M^2/N$ (since $N_S/M_1^2 \mid N/M^2$) and $\sigma^{-1}W\tau = \left(\begin{smallmatrix} M_1 & 0 \\ 0 & N_S/M_1 \end{smallmatrix}\right)n^{-1}$. Furthermore

$$\begin{aligned} \mathrm{Im}(\sigma^{-1}Wz) &= \mathrm{Im}(\sigma^{-1}W\tau z_0) = \mathrm{Im}\left(\begin{pmatrix} M_1 & 0 \\ 0 & N_S/M_1 \end{pmatrix} n^{-1} z_0\right) \\ &= (M_1^2/N_S)\,\mathrm{Im}(z_0) \geq \frac{M^2\,\mathrm{Im}(z_0)}{N} \geq \frac{\sqrt{3}\,M^2}{2N}. \end{aligned}$$

Next, given any pair $(c, d) \neq (0, 0)$, we need to prove that $|c(\sigma^{-1}Wz) + d|^2 \geq 3M^2(c, N/M^2)/(4N)$. It suffices to prove this for c and d coprime. Let $c_1 = c/(c, N_S/M_1^2)$ and $n_2 = N_S/(cM_1^2, N_S)$. Note that c_1 and dn_2 are coprime, and also

$$\frac{1}{n_2} = \frac{(c, N_S/M_1^2)}{N_S/M_1^2} \geq \frac{(c, N/M^2)M^2}{N}.$$

Pick any $\gamma = \left(\begin{smallmatrix} a & b \\ c_1 & dn_2 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and set

$$\gamma' = \begin{pmatrix} a & bM_1^2/N_S \\ cn_2 & dn_2 \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Q}).$$

By the previous lemma, $\mathrm{Im}(z_0) \geq (3/4)\,\mathrm{Im}(\gamma n z_0)$ for all $n \in N(\mathbb{R})$. Also, recall that $\mathrm{Im}(\sigma^{-1}Wz) = (M_1^2/N_S)\,\mathrm{Im}(z_0)$. Then

$$\begin{aligned} \frac{M_1^2\,\mathrm{Im}(z_0)}{n_2 N_S |c(\sigma^{-1}Wz) + d|^2} &= \frac{\mathrm{Im}(\sigma^{-1}Wz)}{n_2|c(\sigma^{-1}Wz) + d|^2} = \mathrm{Im}(\gamma'\sigma^{-1}Wz) \\ &= \mathrm{Im}\left(\gamma' \begin{pmatrix} M_1 & 0 \\ 0 & N_S/M_1 \end{pmatrix} n^{-1} z_0\right) = \frac{M_1^2}{N_S}\,\mathrm{Im}(\gamma n^{-1} z_0) \\ &\leq \frac{4M_1^2}{3N_S}\,\mathrm{Im}(z_0), \end{aligned}$$

giving

$$|c(\sigma^{-1}Wz) + d|^2 \geq \frac{3}{4n_2} \geq \frac{3(c, N/M^2)M^2}{4N},$$

as desired. □

4. Some counting results

Let $1 \leq N = N_2 N_0^2$ with N_2 squarefree and let M be a positive integer that divides N_0 (so $M^2 \mid N$). We define the region $G(N; M) \subset \mathbb{H}$ to consist of the points $z = x + iy \in \mathbb{H}$ with the following properties:

- $y \geq \sqrt{3}\,M^2/2N$.
- For any pair $(c, d) \neq (0, 0)$ of integers, we have $|cz + d|^2 \geq 3M^2(c, N/M^2)/(4N)$.

For $z \in \mathbb{H}$, any $\delta > 0$, and any integer $l \geq 1$, define, with $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$:

$$\begin{aligned} \Delta(l, N; M) &:= \{\gamma \in M_2(\mathbb{Z}) : c \equiv 0 \bmod N,\ a \equiv 1 \bmod M,\ \det(\gamma) = l\}, \\ N_*(z, l, \delta, N; M) &:= |\{\gamma \in \Delta(l, N; M) : u(\gamma z, z) \leq \delta,\ c \neq 0,\ (a + d)^2 \neq 4l\}|, \\ N_u(z, l, \delta, N; M) &:= |\{\gamma = \Delta(l, N; M) : u(\gamma z, z) \leq \delta,\ c = 0,\ (a + d)^2 \neq 4l\}|, \\ N_p(z, l, \delta, N; M) &:= |\{\gamma = \Delta(l, N; M) : u(\gamma z, z) \leq \delta,\ (a + d)^2 = 4l\}|, \\ N(z, l, \delta, N; M) &:= N_*(z, l, \delta, N; M) + N_u(z, l, \delta, N; M) + N_p(z, l, \delta, N; M). \end{aligned}$$

Remark 4.1. These definitions are similar to ones in [HT13] except that we have the added condition $a \equiv 1 \bmod M$.

We will estimate the above quantities, ultimately proving Proposition 4.6 which will be useful for the amplification method to be used later. For the convenience of the reader, we begin by quoting a result that will be frequently used in this section.

Lemma 4.2 ([Sch68, Lemma 2]). *Let L be a lattice in $\mathbb{R}^2$ and $D \subset \mathbb{R}^2$ be a disc of radius R. If λ_1 is the distance from the origin of the shortest vector in L, and d is the covolume of L, then*

$$|L \cap D| \ll 1 + \frac{R}{\lambda_1} + \frac{R^2}{d}.$$

The next lemma, which counts general matrices, is a mild generalization of [HT13, Lemmas 2.2 and 2.3].

Lemma 4.3. *Let $z \in G(N; M)$ and $M^2 \leq L \leq N^{O(1)}$. Let $1 \leq l_1 \equiv 1 \bmod M$ and $l_1 \leq N^{O(1)}$. Then*

$$\sum_{\substack{1 \leq l \leq L \\ l \equiv 1 \bmod M}} N_*(z, l, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)} \left(\frac{L}{MNy} + \frac{L^{3/2}}{M^2\sqrt{N}} + \frac{L^2}{M^2N} \right), \tag{2}$$

$$\sum_{\substack{1 \leq l \leq L \\ l \equiv 1 \bmod M}} N_*(z, l^2, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)} \left(\frac{L}{Ny} + \frac{L^2}{M\sqrt{N}} + \frac{L^3}{MN} \right), \tag{3}$$

$$\sum_{\substack{1 \leq l_2 \leq L \\ l_2 \equiv 1 \bmod M}} N_*(z, l_1 l_2^2, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)} \left(\frac{L^{3/2}}{Ny} + \frac{L^3}{M\sqrt{N}} + \frac{L^{9/2}}{MN} \right). \tag{4}$$

Proof. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ satisfy the conditions $\gamma \in \Delta(l, N; M)$, $u(\gamma z, z) \leq \delta$, $c \neq 0$, $(a+d)^2 \neq 4l$ for some $1 \leq l \leq L$, $l \equiv 1 \bmod M$. As in [HT13], we conclude that there are $\ll_\epsilon N^{O(\epsilon)} L^{1/2}/(Ny)$ possible values for c, and that

$$|-cz^2 + (a-d)z + b|^2 \leq Ly^2 N^\epsilon.$$

We note that $a - d \equiv 0 \bmod M$. Setting $t = (a-d)/M$, and applying Lemma 4.2 to the lattice $\langle 1, Mz \rangle$ (note that $R = \sqrt{L}\, yN^{\epsilon/2}$, $d = My$ and $\lambda_1^2 \gg M^2/N$), we conclude that for each c, the number of pairs (t, b) satisfying the above inequality is $\ll_\epsilon N^\epsilon(1 + \sqrt{LN}\, y/M + Ly/M)$. Moreover, as in [HT13], we conclude that $|a + d| \ll_\epsilon N^\epsilon L^{1/2}$. Since $a + d \equiv 2 \bmod M$, it follows that there are $\ll_\epsilon N^\epsilon(L^{1/2}/M)$ possibilities for $a + d$. This concludes the proof of (2).

