Algebra & Number Theory

Volume 11

2017

No. 5

Hybrid sup-norm bounds for Maass newforms of powerful level

Abhishek Saha

msp

ALGEBRA AND NUMBER THEORY 11:5 (2017)
dx.doi.org/10.2140/ant.2017.11.1009

Hybrid sup-norm bounds for Maass newforms of powerful level

Abhishek Saha

Let f be an L^2-normalized Hecke–Maass cuspidal newform of level N, character χ and Laplace eigenvalue λ. Let N_1 denote the smallest integer such that $N|N_1^2$ and N_0 denote the largest integer such that $N_0^2|N$. Let M denote the conductor of χ and define $M_1 = M/\gcd(M, N_1)$. We prove the bound $\|f\|_\infty \ll_\varepsilon N_0^{1/6+\varepsilon} N_1^{1/3+\varepsilon} M_1^{1/2} \lambda^{5/24+\varepsilon}$, which generalizes and strengthens previously known upper bounds for $\|f\|_\infty$.

This is the first time a hybrid bound (i.e., involving both N and λ) has been established for $\|f\|_\infty$ in the case of nonsquarefree N. The only previously known bound in the nonsquarefree case was in the N-aspect; it had been shown by the author that $\|f\|_\infty \ll_{\lambda,\varepsilon} N^{5/12+\varepsilon}$ provided $M = 1$. The present result significantly improves the exponent of N in the above case. If N is a squarefree integer, our bound reduces to $\|f\|_\infty \ll_\varepsilon N^{1/3+\varepsilon}\lambda^{5/24+\varepsilon}$, which was previously proved by Templier.

The key new feature of the present work is a systematic use of p-adic representation theoretic techniques and in particular a detailed study of Whittaker newforms and matrix coefficients for $\mathrm{GL}_2(F)$ where F is a local field.

1. Introduction

1A. ***The main result.*** Let f be a Hecke–Maass cuspidal newform on the upper half plane of weight 0, level N, character χ, and Laplace eigenvalue λ. We normalize the volume of $Y_0(N)$ to be equal to 1 and assume that $\langle f, f\rangle := \int_{Y_0(N)} |f(z)|^2\, dz = 1$. The problem of bounding the sup-norm $\|f\|_\infty := \sup_{z\in Y_0(N)} |f(z)|$ in terms of the parameters N and λ is interesting from several points of view (quantum chaos, spectral geometry, subconvexity of L-functions, diophantine analysis) and has been much studied recently. For *squarefree* levels N, there were several results, culminating in the best currently known bound due to Templier [2015], which states that

$$\|f\|_\infty \ll_\varepsilon \lambda^{5/24+\varepsilon} N^{1/3+\varepsilon}.$$

The author is supported by EPSRC grant EP/L025515/1.
MSC2010: primary 11F03; secondary 11F41, 11F60, 11F72, 11F85, 35P20.
Keywords: Maass form, sup-norm, automorphic form, newform, amplification.

The exponent $\frac{5}{24}$ above for λ has stayed stable since the pioneering work of Iwaniec and Sarnak [1995] (who proved $\|f\|_\infty \ll_\varepsilon \lambda^{5/24+\varepsilon}$ in the case $N = 1$), and it will likely require some key new idea to improve it. The exponent $\frac{1}{3}$ for N in the above bound also appears difficult to improve, at least for squarefree levels, as it seems that the method used so far, primarily due to Harcos and Templier [2012; 2013] and Templier [2010; 2015] has been pushed to its limit. The purpose of the present paper is to show that the situation is very different for *powerful* (nonsquarefree) levels.

To state our result, we introduce a bit of notation. Let N_1 denote the smallest integer such that $N \mid N_1^2$. Let N_0 be the largest integer such that $N_0^2 \mid N$. Thus N_0 divides N_1 and $N = N_0 N_1$.[1] Note that if N is squarefree, then $N_1 = N$ and $N_0 = 1$. On the other hand, if N is a perfect square or if N is highly powerful (a product of high powers of primes) then $N_1 \asymp N_0 \asymp \sqrt{N}$. Also, let M be the conductor of χ (so M divides N) and put $M_1 = M/\gcd(M, N_1)$. Note that M_1 divides N_0, and in fact M_1 equals 1 if and only if M divides N_1. We will refer to the complementary situation of $M_1 > 1$ (i.e., $M \nmid N_1$) as the case when the character χ is *highly ramified*.

We prove the following result, which generalizes and strengthens previously known upper bounds for $\|f\|_\infty$.

Theorem (see Theorem 3.2). *We have*

$$\|f\|_\infty \ll_\varepsilon N_0^{1/6+\varepsilon} N_1^{1/3+\varepsilon} M_1^{1/2} \lambda^{5/24+\varepsilon}.$$

Thus in the squarefree case, our result reduces to that of Templier. However when $N_1 \asymp N_0 \asymp \sqrt{N}$, and $M \mid N_1$ (i.e., χ is not highly ramified), our result gives

$$\|f\|_\infty \ll_\varepsilon N^{1/4+\varepsilon} \lambda^{5/24+\varepsilon}.$$

The exponent of $\frac{1}{4}$ we obtain in this case is better than the exponent of $\frac{1}{3}$ in the squarefree case.

We note that the only upper bound known before this for general (i.e., possibly nonsquarefree) N is due to the present author, and was proved only very recently [Saha 2014]. It was shown that

$$\|f\|_\infty \ll_{\lambda,\varepsilon} N^{5/12+\varepsilon}$$

when χ is trivial (no dependence on λ was proved). The results of this paper not only substantially improve those of [Saha 2014] but also use quite different methods. We believe that the approach we take in this paper, characterized by a systematic use of adelic language and local representation-theoretic techniques that separate the difficulties place by place, is the right one to take for powerful levels.

[1] If N has the prime factorization $N = \prod_p p^{n_p}$, then N_0 and N_1 have prime factorizations $N_0 = \prod_p p^{\lfloor n_p/2 \rfloor}$ and $N_1 = \prod_p p^{\lceil n_p/2 \rceil}$.

As for the optimum upper bound for $\|f\|_\infty$, it is reasonable to conjecture that

$$\|f\|_\infty \ll_\varepsilon N^\varepsilon \lambda^{1/12+\varepsilon}, \tag{1}$$

if $M_1 = 1$ (i.e., provided χ is *not* highly ramified). If true, (1) is optimal as one can prove *lower bounds* of a similar strength. (But see page 1044.) If χ is highly ramified, we cannot expect (1) to hold, for reasons explained in [Saha 2016]. Roughly speaking, in the highly ramified case, the corresponding *local Whittaker newforms* can have large peaks due to a conspiracy of additive and multiplicative characters. This leads to a lower bound for $\|f\|_\infty$ that is larger than N^ε in the N-aspect. This phenomenon was first observed in Templier [2014] in the case when χ is maximally ramified ($M_1 = N_0$) and extended in [Saha 2016] to cover a much bigger range of M_1. That the factor $M_1^{1/2}$ is present in our main theorem above (giving worse upper bounds in the highly ramified case) fits nicely with this theme.

In the table below we compare the upper bound provided by this paper with the lower bound provided in [Saha 2016]. We consider newforms of level $N = p^n$, for $1 \le n \le 5$. The second column gives the possible values of M in each case, and the next three give the corresponding values of N_0, N_1 and M_1. The penultimate column gives the upper bound provided by the theorem on the previous page and should

N	M	N_0	N_1	M_1	factor $N_0^{1/6}N_1^{1/3}M_1^{1/2}$ in this work's upper bound	factor in lower bound from [Saha 2016]
p	1 or p	1	p	1	$N^{1/3}$	1
p^2	1 or p	p	p	1	$N^{1/4}$	1
	p^2	p	p	p	$N^{1/2}$	$N^{1/4}$
p^3	$1, p$ or p^2	p	p^2	1	$N^{5/18}$	1
	p^3	p	p^2	p	$N^{4/9}$	$N^{1/6}$
p^4	$1, p$ or p^2	p^2	p^2	1	$N^{1/4}$	1
	p^3	p^2	p^2	p	$N^{3/8}$	1
	p^4	p^2	p^2	p^2	$N^{1/2}$	$N^{1/4}$
p^5	$1, p, p^2$ or p^3	p^2	p^3	1	$N^{4/15}$	1
	p^4	p^2	p^3	p	$N^{11/30}$	$N^{1/10}$
	p^5	p^2	p^3	p^2	$N^{7/15}$	$N^{1/5}$

Table 1. A comparison of upper and lower bounds for $\|f\|_\infty$. The penultimate column gives the factor that replaces $\cdots$ in the bound $\|f\|_\infty \ll_\varepsilon N^\varepsilon\lambda^{5/24+\varepsilon} \times \cdots$ given by our main theorem on the previous page. The last column gives the corresponding factor in the previously known lower bound $\|f\|_\infty \gg_\varepsilon N^{-\varepsilon}\lambda^{1/12-\varepsilon} \times \cdots$.

serve as a nice numerical illustration of our result in the *depth aspect* ($N = p^n$, $n \to \infty$). The final column gives the corresponding lower bound proved in Theorem 3.3 of [Saha 2016]. The difference between these last two columns reflects the gap in the state of our current knowledge. As the table makes clear, the larger upper bounds for highly ramified χ are often matched by larger lower bounds.

Finally, we have colored blue all the quantities on the last column that we (optimistically) conjecture to be in fact the *true size* of $\|f\|_\infty$ (up to a factor of $(N\lambda)^\varepsilon$) in those cases.

1B. ***Organization of this paper.*** The remainder of Section 1 is an extended introduction that explains some of the main features of our work. In Section 2, which is the technical heart of this paper, we undertake a detailed analytic study of p-adic Whittaker newforms and matrix coefficients for representations of $\mathrm{GL}_2(F)$ where F is a nonarchimedean local field of characteristic 0. The two main results we prove are related to a) the support and average size of p-adic Whittaker newforms, b) the size of eigenvalues of certain matrix coefficients. These might be of independent interest. In Section 3, we prove the main result, Theorem 3.2. Perhaps surprisingly, and in contrast to our previous work [Saha 2014], no counting arguments are needed in this paper beyond those supplied by Templier for the squarefree case. Also, in contrast to [Saha 2014], we do not need any powerful version of the "gap principle". Instead, we rely almost entirely on the p-adic results of Section 2.

1C. ***Squarefree versus powerful levels.*** The first bound for $\|f\|_\infty$ in the N-aspect was $\|f\|_\infty \ll_{\lambda,\epsilon} N^{216/457+\epsilon}$, proved by Blomer and Holowinsky [2010]. They also proved the hybrid bound $\|f\|_\infty \ll (\lambda^{1/2}N)^{1/2-1/2300}$. These results were only valid under the assumption that N is squarefree. After that, there was fairly rapid progress (again only assuming N squarefree) by Harcos and Templier [Harcos and Templier 2012; 2013; Templier 2010; 2015], culminating in the hybrid bound due to Templier described earlier. Note that the N-exponent in Templier's case is $\frac{1}{3}$, which may be viewed as the "Weyl exponent", as it is a third of the way from the trivial bound of $N^{1/2+\varepsilon}$ towards the expected optimum bound[2] of N^ε.

For a long time, there was no result at all when N is not squarefree. Indeed, all the papers of Harcos and Templier rely crucially on using *Atkin–Lehner operators* to move any point of $\mathbb{H}$ to a point of imaginary part $\geq \frac{1}{N}$ (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 [Saha 2014], the first (and only previous) result for Maass forms of nonsquarefree level was proved; assuming that $M = 1$ we showed that $\|f\|_\infty \ll_{\lambda,\varepsilon} N^{5/12+\varepsilon}$. A key new idea in [Saha 2014] was to look at the behavior of f around *cusps of width 1* and to formulate all the geometric and

[2] As mentioned earlier, this optimum bound is only expected to hold when χ is not highly ramified.

diophantine results around such a cusp. Apart from this, the overall strategy was not that different from the works of Harcos and Templier and the exponent of $\frac{5}{12}$ obtained was weaker than the exponent $\frac{1}{3}$ for the squarefree case.

An initial indication that the exponent $\frac{1}{3}$ in the N-aspect might be beaten for powerful levels was given by Marshall [2016], who showed recently that for a newform g of level N and trivial character on a *compact arithmetic surface* (i.e., coming from a quaternion division algebra) the bound

$$\|g\|_\infty \ll_\varepsilon \lambda^{1/4+\varepsilon} N_1^{1/2+\varepsilon}$$

holds true. In particular, when N is sufficiently powerful, this gives a “sub-Weyl” exponent of $\frac{1}{4}$ in the N-aspect. Marshall’s proof does not work for the usual Hecke–Maass newforms f on the upper-half plane of level N that we consider in this paper (though it does work for certain shifts of these f when restricted to a *fixed compact* set). Finally, our main result, Theorem 3.2, gives (when χ is not highly ramified) the bound $\|f\|_\infty \ll_\varepsilon N_0^{1/6+\varepsilon} N_1^{1/3+\varepsilon} \lambda^{5/24+\varepsilon}$ which may be viewed as a strengthened analogue of Marshall’s result for cusp forms on the upper-half plane.

As indicated already, the powerful level case has been historically more difficult than the squarefree case. It may thus seem surprising that in the powerful case, we succeed in obtaining better exponents than in the squarefree case. However this seems to be a relatively common phenomenon. For example, for the related problem of quantum unique ergodicity in the level aspect, the known results in the squarefree case [Nelson 2011] give mass equidistribution with no power-savings but for powerful levels one obtains mass equidistribution with power savings [Nelson et al. 2014]. Again, for the problem of proving strong subconvexity bounds in the conductor aspect for Dirichlet L-functions, one only has a Weyl exponent $\frac{1}{6}$ when the conductor is squarefree, but Milićević [2016] has shown an improved exponent of $.1645\ldots < \frac{1}{6}$ for high prime powers. The results of this paper continue this surprising pattern (for which we do not attempt to give a general conceptual explanation).

1D. ***Fourier expansions and efficient generating domains.*** It seems worth noting explicitly the following interesting technical aspect of our work: the method of Fourier (Whittaker) expansion, once one chooses a good (adelic) *generating domain*, leads to the rather strong bound $\|f\|_\infty \ll_\varepsilon M_1^{1/2} N_1^{1/2+\varepsilon} \lambda^{1/4+\varepsilon}$. Note that this bound reduces to the “trivial bound” when N is squarefree, but is almost of the same strength (in the N-aspect) as our main theorem when N is sufficiently powerful. In this subsection, we briefly explain the ideas behind this.

It is best to work adelically here. Let ϕ be the automorphic form associated to f, and let $g = g_f g_\infty \in G(\mathbb{A})$, where g_f denotes the finite part of g and g_∞ denotes the infinite component. Then $\|f\|_\infty = \sup_{g \in G(\mathbb{A})} |\phi(g)|$. Because of the invariance properties for ϕ, it suffices to restrict g to a suitable generating domain $D \subset G(\mathbb{A})$.

Roughly speaking, D can be any subset of $G(\mathbb{A})$ such that the natural map from D to $Z(\mathbb{A})G(\mathbb{Q})\backslash G(\mathbb{A})/\overline{K}$ is a surjection where $\overline{K}$ is a subgroup of $G(\mathbb{A})$ generated by a set of elements under which $|\phi|$ is right-invariant.

The Whittaker expansion for ϕ, which we want to exploit to bound $|\phi(g)|$, looks as follows:

$$\phi(g) = \sum_{q \in \mathbb{Q}_{\neq 0}} W_\phi\left(\left[\begin{smallmatrix} q & \\ & 1 \end{smallmatrix}\right] g\right).$$

The above is an infinite sum, but two things make it tractable. First of all there is an integer $Q(g_f)$, depending on g_f, such that the sum is supported only on those q whose denominator divides $Q(g_f)$. Secondly, the sum decays very quickly after a certain point $|q| > T(g_\infty)$ due to the exponential decay of the Bessel function. The upshot is that

$$|\phi(g)| \ll \sum_{\substack{n \in \mathbb{Z}_{\neq 0} \\ |n| < Q(g_f)T(g_\infty)}} W_\phi\left(\left[\begin{smallmatrix} n/Q(g_f) & \\ & 1 \end{smallmatrix}\right] g\right). \tag{2}$$

The key quantity is the *length* $Q(g_f)T(g_\infty)$ of the sum above. Indeed, assuming Ramanujan type bounds on average for the local Whittaker newforms and using Cauchy–Schwartz, the expression (2) leads to the inequality[3] $|\phi(g)| \ll_\varepsilon (Q(g_f)T(g_\infty))^{1/2+\varepsilon}$. The key point therefore, is to choose an efficient generating domain D inside $G(\mathbb{A})$, such that $\sup_{g \in D} Q(g_f)T(g_\infty)$ is as small as possible.

