THE SUBCONVEXITY PROBLEM FOR RANKIN–SELBERG L-FUNCTIONS AND EQUIDISTRIBUTION OF HEEGNER POINTS. II

G. HARCOS AND P. MICHEL

Abstract. We prove a general subconvex bound in the level aspect for Rankin–Selberg L-functions associated with two primitive holomorphic or Maass cusp forms over $\mathbf{Q}$. We use this bound to establish the equidistribution of incomplete Galois orbits of Heegner points on Shimura curves associated with indefinite quaternion algebras over $\mathbf{Q}$.

1. Introduction

In this paper, we pursue the program—initiated in [KMV02] and continued in [M04a]—of solving the subconvexity problem for Rankin–Selberg L-functions in the level aspect. The subconvexity problem is the following: given two primitive cusp forms f and g, we denote by q and D, χ_f and χ_g, $\pi_f = \bigotimes' \pi_{f,p}$ and $\pi_g = \bigotimes' \pi_{g,p}$, respectively, the level, nebentypus, and $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$-automorphic representation attached to f and g; finally, we denote by

$$L(f \otimes g, s) = L(\pi_f \otimes \pi_g, s) = \prod_{p<\infty} L(\pi_{f,p} \otimes \pi_{g,p}, s)$$

the associated (finite) Rankin–Selberg L-function as it is defined in Jacquet's monograph [J72] (see also [C04]).

As in [KMV02, M04a], we are interested in providing non-trivial upper bounds for $L(f \otimes g, s)$ when g is (essentially) fixed, $\Re s = \frac{1}{2}$, and $q \to +\infty$. Using the local Langlands correspondence, one can verify that the conductor $Q(f \otimes g)$ of $L(\pi_f \otimes \pi_g, s)$ satisfies the bound

$$(qD)^2/(q,D)^4 \leqslant Q(f \otimes g) \leqslant (qD)^2/(q,D),$$

hence the subconvexity bound we are seeking is a bound of the form

$$L(f \otimes g, s) \ll q^{\frac{1}{2}-\delta},$$

for $\Re s = \frac{1}{2}$, with some absolute $\delta > 0$, the implied constant depending on $|s|$, g and the parameter at infinity of f (i.e., the weight or the Laplacian eigenvalue).

In [KMV02], the problem was solved under the following assumptions:

- f is a holomorphic cusp form;
- the conductor q^* of $\chi_f\chi_g$ is at most $q^{\frac{1}{2}-\eta}$ for some $\eta > 0$, the corresponding exponent δ then depending on η.

Of the two conditions above, the second is the more serious. It was essentially removed in [M04a] under the assumptions:

- g is a holomorphic cusp form;
- $\chi_f\chi_g$ is non-trivial.

The main objective of the present paper is to remove the assumption of g holomorphic; we prove

2000 *Mathematics Subject Classification.* Primary: 11F66; Secondary: 11F67, 11M41.

Key words and phrases. Automorphic L-functions, Central Values, Subconvexity, Equidistribution.

The second author was supported by the Institut Universitaire de France and by the ACI "Arithmétique des fonctions L".

Theorem 1. *Let f and g be two primitive (either Maass or holomorphic) cusp forms of level q, D and nebentypus χ_f, χ_g, respectively. Suppose that $\chi_f\chi_g$ is not trivial; then for $\Re s = \frac{1}{2}$ one has*

$$L(f \otimes g, s) \ll q^{\frac{1}{2}-\delta}, \tag{1.1}$$

with $\delta = \frac{1}{2648} = 0.000377\ldots$, and with the implied constant depending polynomially on $|s|$, D and on the parameters at infinity of f and g (i.e., the weight or the Laplacian eigenvalue).

Remark 1.1. More precisely, the subconvex exponent $\frac{1}{2} - \delta$ can be replaced by

$$\frac{1}{2} - \frac{(1-2\theta)^2}{1616} + \varepsilon \tag{1.2}$$

for any $\varepsilon > 0$. Here θ stands for an approximation towards the Ramanujan–Petersson conjecture for weight zero Maass forms (see Hypothesis H_θ below). While the current best approximation towards the Ramanujan–Petersson conjecture—due to Kim–Shahidi, Kim and Kim–Sarnak [KS02, K03, KS03]—is rather strong ($\theta = \frac{7}{64}$ is admissible), we see from (1.2) that in order to solve the present subconvexity problem any $\theta < \frac{1}{2}$ would have been sufficient. Note also that for simplicity we decided to exhibit a uniform subconvex exponent: when the conductor of χ_f is small (i.e., smaller than q^η for some fixed $\eta < 1$), better subconvex bounds are admissible (we leave it to the reader to determine the dependency of such bounds in terms of the parameter η: cf. the exponents given in Theorem 4).

Remark 1.2. The method of proof of Theorem 1 can also be applied (with some non-trivial modifications) when g is an Eisenstein series; the most interesting one being the non-holomorphic Eisenstein series of full level:

$$g(z) := \frac{\partial}{\partial s} E(z,s)_{|s=\frac{1}{2}} = 2\sqrt{y}\log(e^\gamma y/4\pi) + 4\sqrt{y}\sum_{n\geqslant 1}\tau(n)\cos(2\pi nx)K_0(2\pi ny);$$

in this case, $L(f \otimes g, s) = L(f,s)^2$ is the square of the standard Hecke L-function of f. When χ_f is primitive (which, in some sense, is the most difficult case), the subconvexity problem for $L(f,s)$ was solved by Duke–Friedlander–Iwaniec, in a series of difficult papers [DFI97a, DFI97b, DFI01, DFI02]: one has

$$L(f,s) \ll q^{\frac{1}{4}-\frac{1}{23041}}, \tag{1.3}$$

for $\Re s = \frac{1}{2}$, where the implied constant depends polynomially on $|s|$ and the parameters at infinity of f. In a future work [BHM05b], we shall use a modification of the methods of the present paper (using the original δ-symbol method of [DFI94a]) to give a fairly different proof of (1.3), valid when χ_f is non-trivial and with a significantly improved subconvexity exponent.

Note that besides achieving a subconvex exponent in the level aspect, a not so small portion of our effort is directed towards checking the polynomial dependency of the bounds of Theorem 1 with respect to the remaining parameters of f and g[1]. This (seemingly minor) aspect turns out to be important in several situations: namely for the problems in which subconvexity is applied not only to an individual L-function but to a whole family so that polynomial control in the remaining parameters indexing the family is crucial. A typical example is the present subconvexity bound (1.1), which relies ultimately on a family of subconvex estimates for twisted Hecke L-functions over a whole Hecke eigenbasis of automorphic forms of small levels (see Sections 5.4 and 5.6). Another example is given in Remark 1.4.

[1] In particular, when g is holomorphic of weight k_g, (1.1) depends polynomially on k_g which is not the case for the bound proven in [M04a].

1.1. **Equidistribution of Heegner points.** A remarkable feature of the subconvexity problem is that in several occasions a subconvex bound is just sufficient to bring a full solution to some natural, apparently unrelated, questions; it is in particular the case with several equidistribution problems. This was first pointed out by Duke [D88] in the context of Linnik's problems on the distribution of integral points on the sphere and on the distribution of Heegner points on the full modular curve (see also [IS00, S94, S01, M04b] for further effective or potential applications of subconvexity to other equidistribution problems).

In fact, one of the main applications of the subconvex bound of Theorem 1 is to provide some further refinements concerning the equidistribution of Heegner points on modular curves. For this introduction, we describe our results in the case of the full modular curve $X_0(1)$, which is the original case treated in [D88]. Recall that as a Riemann surface, $X_0(1) \simeq \mathbb{P}^1$ is defined as the quotient $\mathrm{SL}_2(\mathbf{Z})\backslash\big(\mathbf{H}\cup\mathbf{Q}\cup\{\infty\}\big)$, where $\mathbf{H}$ is the upper half-plane and $\mathrm{SL}_2(\mathbf{Z})$ acts by linear fractional transformations. We denote by $d\mu(z)$ the probability measure on $X_0(1)(\mathbf{C})$ induced by the normalized hyperbolic measure $\frac{3}{\pi}\frac{dxdy}{y^2}$ on $\mathbf{H}$. The map $z \mapsto \mathbf{C}/(\mathbf{Z}+\mathbf{Z}z)$, where $z \in \mathbf{H}$, induces a bijection from $\mathrm{SL}_2(\mathbf{Z})\backslash\mathbf{H}$—the points of $X_0(1)(\mathbf{C})$ excluding the cusp at infinity—to the set of isomorphism classes of elliptic curves over $\mathbf{C}$. We write φ for the inverse of this bijection. For an imaginary quadratic extension $K/\mathbf{Q}$ with maximal order $\mathcal{O}_K$ of discriminant $-D$, we denote by $\mathrm{Ell}(\mathcal{O}_K)$ the set of ($\mathbf{C}$-isomorphism classes of) elliptic curves with complex multiplication by $\mathcal{O}_K$. By the theory of complex multiplication, these curves are in fact defined over H_K, the Hilbert class field of K, and the Hilbert class group $G_K := \mathrm{Gal}(H_K/K) \simeq \mathrm{Pic}(\mathcal{O}_K)$ acts simply transitively on $\mathrm{Ell}(\mathcal{O}_K)$; hence $\mathrm{Ell}(\mathcal{O}_K) = \{E^\sigma,\ \sigma \in G_K\}$ for any $E \in \mathrm{Ell}(\mathcal{O}_K)$. The set of *Heegner points* (of conductor 1) with CM by K is the image of the embedding $\mathrm{Ell}(\mathcal{O}_K) \overset{\varphi}{\hookrightarrow} X_0(1)(\mathbf{C})$. Since by Siegel's theorem,

$$\big|\mathrm{Ell}(\mathcal{O}_K)\big| = \big|\mathrm{Pic}(\mathcal{O}_K)\big| \gg_\varepsilon D^{\frac{1}{2}-\varepsilon} \to +\infty \quad \text{as} \quad D \to +\infty,$$

one may wonder how the Heegner points are distributed on $X_0(1)(\mathbf{C})$ as D grows. This question was investigated by Linnik [L68]: by means of his pioneering ergodic method, Linnik proved, under an additional congruence condition on $-D$ modulo some fixed prime, that $\varphi\big(\mathrm{Ell}(\mathcal{O}_K)\big)$ becomes equidistributed relatively to the measure $d\mu(z)$ as $D \to +\infty$. This restriction was removed by Duke [D88] by using quite different techniques; namely by exploiting a correspondence of Maass to relate the Weyl sums associated to this equidistribution problem to Fourier coefficients of half-integral weight Maass forms and by proving directly non-trivial bounds for them (using a technique introduced by Iwaniec [I87]). The connection of this problem with subconvexity comes from the work of Waldspurger [W81] on the Shimura correspondence, which shows that non-trivial bounds for these Fourier coefficients are in fact equivalent to subconvexity bounds for the central values of twisted L-functions

$$L\left(g \otimes \chi_K, \tfrac{1}{2}\right) \ll_g D^{\frac{1}{2}-\delta},$$

where g ranges over the (weight 0) Maass eigenforms on $X_0(1)$ (possibly Eisenstein series) and χ_K denotes the quadratic character (of conductor D) corresponding to K; the latter bound was proved in [DFI93] for g holomorphic and extended to Maass forms in [H03a, M04a], thus providing another approach to Duke's theorems. Our main application is to strengthen these results by proving the equidistribution of smaller Galois orbits of Heegner points:

Theorem 2. *For any continuous function $g : X_0(1)(\mathbf{C}) \to \mathbf{C}$, there exists a bounded function $\varepsilon_g : \mathbf{R}^+ \to \mathbf{R}^+$ which satisfies*

$$\lim_{x\to 0} \varepsilon_g(x) = 0$$

such that: for any imaginary quadratic field K with discriminant $-D$, any subgroup $G \subset G_K$, and any $E \in \mathrm{Ell}(\mathcal{O}_K)$, one has

$$\left|\frac{1}{|G|}\sum_{\sigma\in G} g\big(\varphi(E^\sigma)\big) - \int_{X_0(1)(\mathbf{C})} g(z)\, d\mu(z)\right| \leqslant \varepsilon_g\big([G_K : G]D^{-\frac{1}{23042}}\big). \tag{1.4}$$

This result states the equidistribution of orbits of Heegner points by subgroups G of G_K of index satisfying $[G_K : G] = o(D^{\frac{1}{23042}})$; this is a special instance of equidistribution for short orbits of Heegner points on Shimura curves associated to indefinite quaternion algebras[2] over $\mathbf{Q}$ and is meaningful in the context of the André–Oort conjectures. At this point, we refer to Section 6 for definitions and proofs of more general cases, and simply say that the equidistribution property is a consequence of deep Gross–Zagier type formulae established by Zhang [Z01a, Z01b, Z04] which connect the Weyl sums associated to this equidistribution problem to central values of Rankin–Selberg L-functions. The proof that these Weyl sums are small follows from the subconvexity bound (1.3) and from the bound (1.1) of Theorem 1.

Remark 1.3. The exponent $\frac{1}{23042}$ in (1.4) comes from the subconvex estimate (1.3) of [DFI02, Theorem 2.4], which is used to bound the Weyl sums corresponding to the Eisenstein spectrum. In particular, for functions g contained in the span of the Maass cusp forms, only Theorem 1 is used and (1.4) holds with the stronger exponent $\frac{1}{5297}$. In Section 6 similar equidistribution problems for compact Shimura curves are considered. As such compact Riemann surfaces have no Eisenstein spectrum, the analogue of (1.4) holds with this stronger exponent as well.

Remark 1.4. The polynomial control can be used to give more quantitative information on the equidistribution of Heegner points, that is, to bound their discrepancy. For instance, it is natural to consider, as in [LS95], the spherical cap discrepancy

$$D_E(G) := \sup_{B \subset X_0(1)} \left| \frac{1}{|G|} \sum_{\substack{\sigma \in G \\ \varphi(E^\sigma) \in B}} 1 - \frac{3}{\pi} \operatorname{Vol}(B) \right|,$$

where the supremum is over all the geodesic balls $B \subset X_0(1)$ and $\operatorname{Vol}(B) = \int_B \frac{dxdy}{y^2}$. Using the bounds of Theorem 1, [DFI02] and the approximation arguments of [LS95, Section 5], one can show that there exists an explicitly computable $\eta > 0$ such that, under the notations and assumptions of Theorem 2, one has

$$D_E(G) \ll [G_K : G] D^{-\frac{1}{23042} - \eta},$$

where the implied constant is absolute (but not effective).

These equidistribution results admit analogues in the case of real quadratic fields. In his recent Ph.D. thesis, Popa [P03] established formulae analogous to those of Gross–Zagier and Zhang: for example, his formulae relate the central values of Rankin–Selberg L-functions (of automorphic forms against theta functions attached to narrow class group characters of a real quadratic field) to some (twisted) integrals of modular forms along geodesic one-cycles on a Shimura curve. The subconvexity bounds of [M04a] and of the present paper then can be used in conjunction with these formulae to study the distribution properties of (orbits of) these cycles either on the curve or in the homology of the curve, generalizing former results of Duke [D88].

1.2. **Outline of the proof of Theorems 1.** The beginning of the proof closely follows [DFI02, M04a]. By a standard approximate functional equation, one is reduced to estimate non-trivially partial sums of the form

$$\sum_{n \geqslant 1} \frac{\lambda_f(n)\lambda_g(n)}{\sqrt{n}} W\left(\frac{n}{N}\right),$$

where $\lambda_f(n)$ and $\lambda_g(n)$ denote the n-th Hecke eigenvalues of f and g, W denotes a rapidly decreasing function (depending essentially on the infinity types of f and g), and N is of size about q. As is customary in the subconvexity problem, we bound such sums using the amplification method, by evaluating their amplified second moment over an orthogonal basis of automorphic forms of level $[q, D]$ containing f (considering a slightly larger family of forms with level $[q, D]$ rather than q is a

[2]For definite quaternion algebras, there is also an analogue (see [DS90]), in which the role of the Shimura curves is played by ellipsoids in $\mathbf{R}^3$ (of positive curvature!); this is closely related to the application discussed in [M04a].

trick very useful to simplifying the computation; this trick also occurred in [M04a] and in a slightly different yet related context in [GZ86, Z01a, Z01b]). Then we analyze the amplified second moment by applying Kuznetsov's trace formula *forwards* (i.e., from sums of Fourier coefficients of modular forms to sums of Kloosterman sums) and by applying Voronoi's summation formulae to the Fourier coefficients of g. We arrive at sums whose archetypical example is of the form

$$\Sigma(\ell,1;q,Y) := \sum_h G_{\chi_f\chi_g}\big(h;[q,D]\big) \sum_{\ell m-n=h} \lambda_g(m)\lambda_g(n)\mathcal{W}\left(\frac{m}{q},\frac{n}{\ell Y}\right), \tag{1.5}$$

where $G_{\chi_f\chi_g}\big(h;[q,D]\big)$ denotes the Gauss sum, $\ell \leqslant L^2$ is an integer coming from the amplifier (of length L), L is a very small (positive) power of q, $Y \leqslant q$, and $\mathcal{W}$ is a bounded rapidly decreasing function. The sum $\Sigma(\ell,1;q,Y)$ splits naturally into the contribution of the $h=0$ term, which in our case is zero by the assumption that $\chi_f\chi_g$ is not trivial, and our main problem is to evaluate the remaining terms. From the bound for Gauss sums, the trivial bound is given by

$$\Sigma(\ell,1;q,Y) \ll_{g,\varepsilon} \sqrt{q^*}q^{2+\varepsilon}$$

for any $\varepsilon > 0$, where q^* denotes the conductor of $\chi_f\chi_g$; on the other hand, in order to solve our given subconvexity problem, we need a bound of the form

$$\Sigma\big(\ell,1;q,Y\big) \ll_g q^{2-\delta} \tag{1.6}$$

for some $\delta > 0$, uniformly in $\ell \leqslant L^2$; thus one sees, from this crude analysis, that the problem gets harder as q^* gets larger. For $h \neq 0$, the innermost sum

$$\sum_{\ell m-n=h} \lambda_g(m)\lambda_g(n)\mathcal{W}\left(\frac{m}{q},\frac{n}{\ell Y}\right)$$

is called a *shifted convolution sum* (the case $h=0$ corresponding to a partial Rankin–Selberg convolution sum), and the problem of bounding non-trivially such a sum (the trivial bound being $\ll_g Y^{1+\varepsilon}$) is known as the *Shifted Convolution Problem*. There are at least two ways to handle this problem:

- The first one is via *the δ-symbol method*[3] which has been used for instance in [DFI93, DFI94a], in [KMV02], and—in a different form—in [J99, H03a, H03b]; the method builds on a formula for the Kronecker symbol $\delta_{\ell m-n=h}$ in terms of additive characters of small moduli (i.e., of size $\sim \sqrt{\ell q}$ instead of ℓq). Plugging this formula into the shifted convolution sum separates the variable m from n; then a double application of Voronoi's summation formula for g yields a linear combination of Kloosterman sums in a short range. In this case, Weil's bound (in fact, any non-trivial bound) for the Kloosterman sums suffices to solve the Shifted Convolution Problem.
- The second one was suggested by Selberg [S65] in 1965; it was worked out in some cases by Good, Jutila and others but was made effective and general by Sarnak in [S01]. It builds on the possibility of decomposing spectrally the shifted convolution sum above, in terms of a basis of automorphic cusp forms (and Eisenstein series) of level $D\ell$ and of their h-th Fourier coefficient. The solution of the Shifted Convolution Problem then follows from fairly non-trivial estimates for certain triple products (cf. [S94, BR99, KSt02]) and from a(ny) non-trivial approximation to the Ramanujan–Petersson conjecture, by which we mean that the following hypothesis is satisfied for some $\theta < \frac{1}{2}$ (in fact, H_θ is now known for $\theta = \frac{7}{64}$ [KS03]):

 Hypothesis H_θ. *For any cuspidal automorphic form π on $\mathrm{GL}_2(\mathbf{Q})\backslash\mathrm{GL}_2(\mathbf{A_Q})$ with local Hecke parameters $\alpha_\pi^{(1)}(p)$, $\alpha_\pi^{(2)}(p)$ for $p<\infty$ and $\mu_\pi^{(1)}(\infty)$, $\mu_\pi^{(2)}(\infty)$, we have the bounds*

$$|\alpha_\pi^{(j)}(p)| \leqslant p^\theta,\ j=1,2$$

[3] a descendent of the classical circle method with the Kloosterman refinement

$$(resp.\ |\Re\mu_\pi^{(j)}(\infty)| \leqslant \theta,\ j=1,2)$$

provided π_p (resp. π_∞) is unramified.

