A NEW BOUND FOR THE SUP NORM OF AUTOMORPHIC FORMS ON NON-COMPACT MODULAR CURVES IN THE LEVEL ASPECT

HARALD HELFGOTT AND GUILLAUME RICOTTA

ABSTRACT. We find a new bound for $||f||_\infty$, where f is a Hecke-Maaß cusp newform (normalised by $||f||_2 = 1$) for the congruence subgroup $\Gamma_0(N)$, $N \to +\infty$ square-free.

Our work is a refinement of [BH10] and especially [Tem10]. The main innovation is a much sharper counting lemma, stating that, under certain broad conditions, the number of images of $z \in \mathbb{H}$ lying close to z under the action of

$$\mathsf{M}(\ell, N) := \left\{\rho = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}), ad - bc = \ell, N \mid c\right\}.$$

is bounded by N^ε for every individual positive relatively small integer ℓ and for all $\varepsilon > 0$. The main ideas involved are diophantine ones. As a result, we can bound a twisted second moment of newforms for each ℓ, leading us to our improved result.

CONTENTS

1. INTRODUCTION AND STATEMENT OF THE RESULT

1.1. **General background.** The correspondence principle in quantum mechanics suggests a way to study a classical system via its semi-classical limit of quantization. For instance, let X be a compact Riemannian manifold. We can choose an orthonormal basis $(f_j)_{j\geqslant 0}$ of $L^2(X)$ satisfying

$$\forall j \geqslant 0, \quad \Delta(f_j) = \lambda_j f_j.$$

where Δ is the Laplace-Beltrami operator on X and $0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \dots$ is its spectrum. If G^t is the geodesic flow on X then its quantization is $-h^2\Delta$, where h is Planck's constant. Thus, it is very natural to attempt to understand the asymptotic behaviour of the eigenfunctions of Δ.

Date: Version of July 17, 2011.

A classical question here – suggested by the correspondence principle – is to bound $||f_j||_\infty$ as $\lambda_j \to \infty$. (See [NTY01] and [Sar95] for more details.) A. Seeger and C. Sogge proved in [SS89] a very general and abstract bound, essentially sharp, in the case of compact Riemannian surfaces.

We will focus on arithmetic surfaces, which are the quotient of the upper-half plane by a congruence subgroup of $SL_2(\mathbb{Z})$. (Such surfaces can be compact or non-compact.) The Laplace-Beltrami operator in this context is the hyperbolic Laplacian. In [IS95], H. Iwaniec and P. Sarnak proved a bound sharper than that of Seeger-Sogge for these surfaces – both in the compact and in the non-compact case; they took advantage of the fact that some additionnal symmetries, the Hecke correspondences, act on these surfaces. S. Koyama investigated the case of quotients of the three-dimensional hyperbolic space by arithmetic subgroups in [Koy95] and proved similar results.

1.2. **Bounds for varying surfaces. Main result.** There is a new direction in the asymptotic study of eigenforms on arithmetic surfaces, in that there are now non-trivial bounds for $|f|_\infty$ as the *surface* changes and the eigenvalues remain bounded (or grow slowly).

V. Blomer and R. Holowinsky [BH10] were the first to prove a (remarkable and difficult) bound for the norm $||f||_\infty$ of non-exceptional Hecke-Maaß eigenforms f on the modular curve of square-free level N. Their bound is $||f||_\infty \ll_T N^{-1/37}$ for forms f of eigenvalue $\lambda \leqslant T$. Note that the trivial bound for $||f||_\infty$ is given by $||f||_\infty \ll_{T,\varepsilon} N^\varepsilon$ for all $\varepsilon > 0$. This follows from very different ideas (see [AU95], [MU98], [JK09] and [JK04]). There is no real evidence for what could be the optimal bound for $||f||_\infty$.

The proof in [BH10] involves many technicalities and relies deeply on the spectral theory of automorphic forms. More recently, N. Templier [Tem10] refined the proof substantially, using geometric arguments instead of very delicate analytical estimates. As a result, [Tem10] gives stronger bounds – both in the non-compact case studied in [BH10] and in the compact case. (The compact case, while in general easier due to the absence of cusps, involves non-trivial manipulations of quaternion algebras.) Furthermore, [Tem10] removes the assumption that the forms studied are non-exceptional.

The bounds given by [Tem10] are better in the compact case ($||f||_\infty \ll_T N^{-1/12}$) than in the non-compact case ($||f||_\infty \ll_T N^{-1/23}$). This suggested to us that the geometric and diophantine arguments in [Tem10] were less than optimal.

Let

$$\mathsf{M}(\ell, N) := \left\{ \rho = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{Z}), ad - bc = \ell, N \mid c \right\} \tag{1.1}$$

where $M_2(\mathbb{Z})$ is the set of 2×2 matrices with integer coefficients. N. Templier shows that, if $z \in \mathbb{H}$ is far from the cusps, there are – on average as ℓ varies – few matrices $\rho \in \mathsf{M}(\ell, N)$ such that $\rho.z$ lies near z. We show that a sharper result holds: the number of such matrices ρ is bounded above by a constant for each individual ℓ. In fact (for c in a dyadic interval $C \leqslant c \leqslant 2C$, and under diophantine conditions slightly stronger than those used by N. Templier) the matrices ρ turn out to be all of the form $\lambda A + \delta I$, where A is a fixed matrix (the same one for *all* ℓ), λ and δ are (small) scalars and I is the identity matrix.

We also derive sharper diophantine conditions than those derived in [Tem10] for points away from the cusps.

As a result, we obtain the following bound.