Next, we prove (3). It suffices to show that

$$\sum_{\substack{1 \leq l \leq L \\ l = m^2 \\ m \equiv 1 \bmod M}} N_*(z, l, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)} \left(\frac{L^{1/2}}{Ny} + \frac{L}{M\sqrt{N}} + \frac{L^{3/2}}{MN} \right).$$

To prove this, we proceed exactly as in the previous case, except that we deal with the number of possibilities for $a + d$ differently. Indeed, we have the equation

$$(a + d - 2m)(a + d + 2m) = (a-d)^2 + 4bc.$$

Hence, given $c, b, a-d$, there are $\ll_\epsilon N^\epsilon$ pairs $(a+d, m)$ satisfying the given constraints. This proves the desired bound.

Finally, we deal with (4). Once again, we proceed as in the first case (note that now $R = L^{3/2}yN^{\epsilon/2}$, $d = My$ and $\lambda_1^2 \gg M^2/N$). However, this time we deal with $a + d$ differently, namely via the observation that the pair $(a+d, l_2)$ satisfies a generalized Pell equation. As the details are identical to [HT13, Lemma 2.3], we omit them.[4] □

Next we count upper-triangular matrices. The lemma below is a mild generalization of [HT13, Lemma 2.4].

Lemma 4.4. *Let $z \in G(N; M)$ and $M \le L \le N^{O(1)}$. Then*

$$\sum_{\substack{1\le l_1,l_2\le L\\ l_1\equiv l_2\equiv 1 \bmod M\\ l_1,l_2 \text{ are primes}}} N_u(z, l_1l_2, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)}\left(\frac{L}{M} + \frac{L^2y\sqrt{N}}{M^2} + \frac{L^3y}{M^2}\right), \tag{5}$$

$$\sum_{\substack{1\le l_1,l_2\le L\\ l_1\equiv l_2\equiv 1 \bmod M\\ l_1,l_2 \text{ are primes}}} N_u(z, l_1l_2^2, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)}\left(\frac{L}{M} + \frac{L^{5/2}y\sqrt{N}}{M^2} + \frac{L^4y}{M^2}\right), \tag{6}$$

$$\sum_{\substack{1\le l_1,l_2\le L\\ l_1\equiv l_2\equiv 1 \bmod M\\ l_1,l_2 \text{ are primes}}} N_u(z, l_1^2l_2^2, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)}\left(1 + \frac{L^2y\sqrt{N}}{M} + \frac{L^4y}{M}\right), \tag{7}$$

$$\sum_{\substack{1\le l_1\le L\\ l_1\equiv 1 \bmod M}} N_u(z, l_1, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)}\left(1 + \frac{L^{1/2}y\sqrt{N}}{M} + \frac{Ly}{M}\right). \tag{8}$$

Proof. Let $\gamma = \left(\begin{smallmatrix} a & b\\ 0 & d\end{smallmatrix}\right)$ satisfy $\gamma \in \Delta(l, N; M)$, $u(\gamma z, z) \le \delta$, $(a+d)^2 \ne 4l$ for some $1 \le l \le \Lambda$, and $l \equiv 1 \bmod M$. As in [HT13], we conclude that

$$|(a-d)z + b|^2 \le \Lambda y^2 N^\epsilon.$$

Also, note that $a - d \equiv 0 \bmod M$. The rest of the proof is identical to the proof of [HT13, Lemma 2.4], with the modifications for M as in Lemma 4.3 above. □

Finally, we count parabolic matrices. The next lemma is a significant extension of [HT12, Lemma 2].

Lemma 4.5. *Let $z \in G(N; M)$ and $1 \le l \equiv 1 \bmod M$. Then:*

(1) *$N_p(z, l, N^\epsilon, N; M) = 0$ if l is not a perfect square.*
(2) *Suppose that $l = m^2$ with $m > 0$ an integer. Suppose also that $l, y \le N^{O(1)}$. Then*

$$N_p(z, l, N^\epsilon, N; M) \ll_\epsilon 1 + N^{O(\epsilon)}\left(\frac{myN_0}{M} + \frac{mN_0}{N}\right).$$

[4] In fact, it turns out that we can completely avoid dealing with this case by making a small adjustment in the proof of our main theorem; see Remark 4.7.

Proof. Let $\gamma \in \Delta(l, N; M)$ be such that $u(\gamma z, z) \leq N^\epsilon$ and $\mathrm{tr}(\gamma)^2 = 4l$. Then γ fixes some point $\tau(\infty)$ where $\tau \in \mathrm{SL}_2(\mathbb{Z})$. Hence $\gamma' = \tau^{-1}\gamma\tau$ fixes the point ∞, and hence is a parabolic upper-triangular matrix with integer coefficients and determinant l. It follows that l must be a perfect square. Writing $\gamma' = \pm\left(\begin{smallmatrix} m & t \\ 0 & m \end{smallmatrix}\right)$ (where $m^2 = l$ and $t \in \mathbb{Z}$) and $\tau^{-1} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we see that $\gamma = \pm\left(\begin{smallmatrix} m+cdt & d^2t \\ -c^2t & m-cdt \end{smallmatrix}\right)$. This shows that $N \mid c^2 t$. Moreover $u(\gamma z, z) = u(\gamma' z', z')$ where $z' = \tau^{-1} z$. Writing $z' = x' + iy'$, we note that

$$N^\epsilon \geq u(\gamma' z', z') = \frac{t^2}{4ly'^2} = \frac{t^2|cz+d|^4}{4ly^2} \gg \frac{t^2(c, N/M^2)^2 M^4}{ly^2N^2}. \tag{9}$$

Next, if $t = 0$ then $\gamma' = \gamma = \pm\left(\begin{smallmatrix} m & 0 \\ 0 & m \end{smallmatrix}\right)$ is the only possibility. So it suffices to consider the case $t > 0$. We claim that if $t > 0$ then

$$\frac{t^2M^4(c, N/M^2)^2}{N^2} \geq \frac{t_0^2M^2}{N_0^2}$$

where $t_0 = t/(t, N^\infty)$ is the N-free part of t. Let p divide N, with $p^{t'} \,||\, t$, $p^{c'} \,||\, c$, $p^{n_p} \,||\, N$, $p^{m_p} \,||\, M$. Then it suffices to show that

$$\min(2t' + 2c' + 4m_p, 2t' + 2n_p) \geq 2n_p + 2m_p - 2\lfloor n_p/2 \rfloor.$$

If $2c' + 4m_p + 2t' \geq 2n_p$ this is immediate (note that $m_p \leq \lfloor n_p/2 \rfloor$). So we assume that $2c' + 4m_p + 2t' < 2n_p$. It suffices in this case to prove that $2t' + 2c' \geq 2n_p - 2\lfloor n_p/2 \rfloor$. But this follows immediately from the fact that $2t' + 2c' \geq n_p$ and $2t' + 2c'$ is even.