When it can be applied, the latter approach is less elementary but (now) more powerful than the δ-symbol method (in part because Weil's bound for Kloosterman sums corresponds to Hypothesis $H_{\frac{1}{4}}$, while $H_{\frac{7}{64}}$ is currently known); it is in particular the case when g is holomorphic [S01, M04a] or for *unbalanced* shifted convolution sums (i.e., when h is small compared to the typical sizes of the variables m and n [S01], or when one of the variables m and n is much smaller than the other). In the case of Maass forms and in a *balanced* situation (i.e., when h is comparable in size with the other variables m and n, as is the case in the present paper), realizing the spectral expansion of the shifted convolution sum is not so immediate (cf. [H03b]). This technical problem is very similar to the kind of difficulty Kuznetsov dealt with when he established his formula with arbitrary test functions on the Kloosterman sum side: in other words, the spectral expansion of the shifted convolution sums has to include contributions from the full automorphic spectrum (i.e., including contribution from the discrete series). In a recent paper, Motohashi [Mo04] has produced such a spectral expansion—confirming the above expectations—but there are several technical issues that need to be addressed before one can apply it to the present problem; we will return to this in the forthcoming work [HM05].

In fact, the two methods discussed above turn out to be closely related and in this paper we exploit this connection to obtain the spectral expansion by following a rather indirect path. Our starting point is the δ-symbol method or, more precisely, another variant of the circle method due to Jutila [J92, J96]: this variant gives us the extra luxury of selecting the moduli of the additive characters to be divisible by $D\ell$ and hence provides us with considerable simplification in the forthcoming argument—it was also employed by the first author in [H03a] to give a simplified solution of the subconvexity problem for the L-function of a Maass form twisted by a character. We then apply the Voronoi summation formula twice getting a sum of Kloosterman sums of moduli divisible by $D\ell$. Finally, we expand spectrally this expression by applying Kuznetsov's formula a second time but *backwards* (i.e., from sums of Kloosterman sums to sums of Fourier coefficients of automorphic forms) transforming the shifted convolution sum into a weighted sum of the h-th Fourier coefficient over a Hecke eigenbasis of automorphic forms of level $D\ell$. Note that this process is (of course) not involutory and that in this way we encounter not only Maass but also holomorphic forms. Of course, this is just another way to realize (up to an error term) *the same* spectral expansion of the shifted convolution sums discussed in the second method above. There are, however, some differences: in particular, triple products of automorphic forms are absent from our argument as well as the delicate issue of showing their exact exponential decay; instead, we use more elementary estimates on Bessel transforms[4].

In any case, the spectral expansion alone is insufficient, even under Hypothesis H_0, to get the bound (1.6) when q^* is large, but now we can proceed as in [M04a] by combining the spectral expansion with the averaging over the h variable and by exploiting the oscillation of the Gauss sum. The problem turns out to be reduced to a collection of subconvexity problems for the twisted L-series $L(\psi\otimes\chi_f\chi_g, s)$ as ψ ranges over a basis of Maass forms, holomorphic forms and Eisenstein series of level $D\ell$, a question which has already been solved (but at this point we need strongly the polynomial dependency with respect to the auxiliary parameters).

Remark 1.5. The process of applying Kuznetsov's formula backwards—after an application of the δ-symbol method—for the resolution of the shifted convolution problem was carried out by Jutila [J99] in the full level case ($D\ell = 1$); for general levels, where exceptional eigenvalues cannot be excluded, this possibility was already considered in [DFI94a], but not pursued further—and in fact, this was not really crucial by comparison with our present problem—cf. *loc. cit.* p.210: "*We shall use Weil's bound for Kloosterman sums rather than the spectral theory of automorphic forms since*

[4]Nevertheless, the two approaches are closely related via representation theory and Bessel models.

the latter approach would require us to deal with the congruence subgroup $\Gamma_0(ab)$ facing intrinsic difficulties with small eigenvalues. The results obtained this way would not be good enough for large a, b." We remark that the last sentence is now obsolete but only because of the progress made in direction of the Ramanujan–Petersson conjecture (more precisely, because Hypothesis H_θ holds for some $\theta < \frac{1}{4}$). We also note that, applied to our present situation, this process is nevertheless robust: Theorem 1 would remain valid (with a weaker subconvex exponent) even under Hypothesis H_θ for any given $\theta < \frac{1}{2}$. The reason is that we need to solve the Shifted Convolution Problem either for a few individual h's but of relatively large size (when q^* is small), or for h's varying in a wider range but then on average (when q^* is large), or for any intermediate case in between these two configurations, thus keeping the distortions created by small eigenvalues limited. In other words, we don't need to resolve an individual Shifted Convolution Problem for a small h. We also note that in a recent preprint [Bl04], V. Blomer gave a striking application of Jutila's method to the subconvexity problem for twisted L-function. His argument does not seem sufficient to solve the present subconvexity problem in full generality, but it can be combined with the one of this paper to obtain a somewhat stronger subconvex exponent in Theorem 1. See [BHM05a] for a first step in this direction.

1.3. **Recent developments regarding the Subconvexity Problem.** During the final stage of preparation of this paper (Summer of 2004), new striking cases of the subconvexity problem have been announced by J. Bernstein and A. Reznikov on the one hand, and by A. Venkatesh on the other hand; these results are now available in print [BR05, V05]. More precisely, Bernstein and Reznikov considered the subconvexity problem in the spectral parameter aspect and established a subconvex bound for the central value $L\left(\pi_\phi \otimes \pi_{\phi'} \otimes \pi_{\phi''}, \frac{1}{2}\right)$ of the triple product L-function associated to three Hecke–Maass cusp forms (over $\mathbf{Q}$), two of which (ϕ' and ϕ'', say) are fixed and the remaining one (ϕ) has large Laplacian eigenvalue. In a different direction, A. Venkatesh considered the subconvexity problem in the level aspect for automorphic L-functions over a general number field F.Amongst other cases, he obtained subconvex bounds for the standard L-function $L(\pi, s)$, the Rankin–Selberg L-function $L(\pi \otimes \pi', s)$, and the central value of the triple product L-function $L\left(\pi \otimes \pi' \otimes \pi'', \frac{1}{2}\right)$, where π' and π'' are some fixed cuspidal automorphic representations of $\mathrm{GL}_2(\mathbf{A}_F)$ and π is a cuspidal automorphic representation of $\mathrm{GL}_2(\mathbf{A}_F)$ with large conductor and *trivial central character*: in particular, these bounds generalize, to any number field, the subconvex bounds obtained in [DFI93, DFI94a, KMV02]. The proof of these new subconvexity cases build on soft but powerful methods which are very different from the ones developed so far: indeed these methods are geometric in nature, and rely on the expression of the central values of automorphic L-functions in terms of periods of automorphic forms; then, the subconvexity problem becomes tantamount to bounding non-trivially these periods. This is achieved by using a beautiful mix of representation theoretic, analytic and/or ergodic theoretic arguments. This is in sharp contrast with the former approaches (like the one presented here) where the relationship between central values and automorphic period is exploited, but only *after* the subconvex bound has been obtained (via the method of amplified moments for instance), to deduce an equidistribution result from subconvexity. Although these alternative approaches are very different from the ones coming from analytic number theory, an astute reader will nevertheless identify some common patterns on each side: this is especially striking when one compares [V05] with say [DFI94a] or [KMV02]. In fact, it seems plausible that one can pursue the analogy further so as, for example, to incorporate some features of the present paper within the framework developed in [V05]: this would yield to a generalization of Theorem 1 (which is not covered by the results of [V05]) to GL_2-automorphic forms on a general number field.

The present paper is organized as follows. In the next section we collect and recall several general facts on the analytic theory of automorphic forms (like Voronoi summation formulae and estimates for Fourier coefficients of modular forms); the material presented there is mostly standard but on several occasions we have not been able to find the proofs of several estimates given in the generality required by the present paper, so we have provided proofs on these occasions. Section 3 reduces our subconvexity problem for Rankin–Selberg L-functions to an estimate for sums of shifted convolution sums of type (1.5). As this reduction is entirely similar to Section 2 of [M04a], we skip most of the details. Sections 4 and 5 constitute the technical core of the paper: there we resolve the given shifted convolution problem by combining Jutila's variant of the circle method with the Kuznetsov trace formula and subconvexity estimates for twisted L-functions. Section 6 contains the application of our subconvexity bounds to the problem of equidistribution of short orbits of Heegner points on Shimura curves associated to indefinite quaternion algebras over $\mathbf{Q}$.

Acknowledgements. This paper was written in part during the first author's doctoral studies at Princeton University under the guidance of Peter Sarnak, and during the second author's visits at the Caltech Mathematics Department, at the Courant Institute and at the Fields Institute on the occasion of the Thematic Program on Automorphic Forms organized by H. Kim and R. Murty. We want to thank these institutions for their support and their pleasant working conditions. We also thank A. Booker, J. Cogdell, P. Cohen, D. Ramakrishnan, E. Ullmo, A. Venkatesh and S. W. Zhang for several helpful discussions related to this paper, as well as the referees for corrections and useful comments. We would like to address a special thank to V. Blomer: during the writing of the joint paper with him [BHM05a], Valentin found an error in the earlier version of this paper (in Lemma 5.1). Then, our attempts to correct the consequences of this error led us to produce a very significant improvement over the previous version of this paper. Finally, this paper would not exist without many conversations with and encouragements from Henryk Iwaniec and Peter Sarnak over the years, starting from the 1999/2000 special year on *Analytic Number theory and L-functions* at the Institute for Advanced Study, and it is a great pleasure to acknowledge them here.

2. Review of automorphic forms

2.1. **Maass forms.** Let k and D be positive integers, and χ be a character of modulus D. An automorphic function of weight k, level D and nebentypus χ is a function $g : \mathbf{H} \to \mathbf{C}$ satisfying, for any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in the congruence subgroup $\Gamma_0(D)$, the automorphy relation

$$g_{|_k\gamma}(z) := j_\gamma(z)^{-k} g(\gamma z) = \chi(d) g(z),$$

where

$$\gamma z := \frac{az+b}{cz+d} \qquad \text{and} \qquad j_\gamma(z) := \frac{cz+d}{|cz+d|} = \exp\bigl(i \arg(cz+d)\bigr).$$

We denote by $\mathcal{L}_k(D,\chi)$ the L^2-space of automorphic functions of weight k with respect to the Petersson inner product

$$\langle g_1, g_2 \rangle := \int\limits_{\Gamma_0(D)\backslash \mathbf{H}} g_1(z)\overline{g_2}(z) \frac{dxdy}{y^2}.$$

By the theory of Maass and Selberg, $\mathcal{L}_k(D,\chi)$ admits a spectral decomposition into eigenspaces of the Laplacian of weight k

$$\Delta_k := -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) + iky \frac{\partial}{\partial x}.$$

The spectrum of Δ_k has two components: the discrete spectrum spanned by the square-integrable smooth eigenfunctions of Δ_k (the Maass cusp forms), and the continuous spectrum spanned by the Eisenstein series $\{E_{\mathfrak{a}}(z,s)\}_{\{\mathfrak{a},\ s \text{ with } \Re s = \frac{1}{2}\}}$: any $g \in \mathcal{L}_k(D,\chi)$ decomposes as

$$g(z) = \sum_{j\geqslant 0} \langle g, u_j \rangle u_j(z) + \sum_{\mathfrak{a}} \frac{1}{4\pi i} \int_{\Re s = \frac{1}{2}} \langle g, E_{\mathfrak{a}}(*,s) \rangle E_{\mathfrak{a}}(z,s)\, ds,$$

where $u_0(z)$ is a constant function of Petersson norm 1, $\mathcal{B}_k(D,\chi) = \{u_j\}_{j\geqslant 1}$ denotes an orthonormal basis of Maass cusp forms and $\{\mathfrak{a}\}$ ranges over the singular cusps of $\Gamma_0(D)$ relative to χ. The

Eisenstein series $E_{\mathfrak{a}}(z,s)$ (which for $\Re s = \frac{1}{2}$ are defined by analytic continuation) are eigenfunctions of Δ_k with eigenvalue $\lambda(s) = s(1-s)$.

A Maass cusp form g decays exponentially near the cusps. It admits a Fourier expansion for each cusp with its zero-th Fourier coefficient vanishing; in particular, for the cusp at ∞, the Fourier expansion takes the form

$$g(z) = \sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} \rho_g(n) W_{\frac{n}{|n|}\frac{k}{2},it}(4\pi|n|y)e(nx), \tag{2.1}$$

where $W_{\alpha,\beta}(y)$ is the Whittaker function, and $\left(\frac{1}{2}+it\right)\left(\frac{1}{2}-it\right)$ is the eigenvalue of g. The Eisenstein series has a similar Fourier expansion

$$E_{\mathfrak{a}}\left(z,\tfrac{1}{2}+it\right) = \delta_{\mathfrak{a}=\infty}y^{\frac{1}{2}+it} + \phi_{\mathfrak{a}}\left(\tfrac{1}{2}+it\right)y^{\frac{1}{2}-it} + \sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} \rho_{\mathfrak{a}}(n,t) W_{\frac{n}{|n|}\frac{k}{2},it}(4\pi|n|y)e(nx),$$

where $\phi_{\mathfrak{a}}\left(\frac{1}{2}+it\right)$ is the entry $(\infty,\mathfrak{a})$ of the scattering matrix.

2.2. **Holomorphic forms.** Let $\mathcal{S}_k(D,\chi)$ denote the space of holomorphic cusp forms of weight k, level D and nebentypus χ, that is, the space of holomorphic functions $g : \mathbf{H} \to \mathbf{C}$ satisfying

$$g(\gamma z) = \chi(\gamma)(cz+d)^k g(z)$$

for every $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(D)$ and vanishing at every cusp. Such a form has a Fourier expansion at ∞ of the form

$$g(z) = \sum_{n\geqslant 1} \rho_g(n)(4\pi n)^{\frac{k}{2}} e(nz).$$

We recall that the cuspidal spectrum of $\mathcal{L}_k(D,\chi)$ is composed of the constant functions (if $k=0$, χ is trivial), Maass cusp forms with eigenvalues $\lambda_g = (\frac{1}{2}+it_g)(\frac{1}{2}-it_g) > 0$ (if k is odd, one has $\lambda_g \geqslant \frac{1}{4}$) which are obtained from the Maass cusp forms of weight $\kappa \in \{0,1\}$, $\kappa \equiv k\,(2)$ by $\frac{k-\kappa}{2}$ applications of the Maass weight raising operator, and of Maass cusp forms with eigenvalues $\lambda = \frac{l}{2}(1-\frac{l}{2}) \leqslant 0$, $0 < l \leqslant k$, $l \equiv k\,(2)$ which are obtained by $\frac{k-l}{2}$ applications of the Maass weight raising operator to weight l Maass cusp forms given by $y^{l/2}g(z)$ for $g \in \mathcal{S}_l(D,\chi)$. In particular, if $g \in \mathcal{S}_k(D,\chi)$, then $y^{k/2}g(z)$ is a Maass form of weight k and eigenvalue $\frac{k}{2}(1-\frac{k}{2})$. Moreover, we note that our two definitions of the Fourier coefficients agree:

$$\rho_g(n) = \rho_{y^{k/2}g}(n).$$

In the sequel, we set

$$\mu_g := \begin{cases} 1+|t_g| & \text{for } g \text{ a Maass cusp form of weight 0 or 1;} \\ 1+\frac{k_g-1}{2} & \text{for } g \text{ a holomorphic cusp form of weight } k_g. \end{cases} \tag{2.2}$$

2.3. **Hecke operators.** We also recall that $\mathcal{L}_k(D,\chi)$ (and its subspace generated by Maass cusp forms) is acted on by the (commutative) algebra $\mathbf{T}$ generated by the Hecke operators $\{T_n\}_{n\geqslant 1}$ which satisfy the multiplicativity relation

$$T_m T_n = \sum_{d|(m,n)} \chi(d) T_{\frac{mn}{d^2}}.$$

We denote by $\mathbf{T}^{(D)}$ the subalgebra generated by $\{T_n\}_{(n,D)=1}$ and call a Maass cusp form which is an eigenform for $\mathbf{T}^{(D)}$ a *Hecke–Maass* cusp form. The elements of $\mathbf{T}^{(D)}$ are normal with respect to the Petersson inner product, therefore the cuspidal subspace admits a basis formed of Hecke–Maass cusp forms. For a Hecke–Maass cusp form g, the following relations hold:

$$\sqrt{n}\rho_g(\pm n) = \rho_g(\pm 1)\lambda_g(n) \quad \text{for } (n,D)=1, \tag{2.3}$$

where $\lambda_g(n)$ denotes the eigenvalue of T_n, and

$$\sqrt{m}\rho_g(m)\lambda_g(n) = \sum_{d|(m,n)} \chi(d)\rho_g\left(\frac{m}{d}\frac{n}{d}\right)\sqrt{\frac{mn}{d^2}}, \tag{2.4}$$

$$\sqrt{mn}\rho_g(mn) = \sum_{d|(m,n)} \chi(d)\mu(d)\rho_g\left(\frac{m}{d}\right)\sqrt{\frac{m}{d}}\lambda_g\left(\frac{n}{d}\right). \tag{2.5}$$

The primitive forms are defined to be the Hecke–Maass cusp forms orthogonal to the subspace of old forms. By Atkin–Lehner theory, these are automatically eigenforms for $\mathbf{T}$ and the relations (2.3) and (2.4) hold for any n. Moreover, if g is a Maass form not coming from a holomorphic form (i.e., if it_g is not of the form $\pm\frac{l-1}{2}$ for $1 \leqslant l \leqslant k$, $l \equiv k\,(2)$), then g is also an eigenform for the involution $Q_{\frac{1}{2}+it_g,k}$ of [DFI02, (4.65)], and one has the following relation between the positive and negative Fourier coefficients:

$$\rho_g(-n) = \varepsilon_g\rho_g(n) \quad \text{for } n \geqslant 1 \tag{2.6}$$

with

$$\varepsilon_g = \pm\frac{\Gamma\left(\frac{1}{2}+it_g+\frac{k}{2}\right)}{\Gamma\left(\frac{1}{2}+it_g-\frac{k}{2}\right)} \tag{2.7}$$

(cf. [DFI02, (4.70)]).

A primitive form g is *arithmetically normalized* if $\rho_g(1) = 1$.