Theorem A– *If f is a L^2-normalised Hecke-Maaß cuspidal newform of square-free level N and bounded Laplace eigenvalue $\lambda \leqslant T$ then*

$$||f||_\infty \ll_{\varepsilon,T} N^{-1/20+\varepsilon}$$

for all $\varepsilon > 0$.

The reader may have noticed that the level N is assumed to be square-free, in which case all the cusps of $\Gamma_0(N)$ belong to the same orbit under the action of the Atkin-Lehner operators. We should say that our work does not shed any new light on how one can remove this geometric assumption.

Acknowledgements. Work towards this paper began while both authors were at the Centre Interfacultaire Bernoulli of the EPFL (Lausanne) on occasion of the GANT special semester. We would like to thank this institution for the hospitable work environment it offered and the organisers of this event, E. Kowalski and P. Michel, for their kind invitation. We are also grateful to E. Royer for conversations on related problems.

The paper was completed while the first author visited Université Bordeaux 1. He thanks the mathematics department at Bordeaux, and in particular K. Belabas, for the invitation and for good working conditions.

2. NOTATION AND PARAMETERS

As is usual, instead of working directly with the hyperbolic distance on the upper half plane $\mathbb{H} := \{z = x + iy, x \in \mathbb{R}, y > 0\}$, we will work with the function

$$u(z_1, z_2) = \frac{|z_1 - z_2|^2}{\Im m\,(z_1)\,\Im m\,(z_2)} \tag{2.1}$$

for z_1 and z_2 in $\mathbb{H}$. (Note that $u = 4\cosh(d) - 1$), where d is the hyperbolic distance on $\mathbb{H}$.)

Our main parameter will be a positive integer N. We will work with the sets of matrices $\mathsf{M}(\ell, N) \subset M_2(\mathbb{Z})$ defined in (1.1). As in [Tem10, Section 2.2], let $A_0(N)$ be the subgroup of $SL_2(\mathbb{R})$ generated by Atkin-Lehner operators.

Define the *Siegel set*

$$\sigma_v := \left\{ z = x + iy, 0 \leqslant x < 1, y \geqslant \frac{\sqrt{3}}{2v} \right\} \subset \mathbb{H}.$$

for $v > 0$. We will use a parameter η, set to be a negative power of N, to control the position of a point z with relation to the cusps: we will think of z as being near the cusp at ∞ if $z \in \sigma_{\eta N}$.

Given any $x \in \mathbb{R}$, there exist two coprime integers e and $1 \leqslant q \leqslant H$ such that

$$\left| x - \frac{e}{q} \right| \leqslant \frac{1}{qH} \qquad \text{(Dirichlet approximation).} \tag{2.2}$$

We will set H equal to a positive power of N, and $Q \leqslant H$ equal to a (smaller) positive power of N. N. Templier calls $x \in \mathbb{R}$ *well approximable* if it satisfies (2.2) for some $q \leqslant Q$, and *poorly approximable* otherwise.

3. CUSPS AND DIOPHANTINE CONDITIONS

N. Templier proved in [Tem10, Lemma 2.2] that the well-approximable points are those that lie high in the cusps. We prove a slightly better version of his lemma.

Lemma 3.1—*Assume that* $H^2 \geqslant \frac{2N}{\eta}$ *and that* N *is square-free. If* $z = x + iy$ *belongs to* $\sigma_N \backslash \cup_{\delta \in A_0(N)} \delta.\sigma_{\eta N}$ *then any approximation* e/q *of* x *in the sense of* (2.2) *satisfies* $q \geqslant \frac{\sqrt{2}\eta\sqrt{N}}{\sqrt{3}}$.

Proof of Lemma 3.1. By [Tem10, Lemma 2.2] and the fact that N is square-free, there exist $b, d \in \mathbb{Z}$, a positive integer $M \mid N$ and a matrix $\gamma \in \Gamma_0(N)$ such that

$$\begin{pmatrix} e & b \\ q & d \end{pmatrix} = \gamma W_M \begin{pmatrix} M^{-1} & 0 \\ 0 & 1 \end{pmatrix}$$

where W_M is an Atkin-Lehner matrix namely an element of $M_2(\mathbb{Z})$ of determinant M satisfying

$$W_M \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod N, \quad W_M \equiv \begin{pmatrix} 0 & * \\ 0 & 0 \end{pmatrix} \pmod M.$$

Note that

$$\left(\gamma W_M\right)^{-1}.z = \begin{pmatrix} M^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d & -b \\ -q & e \end{pmatrix}.z \text{ does not belong to } \begin{pmatrix} 1 & \mathbb{Z} \\ 0 & 1 \end{pmatrix}.\sigma_{\eta N}$$

by the assumption on the position of z. In other words,

$$\frac{\sqrt{3}}{2\eta N} > \Im m\left(\left(\gamma W_M\right)^{-1}.z\right) = \frac{1}{M}\frac{y}{(e-qx)^2 + q^2y^2},$$

which implies

$$\frac{\sqrt{3}}{2\eta N} > \frac{1}{N}\frac{y}{H^{-2} + q^2y^2} = \frac{1}{N}\varphi(y) \geqslant \frac{1}{N}\min_{\frac{\sqrt{3}}{2N} \leqslant t < \frac{\sqrt{3}}{2\eta N}} \varphi(t) \tag{3.1}$$

where φ is the function on $\mathbb{R}_+$ defined by $\varphi(t) := \frac{t}{H^{-2}+q^2t^2}$. So far, we have proceeded as in [Tem10, Lemma 4.2]. Now, the function φ satisfies

$$\varphi'(t) = \frac{H^{-2} - q^2t^2}{(H^{-2} + q^2t^2)^2} \geqslant 0 \text{ if and only if } t \leqslant (qH)^{-1}.$$