So we have proved that if $t \neq 0$ then

$$N^\epsilon \geq \frac{t^2(c, N/M^2)^2M^4}{ly^2N^2} \geq \frac{t_0^2M^2}{ly^2N_0^2}.$$

Hence

$$t_0 \leq N^\epsilon \frac{yN_0\sqrt{l}}{M}.$$

On the other hand, (9) implies that

$$t \leq N^\epsilon \frac{y\sqrt{l}}{|cz+d|^2} \ll \frac{yN\sqrt{l}}{M^2}.$$

Write $t = t_0t_1$ where $t_1 \mid N^\infty$. Given any such $t = t_0t_1$, let us count the number of admissible γ; this reduces to counting the number of admissible c, d. Given any integer $f = \prod p_i^{a_i}$, we define (temporarily) $\{\sqrt{f}\} = \prod p_i^{\lceil a_i/2 \rceil}$. Note that if f divides a^2 for some integer a, then $\{\sqrt{f}\}$ divides a. Note also that $\{\sqrt{N}\} = N_2N_0 = N/N_0$. We have already proved that $N \mid c^2t_1$. It follows that $\{\sqrt{g}\}$ divides c where $g = N/(t_1, N)$. Note also that $\{\sqrt{g}\} \geq N_2N_0/(t_1, N) \geq N/(N_0t_1)$.

Let us count the number of pairs of integers c, d such that $|cz + d|^2 \ll N^\epsilon y\sqrt{l}/t$ and $\{\sqrt{g}\}$ divides c. Considering the lattice $\langle 1, \{\sqrt{g}\}z\rangle$ we see that the quantity λ_1 for this lattice satisfies

$$\lambda_1^2 \geq \frac{(\{\sqrt{g}\}M^2, N)}{N} \geq \frac{\{\sqrt{g}\}M}{N} \geq \frac{M}{N_0 t_1}$$

(we have used the fact that $\{\sqrt{g}\}M$ divides N, which follows as $\{\sqrt{g}\}$ divides N_2N_0 and M divides N_0). Furthermore, the covolume d of this lattice satisfies $d \geq Ny/(N_0t_1)$. Hence, by Lemma 4.2, the total number of admissible c, d for each fixed $t = t_0t_1$ is

$$\ll_\epsilon 1 + \frac{N^\epsilon((N_0t_1)^{1/2}y^{1/2}l^{1/4}}{t^{1/2}\sqrt{M}} + \frac{N^\epsilon\sqrt{l}\,N_0}{t_0N}.$$

Hence, the total number of parabolic matrices $\gamma \in \Delta(l, N; M)$ such that $u(\gamma z, z) \leq N^\epsilon$ is

$$\begin{aligned}
&\ll_\epsilon 1 + N^\epsilon \sum_{\substack{1\leq t_0\leq yN_0\sqrt{l}/M\\ (t_0,N)=1}} \sum_{\substack{1\leq t_1\leq yN\sqrt{l}/(M^2t_0)\\ t_1|N^\infty}} \left(1 + \frac{(N_0ym)^{1/2}}{(t_0M)^{1/2}} + \frac{mN_0}{t_0N}\right)\\
&\ll_\epsilon 1 + N^{O(\epsilon)}\left(\frac{yN_0m}{M} + \frac{mN_0}{N}\right).
\end{aligned}$$

In the last step above, we have used a fact that will also be used a few times later: for all positive integers X, N, one has $\sum_{t_1\leq X,\, t_1|N^\infty} 1 \ll_\epsilon (NX)^\epsilon$. This follows from Rankin's trick.[5] □

Combining all the above bounds, we get the following proposition, which is all that we will use later.

Proposition 4.6. *Let $1 \leq N = N_2N_0^2$ with N_2 squarefree and let M be a positive integer that divides N_0. Suppose that $z = x+iy \in G(N; M)$ and assume further that $y \leq N^{-1/2}$. Let $M^2 \leq \Lambda \leq N^{O(1)}$. Define*

$$y_l := \begin{cases} \Lambda/M, & l = 1,\\ 1, & l \in \{l_1, l_1l_2, l_1l_2^2, l_1^2l_2^2\} \text{ with } \Lambda < l_1, l_2 < 2\Lambda \text{ primes, } l_1 \equiv l_2 \equiv 1 \bmod M,\\ 0 & \text{otherwise.}\end{cases}$$

Then

$$\sum_{l\geq 1} \frac{y_l}{\sqrt{l}} N(z, l, N^\epsilon, N; M) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^2yN_0}{M^3} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}} + \frac{\Lambda^4}{MN}\right). \quad (10)$$

Proof. The contribution to the LHS of (10) from the parabolic matrices is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{M^2}\left(1 + \frac{\Lambda^2yN_0}{M} + \frac{\Lambda^2\sqrt{N}}{N}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^2yN_0}{M^3} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}}\right)$$

in view of Lemma 4.5 and $\Lambda \geq M$.

[5] Observe that $\sum_{t_1\leq X,\, t_1|N^\infty} 1 \leq X^\epsilon \prod_{t_1|N^\infty} t_1^{-\epsilon} = X^\epsilon \prod_{p|N}(1 - p^{-\epsilon})^{-1}$ and then apply the divisor bound.

The contribution to the LHS of (10) from the upper-triangular matrices with $l = 1$ is

$$\ll_\epsilon \frac{N^{O(\epsilon)}\Lambda}{M}\left(1 + y\frac{\sqrt{N}}{\sqrt{M}}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M}\right)$$

by (8). For $\Lambda < l < 2\Lambda$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\sqrt{\Lambda}}\left(1 + y\frac{\sqrt{\Lambda N}}{M} + \frac{\Lambda y}{M}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M}\right)$$

by (8). For $\Lambda^2 < l < 4\Lambda^2$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\Lambda}\left(\frac{\Lambda}{M} + \frac{\Lambda^2 y\sqrt{N}}{M^2} + \frac{\Lambda^3 y}{M^2}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}}\right)$$

by (5). For $\Lambda^3 < l < 8\Lambda^3$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\Lambda^{3/2}}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2} y\sqrt{N}}{M^2} + \frac{\Lambda^4 y}{M^2}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}}\right)$$

by (6). For $\Lambda^4 < l$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\Lambda^2}\left(1 + \frac{\Lambda^2 y\sqrt{N}}{M} + \frac{\Lambda^4 y}{M}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}}\right)$$

by (7).