Let us look at some examples. Suppose ϕ corresponds to a Hecke–Maass cusp form for $\mathrm{SL}_2(\mathbb{Z})$. Then it is natural to take D to be the subset of $G(\mathbb{A})$ consisting of the elements g with $g_f = 1$ and $g_\infty = \left[\begin{smallmatrix} y & x \\ & 1 \end{smallmatrix}\right]$ such that $-\frac{1}{2} \le x \le \frac{1}{2}$ and $y \ge \frac{\sqrt{3}}{2}$. In this case $Q(g_f) = 1$ and $T(g_\infty) = \lambda^{1/2}/y$, leading to the bound $|\phi(g)| \ll_\varepsilon \lambda^{1/4+\varepsilon}$ as expected. Next, suppose ϕ corresponds to a newform of level N where N is squarefree. In this case one can include the Atkin–Lehner operators inside the symmetry group $\overline{K}$ above. Harcos and Templier showed that one can take D to be the subset of $G(\mathbb{A})$ consisting of the elements with $g_f = 1$ and for which $g_\infty = \left[\begin{smallmatrix} y & x \\ & 1 \end{smallmatrix}\right]$ such that $y \ge \sqrt{3}/(2N)$ (and some additional properties). For such an element, one again has $Q(g_f) = 1$ and $T(g_\infty) = \lambda^{1/2}/y$ leading to the bound $|\phi(g)| \ll_\varepsilon (\lambda^{1/2}/y)^{1/2+\varepsilon} \ll N^{1/2+\varepsilon}\lambda^{1/4+\varepsilon}$.

When N is nonsquarefree, it is not possible to construct a generating domain D with a finite value of $\sup_{g \in D} T(g_\infty)$ for which all points have $g_f = 1$. Classically, this means that any fundamental domain (for the full symmetry group generated by $\Gamma_0(N)$ and the Atkin–Lehner operators) must touch the real line. The idea used in [Saha 2014] was to take the infinite part of D essentially the same as in the squarefree case and take the finite part to be a certain nice subset of $\prod_{p|N} \mathrm{GL}_2(\mathbb{Z}_p)$.

[3] Strictly speaking, this inequality is not completely accurate as one has to add an (usually smaller) error term coming from peaks of the local Whittaker and K-Bessel functions.

Classically, our choice of generating domain in [Saha 2014] corresponded to taking discs around cusps of width 1. Assuming $\chi = 1$, this choice again gave $Q(g_f) = 1$ and $T(g_\infty) = \lambda^{1/2}/y$ leading to the same bound $|\phi(g)| \ll_\varepsilon N^{1/2+\varepsilon}\lambda^{1/4+\varepsilon}$ as earlier. Thus, in all the above papers, the worst case bound obtained by the Whittaker expansion (i.e., for smallest y) was just the *trivial bound* $N^{1/2+\varepsilon}\lambda^{1/4+\varepsilon}$ (also, all these papers restricted to $\chi = 1$).

In this paper we choose a somewhat different generating domain from that of [Saha 2014]. For simplicity, we describe this domain here in the special case when $N = p^{2n_0}$ and $M = p^m$, for some prime p and some nonnegative integers n_0 and m. Take D to consist of the elements $g_p g_\infty$, where $g_\infty = \left[\begin{smallmatrix} y & x \\ & 1 \end{smallmatrix}\right]$ with $y \geq \sqrt{3}/2$ and $g_p \in \mathrm{GL}_2(\mathbb{Z}_p)\left[\begin{smallmatrix} p^{n_0} & \\ & 1 \end{smallmatrix}\right]$. It is easy to prove this is a generating domain. The difficulties lie in computing $Q(g_f)$ and in proving that the required Ramanujan type bounds on average hold. These key technical local results involve intricate calculations that take up a good part of Section 2. We are able to prove that $\sup_{g\in D} Q(g_f) = p^{\max(m,n_0)} = M_1\sqrt{N}$. Also, $T(g_\infty) = \lambda^{1/2}/y$ as usual. This leads to the surprisingly strong Whittaker expansion bound of

$$|\phi(g)| \ll_\varepsilon M_1^{1/2} N^{1/4+\varepsilon}\lambda^{1/4+\varepsilon}$$

in this case. Classically, the generating domain described above corresponds to taking discs around the cusps of the group

$$\Gamma_0(p^{n_0}, p^{n_0}) = \left\{\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z}) : p^{n_0} \mid b,\ p^{n_0} \mid c\right\}.$$

We remark here that the function $f'(z) := f(z/p^{n_0})$ is a Maass form for $\Gamma_0(p^{n_0}, p^{n_0})$.

When N is not a perfect square, the generating domain we actually use is slightly different than described above. Roughly speaking, we exploit the existence of Atkin–Lehner operators at primes that divide N to an odd power. This does not change the value of $\sup_{g\in D} Q(g_f)T(g_\infty)$ and so does not really affect the Whittaker expansion analysis; however it makes it easier to count lattice points for amplification (described in the next subsection). In any case, the Whittaker expansion bound we prove ultimately (see Section 3D) is $|\phi(g)| \ll_\varepsilon (N_0 M_1 \lambda^{1/2}/y)^{1/2+\varepsilon}$ where $y \geq N_0/N_1$, leading to the worst case bound of $|\phi(g)| \ll_\varepsilon M_1^{1/2} N_1^{1/2+\varepsilon}\lambda^{1/4+\varepsilon}$. This, as mentioned earlier, is essentially of the same strength (in the N-aspect) as our main theorem when N is sufficiently powerful.

It bears repeating that the main tools used for the above bound are local, relating to the representation theory of p-adic Whittaker functions. This supports the assertion of Marshall [2016] that $N_1^{1/2+\varepsilon}$ should be viewed as the correct *local bound* in the level aspect (when χ is not highly ramified). Our analysis of these p-adic Whittaker functions also lead to other interesting questions. For example, one can ask for a sup-norm bound for these local Whittaker newforms, and in (1), we predict a Lindelöf type bound when χ is not highly ramified (this conjecture was originally made in [Saha 2016]). One of the key technical results in Section 2 essentially proves an *averaged* version of this conjecture (this is the Ramanujan type bound on average alluded to earlier).

1E. ***The pretrace formula and amplification.*** Recall our main theorem:

$$\|f\|_\infty \ll_\varepsilon N_0^{1/6+\varepsilon} N_1^{1/3+\varepsilon} M_1^{1/2} \lambda^{5/24+\varepsilon}.$$

As we have seen above, the method of Fourier (Whittaker) expansion gives us the bound $\|f\|_\infty \ll_\varepsilon N_1^{1/2+\varepsilon} M_1^{1/2} \lambda^{1/4+\varepsilon}$ (with even better bounds when the relevant point on our generating domain has a large value for y) so we need to save a further factor of $(N_1/N_0)^{1/6}\lambda^{1/24}$. This is done by *amplification*, whereby we choose suitable test functions at each prime to obtain a *pretrace formula* and then estimate its geometric side via some point counting results from [Harcos and Templier 2013; Templier 2015]. The basic idea is that by choosing these local test functions carefully (constructing an amplifier) one should be able to boost the contribution of the newform f to the resulting pretrace formula. The details for this are given (in a fairly flexible adelic framework) in Sections 3E–3G.

The unramified local test functions that we use in this paper are standard and essentially go back to Iwaniec–Sarnak (the key point is to exploit a simple identity relating the eigenvalues for the Hecke operators $T(\ell)$ and $T(\ell^2)$). However, our ramified local test functions are very different from the papers of Harcos and Templier or our previous paper [Saha 2014]. In all those past papers, the ramified test functions had been simply chosen to be the characteristic functions of the relevant congruence subgroups. In contrast, we use a variant of the local test function used by Marshall [2016]. The main results about this test function are proved in Sections 2F–2H. Roughly speaking, it is (the restriction to a large compact subgroup of) the *matrix coefficient* for a local vector v' obtained by translating the local newform. The key property of this test function is that its unique nonzero eigenvalue is fairly large (and v' is an eigenvector with this eigenvalue).

Our choice of test functions at ramified primes ensures that any pretrace formula involving them averages over relatively few representations of level N. It may be useful to view this as a ramified analogue of the classical (unramified) amplifier. Indeed, the resulting "trivial bound" obtained via the pretrace formula (by choosing the unramified test functions trivially) matches exactly (on compact subsets) with the strong local bounds obtained via Whittaker expansion. This is an important point because it means that we only need to save a further factor of $(N_1/N_0)^{1/6}\lambda^{1/24}$ by putting in the unramified amplifier and counting lattice points. This is carried out in Section 3G.

It is worth noting that we do not need any new counting results in this paper beyond those proved by Harcos and Templier. This is because the counting part of our paper is only concerned with the squarefree integer N_1/N_0. In particular, the role of amplification in this paper to improve the N exponent is relatively minor when N is highly powerful (note that N_1/N_0 approaches a negligible power of N as N gets more powerful). Indeed, when N is a perfect square ($N_1 = N_0$), all our savings in the N-aspect come from Whittaker expansion and we do not gain anything further by amplification.[4] In contrast, in [Saha 2014] we had a relatively poor bound coming from Whittaker expansion but we then saved a nontrivial power of N via amplification.

The technical reason why the method of amplification does not improve the N-aspect too much beyond our strong local bounds is that our ramified test functions have relatively large support. Consequently, we do not have many global congruences related to N, and congruences are essential for savings via counting. More precisely, our ramified test functions are supported on the maximal compact subgroup at primes that divide N to an even power, and supported on a (slightly) smaller subgroup at primes that divide N to an odd power (it is the latter case that leads to the savings of $(N_1/N_0)^{1/6}$). If we were to reduce the support of our test functions further and thus force new congruences, the resulting savings via counting would be eclipsed by the resulting loss due to the fact that our pretrace formula would now be averaging over more representations of level N. Somehow the ramified and unramified parts of the amplifier seem to work against each other and the key point is to strike the right balance.

It would be an interesting and challenging problem to detect any additional cancellation on the geometric side of our pretrace formula by going beyond counting lattice points and perhaps taking into account the *phases* of the matrix coefficient used to construct the ramified test function. Such a result could potentially push the upper-bound for $\|f\|_\infty$ below $N^{1/4}$.

1F. *Notations.* We collect here some general notations that will be used throughout this paper. Additional notations will be defined where they first appear in the paper.

Given two integers a and b, we use $a \mid b$ to denote that a divides b, and we use $a \mid b^\infty$ to denote that $a \mid b^n$ for some positive integer n. For any real number α, we let $\lfloor \alpha \rfloor$ denote the greatest integer less than or equal to α and we let $\lceil \alpha \rceil$ denote the smallest integer greater than or equal to α. The symbol $\mathbb{A}$ denotes the ring of adeles of $\mathbb{Q}$ and $\mathbb{A}_f$ denotes the subset of finite adeles. For any two complex numbers α and z we let $K_\alpha(z)$ denote the modified Bessel function of the second kind.

The groups GL_2, SL_2, PSL_2, $\Gamma_0(N)$ and $\Gamma_1(N)$ have their usual meanings. The letter G always stands for the group GL_2. If H is any subgroup of G, and R is any subring of $\mathbb{R}$, then $H(R)^+$ denotes the subgroup of $H(R)$ consisting of matrices with positive determinant.

[4] However, we always gain a nontrivial savings in the λ aspect via amplification.

We let $\mathbb{H} = \{x+iy : x \in \mathbb{R},\ y \in \mathbb{R},\ y > 0\}$ 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 $(az+b)/(cz+d)$. This action of $\mathrm{GL}_2(\mathbb{R})^+$ on $\mathbb{H}$ extends naturally to the boundary of $\mathbb{H}$.

We say that a function f on $\mathbb{H}$ is a Hecke–Maass cuspidal newform of weight 0, level N, character χ and Laplace eigenvalue λ if it has the following properties:

- f is a smooth real analytic function on $\mathbb{H}$.
- f satisfies $(\Delta + \lambda) f = 0$ where $\Delta := y^{-2}(\partial_x^2 + \partial_y^2)$.
- For all $\gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \Gamma_0(N)$, $f(\gamma z) = \chi(d) f(z)$.
- f decays rapidly at the cusps of $\Gamma_1(N)$.
- f is orthogonal to all oldforms.
- f is an eigenfunction of all the Hecke and Atkin–Lehner operators.[5]

The study of newforms f as above is equivalent to the study of corresponding adelic newforms ϕ which are certain functions on $G(\mathbb{A})$. For the details of this correspondence, see Remark 3.1.

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, so that $|A| \le C|B|$. The symbol ϵ will denote a small positive quantity. The values of ϵ and that of the constant implicit in $\ll_{\epsilon,\ldots}$ may change from line to line.

2. Local calculations

2A. ***Preliminaries.*** We begin with fixing some notations that will be used throughout this section. Let F be a nonarchimedean local field of characteristic zero whose residue field has cardinality q. Let $\mathfrak{o}$ be its ring of integers, and $\mathfrak{p}$ its maximal ideal. Fix a generator ϖ of $\mathfrak{p}$. Let $|\cdot|$ denote the absolute value on F normalized so that $|\varpi| = q^{-1}$. For each $x \in F^\times$, let $v(x)$ denote the integer such that $|x| = q^{-v(x)}$. For a nonnegative integer m, we define the subgroup U_m of $\mathfrak{o}^\times$ to be the set of elements $x \in \mathfrak{o}^\times$ such that $v(x-1) \ge m$.

Let $G = \mathrm{GL}_2(F)$ and $K = \mathrm{GL}_2(\mathfrak{o})$. For each integral ideal $\mathfrak{a}$ of $\mathfrak{o}$, let

$$K_0(\mathfrak{a}) = K \cap \begin{bmatrix} \mathfrak{o} & \mathfrak{o} \\ \mathfrak{a} & \mathfrak{o} \end{bmatrix}, \quad K_1(\mathfrak{a}) = K \cap \begin{bmatrix} 1+\mathfrak{a} & \mathfrak{o} \\ \mathfrak{a} & \mathfrak{o} \end{bmatrix}, \quad K^0(\mathfrak{a}) = K \cap \begin{bmatrix} \mathfrak{o} & \mathfrak{a} \\ \mathfrak{o} & \mathfrak{o} \end{bmatrix}.$$

For $x \in F$, $y \in F^\times$ and $t \in F^\times$, write

$$w = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad a(y) = \begin{bmatrix} y & \\ & 1 \end{bmatrix}, \quad n(x) = \begin{bmatrix} 1 & x \\ & 1 \end{bmatrix}, \quad z(t) = \begin{bmatrix} t & \\ & t \end{bmatrix},$$

[5] Assuming the previous properties, this last property is equivalent to the weaker condition that f is an eigenfunction of almost all Hecke operators.

Define subgroups $N = \{n(x) : x \in F\}$, $A = \{a(y) : y \in F^\times\}$, $Z = \{z(t) : t \in F^\times\}$, $B_1 = NA$ and $B = ZNA = G \cap \left[\begin{smallmatrix} * & * \\ & * \end{smallmatrix}\right]$ of G.

We normalize Haar measures as follows. The measure dx on the additive group F assigns volume 1 to $\mathfrak{o}$, and transports to a measure on N. The measure $d^\times y$ on the multiplicative group $F^\times$ assigns volume 1 to $\mathfrak{o}^\times$, and transports to measures on A and Z. We obtain a left Haar measure $d_L b$ on B via $d_L(z(u)n(x)a(y)) = |y|^{-1}\, d^\times u\, dx\, d^\times y$. Let dk be the probability Haar measure on K. The Iwasawa decomposition $G = BK$ gives a left Haar measure $dg = d_L b\, dk$ on G.