2.4. **Voronoi summation formulae.** The modular properties of a cusp form $g \in \mathcal{L}_k(D,\chi)$ translate into various functional equations for Dirichlet series

$$D(g,x,s) := \sum_{n\geqslant 1}\sqrt{n}\rho_g(n)e(nx)n^{-s}$$

attached to additive twists of the Fourier coefficients $\rho_g(n)$. When $x = \frac{a}{c}$ is a rational number in lowest terms with denominator c divisible by the level D, the functional equation is particularly simple.

If g is induced from a holomorphic form of weight l, then by Appendix A.3 of [KMV02] (see also [DI90]),

$$D\left(g,\frac{a}{c},s\right) = i^l\chi(\overline{a})\left(\frac{c}{2\pi}\right)^{1-2s}\frac{\Gamma\left(1-s+\frac{l-1}{2}\right)}{\Gamma\left(s+\frac{l-1}{2}\right)}D\left(g,-\frac{\overline{a}}{c},1-s\right).$$

If g is not induced from a holomorphic form, then

$$D\left(g,\frac{a}{c},s\right) = i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\left\{\Psi^+_{k,it}(s)D\left(g,-\frac{\overline{a}}{c},1-s\right)+\Psi^-_{k,it}(s)D\left(Qg,\frac{\overline{a}}{c},1-s\right)\right\}, \tag{2.8}$$

where $\Psi^\pm_{k,it}(s)$ are meromorphic functions depending at most on k and it, $\frac{1}{4}+t^2$ is the Laplacian eigenvalue of g, and $Q = Q_{\frac{1}{2}+it,k}$ is the involution given in (4.65) of [DFI02]. In fact, we can assume that $Qg = \varepsilon g$ for some $\varepsilon = \pm 1$, and reduce the above to

$$D\left(g,\frac{a}{c},s\right) = i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\left\{\Psi^+_{k,it}(s)D\left(g,-\frac{\overline{a}}{c},1-s\right)+\varepsilon\Psi^-_{k,it}(s)D\left(g,\frac{\overline{a}}{c},1-s\right)\right\}. \tag{2.9}$$

For $k = 0$, $\Psi^\pm_{k,it}(s)$ are determined in Appendix A.4 of [KMV02] (see also [M88]):

$$\Psi^\pm_{0,it}(s) = \frac{\Gamma\left(\frac{1-s+it}{2}\right)\Gamma\left(\frac{1-s-it}{2}\right)}{\Gamma\left(\frac{s-it}{2}\right)\Gamma\left(\frac{s+it}{2}\right)} \mp \frac{\Gamma\left(\frac{2-s+it}{2}\right)\Gamma\left(\frac{2-s-it}{2}\right)}{\Gamma\left(\frac{1+s-it}{2}\right)\Gamma\left(\frac{1+s+it}{2}\right)}. \tag{2.10}$$

For $k \neq 0$, we will express $\Psi^{\pm}_{k,it}(s)$ in terms of the functions $\Phi^{\pm}_{k}(s,it)$ defined by (8.25) of [DFI02]:

$$\Phi_k^{\pm 1}(s,it) := \frac{\sqrt{\pi}}{4}\int_0^\infty \left\{ W_{\frac{k}{2},it}(4y) \pm \frac{\Gamma\left(\frac12+it+\frac{k}{2}\right)}{\Gamma\left(\frac12+it-\frac{k}{2}\right)} W_{-\frac{k}{2},it}(4y)\right\} y^{s-\frac12}\frac{dy}{y}. \tag{2.11}$$

Our starting point for establishing the functional equation is the identity

$$\frac{\pi^s}{4}\int_0^\infty g(x+iy)y^{s-\frac12}\frac{dy}{y} = \Phi_k^{\varepsilon}(s,it)D^{+1}(g,x,s)+\Phi_k^{-\varepsilon}(s,it)D^{-1}(g,x,s), \tag{2.12}$$

where

$$2D^{\pm 1}(g,x,s) = D(g,x,s) \pm D(g,-x,s).$$

In deriving this identity we use (2.1), (2.6), and (2.7) with the sign $\varepsilon = \pm 1$. The modularity of g implies, for any $y>0$,

$$g\left(\frac{a}{c}+\frac{iy}{c}\right) = i^k\chi(\overline{a})g\left(-\frac{\overline{a}}{c}+\frac{i}{cy}\right).$$

We integrate both sides against $y^{s-\frac12}\frac{dy}{y}$ to obtain, by (2.12),

$$\sum_{\pm}\Phi_k^{\pm\varepsilon}(s,it)D^{\pm 1}\left(g,\frac{a}{c},s\right) = i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\sum_{\pm}\Phi_k^{\pm\varepsilon}(1-s,it)D^{\pm 1}\left(g,-\frac{\overline{a}}{c},1-s\right).$$

The analogous equation holds when a is replaced by $-a$:

$$\sum_{\pm}\Phi_k^{\pm\varepsilon}(s,it)D^{\pm 1}\left(g,-\frac{a}{c},s\right) = i^k\chi(-\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\sum_{\pm}\Phi_k^{\pm\varepsilon}(1-s,it)D^{\pm 1}\left(g,\frac{\overline{a}}{c},1-s\right).$$

Using that $D^{\pm1}(g,-x,s) = \pm D^{\pm1}(g,x,s)$, and also that $\chi(-1)=(-1)^k$, we can infer that

$$\Phi_k^{\pm\varepsilon}(s,it)D^{\pm 1}\left(g,\frac{a}{c},s\right) = i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\Phi_k^{\pm\varepsilon(-1)^k}(1-s,it)D^{\pm(-1)^k}\left(g,-\frac{\overline{a}}{c},1-s\right).$$

It is important to note that the functions $\Phi_k^{\pm\varepsilon}(s,it)$ are not identically zero by $k\neq 0$ and Lemma 8.2 of [DFI02] (cf. (8.32) and (8.33) of [DFI02]). Therefore we can conclude that

$$\begin{aligned} D\left(g,\frac{a}{c},s\right) &= \sum_{\pm} D^{\pm 1}\left(g,\frac{a}{c},s\right) \\ &= i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\sum_{\pm}\frac{\Phi_k^{\pm\varepsilon(-1)^k}(1-s,it)}{\Phi_k^{\pm\varepsilon}(s,it)}D^{\pm(-1)^k}\left(g,-\frac{\overline{a}}{c},1-s\right). \end{aligned}$$

Combining this equation with

$$2D^{\pm 1}\left(g,-\frac{\overline{a}}{c},1-s\right) = D\left(g,-\frac{\overline{a}}{c},1-s\right) \pm D\left(g,\frac{\overline{a}}{c},1-s\right),$$

we find that (2.9) indeed holds with the following definition of $\Psi^{\pm}_{k,it}(s)$:

$$\Psi^{\pm}_{k,it}(s) = \frac{\Phi_k^{1}(1-s,it)}{\Phi_k^{(-1)^k}(s,it)} \pm \frac{\Phi_k^{-1}(1-s,it)}{\Phi_k^{-(-1)^k}(s,it)}.$$

This formula works for $k\neq 0$ and complements (2.10) which corresponds to $k=0$.

Using the calculations of [DFI02] we can express $\Psi^{\pm}_{k,it}(s)$ in more explicit terms. First, we use (8.34) of [DFI02] to see that

$$\Psi^{\pm}_{k,it}(s) = \frac{\Phi_k^{1}(1-s,it)}{\Phi_k^{1}(s,-it)} \pm \frac{\Phi_k^{-1}(1-s,it)}{\Phi_k^{-1}(s,-it)}.$$

Then we refer to Lemma 8.2 of [DFI02], the functional equation (8.36) of [DFI02], and the determination of the constant $\nu = \nu_k^\varepsilon = \pm 1$ in that functional equation (p.534 of [DFI02]) to derive that

$$\begin{aligned}\Psi^{\pm}_{k,it}(s) &= i^k \frac{\Gamma\left(\frac{1-s+it}{2}\right)\Gamma\left(\frac{1-s-it}{2}\right)}{\Gamma\left(\frac{s-it}{2}\right)\Gamma\left(\frac{s+it}{2}\right)} \mp i^k \frac{\Gamma\left(\frac{2-s+it}{2}\right)\Gamma\left(\frac{2-s-it}{2}\right)}{\Gamma\left(\frac{1+s-it}{2}\right)\Gamma\left(\frac{1+s+it}{2}\right)}, \qquad k \text{ even};\\ \Psi^{\pm}_{k,it}(s) &= i^{k-1} \frac{\Gamma\left(\frac{1-s+it}{2}\right)\Gamma\left(\frac{2-s-it}{2}\right)}{\Gamma\left(\frac{s-it}{2}\right)\Gamma\left(\frac{1+s+it}{2}\right)} \pm i^{k-1} \frac{\Gamma\left(\frac{2-s+it}{2}\right)\Gamma\left(\frac{1-s-it}{2}\right)}{\Gamma\left(\frac{1+s-it}{2}\right)\Gamma\left(\frac{s+it}{2}\right)}, \quad k \text{ odd}.\end{aligned}$$

Note that by (2.10) this formula is also valid for $k = 0$.

We can simplify the above expressions for $\Psi^{\pm}_{k,it}(s)$ using the functional equation and the duplication formula for Γ:

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}, \qquad \Gamma(s)\Gamma(\tfrac{1}{2}+s) = \sqrt{\pi}2^{1-2s}\Gamma(2s).$$

For even k, we obtain

$$\begin{aligned}\Psi^{+}_{k,it}(s) &= i^k\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\bigl\{-\cos(\pi s)\bigr\};\\ \Psi^{-}_{k,it}(s) &= i^k\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\bigl\{\cos(\pi it)\bigr\}.\end{aligned} \tag{2.13}$$

For odd k, we obtain

$$\begin{aligned}\Psi^{+}_{k,it}(s) &= i^{k-1}\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\bigl\{\sin(\pi s)\bigr\};\\ \Psi^{-}_{k,it}(s) &= i^{k-1}\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\bigl\{-\sin(\pi it)\bigr\}.\end{aligned} \tag{2.14}$$

These identities enable us to derive a general Voronoi-type summation formula for the coefficients $\rho_g(n)$ of an arbitrary cusp form $g \in \mathcal{L}_k(D,\chi)$. Special cases of this formula already appeared in [M88, DI90, KMV02].

Proposition 2.1. *Let D be a positive integer, χ be a character of modulus D, and $g \in \mathcal{L}_k(D,\chi)$ be a cusp form with spectral parameter $t = t_g$. Let $c \equiv 0\,(D)$ and a be an integer coprime to c. If $F \in C^\infty(\mathbf{R}^{\times,+})$ is a Schwartz class function vanishing in a neighborhood of zero, then*

$$\sum_{n\geqslant 1}\sqrt{n}\rho_g(n)e\Bigl(n\frac{a}{c}\Bigr)F(n) = \frac{\chi(\overline{a})}{c}\sum_{\pm}\sum_{n\geqslant 1}\sqrt{n}\rho_g^{\pm}(n)e\Bigl(\mp n\frac{\overline{a}}{c}\Bigr)\mathcal{F}^{\pm}\Bigl(\frac{n}{c^2}\Bigr). \tag{2.15}$$

In this formula,

$$\rho_g^{+}(n) := \rho_g(n), \qquad \rho_g^{-}(n) := \rho_{Qg}(n) = \frac{\Gamma\bigl(\frac{1}{2}+it-\frac{k}{2}\bigr)}{\Gamma\bigl(\frac{1}{2}+it+\frac{k}{2}\bigr)}\rho_g(-n),$$

and

$$\mathcal{F}^{\pm}(y) := \int_0^\infty F(x)J_g^{\pm}\bigl(4\pi\sqrt{xy}\bigr)\,dx, \tag{2.16}$$

where

- $$J_g^{+}(x) := 2\pi i^l J_{l-1}(x), \qquad J_g^{-}(x) := 0,$$
  if g is induced from a holomorphic form of weight l;
- $$J_g^{+}(x) := \frac{-\pi}{\operatorname{ch}(\pi t)}\bigl\{Y_{2it}(x)+Y_{-2it}(x)\bigr\}, \qquad J_g^{-}(x) := 4\operatorname{ch}(\pi t)K_{2it}(x),$$
  if k is even, and g is not induced from a holomorphic form;
- $$J_g^{+}(x) := \frac{\pi}{\operatorname{sh}(\pi t)}\bigl\{Y_{2it}(x)-Y_{-2it}(x)\bigr\}, \qquad J_g^{-}(x) := -4i\operatorname{sh}(\pi t)K_{2it}(x),$$
  if k is odd, and g is not induced from a holomorphic form.

We outline the proof for non-holomorphic forms g. We represent the left hand side of (2.15) as an inverse Mellin transform

$$\sum_{n\geqslant 1}\sqrt{n}\rho_g(n)e\Big(n\frac{a}{c}\Big)F(n)=\frac{1}{2\pi i}\int_{(2)}\hat{F}(s)D\left(g,\frac{a}{c},s\right)ds.$$

By the functional equation (2.8), the right hand side can be rewritten as

$$i^k\chi(\overline{a})\frac{1}{2\pi i}\int_{(2)}\hat{F}(s)\left(\frac{c}{\pi}\right)^{1-2s}\Psi^+_{k,it}(s)D\left(g,-\frac{\overline{a}}{c},1-s\right)ds$$
$$+i^k\chi(\overline{a})\frac{1}{2\pi i}\int_{(2)}\hat{F}(s)\left(\frac{c}{\pi}\right)^{1-2s}\Psi^-_{k,it}(s)D\left(Qg,\frac{\overline{a}}{c},1-s\right)ds.$$

By changing s to $1-\frac{s}{2}$ and shifting the contour, we see that this is the same as

$$\begin{aligned}(2.17)\qquad &i^k\chi(\overline{a})\frac{1}{2\pi i}\int_{(2)}\hat{F}\left(1-\frac{s}{2}\right)\left(\frac{c}{\pi}\right)^{s-1}\Psi^+_{k,it}\left(1-\frac{s}{2}\right)D\left(g,-\frac{\overline{a}}{c},\frac{s}{2}\right)\frac{ds}{2}\\ &+i^k\chi(\overline{a})\frac{1}{2\pi i}\int_{(2)}\hat{F}\left(1-\frac{s}{2}\right)\left(\frac{c}{\pi}\right)^{s-1}\Psi^-_{k,it}\left(1-\frac{s}{2}\right)D\left(Qg,\frac{\overline{a}}{c},\frac{s}{2}\right)\frac{ds}{2}.\end{aligned}$$

Using (2.13) and (2.14) it is straightforward to check that

$$i^k\Psi^\pm_{k,it}\left(1-\frac{s}{2}\right)=\frac{2}{\pi}\widehat{J^\pm_g(4x)}(s),$$

so that

$$\hat{F}\left(1-\frac{s}{2}\right)i^k\Psi^\pm_{k,it}\left(1-\frac{s}{2}\right)=2\pi^{s-1}\widehat{\mathcal{F}^\pm(y)}\left(\frac{s}{2}\right)=2\pi^{s-1}\widehat{\mathcal{F}^\pm(y^2)}(s),$$

where $\mathcal{F}^\pm$ is the Hankel-type transform of F given by (2.16). In particular,

$$i^k\frac{1}{2\pi i}\int_{(2)}\hat{F}\left(1-\frac{s}{2}\right)\left(\frac{c}{\pi}\right)^{s-1}\Psi^\pm_{k,it}\left(1-\frac{s}{2}\right)n^{-\frac{s}{2}}\frac{ds}{2}=\frac{1}{c}\mathcal{F}^\pm\Big(\frac{n}{c^2}\Big),$$

and this shows that (2.17) is equal to the right hand side of (2.15). But (2.17) is also equal to the left hand side of (2.15), therefore the proof is complete.

2.5. **Spectral summation formulae.** The following spectral summation formulae form an important tool for the analytic theory and the harmonic analysis on spaces of modular forms.

The first one is Petersson's formula concerning the case of holomorphic forms (cf. Theorem 9.6 in [I95]): let $\mathcal{B}^h_k(D,\chi)$ denote an orthonormal basis of the space of holomorphic cusp forms of weight $k\geqslant 1$, level D and nebentypus χ.

Proposition 2.2. *For any positive integers m,n, one has*

$$(2.18)\quad 4\pi\Gamma(k-1)\sqrt{mn}\sum_{f\in\mathcal{B}^h_k(D,\chi)}\overline{\rho_f}(m)\rho_f(n)=\delta_{m,n}+2\pi i^{-k}\sum_{c\equiv 0\,(D)}\frac{S_\chi(m,n;c)}{c}J_{k-1}\left(\frac{4\pi\sqrt{mn}}{c}\right).$$

Here $S_\chi(m,n;c)$ is the twisted Kloosterman sum

$$S_\chi(m,n;c):=\sum_{\substack{x(c)\\(x,c)=1}}\chi(x)e\left(\frac{mx+n\overline{x}}{c}\right).$$

Let $\mathcal{B}_k(D,\chi)=\{u_j\}_{j\geqslant 1}$ be an orthonormal basis of the cuspidal part of $\mathcal{L}_k(D,\chi)$ formed of Maass forms with eigenvalues $\lambda_j=\frac{1}{4}+t_j^2$ and Fourier coefficients $\rho_j(n)$. The following spectral summation formula is a combination of [DFI02, Proposition 5.2], a slight refinement of [DFI02, (14.7)], [DFI02, Proposition 17.1], and [DFI02, Lemma 17.2].

Proposition 2.3. *For any integer $k \geqslant 0$ and any $A > 0$, there exist functions $\mathcal{H}(t) : \mathbf{R} \cup i\mathbf{R} \to (0, \infty)$ and $\mathcal{I}(x) : (0, \infty) \to \mathbf{R} \cup i\mathbf{R}$ depending on k and A such that*

$$\mathcal{H}(t) \gg (1 + |t|)^{k-16} e^{-\pi|t|}; \tag{2.19}$$

for any integer $j \geqslant 0$,

$$x^j \mathcal{I}^{(j)}(x) \ll \left(\frac{x}{1+x}\right)^{A+1} (1+x)^{1+j}; \tag{2.20}$$

(here the implied constants depend only on A and j, that is, they are independent of k) and for any positive integers m, n,

$$\sqrt{mn} \sum_{j \geqslant 1} \mathcal{H}(t_j) \overline{\rho_j}(m) \rho_j(n) + \sqrt{mn} \sum_{\mathfrak{a}} \frac{1}{4\pi} \int_{-\infty}^{+\infty} \mathcal{H}(t) \overline{\rho_{\mathfrak{a}}}(m,t) \rho_{\mathfrak{a}}(n,t)\, dt$$

$$= c_A \delta_{m,n} + \sum_{c \equiv 0\,(D)} \frac{S_\chi(m,n;c)}{c} \mathcal{I}\left(\frac{4\pi\sqrt{mn}}{c}\right).$$

Here $c_A > 0$ depends only on A.

It will also be useful to have an even more general form of the summation formulae above, namely when $\mathcal{I}(x)$ is replaced by an arbitrary test function. This is one of Kuznetsov's main results (in the case of full level). His formula was generalized in various ways, mainly by Deshouillers–Iwaniec (to arbitrary levels), by Proskurin (for arbitrary integral or half-integral weights); see also [CPS90] for an illuminating discussion of this formula from the representation theoretic point of view. In this paper, we will only need the trivial nebentypus case [I87, Theorems 9.4, 9.5, 9.7]. To save notation, we set $\mathcal{B}_0(D) = \{u_j\}_{j \geqslant 1}$ (resp. $\mathcal{B}_k^h(D)$) for an orthonormal basis of the space of weight 0 Maass cusp forms (resp. of the space of holomorphic cusp forms of weight k) of level D, and trivial nebentypus.