The contribution to the LHS of (10) from the general matrices with $l = 1$ is

$$\ll_\epsilon \frac{N^{O(\epsilon)}\Lambda}{M}\left(\frac{M}{Ny} + \frac{M}{\sqrt{N}} + \frac{M^2}{N}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M}\right)$$

by (2). For $\Lambda < l < 2\Lambda$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\sqrt{\Lambda}}\left(\frac{\Lambda}{MNy} + \frac{\Lambda^{3/2}}{M^2\sqrt{N}} + \frac{\Lambda^2}{M^2 N}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}}\right)$$

by (2). For $\Lambda^2 < l < 4\Lambda^2$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\Lambda}\left(\frac{\Lambda^2}{MNy} + \frac{\Lambda^3}{M^2\sqrt{N}} + \frac{\Lambda^4}{M^2 N}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}} + \frac{\Lambda^4}{MN}\right)$$

by (2). For $\Lambda^3 < l < 8\Lambda^3$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}\Lambda}{\Lambda^{3/2}M}\left(\frac{\Lambda^{3/2}}{Ny} + \frac{\Lambda^3}{M\sqrt{N}} + \frac{\Lambda^{9/2}}{MN}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}} + \frac{\Lambda^4}{MN}\right)$$

by (4). For $\Lambda^4 < l$, it is

$$\ll_\epsilon \frac{N^{O(\epsilon)}}{\Lambda^2}\left(\frac{\Lambda^2}{Ny} + \frac{\Lambda^4}{M\sqrt{N}} + \frac{\Lambda^6}{MN}\right) \ll_\epsilon N^{O(\epsilon)}\left(\frac{\Lambda}{M} + \frac{\Lambda^{5/2}}{M^2\sqrt{N}} + \frac{\Lambda^4}{MN}\right)$$

by (3).

The proof is complete. □

Remark 4.7. Gergely Harcos and Guillaume Ricotta have pointed out to the author the possibility of using an improved amplifier, as in [BHM16], in the proof of our main result. With this modification, we would only need to prove a weaker version of the above proposition, where the terms corresponding to $l = l_1$ or $l_1 l_2^2$ are removed.

5. Hecke operators on $\Gamma_0(N; M)$

We begin by recalling the usual Hecke algebra on $\Gamma_0(N)$. Define

$$\Delta_0(N) := \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}) : c \equiv 0 \bmod N,\ \det(\gamma) > 0 \right\},$$

$$\mathcal{H}_0(N) := \left\{ \sum_{\alpha \in \Delta_0(N)} t_\alpha \Gamma_0(N) \alpha \Gamma_0(N) : t_\alpha \in \mathbb{Z},\ t_\alpha = 0 \text{ for almost all } \alpha \right\}.$$

Next, for any divisor M of N, define

$$\Delta_0(N; M) := \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}) : c \equiv 0 \bmod N,\ a \equiv 1 \bmod M,\ \det(\gamma) > 0 \right\},$$

$$\mathcal{H}_0(N; M) := \left\{ \sum_{\alpha \in \Delta_0(N; M)} t_\alpha \Gamma_0(N; M) \alpha \Gamma_0(N; M) : t_\alpha \in \mathbb{Z},\ t_\alpha = 0 \text{ for almost all } \alpha \right\}.$$

Elements of $\mathcal{H}_0(N)$ (resp. $\mathcal{H}_0(N; M)$) act on Maass cusp forms f on the group $\Gamma_0(N)$ (resp. $\Gamma_0(N; M)$) in the usual manner. We will normalize this action as follows:

$$f | \Gamma \alpha \Gamma = \det(\alpha)^{-1/2} \sum_{\gamma \in \Gamma \backslash \Gamma \alpha \Gamma} f | \gamma, \qquad \text{where } \Gamma = \Gamma_0(N) \text{ or } \Gamma_0(N; M).$$

Consider the natural map from $\mathcal{H}_0(N; M)$ to $\mathcal{H}_0(N)$ defined via

$$\Gamma_0(N; M) \alpha \Gamma_0(N; M) \mapsto \Gamma_0(N) \alpha \Gamma_0(N).$$

Standard arguments (see [Miy06, remarks above Thm. 4.5.19]) imply that this map is an *isomorphism* of Hecke algebras. For $T \in \mathcal{H}_0(N)$, let T' denote its image in $\mathcal{H}_0(N; M)$ under this isomorphism. Then, given any Maass cusp form f on the group $\Gamma_0(N)$ (which can therefore also be thought of as a cusp form on the group $\Gamma_0(N; M)$) and any $T \in \mathcal{H}_0(N)$, one has the compatibility relation

$$f | T = f | T'.$$

For any integer $l \geq 1$, we let $T(l) \in \mathcal{H}_0(N)$ be the $\mathbb{Z}$-linear span of the double cosets $\Gamma_0(N) \alpha \Gamma_0(N)$ for $\alpha \in \Delta_0(N)$ of determinant l. Then, since the above-defined isomorphism of Hecke algebras is determinant-preserving on double cosets, it follows that $T(l)'$ is the $\mathbb{Z}$-linear span of the double cosets $\Gamma_0(N) \alpha \Gamma_0(N)$ where $\alpha \in \Delta_0(N; M)$ has determinant l. Recall the definition from the previous section:

$$\Delta(l, N; M) := \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}) : c \equiv 0 \bmod N,\ a \equiv 1 \bmod M,\ \det(\gamma) = l \right\}.$$

The comments above imply that if f is a Maass cusp form on the group $\Gamma_0(N)$ that is an eigenform for the Hecke operator $T(l)$ with (normalized) eigenvalue $\lambda_f(l)$, then

$$\sum_{\gamma\in\Gamma_0(N;M)\backslash\Delta(l,N;M)} f|\gamma = l^{1/2}\lambda_f(l)f. \tag{11}$$

Now, suppose that M^2 divides N, and $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ satisfies $C(\sigma) = N/M$. Then, by Proposition 2.2, we know that the map $g \mapsto g|\sigma$ is an *endomorphism* of the space of Maass cusp forms for the group $\Gamma_0(N; M)$. It is a natural question if this endomorphism commutes with the Hecke algebra action on the same space. While this is not true in general, it is indeed true for the Hecke operators $T(l)'$ for $l \equiv 1 \bmod M$.

Proposition 5.1. *Let M be a positive integer such that M^2 divides N and let $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ satisfy $C(\sigma) = N/M$. Let g be a Maass cusp form for the group $\Gamma_0(N; M)$. Then, for any positive integer l such that $l \equiv 1 \bmod M$,*

$$g|T(l)'|\sigma = g|\sigma|T(l)'.$$

Proof. Recall that for any Maass cusp form h for the group $\Gamma_0(N; M)$, we have

$$h|T(l)' = l^{-1/2} \sum_{\gamma\in\Gamma_0(N;M)\backslash\Delta(l,N;M)} h|\gamma.$$

So it suffices to prove that

$$\sigma\Delta(l, N; M)\sigma^{-1} = \Delta(l, N; M).$$

But this follows from equation (1). □

This gives us the following corollary, which is all that we will use.