For each irreducible admissible representation σ of G (resp. $F^\times$) we define $a(\sigma)$ to be the smallest nonnegative integer such that σ has a $K_1(\mathfrak{p}^{a(\sigma)})$-fixed (resp. $U_{a(\sigma)}$-fixed) vector.

2B. ***Some matrix invariants.*** *From now on, fix π to be a generic irreducible admissible unitary representation of G. Let $n = a(\pi)$, and let ω_π denote the central character of π.*

It is convenient now to introduce some notation. Define

- $n_1 := \left\lceil \frac{n}{2} \right\rceil$,
- $n_0 := n - n_1 = \left\lfloor \frac{n}{2} \right\rfloor$,
- $m = a(\omega_\pi)$,
- $m_1 = \max(0, m - n_1)$.

Note that $m_1 = 0$ if and only if $m \le n_1$; this can be viewed as the case when ω_π is *not* highly ramified.

Next, for any $g \in G$, we define two integers $t(g)$ and $l(g)$ which depend on g and n. Recall the disjoint double coset decomposition [Saha 2016, Lemma 2.13]:

$$G = \bigsqcup_{t \in \mathbb{Z}} \bigsqcup_{0 \le l \le n} \bigsqcup_{v \in \mathfrak{o}^\times / U_{\min(l, n-l)}} ZNa(\varpi^t) w n(\varpi^{-l} v) K_1(\mathfrak{p}^n). \tag{3}$$

Accordingly, given any matrix $g \in G$, we define $t(g)$ and $l(g)$ to be the unique integers such that

- $0 \le l(g) \le n$,
- $g \in ZNa(\varpi^{t(g)}) w n(\varpi^{-l(g)} v) K_1(\mathfrak{p}^n)$ for some $v \in \mathfrak{o}^\times$.

Remark 2.1. It is illuminating to restate these matrix invariants slightly differently. Let g in G. The Iwasawa decomposition tells us that $g \in ZNa(y)k$ where $k = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in K$. Then one can check that $l(g) = \min(v(c), n)$, and $t(g) = v(y) - 2l(g)$.

In the sequel, we will often consider matrices g lying in the set $Ka(\varpi^{n_1})$. The next few lemmas explicate some key properties of this set.

Lemma 2.2. *Suppose that $k \in K$ and n is odd (so $n_1 = n_0 + 1$). Then*:

(1) *$l(ka(\varpi^{n_1})) \geq n_1$ if and only if $k \in N(\mathfrak{o})K^0(\mathfrak{p})$.*

(2) *$l(ka(\varpi^{n_1})) \leq n_0$ if and only if $k \in wK^0(\mathfrak{p})$.*

Proof. We first assume that $l(ka(\varpi^{n_1})) \geq n_1$ and prove that $k \in BK^0(\mathfrak{p})$. For brevity, put $l = l(ka(\varpi^{n_1}))$. So we can write $ka(\varpi^{n_1}) = bwn(\varpi^{-l}v)k'$, where $b \in B$, $k' \in K_1(\mathfrak{p}^n)$ and $n \geq l \geq n_1$. Therefore $k = b'wn(\varpi^{n_1-l}v)k_2$, where $k_2 = a(\varpi^{n_1})k'a(\varpi^{-n_1}) \in K^0(\mathfrak{p}^{n_1})$ and $b' \in B$. To complete the proof that $k \in BK^0(\mathfrak{p})$, it suffices to check that there exists a matrix $b_2 \in B$ such that $b_2wn(\varpi^{n_1-l}v) \in K^0(\mathfrak{p})$. By explicit verification, $b_2 = \left[\begin{smallmatrix} \varpi^{n_1-l}v & 1 \\ 0 & \varpi^{l-n_1}v^{-1} \end{smallmatrix}\right]$ works. Once we have $k \in BK^0(\mathfrak{p})$, it follows immediately that $k \in B(\mathfrak{o})K^0(\mathfrak{p}) = N(\mathfrak{o})K^0(\mathfrak{p})$.

The proof that $l(ka(\varpi^{n_1})) \leq n_0$ implies $k \in wK^0(\mathfrak{p})$ is similar. The reverse implications follow using $N(\mathfrak{o})K^0(\mathfrak{p}) \cap wK^0(\mathfrak{p}) = \varnothing$. $\square$

Lemma 2.3. *Suppose that $k \in K_0(\mathfrak{p})$, n is odd, and $g \in \left\{1, \left[\begin{smallmatrix} & 1 \\ \varpi & \end{smallmatrix}\right]\right\}$. Then*

$$kgwa(\varpi^{n_1}) = k'a(\varpi^{n_1})g'z,$$

where $k' \in K$, $l(k'a(\varpi^{n_1})) \leq n_0$, $g' \in \left\{1, \left[\begin{smallmatrix} & 1 \\ \varpi^n & \end{smallmatrix}\right]\right\}$, and $z \in Z$.

Proof. If $g = 1$, then $kgwa(\varpi^{n_1}) = w(w^{-1}kw)a(\varpi^{n_1})$. If $g = \left[\begin{smallmatrix} & 1 \\ \varpi & \end{smallmatrix}\right]$, then $kgwa(\varpi^{n_1}) = w(w^{-1}kw)a(\varpi^{n_1})\left[\begin{smallmatrix} & 1 \\ \varpi^n & \end{smallmatrix}\right]z(-\varpi^{n_1-n})$. Since $(w^{-1}kw) \in K^0(\mathfrak{p})$, the result now follows from Lemma 2.2. $\square$

Lemma 2.4. *Suppose that $g \in Ka(\varpi^{n_1})$. Then $t(g) = \min(n_1 - 2l(g), -n_1)$.*

Proof. This follows by an explicit computation similar to the proof of Lemma 2.2. We omit the details. $\square$

2C. *Our goal.* It may be worthwhile to declare at this point the output from the rest of Section 2 that will be needed for our main theorem.

In Sections 2D–2E, we will study the local Whittaker newform W_π, which is a certain function on G. Given a compact subset $\mathcal{J}$ of G, we are interested in the following questions:

(1) For each $g \in \mathcal{J}$, provide a good upper bound for the quantity

$$\sup\{|y| : W_\pi(a(y)g) \neq 0\}.^6$$

(2) Prove an average Ramanujan-type bound for the function $|W_\pi(a(y)g)|$ whenever $g \in \mathcal{J}$ and $W_\pi(a(y)g) \neq 0$.

For our global applications, it will be useful to have the set $\mathcal{J}$ to be relatively large (so that we can create a generating domain out of it with a relatively small archimedean component) while also making sure that the supremum of the upper

[6] This is essentially the local analogue of the quantity $Q(g_f)$ described in the introduction.

bound above (as g varies in $\mathcal{J}$) is fairly small (so as to optimize the Whittaker expansion bound). We will choose $\mathcal{J}$ to equal the set $Ka(\varpi^{n/2})$ if n is even and equal to $\{g \in Ka(\varpi^{n_1}) : l(g) \le n_0\}$ if n is odd. For this set $\mathcal{J}$ we will answer the two questions above in Proposition 2.11. This proposition will be of key importance for our global Whittaker expansion bound.

Next, in Sections 2F–2H, we will study a certain test function Φ'_π. This test function, viewed as a convolution operator, is essentially idempotent, and therefore has exactly one nonzero positive eigenvalue. In Proposition 2.13, we determine the size of this nonzero eigenvalue, and we also prove that $a(\varpi^{n_1}) \cdot W_\pi$ is an eigenvector with this eigenvalue. This proposition will be of key importance for our global bound coming from the amplified trace formula.

In view of the technical material coming up, it is worth emphasizing that Propositions 2.11 and 2.13 are the *only* results from the rest of Section 2 that will be used in Section 3.

2D. *The Whittaker newform.* Fix an additive character $\psi : F \to S^1$ with conductor $\mathfrak{o}$. Then π can be realized as a unique subrepresentation of the space of functions W on G satisfying $W(n(x)g) = \psi(x)W(g)$. This is the Whittaker model of π and will be denoted $\mathcal{W}(\pi, \psi)$.

Definition 2.5. The *normalized Whittaker newform* W_π is the unique function in $\mathcal{W}(\pi, \psi)$ invariant under $K_1(\mathfrak{p}^n)$ that satisfies $W_\pi(1) = 1$.

The following lemma is well known and so we omit its proof.

Lemma 2.6. *Suppose that $W_\pi(a(y)) \neq 0$. Then $|y| \le 1$, i.e., $y \in \mathfrak{o}$.*

Lemma 2.7 [Saha 2016, Proposition 2.28]. *Let $\tilde{\pi}$ denote the contragredient representation of π. Let $t \in \mathbb{Z}$, $0 \le l \le n$, $v \in \mathfrak{o}^\times$, and assume*[7] *$\omega_\pi(\varpi) = 1$. We have*

$$\begin{aligned} &W_{\tilde{\pi}}(a(\varpi^t)wn(\varpi^{-l}v)) \\ &\qquad = \varepsilon\left(\tfrac{1}{2}, \pi\right)\omega_\pi(v)\psi(-\varpi^{t+l}v^{-1})W_\pi(a(\varpi^{t+2l-n})wn(-\varpi^{l-n}v)). \end{aligned}$$

Define $g_{t,l,v} := a(\varpi^t)wn(\varpi^{-l}v)$. Let $\widetilde{X}$ denote the group of characters μ of $F^\times$ such that $\mu(\varpi) = 1$. For each $\mu \in \widetilde{X}$ and each $x \in F$, define the Gauss sum $G(x, \mu) = \int_{\mathfrak{o}^\times} \psi(xy)\mu(y)\, d^\times y$.

We will need two additional results for the results of the next subsection. The first one is a key formula from [Saha 2016].

Lemma 2.8 [Saha 2016, Proposition 2.23]. *Assume that $\omega_\pi \in \widetilde{X}$, we have*

$$W_\pi(g_{t,l,v}) = \sum_{\substack{\mu : a(\mu) \le l, \\ \mu \in \widetilde{X}}} c_{t,l}(\mu)\mu(v),$$

[7] There is no loss of generality in this assumption as we can always twist π by a character of the form $|\;|^{ir}$ to ensure this.

where the coefficients $c_{t,l}(\mu)$ can be read off from the following identity

$$\varepsilon\big(\tfrac{1}{2},\mu\pi\big)\bigg(\sum_{t=-\infty}^{\infty} q^{(t+a(\mu\pi))(1/2-s)}c_{t,l}(\mu)\bigg)L(s,\mu\pi)^{-1}$$
$$=\omega_\pi(-1)\bigg(\sum_{a=0}^{\infty} W_\pi(a(\varpi^a))q^{-a(1/2-s)}G(\varpi^{a-l},\mu^{-1})\bigg)L(1-s,\mu^{-1}\omega_\pi^{-1}\pi)^{-1}. \quad (4)$$

The next result deals with conductors of character twists. While the proof is quite easy, it involves a question that comes up frequently in such problems, see, e.g., Remark 1.9 of [Nelson et al. 2014].

Lemma 2.9. *Let $l \le n_0$ be a nonnegative integer. For each character μ with $a(\mu)=l$, we have $a(\mu\pi)\le \max(n,l+m)$. Furthermore, for each $r\ge 0$,*

$$\big|\{\mu\in\widetilde{X}: a(\mu)=l,\, a(\mu\pi)=\max(n,l+m)-r\}\big|\le q^{l-r/2}.$$

Proof. If π is supercuspidal we have $l+m\le n$. Writing π as a twist of a minimal supercuspidal, the result follows from Tunnell's theorem [1978, Proposition 3.4] on conductors of twists of supercuspidal representations. If π is principal series, then it follows from the well-known formula $a(\chi_1\boxplus\chi_2)=a(\chi_1)+a(\chi_2)$. If π is a twist of the Steinberg representations, it follows from the formula $a(\chi\,\mathrm{St})=\max(2\mathrm{a}(\chi),1)$. □

2E. *The support and average size of W_π.* In this subsection we will prove an important technical result (Proposition 2.10) about the size and support of W_π. This will then be combined with the results of the previous subsection to deduce Proposition 2.11 which will be needed for our global application. To motivate all these results, we first recall a conjecture made in [Saha 2016].

Conjecture 1 (local Lindelöf hypothesis for Whittaker newforms). *Suppose that $a(\omega_\pi)\le n_1$ (i.e., $m_1=0$). Then*

$$1\ll \sup_{g\in G}|W_\pi(g)|\ll_\varepsilon q^{n\varepsilon}.$$

This conjecture (originally stated as [Saha 2016, Conjecture 2]) seems to be quite hard as it implies square-root cancellation in sums of twisted GL_2-ε-factors. However, for the purpose of this paper, we can prove a bound that is (at least) as strong as the above conjecture on *average*. This is achieved by the second part of the next proposition, which generalizes some results obtained in [Nelson et al. 2014, Section 2], which considered the special case $\omega_\pi=1$.

Proposition 2.10. (1) *If $W_\pi(g)\ne 0$, then $t(g)\ge -\max(2l(g),l(g)+m,n)$.*

(2) *Suppose* $t(g) = -\max(2l(g), l(g)+m, n)+r$ *where* $r \geq 0$. *Then we have*

$$\left(\int_{v\in\mathfrak{o}^\times} \left|W_\pi(a(v)g)\right|^2 d^\times v\right)^{1/2} \ll q^{-r/4}.$$

Proof. By twisting π with a character of the form $|\ |^{ir}$ if necessary (which does not change $|W_\pi|$), we may assume $\omega_\pi \in \widetilde{X}$. Also assume $n \geq 1$, as the case $n = 0$ is trivial. Because of the coset decomposition from earlier, we may further assume that $g = g_{t,l,v} := a(\varpi^t)wn(\varpi^{-l}v))$. Finally, because of Lemma 2.7, we can assume (by replacing π by $\tilde{\pi}$ if necessary) that $0 \leq l \leq n_0$. The desired result then is the following:

- Let $0 \leq l \leq n_0$. If $W_\pi(g_{t,l,v}) \neq 0$, then $t \geq -\max(n, l+m)$. Further if $t = -\max(n, l+m)+r$ where $r \geq 0$ then

$$\left(\int_{v\in\mathfrak{o}^\times} \left|W_\pi(g_{t,l,v})\right|^2 d^\times v\right)^{1/2} \ll q^{-r/4}.$$

In the notation of (4), the above is equivalent to:

Claim 1. *Let* $0 \leq l \leq n_0$. *If there exists* $\mu \in \widetilde{X}$ *such that* $a(\mu) \leq l$ *and* $c_{t,l}(\mu) \neq 0$ *then* $t \geq -\max(n, l+m)$. *Further if* $t = -\max(n, l+m)+r$ *where* $r \geq 0$ *then* $\sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)\leq l}} |c_{t,l}(\mu)|^2 \ll q^{-r/2}$.

Define the quantities $d_{t,l}(\mu)$ via the following identity (of polynomials in $q^{\pm s}$):

$$\varepsilon\left(\tfrac{1}{2}, \mu\pi\right)\left(\sum_{t=-\infty}^{\infty} q^{(t+a(\mu\pi))(1/2-s)} c_{t,l}(\mu)\right) L(s, \mu\pi)^{-1} = \left(\sum_{t=-\infty}^{\infty} q^{(t+a(\mu\pi))(1/2-s)} d_{t,l}(\mu)\right). \quad (5)$$

Note that (for fixed l and μ) $d_{t,l}(\mu)$ is nonzero for only finitely many t. Furthermore, $c_{t,l}(\mu) = \sum_{i=0}^{\infty} \alpha_i d_{t-i,l}(\mu)$ where $|\alpha_0| = 1$ and $|\alpha_i| \ll q^{-i/2}$. (In fact, if π is supercuspidal, $\alpha_i = 0$ for all $i > 0$). Hence it suffices to prove Claim 1 for the quantities $d_{t,l}(\mu)$ rather than $c_{t,l}(\mu)$. Therefore using (4) it suffices to prove this:

Claim 2. *Let* $0 \leq l \leq n_0$. *Define the quantities* $d_{t,l}(\mu)$ *via the identity*

$$\left(\sum_{t=-\infty}^{\infty} q^{(t+a(\mu\pi))(1/2-s)} d_{t,l}(\mu)\right) = \omega_\pi(-1)\left(\sum_{a=0}^{\infty} W_\pi(a(\varpi^a)) q^{-a(1/2-s)} G(\varpi^{a-l}, \mu^{-1})\right) L(1-s, \mu^{-1}\omega_\pi^{-1}\pi)^{-1}. \quad (6)$$

If there exists $\mu \in \widetilde{X}$ *such that* $a(\mu) \leq l$ *and* $d_{t,l}(\mu) \neq 0$ *then* $t \geq -\max(n, l+m)$. *Further, if* $t = -\max(n, l+m)+r$ *with* $r \geq 0$, *then* $\sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)\leq l}} |d_{t,l}(\mu)|^2 \ll q^{-r/2}$.