As a consequence,

$$\min_{\frac{\sqrt{3}}{2N} \leqslant t < \frac{\sqrt{3}}{2\eta N}} \varphi(t) = \begin{cases} \varphi\left(\frac{\sqrt{3}}{2\eta N}\right) & \text{if } q \geqslant \frac{2N}{\sqrt{3}H}, \\ \min\left(\varphi\left(\frac{\sqrt{3}}{2\eta N}\right), \varphi\left(\frac{\sqrt{3}}{2N}\right)\right) & \text{if } \frac{2\eta N}{\sqrt{3}H} \leqslant q < \frac{2N}{\sqrt{3}H}, \\ \varphi\left(\frac{\sqrt{3}}{2N}\right) & \text{if } q < \frac{2\eta N}{\sqrt{3}H}. \end{cases}$$

A direct computation ensures that

$$\min\left(\varphi\left(\frac{\sqrt{3}}{2\eta N}\right), \varphi\left(\frac{\sqrt{3}}{2N}\right)\right) = \begin{cases} \varphi\left(\frac{\sqrt{3}}{2\eta N}\right) & \text{if } \frac{2\sqrt{\eta}N}{\sqrt{3}H} \leqslant q < \frac{2N}{\sqrt{3}H}, \\ \varphi\left(\frac{\sqrt{3}}{2N}\right) & \text{if } \frac{2\eta N}{\sqrt{3}H} \leqslant q < \frac{2\sqrt{\eta}N}{\sqrt{3}H} \end{cases}$$

and so

$$\min_{\frac{\sqrt{3}}{2N} \leqslant t < \frac{\sqrt{3}}{2\eta N}} \varphi(t) = \begin{cases} \varphi\left(\frac{\sqrt{3}}{2\eta N}\right) & \text{if } q \geqslant \frac{2\sqrt{\eta}N}{\sqrt{3}H}, \\ \varphi\left(\frac{\sqrt{3}}{2N}\right) & \text{if } q < \frac{2\sqrt{\eta}N}{\sqrt{3}H}. \end{cases} \tag{3.2}$$

By (3.1) and (3.2)

$$\frac{\sqrt{3}}{2\eta N} > \frac{1}{N} \times \begin{cases} \varphi\left(\frac{\sqrt{3}}{2\eta N}\right) & \text{if } q \geqslant \frac{2\sqrt{\eta}N}{\sqrt{3}H}, \\ \varphi\left(\frac{\sqrt{3}}{2N}\right) & \text{if } q < \frac{2\sqrt{\eta}N}{\sqrt{3}H}. \end{cases} \tag{3.3}$$

Clearly

$$\varphi(t) \geqslant \frac{t}{2\max\left(H^{-2}, q^2t^2\right)} = \frac{t}{2}\min\left(H^2, q^{-2}t^{-2}\right) = \frac{\sqrt{3}}{2\eta}\min\left(\frac{\eta t H^2}{\sqrt{3}}, \frac{\eta}{\sqrt{3}q^2t}\right) \geqslant \frac{\sqrt{3}}{2\eta}$$

if $H^2 \geqslant \frac{\sqrt{3}}{\eta t}$ and $q \leqslant \frac{\sqrt{\eta}}{3^{1/4}\sqrt{t}}$. In particular, if $H^2 \geqslant \frac{2N}{\eta}$ and $q \leqslant \frac{\sqrt{2\eta N}}{\sqrt{3}}$ then

$$\varphi\left(\frac{\sqrt{3}}{2N}\right) \geqslant \frac{\sqrt{3}}{2\eta} \tag{3.4}$$

whereas if $H^2 \geqslant 2N$ and $q \leqslant \frac{\sqrt{2}\eta\sqrt{N}}{\sqrt{3}}$ then

$$\varphi\left(\frac{\sqrt{3}}{2\eta N}\right) \geqslant \frac{\sqrt{3}}{2\eta}. \tag{3.5}$$

Let us assume that $H^2 \geqslant \frac{2N}{\eta}$, in which case $\frac{2\sqrt{\eta}N}{\sqrt{3}H} \leqslant \frac{\sqrt{2}\eta\sqrt{N}}{\sqrt{3}} < \frac{\sqrt{2\eta N}}{\sqrt{3}}$.

We would like to prove that $q \geqslant \frac{2\sqrt{\eta}N}{\sqrt{3}H}$. If this were not the case, the second inequality in (3.3) and (3.4) would imply $\frac{\sqrt{3}}{2\eta N} > \frac{\sqrt{3}}{2\eta N}$.

Now let us prove that $q \geqslant \frac{\sqrt{2}\eta\sqrt{N}}{\sqrt{3}}$. If this were note the case, the first inequality in (3.3) and (3.5) would imply $\frac{\sqrt{3}}{2\eta N} > \frac{\sqrt{3}}{2\eta N}$.

□

4. THE COUNTING LEMMA

The section is devoted to the proof of sharp estimates for the cardinality of

$$\mathcal{M}(\ell, N; z) := \left\{\rho \in \mathsf{M}(\ell, N), u(\rho.z, z) \leqslant N^{\varepsilon}\right\}.$$

Here, $z = x + iy$ belongs to Poincaré upper-half plane, ℓ is a positive integer and $\mathsf{M}(\ell, N)$ is as in (1.1). We can split $\mathcal{M}(\ell, N; z)$ into

$$\mathcal{M}(\ell, N; z) := \mathcal{M}_0(\ell, N; z) + \mathcal{M}_*(\ell, N; z)$$

where

$$\mathcal{M}_0(\ell, N; z) := \left\{\rho = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathsf{M}(\ell, N), u(\rho.z, z) \leqslant N^{\varepsilon}\right\}.$$

We begin by estimating the cardinality of $\mathcal{M}_0(\ell, N; z)$ in the following proposition, which is a refinement of [Tem10, Lemma 4.2].