Theorem 3. *Let m, n, D be positive integers and $\varphi \in C_c^\infty(\mathbf{R}^{\times,+})$. One has*

$$\frac{1}{4\sqrt{mn}} \sum_{c \equiv 0\,(D)} \frac{S(m,n;c)}{c} \varphi\left(\frac{4\pi\sqrt{mn}}{c}\right) = \sum_{k \equiv 0\,(2)} \Gamma(k) \tilde{\varphi}(k-1) \sum_{f \in \mathcal{B}_k^h(D)} \overline{\rho_f}(m) \rho_f(n) \tag{2.21}$$

$$+ \sum_{j \geqslant 1} \frac{\hat{\varphi}(t_j)}{\mathrm{ch}(\pi t_j)} \overline{\rho_j}(m) \rho_j(n) + \frac{1}{4\pi} \sum_{\mathfrak{a}} \int_{-\infty}^{+\infty} \frac{\hat{\varphi}(t)}{\mathrm{ch}(\pi t)} \overline{\rho_{\mathfrak{a}}}(m,t) \rho_{\mathfrak{a}}(n,t)\, dt,$$

and

$$\frac{1}{4\sqrt{mn}} \sum_{c \equiv 0\,(D)} \frac{S(m,-n;c)}{c} \varphi\left(\frac{4\pi\sqrt{mn}}{c}\right) =$$

$$\sum_{j \geqslant 1} \frac{\check{\varphi}(t_j)}{\mathrm{ch}(\pi t_j)} \overline{\rho_j}(m) \rho_j(n) + \frac{1}{4\pi} \sum_{\mathfrak{a}} \int_{-\infty}^{+\infty} \frac{\check{\varphi}(t)}{\mathrm{ch}(\pi t)} \overline{\rho_{\mathfrak{a}}}(m,t) \rho_{\mathfrak{a}}(n,t)\, dt,$$

where the Bessel transforms are defined by

$$\tilde{\varphi}(k-1) := \int_0^\infty \varphi(x) i^k J_{k-1}(x) \frac{dx}{x};$$

$$\hat{\varphi}(t) := \int_0^\infty \varphi(x) \frac{-\pi}{2\,\mathrm{ch}(\pi t)} \left\{Y_{2it}(x) + Y_{-2it}(x)\right\} \frac{dx}{x}; \tag{2.22}$$

$$\check{\varphi}(t) := \int_0^\infty \varphi(x) 4\,\mathrm{ch}(\pi t) K_{2it}(x) \frac{dx}{x}. \tag{2.23}$$

Remark 2.1. The kernels in (2.22) and (2.23) can be expressed alternately as

$$\frac{-\pi}{2\operatorname{ch}(\pi t)}\{Y_{2it}(x)+Y_{-2it}(x)\}=\frac{\pi i}{2\operatorname{sh}(\pi t)}\{J_{2it}(x)-J_{-2it}(x)\};$$
$$4\operatorname{ch}(\pi t)K_{2it}(x)=\frac{\pi i}{\operatorname{sh}(\pi t)}\{I_{2it}(x)-I_{-2it}(x)\}.$$

2.6. **Bounds for the Fourier coefficients of cusp forms.** In this section we recall several (now) standard bounds for the Fourier coefficients of cusp forms; references to proofs can be found in Section 2.5 of [M04a].

If g is an L^2-normalized primitive Maass cusp form of level D, weight $\kappa\in\{0,1\}$ and eigenvalue $\frac14+t_g^2$, then from [DFI02] and [HL94] we have for any $\varepsilon>0$ (cf. (2.2)),

$$(D\mu_g)^{-\varepsilon}\left(\frac{\operatorname{ch}(\pi t_g)}{D\mu_g^\kappa}\right)^{1/2}\ll_\varepsilon|\rho_g(1)|\ll_\varepsilon(D\mu_g)^{\varepsilon}\left(\frac{\operatorname{ch}(\pi t_g)}{D\mu_g^\kappa}\right)^{1/2}. \tag{2.24}$$

If $g\in\mathcal{S}_k(D,\chi)$ is an L^2-normalized primitive holomorphic cusp form, then

$$\frac{(Dk)^{-\varepsilon}}{(D\Gamma(k))^{1/2}}\ll_\varepsilon|\rho_g(1)|\ll_\varepsilon\frac{(Dk)^{\varepsilon}}{(D\Gamma(k))^{1/2}}. \tag{2.25}$$

For Hecke eigenvalues, Hypothesis H_θ gives in general the individual bound

$$|\lambda_g(n)|\leqslant\tau(n)n^\theta. \tag{2.26}$$

Note that this bound remains true when n is divisible by ramified primes. Moreover, if g is holomorphic, it follows from Deligne's proof of the Ramanujan–Petersson conjecture that (2.26) holds with $\theta=0$. Hence for all $n\geqslant1$ and for any $\varepsilon>0$ we have by (2.3)

$$\sqrt{n}\rho_g(n)\ll_\varepsilon\begin{cases}(nD\mu_g)^\varepsilon\left(\dfrac{\operatorname{ch}(\pi t_g)}{D\mu_g^\kappa}\right)^{1/2}n^\theta & \text{for } g\in\mathcal{L}_\kappa(D,\chi),\ \kappa\in\{0,1\};\\ \dfrac{(nDk)^\varepsilon}{(D\Gamma(k))^{1/2}} & \text{for } g\in\mathcal{S}_k(D,\chi).\end{cases} \tag{2.27}$$

The implied constant depends at most on ε and is effective. In fact, for a Maass cusp form g of weight $\kappa\in\{0,1\}$, Rankin–Selberg theory implies that the Ramanujan–Petersson bound holds on average: one has, for all $X\geqslant1$ and all $\varepsilon>0$,

$$\sum_{n\leqslant N}|\lambda_g(n)|^2\ll_\varepsilon(D\mu_gN)^\varepsilon N. \tag{2.28}$$

It will also be useful to introduce the function

$$\sigma_g(n):=\sum_{d|n}|\lambda_g(d)|.$$

This function is almost multiplicative,

$$(mn)^{-\varepsilon}\sigma_g(mn)\ll\sigma_g(m)\sigma_g(n)\ll(mn)^\varepsilon\sigma_g(mn),$$

and satisfies (from (2.28))

$$\sum_{n\leqslant N}\sigma_g(n)^2\ll(D\mu_gN)^\varepsilon N,$$

for all $N,\varepsilon>0$. In the above estimates, the implied constants depend on ε but not on g.

In several occasions, we will need a substitute for (2.27) when g is an L^2-normalized but not necessarily primitive Hecke–Maass cusp form. This estimate can be achieved on average over an orthonormal basis, and this is sufficient for our application.

Lemma 2.1. *Assume that Hypothesis H_θ holds. Let $\mathcal{B}_0(D,\chi)=\{u_j\}_{j\geqslant 0}$ denote an orthonormal Hecke eigenbasis of the space of Maass cusp forms of weight 0, level D and nebentypus χ. For any $n,T\geqslant 1$, one has*

$$\sum_{\substack{u_j\in\mathcal{B}_0(D,\chi)\\ |t_j|\leqslant T}}\frac{n|\rho_j(n)|^2}{\mathrm{ch}(\pi t_j)}\ll (nDT)^\varepsilon T^2 n^{2\theta}, \tag{2.29}$$

and for any $m,N,T\geqslant 1$, one has

$$\sum_{\substack{u_j\in\mathcal{B}_0(D,\chi)\\ |t_j|\leqslant T}}\sum_{n\leqslant N}\frac{mn|\rho_j(mn)|^2}{\mathrm{ch}(\pi t_j)}\ll (mNDT)^\varepsilon T^2 N m^{2\theta}. \tag{2.30}$$

Here the implied constants depend at most on ε.

Proof. Inequality (2.29) is a straightforward generalization of Lemma 2.3 in [M04a]. For (2.30), combine (2.5), (2.28), (2.29), and note that $\theta<\frac14$:

$$\begin{aligned}\sum_{\substack{u_j\in\mathcal{B}_0(D,\chi)\\ |t_j|\leqslant T}}\sum_{n\leqslant N}\frac{mn|\rho_j(mn)|^2}{\mathrm{ch}(\pi t_j)}&=\sum_{\substack{d|(mD)^\infty\\ d\leqslant N}}\sum_{\substack{u_j\in\mathcal{B}_0(D,\chi)\\ |t_j|\leqslant T}}\frac{md|\rho_j(md)|^2}{\mathrm{ch}(\pi t_j)}\sum_{\substack{n\leqslant N/d\\ (n,mD)=1}}|\lambda_j(n)|^2\\ &\ll_\varepsilon (mNDT)^\varepsilon T^2Nm^{2\theta}\sum_{\substack{d|(mD)^\infty\\ d\leqslant N}}d^{2\theta-1}\ll_\varepsilon (mNDT)^{2\varepsilon}T^2Nm^{2\theta}.\end{aligned}$$

□

Lemma 2.2. *For $k\geqslant 1$, let $\mathcal{B}_k^h(D,\chi)\subset\mathcal{S}_k(D,\chi)$ denote an orthonormal basis of the space of holomorphic cusp forms of weight k, level D and nebentypus χ. For any $T\geqslant 1$, any $n\geqslant 1$ and any $\varepsilon>0$, one has*

$$\sum_{k\leqslant K}\Gamma(k)\sum_{f\in\mathcal{B}_k^h(D,\chi)}n|\rho_f(n)|^2\ll (nDK)^\varepsilon K^2, \tag{2.31}$$

where the implied constant depends at most on ε.

2.7. **Bounds for exponential sums associated to cusp forms.** In this section we prove uniform bounds for exponential sums

$$S_g(\alpha,X):=\sum_{n\leqslant X}\lambda_g(n)e(n\alpha) \tag{2.32}$$

associated to a primitive cusp form g. Our goal is to arrive at

Proposition 2.4. *Let g be a primitive Maass cusp form of level D, weight $\kappa\in\{0,1\}$ and Laplacian eigenvalue $\frac14+t_g^2$. Then we have, uniformly for $X\geqslant 1$ and $\alpha\in\mathbf{R}$,*

$$\sum_{n\leqslant X}\lambda_g(n)e(n\alpha)\ll (D\mu_gX)^\varepsilon D\mu_g^2X^{1/2},$$

where the implied constant depends at most on ε.

Remark 2.2. This bound is a classical estimate and due to Wilton in the case of holomorphic forms of full level. However, we have not found it in this generality in the existing literature. One of our goals here is to achieve a polynomial control in the parameters of g (the level or the weight or the eigenvalue). The latter will prove necessary in order to achieve polynomial control in the remaining parameters in the subconvexity problem. Note that the exponents we provide here for D and μ_g are not optimal: with more work, one could replace the factor $D\mu_g^2X^{1/2}$ above by $(D\mu_g^2X)^{1/2}$, and in the D and μ_g aspects it should be possible to go even further by using the amplification method.

First we derive uniform bounds for $g(x+iy)$.

If g is an L^2-normalized primitive Maass cusp form of level D, weight $\kappa \in \{0,1\}$ and spectral parameter $it = it_g$, then we have the Fourier expansion

$$g(x+iy) = \sum_{n\geqslant 1} \rho_g(n)\big\{W_{\frac{\kappa}{2},it}(4\pi ny)e(nx) + \varepsilon_g W_{-\frac{\kappa}{2},it}(4\pi ny)e(-nx)\big\}, \tag{2.33}$$

where $\varepsilon_g = \pm(it)^\kappa$ is the constant in (2.7). The Whittaker functions here can be expressed explicitly from K-Bessel functions:

$$\begin{aligned} W_{0,it}(4y) &= \frac{2y^{1/2}}{\sqrt{\pi}} K_{it}(2y);\\ W_{\frac12,it}(4y) &= \frac{2y}{\sqrt{\pi}}\big\{K_{\frac12+it}(2y) + K_{\frac12-it}(2y)\big\};\\ itW_{-\frac12,it}(4y) &= \frac{2y}{\sqrt{\pi}}\big\{K_{\frac12+it}(2y) - K_{\frac12-it}(2y)\big\}. \end{aligned} \tag{2.34}$$

By the Cauchy–Schwarz inequality, we have

$$y^{2\varepsilon}|g(x+iy)|^2 \ll \sum_{m\geqslant 1}\frac{|\rho_g(m)|^2}{m^{2\varepsilon}} \sum_{n\geqslant 1}(4\pi ny)^{2\varepsilon}\big\{|W_{\frac{\kappa}{2},it}(4\pi ny)|^2 + |\varepsilon_g W_{-\frac{\kappa}{2},it}(4\pi ny)|^2\big\}.$$

Combining this estimate with (2.3), (2.24), (2.28), (2.34) and the uniform bounds of Proposition 7.2, we can conclude that

$$y^\varepsilon g(x+iy) \ll_\varepsilon (D\mu_g)^{2\varepsilon} D^{-1/2}\mu_g y^{-1/2}. \tag{2.35}$$

For small values of y, we improve upon this bound by a variant of the same argument. Namely, we know that every $z = x+iy$ can be represented as βv, where $\beta \in \mathrm{SL}_2(\mathbf{Z})$ and $\Im v \geqslant \frac{\sqrt{3}}{2}$. If $y < \frac{\sqrt{3}}{2}$, as we shall from now on assume, β does not fix the cusp ∞, hence the explicit knowledge of the cusps of $\Gamma_0(D)$ tells us that it factors as $\beta = \gamma\delta$, where $\gamma \in \Gamma_0(D)$ and $\delta = \begin{pmatrix} a & * \\ c & * \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$ with $c \neq 0$ and $c \mid D$. We further factor δ as $\sigma_{\mathfrak{a}}\tau$, where $\sigma_{\mathfrak{a}}$ is a scaling matrix for the cusp $\mathfrak{a} = a/c$ (see Section 2.1 of [I95]) and τ fixes ∞. An explicit choice for $\sigma_{\mathfrak{a}}$ is given by (2.3) of [DI82]:

$$\sigma_{\mathfrak{a}} := \begin{pmatrix} \mathfrak{a}\sqrt{[c^2,D]} & 0 \\ \sqrt{[c^2,D]} & 1/\mathfrak{a}\sqrt{[c^2,D]} \end{pmatrix}.$$

This also implies that

$$\tau = \begin{pmatrix} c/\sqrt{[c^2,D]} & * \\ 0 & \sqrt{[c^2,D]}/c \end{pmatrix},$$

therefore the point $w := \tau v$ has imaginary part

$$\Im w \gg c^2/[c^2,D]. \tag{2.36}$$

Observe that

$$|g(z)| = |g(\delta v)| = |g(\sigma_{\mathfrak{a}} w)| = |h(w)|, \tag{2.37}$$

where $h := g_{|_\kappa \sigma_{\mathfrak{a}}}$ is a cusp form for the congruence subgroup $\sigma_{\mathfrak{a}}^{-1}\Gamma_0(D)\sigma_{\mathfrak{a}}$ of level D, weight κ and spectral parameter $it_h = it_g$. We argue now for h exactly as we did for g, except that in place of (2.3), (2.24), (2.28) we use the uniform bound

$$\sum_{1\leqslant n\leqslant X} n|\rho_h(n)|^2 \ll \mu_h^{1-\kappa}\,\mathrm{ch}(\pi t_h)X.$$

This bound follows exactly as Lemma 19.3[5] in [DFI02] upon noting that $c_{\mathfrak{a}}$ for the cusp $\mathfrak{a} = a/c$ (see Section 2.6 of [I95]) is at least $[c, D/c] \geqslant 1$ (cf. Lemma 2.4 of [DI82]). The analogue of (2.35) that we can derive this way is

$$(\Im w)^{\varepsilon} h(w) \ll_{\varepsilon} \mu_h^{3/2+2\varepsilon} (\Im w)^{-1/2}.$$

By (2.36) and (2.37), this implies that

$$g(x+iy) \ll_{\varepsilon} (D\mu_g)^{\varepsilon} D^{1/2} \mu_g^{3/2}. \tag{2.38}$$

Note that this estimate was derived for $y < \frac{\sqrt{3}}{2}$, but it also holds for all other values of y in the light of (2.35).

With the uniform bounds (2.35) and (2.38) at hand we proceed to estimate the exponential sums $S_g(\alpha, X)$. By applying Fourier inversion to (2.33), we obtain, for any $\alpha \in \mathbf{R}$,

$$\rho_g(n) \left\{ W_{\frac{\kappa}{2}, it}(4\pi ny) + \frac{\Gamma\left(\frac{1}{2} + it + \frac{\kappa}{2}\right)}{\Gamma\left(\frac{1}{2} + it - \frac{\kappa}{2}\right)} W_{-\frac{\kappa}{2}, it}(4\pi ny) \right\} e(n\alpha) = \int_0^1 \bigl\{ g(\alpha+\beta+iy) \pm g(-\alpha-\beta+iy) \bigr\} e(-n\beta)\, d\beta,$$

where the $\pm$ on the right hand side matches the one in (2.7). Then we integrate both sides against $(\pi y)^{\varepsilon} \frac{dy}{y}$ to see that

$$\frac{\lambda_g(n) e(n\alpha)}{n^{1/2+\varepsilon}} = \int_0^1 G_\alpha(\beta) e(-n\beta)\, d\beta, \tag{2.39}$$

where

$$G_\alpha(\beta) := \frac{\pi^{1/2+\varepsilon}}{4\rho_g(1)\Phi_\kappa^1\left(\frac{1}{2}+\varepsilon, it_g\right)} \int_0^\infty \bigl\{ g(\alpha+\beta+iy) \pm g(-\alpha-\beta+iy) \bigr\} y^{\varepsilon} \frac{dy}{y}. \tag{2.40}$$

The function $\Phi_\kappa^1(s, it)$ is defined in (2.11), and is determined explicitly by Lemma 8.2 of [DFI02]. For $\kappa \in \{0, 1\}$, this result can be seen more directly from the explicit formulae (2.34). At any rate,

$$\Phi_\kappa^1\left(\tfrac{1}{2}+\varepsilon, it_g\right) \asymp \widehat{K_{\frac{\kappa}{2}+it_g}}\left(\tfrac{1+\kappa}{2}+\varepsilon\right) \asymp \mu_g^{(\kappa-1)/2+\varepsilon} \operatorname{ch}^{-1/2}(\pi t_g),$$

so that by (2.24) we also have

$$\rho_g(1)\Phi_\kappa^1\left(\tfrac{1}{2}+\varepsilon, it_g\right) \gg_{\varepsilon} (D\mu_g)^{-1/2-\varepsilon}.$$

The integral in (2.40) is convergent by (2.35) and (2.38). Moreover,

$$\int_0^\infty \bigl\{ g(\alpha+\beta+iy) \pm g(-\alpha-\beta+iy) \bigr\} y^{\varepsilon} \frac{dy}{y} \ll_{\varepsilon} (D\mu_g)^{2\varepsilon} D^{1/2} \mu_g^{3/2}.$$

Altogether we have obtained the uniform bound

$$G_\alpha(\beta) \ll_{\varepsilon} (D\mu_g)^{\varepsilon} D\mu_g^2, \quad \alpha \in \mathbf{R}. \tag{2.41}$$

For $X \geqslant 1$, we introduce the modified Dirichlet kernel

$$D(\beta, X) := \sum_{1 \leqslant n \leqslant X} e(-n\beta).$$

It follows from (2.39) that

$$\sum_{n \leqslant X} \frac{\lambda_g(n) e(n\alpha)}{n^{1/2+\varepsilon}} = \int_0^1 G_\alpha(\beta) D(\beta, X)\, d\beta.$$

[5] In this lemma, $|s_j|$ should really be $|s_j|^{1-k}$. In fact, this is the dependence that follows from Lemma 19.2 of [DFI02]. We also note that the proof of the latter lemma is not entirely correct. Namely, (19.12) in [DFI02] does not follow from the bound preceding it. Nevertheless, it does follow from the exponential decay of the Whittaker functions (cf. our (2.34) and Proposition 7.2).