We only consider the case $L(s,\pi)=1$, as the case $L(s,\pi)\neq 1$ is similar but easier. (Note that $L(s,\pi)\neq 1$ if and only if either $m=n$ or $n=1$.)

Let $\mu\in\widetilde{X}$ be such that $a(\mu)\le l$. As $L(s,\pi)=1$, we can use the well-known formulas stated in [Saha 2016, Equation (6) and Lemma 2.5] to deduce that the quantity on the RHS of (6) lying inside the bracket is a constant of absolute value $\ll q^{-l/2}$ if $a(\mu)=l$ or if $a(\mu)=0$ and $l=1$; and is equal to 0 otherwise. Furthermore, there are at most 2 characters $\mu\in\widetilde{X}$ with $a(\mu)\le n_0$ and $L(s,\mu^{-1}\omega_\pi^{-1}\pi)\neq 1$ (this can be checked, for example, using the classification written down in [Saha 2016, Section 2.2]).

We henceforth assume that $a(\mu)=l$ or $a(\mu)=0$ and $l=1$; else there is nothing to prove as $d_{t,l}(\mu)=0$. Suppose first that $L(s,\mu^{-1}\omega_\pi^{-1}\pi)=1$. Then, by equating coefficients on both sides of (6), we see that $d_{t,l}(\mu)\neq 0$ implies $t=-a(\mu\pi)\ge -\max(n,a(\mu)+m)\ge -\max(n,l+m)$, using Lemma 2.9. Furthermore if $t=-\max(n,l+m)+r$, then

$$\sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)\in\{l,0\}\\ L(s,\mu\pi)=1}} \left|d_{t,l}(\mu)\right|^2 \ll \sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)\in\{l,0\}\\ a(\mu\pi)=\max(n,l+m)-r}} q^{-l} \ll q^{-r/2},$$

again using Lemma 2.9.

Suppose next that $L(s,\mu^{-1}\omega_\pi^{-1}\pi)\neq 1$. In this case $\mu\neq 1$ so if $d_{t,l}(\mu)\neq 0$ we must have $a(\mu)=l$. Also, the right side of (6) is of the form $\alpha_0+\alpha_1 q^{-1(1/2-s)}+\alpha_2 q^{-2(1/2-s)}$ with $\alpha_i\ll q^{-(l+i)/2}$. Furthermore if $\alpha_2\neq 0$ then $a(\mu\pi)=0\le n-2$ and if $\alpha_2=0$ then $\alpha_1\neq 0$ and $a(\mu\pi)\le\max(n_0,m)\le n-1$. So again equating coefficients and using Lemma 2.9, we see that $d_{t,l}(\mu)\neq 0\Rightarrow t\ge -n\ge -\max(n,l+m)$, and furthermore if $t=-\max(n,l+m)+r$, then

$$\sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)=l\\ L(s,\mu\pi)\neq 1}} \left|d_{t,l}(\mu)\right|^2 \ll \sum_{i=0}^{2} \sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)=l\\ a(\mu\pi)=\max(n,l+m)-r-i}} q^{-l-i} \ll q^{-r/2},$$

again using Lemma 2.9.

Putting everything together, the proof of Claim 2 is complete. $\square$

Next, for any $g\in G$, define

$$n_0(g)=\min(l(g),n-l(g)), \quad \text{and} \quad q(g)=\max(n_0,n_0(g)-n_1+m).$$

We note the useful bounds $0\le n_0(g)\le n_0$ and $n_0\le q(g)\le n_0+m_1$.

Proposition 2.11. *Suppose that $g\in Ka(\varpi^{n_1})$. Assume further that either n is even or $l(g)\le n_0$.*

(1) *If for some $y\in F^\times$, we have $W_\pi(a(y)g)\neq 0$, then $v(y)\ge -q(g)$.*

(2) *Suppose $b = -q(g) + r$ where $r \geq 0$. Then we have*

$$\left(\int_{v \in \mathfrak{o}^\times} \left| W_\pi(a(\varpi^b v)g) \right|^2 d^\times v \right)^{1/2} \ll q^{-r/4}.$$

Proof. This follows immediately from Lemma 2.4 and Proposition 2.10. □

Remark 2.12. Note that the map on $\mathfrak{o}^\times$ given by $v \mapsto |W_\pi(a(vy)g)|$ is $U_{n_0(g)}$ invariant for all $y \in F^\times$ and $g \in G$. Hence the second part of Proposition 2.11 is equivalent to

$$\frac{1}{|\mathfrak{o}^\times / U_{n_0(g)}|} \sum_{v \in \mathfrak{o}^\times / U_{n_0(g)}} \left| W_\pi(a(\varpi^b v)g) \right|^2 d^\times v \ll q^{-r/2}.$$

2F. *Test functions.* We now change gears and start looking at certain local test functions (related to matrix coefficients) that will be used later in the trace formula. We begin with some definitions. Let $C_c^\infty(G, \omega_\pi^{-1})$ be the space of functions κ on G with the following properties:

(1) $\kappa(z(y)g) = \omega_\pi^{-1}(y)\kappa(g)$.

(2) κ is locally constant.

(3) $|\kappa|$ is compactly supported on $Z\backslash G$.

Given $\kappa_1, \kappa_2 \in C_c^\infty(G, \omega_\pi^{-1})$ we define the convolution $\kappa_1 * \kappa_2 \in C_c^\infty(G, \omega_\pi^{-1})$ via

$$(\kappa_1 * \kappa_2)(h) = \int_{Z\backslash G} \kappa_1(g^{-1})\kappa_2(gh)\, dg, \tag{7}$$

which turns $C_c^\infty(G, \omega_\pi^{-1})$ into an associative algebra.

Next let σ be a representation of G with central character equal to ω_π. Then, for any $\kappa \in C_c^\infty(G, \omega_\pi^{-1})$ and any vector $v \in \sigma$, we define $R(\kappa)v$ to be the vector in σ given by

$$R(\kappa)v = \int_{Z\backslash G} \kappa(g)(\sigma(g)v)\, dg. \tag{8}$$

Let v_π be any *newform* in the space of π, i.e., any nonzero vector fixed by $K_1(\mathfrak{p}^n)$. Equivalently v_π can be any vector in π corresponding to W_π under some isomorphism $\pi \simeq \mathcal{W}(\pi, \psi)$. Thus v_π is unique up to multiples. Put $v'_\pi = \pi(a(\varpi^{n_1}))v_\pi$. Note that v'_π is, up to multiples, the unique nonzero vector in π that is invariant under the subgroup $a(\varpi^{n_1})K_1(\mathfrak{p}^n)a(\varpi^{-n_1})$.

Let $\langle\, ,\, \rangle$ be any G-invariant inner product on π (this is also unique up to multiples). Define a matrix coefficient Φ_π on G as follows:

$$\Phi_\pi(g) = \frac{\langle v_\pi, \pi(g)v_\pi \rangle}{\langle v_\pi, v_\pi \rangle};$$

this is clearly independent of the choice of v_π or the normalization of the inner product. Further, let

$$K^0 := K^0(\mathfrak{p}^{n_1-n_0}) = \begin{cases} K & \text{if } n \text{ is even}, \\ K^0(\mathfrak{p}) & \text{if } n \text{ is odd}. \end{cases}$$

Put

$$\Phi'_\pi(g) = \begin{cases} \Phi_\pi(a(\varpi^{-n_1})ga(\varpi^{n_1})) = \dfrac{\langle v'_\pi, \pi(g)v'_\pi\rangle}{\langle v'_\pi, v'_\pi\rangle} & \text{if } g \in ZK^0, \\ 0 & \text{if } g \notin ZK^0. \end{cases}$$

Then it follows that $\Phi'_\pi \in C_c^\infty(G, \omega_\pi^{-1})$ and $\Phi'_\pi(g^{-1}) = \overline{\Phi'_\pi(g)}$. In particular, the operator $R(\Phi'_\pi)$ is self-adjoint.

Proposition 2.13. *There exists a positive real constant δ_π depending only on π and satisfying $\delta_\pi \gg q^{-n_1-m_1}$ such that the following hold*:

(1) $R(\Phi'_\pi)v'_\pi = \delta_\pi v'_\pi$.

(2) $\Phi'_\pi * \Phi'_\pi = \delta_\pi \Phi'_\pi$.

Remark 2.14. This is a refinement of a result of Marshall [2016] that holds in the special case $\omega_\pi = 1$; he used a slightly different test function which does not differentiate between n odd and even.

Remark 2.15. In fact with some additional work one can prove $\delta_\pi \asymp q^{-n_1-m_1}$.

The rest of this section will be devoted to proving this proposition. We note a useful corollary.

Corollary 2.16. *Let σ be a generic irreducible admissible unitarizable representation of G such that $\omega_\sigma = \omega_\pi$ and let v_σ be any vector in the space of σ. Suppose that $R(\Phi'_\pi)v_\sigma = \delta v_\sigma$ for some complex number δ. Then $\delta \in \{0, \delta_\pi\}$*; *in particular*, *$\delta$ is a nonnegative real number.*

Proof. We have

$$\delta\delta_\pi v_\sigma = \delta_\pi R(\Phi'_\pi)v_\sigma = R(\Phi'_\pi * \Phi'_\pi)v_\sigma = R(\Phi'_\pi)R(\Phi'_\pi)v_\sigma = \delta^2 v_\sigma,$$

implying that $\delta \in \{0, \delta_\pi\}$. □

2G. *Some preparatory lemmas.*

Lemma 2.17. *Consider the representation $\pi|_{K^0}$ of K^0 and let π' be the subrepresentation of $\pi|_{K^0}$ generated by v'_π. Then π' is a finite dimensional irreducible representation of K^0.*

Proof. We know that π' is isomorphic to a direct sum of irreducible representations of K^0. However if there were more than one summand in the decomposition of π', then the representation $\pi|_{K^0}$ (and hence the representation π) would contain a $a(\varpi^{n_1})K_1(\mathfrak{p}^n)a(\varpi^{-n_1})$-fixed subspace of dimension greater than one; by newform theory this is impossible. Hence π' is irreducible. The finite dimensionality of π' follows from the admissibility of π. $\square$

Lemma 2.18. *Let π' be as in Lemma 2.17. Then both claims of Proposition 2.13 hold with the quantity δ_π defined as follows*:

$$\delta_\pi = \int_{Z\backslash G} |\Phi'_\pi(g)|^2 \, dg = \int_{K^0} |\Phi'_\pi(g)|^2 \, dg = \frac{1}{[K : K^0] \, \dim(\pi')}.$$

Proof. Note that $\langle \, , \rangle$ is an invariant inner product for π'. It follows immediately (from the orthonormality of matrix coefficients) that the last two quantities are equal. The equality of the middle two quantities is immediate from our normalization of Haar measures.

We now show that this quantity satisfies the claims of Proposition 2.13. First of all, $R(\Phi'_\pi)v'_\pi$ is a vector in π that is invariant under the subgroup $a(\varpi^{n_1})K_1(\mathfrak{p}^n)a(\varpi^{-n_1})$. It follows that $R(\Phi'_\pi)v'_\pi = \delta v'_\pi$ for some constant δ. Taking inner products with v'_π immediately shows that $\delta = \delta_\pi$. This proves the first assertion of Proposition 2.13. The second assertion is a standard property of convolutions of matrix coefficients. $\square$

Proof of Proposition 2.13 in the case of nonsupercuspidal representations. For any nonsupercuspidal representation π it suffices to show that

$$\dim(\pi') \ll q^{n_0+m_1},$$

where π' is as in Lemma 2.17.

We can embed π inside a representation $\chi_1 \boxplus \chi_2$, consisting of smooth functions f on G satisfying

$$f\left(\left[\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right]g\right) = \left|\tfrac{a}{d}\right|^{1/2} \chi_1(a)\chi_2(d)f(g).$$

Here χ_1 and χ_2 are two (not necessarily unitary) characters. Let f' be the function in $\chi_1 \boxplus \chi_2$ that corresponds to v'_π. Let K' be the (normal) subgroup of K^0 consisting of matrices $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ such that $a \equiv d \equiv 1 \pmod{\mathfrak{p}^{n_0+m_1}}$, $b \equiv 0 \pmod{\mathfrak{p}^{n_1+m_1}}$ and $c \equiv 0 \pmod{\mathfrak{p}^{n_0+m_1}}$. Let $V_{K'}$ be the subspace of $\chi_1 \boxplus \chi_2$ consisting of the functions f that satisfy $f(gk) = \omega_\pi(a)f(g)$ for all $k = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in K'$. Then $f' \in V_{K'}$. Moreover (and this is the key fact!) if $k \in K^0$ and $k' \in K'$, then the top left entries of k' and $kk'k^{-1}$ (both these matrices are elements of K') are equal modulo $\mathfrak{p}^m$. Hence the space $V_{K'}$ is stable under the action of K^0. So it suffices to prove that $\dim(V_{K'}) \ll q^{n_0+m_1}$.

Using the Iwasawa decomposition, it follows that $|B(F)\backslash G(F)/K'| \asymp q^{n_0+m_1}$. Fix a set of double coset representatives S for $B(F)\backslash G(F)/K'$. Since any element of $V_{K'}$ is uniquely determined by its values on S, it follows that $\dim(V_{K'}) \ll q^{n_0+m_1}$. The proof is complete. $\square$

2H. ***Proof of Proposition 2.13 in the case of supercuspidal representations.*** We now assume that π is supercuspidal. In this case, $m \le n_0$, hence $m_1 = 0$. So it suffices to prove that

$$\int_{K^0} |\Phi_\pi(a(\varpi^{-n_1})ga(\varpi^{n_1}))|^2\, dg \gg q^{-n_1}. \tag{9}$$

The next proposition gives a formula for Φ_π, which may be of independent interest.

Proposition 2.19. *For* $0 \le l < n$, *we have*

$$\begin{aligned}\Phi_\pi(n(x)g_{t,l,v}) &= G(-\varpi^{l-n}, 1)G(\varpi^{t+l}v^{-1} - x, 1)\omega_\pi(-v)\delta_{t,-2l} \\ &\quad + \varepsilon\big(\tfrac{1}{2}, \pi\big)\omega_\pi(v) \sum_{\substack{\mu \in \widetilde{X} \\ a(\mu)=n-l \\ a(\mu\tilde{\pi})=n-2l-t}} G(\varpi^{l-n}, \mu)G(vx - \varpi^{t+l}, \mu)\varepsilon\big(\tfrac{1}{2}, \mu\tilde{\pi}\big).\end{aligned} \tag{10}$$

Proof. Using the usual inner product in the Whittaker model, and the fact that $W_\pi(a(t))$ is supported on $t \in \mathfrak{o}^\times$, (as π is supercuspidal) it follows that

$$\Phi_\pi(n(x)g_{t,l,v}) = \int_{\mathfrak{o}^\times} \psi(-ux)\omega_\pi(u)\overline{W_\pi(g_{t,l,vu^{-1}})}\, d^\times u. \tag{11}$$

On the other hand, by the formula for W_π in [Saha 2016, Proposition 2.30] and using Lemma 2.7 we have

$$\begin{aligned}W_\pi(a(\varpi^t)wn(\varpi^{-l}v)) &= \omega_\pi(-v^{-1})\psi(-\varpi^{t+l}v^{-1})G(\varpi^{l-n}, 1)\delta_{t,-2l} \\ &\quad + \Bigg(\varepsilon\big(\tfrac{1}{2}, \tilde{\pi}\big)\omega_\pi(v^{-1})\psi(-\varpi^{t+l}v^{-1}) \\ &\qquad \times \sum_{\substack{\mu \in \widetilde{X} \\ a(\mu)=n-l \\ a(\mu\pi)=n-t-2l}} G(\varpi^{l-n}, \mu^{-1})\varepsilon\big(\tfrac{1}{2}, \mu^{-1}\pi\big)\mu(-v).\Bigg)\end{aligned}$$

Substituting this into (11), we immediately get the required result. □

To obtain (9), we will need to substitute the formula from Proposition 2.19 and integrate. The following elementary lemma (which is similar to Lemma 2.6 of [Hu 2017]) will be useful; we omit its proof.