***Proposition* 4.1**– *Let $z = x + iy$ in $\mathbb{H}$. Then*

$$\left|\left\{\rho = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(\mathbb{Z}), ad = \ell, u(\rho.z, z) \leqslant N^{\varepsilon}\right\}\right| \leqslant \tau(\ell)(1 + N^{\varepsilon/2}\sqrt{\ell}y)$$

for all $\varepsilon > 0$. In particular, if $\ell \leqslant \frac{4\eta^2N^2}{3}$ and $z \in \sigma_N \setminus \sigma_{\eta N}$ then

$$\left|\left\{\rho = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(\mathbb{Z}), ad = \ell, u(\rho.z, z) \leqslant N^{\varepsilon}\right\}\right| \leqslant \tau(\ell)(1 + N^{\varepsilon/2}).$$

Proof of Proposition 4.1. By (2.1),

$$u(\rho.z,z)=\frac{|b-(a-d)z|^2}{\ell y^2}\leqslant N^{\varepsilon}.$$

Taking the real part implies that

$$|b-(a-d)x|\leqslant N^{\varepsilon/2}\sqrt{\ell}y.$$

Thus, b belongs to an interval of length at most $N^{\varepsilon/2}\sqrt{\ell}y$. This fact, together with $ad=\ell$ implies the desired result. □

Now, we can estimate the cardinality of $\mathscr{M}_*(\ell,N;z)$.

Proposition 4.2– *Let $z\in\sigma_N\setminus\sigma_{\eta N}$, N a positive integer. Let $H,\eta,\mathscr{L},Q$ satisfy*

$$H^2\geqslant\frac{2N}{\eta},\tag{4.1}$$

$$\frac{16}{\sqrt{3}}N^{\varepsilon/2}(2N^{\varepsilon/2}+1)\mathscr{L}\leqslant Q\leqslant\frac{\sqrt{2}\eta\sqrt{N}}{\sqrt{3}},\tag{4.2}$$

$$\frac{32}{\sqrt{3}}(1+2N^{\varepsilon/2})^3\mathscr{L}\leqslant\frac{N}{H}.\tag{4.3}$$

Then, for any $C>0$, the set

$$\left\{\rho=\begin{pmatrix}a&b\\c&d\end{pmatrix}\in M_2(\mathbb{Z}),|ad-bc|\leqslant\mathscr{L},N\mid c,C\leqslant c<2C,u(\rho.z,z)\leqslant N^{\varepsilon}\right\}.$$

is a subset of

$$\left\{\lambda A+\delta I:\lambda\in\mathbb{Z},\delta\in\mathbb{Z}/2,|\lambda|\leqslant\frac{2}{\sqrt{3}}(1+2N^{\varepsilon/2})\sqrt{\mathscr{L}},|\delta|\leqslant\left(1+\frac{1}{\sqrt{3}}\right)(1+2N^{\varepsilon/2})\sqrt{\mathscr{L}}\right\}$$

for some matrix A in $M_2(\mathbb{Z})$.

Remark 4.3– This description cannot be made any tighter – if $u(A.z,z)$ is small, then $u((\lambda A+\delta I).z,z)$ is small for all $\lambda,\delta\in\mathbb{R}$ with $|\lambda|,|\delta|\ll\sqrt{L}$. This is easy to show. We get from (2.1) that

$$u(\rho.z,z)=\frac{|(az+b)-(cz+d)z|^2}{\det(\rho)y^2}\text{ if }\rho=\begin{pmatrix}a&b\\c&d\end{pmatrix}.$$

The numerator of the expression on the right is proportional to λ^2 and independent of δ; note also that $\det(\lambda A+\delta I)=\lambda^2\det(A)+\delta^2+\lambda\delta\mathrm{tr}(A)$, and that, as we will see in the proof, $u(\rho.z,z)$ small implies $\mathrm{tr}(\rho)$ small.

Proof of Proposition 4.2. We can assume without loss of generality that $c>0$ because $\rho.z=(-\rho).z$ and thus $u(z,\rho.z)=u(z,(-\rho).z)$. An easy computation (starting from (2.1)) gives us that

$$u(\rho.z,z)=\frac{|\ell-|cz+d|^2+(cz+d)(2cx+d-a)|^2}{\ell c^2y^2}.\tag{4.4}$$

Considering the imaginary part, we get

$$|2cx+d-a|\leqslant N^{\varepsilon/2}\sqrt{\ell}.\tag{4.5}$$

Now

$$\begin{aligned}|cz+d|-\sqrt{\ell} &= \frac{|cz+d|^2-\ell}{|cz+d|+\sqrt{\ell}} \\ &= \frac{|cz+d|^2-\ell-(cz+d)(2cx+d-a)}{|cz+d|+\sqrt{\ell}} + \frac{(cz+d)(2cx+d-a)}{|cz+d|+\sqrt{\ell}}\end{aligned}$$

and so

$$\begin{aligned}||cz+d|-\sqrt{\ell}| &\leqslant \frac{||cz+d|^2-\ell-(cz+d)(2cx+d-a)|}{cy} + \frac{|cz+d||2cx+d-a|}{|cz+d|} \\ &\leqslant 2N^{\varepsilon/2}\sqrt{\ell}\end{aligned}$$