Corollary 5.2. *Let f be a Hecke–Maass cuspidal newform for the group $\Gamma_0(N)$. For any $n \geq 1$, let $\lambda_f(n)$ denote the (normalized) nth Hecke eigenvalue for f. Let M be a positive integer such that M^2 divides N, let $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ satisfy $C(\sigma) = N/M$, and l be a positive integer such that $l \equiv 1 \bmod M$. Then, if $g := f|\sigma$, then*

$$\sum_{\gamma\in\Gamma_0(N;M)\backslash\Delta(l,N;M)} g|\gamma = \lambda_f(l)l^{1/2}g.$$

Proof. This follows by combining (11) and Proposition 5.1. □

Remark 5.3. The results of this section continue to hold in the holomorphic case.

Remark 5.4. The methods and proofs of this section are similar in spirit to those in [Shi71, Section 3.5].

6. The bound via Fourier expansions at width 1 cusps

We will prove the following proposition.

Proposition 6.1. *Let f be a Hecke–Maass cuspidal newform for the group $\Gamma_0(N)$ with Laplace eigenvalue $\lambda = 1/4 + r^2$. Let $M \geq 1$ be an integer such that $M^2 \mid N$ and let $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ satisfy $C(\sigma) = N/M$. Assume further that $\langle f, f\rangle_{\Gamma_0(N)} = 1$ and $|r| \leq R$. Then*

$$|f(\sigma z)| = |(f|\sigma)(z)| \ll_{R,\epsilon} N^{\epsilon} \cdot \begin{cases} \frac{1}{(Ny)^{1/2}}, & 1/N \leq y \leq 1/M^2, \\ \frac{M^{1/2}}{N^{1/2}y^{1/4}}, & 1/M^2 \leq y. \end{cases}$$

The proof will follow from a careful analysis of the Fourier expansion at the cusp $\sigma(\infty)$. Let f be as in the proposition. Then f has the usual Fourier expansion at ∞,

$$f(z) = y^{1/2} \sum_{n\neq 0} \rho(n) K_{ir}(2\pi|n|y) e(nx).$$

We have $|\lambda_f(|n|)| = |\rho(n)/\rho(1)|$, where for each $l \geq 1$, $\lambda_f(l)$ denotes the (normalized) lth Hecke eigenvalue for f. Let $\sigma \in \mathrm{SL}_2(\mathbb{Z})$ satisfy $C(\sigma) = N/M$ (so $W(\sigma) = 1$) and let $h = f|\sigma$. Then h is a Maass cusp form for the congruence subgroup $\Gamma_0(N; M)$. It has a Fourier expansion

$$h(z) = y^{1/2} \sum_{n\neq 0} \rho_\sigma(n) K_{ir}(2\pi|n|y) e(nx).$$

The coefficients $\rho_\sigma(n)$ are the Fourier coefficients of f at the cusp $\sigma(\infty)$; unlike the coefficients at infinity, these cannot be understood simply in terms of Hecke eigenvalues (in fact, they are not even multiplicative). These coefficients were studied adelically in [NPS14, Sec. 3.4.2], and we will use some calculations from there in what follows.[6]

The adelization of the form f gives rise to a cuspidal automorphic representation $\pi = \otimes_{p\leq\infty}\pi_p$ of GL! 2($\mathbb{A}$). Let $W = \prod_{p\leq\infty} W_p$ be the global Whittaker newform in π (with respect to the standard additive character $\psi = \prod_{p\leq\infty} \psi_p$), where we normalize at the non-archimedean places so that $W_p(1) = 1$ for all finite primes p. Fix an integer a such that the cusp $\sigma(\infty)$ contains a representative of the form $\frac{a}{N/M}$ with $(a, N) = 1$. For each integer n, define

$$\lambda_{\sigma,N}(n) = \prod_{p|N} \left((n, p^\infty)^{1/2} W_p \left(\begin{pmatrix} nM^2/N^2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -aM/N \\ 0 & 1 \end{pmatrix} \right) \right).$$

Using the usual adelic intepretation of Fourier coefficients as Whittaker functions, one observes (see [NPS14, discussion following (48)]) that[7]

$$|\rho_\sigma(n)| = \left| \lambda_{\sigma,N}(n)\rho(1)\lambda_f\left(\frac{|n|}{(|n|, N^\infty)}\right) \right|. \tag{12}$$

[6] In [NPS14, Sec. 3.4.2], we restricted ourselves to the holomorphic case but this does not matter because we will only use some local non-archimedean calculations from there which are the same for Maass and holomorphic forms.

[7] A comparison with [NPS14] reveals a conflict between (12) and [NPS14, (49)]. This reflects a typo in [NPS14]; the version stated here is correct.

Lemma 6.2. *For all $X > 0$, we have*

$$\sum_{0\le |n|\le X} |\lambda_{\sigma,N}(n)|^2 \ll_\epsilon (NX)^\epsilon (X + M\sqrt{X}).$$

Proof. The proof is rather involved. Write $n = n_0 n_1$ where $n_1 := (n, N^\infty)$. Let us first show that[8]

$$|\lambda_{\sigma,N}(n_0 n_1)| = |\lambda_{\sigma,N}(n_0' n_1)| \quad \text{if } (n_0, N) = (n_0', N) = 1,\ n_0 \equiv n_0' \bmod M. \tag{13}$$

Indeed, to prove (13), it suffices to show that for each $p \mid N$, $r \in \mathbb{Z}$, and $u_i \in \mathbb{Z}_p^\times$ with $u_1 \equiv u_2 \bmod p^{m_p}$, one has

$$\left| W_p\left(\begin{pmatrix} u_1 p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -ap^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \right) \right| \\ = \left| W_p\left(\begin{pmatrix} u_2 p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -ap^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \right) \right|.$$

Set

$$\nu = \begin{pmatrix} 1 & p^{r-m_p+n_p}(u_1-u_2)a^{-1} \\ 0 & 1 \end{pmatrix}, \qquad k = \begin{pmatrix} u_1/u_2 & 0 \\ p^{-m_p+n_p}(u_2-u_1)a^{-1} & u_2/u_1 \end{pmatrix}.$$

Note that $k \in K_0(p^{n_p})$. We can check that

$$\nu \begin{pmatrix} p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -au_1^{-1}p^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \\ = \begin{pmatrix} p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -au_2^{-1}p^{m_p-n_p} \\ 0 & 1 \end{pmatrix} k.$$

Now (13) follows from the following calculation:

$$\begin{aligned} &W_p\left(\begin{pmatrix} u_1 p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -ap^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \right) \\ &\quad = W_p\left(\begin{pmatrix} p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -au_1^{-1}p^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \right) \\ &\quad = \epsilon W_p\left(\nu \begin{pmatrix} p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -au_1^{-1}p^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \right) \\ &\quad = \epsilon W_p\left(\begin{pmatrix} p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -au_2^{-1}p^{m_p-n_p} \\ 0 & 1 \end{pmatrix} k \right) \\ &\quad = \epsilon W_p\left(\begin{pmatrix} u_2 p^r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -ap^{m_p-n_p} \\ 0 & 1 \end{pmatrix} \right) \end{aligned}$$

where $\epsilon = \psi_p(p^{r-m_p+n_p}(u_1-u_2)a^{-1}) \in S^1$.