Lemma 2.20. *Let f be a function on G that is right $K_1(\mathfrak{p}^n)$-invariant. Then*

$$\int_G f(g)\, dg = \sum_{k=0}^{n} A_k \int_B f(bwn(\varpi^{-k}))\, db,$$

where $A_0 = (1+q^{-1})^{-1}$, $A_n = q^n(1+q^{-1})^{-1}$ *and, for* $0 < k < n$, $A_k = q^k(1-q^{-1})(1+q^{-1})^{-1}$.

We now complete the proof of (9). Using Lemma 2.20, it suffices to prove that

$$\int_{\substack{b\in B\\ bwn(\varpi^{-n_1})\in a(\varpi^{-n_1})K^0a(\varpi^{n_1})}} |\Phi_\pi(bwn(\varpi^{-n_1}))|^2\, db \gg q^{-2n_1}. \tag{12}$$

Now, note that the quantity $z(u)n(x)a(y)wn(\varpi^{-n_1})$ lies in $a(\varpi^{-n_1})K^0a(\varpi^{n_1})$ if and only if

$$u = \varpi^{n_1}u', \quad y = \varpi^{-2n_1}y' \quad \text{and} \quad x = \varpi^{-n_1}x',$$

for $y' \in \mathfrak{o}^\times$, $u' \in \mathfrak{o}^\times$, $x' \in \mathfrak{o}$ and $y' - x' \in \mathfrak{p}^{n_1-n_0}$. Hence the left side of (12) is equals

$$q^{-n_1}\int_{\substack{y'\in\mathfrak{o}^\times, x'\in\mathfrak{o}\\ x'\in y'+\mathfrak{p}^{n_1-n_0}}} |\Phi_\pi(n(\varpi^{-n_1}x')g_{-2n_1,-n_1,y'^{-1}})|^2\, dx'\, d^\times y'. \tag{13}$$

Now, we can exactly evaluate the integral in (13) using Proposition 2.19. We expand $|\Phi_\pi(n(\varpi^{-n_1}x')g_{-2n_1,n_1,y'^{-1}})|^2$ and observe that the main (diagonal) terms are simple to evaluate as we know the modulus-squared of Gauss sums. Indeed, the contribution to (13) from the diagonal terms is simply

$$q^{-n_1}\int_{\substack{y'\in\mathfrak{o}^\times,\ x'\in\mathfrak{o}\\ x'\in y'+\mathfrak{p}^{n_1-n_0}}} \sum_{\substack{\mu\in\widetilde{X}\\ a(\mu)=n_0\\ a(\mu\tilde{\pi})=n}} q^{-2n_0} \asymp q^{-2n_1}.$$

On the other hand, the contribution from the cross terms is zero. Indeed, each cross term involves an integral like

$$\int_{\substack{y'\in\mathfrak{o}^\times,\ x'\in\mathfrak{o}\\ y'^{-1}x'-1\in\mathfrak{p}^{n_1-n_0}\mathfrak{o}^\times}} \mu_1^{-1}\mu_2((y'^{-1}x'-1)\varpi^{n_0-n_1}),$$

which equals 0 because of the orthogonality of characters. This completes the proof of (12). □

3. Sup-norms of global newforms

From now on, we move to a global setup and consider newforms on $\mathrm{GL}_2(\mathbb{A})$ where $\mathbb{A}$ is the ring of adeles over $\mathbb{Q}$. For any place v of $\mathbb{Q}$, we will use the notation X_v for each *local object* X introduced in the previous section. The corresponding global objects will be typically denoted without the subscript v. The archimedean place will be denoted by $v = \infty$. We will usually denote a nonarchimedean place v by p, where p is a rational prime. The set of all nonarchimedean places (primes) will be denoted by $\boldsymbol{f}$.

We fix measures on all our adelic groups (like $\mathbb{A}$, $\mathrm{GL}_2(\mathbb{A})$, etc.) by taking the product of the local measures over all places (for the nonarchimedean places, these local measures were normalized in Section 2A; at the archimedean place we fix once and for all a suitable Haar measure). We normalize the Haar measure on $\mathbb{R}$ to be the usual Lebesgue measure. We give all discrete groups the counting measure and thus obtain a measure on the appropriate quotient groups.

3A. ***Statement of result.*** As usual, let $G = \mathrm{GL}_2$. Let $\pi = \otimes_v \pi_v$ be an irreducible, unitary, cuspidal automorphic representation of $G(\mathbb{A})$ with central character $\omega_\pi = \prod_v \omega_{\pi_v}$. For each prime p, let the integers n_p, $n_{1,p}$, $n_{0,p}$, m_p, and $m_{1,p}$ be defined as in Section 2B. We put $N = \prod_p p^{n_p}$, $N_0 = \prod_p p^{n_{0,p}}$, $N_1 = \prod_p p^{n_{1,p}}$, $M = \prod_p p^{m_p}$, and $M_1 = \prod_p p^{m_{1,p}}$. Thus, N is the conductor of π, M is the conductor of ω_π, N_0 is the largest integer such that $N_0^2 \mid N$, and $N_1 = N/N_0$ is the smallest integer such that $N \mid N_1^2$. Let $N_2 = N_1/N_0 = N/N_0^2$. Note that N_2 is a squarefree integer and is the product of all the primes p such that p divides N to an odd power. If N is squarefree, then $N_2 = N_1 = N$ and $N_0 = 1$ while if N is a perfect square then $N_0 = N_1 = \sqrt{N}$ and $N_2 = 1$. Note also that $M_1 = M/\gcd(M, N_1)$.

We assume that π_∞ is a spherical principal series representation whose central character is trivial on $\mathbb{R}^+$. This means that $\pi_\infty \simeq \chi_1 \boxplus \chi_2$,[8] where for $i = 1, 2$, we have $\chi_1 = |y|^{it} \operatorname{sgn}(y)^m$, $\chi_2 = |y|^{-it} \operatorname{sgn}(y)^m$, with $m \in \{0, 1\}$, and $t \in \mathbb{R} \cup \left(-\frac{i}{2}, \frac{i}{2}\right)$.

Let $K_1(N) = \prod_{p \in f} K_{1,p}(p^{n_p}) = \prod_{p \nmid N} G(\mathbb{Z}_p) \prod_{p \mid N} K_{1,p}(p^{n_p})$ be the standard congruence subgroup of $G(\widehat{\mathbb{Z}}) = \prod_{p \in f} G(\mathbb{Z}_p)$; note that $K_1(N) G(\mathbb{R})^+ \cap G(\mathbb{Q})$ is equal to the standard congruence subgroup $\Gamma_1(N)$ of $\mathrm{SL}_2(\mathbb{Z})$. Let $K_\infty = \mathrm{SO}_2(\mathbb{R})$ be the maximal connected compact subgroup of $G(\mathbb{R})$ (equivalently, the maximal compact subgroup of $G(\mathbb{R})^+$). We say that a nonzero automorphic form $\phi \in V_\pi$ is a *newform* if ϕ is $K_1(N)K_\infty$-invariant. It is well-known that a newform ϕ exists and is unique up to multiples, and corresponds to a factorizable vector $\phi = \otimes_v \phi_v$. We define

$$\|\phi\|_2 = \int_{Z(\mathbb{A})G(F)\backslash G(\mathbb{A})} |\phi(g)|^2 \, dg.$$

Remark 3.1. If ϕ is a newform, then the function f on $\mathbb{H}$ defined by $f(g(i)) = \phi(g)$ for each $g \in \mathrm{SL}_2(\mathbb{R})$ is a Hecke–Maass cuspidal newform of level N (and character ω_π). Precisely, it satisfies the relation

$$f\left(\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] z\right) = \left(\prod_{p \mid N} \omega_{\pi,p}(d)\right) f(z), \quad \text{for all } \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \Gamma_0(N). \tag{14}$$

[8] For two characters χ_1 and χ_2 on $\mathbb{R}^\times$, we let $\chi_1 \boxplus \chi_2$ denote the principal series representation on $G(\mathbb{R})$ that is unitarily induced from the corresponding representation of $B(\mathbb{R})$; this consists of smooth functions f on $G(\mathbb{R})$ satisfying

$$f\left(\left[\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right] g\right) = \left|\tfrac{a}{d}\right|^{1/2} \chi_1(a) \chi_2(d) f(g).$$

The Laplace eigenvalue λ for f is given by $\lambda = \frac{1}{4} + t^2$ where t is as above. (Note that $\lambda \asymp (1+|t|)^2$.)

Furthermore, any Hecke–Maass cuspidal newform f is obtained in the above manner from a newform ϕ in a suitable automorphic representation π. The newform ϕ can be directly constructed from f via strong approximation. It is clear that $\sup_{g\in G(\mathbb{A})}|\phi(g)| = \sup_{z\in\Gamma_0(N)\backslash\mathbb{H}}|f(z)|$.

Our main result is as follows.

Theorem 3.2. *Let π be an irreducible, unitary, cuspidal automorphic representation of $G(\mathbb{A})$ such that $\pi_\infty \simeq \chi_1 \boxplus \chi_2$, where, for $i = 1, 2$, we have $\chi_1 = |y|^{it}\,\mathrm{sgn}(y)^m$ and $\chi_2 = |y|^{-it}\,\mathrm{sgn}(y)^m$, with $m \in \{0, 1\}$ and $t \in \mathbb{R} \cup \left(-\frac{i}{2}, \frac{i}{2}\right)$. Let the integers N_0, N_1 and M_1 be defined as above and let $\phi \in V_\pi$ be a newform satisfying $\|\phi\|_2 = 1$. Then*

$$\sup_{g\in G(\mathbb{A})} |\phi(g)| \ll_\varepsilon N_0^{1/6+\varepsilon} N_1^{1/3+\varepsilon} M_1^{1/2} (1+|t|)^{5/12+\varepsilon}.$$

Remark 3.3. If π has trivial central character and $N_1 \asymp \sqrt{N}$ (this is the case whenever N is sufficiently "powerful"), then we get $\sup_{g\in G(\mathbb{A})}|\phi(g)| \ll_{t,\varepsilon} N^{1/4+\varepsilon}$, which is a considerable improvement over the best previously known result [Saha 2014] $\sup_{g\in G(\mathbb{A})}|\phi(g)| \ll_{t,\varepsilon} N^{5/12+\varepsilon}$, due to the author.

3B. ***Atkin–Lehner operators and a generating domain.*** Let π be as in Section 3A and $\phi \in V_\pi$ a newform. In order to prove Theorem 3.2 we will restrict the variable g to a carefully chosen generating domain inside $G(\mathbb{A})$. In order to do this, we will have to consider the newform ϕ along with some of its Atkin–Lehner translates. The object of this section is to explain these ideas and describe our generating domain. The main result in this context is Proposition 3.6 below.

We begin with some definitions. For any integer L, let $\mathcal{P}(L)$ denote the set of distinct primes dividing L. For any subset S of $\mathcal{P}(N)$, let $\eta_S, h_S \in G(\mathbb{A}_f)$ be defined as follows: $\eta_{S,p} = \left[\begin{smallmatrix} & 1 \\ p^{n_p} & \end{smallmatrix}\right]$ if $p \in S$ and $\eta_{S,p} = 1$ otherwise; $h_{S,p} = a(p^{n_{1,p}})$ if $p \in S$ and $h_{S,p} = 1$ otherwise. Define

$$K_S = \prod_{p\in S} G(\mathbb{Z}_p) \subset G(\mathbb{A}_f), \qquad J_S = K_S h_S \subset G(\mathbb{A}_f).$$

Finally, define

$$\mathcal{J}_S = \{g \in J_S : l(g_p) \le n_{0,p} \text{ for all } p \in S \text{ such that } n_p \text{ is odd}\}.$$

Using Lemma 2.2, we see that $g \in \prod_{p\in S} G(\mathbb{Q}_p)$ belongs to $\mathcal{J}_S$ if and only if $g_p \in wK_p^0(p)a(p^{n_{1,p}})$ for all $p \in S$ for which n_p is odd. If L divides N, we abuse notation by denoting

$$h_L = h_{\mathcal{P}(L)}, \quad K_L = K_{\mathcal{P}(L)}, \quad J_L = J_{\mathcal{P}(L)} \quad \text{and} \quad \mathcal{J}_L = \mathcal{J}_{\mathcal{P}(L)}.$$

For any $0 < c < \infty$, let D_c be the subset of $B_1(\mathbb{R})^+ \simeq \mathbb{H}$ defined by

$$D_c := \{n(x)a(y) : x \in \mathbb{R},\, y \geq c\}.$$

Finally, for $L > 0$, define

$$\mathcal{F}_L = \{n(x)a(y) \in D_{\sqrt{3}/(2L)} : z = x + iy \text{ satisfies } |cz+d|^2 \geq 1/L,\, \forall (0,0) \neq (c,d) \in \mathbb{Z}^2\}.$$

Next, for any subset S of $\mathcal{P}(N)$, let $\omega_\pi^S = \prod_v \omega_{\pi,v}^S$ be the unique character[9] on $\mathbb{Q}^\times \backslash \mathbb{A}^\times$ with the following properties:

(1) $\omega_{\pi,\infty}^S$ is trivial on $\mathbb{R}^+$.

(2) $\omega_{\pi,p}^S|_{\mathbb{Z}_p^\times}$ is trivial if $p \in S$ and equals $\omega_{\pi,p}|_{\mathbb{Z}_p^\times}$ if $p \notin S$.

Note that $\omega_\pi^{\mathcal{P}(N)} = 1$, $\omega_\pi^\varnothing = \omega_\pi$ and, for each S, ω_π^S has conductor $\prod_{p \notin S} p^{m_p}$. Define the irreducible, unitary, cuspidal automorphic representation π^S by $\pi^S = \tilde{\pi} \otimes \omega_\pi^S = \pi \otimes (\omega_\pi^{-1}\omega_\pi^S)$. A key observation is that for every S, the representation π^S has conductor N and its central character $\omega_{\pi^S} = \omega_\pi^{-1}(\omega_\pi^S)^2$ has conductor M. We have $\pi^\varnothing = \pi$ and $\pi^{\mathcal{P}(N)} = \tilde{\pi}$.

Lemma 3.4. *The function ϕ^S on $G(\mathbb{A})$ given by $\phi^S(g) := (\omega_\pi^{-1}\omega_\pi^S)(\det(g))\phi(g\eta_S)$ is a newform in π^S.*

Proof. It is clear that ϕ^S is a vector in π^S, and one can easily check from the defining relation that it is $K_1(N)K_\infty$ invariant. □

Remark 3.5. In the special case $\omega_\pi = 1$, one has $\pi^S = \pi$ for every subset S of $\mathcal{P}(N)$. In this case, for each S, the involution $\pi(\eta_S)$ on V_π corresponds to a classical Atkin–Lehner operator, and $\phi^S = \pm\phi$ with the sign equal to the Atkin–Lehner eigenvalue. We will call the natural map on $Z(\mathbb{A})G(\mathbb{Q})\backslash G(\mathbb{A})/K_1(N)K_\infty$ induced by $g \mapsto g\eta_S$ the adelic Atkin–Lehner operator associated to S.