by (4.4) and (4.5). Thus

$$|cz+d| \leqslant (1+2N^{\varepsilon/2})\sqrt{\ell} \tag{4.6}$$

and, considering the imaginary part once again, we get

$$cy \leqslant (1+2N^{\varepsilon/2})\sqrt{\ell}. \tag{4.7}$$

Equation (4.6) applied with $\ell\rho^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ instead of ρ gives

$$|-cz+a| \leqslant (1+2N^{\varepsilon/2})\sqrt{\ell}; \tag{4.8}$$

this is legitimate because

$$u(\rho.z,z) = u(z,\rho^{-1}.z) = u(z,(\ell\rho^{-1}).z) = u((\ell\rho^{-1}).z,z).$$

Hence, by (4.6) and (4.8),

$$|a+d| \leqslant 2(1+2N^{\varepsilon/2})\sqrt{\ell} \tag{4.9}$$

since $a+d = -cz+a+cz+d$. Setting $s = a-d$ and $t = a+d$, we are reduced to counting the number of quadruples of integers (s,t,b,c) satisfying

$$\begin{cases} |s-2cx| \leqslant N^{\varepsilon/2}\sqrt{\ell} \\ N \leqslant c = Nc' \leqslant (1+2N^{\varepsilon/2})y^{-1}\sqrt{\ell} \\ |t| \leqslant 2(1+2N^{\varepsilon/2})\sqrt{\ell} \\ s^2 = t^2-4\ell-4bc. \end{cases} \tag{4.10}$$

according to (4.5), (4.7) and (4.9). Note also that $y^{-1} \leqslant (2/\sqrt{3})N$ (because $z \in \sigma_N$) and so $c' \leqslant (2/\sqrt{3})(1+2N^{\varepsilon/2})\sqrt{l}$.

The last equation in (4.10) implies immediately that

$$s^2 \equiv t^2-4\ell \pmod{Nc'}. \tag{4.11}$$

(So far, we have proceeded as in [Tem10, p. 520], [BH10, pp. 675–676] and [IS95, pp. 317–318].)

By the first line of (4.10), we can write

$$s = 2Nc'x+r, \quad |r| \leqslant N^{\varepsilon/2}\sqrt{\ell}. \tag{4.12}$$

Note that r is not in general an integer. Equations (4.11) and (4.12) entail

$$4(Nc'x^2+rx) \equiv \frac{t^2-4\ell-r^2}{Nc'} \pmod 1. \tag{4.13}$$

Let $\rho_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ Nc_1' & d_1 \end{pmatrix}$ in $\mathcal{M}(\ell_1, N; z)$, and $\rho_2 = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} = \begin{pmatrix} a_2 & b_2 \\ Nc_2' & d_2 \end{pmatrix}$ in $\mathcal{M}(\ell_2, N; z)$ be two matrices with $c_1, c_2 \geqslant 1$, $\ell_1, \ell_2 \leqslant \mathcal{L}$. One can define as previously s_i, t_i, r_i for $i = 1, 2$. In particular,

$$\begin{aligned} 4(Nc_1'x^2 + r_1 x) &\equiv \frac{t_1^2 - 4\ell_1 - r_1^2}{Nc_1'} \pmod 1, \\ 4(Nc_2'x^2 + r_2 x) &\equiv \frac{t_2^2 - 4\ell_2 - r_2^2}{Nc_2'} \pmod 1 \end{aligned}$$

according to (4.13). Multiplying the first congruence by c_2' and the second one by c_1' and substracting, one gets

$$4(c_2' r_1 - c_1' r_2)x \equiv \frac{c_2'(t_1^2 - 4\ell_1 - r_1^2)}{Nc_1'} - \frac{c_1'(t_2^2 - 4\ell_2 - r_2^2)}{Nc_2'} \pmod 1.$$

Note that according to (4.12)

$$c_2' r_1 - c_1' r_2 = c_2'(s_1 - 2Nc_1'x) - c_1'(s_2 - 2Nc_2'x) = c_2's_1 - c_1's_2 \in \mathbb{Z}.$$

Thus $q' := 4(c_2'r_1 - c_1'r_2)$ is an integer and

$$q'x = e' + w'$$

for some integer e', where

$$w' = \frac{c_2'(t_1^2 - 4\ell_1 - r_1^2)}{Nc_1'} - \frac{c_1'(t_2^2 - 4\ell_2 - r_2^2)}{Nc_2'} \tag{4.14}$$

is a real number.

Let us prove that $q' = 0$. If this were not the case, we would get

$$x = \frac{e}{q} + \frac{w}{q}$$

where $q = q'/(q', e')$, $e = e'/(q', e')$ and $w = w'/(q', e')$. Note that e and q have been made coprime and that

$$1 \leqslant |q| \leqslant \frac{16}{\sqrt{3}} N^{\varepsilon/2}(2N^{\varepsilon/2} + 1)\sqrt{\ell_1 \ell_2} \leqslant Q.$$

according to (4.2). In addition,

$$|w| \leqslant \frac{\frac{32}{\sqrt{3}}(1 + 2N^{\varepsilon/2})^3 \max(\ell_1, \ell_2)}{N} \leqslant \frac{1}{H}$$

by (4.3), (4.10), (4.12) and (4.14). By [Tem10, Lemma 2.2] and equations (4.1), (4.2), this contradicts the assumption that $z \in \mathbb{H} \setminus \cup_{\delta \in A_0(N)} \delta.\sigma_{\eta N}$. We conclude that $q' = 0$.