8 This fact was implicitly proved in [NPS14] but we give a proof here for completeness.

So, for each $n_1 \mid N^\infty$, we can define the quantity

$$\lambda_{[N/M],N}(n_1) := \left(\frac{1}{\phi(M)} \sum_{\substack{n_0 \bmod M \\ (n_0,M)=(n_0,N)=1}} |\lambda_{\sigma,N}(n_0 n_1)|^2\right)^{1/2},$$

where the sum is taken over any set of integers n_0 which form a reduced residue system modulo M and each n_0 is coprime to N (e.g., if $M = 5$, $N = 50$, we can sum over the elements 1, 3, 7, 9).

Next, note that

$$\begin{aligned}
\lambda_{[N/M],N}(n_1)^2 &= \frac{1}{\phi(M)} \sum_{\substack{n_0 \bmod M \\ (n_0,M)=(n_0,N)=1}} |\lambda_{\sigma,N}(n_0 n_1)|^2 = \frac{1}{\phi(N)} \sum_{\substack{n_0 \bmod N \\ (n_0,N)=1}} |\lambda_{\sigma,N}(n_0 n_1)|^2 \\
&= \frac{n_1^{1/2}}{\phi(N)} \sum_{\substack{n_0 \bmod N \\ (n_0,N)=1}} \prod_{p|N} \left| W_p\left(\begin{pmatrix} n_0 n_1 M^2/N^2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & -aM/N \\ 0 & 1 \end{pmatrix}\right)\right|^2 \\
&= \frac{n_1^{1/2}}{\phi(N)} \sum_{\substack{n_0 \bmod N \\ (n_0,N)=1}} \prod_{p|N} \left| W_p\left(\begin{pmatrix} -a^{-1} n_0 n_1 M^2/N^2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & M/N \\ 0 & 1 \end{pmatrix}\right)\right|^2 \\
&= \prod_{p|N} (n_1, p^\infty) \int_{n_0 \in \mathbb{Z}_p^\times} \left| W_p\left(\begin{pmatrix} n_0 n_1 M^2/N^2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & M/N \\ 0 & 1 \end{pmatrix}\right)\right|^2
\end{aligned}$$

where the last step follows from the invariance properties of W_p and an application of the Chinese Remainder Theorem. This shows that

$$\lambda_{[N/M],N}(n_1) = \prod_{p|N} \lambda_{[N/M],p}(n_1),$$

where for each integer n_1, each prime $p \mid N$, and each integer c such that $c \mid N \mid c^2$,

$$\lambda_{[c],p}(n_1) := (n_1, p^\infty)^{1/2} \left(\int_{\mathbb{Z}_p^\times} \left| W_p\left(\begin{pmatrix} u n_1/c^2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1/c \\ 0 & 1 \end{pmatrix}\right)\right|^2 du \right)^{1/2}.$$

These local functions $\lambda_{[c],p}(n_1)$ were studied in detail in [NPS14], and completely explicit (and remarkably simple) expressions for them were proved. For the purposes of this lemma, we only need the following bound, which follows from [NPS14, Prop. 3.12(b)] and [NPS14, Cor. 3.13]:

$$\lambda_{[N/M],p}(n_1) \ll (n_1, p^\infty)^{1/4}.$$

This gives $\lambda_{[N/M],N}(n_1) \ll n_1^{1/4}$, which implies that for each $n_1 \mid N^\infty$,

$$\sum_{\substack{n_0 \bmod M \\ (n_0,M)=(n_0,N)=1}} |\lambda_{\sigma,N}(n_0 n_1)|^2 \ll M n_1^{1/2}.$$

In particular, if Y is any positive *integer*, we deduce (by completing the residue classes) that

$$\sum_{\substack{1\le|n_0|\le MY\\(n_0,N)=1}} |\lambda_{\sigma,N}(n_0n_1)|^2 \ll YMn_1^{1/2}. \tag{14}$$

Hence

$$\begin{aligned}
\sum_{1\le|n|\le X}|\lambda_{\sigma,N}(n)|^2 &= \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}\ \sum_{\substack{1\le|n_0|\le X/n_1\\(n_0,N)=1}}|\lambda_{\sigma,N}(n_0n_1)|^2\\
&\le \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}\ \sum_{\substack{1\le|n_0|\le M\lceil X/(Mn_1)\rceil\\(n_0,N)=1}}|\lambda_{\sigma,N}(n_0n_1)|^2\\
&\ll \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}\left\lceil\frac{X}{Mn_1}\right\rceil Mn_1^{1/2}\qquad\text{(by (14))}\\
&\le \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}\left(\frac{X}{Mn_1}+1\right)Mn_1^{1/2}\le \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}(X+M\sqrt{X})\\
&\ll_\epsilon (NX)^\epsilon(X+M\sqrt{X}),
\end{aligned}$$

as required. In the last step above, we have used the fact that there are $\ll_\epsilon (NX)^\epsilon$ integers n_1 satisfying $1\le n_1\le X$ and $n_1 \mid N^\infty$. □

Remark 6.3. It is possible that the error term $M\sqrt{X}$ might be sharpened with more delicate analysis. However, this will not lead to any improvement in our main theorem, and so we do not attempt to do it.

Lemma 6.4. *Suppose $y\ge 1/N$. We have the bound*

$$\sum_{1\le n\le X}\left|\lambda_f\left(\frac{|n|}{(|n|,N^\infty)}\right)K_{ir}(2\pi|n|y)\right|^2 \ll_{R,\epsilon} X^{1-2\operatorname{Im}(r)}y^{-2\operatorname{Im}(r)}(NX)^\epsilon.$$

Proof. Recall that $\operatorname{Im}(r)\in[0,1/2]$. Using the well-known bound $K_{ir}(u)\ll u^{-\operatorname{Im}(r)-\epsilon}$ for $u>0$, we see that it suffices to prove that

$$\sum_{1\le n\le X}\left|\lambda_f\left(\frac{|n|}{(|n|,N^\infty)}\right)\right|^2|n|^{-2\operatorname{Im}(r)} \ll_{R,\epsilon} X^{1-2\operatorname{Im}(r)}(NX)^\epsilon. \tag{15}$$