Recall that $\mathcal{J}_N = \prod_{p \mid N_2} wK_p^0(p)a(p^{n_{1,p}}) \prod_{p \mid N,\ p \nmid N_2} G(\mathbb{Z}_p)a(p^{n_{1,p}}) \subset G(\mathbb{A}_f)$. The next proposition tells us that any point in $Z(\mathbb{A})G(\mathbb{Q})\backslash G(\mathbb{A})/K_1(N)K_\infty$ can be moved by an adelic Atkin–Lehner operator to a point whose finite part lies in $\mathcal{J}_N$ and whose infinite component lies in $\mathcal{F}_{N_2}$.

Proposition 3.6. *Suppose that $g \in G(\mathbb{A})$. Then there exists a subset S of $\mathcal{P}(N_2)$ such that*

$$g \in Z(\mathbb{A})G(\mathbb{Q})(\mathcal{J}_N \times \mathcal{F}_{N_2})\eta_S K_1(N)K_\infty.$$

Proof. Let w_N be the diagonal embedding of $w = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ into K_N. The determinant map from $w_N h_N K_1(N) h_N^{-1} w_N^{-1}$ is surjective onto $\prod_p \mathbb{Z}_p^\times$. Hence by strong

[9]The existence, as well as uniqueness, of the character ω_π^S follows from the identity $\mathbb{A}^\times = \mathbb{Q}^\times \mathbb{R}^+ \prod_p \mathbb{Z}_p^\times$.

approximation for $gh_N^{-1}w_N^{-1}$, we can write $gh_N^{-1}w_N^{-1} = zg_{\mathbb{Q}}g_\infty^{+}(w_N h_N k h_N^{-1} w_N^{-1})$ where $z \in Z(\mathbb{A})$, $g_{\mathbb{Q}} \in G(\mathbb{Q})$, $g_\infty^{+} \in G(\mathbb{R})^{+}$ and $k \in K_1(N)$. In other words,

$$g \in Z(\mathbb{A})G(\mathbb{Q})g_\infty^{+} w_N h_N K_1(N). \tag{15}$$

Using Lemma 1 from [Harcos and Templier 2012], we can find a divisor N_3 of N_2, and a matrix $W \in M_2(\mathbb{Z})$ such that

$$W \equiv \left[\begin{smallmatrix} 0 & * \\ 0 & 0 \end{smallmatrix}\right] \bmod N_3, \quad W \equiv \left[\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right] \bmod N_2, \quad \det(W) = N_3, \quad W_\infty g_\infty^{+} \in \mathcal{F}_{N_2} K_\infty.$$

W_∞ denotes the element W considered as an element of $G(\mathbb{R})^{+}$. Let S be the set of primes dividing N_3. Note that $W_p \in K_{0,p}(p)\left[\begin{smallmatrix} 0 & 1 \\ p & 0 \end{smallmatrix}\right]$ if $p \in S$, $W_p \in K_{0,p}(p)$ if $p \mid N_2$ but $p \notin S$, and $W_p \in G(\mathbb{Z}_p)$ if $p \nmid N_2$. Since $W \in G(\mathbb{Q})$, it follows from above and from (15) that g is an element of

$$\begin{aligned} Z(\mathbb{A})G(\mathbb{Q})\mathcal{F}_{N_2}K_\infty \Bigl(\prod_{p \in S} K_{0,p}(p)\left[\begin{smallmatrix} 0 & 1 \\ p & 0 \end{smallmatrix}\right]w\Bigr)\Bigl(\prod_{\substack{p \mid N_2 \\ p \notin S}} K_{0,p}(p)w\Bigr)\Bigl(\prod_{\substack{p \mid N \\ p \nmid N_2}} G(\mathbb{Z}_p)\Bigr)h_N K_1(N) \\ = Z(\mathbb{A})G(\mathbb{Q})(\mathcal{J}_N \times \mathcal{F}_{N_2})\eta_S K_1(N) K_\infty, \end{aligned}$$

where in the last step we have used Lemma 2.3. □

Corollary 3.7. *Let π and ϕ be as in Theorem 3.2. Suppose that for all subsets S of $\mathcal{P}(N_2)$ and all $g \in \mathcal{J}_N$, $n(x)a(y) \in \mathcal{F}_{N_2}$, we have*

$$|\phi^S(gn(x)a(y))| \ll_\varepsilon N_1^{1/2+\varepsilon} M_1^{1/2} N_2^{-1/6}(1+|t|)^{5/12+\varepsilon}.$$

Then the conclusion of Theorem 3.2 is true.

Proof. This follows from Proposition 3.6 and the fact that

$$|\phi^S(gn(x)a(y))| = |\phi(gn(x)a(y)\eta_S)|. \qquad \square$$

3C. ***Sketch of proof modulo technicalities.*** We now prove Theorem 3.2 assuming some key bounds whose proofs will take the rest of this paper. For brevity we put $T = 1+|t|$. Also, recall that $N_2 = N_1/N_0$. We need to show that for each $g \in G(\mathbb{A})$,

$$|\phi(g)| \ll_\varepsilon N_1^{1/2+\varepsilon} M_1^{1/2} N_2^{-1/6} T^{5/12+\varepsilon}.$$

By letting ϕ run over all its various Atkin–Lehner translates ϕ^S, for $S \subseteq \mathcal{P}(N_2)$, we may assume (by Corollary 3.7) that $g \in \mathcal{J}_N \mathcal{F}_{N_2}$. Therefore, in what follows, we will not explicitly keep track of the set S, but instead prove the following: *Given an automorphic representation π as in Section 3A* (*with associated quantities N_1, N_2, T and M_1 as defined earlier*), *a newform $\phi \in V_\pi$ satisfying $\|\phi\|_2 = 1$, and elements $g \in \mathcal{J}_N$ and $n(x)a(y) \in \mathcal{F}_{N_2}$, we have*

$$|\phi(gn(x)a(y))| \ll_\varepsilon N_1^{1/2+\varepsilon} M_1^{1/2} N_2^{-1/6} T^{5/12+\varepsilon}. \tag{16}$$

As noted, the above statement implies Theorem 3.2. Implicit here is the fact that we are letting π vary among the various π^S, which all have exactly the same values of N, N_1, N_2, M_1 and T as π, and moreover the corresponding newforms ϕ^S all satisfy $\|\phi^S\|_2 = \|\phi\|_2$.

We prove (16) by a combination of two methods. In Proposition 3.8, we use the Whittaker expansion to bound this quantity. Precisely, we prove the following bound:

$$|\phi(gn(x)a(y))| \ll_\varepsilon (NT)^\varepsilon \Big(\Big(\frac{N_1 M_1 T}{N_2 y} \Big)^{1/2} + \Big(\frac{N_1 T^{1/3}}{N_2} \Big)^{1/2} \Big). \tag{17}$$

To prove (17) we rely on Proposition 2.11. Next, in Proposition 3.16, we use the amplification method to bound this quantity. We prove that for each $\Lambda \geq 1$, we have

$$\begin{aligned} &|\phi(gn(x)a(y))|^2 \\ &\ll_\varepsilon (NT\Lambda)^\varepsilon N_1 M_1 \Big[\frac{T + N_2^{1/2} T^{1/2} y}{\Lambda} + \Lambda^{1/2} T^{1/2} (N_2^{-1/2} + y) + \Lambda^2 T^{1/2} N_2^{-1} \Big]. \end{aligned} \tag{18}$$

The proof of this bound relies on Proposition 2.13 and some counting arguments due to Harcos and Templier. Combining the two bounds leads to Theorem 3.2, as we explain now.

Choose $\Lambda = T^{1/6} N_2^{1/3}$. Then (18) becomes

$$|\phi(gn(x)a(y))|^2 \ll_\varepsilon (NT)^\varepsilon N_1 M_1 [T^{5/6} N_2^{-1/3} + T^{7/12} N_2^{-1/6} y]. \tag{19}$$

If $y \leq T^{1/4} N_2^{-1/6}$, then we use (19) to immediately deduce (16). If $y \geq T^{1/4} N_2^{-1/6}$, then we use (17) to obtain the bound

$$|\phi(gn(x)a(y))| \ll_\varepsilon (NT)^\varepsilon M_1^{1/2} N_1^{1/2} N_2^{-5/12} T^{3/8}, \tag{20}$$

which is much stronger than (16)! This completes the proof.

3D. *The bound via the Whittaker expansion.* Let π and ϕ be as in Section 3A with $\|\phi\|_2 = 1$. The object of this section is to prove the following result.

Proposition 3.8. *Let $x \in \mathbb{R}$, $y \in \mathbb{R}^+$ and $g \in \mathcal{J}_N$. Then*

$$|\phi(gn(x)a(y))| \ll_\varepsilon (NT)^\varepsilon \Big(\Big(\frac{N_0 M_1 T}{y} \Big)^{1/2} + (N_0 T^{1/3})^{1/2} \Big).$$

Remark 3.9. If we assume Conjecture 1, then we can improve the bound in Proposition 3.8 to

$$(NT)^\varepsilon \Big(\Big(\frac{N_0 M_1 T}{y} \Big)^{1/2} + (T^{1/3})^{1/2} \Big).$$

We now begin the proof of Proposition 3.8. One has the usual Fourier expansion at infinity

$$\phi(n(x)a(y)) = y^{1/2} \sum_{n \in \mathbb{Z}_{\neq 0}} \rho_\phi(n) K_{it}(2\pi |n| y) e(nx). \tag{21}$$

Lemma 3.10 notes some key properties about the Fourier coefficients in (21).

Lemma 3.10. *The Fourier coefficients $\rho_\phi(n)$ satisfy the following properties*:

(1) $|\rho_\phi(n)| = |\rho_\phi(1)\lambda_\pi(n)|$ *where* $\lambda_\pi(n)$ *are the coefficients of the* L*-function of* π.

(2) $|\rho_\phi(1)| \ll_\varepsilon (NT)^\varepsilon e^{\pi t/2}$.

(3) $\sum_{1\le|n|\le X} |\lambda_\pi(n)|^2 \ll X(NTX)^\varepsilon$.

Proof. All the parts are standard. The first part is a basic well-known relation between the Fourier coefficients and Hecke eigenvalues. The second part is due to Hoffstein and Lockhart [1994]. The last part follows from the analytic properties of the Rankin–Selberg L-function (e.g., see [Harcos and Michel 2006]). □

The Fourier expansion (21) is a special case of the more general Whittaker expansion that we describe now. Let $g_f \in G(\mathbb{A}_f)$. Then the Whittaker expansion for ϕ says that

$$\phi(g_f n(x)a(y)) = \sum_{q\in\mathbb{Q}_{\neq 0}} W_\phi(a(q)g_f n(x)a(y)), \tag{22}$$

where W_ϕ is a global Whittaker newform corresponding to ϕ given explicitly by

$$W_\phi(g) = \int_{x\in\mathbb{A}/\mathbb{Q}} \phi(n(x)g)\psi(-x)\,dx.$$

Putting $g_f = 1$ in (22) gives us the expansion (21). On the other hand, the function W_ϕ factors as $W_\phi(g) = c\prod_v W_v(g_v)$ where

(1) $W_p = W_{\pi_p}$ at all finite primes p,

(2) $|W_\infty(a(q)n(x)a(y))| = |qy|^{1/2}|K_{it}(2\pi|q|y)|$.

The constant c is related to $L(1,\pi,\mathrm{Ad})$; for further details on this constant, see [Saha 2016, Section 3.4].

For any $g = \prod_{p\mid N} g_p \in \mathcal{J}_N$, define

$$N_0^g = \prod_{p\mid N} p^{n_0(g_p)} \quad\text{and}\quad Q^g = \prod_{p\mid N} p^{q(g_p)},$$

where the integers $n_0(g_p)$ and $q(g_p)$ are as defined just before Proposition 2.11. Note that the "useful bounds" stated there imply that $N_0^g \mid N_0$ and $Q^g \mid N_0 M_1$.

Lemma 3.11. *Suppose that* $g \in \mathcal{J}_N$ *and* $W_\phi(a(q)gn(x)a(y)) \neq 0$ *for some* $q \in \mathbb{Q}$. *Then we have* $q = n/Q^g$ *for some* $n \in \mathbb{Z}$.

Proof. We have $W_{\pi_p}(a(q)g_p) \neq 0$ for each $p \mid N$ and $W_{\pi_p}(a(q)) \neq 0$ for each $p \nmid N$. Now the result follows from Proposition 2.11 and Lemma 2.6. □

Henceforth we fix some $g \in \mathcal{J}_N$. By comparing the expansion (22) for $g_f = g$ with the trivial case $g_f = 1$, we conclude that

$$\begin{aligned}
&\phi(gn(x)a(y)) \\
&\quad = \sum_{n\in\mathbb{Z}_{\neq 0}} W_\phi(a(n/Q^g)gn(x)a(y)) \\
&\quad = \sum_{n\in\mathbb{Z}_{\neq 0}} \Big(\prod_{p\mid N} W_{\pi_p}(a(n/Q^g)g)\Big)\Big(c\prod_{p\nmid N} W_{\pi_p}(a(n))\Big) W_\infty(a(n/Q^g)n(x)a(y)) \\
&\quad = \Big(\frac{y}{Q^g}\Big)^{1/2} \sum_{n\in\mathbb{Z}_{\neq 0}} (|n|, N^\infty)^{1/2} \rho_\phi\Big(\frac{n}{(|n|, N^\infty)}\Big) \lambda_{\pi_N}(n; g) K_{it}\Big(\frac{2\pi|n|y}{Q^g}\Big)\chi_n, \qquad (23)
\end{aligned}$$

where χ_n is some complex number of absolute value 1, and for each nonnegative integer n we define

$$\lambda_{\pi_N}(n; g) := \prod_{p\mid N} W_{\pi_p}(a(np^{-q(g_p)})g_p).$$

The tail of the sum (23) consisting of the terms with $2\pi|n|y/Q^g > T + T^{1/3+\varepsilon}$ is negligible because of the exponential decay of the Bessel function. Put

$$R = Q^g\Big(\frac{T+T^{1/3+\varepsilon}}{2\pi y}\Big) \asymp \frac{Q^g T}{y}.$$

Using the Cauchy–Schwarz inequality and Lemma 3.10, we therefore have

$$\begin{aligned}
|\phi(gn(x)a(y))|^2 \ll_\varepsilon (NT)^\varepsilon e^{\pi t}\Big(\frac{y}{Q^g}\Big)\Big(\sum_{0<n<R} (|n|, N^\infty)\Big|\lambda_\pi\Big(\frac{n}{(|n|, N^\infty)}\Big)\Big|^2\Big) \\
\times \Big(\sum_{0<n<R} \Big|\lambda_{\pi_N}(n; g) K_{it}\Big(\frac{2\pi|n|y}{Q^g}\Big)\Big|^2\Big). \qquad (24)
\end{aligned}$$

Lemma 3.12. *The function $\lambda_{\pi_N}(n; g)$ satisfies the following properties*:

(1) *Suppose that n_1 is a positive integer such that $n_1 \mid N^\infty$, and n_0, n_0' are two integers coprime to N such that $n_0 \equiv n_0' \pmod{N_0^g}$. Then*

$$|\lambda_{\pi_N}(n_0 n_1; g)| = |\lambda_{\pi_N}(n_0' n_1; g)|.$$

(2) *For any integer r and any $n_1 \mid N^\infty$,*

$$\sum_{\substack{rN_0^g \le |n_0| < (r+1)N_0^g \\ (n_0, N)=1}} |\lambda_{\pi_N}(n_0 n_1; g)|^2 \ll N_0^g n_1^{-1/2}.$$

Proof. Let $p \mid N$ and $u_1, u_2 \in \mathbb{Z}_p^\times$. Then by (3) it follows that, for all $w \in \mathbb{Q}_p^\times$,

$$|W_{\pi_p}(a(wu_1)g_p)| = |W_{\pi_p}(a(wu_2)g_p)|,$$

whenever $u_1 \equiv u_2 \mod (p^{n_0(g_p)})$. It follows that if $n_1 \mid N^\infty$, then

$$|\lambda_{\pi_N}(n_0 n_1; g)| = |\lambda_{\pi_N}(n_0' n_1; g)| \quad \text{if } n_0 \equiv n_0' \pmod{N_0^g}. \qquad (25)$$

Furthermore using the above and the Chinese remainder theorem,

$$\frac{1}{N_0^g}\sum_{\substack{n_0 \bmod N_0^g\\(n_0,N)=1}}|\lambda_{\pi_N}(n_0n_1;g)|^2=\prod_{p\mid N}\biggl(\int_{\mathbb{Z}_p^\times}\bigl|W_{\pi_p}(a(n_1vp^{-q(g_p)})g_p)\bigr|^2\,d^\times v\biggr),$$

and hence by Proposition 2.11,

$$\frac{1}{N_0^g}\sum_{\substack{n_0 \bmod N_0^g\\(n_0,N)=1}}|\lambda_{\pi_N}(n_0n_1;g)|^2\ll n_1^{-1/2}. \qquad \square$$

Lemma 3.13. *We have*

$$\sum_{0<n<R}te^{\pi t}\Bigl|\lambda_{\pi_N}(n;g)K_{it}\Bigl(\frac{2\pi|n|y}{Q^g}\Bigr)\Bigr|^2\ll_\epsilon (NT)^\varepsilon\Bigl(N_0^gT^{1/3}+\frac{Q^gT}{y}\Bigr).$$

Remark 3.14. If we assume Conjecture 1, then the bound on the right side can be improved to $(NT)^\varepsilon\bigl(T^{1/3}+(Q^gT)/y\bigr)$.