Combining (2.41) with the fact that the L^1-norm of $D(\beta, X)$ is $\ll \log(2X)$, we can conclude that

$$\sum_{n\leqslant X} \frac{\lambda_g(n)e(n\alpha)}{n^{1/2+\varepsilon}} \ll_\varepsilon (D\mu_g X)^\varepsilon D\mu_g^2.$$

Finally, by partial summation we arrive to Proposition 2.4.

For completeness, we display the analogous result for holomorphic forms that can be proved along the same lines.

Proposition 2.5. *Let g be a primitive holomorphic cusp form of level D and weight k. Then we have, uniformly for $X \geqslant 1$ and $\alpha \in \mathbf{R}$,*

$$\sum_{n\leqslant X} \lambda_g(n)e(n\alpha) \ll (DkX)^\varepsilon Dk^{3/2}X^{1/2},$$

where the implied constant depends at most on ε.

These estimates are useful to derive bounds for shifted convolution sums on average which will be used later on: the following lemma is similar to Lemma 3 of [J96] (see also Lemma 3.2 of [Bl04]).

Lemma 2.3. *Let g be a primitive (either Maass or holomorphic) cusp form of level D. For any $X, Y \geqslant 1$, for any nonzero integers ℓ_1, ℓ_2, for any sequence of complex numbers $\mathbf{a} = (a_h)_{h\in\mathbf{Z}}$, and for any $\varepsilon > 0$, one has*

$$\sum_{h\in\mathbf{Z}} a_h \sum_{\substack{m\leqslant X,\ n\leqslant Y\\ \ell_1 m-\ell_2 n=h}} \overline{\lambda_g}(m)\lambda_g(n) \ll_\varepsilon (XYD\mu_g)^\varepsilon D^2\mu_g^4(XY)^{1/2}\|\mathbf{a}\|_2.$$

Remark 2.3. Of course this lemma will be applied to sequences $\mathbf{a}$ supported in $[-H, H]$ with $H = |\ell_1|X + |\ell_2|Y$.

Proof. The estimate follows by combining Proposition 2.4 or 2.5 with the Cauchy—Schwarz inequality and the Parseval identity:

$$\begin{aligned}\sum_{h\in\mathbf{Z}} a_h \sum_{\substack{m\leqslant X,\ n\leqslant Y\\ \ell_1 m-\ell_2 n=h}} \overline{\lambda_g}(m)\lambda_g(n) &= \int_0^1 \sum_{h\in\mathbf{Z}} a_h e(\alpha h)\overline{S_g(\ell_1\alpha, X)}S_g(\ell_2\alpha, Y)\,d\alpha\\ &\ll_\varepsilon (XY\mu_g D)^\varepsilon D^2\mu_g^4(XY)^{1/2}\int_0^1\left|\sum_{h\in\mathbf{Z}} a_h e(\alpha h)\right|d\alpha\\ &\ll_\varepsilon (XY\mu_g D)^\varepsilon D^2\mu_g^4(XY)^{1/2}\|\mathbf{a}\|_2.\end{aligned}$$

□

3. Rankin–Selberg L-functions

Given two primitive cusp forms f and g, we denote by q and D, χ_f and χ_g, $\pi_f = \bigotimes' \pi_{f,p}$ and $\pi_g = \bigotimes' \pi_{g,p}$, respectively, the level, nebentypus, and $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$-automorphic representation attached to f and g; the Rankin–Selberg L-function of f and g is the Euler product of degree 4 given by

$$L(f\otimes g, s) = L(\pi_f\otimes\pi_g, s) = \prod_{p<\infty} L(\pi_{f,p}\otimes\pi_{g,p}, s) = \prod_{p<\infty}\prod_{i=1}^4\left(1 - \frac{\alpha_{\pi_f\otimes\pi_g,i}(p)}{p^s}\right)^{-1}$$

(say); it is the associated (finite) Rankin–Selberg L-function as defined in Jacquet's monograph [J72] (see also [C04]).

Remark 3.1. Although we will not use this fact, it was proved by Ramakrishnan [R00] that $L(\pi_f\otimes\pi_g, s)$ is automorphic: there exists a $\mathrm{GL}_4(\mathbf{A}_\mathbf{Q})$-automorphic isobaric representation $\pi_f\boxtimes\pi_g$ whose L-function coincides with $L(\pi_f\otimes\pi_g, s)$.

For $p \nmid (q, D)$, the local parameters $\{\alpha_{\pi_f\otimes\pi_g,i}(p)\}_{i=1,\dots,4}$ have a simple expression in terms of the local numerical parameters of π_f and π_g at p. Namely,

$$L(\pi_{f,p}\otimes\pi_{g,p},s)=\prod_{i=1}^{2}\prod_{j=1}^{2}\left(1-\frac{\alpha_{\pi_f,i}(p)\alpha_{\pi_g,j}(p)}{p^s}\right)^{-1},$$

where

$$L(\pi_{f,p},s)=\prod_{i=1}^{2}\left(1-\frac{\alpha_{\pi_f,i}(p)}{p^s}\right)^{-1}=\sum_{n\mid p^\infty}\frac{\lambda_f(n)}{n^s},$$
$$L(\pi_{g,p},s)=\prod_{j=1}^{2}\left(1-\frac{\alpha_{\pi_g,j}(p)}{p^s}\right)^{-1}=\sum_{n\mid p^\infty}\frac{\lambda_g(n)}{n^s}. \tag{3.1}$$

In general, inspecting the possible cases one can verify that for $i=1,\dots,4$,

$$\left|\alpha_{\pi_f\otimes\pi_g,i}(p)\right|\leqslant p^{2\theta}.$$

The local factors at the finite places are completed by a local factor at the infinite place of the form

$$L_\infty(f\otimes g,s)=L(\pi_{f,\infty}\otimes\pi_{g,\infty},s)=\prod_{i=1}^{4}\Gamma_{\mathbf{R}}\big(s+\mu_{\pi_f\otimes\pi_g,i}(\infty)\big),\quad \Gamma_{\mathbf{R}}(s):=\pi^{-s/2}\Gamma(s/2).$$

Again, one can verify that for $i=1,\dots,4$,

$$\Re\mu_{\pi_f\otimes\pi_g,i}(\infty)\geqslant-2\theta. \tag{3.2}$$

The L-function $L(f\otimes g,s)$ satisfies a functional equation of the form

$$\Lambda(f\otimes g,s)=\varepsilon(f\otimes g)\overline{\Lambda(f\otimes g,1-\overline{s})}$$

with $|\varepsilon(f\otimes g)|=1$ and

$$\Lambda(f\otimes g,s)=Q(f\otimes g)^{s/2}L_\infty(f\otimes g,s)L(f\otimes g,s),$$

where $Q(f\otimes g)=Q(\pi_f\otimes\pi_g)$ is the conductor of the Rankin–Selberg L-function. By the local Langlands correspondence, one can verify that $Q(f\otimes g)$ satisfies

$$(qD)^2/(q,D)^4\leqslant Q(f\otimes g)\leqslant(qD)^2/(q,D).$$

3.1. **Approximate functional equation.** For s on the critical line $\Re s=\frac{1}{2}$, we set

$$P:=\prod_{i=1}^{4}\left|s+\mu_{\pi_f\otimes\pi_g,i}(\infty)\right|^{1/2};$$

The local parameters $\mu_{\pi_f\otimes\pi_g,i}(\infty)$ can be computed in terms of the local parameters at ∞ of π_f and π_g; in particular, one can check that (cf. (2.2))

$$P\leqslant(|s|+\mu_f+\mu_g)^2.$$

Note that (3.2) and $\theta<\frac{1}{4}$ imply that $P>0$. By standard techniques (see [M04a] for instance), one can show that for s with $\Re s=\frac{1}{2}$ and for any $A\geqslant1$, one has a bound of the form

$$L(f\otimes g,s)\ll_A\log^2(qDP+1)\sum_N\frac{\left|L_{f\otimes g}(N)\right|}{\sqrt{N}}\left(1+\frac{N}{PQ(f\otimes g)^{1/2}}\right)^{-A}, \tag{3.3}$$

where N ranges over the reals of the form $N=2^\nu$, $\nu\geqslant-1$, and $L_{f\otimes g}(N)$ are sums of type

$$L_{f\otimes g}(N)=\sum_n\lambda_f(n)\lambda_g(n)W(n)$$

for some smooth function $W(x)=W_{N,A}(x)$ supported on $[N/2,5N/2]$ such that

$$x^jW^{(j)}(x)\ll_{j,A}P^j \tag{3.4}$$

for all $j \geqslant 0$. In particular, Theorem 1 follows from

Proposition 3.1. *Assume Hypothesis H_θ and that $\chi_f\chi_g$ is nontrivial. For any $0 < \varepsilon \leqslant 10^{-3}$ and any $N \geqslant 1$ satisfying*

$$N \leqslant (qDP)^{1+\varepsilon}, \tag{3.5}$$

one has

$$\frac{L_{f\otimes g}(N)}{\sqrt{N}} \ll_\varepsilon q^{100\varepsilon} q^{\frac{1}{2}-\frac{(1-2\theta)\delta_{\mathrm{tw}}}{202}}.$$

The implied constant depends on ε and polynomially on μ_f, μ_g, D and P.

Indeed, for any $0 < \varepsilon \leqslant 10^{-3}$ by a trivial estimate and by taking A sufficiently large, we see that the contribution to (3.3) of the N's such that $N \geqslant (qDP)^{1+\varepsilon/200}$ is bounded by

$$\ll_\varepsilon (qDP)^\varepsilon.$$

For the remaining terms, we apply Proposition 3.1, getting

$$\begin{aligned} L(f\otimes g, s) &\ll_{\varepsilon,|s|,\mu_f,\mu_g,D} q^{\frac{\varepsilon}{2}} \log^3(qDP+1) q^{\frac{1}{2}-\frac{(1-2\theta)\delta_{\mathrm{tw}}}{202}} \\ &\ll_{\varepsilon,|s|,\mu_f,\mu_g,D} q^{\varepsilon+\frac{1}{2}-\frac{(1-2\theta)\delta_{\mathrm{tw}}}{202}}. \end{aligned}$$

3.2. **Amplification.** As usual, the bound for $L_{f\otimes g}(N)$ in Proposition 3.1 follows from an application of the amplification method. For this one has to embed f into an appropriate family. In preparation of this, we change the notation slightly and write χ for the nebentypus of f and rename our original primitive form f to f_0. We note that when f_0 is a holomorphic form of weight $k \geqslant 1$, then $F_0(z) := y^{k/2} f_0(z)$ is a Maass form of weight k and of course $L_{f_0\otimes g}(N) = L_{F_0\otimes g}(N)$, so we may treat f_0 as a Maass form of some weight $k \geqslant 0$. As an appropriate family we choose an orthonormal basis $\mathcal{B}_k([q,D],\chi) = \{u_j\}_{j\geqslant 1}$ of Maass cusp forms of level $[q,D]$ and nebentypus χ containing (the old form) $f_0/\langle f_0, f_0\rangle_{[q,D]}^{1/2}$ (note the enlargement of the level from q to the l.c.m. of q and D) .

Remark 3.2. As was emphasized in [DFI02], the replacement of the holomorphic form f_0 by its associated weight k Maass form is not a cosmetic artefact but turns out to be crucial when k is small. Indeed, for small k, the c-summation in the Petersson formula (2.18) does not converge quickly enough (and Petersson's formula does not even exist when $k = 1$!): the reason is that when k is small, the holomorphic forms of weight k are too close to the continuous spectrum. On the other hand, when k is large ($k \geqslant 10^6$ say), we could have chosen for family an orthonormal basis of the space of holomorphic cusp forms of level $[q,D]$ and nebentypus χ containing (the old form) $f_0/\langle f_0, f_0\rangle_{[q,D]}^{1/2}$, see Remark 3.3 below.

For $L \geqslant 1$ (a small positive power of q), let $\vec{x} = (x_1, \ldots, x_\ell, \ldots, x_L)$ be any complex vector whose entries x_ℓ satisfy

$$(\ell, qD) \neq 1 \quad \Longrightarrow \quad x_\ell = 0. \tag{3.6}$$

For $f(z) \in \mathcal{L}_k([q,D],\chi)$ either a Maass cusp form or an Eisenstein series $E_{\mathfrak{a}}(z,s)$, we consider the following linear form:

$$L_{f\otimes g}(\vec{x}, N) := \sum_\ell x_\ell \sum_{de=\ell} \chi(d) \sum_{ab=d} \mu(a)\chi_g(a)\lambda_g(b) \sum_n W(adn)\lambda_g(n)\sqrt{aen}\rho_f(aen).$$

As explained in Section 4 of [M04a], it follows from (2.3) and (2.4) that for our original primitive form $f = f_0$, $L_{f\otimes g}(\vec{x}, N)$ factors as

$$L_{f_0\otimes g}(\vec{x}, N) = \rho_{f_0}(1) \left(\sum_{\ell\leqslant L} x_\ell \lambda_{f_0}(\ell)\right) L_{f_0\otimes g}(N). \tag{3.7}$$

Thus we form the "spectrally complete" quadratic form

$$Q(\vec{x},N) := \sum_j \mathcal{H}(t_j)\big|L_{u_j\otimes g}(\vec{x},N)\big|^2 + \sum_{\mathfrak{a}} \frac{1}{4\pi}\int_{\mathbf{R}} \mathcal{H}(t)\big|L_{\mathfrak{a},t,g}(\vec{x},N)\big|^2 dt,$$

where $\mathcal{H}(t)$ is as in Proposition 2.3, and the parameter A used to define $\mathcal{H}(t)$ will be chosen sufficiently large. Our goal is the following estimate for the complete quadratic form.

Proposition 3.2. *Assume Hypothesis H_θ for any $0 \leqslant \theta \leqslant \frac{1}{2}$. With the above notation, suppose that $\chi\chi_g$ is nontrivial and let $q^* > 1$ denote its conductor; moreover, suppose that g satisfies*

$$w \mid D \quad\Longrightarrow\quad q^* \nmid (w, D/w), \tag{3.8}$$

then for any $1 \leqslant L \leqslant q$, any $0 < \varepsilon \leqslant 10^{-3}$ and any N satisfying (3.5), there is $A = A(\varepsilon)$ as in Proposition 2.3 such that

$$Q(\vec{x},N) \ll_\varepsilon q^{100\varepsilon} N \left\{\|\vec{x}\|_2^2 + \|\vec{x}\|_1^2 L^{\delta_L} q^{-\delta_q}\right\}$$

with

$$\delta_L := \frac{46 - 9\theta - 22\theta^2}{9}, \qquad \delta_q := \frac{1-2\theta}{9}\,\delta_{\mathrm{tw}}.$$

Here

$$\|\vec{x}\|_1 := \sum_{\ell\leqslant L} |x_\ell|, \qquad \|\vec{x}\|_2^2 := \sum_{\ell\leqslant L} |x_\ell|^2,$$

and $\delta_{\mathrm{tw}} := \frac{1-2\theta}{8}$ is the convexity breaking exponent of Theorem 5. If g does not satisfy (3.8), then

$$Q(\vec{x},N) \ll_\varepsilon q^{100\varepsilon} N \left\{\|\vec{x}\|_2^2 + \|\vec{x}\|_1^2 \left(L^{\delta_L} q^{-\delta_q} + L^{\delta_{3L}} q^{-\delta_{3q}} + L^{\delta_{4L}} q^{-\delta_{4q}}\right)\right\}$$

with

$$\delta_{3L} := 9 + 4\theta, \qquad \delta_{3q} := \frac{1}{2} - \theta,$$

$$\delta_{4L} := \frac{7 + 10\theta + 4\theta^2}{2(1+\theta)}, \qquad \delta_{4q} := \frac{1+4\theta}{4(1+\theta)}.$$

In these inequalities the implied constant depends on ε and polynomially on μ_g, D and P.

Proof of Proposition 3.1. As explained in Section 4 of [M04a], Proposition 3.1 now follows from Proposition 3.2. Indeed, by (3.7) and by positivity, in particular by (2.19), one has

$$\frac{\mathcal{H}(t_{f_0})|\rho_{f_0}(1)|^2}{\langle f_0, f_0\rangle_q[\Gamma_0(q):\Gamma_0([q,D])]} \left|\sum_{\ell\leqslant L} x_\ell \lambda_{f_0}(\ell)\right|^2 \big|L_{f_0\otimes g}(N)\big|^2 \leqslant Q(\vec{x},N).$$

Moreover, for a Maass cusp form f_0 of weight $k \in \{0,1\}$, we have, by (3.7), (2.19) and (2.24),

$$\frac{\mathcal{H}(t_{f_0})|\rho_{f_0}(1)|^2}{\langle f_0, f_0\rangle_q[\Gamma_0(q):\Gamma_0([q,D])]} \gg \frac{(qD + |t_{f_0}|)^{-\varepsilon}}{[q,D](1+|t_{f_0}|)^{16}},$$

where the implied constant depends at most on ε. When f_0 comes from a holomorphic form of weight k (i.e., $t_{f_0} = \pm i\frac{k-1}{2}$), we have, by (2.19) and (2.25),

$$\frac{\mathcal{H}(t_{f_0})|\rho_{f_0}(1)|^2}{\langle f_0, f_0\rangle_q[\Gamma_0(q):\Gamma_0([q,D])]} \gg \frac{(qD + |t_{f_0}|)^{-\varepsilon} e^{-Ck}}{[q,D](1+|t_{f_0}|)^{16}},$$

for some absolute positive constant $C > 0$, the implied constant depending at most on ε. We suppose first that g satisfies (3.8); by Proposition 3.2, we have

$$\left|\sum_{\ell\leqslant L} x_\ell \lambda_{f_0}(\ell)\right|^2 \big|L_{f_0\otimes g}(N)\big|^2 \ll_{\mu_g,D,P,\varepsilon} D^\varepsilon q^{101\varepsilon}(1+|t_{f_0}|)^{16} e^{Ck}[q,D] N \left\{\|\vec{x}\|_2^2 + \|\vec{x}\|_1^2 L^{\delta_L} q^{-\delta_q}\right\},$$

where the implied constant depends at most polynomially on μ_g, D and P. The result follows by choosing the (standard) amplifier $(x_1, \dots, x_L)$ given by

$$x_\ell := \begin{cases} \lambda_{f_0}(p)\overline{\chi}(p) & \text{if} \quad \ell = p, \quad (p, qD) = 1, \quad \sqrt{L}/2 < p \leqslant \sqrt{L}; \\ -\overline{\chi}(p) & \text{if} \quad \ell = p^2, \quad (p, qD) = 1, \quad \sqrt{L}/2 < p \leqslant \sqrt{L}; \\ 0 & \text{else.} \end{cases}$$

Indeed, from the relation $\lambda_{f_0}(p)^2 - \lambda_{f_0}(p^2) = \chi(p)$, we have

$$\left| \sum_{\ell \leqslant L} x_\ell \lambda_{f_0}(\ell) \right| \gg \frac{\sqrt{L}}{\log L}$$

for $L \geqslant 5(\log qD)^2$, and from (2.28) we have

$$\|\vec{x}\|_1 + \|\vec{x}\|_2^2 \ll \big(qD(1 + |t_{f_0}|)L\big)^\varepsilon L^{1/2},$$

where the implied constants depend at most on ε. Hence we have

$$L_{f_0 \otimes g}(N) \ll_{\mu_g, D, P, \varepsilon} (1 + |t_{f_0}|)^8 e^{Ck/2} q^{52\varepsilon} (qN)^{1/2} \left(L^{-1/4} + L^{\delta_L/2} q^{-\delta_q/2} \right).$$

By choosing

$$L = L_0 := q^{\frac{2\delta_q}{1+2\delta_L}} \geqslant q^{\frac{18}{101}\delta_q},$$

we conclude the proof of Proposition 3.1 when g satisfies (3.8). When g does not satisfy (3.8), one can check that for $0 \leqslant \theta \leqslant \frac{1}{2}$ and for $\delta_{\text{tw}} = \frac{1-2\theta}{8}$ one has

$$L_0^{\delta_{3L}/2} q^{-\delta_{3q}/2} + L_0^{\delta_{4L}/2} q^{-\delta_{4q}/2} \leqslant L_0^{-1/4},$$

so that Proposition 3.1 holds in that case, too. □

Remark 3.3. The above estimates prove Proposition 3.1 with a polynomial dependency in μ_{f_0}, μ_g, D, P except possibly when f_0 is a holomorphic form of weight k in which case the dependency in $\mu_{f_0} = 1 + \frac{k-1}{2}$ is only proven to be at most exponential. This comes from the fact that $\Gamma(k)/\mathcal{H}\left(i\frac{k-1}{2}\right)$ is bounded exponentially in k rather than polynomially. We could probably remedy this by making a different choice for the weight function $\mathcal{H}(t)$; another—more natural—way is to consider, instead of $Q(\vec{x}, N)$, the quadratic form

$$Q^h(\vec{x}, N) := \sum_{f \in \mathcal{B}_k^h([q,D],\chi)} \Gamma(k) |L_{f \otimes g}(\vec{x}, N)|^2,$$

where $\mathcal{B}_k^h\big([q, D], \chi\big)$ is an orthonormal basis of the space of holomorphic cusp forms of level $[q, D]$ and nebentypus χ containing (the old form) $f_0/\langle f_0, f_0\rangle_{[q,D]}^{1/2}$. If k is large enough ($k \geqslant 10^6$ say), $Q^h(\vec{x}, N)$ can be analyzed (by means of the holomorphic Petersson formula Proposition 2.2) exactly as in the next section, and Proposition 3.2 can be shown to hold for $Q^h(\vec{x}, N)$ with the same (polynomial) dependencies in μ_g, P and D only. Then the argument above (using (2.25)) yields Proposition 3.1 with a polynomial dependency in k_{f_0} as well.