Since $4(c_2'r_1 - c_1'r_2) = q' = 0$, we see that (c_1, r_1) and (c_2, r_2) are proportional to each other. Thus $(c_1, s_1) = (c_1, 2c_1 x + r_1)$ and $(c_2, s_2) = (c_2, 2c_2 x + r_2)$ are proportional to each other.

Let $(c_1, s_1), (c_2, s_2), (c_3, s_3), \ldots \in \mathbb{Z}^2$ be all the pairs coming from solutions to (4.10); by what we have just shown, these pairs are all proportional to each other. Let c_0 be the greatest common divisor of all values of c_i. Then there is an integer s_0 and integers λ_i such that $(c_i, s_i) = \lambda_i(c_0, s_0)$ for every i. (Write c_0 as a linear combination $c_0 = \gamma_1 c_1 + \gamma_2 c_2 + \ldots + \gamma_m c_m$, $\gamma_i \in \mathbb{Z}$; then s_0 is given by

$s_0 = \gamma_1 s_1 + \gamma_2 s_2 + \ldots + \gamma_m s_m$.) Clearly $N|c_0$, and so, by $c_i \leqslant (2/\sqrt{3})(1+2N^{\varepsilon/2})\sqrt{\ell_i}N$, we have $\lambda_i \leqslant (2/\sqrt{3})(1+2N^{\varepsilon/2})\sqrt{\ell_i}$.

Let i, j be arbitrary. By the last line of (4.10),

$$\lambda_i^2 s_0^2 = s_i^2 = t_i^2 - 4\ell_i - 4\lambda_i b_i c_0,$$
$$\lambda_j^2 s_0^2 = s_j^2 = t_j^2 - 4\ell_j - 4\lambda_j b_j c_0$$

and thus

$$\lambda_j^2 \lambda_i^2 s_0^2 = \lambda_j^2 (t_i^2 - 4\ell_i - 4\lambda_i b_i c_0),$$
$$\lambda_i^2 \lambda_j^2 s_0^2 = \lambda_i^2 (t_j^2 - 4\ell_j - 4\lambda_j b_j c_0).$$

Substracting, we obtain

$$0 = \lambda_j^2 (t_i^2 - 4\ell_i) - \lambda_i^2 (t_j^2 - 4\ell_j) - 4\lambda_i\lambda_j c_0 (\lambda_j b_i - \lambda_i b_j).$$

Now

$$\begin{aligned}|\lambda_j^2(t_i^2 - 4\ell_i) - \lambda_i^2(t_j^2 - 4\ell_j)| &\leqslant \max(\lambda_j^2 t_i^2 + 4\ell_j\lambda_i^2, \lambda_i^2 t_j^2 + 4\ell_i\lambda_j^2)\\ &\leqslant \frac{8}{\sqrt{3}}(1+2N^{\varepsilon/2})((1+2N^{\varepsilon/2})^2+1)\mathscr{L}^{3/2}\max(\lambda_i,\lambda_j)\\ &< N\max(\lambda_i,\lambda_j),\end{aligned}$$

where we use (4.2). On the other hand, $c_0 \geqslant N$ (because $N|c_0$) and so

$$4\lambda_i\lambda_j c_0 \geqslant 4\lambda_i\lambda_j N \geqslant N\max(\lambda_i,\lambda_j).$$

Thus we must have $(\lambda_j b_i - \lambda_i b_j) = 0$ (as otherwise we would have a contradiction). In other words, the tuples (b_i, c_i, s_i) are all proportional to each other. Write $(b_i, c_i, s_i) = \lambda_i(b_0, c_0, s_0)$, where (by the same reasoning we used for s_0) b_0 is an integer.

Define

$$A = \begin{pmatrix} (s_0+\varepsilon_0)/2 & b_0 \\ c_0 & -(s_0-\varepsilon_0)/2 \end{pmatrix} \text{ with } \varepsilon_0 := \begin{cases} 0 & \text{if } s_0 \text{ is even,} \\ 1 & \text{otherwise.}\end{cases}$$

The statement $(b_i, c_i, s_i) = \lambda_i(b_0, c_0, s_0)$ implies that $\rho_i = \lambda_i A + \delta_i I$ for some $\delta_i \in \mathbb{Z}/2$. Moreover, $t_i = 2\delta_i + \varepsilon_0\lambda_i$, and thus, by (4.10), $|\delta_i| \leqslant (1+3^{-1/2})(1+2N^{\varepsilon/2})\sqrt{\mathscr{L}}$. □

Proposition 4.4– *Let $z \in \sigma_N \setminus \cup_{\delta\in A_0(N)}\delta.\sigma_{\eta N}$, N a square-free positive integer. Let* (4.1), (4.2) *and* (4.3) *hold for $\mathscr{L} = \ell$. Then*

$$|\mathscr{M}_*(\ell, N; z)| \ll \tau(\ell)(\log(\ell) + \varepsilon\log(N))$$

for all $\varepsilon > 0$ and where $\tau(\ell)$ is the number of divisors of ℓ.

Remark 4.5– In our applications, ℓ will be always the product of two numbers each equal to 1, a prime or the square of a prime. In that case, $\tau(\ell) \ll 1$.

Remark 4.6– It is tempting to believe it should be possible to somehow relax the quite strict diophantine constraints imposed in (4.2) and (4.3) when ℓ is a perfect square. This would improve the bound given in Theorem A.