Proof. Let $f(y)=\min(T^{1/3},|y/T-1|^{-1/2})$. Then it is known that

$$te^{\pi t}|K_{it}(y)|^2\ll f(y),$$

see, e.g., [Templier 2015, (3.1)]. Using the previous lemma, we may write

$$\begin{aligned}
te^{\pi t}&\sum_{0<n<R}\Bigl|\lambda_{\pi_N}(n;g)K_{it}\Bigl(\frac{2\pi|n|y}{Q^g}\Bigr)\Bigr|^2\\
&\ll\sum_{\substack{1\le n_1\le R\\ n_1\mid N^\infty}}\ \sum_{\substack{1\le|n_0|\le R/n_1\\(n_0,N)=1}}|\lambda_{\pi_N}(n_0n_1;g)|^2f\Bigl(\frac{2\pi|n_0n_1|y}{Q^g}\Bigr)\\
&\ll\sum_{\substack{1\le n_1\le R\\ n_1\mid N^\infty}}\ \sum_{0\le r\le\lfloor R/(n_1N_0^g)\rfloor}\ \sum_{\substack{rN_0^g\le|n_0|\le(r+1)N_0^g\\(n_0,N)=1}}|\lambda_{\pi_N}(n_0n_1;g)|^2f\Bigl(\frac{2\pi|n_0n_1|y}{Q^g}\Bigr)\\
&\ll N_0^g\sum_{\substack{1\le n_1\le R\\ n_1\mid N^\infty}}n_1^{-1/2}\sum_{0\le r\le\lfloor R/(n_1N_0^g)\rfloor}f\Bigl(\frac{2\pi|n_0^{(r)}n_1|y}{Q^g}\Bigr)
\end{aligned}$$

(where $n_0^{(r)}\in[rN_0^g,(r+1)N_0^g]$ is the point where $f(2\pi n_0^{(r)}n_1y/Q^g)$ is maximum, and where we have used Lemma 3.12)

$$\ll\sum_{\substack{1\le n_1\le R\\ n_1\mid N^\infty}}n_1^{-1/2}N_0^g\biggl(T^{1/3}+\int_0^{R/(N_0^gn_1)}f\Bigl(\frac{2\pi N_0^grn_1y}{Q^g}\Bigr)dr\biggr)$$

(as f has $\ll 1$ turning points)

$$\ll \sum_{\substack{1\le n_1\le R\\ n_1\mid N^\infty}} \Big(N_0^g T^{1/3} n_1^{-1/2} + n_1^{-3/2} \frac{Q^g}{y} \int_0^{T+T^{1/3+\varepsilon}} \Big|\frac{s}{T}-1\Big|^{-1/2} ds \Big)$$

$$\ll_\varepsilon (NT)^\varepsilon \Big(N_0^g T^{1/3} + \frac{Q^g T}{y} \Big). \qquad \square$$

Lemma 3.15. *For all $X > 0$, we have*

$$\sum_{0<n<X} (|n|, N^\infty) \Big| \lambda_\pi \Big(\frac{n}{(|n|, N^\infty)} \Big) \Big|^2 \ll_\epsilon (NTX)^\epsilon X.$$

Proof. This follows from the last part of Lemma 3.10 using a similar (but simpler) argument to Lemma 3.13. $\square$

Finally, by combining (24), Lemma 3.13 and Lemma 3.15, we get the bound

$$|\phi(gn(x)a(y))|^2 \ll_\varepsilon (NT)^\varepsilon \Big(\frac{Q^g T}{y} + N_0^g T^{1/3} \Big). \tag{26}$$

Taking square roots, and using that $Q^g \le N_0 M_1$ and $N_0^g \le N_0$, we get the conclusion of Proposition 3.8.

3E. *Preliminaries on amplification.* Our aim for the rest of this paper is to prove the following proposition. As explained in Section 3C, this will complete the proof of our main result.

Proposition 3.16. *Let $\Lambda \ge 1$ be a real number. Let $n(x)a(y) \in \mathcal{F}_{N_2}$, $g \in \mathcal{J}_N$. Then*

$$|\phi(gn(x)a(y))|^2$$
$$\ll_\varepsilon (\Lambda NT)^\varepsilon N_1 M_1 \bigg[\frac{T + N_2^{1/2} T^{1/2} y}{\Lambda} + \Lambda^{1/2} T^{1/2} (N_2^{-1/2} + y) + \Lambda^2 T^{1/2} N_2^{-1} \bigg]. \tag{27}$$

Recall that $h_N = \prod_{p\mid N} a(p^{n_{1,p}})$. Define the vector $\phi' \in V_\pi$ by

$$\phi'(g) = \phi(gh_N).$$

Then the problem becomes equivalent to bounding the quantity $\phi'(k_N n(x)a(y))$ where $k_N \in K_N = \prod_{p\mid N} G(\mathbb{Z}_p)$ and $k_N h_N \in \mathcal{J}_N$. Note that ϕ' is $K_1'(N)K_\infty$-invariant where $K_1'(N) := h_N K_1(N) h_N^{-1}$.

Define the function Φ_N' on $\prod_{p\mid N} G(\mathbb{Q}_p)$ by $\Phi_N' = \prod_{p\mid N} \Phi_{\pi_p}'$, with the functions Φ_{π_p}' defined in Section 2F. By Proposition 2.13, it follows that

$$R(\Phi_N')\phi' := \int_{(Z\backslash G)(\prod_{p\mid N}\mathbb{Q}_p)} \Phi_N'(g)(\pi(g)\phi')\, dg = \delta_N \phi',$$

where $\delta_N \gg N_1^{-1}M_1^{-1}$. Note also that if $g \in \prod_{p\mid N} G(\mathbb{Q}_p)$ and $\Phi'_N(g) \neq 0$, then $g \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)$ for each prime p dividing N and $g \in Z(\mathbb{Q}_p)K_p^0(p)$ for each prime p dividing N_2. Also, recall that $R(\Phi'_N)$ is a self-adjoint, essentially idempotent operator.

Next, we consider the primes not dividing N. Let $\mathcal{H}_{\mathrm{ur}}$ be the usual global (unramified) convolution Hecke algebra; it is generated by the set of all functions κ_{ur} on $\prod_{p\nmid N} G(\mathbb{Q}_p)$ such that for each finite prime p not dividing N,

(1) $\kappa_p \in C_c^\infty(G(\mathbb{Q}_p), \omega_{\pi_p}^{-1})$,

(2) κ_p is bi-$G(\mathbb{Z}_p)$-invariant.

It is well-known that $\mathcal{H}_{\mathrm{ur}}$ is a commutative algebra and is generated by the various functions κ_ℓ (as ℓ varies over integers coprime to N) where $\kappa_\ell = \prod_{p\nmid N} \kappa_{\ell,p}$ and the function $\kappa_{\ell,p}$ in $C_c^\infty(G(\mathbb{Q}_p), \omega_{\pi_p}^{-1})$ is defined as follows:

(1) $\kappa_{\ell,p}(zka(\ell)k) = |\ell|^{-1/2}\omega_{\pi_p}^{-1}(z)$ for all $z \in Z(\mathbb{Q}_p)$ and $k \in G(\mathbb{Z}_p)$.

(2) $\kappa_{\ell,p}(g) = 0$ if $g \notin Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p)$.

Then, it follows that for each $\kappa_{\mathrm{ur}} \in \mathcal{H}_{\mathrm{ur}}$,

$$R(\kappa_{\mathrm{ur}})\phi' := \int_{\prod_{p\nmid N}(Z\backslash G)(\mathbb{Q}_p)} \kappa_{\mathrm{ur}}(g)(\pi(g)\phi')\,dg = \delta_{\mathrm{ur}}\phi',$$

where δ_{ur} is a complex number (depending linearly on κ_{ur}). Furthermore,

$$R(\kappa_\ell)\phi' = \lambda_\pi(\ell)\phi',$$

where the Hecke eigenvalues $\lambda_\pi(\ell)$ were defined earlier in Lemma 3.10. Moreover, we note that as κ_{ur} varies over $\mathcal{H}_{\mathrm{ur}}$, the corresponding operators $R(\kappa_{\mathrm{ur}})$ form a commuting system of normal operators. Indeed, if we define $\kappa_\ell^* = \bigl(\prod_{p\mid\ell} \omega_{\pi_p}^{-1}(\ell)\bigr)\kappa_\ell$, and extend this via multiplicativity and antilinearity to all of $\mathcal{H}_{\mathrm{ur}}$, then we have an involution $\kappa \mapsto \kappa^*$ on all of $\mathcal{H}_{\mathrm{ur}}$. It is well-known that $\kappa^*(g) = \overline{\kappa(g^{-1})}$ and hence $R(\kappa^*)$ is precisely the adjoint of $R(\kappa)$.

Finally, we consider the infinite place. For $g \in G(\mathbb{R})^+$, let $u(g)$ denote the hyperbolic distance from $g(i)$ to i; precisely $u(g) = 3D|g(i) - i|^2/(4\,\mathrm{Im}(g(i)))$. Each bi-$Z(\mathbb{R})K_\infty$-invariant function κ_∞ in $C_c^\infty(Z(\mathbb{R})\backslash G(\mathbb{R})^+)$, can be viewed as a function on $\mathbb{R}^+$ via $\kappa_\infty(g) = \kappa_\infty(u(g))$. For each irreducible spherical unitary principal series representation σ of $G(\mathbb{R})$, we define the Harish-Chandra–Selberg transform $\hat{\kappa}_\infty(\sigma)$ via

$$\hat{\kappa}_\infty(\sigma) = \int_{Z(\mathbb{R})\backslash G(\mathbb{R})^+} \kappa_\infty(g)\frac{\langle \sigma(g)v_\sigma, v_\sigma\rangle}{\langle v_\sigma, v_\sigma\rangle}\,dg,$$

where v_σ is the unique (up to multiples) spherical vector in the representation σ. For all such σ it is known that $R(\kappa_\infty)v_\sigma = \hat{\kappa}_\infty(\sigma)v_\sigma$; in particular $R(\kappa_\infty)\phi' = \hat{\kappa}_\infty(\pi_\infty)\phi'$.

By [Templier 2015, Lemma 2.1] there exists such a function κ_∞ on $G(\mathbb{R})$ with the following properties:

(1) $\kappa_\infty(g) = 0$ unless $g \in G(\mathbb{R})^+$ and $u(g) \le 1$.

(2) $\hat{\kappa}_\infty(\sigma) \ge 0$ for all irreducible spherical unitary principal series representations σ of $G(\mathbb{R})$.

(3) $\hat{\kappa}_\infty(\pi_\infty) \gg 1$.

(4) For all $g \in G(\mathbb{R})^+$, $|\kappa_\infty(g)| \le T$. Moreover, if $u(g) \ge T^{-2}$, then $|\kappa_\infty(g)| \le T^{1/2}/u(g)^{1/4}$.

Henceforth, we fix a function κ_∞ as above.

3F. ***The amplified pretrace formula.*** In this subsection, we will use $L^2(X)$ as a shorthand for $L^2(G(\mathbb{Q})\backslash G(\mathbb{A})/K_1'(N)K_\infty, \omega_\pi)$.

Let the functions Φ_N', κ_∞ be as defined in the previous subsection. Consider the space of functions κ on $G(\mathbb{A})$ such that $\kappa = \Phi_N'\kappa_{\mathrm{ur}}\kappa_\infty$ with κ_{ur} in $\mathcal{H}_{\mathrm{ur}}$. We fix an orthonormal basis $\mathcal{B} = \{\psi\}$ of the space $L^2(X)$ such that $\phi' \in \mathcal{B}$ and consisting of eigenfunctions for all the operators $R(\kappa)$ with κ as above; i.e., for all $\psi \in \mathcal{B}$, there exists a complex number λ_ψ satisfying

$$R(\Phi_N')R(\kappa_{\mathrm{ur}})R(\kappa_\infty)\psi = R(\kappa)\psi := \int_{Z(\mathbb{A})\backslash G(\mathbb{A})} \kappa(g)(\pi(g)\psi)\,dg = \lambda_\psi\psi.$$

Such a basis exists because the set of all $R(\kappa)$ as above form a commuting system of normal operators. The basis $\mathcal{B}$ naturally splits into a discrete and continuous part, with the continuous part consisting of Eisenstein series and the discrete part consisting of cusp forms and residual functions.

Given a $\kappa = \Phi_N'\kappa_{\mathrm{ur}}\kappa_\infty$ as above, we define the automorphic kernel $K_\kappa(g_1, g_2)$ for $g_1, g_2 \in G(\mathbb{A})$ via

$$K_\kappa(g_1, g_2) = \sum_{\gamma \in Z(\mathbb{Q})\backslash G(\mathbb{Q})} \kappa(g_1^{-1}\gamma g_2).$$

A standard calculation tells us that if $\psi = \otimes_v \psi_v$ is an element of $L^2(X)$ such that ψ_v is an eigenfunction for $R(\kappa_v)$ with eigenvalue λ_v, for each place v, then

$$\int_{Z(\mathbb{A})G(\mathbb{Q})\backslash G(\mathbb{A})} K_\kappa(g_1, g_2)\psi(g_2)\,dg_2 = \left(\prod_v \lambda_v\right)\psi(g_1). \tag{28}$$

Lemma 3.17. *Suppose that $\kappa_{\mathrm{ur}} = \kappa_{\mathrm{ur}}' * (\kappa_{\mathrm{ur}}')^*$ for some $\kappa_{\mathrm{ur}}' \in \mathcal{H}_{\mathrm{ur}}$. Put $\kappa = \Phi_N'\kappa_{\mathrm{ur}}\kappa_\infty$. If $\psi \in \mathcal{B}$ then*

$$\int_{Z(\mathbb{A})G(\mathbb{Q})\backslash G(\mathbb{A})} K_\kappa(g_1, g_2)\psi(g_2)\,dg_2 = \lambda_\psi\psi(g_1),$$

for some $\lambda_\psi \geq 0$. Moreover $\lambda_{\phi'} \geq M_1^{-1}N_1^{-1}|\lambda'_{\mathrm{ur}}|^2\hat{\kappa}_\infty(\pi_\infty)$ where the quantity λ'_{ur} is defined by $R(\kappa'_{\mathrm{ur}})\phi' = \lambda'_{\mathrm{ur}}\phi'$.