3.3. **Analysis of the quadratic form.** We compute the quadratic form $Q(\vec{x}, N)$ by applying the spectral summation formula of Proposition 2.3. $Q(\vec{x}, N)$ decomposes into a diagonal part and a non-diagonal one:

$$\begin{aligned} Q(\vec{x}, N) = \sum_{\ell_1, \ell_2} \overline{x_{\ell_1}} x_{\ell_2} &\sum_{\substack{a_1 b_1 e_1 = \ell_1 \\ a_2 b_2 e_2 = \ell_2}} \mu(a_1)\mu(a_2)\chi\chi_g(\overline{a_1} a_2)\chi(\overline{b_1} b_2)\overline{\lambda_g}(b_1)\lambda_g(b_2) \\ &\times \left\{ S^{\mathrm{D}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N \right) + \sum_{c \equiv 0\,([q,D])} \frac{1}{c^2} S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N; c \right) \right\} \\ &= c_A Q^{\mathrm{D}}(\vec{x}, N) + Q^{\mathrm{ND}}(\vec{x}, N), \end{aligned} \tag{3.9}$$

say, with

$$S^{\mathrm{D}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N\right) := \sum_{a_1e_1m=a_2e_2n} \overline{\lambda_g}(m)\lambda_g(n)\overline{W}(a_1d_1m)W(a_2d_2n),$$

and

$$S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N; c\right) := c\sum_{m,n} \overline{\lambda_g}(m)\lambda_g(n)S_\chi(a_1e_1m, a_2e_2n; c)\mathcal{I}\left(\frac{4\pi\sqrt{a_1a_2e_1e_2mn}}{c}\right)\overline{W}(a_1d_1m)W(a_2d_2n). \tag{3.10}$$

Here we have put $d_1 := a_1b_1$ and $d_2 := a_2b_2$. The diagonal term is easy to bound and the arguments of [M04a, Section 4.1.1] yield

$$Q^{\mathrm{D}}(\vec{x}, N) \ll_\varepsilon (qNP)^\varepsilon N \sum_{d,\ell_1,\ell_2} |x_{d\ell_1}||x_{d\ell_2}| \frac{\sigma_g(\ell_1)\sigma_g(\ell_2)}{\sqrt{\ell_1\ell_2}} \ll_{g,\varepsilon} (qNP)^{2\varepsilon} N\|\vec{x}\|_2^2 \tag{3.11}$$

for any $\varepsilon > 0$.

3.3.1. *The non-diagonal term.* We transform (3.10) further by applying the Voronoi summation formula of Proposition 2.1 to the n variable. We set $e := (a_2e_2, c)$, $c^* := c/e$, $e^* := a_2e_2/e$, so that $(c^*, e^*) = 1$, and by (3.6) we have $(e, qD) = 1$, hence $D|[q, D]|c^*$. Again, the arguments of [M04a, Section 4.1.2] yield, for a cusp form g,

$$S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N; c\right) = e\overline{\chi_g}(e^*)\sum_{\pm} \varepsilon_g^{\pm} \sum_{m,n\geqslant 1} \overline{\lambda_g}(m)\lambda_g(n)G_{\chi\chi_g}(a_1e_1m \mp e\overline{e^*}n; c)\mathcal{J}^{\pm}(m, n),$$

where $\varepsilon_g^+ = 1$ and $\varepsilon_g^- = \pm 1$ is the sign in (2.7) (for g not induced from a holomorphic form) and

$$\mathcal{J}^{\pm}(x, y) := \overline{W}(a_1d_1x)\int_0^\infty W(a_2d_2u)\mathcal{I}\left(\frac{4\pi\sqrt{a_1a_2e_1e_2xu}}{c}\right) J_g^{\pm}\left(\frac{4\pi e\sqrt{yu}}{c}\right) du.$$

We consider the following (unique) factorization of c:

$$c = c^\sharp c^\flat, \quad \text{where} \quad c^\flat := \prod_{\substack{p|c \\ v_p(c) < v_p(a_2e_2)}} p^{v_p(c)}.$$

Then

$$(c^\sharp, c^\flat) = 1, \qquad c^\flat | e, \qquad (c^\sharp, e^*) = 1,$$

and a calculation in [M04a, Section 4.1.2] yields

$$S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N; c\right) = \chi(e^*)\chi\chi_g(c^\flat)e\sum_{f|c^\flat} f\mu\left(\frac{c^\flat}{f}\right) \sum_{f'|f^*} \mu(f')\overline{\chi_g(f')\lambda_g(f^*/f')} \sum_{\pm} \varepsilon_g^{\pm}\Sigma^{\pm}(a_1e_1e^*f'f^*, e), \tag{3.12}$$

where $f^* := f/(a_1e_1, f)$ and

$$\Sigma^{\pm}(a_1e_1e^*f'f^*, e) := \sum_h G_{\chi\chi_g}(h; c^\sharp)S_h^{\pm}(a_1e_1e^*f'f^*, e) \tag{3.13}$$

with

$$S_h^{\pm}(a_1e_1e^*f'f^*, e) := \sum_{a_1e_1e^*f'f^*m \mp en = h} \overline{\lambda_g}(m)\lambda_g(n)\mathcal{J}^{\pm}(f'f^*m, n).$$

3.3.2. *Bounding* $\Sigma^{\pm}(a_1e_1e^*f'f^*, e)$. Since $\chi\chi_g$ is not the trivial character, $G_{\chi\chi_g}(0;c)S_0 = 0$, and we are left to evaluate (3.13) over the frequencies $h \neq 0$. This will be done in Theorem 4.

First we analyze the properties of $\mathcal{J}^{\pm}(x, y)$; to simplify the notation we set

$$a := a_1d_1, \qquad b := a_2d_2, \qquad d := a_1a_2e_1e_2.$$

Lemma 3.1. *Let*

$$\Theta := \left(\frac{d}{ab}\right)^{1/2}\frac{N}{c}, \qquad Z := P + \Theta, \qquad W_0 := \frac{bc^2}{e^2N}, \qquad X_0 := \frac{N}{a},$$

$$Y_0 := P^2W_0(1+\Theta^2) = P^2\left(\frac{bc^2}{e^2N} + \frac{dN}{ae^2}\right) = P^2\frac{d}{e^2}\left(\frac{1+\Theta^2}{\Theta^2}\right)X_0.$$

For any $i, j, k \geqslant 0$*, any* $\varepsilon > 0$ *we have*

$$x^iy^i\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j}\mathcal{J}^{\pm}(x,y) \ll Z^{i+j}\frac{N}{b}(1+\Theta)\left(\frac{\Theta}{1+\Theta}\right)^{A+1}\left(1+\frac{y}{Y_0}\right)^{-k}\left(\frac{Y_0}{y}\right)^{\theta_g+\varepsilon},$$

where

$$\theta_g := \begin{cases} |\Im t_g| & \text{if } g \text{ is a weight } 0 \text{ Maass cusp form;} \\ 0 & \text{otherwise.}\end{cases}$$

Here the implied constant depends on ε, i, j, A *and polynomially on* μ_g; A *is the constant which appears in* (2.20)*. Note that* $\Im t_g = 0$ *when* g *is a Maass form of weight* 1*. Recall also that as a function of* x, $\mathcal{J}(x, y)$ *is supported on* $[X_0/2, 5X_0/2]$.

Proof. We have

$$\mathcal{J}^{\pm}(x,y) = \overline{W}(ax)\int_0^{\infty}W(bu)\mathcal{I}\left(W_1\right)J_g^{\pm}\left(W_2\right)du$$

with

$$W_1 := \frac{4\pi\sqrt{dxu}}{c} \sim \Theta, \qquad W_2 := \frac{4\pi\sqrt{e^2yu}}{c} \sim \left(\frac{y}{W_0}\right)^{1/2} \geqslant \left(\frac{y}{Y_0}\right)^{1/2}.$$

The latter integral can be written as a linear combination (with constant coefficients) of integrals of the form

$$\overline{W}(ax)\int_0^{\infty}\left\{W(bu)\mathcal{I}\left(W_1\right)W_2^{-\nu}\right\}W_2^{\nu}J_{\nu}\left(W_2\right)du,$$

where

$$J_{\nu}(x) \in \left\{\frac{Y_{\nu}(x)}{\operatorname{ch}(\pi t)},\ \operatorname{ch}(\pi t)K_{\nu}(x)\right\}$$

with $\nu \in \{\pm 2it_g\}$ if g is a Maass form of weight 0; or

$$J_{\nu}(x) \in \left\{\frac{Y_{\nu}(x)}{\operatorname{sh}(\pi t)},\ \operatorname{sh}(\pi t)K_{\nu}(x)\right\}$$

with $\nu \in \{\pm 2it_g\}$ if g is a Maass form of weight 1; or

$$J_{\nu}(x) = J_{k_g-1}(x),$$

if g is a holomorphic form of weight k_g. Using (7.1) we integrate by parts $2k$ times (where we may assume that $k = 0$ for $y \leqslant Y_0$). We obtain, using also Propositions 7.1 and 7.2, (2.20), (3.4) and that $u \sim N/b$,

$$\mathcal{J}^{\pm}(x,y) \ll_{A,\varepsilon} \frac{N}{b}(1+\Theta)\left(\frac{\Theta}{1+\Theta}\right)^{A+1}\left(1+\frac{y}{Y_0}\right)^{-k-1/4} \times \begin{cases}\left(\frac{W_2}{1+W_2}\right)^{-2|\Im t_g|-\varepsilon} & \text{if } g \text{ is a Maass form;}\\ \left(\frac{W_2}{1+W_2}\right)^{k_g-1} \leqslant 1 & \text{if } g \text{ is holomorphic.}\end{cases} \tag{3.14}$$

For the higher derivatives, the proof is similar after several derivations with respect to the variables x, y. $\square$

We proceed now by bounding $\Sigma^{\pm}(a_1e_1e^*f'f^*, e)$. We set

$$\mathbf{l_1} := a_1e_1e^*f'f^* = \frac{d}{e}f'f^*, \qquad \mathbf{l_2} := e;$$

$$\mathbf{X} := \frac{\mathbf{l_1}}{f'f^*}X_0 = \frac{d}{e}X_0, \qquad \mathbf{Y} := \mathbf{l_2}Y_0 = P^2\left(\frac{1+\Theta^2}{\Theta^2}\right)\mathbf{X};$$

$$\mathbf{q} := \mathrm{Cond}(\chi\chi_g) = q^*, \qquad \mathbf{c} := c^\sharp, \qquad \mathbf{F}(x,y) := \mathcal{J}^{\pm}(f'f^*x/\mathbf{l_1}, y/\mathbf{l_2}).$$

By a smooth dyadic partition of unity, we have the decomposition

$$\mathbf{F}(x,y) = \frac{N}{b}(1+\Theta)\left(\frac{\Theta}{1+\Theta}\right)^{A+1}\sum_{Y\geqslant 1}F_Y(x,y),$$

where Y is of the form 2^ν, $\nu \in \mathbf{N}$, $F_Y(x,y)$ is supported on $[\mathbf{X}/4, 4\mathbf{X}]\times[Y/4,4Y]$ and satisfies

$$x^iy^i\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j}F_Y(x,y) \ll_{i,j,k,\varepsilon} Z^{i+j}\left(1+\frac{Y}{\mathbf{Y}}\right)^{-k}\left(\frac{\mathbf{Y}}{Y}\right)^{\theta_g+\varepsilon} \tag{3.15}$$

for any $i,j,k\geqslant 0$ and any $\varepsilon>0$. The sum $\Sigma^{\pm}(\mathbf{l_1},\mathbf{l_2})$ decomposes accordingly as

$$\Sigma^{\pm}(\mathbf{l_1},\mathbf{l_2}) := \frac{N}{b}(1+\Theta)\left(\frac{\Theta}{1+\Theta}\right)^{A+1}\sum_{Y\geqslant 1}\sum_h G_{\chi\chi_g}(h;\mathbf{c})S_{h,Y}(\mathbf{l_1},\mathbf{l_2})$$

with

$$S_{h,Y}(\mathbf{l_1},\mathbf{l_2}) := \sum_{\mathbf{l_1}m\pm\mathbf{l_2}n=h}\overline{\lambda_g}(m)\lambda_g(n)F_Y(\mathbf{l_1}m,\mathbf{l_2}n).$$

We want to apply Theorem 4 (to be proved in the forthcoming section) to the h-sums above. Given $\varepsilon>0$ very small, we see by trivial estimation and by choosing A large enough (we will take $A = 1000/\varepsilon+100$), that the total contribution of the $S^{\mathrm{ND}}(\dots,N;c)$ such that $\Theta<q^{-\varepsilon}$ is negligible; hence in the remaining case we have the easy inequalities

$$\Theta^{-1}\leqslant q^\varepsilon, \qquad \Theta\leqslant LN/c, \qquad 1+\Theta\leqslant 2q^\varepsilon LN/c, \qquad \mathbf{l_1}\mathbf{l_2}\leqslant (Lc^\flat)^2,$$

$$\mathbf{X}\leqslant dN/e\leqslant L^2N/e, \qquad \mathbf{Y}/\mathbf{X}\leqslant q^{2\varepsilon}P^2, \qquad \mathbf{Y}\leqslant q^{2\varepsilon}P^2L^2N/e.$$

We will also use the trivial bound $\Theta/(1+\Theta)\leqslant 1$. We introduce a parameter $Y_{\min}$ to be determined later, and denote by $\Sigma^{\pm}_{Y\leqslant Y_{\min}}(\mathbf{l_1},\mathbf{l_2})$ (resp. $\Sigma^{\pm}_{Y>Y_{\min}}(\mathbf{l_1},\mathbf{l_2})$) the contribution to $\Sigma^{\pm}(\mathbf{l_1},\mathbf{l_2})$ of $Y\leqslant Y_{\min}$ (resp. $Y>Y_{\min}$). For $Y\leqslant Y_{\min}$, we apply the "trivial" bound (4.4) to the sums $\sum_h G_{\chi\chi_g}(h;\mathbf{c})S_{h,Y}(\mathbf{l_1},\mathbf{l_2})$, and find that (since $\mathbf{l_1}\mathbf{l_2}=df'f^*$ and $\theta_g\leqslant\theta$)

$$\begin{aligned}\Sigma^{\pm}_{Y\leqslant Y_{\min}}(\mathbf{l_1},\mathbf{l_2}) &\ll_{P,g,\varepsilon} q^{10\varepsilon}N\frac{LN}{c}c^{1/2}\left(\frac{L^2N}{e^2}\right)^{1/2}\left(\frac{dNY_{\min}}{df'f^*}\right)^{1/2}\left(\frac{\mathbf{Y}}{Y_{\min}}\right)^{\theta}\\ &\ll_{P,g,\varepsilon} q^{10\varepsilon}N\frac{L^{2+2\theta}N^{2+\theta}}{(f^*)^{1/2}e^{1+\theta}c^{1/2}}Y_{\min}^{1/2-\theta}.\end{aligned}\tag{3.16}$$

For $Y>Y_{\min}$, we apply Theorem 4 and for this we set (cf. the next section)

$$\eta_{1Z} := 13+3\theta, \qquad \eta_{1L} := 1, \qquad \eta_{1X} := 0, \qquad \eta_{1Y} := 1, \qquad \eta_{1Y/X} := 4,$$

$$\eta_{1\mathbf{c}} := \frac{1+2\theta}{2}, \qquad \eta_{1q} := \frac{1-2\theta-2\delta_{\mathrm{tw}}}{2},$$

and

$$D_Z := -\frac{11+3\theta}{2(1+\theta)}, \quad D_L := -\frac{1}{4(1+\theta)}, \quad D_X := \frac{1}{4(1+\theta)}, \quad D_Y := 0, \quad D_{Y/X} := -\frac{2}{1+\theta},$$

$$D_{\mathbf{c}} := -\frac{\theta}{2(1+\theta)}, \qquad D_q := -\frac{1-2\theta-2\delta_{\mathrm{tw}}}{4(1+\theta)} = -\frac{\eta_{1q}}{2(1+\theta)},$$

$$\eta_{2Z} := \eta_{1Z} + (1+2\theta)D_Z = \frac{15+7\theta}{2(1+\theta)}, \quad \eta_{2L} := \eta_{1L} + (1+2\theta)D_L = \frac{3+2\theta}{4(1+\theta)},$$

$$\eta_{2X} := \eta_{1X} + (1+2\theta)D_X = \frac{1+2\theta}{4(1+\theta)}, \quad \eta_{2Y} := \eta_{1Y} = 1, \quad \eta_{2Y/X} := \eta_{1Y/X} + (1+2\theta)D_{Y/X} = \frac{2}{1+\theta},$$

$$\eta_{2\mathbf{c}} := \eta_{1\mathbf{c}} + (1+2\theta)D_{\mathbf{c}} = \frac{1+2\theta}{2(1+\theta)}, \quad \eta_{2q} := \eta_{1q} + (1+2\theta)D_q = \frac{1-2\theta-2\delta_{\mathrm{tw}}}{4(1+\theta)} = \frac{\eta_{1q}}{2(1+\theta)}.$$