Proof of Proposition 4.4. We have $N \leqslant c \leqslant (2/\sqrt{3})(1+2N^{\varepsilon/2})\sqrt{\ell}N$ (by (4.10)), an interval which we can split into $O(\varepsilon\log(N) + \log(\ell))$ dyadic intervals $C \leqslant c < 2C$.

We apply Proposition 4.2 to each such dyadic interval. We obtain that there are integers b_0, c_0, s_0 such that, for every solution $\rho = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with c in our interval, there is a $\lambda \in \mathbb{Z}$ such that $(b, c, s) = \lambda(b_0, c_0, s_0)$.

Now recall that $s^2 = t^2 - 4\ell - 4bc$ (last line of (4.10); this is simply the determinant equation), and so $t^2 - 4\ell = \lambda^2(s_0^2 + 4b_0c_0)$. Define $d_0 = s_0^2 + 4b_0c_0 \in \mathbb{Z}$. Then

$$4\ell = t^2 - \lambda^2 d_0 = (t - \lambda\sqrt{d_0})(t + \lambda\sqrt{d_0}).$$

This is a factorisation of 4ℓ into two (principal) ideals of $\mathbb{Q}(\sqrt{d_0})$ of equal norm (or, if d_0 is a square, simply a factorisation of 4ℓ in $\mathbb{Z}$). There are at most $\tau(4\ell) \ll \tau(\ell)$ such factorisations for given ℓ, and so the bound follows. □

5. THE TWISTED SECOND MOMENT

Following [Tem10, Section 2.4], we define

$$h(r) := \left(\cosh\left(\frac{\pi r}{2} + 2\right)\right)^{-1}, \quad r \in \mathbb{R} \cup i\mathbb{R}.$$

This function h is an even positive function on $\mathbb{R} \cup i\mathbb{R}$. It turns out that h is the Selberg transform of a smooth point-pair invariant $k : \mathbb{R}_+ \to \mathbb{R}$ satisfying

$$k(u) \ll_A (1+u)^{-A} \tag{5.1}$$

for all $A > 0$ and $u \geqslant 0$ (here [Tem10] cites the survey paper [Mar, §5, Prop. 3]). The twisted second moment is defined by

$$M_2(\ell; z) := \sum_{j \geqslant 0} \lambda_j(\ell) h(r_j) |f_j(z)|^2 + * * *.$$

Here and from now on, $\beta_N = (f_j)_{j \geqslant 0}$ is an orthonormal basis of Hecke-Maaß eigenforms with f_0 the constant function and f_j cuspidal otherwise. The Laplace eigenvalue of f_j is $1/4 + r_j^2$ and $\lambda_j(\ell)$ is its ℓ-th Hecke eigenvalue. Lastly, $* * *$ stands for the contribution of the continuous spectrum and will be eliminated by positivity in the amplification step (see (6.1)). We would like to bound $M_2(\ell; z)$ following the strategy in [IS95].

***Proposition* 5.1**– *If $\ell \leqslant 4\eta^2 N^2/3$ then, under the assumptions of Proposition 4.4,*

$$M_2(\ell; z) \ll_\varepsilon \frac{N^\varepsilon}{\sqrt{\ell}}$$

for all $\varepsilon > 0$.

Remark 5.2– It should be mentioned that N. Templier got the same bound for the twisted second moment in the case of compact arithmetic surfaces (see [Tem10, Proposition 6.6]) but with less restrictive constraints on ℓ. This partly explains why his bound for the sup norm in the compact case is better than ours.

Remark 5.3– Averaging the previous result over ℓ improves the bound proved by N. Templier for the averaged twisted second moment (see [Tem10, Proposition 3.2]. This partly explains why our bound for the sup norm in the non-compact case is better than Templier's.

Proof of Proposition 5.1. The pre-trace formula (see [IS95]) says that

$$M_2(\ell;z) = \frac{1}{\sqrt{\ell}} \sum_{\rho \in \mathrm{M}(\ell,N)} k\big(u(\rho.z,z)\big).$$

By (5.1),

$$\begin{aligned} |M_2(\ell;z)| &\leqslant \frac{1}{\sqrt{\ell}} \sum_{\rho \in \mathrm{M}(\ell,N)} |k\big(u(\rho.z,z)\big)| \\ &= \frac{1}{\sqrt{\ell}} \sum_{\rho \in \mathcal{M}(\ell,N;z)} |k\big(u(\rho.z,z)\big)| + O_A(N^{-A}) \\ &\ll_A \frac{1}{\sqrt{\ell}} |\mathcal{M}(\ell,N;z)| + \frac{1}{N^A} \end{aligned}$$

for any $A > 0$. Propositions 4.1 and 4.4 are then used. □

6. End of the proof

Let f be a newform of square-free level N. We want to estimate $|f(z)|$. We can assume that $z \in \sigma_N$ by [Tem10, Lemma 2.1] since N is square-free and newforms are eigenvectors of $A_0(N)$ with eigenvalues ± 1. We can also assume that f belongs to the orthonormal basis β_N. Let $\Lambda := \{p \text{ prime}, p \nmid N, L \leqslant p \leqslant 2L\}$ for some integer L. Iwaniec's classical amplifier is defined by

$$x_\ell := \begin{cases} -\lambda_f(\ell) & \text{if } \ell \in \Lambda, \\ 1 & \text{if } \ell \in \Lambda^2, \\ 0 & \text{otherwise.} \end{cases}$$

This amplifier satisfies

$$\left| \sum_{\ell \geqslant 1} x_\ell \lambda_f(\ell) \right| \gg_\varepsilon L^{1-\varepsilon} \tag{6.1}$$

since $\lambda_f(p)^2 - \lambda_f(p^2) = 1$ for all prime $p \nmid N$.