Proof. By our assumption that $\psi \in \mathcal{B}$, a complex number λ_ψ as above exists. We can write $\lambda_\psi = \lambda_{\psi,N}\lambda_{\psi,\mathrm{ur}}\lambda_{\psi,\infty}$ using the decomposition $R(\kappa) = R(\Phi'_N)R(\kappa_{\mathrm{ur}})R(\kappa_\infty)$. We have $\lambda_{\psi,\infty} \geq 0$ by our assumption $\hat{\kappa}_\infty(\sigma) \geq 0$ for all irreducible spherical unitary principal series representations σ of $G(\mathbb{R})$. We have $\lambda_{\psi,N} \geq 0$ by Corollary 2.16. Finally if $R(\kappa'_{\mathrm{ur}})\psi = \lambda'_\psi\psi$ then $\lambda_{\psi,\mathrm{ur}} = |\lambda'_\psi|^2 \geq 0$. Hence $\lambda_\psi \geq 0$. The last assertion is immediate from the results of the previous subsection. □

Henceforth we assume that $\kappa_{\mathrm{ur}} = \kappa'_{\mathrm{ur}} * (\kappa'_{\mathrm{ur}})^*$ for some $\kappa'_{\mathrm{ur}} \in \mathcal{H}_{\mathrm{ur}}$ and we put $\kappa = \Phi'_N\kappa_{\mathrm{ur}}\kappa_\infty$. Then spectrally expanding $K_\kappa(g,g)$ along $\mathcal{B}$ and using Lemma 3.17 we get, for all $g \in G(\mathbb{A})$,

$$M_1^{-1}N_1^{-1}\hat{\kappa}_\infty(\pi_\infty)|\lambda'_{\mathrm{ur}}\phi'(g)|^2 \leq K_\kappa(g,g).$$

Note that $\hat{\kappa}_\infty(\pi_\infty) \geq 1$. Next we look at the quantity $K_\kappa(g,g)$. Assume that $g = k_N n(x)a(y)$ with $k_N = \prod_{p\,|\,N} k_p \in K_N$ and $k_N h_N \in \mathcal{J}_N$. The second condition means that $k_p \in wK_p^0(p)$ for all $p \mid N_2$. We have

$$K_\kappa(g,g) = \sum_{\gamma \in Z(\mathbb{Q})\backslash G(\mathbb{Q})} \Phi'_N(k_N^{-1}\gamma k_N)\kappa_{\mathrm{ur}}(\gamma)\kappa_\infty((n(x)a(y))^{-1}\gamma n(x)a(y)).$$

Above we have $\Phi'_N(k_N^{-1}\gamma k_N) \leq 1$, moreover if $\Phi'_N(k_N^{-1}\gamma k_N) \neq 0$ then we must have

(a) $k_p^{-1}\gamma k_p \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)$ for all primes p dividing N, and

(b) $k_p^{-1}\gamma k_p \in Z(\mathbb{Q}_p)K_p^0(p)$ for all primes p dividing N_2.

Condition (a) implies that $\gamma \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)$ for all primes p dividing N. Condition (b), together with the fact that $k_p \in wK_p^0(p)$ for all $p \mid N_2$, implies that $\gamma \in Z(\mathbb{Q}_p)K_{0,p}(p)$ for all primes p dividing N_2.

Finally we have $\kappa_\infty(g) = 0$ if $\det(g) < 0$, and if $\det(g) > 0$ we can write $\kappa_\infty(g) = \kappa_\infty(u(g))$ as explained earlier, whence

$$\kappa_\infty((n(x)a(y))^{-1}\gamma n(x)a(y)) = \kappa_\infty(u(z,\gamma z)), \quad z = x+iy,$$

where, for any two points z_1 and z_2 on the upper-half plane, $u(z_1,z_2)$ denotes the hyperbolic distance between them, i.e., $u(z_1,z_2) = |z_1-z_2|^2/(4\,\mathrm{Im}(z_1)\,\mathrm{Im}(z_2))$. Putting everything together, we get the following Proposition.

Proposition 3.18. *Let $\kappa'_{\mathrm{ur}} \in \mathcal{H}_{\mathrm{ur}}$ and suppose that $R(\kappa'_{\mathrm{ur}})\phi' = \lambda'_{\mathrm{ur}}\phi'$. Let $\kappa_{\mathrm{ur}} = \kappa'_{\mathrm{ur}} * (\kappa'_{\mathrm{ur}})^*$ and $\kappa = \Phi'_N\kappa_{\mathrm{ur}}\kappa_\infty$. Then for all $z = x+iy$ and all $k \in K_N$ such that $kh_N \in \mathcal{J}_N$, we have*

$$|\phi'(kn(x)a(y))|^2 \leq \frac{M_1N_1}{|\lambda'_{\mathrm{ur}}|^2} \sum_{\substack{\gamma \in Z(\mathbb{Q})\backslash G(\mathbb{Q})^+, \\ \gamma \in Z(\mathbb{Q}_p)K_{0,p}(p)\forall p\,|\,N_2 \\ \gamma \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)\forall p\,|\,N}} |\kappa_{\mathrm{ur}}(\gamma)\kappa_\infty(u(z,\gamma z))|.$$

3G. ***Conclusion.*** We now make a specific choice for κ_{ur}. Let $\Lambda \geq 1$ be a real number. We let

$$S = \{\ell : \ell \text{ prime}, (\ell, N) = 1, \Lambda \leq \ell \leq 2\Lambda\}.$$

Define for each integer r,

$$c_r = \begin{cases} |\lambda_\pi(r)|/\lambda_\pi(r) & \text{if } r = \ell \text{ or } r = \ell^2, \ell \in S, \\ 0 & \text{otherwise.} \end{cases}$$

We set $\kappa'_{\mathrm{ur}} = \sum_r c_r \kappa_r$, and $\kappa_{\mathrm{ur}} = \kappa'_{\mathrm{ur}} * (\kappa'_{\mathrm{ur}})^*$. Given this, let us estimate the quantities appearing in Proposition 3.18.

First of all, we have $\lambda'_{\mathrm{ur}} = \sum_{\ell \in S}(|\lambda_\pi(\ell)| + |\lambda_\pi(\ell^2)|)$. By the well-known relation $\lambda_\pi(\ell)^2 - \lambda_\pi(\ell^2) = \omega_{\pi_\ell}(\ell)$, it follows that $|\lambda_\pi(\ell)| + |\lambda_\pi(\ell^2)| \geq 1$. Hence $\lambda'_{\mathrm{ur}} \gg_\varepsilon \Lambda^{1-\varepsilon}$.

Next, using the well-known relation

$$\kappa_m * \kappa_n^* = \sum_{t \,|\, \gcd(m,n)} \biggl(\prod_{p \,|\, t} \omega_{\pi_p}(t)\biggr)\biggl(\prod_{p \,|\, n} \omega_{\pi_p}^{-1}(n)\biggr)\kappa_{mn/t^2},$$

we see that

$$\kappa_{\mathrm{ur}} = \sum_{1 \leq l \leq 16\Lambda^4} y_l \kappa_l$$

where the complex numbers y_l satisfy:

$$|y_l| \ll \begin{cases} \Lambda, & l = 1, \\ 1, & l = \ell_1 \text{ or } l = \ell_1\ell_2 \text{ or } l = \ell_1\ell_2^2 \text{ or } l = \ell_1^2\ell_2^2 \text{ with } \ell_1, \ell_2 \in S, \\ 0, & \text{otherwise.} \end{cases}$$

We have $|\kappa_l(\gamma)| \leq l^{-1/2}$. Moreover $\kappa_l(\gamma) = 0$ unless $\gamma \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p)$ for all $p \nmid N$. We deduce the following bound:

$$|\phi'(kn(x)a(y))|^2 \ll_\varepsilon \Lambda^{-2+\varepsilon} M_1 N_1 \sum_{1 \leq l \leq 16\Lambda^4} \frac{y_l}{\sqrt{l}} \sum_{\substack{\gamma \in Z(\mathbb{Q})\backslash G(\mathbb{Q})^+, \\ \gamma \in Z(\mathbb{Q}_p)K_{0,p}(p) \forall p \,|\, N_2 \\ \gamma \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p) \forall p \nmid N_2}} |\kappa_\infty(u(z, \gamma z))|. \tag{29}$$

Define

$$M(\ell, N_2) = \bigl\{\bigl[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\bigr], a, b, c, d \in \mathbb{Z}, a > 0, N_2 \,|\, c, ad - bc = \ell\bigr\}.$$

The following lemma follows immediately from strong approximation.

Lemma 3.19. *Let $\gamma \in G(\mathbb{Q})^+$ and ℓ be a positive integer coprime to N_2. Suppose for each prime p that $\gamma \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p)$. Suppose also for each prime $p \mid N_2$ that $\gamma \in Z(\mathbb{Q}_p)K_{0,p}(p)$. Then there exists $z \in Z(\mathbb{Q})$ such $z\gamma \in M(\ell, N_2)$.*

Proof. Omitted. □

Let us take another look at (29) in view of Lemma 3.19. The sum in (29) is over all matrices γ in $Z(\mathbb{Q})\backslash G(\mathbb{Q})^+$ such that $\gamma \in Z(\mathbb{Q}_p)K_{0,p}(p)$ for $p \mid N_2$ and $\gamma \in Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p)$ for $p \nmid N_2$. The latter condition can equally well be taken over all p as

$$Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p) = Z(\mathbb{Q}_p)G(\mathbb{Z}_p)a(\ell)G(\mathbb{Z}_p) = Z(\mathbb{Q}_p)G(\mathbb{Z}_p)$$

if ℓ and p are coprime, which is the case when $p \mid N$. Therefore Lemma 3.19, together with the fact that the natural map from $M(\ell, N_2)$ to $Z(\mathbb{Q})\backslash G(\mathbb{Q})^+$ is an injection, implies that the sum in (29) can replaced by a sum over the set $M(\ell, N_2)$. Hence, writing $g = kh_N \in \mathcal{J}_N$ as before, we get

$$\begin{aligned} |\phi(gn(x)a(y))|^2 &= |\phi'(kn(x)a(y))|^2 \\ &\ll_\varepsilon \Lambda^{-2+\varepsilon} M_1 N_1 \sum_{1 \le l \le 16\Lambda^4} \tfrac{y_l}{\sqrt{l}} \sum_{\gamma \in M(\ell, N_2)} |\kappa_\infty(u(z, \gamma z))|. \end{aligned} \tag{30}$$

For any $\delta > 0$, we define

$$N(z, \ell, \delta, N_2) = |\{\gamma \in M(\ell, N_2) : u(z, \gamma z) \le \delta\}|.$$

We have the following counting result due to Templier [2015, Proposition 6.1].

Proposition 3.20. *Let $z = x + iy \in \mathcal{F}_{N_2}$. For any $0 < \delta < 1$ and positive integer ℓ coprime to N_2, let the number $N(z, \ell, \delta, N_2)$ be defined as above.*

For $\Lambda \ge 1$, define

$$A(z, \Lambda, \delta, N_2) = \sum_{1 \le l \le 16\Lambda^4} \tfrac{y_l}{\sqrt{l}} N(z, \ell, \delta, N_2).$$

Then

$$A(z,\Lambda,\delta,N_2) \ll_\varepsilon \Lambda^\varepsilon N_2^\varepsilon \big[\Lambda + \Lambda N_2^{1/2}\delta^{1/2} y + \Lambda^{5/2}\delta^{1/2} N_2^{-1/2} + \Lambda^{5/2}\delta^{1/2} y + \Lambda^4 \delta N_2^{-1}\big].$$

Proof. This is Proposition 6.1 of [Templier 2015]. □

Now (30) gives us

$$|\phi(gn(x)a(y))|^2 \ll_\varepsilon \Lambda^{-2+\varepsilon} M_1 N_1 \int_0^1 |\kappa_\infty(\delta)| \, dA(z, \Lambda, \delta, N_2). \tag{31}$$

Using Proposition 3.20 and the property $|\kappa_\infty(\delta)| \le \min(T, T^{1/2}/\delta^{1/4})$, we immediately deduce Proposition 3.16 after a simple integration, as in [Templier 2015, Section 6.2].

Acknowledgements

I would like to thank Edgar Assing, Simon Marshall, Paul Nelson, and Nicolas Templier for useful discussions and feedback.

Note added in proof. Recent work of the author with Yueke Hu suggests that (1) may not hold in general in the case of powerful levels. This is due to the failure of Conjecture 1 in certain cases, which we have recently discovered.

References

[Blomer and Holowinsky 2010] V. Blomer and R. Holowinsky, "Bounding sup-norms of cusp forms of large level", *Invent. Math.* **179**:3 (2010), 645–681. MR Zbl

[Harcos and Michel 2006] G. Harcos and P. Michel, "The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points II", *Invent. Math.* **163**:3 (2006), 581–655. MR Zbl

[Harcos and Templier 2012] G. Harcos and N. Templier, "On the sup-norm of Maass cusp forms of large level: II", *Int. Math. Res. Not.* **2012**:20 (2012), 4764–4774. MR Zbl

[Harcos and Templier 2013] G. Harcos and N. Templier, "On the sup-norm of Maass cusp forms of large level III", *Math. Ann.* **356**:1 (2013), 209–216. MR Zbl

[Hoffstein and Lockhart 1994] J. Hoffstein and P. Lockhart, "Coefficients of Maass forms and the Siegel zero", *Ann. of Math.* (2) **140**:1 (1994), 161–181. MR Zbl

[Hu 2017] Y. Hu, "Triple product formula and mass equidistribution on modular curves of level N", *Int. Math. Res. Not.* (2017), art. id. rnw322.

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

[Marshall 2016] S. Marshall, "Local bounds for L^p norms of Maass forms in the level aspect", *Algebra Number Theory* **10**:4 (2016), 803–812. MR Zbl

[Milićević 2016] D. Milićević, "Sub-Weyl subconvexity for Dirichlet L-functions to prime power moduli", *Compos. Math.* **152**:4 (2016), 825–875. MR

[Nelson 2011] P. D. Nelson, "Equidistribution of cusp forms in the level aspect", *Duke Math. J.* **160**:3 (2011), 467–501. MR Zbl

[Nelson et al. 2014] P. D. Nelson, A. Pitale, and A. Saha, "Bounds for Rankin–Selberg integrals and quantum unique ergodicity for powerful levels", *J. Amer. Math. Soc.* **27**:1 (2014), 147–191. MR Zbl

[Saha 2014] A. Saha, "On sup-norms of cusp forms of powerful level", Preprint, 2014. arXiv

[Saha 2016] A. Saha, "Large values of newforms on GL(2) with highly ramified central character", *Int. Math. Res. Not.* **2016**:13 (2016), 4103–4131. MR

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

[Templier 2014] N. Templier, "Large values of modular forms", *Camb. J. Math.* **2**:1 (2014), 91–116. MR Zbl

[Templier 2015] N. Templier, "Hybrid sup-norm bounds for Hecke–Maass cusp forms", *J. Eur. Math. Soc.* (*JEMS*) **17**:8 (2015), 2069–2082. MR Zbl

[Tunnell 1978] J. B. Tunnell, "On the local Langlands conjecture for $GL(2)$", *Invent. Math.* **46**:2 (1978), 179–200. MR Zbl

Communicated by Peter Sarnak
Received 2015-10-13 Revised 2016-10-25 Accepted 2016-12-16

abhishek.saha@bristol.ac.uk *Department of Mathematics, University of Bristol, Bristol, BS81SN, United Kingdom*

Algebra & Number Theory

msp.org/ant

See inside back cover or msp.org/ant for submission instructions.

The subscription price for 2017 is US $325/year for the electronic version, and $520/year (+$55, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebra & Number Theory

Volume 11 No. 5 2017

Hybrid sup-norm bounds for Maass newforms of powerful level 1009
ABHISHEK SAHA

Collinear CM-points 1047
YURI BILU, FLORIAN LUCA and DAVID MASSER

A uniform classification of discrete series representations of affine Hecke algebras 1089
DAN CIUBOTARU and ERIC OPDAM

An explicit bound for the least prime ideal in the Chebotarev density theorem 1135
JESSE THORNER and ASIF ZAMAN

Modular curves of prime-power level with infinitely many rational points 1199
ANDREW V. SUTHERLAND and DAVID ZYWINA

Some sums over irreducible polynomials 1231
DAVID E. SPEYER

1937-0652(2017)11:5;1-G