It follows from (3.6) that $\mathbf{l_1 l_2}$ is coprime with $\mathbf{q} = q^* \mid q$ and also with D, therefore if the cusp form g satisfies (3.8) then Theorem 4 yields, by (3.15),

$$\begin{aligned}\Sigma^{\pm}_{Y>Y_{\min}}(\mathbf{l_1},\mathbf{l_2}) \ll_{P,g,\varepsilon} q^{10\varepsilon} N \frac{LN}{c} & \\ \times\Bigg(\left(\frac{LN}{c}\right)^{\eta_{1Z}} (Lc^{\flat})^{2\eta_{1L}} \left(\frac{L^2N}{e}\right)^{\eta_{1Y}+\eta_{1Y/X}+\theta} & Y_{\min}^{\eta_{1X}-\eta_{1Y/X}-\theta} \left(\frac{c}{c^{\flat}}\right)^{\eta_{1\mathbf{c}}} (q^*)^{\eta_{1q}} \\ + \left(\frac{LN}{c}\right)^{\eta_{2Z}} (Lc^{\flat})^{2\eta_{2L}} \left(\frac{L^2N}{e}\right)^{\eta_{2Y}+\eta_{2Y/X}+\theta} & Y_{\min}^{\eta_{2X}-\eta_{2Y/X}-\theta} \left(\frac{c}{c^{\flat}}\right)^{\eta_{2\mathbf{c}}} (q^*)^{\eta_{2q}} \Bigg),\end{aligned}$$

i.e.,

$$\begin{aligned}\Sigma^{\pm}_{Y>Y_{\min}}(\mathbf{l_1},\mathbf{l_2}) \ll_{P,g,\varepsilon} q^{10\varepsilon} N \Bigg(L^{26+5\theta} N^{19+4\theta} \frac{(c^{\flat})^{\frac{3-2\theta}{2}}}{e^{5+\theta} c^{\frac{27+4\theta}{2}}} (q^*)^{\eta_{1q}} Y_{\min}^{-(4+\theta)} & \\ + L^{\frac{32+19\theta+4\theta^2}{2(1+\theta)}} N^{\frac{23+13\theta+2\theta^2}{2(1+\theta)}} \frac{(c^{\flat})^{\frac{1}{1+\theta}}}{e^{\frac{3+2\theta+\theta^2}{1+\theta}} c^{\frac{16+7\theta}{2(1+\theta)}}} (q^*)^{\eta_{2q}} Y_{\min}^{-\frac{7+2\theta+4\theta^2}{4(1+\theta)}} \Bigg).\end{aligned}$$

A comparison of the second portion of this bound with (3.16) suggests the choice

$$Y_{\min} := L^{\frac{56+22\theta}{9}} N^{\frac{38+14\theta}{9}} (c^{\flat})^{\frac{4}{9}} e^{-\frac{8}{9}} (f^*)^{\frac{2(1+\theta)}{9}} c^{-\frac{10+4\theta}{3}} (q^*)^{\frac{1-2\theta-2\delta_{\mathrm{tw}}}{9}}.$$

Note that $c \leqslant q^{\varepsilon} LN$ and $q^* \leqslant c$ imply that $Y_{\min} \geqslant 1$. With this choice, one has

$$\begin{aligned}\Sigma^{\pm}(\mathbf{l_1},\mathbf{l_2}) \ll_{P,g,\varepsilon} q^{10\varepsilon} N \Bigg(L^{\frac{10-99\theta-22\theta^2}{9}} N^{\frac{19-58\theta-14\theta^2}{9}} \frac{(c^{\flat})^{\frac{-5-26\theta}{18}}}{(f^*)^{\frac{8+10\theta+2\theta^2}{9}} e^{\frac{13+\theta}{9}} c^{\frac{1-40\theta-8\theta^2}{6}}} (q^*)^{\delta_{q^*}} & \\ + L^{\frac{46-27\theta-22\theta^2}{9}} N^{\frac{37-22\theta-14\theta^2}{9}} \frac{(c^{\flat})^{\frac{2-4\theta}{9}}}{(f^*)^{\frac{7+2\theta+4\theta^2}{18}} e^{\frac{13+\theta}{9}} c^{\frac{13-16\theta-8\theta^2}{6}}} (q^*)^{\delta_{q^*}} \Bigg),\end{aligned}$$

where

$$\delta_{q^*} := (1-2\theta)\frac{1-2\theta-2\delta_{\mathrm{tw}}}{18}.$$

We note that by $f^*|f|c^{\flat}|e|c$ and $(e, qD) = 1$,

$$\sum_{c \equiv 0\,([q,D])} \sum_{f|c^{\flat}} \frac{ef(f^*)^{\theta}}{c^2} \frac{(c^{\flat})^{\frac{-5-26\theta}{18}}}{(f^*)^{\frac{8+10\theta+2\theta^2}{9}} e^{\frac{13+\theta}{9}} c^{\frac{1-40\theta-8\theta^2}{6}}} \ll_{\varepsilon} q^{\varepsilon - \frac{13-40\theta-8\theta^2}{6}}$$

and

$$\sum_{c \equiv 0\,([q,D])} \sum_{f|c^{\flat}} \frac{ef(f^*)^{\theta}}{c^2} \frac{(c^{\flat})^{\frac{2-4\theta}{9}}}{(f^*)^{\frac{7+2\theta+4\theta^2}{18}} e^{\frac{13+\theta}{9}} c^{\frac{13-16\theta-8\theta^2}{6}}} \ll_{\varepsilon} q^{\varepsilon - \frac{25-16\theta-8\theta^2}{6}}.$$

Collecting all the terms (see (3.9), (3.12), (3.13)) and using also $q^* \leqslant q$, we deduce that for g satisfying (3.8) and for $N \leqslant (qDP)^{1+\varepsilon}$,

$$Q^{\mathrm{ND}}(\vec{x}, N) \ll_{P,g,\varepsilon} q^{100\varepsilon} \|\vec{x}\|_1^2 N L^{\delta_L} q^{-\delta_q}$$

with

$$\delta_L := 2\theta + \frac{46 - 27\theta - 22\theta^2}{9} = \frac{46 - 9\theta - 22\theta^2}{9},$$

$$\begin{aligned}\delta_q &:= \frac{13 - 40\theta - 8\theta^2}{6} - \frac{19 - 58\theta - 14\theta^2}{9} - \delta_{q^*}\\ &= \frac{25 - 16\theta - 8\theta^2}{6} - \frac{37 - 22\theta - 14\theta^2}{9} - \delta_{q^*}\\ &= \frac{1-2\theta}{9}\,\delta_{\mathrm{tw}}.\end{aligned}$$

For g not satisfying (3.8), an additional term occurs whose contribution to $Q^{\mathrm{ND}}(\vec{x}, N)$ is bounded by (cf. Theorem 4)

$$\ll_{P,g,\varepsilon} q^{100\varepsilon}\|\vec{x}\|_1^2 N \left(L^{\delta_{3L}} q^{-\delta_{3q}} + L^{\delta_{4L}} q^{-\delta_{4q}}\right)$$

with

$$\delta_{3L} := 9 + 4\theta, \qquad \delta_{3q} := \frac{1}{2} - \theta,$$

$$\delta_{4L} := \frac{7 + 10\theta + 4\theta^2}{2(1+\theta)}, \qquad \delta_{4q} := \frac{1+4\theta}{4(1+\theta)}.$$

The above estimates together with (3.11) conclude the proof of Proposition 3.2.

4. A shifted convolution problem

Our main point is to solve the following instance of the shifted convolution problem: let χ be a primitive character of modulus $q > 1$, $1 < \mathbf{c} \equiv 0\,(q)$, $\ell_1, \ell_2 \geqslant 1$ be two integers, and g be a primitive cusp form of level D and nebentypus χ_g. We assume that g is arithmetically normalized by which we mean that its first Fourier coefficient (see (2.1)) $\rho_g(1)$ equals one and consequently, by (2.3), that

$$\lambda_g(n) = \sqrt{n}\rho_g(n)$$

for any $n \geqslant 1$.

Given $X, Y, Z \geqslant 1$ and a smooth function $f(u,v)$ supported on $[1/4, 4] \times [1/4, 4]$ satisfying $\|f\|_\infty = 1$ and

$$\frac{\partial^i}{\partial u^i}\frac{\partial^j}{\partial v^j} f(u,v) \ll Z^{i+j}$$

for all $i, j \geqslant 0$, where the implied constant depends only on i and j , we consider $F(x,y) := f(\frac{x}{X}, \frac{y}{Y})$ which is supported on $[X/4, 4X] \times [Y/4, 4Y]$ and satisfies

$$x^i y^j \frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j} F(x,y) \ll Z^{i+j} \tag{4.1}$$

for all $i, j \geqslant 0$, the implied constant depending at most on i and j.

We consider the sum

$$\Sigma_\chi^{\pm}(\ell_1, \ell_2; \mathbf{c}) := \sum_{h \neq 0} G_\chi(h; \mathbf{c}) S_h^{\pm}(\ell_1, \ell_2),$$

where $G_\chi(h; \mathbf{c})$ is the Gauss sum of the (induced) character χ (mod $\mathbf{c}$) and

$$S_h^{\pm}(\ell_1, \ell_2) := \sum_{\ell_1 m \mp \ell_2 n = h} \overline{\lambda_g}(m)\lambda_g(n) F(\ell_1 m, \ell_2 n). \tag{4.2}$$

Our goal is

Theorem 4. *Assume Hypothesis H_θ. Set*

$$P := \mathbf{c} D \mu_g \ell_1 \ell_2 Z(X+Y),$$

and assume (as one may by symmetry) that $Y \geqslant X$. Let $\delta_{\rm tw} := \frac{1-2\theta}{8}$ be the power saving exponent of Theorem 5, and let $D_{\rm opt}$ be

$$D_{\rm opt} := Z^{-\frac{11+3\theta}{2(1+\theta)}} (\ell_1\ell_2)^{-\frac{1}{4(1+\theta)}} \mathbf{c}^{-\frac{\theta}{2(1+\theta)}} q^{-\frac{1-2\theta-2\delta_{\rm tw}}{4(1+\theta)}} (X/Y)^{\frac{2}{1+\theta}} X^{\frac{1}{4(1+\theta)}}.$$

Suppose that

$$w | D\ell_1\ell_2 \quad \Longrightarrow \quad q \nmid (w, D\ell_1\ell_2/w),$$

then the following upper bound holds:

$$\Sigma^{\pm}_{\chi}(\ell_1,\ell_2;\mathbf{c}) \ll_{g,\varepsilon} P^{\varepsilon} Z^{13+3\theta} (\ell_1\ell_2) \mathbf{c}^{\frac12+\theta} q^{\frac12-\theta-\delta_{\rm tw}} (Y/X)^4 Y (1+D_{\rm opt})^{1+2\theta}.$$

On the other hand, if $q | (w, D\ell_1\ell_2/w)$ for some $w | D\ell_1\ell_2$ (in which case $q \leqslant (D\ell_1\ell_2)^{\frac12}$), the above bound holds up to an additional term bounded by

$$\ll P^{\varepsilon} Z^4 (\mathbf{c},\ell_1\ell_2)^{\frac12} (\ell_1\ell_2)^{\frac12} Y^{\frac32} (1+D_{\rm opt}).$$

In these bounds, the implied constants depend at most on ε and g. The latter dependence is at most polynomial in D and μ_g, where D (resp. μ_g) denotes the level (resp. spectral parameter given in (2.2)) of g.

Remark 4.1. It is crucial for applications to subconvexity that the sums of the exponents in the X, Y, $\mathbf{c}$, q variables are strictly smaller than 2: indeed, one has

$$1 + \left(\frac12 + \theta\right) + \left(\frac12 - \theta - \delta_{\rm tw}\right) = 2 - \delta_{\rm tw}$$

and

$$\frac{1+2\theta}{4(1+\theta)} + 1 + (1+2\theta)\left(\frac12 - \frac{\theta}{2(1+\theta)}\right) + \left(\frac12 - \theta - \delta_{\rm tw}\right)\left(1 - \frac{1+2\theta}{2(1+\theta)}\right) = 2 - \frac{\delta_{\rm tw}}{2(1+\theta)}.$$

The proof of this theorem will take us the next two sections. For the rest of this section and in the next one, we use the following convention: $\dots \ll_g \dots$ means implicitly that the implied constant in the Vinogradov symbol depends at most polynomially on D and μ_g.

By symmetry, we assume that $Y \geqslant X$. Considering the unique factorization

$$\mathbf{c} = qq'\mathbf{c}', \qquad (\mathbf{c}', q) = 1, \qquad q' | q^{\infty},$$

we have

$$G_\chi(h;\mathbf{c}) = \chi(\mathbf{c}') G_\chi(h; qq') r(h;\mathbf{c}'),$$

where

$$r(h;\mathbf{c}') = \sum_{d | (\mathbf{c}', h)} d\mu(\mathbf{c}'/d)$$

denotes the Ramanujan sum. Moreover, $G_\chi(h;qq') = 0$ unless $q' | h$ in which case

$$G_\chi(h;qq') = \overline{\chi}(h/q') q' G_\chi(1;q),$$

hence we have

$$\Sigma^{\pm}_{\chi}(\ell_1,\ell_2;\mathbf{c}) = \chi(\mathbf{c}') q' G_\chi(1;q) \sum_{d|\mathbf{c}'} d\mu(\mathbf{c}'/d)\overline{\chi}(d) \sum_{h\neq 0} \overline{\chi}(h) S^{\pm}_{hq'd}(\ell_1,\ell_2). \tag{4.3}$$

Observe that by (4.1), (4.2) and (2.28) this implies the trivial bound

$$\begin{aligned}\Sigma^{\pm}_{\chi}(\ell_1,\ell_2;\mathbf{c}) &\ll q'q^{1/2}\sum_{d|\mathbf{c}'} d \sum_{\substack{m\ll X/\ell_1\\ n\ll Y/\ell_2\\ q'd|\ell_1 m\mp\ell_2 n}} |\lambda_g(m)||\lambda_g(n)|\\ &\leqslant q'q^{1/2}\sum_{d|\mathbf{c}'} d \sum_{\substack{m\ll X/\ell_1\\ n\ll Y/\ell_2\\ q'd|\ell_1 m\mp\ell_2 n}} \left(|\lambda_g(m)|^2+|\lambda_g(n)|^2\right)\\ &\ll_{\varepsilon} P^{\varepsilon}q^{1/2}\frac{(\ell_1\ell_2,\mathbf{c})}{\ell_1\ell_2}XY.\end{aligned}$$

When q is large a better bound can be obtained from an application of Lemma 2.3: integrating by parts, we obtain

$$\begin{aligned}\Sigma^{\pm}_{\chi}(\ell_1,\ell_2;\mathbf{c}) &\leqslant q'q^{1/2}\sum_{d|\mathbf{c}'} d \iint_{(\mathbf{R}^+)^2}\ell_1\ell_2\left|\frac{\partial^2}{\partial x\partial y}F(\ell_1 x,\ell_2 y)\right|\sum_{h\neq 0}\Bigg|\sum_{\substack{m\leqslant x,\ n\leqslant y\\ \ell_1 m\mp\ell_2 n=dq'h}}\overline{\lambda_g}(m)\lambda_g(n)\Bigg|\,dxdy\\ &\leqslant Z^2 q'q^{1/2}\sum_{d|\mathbf{c}'} d \max_{\substack{x\ll X/\ell_1\\ y\ll Y/\ell_2}}\sum_{|dq'h|\leqslant \ell_1 x+\ell_2 y}\Bigg|\sum_{\substack{m\leqslant x,\ n\leqslant y\\ \ell_1 m\mp\ell_2 n=h}}\overline{\lambda_g}(m)\lambda_g(n)\Bigg|\\ &\ll_{\varepsilon} P^{\varepsilon}D^2\mu_g^4 Z^2 q'q^{1/2}\sum_{d|\mathbf{c}'} d\left(\frac{X+Y}{dq'}\right)^{1/2}\left(\frac{XY}{\ell_1\ell_2}\right)^{1/2}\\ &\ll_{\varepsilon} P^{2\varepsilon}D^2\mu_g^4 Z^2 \mathbf{c}^{1/2}(X+Y)^{1/2}\left(\frac{XY}{\ell_1\ell_2}\right)^{1/2}.\end{aligned}\tag{4.4}$$

On the other hand, an application of the δ-symbol method of [DFI94b] yields (cf. [M04a, Section 7.1], [H03b, Theorem 3.1], [KMV02, Appendix B])

$$\Sigma^{\pm}_{\chi}(\ell_1,\ell_2;\mathbf{c}) \ll_{g,\varepsilon} P^{\varepsilon}Z^{5/4}q^{1/2}X^{1/4}Y^{3/2}.$$

For our given subconvexity problem, one typically has $\mathbf{c}\sim\sqrt{XY}$, $X\sim Y$ and $\ell_1\ell_2$ is a very small power of Y.

4.1. **A variant of the δ-symbol method.** In the next two sections, we only treat the case of the "+" sums (i.e., $\Sigma^{+}_{\chi}(\ell_1,\ell_2;\mathbf{c})$ and $S^{+}_{h}(\ell_1,\ell_2)$), the case of the "−" sums being identical; consequently, we simplify notation by omitting the "+" sign from $\Sigma^{+}_{\chi}(\dots)$ and $S^{+}_{h}(\dots)$.

We shall assume that

$$Y\geqslant(4D\ell_1\ell_2)^2,$$

for otherwise the bound of Theorem 4 follows from (4.4). We denote by $D(g,\ell_1,\ell_2,q'd)$ the h-sum in (4.3); to simplify notation further, we change it slightly and replace $\overline{\chi}$ by χ and $q'd$ by d and set

$$D(g,\ell_1,\ell_2,d):=\sum_{h\neq 0}\chi(h)S_{hd}(\ell_1,\ell_2)=\sum_{h\neq 0}\chi(h)\sum_{\ell_1 m-\ell_2 n=dh}\overline{\lambda_g}(m)\lambda_g(n)F(\ell_1 m,\ell_2 n)\phi(dh).$$

As in [DFI94a], we have multiplied $F(\ell_1 m,\ell_2 n)$ by a redundancy factor $\phi(dh)$, where ϕ is a smooth even function satisfying $\phi_{|[-2Y,2Y]}\equiv 1$, $\mathrm{supp}\phi\subset[-4Y,4Y]$ and $\phi^{(i)}(x)\ll_i Y^{-i}$. Of course, this extra factor does not change the value of $D(g,\ell_1,\ell_2,d)$, but will prove to be useful in the forthcoming computations.