Remark 6.1– Note that N. Templier in [Tem10, Section 3.4] uses Venkatesh's variation (see [Ven10]) of Iwaniec's amplifier since he only gets a bound for the twisted second moment on average over ℓ. This enables him to remove the assumption f non-exceptional, which occurs in [BH10]. In our case, we can use Iwaniec's clasical amplifier and appeal to Rankin-Selberg theory to bound on average the Hecke eigenvalues.

Let η, H, Q be some parameters, which satisfy all the constraints given in Proposition 5.1 for all $\ell \leqslant (2L)^4$.

Let us assume first that $z \in \sigma_N \setminus \cup_{\delta \in A_0(N)} \delta.\sigma_{\eta N}$. We successively have

$$\begin{aligned}
|f(z)|^2 &\ll \frac{1}{L^{2-2\varepsilon}} \left| \sum_{\ell \geqslant 1} x_\ell \lambda_f(\ell) \right|^2 h(r_f)|f(z)|^2 \\
&\leqslant \frac{1}{L^{2-2\varepsilon}} \left\{ \sum_{j \geqslant 0} \left| \sum_{\ell \geqslant 1} x_\ell \lambda_j(\ell) \right|^2 h(r_j)|f_j(z)|^2 + \mathrm{cont} \right\} \\
&= \frac{1}{L^{2-2\varepsilon}} \left\{ \sum_{j \geqslant 0} \sum_{\ell_1, \ell_2 \geqslant 1} x_{\ell_1} x_{\ell_2} \lambda_j(\ell_1) \lambda_j(\ell_2) h(r_j) |f_j(z)|^2 + \mathrm{cont} \right\} \\
&= \frac{1}{L^{2-2\varepsilon}} \left\{ \sum_{j \geqslant 0} \sum_{\ell_1, \ell_2 \geqslant 1} x_{\ell_1} x_{\ell_2} \left[\sum_{d | (\ell_1, \ell_2)} \lambda_j \left(\frac{\ell_1 \ell_2}{d^2} \right) \right] h(r_j) |f_j(z)|^2 + \mathrm{cont} \right\} \\
&\leqslant \frac{1}{L^{2-2\varepsilon}} \sum_{\ell_1, \ell_2 \geqslant 1} |x_{\ell_1}||x_{\ell_2}| \sum_{d | (\ell_1, \ell_2)} \left| M_2 \left(\frac{\ell_1 \ell_2}{d^2} \right) \right| \\
&\ll_\varepsilon \frac{1}{L^{2-2\varepsilon}} L^\varepsilon ||x||_2^2 \\
&\ll_\varepsilon \frac{L^{3\varepsilon}}{L}
\end{aligned}$$

according to the fact that $h(r_f) \gg 1$, (6.1), the positivity of h, the multiplicative properties of Hecke eigenvalues, Proposition 5.1 and by Rankin-Selberg theory.

If z belongs to $\cup_{\delta \in A_0(N)} \delta.\sigma_{\eta N}$ then

$$|f(z)|^2 \ll_\varepsilon N^\varepsilon \eta$$

by [Tem10, Lemma 3.1].

Finally, the following choice for the parameters

$$(H, Q, L, \eta) = \left(N^{5/9}, N^{2/5-\varepsilon/2}, N^{1/10-\varepsilon/2}, N^{-1/10} \right)$$

is both optimal (up to a factor of $N^{O(\varepsilon)}$) and admissible (for ε smaller than an absolute positive constant and N larger than an absolute constant.) This choice of parameters gives us

$$|f(z)| \ll_\varepsilon N^{-1/20+O(\varepsilon)}.$$

REFERENCES

[AU95] Ahmed Abbes and Emmanuel Ullmo. Comparaison des métriques d'Arakelov et de Poincaré sur $X_0(N)$. *Duke Math. J.*, 80(2):295–307, 1995.

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

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

[JK04] J. Jorgenson and J. Kramer. Bounding the sup-norm of automorphic forms. *Geom. Funct. Anal.*, 14(6):1267–1277, 2004.

[JK09] Jay Jorgenson and Jürg Kramer. Bounds on Faltings's delta function through covers. *Ann. of Math. (2)*, 170(1):1–43, 2009.

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

[Mar] J. Marklof. Selberg's trace formula: an introduction. Proceedings of the International School "Quantum Chaos on Hyperbolic Manifolds" (to appear).

[MU98] P. Michel and E. Ullmo. Points de petite hauteur sur les courbes modulaires $X_0(N)$. *Invent. Math.*, 131(3):645–674, 1998.

[NTY01] N. Nadirashvili, Dzh. Tot, and D. Yakobson. Geometric properties of eigenfunctions. *Uspekhi Mat. Nauk*, 56(6(342)):67–88, 2001.

[Sar95] Peter Sarnak. Arithmetic quantum chaos. In *The Schur lectures (1992) (Tel Aviv)*, volume 8 of *Israel Math. Conf. Proc.*, pages 183–236. Bar-Ilan Univ., Ramat Gan, 1995.

[SS89] A. Seeger and C. D. Sogge. Bounds for eigenfunctions of differential operators. *Indiana Univ. Math. J.*, 38(3):669–682, 1989.

[Tem10] Nicolas Templier. On the sup-norm of Maass cusp forms of large level. *Selecta Math. (N.S.)*, 16(3):501–531, 2010.

[Ven10] Akshay Venkatesh. Sparse equidistribution problems, period bounds and subconvexity. *Ann. of Math. (2)*, 172(2):989–1094, 2010.