The left side of (15) equals

$$\begin{aligned}
\sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}\ \sum_{\substack{1\le n_0\le X/n_1\\(n_0,N)=1}}\lambda_f(|n_0|)^2(n_0n_1)^{-2\operatorname{Im}(r)} &\le \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}}\ \sum_{1\le n_0\le X/n_1}\lambda_f(|n_0|)^2(n_0n_1)^{-2\operatorname{Im}(r)}\\
&\ll_\epsilon \sum_{\substack{1\le n_1\le X\\ n_1|N^\infty}} X^{1-2\operatorname{Im}(r)}(NRX)^\epsilon \ll_\epsilon X^{1-2\operatorname{Im}(r)}(NRX)^\epsilon
\end{aligned}$$

where we have used the bound

$$\sum_{1\le n\le X} |\lambda_f(n)|^2 |n|^{-2\,\mathrm{Im}(r)} \ll_\epsilon X^{1-2\,\mathrm{Im}(r)}(NRX)^\epsilon, \tag{16}$$

which follows from the analytic properties of the Rankin–Selberg L-function (e.g., combine [HM06, (2.28)] with the usual partial summation). □

We can now prove Proposition 6.1. Recall the Fourier expansion

$$h(z) = y^{1/2} \sum_{n\neq 0} \rho_\sigma(n) K_{ir}(2\pi|n|y) e(nx).$$

The tail of the sum, with $|n|y > N^\epsilon$, is negligible because of the decay of the Bessel function. For the remaining terms, we apply the Cauchy–Schwarz inequality:

$$|h(z)|^2$$
$$\ll_{R,\epsilon} N^\epsilon y|\rho(1)|^2 \Bigg(\sum_{1\le|n|\le N^\epsilon/y} |\lambda_{\sigma,N}(n)|^2\Bigg)\Bigg(\sum_{1\le|n|\le N^\epsilon/y}\left|\lambda_f\left(\frac{|n|}{(|n|,N^\infty)}\right)K_{ir}(2\pi|n|y)\right|^2\Bigg)$$
$$\ll_{R,\epsilon} N^\epsilon\left(\frac{1}{Ny} + \frac{M}{N\sqrt{y}}\right)$$

where we have used Lemma 6.2, Lemma 6.4 and the estimate $|\rho(1)|^2 \ll N^{\epsilon-1}$ due to Hoffstein–Lockhart [HL94].

7. Proof of the main result

Recall that f is a Hecke–Maass cuspidal newform for the group $\Gamma_0(N)$ with Laplace eigenvalue $\lambda = 1/4 + r^2$ such that $\langle f, f\rangle_{\Gamma_0(N)} = 1$ and $|r| \le R$.

We first deal with the question of proving

$$\|f\|_\infty \ll_{R,\epsilon} N^{-1/12+\epsilon}.$$

Let $z \in \mathbb{H}$. We need to prove that $|f(z)| \ll_{R,\epsilon} N^{-1/12+\epsilon}$. Let M, W, σ be as in Proposition 3.1 and set $x' + iy' = z' := \sigma^{-1}Wz$. Let $g := f|\sigma$. Then, as $f|W = \pm f$, it follows that $|g(z')| = |f(z)|$. So it suffices to prove that $|g(z')| \ll_{R,\epsilon} N^{-1/12+\epsilon}$.

We first consider the case $M \ge N^{1/12}$. Then by Proposition 3.1 we have $y' \gg M^2/N \ge N^{-5/6}$. Using Proposition 6.1, we conclude that

$$g(z') \ll_{R,\epsilon} N^\epsilon \max\left(\frac{1}{(N\cdot N^{-5/6})^{1/2}}, \frac{M^{1/2}}{N^{1/2}(M^2/N)^{1/4}}\right) \ll_{R,\epsilon} N^{-1/12+\epsilon}.$$

So we may henceforth assume that $M < N^{1/12}$. Furthermore, we may assume that $y' < N^{-5/6}$, for otherwise Proposition 6.1 finishes the job again. For future reference, we record this as follows:

$$1 \le M \ll N^{1/12}, \qquad M^2/N \ll y' \ll N^{-5/6}. \tag{17}$$

Set $\Gamma = \Gamma_0(N; M)$. We note that g is a Maass cusp form on Γ that satisfies

$$M^{1-\epsilon} \ll_\epsilon \phi(M)/(M, 2) = \langle g, g\rangle_\Gamma \le M.$$

Let $g' = g/\langle g, g\rangle_\Gamma^{1/2}$. Then $\langle g', g'\rangle_\Gamma = 1$. It suffices to show that

$$|g'(z')|^2 \ll_{R,\epsilon} M^{-1}N^{-1/6+\epsilon}. \tag{18}$$

By Corollary 5.2, g' satisfies, for all $l \equiv 1 \bmod M$,

$$\sum_{\gamma\in\Gamma_0(N;M)\backslash\Delta(l,N;M)} g'|\gamma = \lambda_f(l)l^{1/2}g.$$

Define

$$\mathcal{P} := \{p \text{ prime} : p \equiv 1 \bmod M,\ \Lambda < p < 2\Lambda\}, \quad \mathcal{P}^2 := \{p^2 : p \in \mathcal{P}\},$$
$$x_l := \begin{cases} \operatorname{sgn}(\lambda_f(l)), & l \in \mathcal{P}\cup\mathcal{P}^2,\\ 0 & \text{otherwise.}\end{cases}$$

By embedding the cusp form g' into an orthonormal basis of Maass cusp forms for Γ and then using the amplifier method as in [HT12] (with the amplifier x_l defined above), we obtain the inequality

$$\frac{\Lambda^2}{M^2}|g'(z')|^2 \ll_{R,\epsilon} (N\Lambda)^\epsilon \sum_{l\ge1}\frac{y_l}{\sqrt{l}}N(z, l, N; M),$$

where y_l is defined as in Proposition 4.6. Using (10) and (18), we conclude that it suffices to prove that for some $\Lambda \ge M^2$,

$$\frac{M^2}{\Lambda} + y'N_0 + \frac{M\Lambda^{1/2}}{\sqrt{N}} + \frac{M^2\Lambda^2}{N} \ll N^{-1/6}.$$

Choosing $\Lambda = N^{1/3}$, and using (17) and $N_0 \le N^{1/2}$, yields the above inequality.

Next, we suppose that there is no integer M' in the range $1 < M' < N^{1/6}$ such that M'^2 divides N. We need to prove that

$$\|f\|_\infty \ll_{R,\epsilon} N^\epsilon \max(N^{-1/6}, N^{-1/4}N_0^{1/4}).$$

Let $z \in \mathbb{H}$. We need to show that $|f(z)| \ll_{R,\epsilon} N^\epsilon \max(N^{-1/6}, N^{-1/4}N_0^{1/4})$. Let M, W, σ be as in Proposition 3.1 and set $x' + iy' = z' := \sigma^{-1}Wz$. Let $g := f|\sigma$. Then, as $f|W = \pm f$, it follows that $|g(z')| = |f(z)|$. So it suffices to prove that $|g(z')| \ll_{R,\epsilon} N^\epsilon \max(N^{-1/6}, N^{-1/4}N_0^{1/4})$.

As before, we can reduce to the case $M < N^{1/6}$ using Proposition 6.1. *By our assumption, it follows that* $M = 1$. Furthermore, we may assume that $y' \le 1/\sqrt{NN_0}$, for otherwise Proposition 6.1 finishes the job again. By proceeding exactly as before, the amplification method reduces our task to proving that

$$\frac{1}{\Lambda} + y' N_0 + \frac{\Lambda^{1/2}}{\sqrt{N}} + \frac{\Lambda^2}{N} \ll_\epsilon N^\epsilon \max(N^{-1/3}, N^{-1/2} N_0^{1/2}).$$

Choosing $\Lambda = N^{1/3}$ and using $y' \leq 1/\sqrt{N N_0}$ yields the above inequality. The proof is complete.

Acknowledgments. I would like to thank Martin Dickson, Gergely Harcos, Emmanuel Kowalski, Paul Nelson and Guillaume Ricotta for helpful comments.

References

[AU95] Abbes, A., Ullmo, E.: Comparaison des métriques d'Arakelov et de Poincaré sur $X_0(N)$. Duke Math. J. **80**, 295–307 (1995) Zbl 0895.14007 MR 1369394

[AL70] Atkin, A. O. L., Lehner, J.: Hecke operators on $\Gamma_0(m)$. Math. Ann. **185**, 134–160 (1970) Zbl 0177.34901 MR 0268123

[BHM16] Blomer, V., Harcos, G., Milićević, D.: Bounds for eigenforms on arithmetic hyperbolic 3-manifolds. Duke Math. J. **165**, 625–659 (2016) Zbl 1339.11062 MR 3474814

[BH10] Blomer, V., Holowinsky, R.: Bounding sup-norms of cusp forms of large level. Invent. Math. **179**, 645–681 (2010) Zbl 1243.11059 MR 2587342

[BM15] Blomer, V., Maga, P.: The sup-norm problem for PGL(4). Int. Math. Res. Notices **2015**, 5311–5332 Zbl 06513139 MR 3384442

[BM13] Blomer, V., Michel, P.: Hybrid bounds for automorphic forms on ellipsoids over number fields. J. Inst. Math. Jussieu **12**, 727–758 (2013) Zbl 1355.11062 MR 3103131

[BP16] Blomer, V., Pohl, A.: The sup-norm problem on the Siegel modular space of rank two. Amer. J. Math. **138**, 999–1027 (2016) Zbl 1350.11061 MR 3538149

[BT17] Brumley, F., Templier, N.: Large values of cusp forms on GL_n. arXiv:1411.4317v3 (2017)

[DS15] Das, S., Sengupta, J.: L^∞ norms of holomorphic modular forms in the case of compact quotient. Forum Math. **27**, 1987–2001 (2015) Zbl 06458904 MR 3365786

[Don01] Donnelly, H.: Bounds for eigenfunctions of the Laplacian on compact Riemannian manifolds. J. Funct. Anal. **187**, 247–261 (2001) Zbl 0991.58006 MR 1867351

[HM06] Harcos, G., Michel, P.: The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points. II. Invent. Math. **163**, 581–655 (2006) Zbl 1111.11027 MR 2207235

[HT12] Harcos, G., Templier, N.: On the sup-norm of Maass cusp forms of large level: II. Int. Math. Res. Notices **2012**, 4764–4774 Zbl 1332.11048 MR 2989618

[HT13] Harcos, G., Templier, N.: On the sup-norm of Maass cusp forms of large level. III. Math. Ann. **356**, 209–216 (2013) Zbl 1332.11049 MR 3038127

[HL94] Hoffstein, J., Lockhart, P.: Coefficients of Maass forms and the Siegel zero (with an appendix by D. Goldfeld, J. Hoffstein and D. Lieman). Ann. of Math. (2) **140**, 161–181 (1994) Zbl 0814.11032 MR 1289494

[HRR14] Holowinsky, R., Ricotta, G., Royer, E.: On the sup-norm of SL_3 Hecke–Maass cusp form. arXiv:1404.3622 (2014)

[IS95] Iwaniec, H., Sarnak, P.: L^∞ norms of eigenfunctions of arithmetic surfaces. Ann. of Math. (2) **141**, 301–320 (1995) Zbl 0833.11019 MR 1324136

[JK04] Jorgenson, J., Kramer, J.: Bounding the sup-norm of automorphic forms. Geom. Funct. Anal. **14**, 1267–1277 (2004) Zbl 1078.11027 MR 2135167

[Kir14] Kiral, E. M.: Bounds on sup-norms of half-integral weight modular forms. Acta Arith. **165**, 385–400 (2014) Zbl 1329.11032 MR 3268710

[Koy95] Koyama, S.: L^∞-norms on eigenfunctions for arithmetic hyperbolic 3-manifolds. Duke Math. J. **77**, 799–817 (1995) Zbl 0869.11050 MR 1324641

[Lau10] Lau, Y.-K.: An omega result for supremum norms of Hecke-eigenforms in the level aspect. Sci. China Math. **53**, 223–242 (2010) Zbl 1269.11041 MR 2594762

[Miy06] Miyake, T.: Modular Forms. Springer Monogr. Math., Springer, Berlin (2006) Zbl 1159.11014 MR 2194815

[NPS14] Nelson, P. D., Pitale, A., Saha, A.: Bounds for Rankin–Selberg integrals and quantum unique ergodicity for powerful levels. J. Amer. Math. Soc. **27**, 147–191 (2014) Zbl 1322.11051 MR 3110797

[Rud05] Rudnick, Z.: On the asymptotic distribution of zeros of modular forms. Int. Math. Res. Notices **2005**, 2059–2074 Zbl 1162.11345 MR 2181743

[Sah16] Saha, A.: Large values of newforms on GL(2) with highly ramified central character. Int. Math. Res. Notices **2016**, 4103–4131 MR 3544630

[Sah17] Saha, A.: Hybrid sup-norm bounds for Maass newforms of powerful level. Algebra Number Theory **11**, 1009–1045 (2017) Zbl 06748165 MR 3671430

[Sch68] Schmidt, W. M.: Asymptotic formulae for point lattices of bounded determinant and subspaces of bounded height. Duke Math. J. **35**, 327–339 (1968) Zbl 0172.06304 MR 0224562

[Shi71] Shimura, G.: Introduction to the Arithmetic Theory of Automorphic Functions. Publ. Math. Soc. Japan 11, Iwanami Shoten, Tokyo (1971) Zbl 0872.11023 MR 1291394

[Tem10] Templier, N.: On the sup-norm of Maass cusp forms of large level. Selecta Math. (N.S.) **16**, 501–531 (2010) Zbl 1260.11032 MR 2734340

[Tem14] Templier, N.: Large values of modular forms. Cambridge J. Math. **2**, 91–116 (2014) Zbl 1307.11062 MR 3272013

[Tem15] Templier, N.: Hybrid sup-norm bounds for Hecke–Maass cusp forms. J. Eur. Math. Soc. **17**, 2069–2082 (2015) Zbl 06481085 MR 3372076

[Van97] VanderKam, J. M.: L^∞ norms and quantum ergodicity on the sphere. Int. Math. Res. Notices **1997**, 329–347 Zbl 0877.58056 MR 1440572

[Xia07] Xia, H.: On L^∞ norms of holomorphic cusp forms. J. Number Theory **124**, 325–327 (2007) Zbl 1121.11038 MR 2321365