We detect the summation condition $\ell_1 m - \ell_2 n - dh = 0$ by means of additive characters:

$$D(g,\ell_1,\ell_2,d) = \int_{\mathbf{R}} G(\alpha) 1_{[0,1]}(\alpha)\, d\alpha$$

with

$$G(\alpha) := \sum_{h\neq 0} \chi(h) \sum_{m,n\geqslant 1} \overline{\lambda_g}(m)\lambda_g(n) F(\ell_1 m, \ell_2 n)\phi(dh) e\big(\alpha(\ell_1 m - \ell_2 n - dh)\big).$$

As in [H03a], we apply Jutila's method of overlapping intervals [J92, J96] to approximate the characteristic function of the unit interval $I(\alpha) = 1_{[0,1]}(\alpha)$ by sums of characteristic functions of intervals centered at well chosen rationals: for this we consider some C satisfying

$$Y^{1/2} \leqslant C \leqslant Y$$

and a smooth function w supported on $[C/2, 3C]$ with values in $[0,1]$ equal to 1 on $[C, 2C]$ such that $w^{(i)}(x) \ll_i C^{-i}$; we also set

$$\delta := Y^{-1}, \qquad N := D\ell_1\ell_2, \qquad L := \sum_{c\equiv 0\,(N)} w(c)\varphi(c).$$

Note that $C \geqslant 4D\ell_1\ell_2$, hence L satisfies the inequality

$$L \gg_\varepsilon C^{2-\varepsilon}/N \gg_{g,\varepsilon} C^{2-\varepsilon}/(\ell_1\ell_2) \tag{4.5}$$

for any $\varepsilon > 0$. The approximation to $I(\alpha)$ is provided by

$$\tilde{I}(\alpha) := \frac{1}{2\delta L} \sum_{c\equiv 0\,(N)} w(c) \sum_{\substack{a(c)\\(a,c)=1}} 1_{[\frac{a}{c}-\delta, \frac{a}{c}+\delta]}(\alpha)$$

(which is supported in $[-1,2]$), and by the main theorem in [J92] one has

$$\int_{[-1,2]} |I(\alpha) - \tilde{I}(\alpha)|^2 d\alpha \ll_\varepsilon \frac{C^{2+\varepsilon}}{\delta L^2} \ll_{g,\varepsilon} C^{2\varepsilon}(\ell_1\ell_2)^2 \frac{Y}{C^2}. \tag{4.6}$$

Next, we introduce the corresponding approximation of $D(g,\ell_1,\ell_2,d)$:

$$\tilde{D}(g,\ell_1,\ell_2,d) := \int_{[-1,2]} G(\alpha)\tilde{I}(\alpha) d\alpha,$$

then it follows from (4.6) that

$$|D(g,\ell_1,\ell_2,d) - \tilde{D}(g,\ell_1,\ell_2,d)| \leqslant \|I - \tilde{I}\|_2 \|G\|_2 \ll_{g,\varepsilon} C^\varepsilon \ell_1\ell_2 \frac{Y^{1/2}}{C}\|G\|_2.$$

4.2. **Bounding $\|G\|_2$.** We factor $G(\alpha)$ as

$$G(\alpha) = \sum_{h\neq 0} \chi(h)\phi(dh) e\big(-\alpha dh\big) \times \sum_{m,n\geqslant 1} \overline{\lambda_g}(m)\lambda_g(n) F(\ell_1 m, \ell_2 n) e\big(\alpha(\ell_1 m - \ell_2 n)\big) =: H(\alpha)K(\alpha),$$

say. By Parseval, we have

$$\|G\|_2 \leqslant \|H\|_2 \|K\|_\infty \ll \left(\frac{Y}{d}\right)^{1/2} \|K\|_\infty.$$

Integrating by parts shows that (cf. (2.32))

$$K(\alpha) = \ell_1\ell_2 \iint_{(\mathbf{R}^+)^2} F^{(1,1)}(\ell_1 x, \ell_2 y) \overline{S_g(-\ell_1\alpha, x)} S_g(-\ell_2\alpha, y)\, dxdy,$$

where by (4.1),

$$F^{(1,1)}(\ell_1 x, \ell_2 y) \ll \frac{Z^2}{XY},$$

and by Proposition 2.4,

$$\overline{S_g(-\ell_1\alpha, x)} S_g(-\ell_2\alpha, y) \ll_{g,\varepsilon} (xy)^{1/2+\varepsilon},$$

so that

$$\|K\|_\infty \ll_{g,\varepsilon} (XY)^\varepsilon Z^2 \left(\frac{XY}{\ell_1\ell_2}\right)^{1/2}.$$

Collecting the above estimates, we find that

$$D - \tilde{D} \ll_{g,\varepsilon} P^\varepsilon Z^2 \left(\ell_1\ell_2 XY\right)^{1/2} \left(\frac{Y}{d}\right)^{1/2} \frac{Y^{1/2}}{C},$$

therefore the contribution of this difference to $\Sigma_\chi(\ell_1, \ell_2; \mathbf{c})$ is bounded by

$$\ll_{g,\varepsilon} P^{2\varepsilon} Z^2 (\ell_1\ell_2)^{1/2} \mathbf{c}^{1/2} \frac{X^{1/2}Y^{3/2}}{C}. \tag{4.7}$$

4.3. **Bounding $\tilde{D}$.** We have

$$\tilde{D} = \frac{1}{L} \sum_{c\equiv 0\,(N)} w(c) \sum_{\substack{a(c)\\(a,c)=1}} \mathfrak{I}_{\delta,\frac{a}{c}},$$

where

$$\mathfrak{I}_{\delta,\frac{a}{c}} := \sum_h \chi(h) e\left(\frac{-adh}{c}\right) \sum_{m,n} \overline{\lambda_g}(m)\lambda_g(n) e\left(\frac{a\ell_1 m}{c}\right) e\left(\frac{-a\ell_2 n}{c}\right) E(m,n,h)$$

and

$$E(x,y,z) := F(\ell_1 x, \ell_2 y)\phi(dz) \frac{1}{2\delta} \int_{-\delta}^{\delta} e\big(\alpha(\ell_1 x - \ell_2 y - dz)\big)\, d\alpha.$$

By applying Proposition 2.1 to the variables m, n (cf. (2.3), (2.6), (2.7)) and by summing over a, c, we get (observe that the factor $\overline{\chi_g}(\overline{a})$ from the m-sum is cancelled by $\chi_g(\overline{a})$ coming from the n-sum)

$$\tilde{D} = \sum_{\pm,\pm} \varepsilon_g^{\pm} \varepsilon_g^{\pm} \tilde{D}^{\pm,\pm},$$

where $\varepsilon_g^+ = 1$ and $\varepsilon_g^- = \pm 1$ is the sign in (2.7) (for g not induced from a holomorphic form),

$$\tilde{D}^{\pm,\pm} := \frac{1}{L} \sum_{m,n} \overline{\lambda_g}(m)\lambda_g(n) \sum_{c\equiv 0\,(N)} \sum_h \chi(h) \frac{S(dh, \mp\ell_1 m \pm \ell_2 n; c)}{c} \mathcal{E}^{\pm,\pm}(m,n,h;c) \tag{4.8}$$

and

$$\mathcal{E}^{\pm,\pm}(m,n,h;c) := \frac{\ell_1\ell_2 w(c)}{c} \iint_{(\mathbf{R}^+)^2} E(x,y,h) J_g^{\pm}\left(\frac{4\pi\ell_1\sqrt{mx}}{c}\right) J_g^{\pm}\left(\frac{4\pi\ell_2\sqrt{ny}}{c}\right) dxdy.$$

4.4. **Estimates for $\mathcal{E}^{\pm,\pm}$ and its derivatives.** Notice that the definition of E and the various assumptions made so far imply that

$$E(x,y,z) = 0 \ \text{ unless } \ x \sim X/\ell_1,\ y \sim Y/\ell_2,\ |dz| \leqslant 4Y. \tag{4.9}$$

Moreover,

$$E^{(i,j,k)}(x,y,z) \ll_{i,j,k} \frac{Z^{i+j}\ell_1^i \ell_2^j d^k}{X^i Y^{j+k}}, \tag{4.10}$$

so that for any fixed h

$$\|E^{(i,j,k)}(*,*,h)\|_1 \ll_{i,j,k} \frac{Z^{i+j}\ell_1^{i-1}\ell_2^{j-1} d^k XY}{X^i Y^{j+k}}, \tag{4.11}$$

and therefore

$$\|E^{(i,j,k)}\|_1 \ll_{i,j,k} \frac{Z^{i+j}\ell_1^{i-1}\ell_2^{j-1} d^{k-1} XY^2}{X^i Y^{j+k}}.$$

Next, we evaluate $\mathcal{E}^{\pm,\pm}(m,n,h;c)$ and its partial derivatives: depending on the case, $\mathcal{E}^{\pm,\pm}(m,n,h;c)$ can be written as a linear combination (with constant coefficients) of integrals of the form

$$\frac{\ell_1\ell_2 w(c)}{c}\iint_{(\mathbf{R}^+)^2} E(x,y,h) J_{1,\nu_1}\left(\frac{4\pi\ell_1\sqrt{mx}}{c}\right) J_{2,\nu_2}\left(\frac{4\pi\ell_2\sqrt{ny}}{c}\right) dxdy, \tag{4.12}$$

where

$$\big\{J_{1,\nu}(x), J_{2,\nu}(x)\big\} \subset \left\{\frac{Y_\nu(x)}{\mathrm{ch}(\pi t)},\ \mathrm{ch}(\pi t)K_\nu(x)\right\}$$

with $\nu\in\{\pm 2it_g\}$ if g is a Maass form of weight 0; or

$$\big\{J_{1,\nu}(x), J_{2,\nu}(x)\big\} \subset \left\{\frac{Y_\nu(x)}{\mathrm{sh}(\pi t)},\ \mathrm{sh}(\pi t)K_\nu(x)\right\}$$

with $\nu\in\{\pm 2it_g\}$ if g is a Maass form of weight 1; or

$$J_{1,\nu}(x) = J_{2,\nu}(x) = J_{k_g-1}(x),$$

if g is a holomorphic form of weight k_g.

In order to estimate (4.12) efficiently, we integrate by parts i (resp. j) times with respect to x (resp. y), where i (resp. j) will be determined later in terms of m (resp. n) and ε. Using (7.1), we see that $\mathcal{E}^{\pm,\pm}(m,n,h;c)$ can be written as a linear combination (with constant coefficients) of expressions of the form

$$\begin{aligned}\frac{\ell_1\ell_2 w(c)}{c}\left(\frac{\ell_1\sqrt{m}}{c}\right)^{-2i}\left(\frac{\ell_2\sqrt{n}}{c}\right)^{-2j}\iint_{(\mathbf{R}^+)^2}&\frac{\partial^{i+j}}{\partial x^i\partial y^j}\big\{E(x,y,h)W_1^{-\nu_1}W_2^{-\nu_2}\big\}\\ &\times W_1^{\nu_1+i}W_2^{\nu_2+j}J_{1,\nu_1+i}(W_1)J_{2,\nu_2+j}(W_2)\,dxdy,\end{aligned}$$

where $\{\nu_1,\nu_2\}\subset\{\pm 2it_g\}$ (or $\nu_1,\nu_2=k_g-1$) and

$$W_1 := \frac{4\pi\ell_1\sqrt{mx}}{c}\sim\frac{\sqrt{m\ell_1X}}{C},\qquad W_2 := \frac{4\pi\ell_2\sqrt{ny}}{c}\sim\frac{\sqrt{n\ell_2Y}}{C},$$

in view of (4.9). Using (4.11) and Proposition 7.2 in the slightly weaker form

$$\begin{aligned}J_{1,\nu_1+i}(W_1)&\ll_{i,\varepsilon}\mu_g^{i+\varepsilon}\big(1+W_1^{-1}\big)^{i+2|\Im t_g|+\varepsilon}\big(1+W_1\big)^{-1/2},\\ J_{2,\nu_2+j}(W_2)&\ll_{j,\varepsilon}\mu_g^{j+\varepsilon}\big(1+W_2^{-1}\big)^{j+2|\Im t_g|+\varepsilon}\big(1+W_2\big)^{-1/2},\end{aligned}$$

we can deduce for any $i,j\geqslant 0$ that

$$\mathcal{E}^{\pm,\pm}(m,n,h;c)\ll_{i,j,\varepsilon}P^\varepsilon(\mu_g^2Z)^{i+j}\left\{\frac{C^2}{\ell_1mX}+\left(\frac{C^2}{\ell_1mX}\right)^{1/2}\right\}^i\left\{\frac{C^2}{\ell_2nY}+\left(\frac{C^2}{\ell_2nY}\right)^{1/2}\right\}^j\Xi(m,n),$$

where

$$\Xi(m,n) := \frac{XY}{C}\left\{\left(1+\frac{C^2}{\ell_1mX}\right)\left(1+\frac{C^2}{\ell_2nY}\right)\right\}^{|\Im t_g|}\left\{\left(1+\frac{\ell_1mX}{C^2}\right)\left(1+\frac{\ell_2nY}{C^2}\right)\right\}^{-1/4}. \tag{4.13}$$

This shows, upon choosing i and j appropriately, that $\mathcal{E}^{\pm,\pm}(m,n,h;c)$ is very small unless

$$d|h|\leqslant 4Y,\qquad c\sim C,\qquad m\ll_\varepsilon P^\varepsilon\frac{\mu_g^4Z^2C^2}{\ell_1X},\qquad n\ll_\varepsilon P^\varepsilon\frac{\mu_g^4Z^2C^2}{\ell_2Y}, \tag{4.14}$$

and in this range we retain the bound (by taking $i=j=0$)

$$\mathcal{E}^{\pm,\pm}(m,n,h;c)\ll_\varepsilon P^\varepsilon\Xi(m,n). \tag{4.15}$$

The partial derivatives

$$h^jc^k\frac{\partial^{j+k}}{\partial h^j\partial c^k}\mathcal{E}^{\pm,\pm}(m,n,h;c)$$

can be estimated similarly. We shall restrict our attention to the range (4.14); the argument also yields that outside this range the partial derivatives are very small. By (7.2), the above partial derivative is a linear combination of integrals of the form

$$R_k(it_g)c^{k_3}\frac{\partial^{k_3}}{\partial c^{k_3}}\left(\frac{w(c)}{c}\right)\iint_{(\mathbf{R}^+)^2} h^j\frac{\partial^j}{\partial h^j}E(x,y,h)W_1^{k_1}W_2^{k_2}J_{1,\nu_1-k_1}(W_1)J_{2,\nu_2-k_2}(W_2)\,dxdy,$$

where R_k is a polynomial of degree $\leqslant k$ and the integers k_1,k_2,k_3 satisfy

$$k_1+k_2+k_3\leqslant k.$$

Therefore we obtain

$$\begin{aligned} h^jc^k(\mathcal{E}^{\pm,\pm})^{(0,0,j,k)}(m,n,h;c) &\ll_{j,k,\varepsilon} P^\varepsilon\left(\frac{d|h|}{Y}\right)^j\mu_g^k\left(1+\frac{\sqrt{\ell_1mX}}{C}+\frac{\sqrt{\ell_2nY}}{C}\right)^k\Xi(m,n)\\ &\ll_{j,k,\varepsilon} P^\varepsilon(P^\varepsilon\mu_g^3Z)^k\Xi(m,n). \end{aligned} \tag{4.16}$$

5. Expanding the c-sum

Our next step will be to expand spectrally the c-sum in (4.8) as a sum over a basis of Maass and holomorphic forms on $\Gamma_0(N)$. To do this we use the complete version of the Petersson–Kuznetsov formulae given in Theorem 3. We only treat $\tilde{D}^{-,-}$, the other terms being similar. To simplify notation further, we denote $\tilde{D}^{-,-}$ by $\tilde{D}$ and $\mathcal{E}^{-,-}$ by $\mathcal{E}$. The shape of the sum formula depends on the sign of the product $h(\ell_1m-\ell_2n)$ when it is nonzero. So our first step will be to isolate the contribution of the m,n such that $\ell_1m-\ell_2n=0$ (the contribution of the $h=0$ is void since we assume that χ is non-trivial). Thus we have the splitting

$$\tilde{D}=\tilde{D}^0+\tilde{D}^++\tilde{D}^-,$$

where

$$\tilde{D}^0:=\frac{1}{L}\sum_{\ell_1m=\ell_2n}\overline{\lambda_g}(m)\lambda_g(n)\sum_{c\equiv 0\,(N)}\sum_h\chi(h)\frac{r(dh;c)}{c}\mathcal{E}(m,n,h;c)$$

with

$$r(dh;c)=S(dh,0;c)=\sum_{c'|(dh,c)}\mu(c/c')c'$$

the Ramanujan sum, and

$$\begin{aligned}\tilde{D}^\pm:=&\frac{1}{L}\sum_{\ell_1m-\ell_2n\neq 0}\overline{\lambda_g}(m)\lambda_g(n)\sum_{c\equiv 0\,(N)}\sum_{\substack{h\\ \pm h(\ell_1m-\ell_2n)>0}}\chi(h)\frac{S(dh,\ell_1m-\ell_2n;c)}{c}\mathcal{E}(m,n,h;c)\\ =&\frac{1}{L}\sum_{\ell_1m-\ell_2n\neq 0}\overline{\lambda_g}(m)\lambda_g(n)\tilde{D}^\pm(m,n)\end{aligned}$$

with

$$\tilde{D}^\pm(m,n):=\sum_{c\equiv 0\,(N)}\sum_{\pm hh'>0}\chi(h)\frac{S(dh,h';c)}{c}\mathcal{E}(m,n,h;c);$$

here we have set $h':=\ell_1m-\ell_2n\neq 0$.

5.1. **Bounding $\tilde{D}^0$.** We set $\ell_1' := \ell_1/(\ell_1,\ell_2)$, $\ell_2' := \ell_2/(\ell_1,\ell_2)$, then

$$\tilde{D}^0 = \frac{1}{L}\sum_{m\geqslant 1}\overline{\lambda_g}(\ell_2' m)\lambda_g(\ell_1' m)\sum_{c\equiv 0\,(N)}\frac{1}{c}\sum_h \chi(h)r(dh;c)\mathcal{E}(\ell_2' m,\ell_1' m,h;c),$$

and the c-sum equals

$$\sum_{c''}\frac{\mu(c'')}{c''}\sum_{c'\equiv 0\,(N/(c'',N))}\chi\left(\frac{c'}{(c',d)}\right)\sum_h \chi(h)\mathcal{E}\left(\ell_2' m,\ell_1' m,\frac{c'}{(c',d)}h;c'c''\right).$$

Combining partial summation with (4.16) and Burgess' bound

$$\sum_{h\leqslant H}\chi(h)\ll_\varepsilon H^{1/2}q^{3/16+\varepsilon},$$

we find that the h-sum is bounded by

$$\begin{aligned}\sum_h \chi(h)\ldots &\ll_\varepsilon P^\varepsilon\left(\frac{(c',d)}{c'}\right)^{1/2}\frac{Y^{1/2}}{d^{1/2}}q^{3/16}\Xi(\ell_2' m,\ell_1' m)\frac{(c',d)Y}{c'd}\,\frac{c'd}{(c',d)Y}\\ &\ll_\varepsilon P^\varepsilon\left(\frac{(c',d)}{c'}\right)^{1/2}\frac{Y^{1/2}}{d^{1/2}}q^{3/16}\Xi(\ell_2' m,\ell_1' m),\end{aligned}$$

and the c-sum is bounded by

$$\sum_{c\equiv 0\,(N)}\frac{1}{c}\sum_h \chi(h)\cdots\ll_\varepsilon P^{2\varepsilon}\frac{(d,\ell_1\ell_2)^{1/2}}{\ell_1\ell_2}\frac{Y^{1/2}}{d^{1/2}}C^{1/2}q^{3/16}\Xi(\ell_2' m,\ell_1' m).$$

In summing over the m variable we may restrict ourselves to some range

$$[\ell_1,\ell_2]m\ll_{g,\varepsilon} P^\varepsilon Z^2(C^2/Y),$$

as the remaining contribution is negligible. If $Y/X\ll_{g,\varepsilon} P^\varepsilon Z^2$, then we split the m-sum into three parts,

$$\sum_{[\ell_1,\ell_2]m<C^2/Y}\cdots\quad+\sum_{C^2/Y\leqslant[\ell_1,\ell_2]m<C^2/X}\cdots\quad+\sum_{C^2/X\leqslant[\ell_1,\ell_2]m\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/Y)}\cdots\quad,$$

and combine (2.5), (2.28), (4.5) and (4.13) to infer that

$$\tilde{D}^0\ll_{g,\varepsilon} P^{3\varepsilon}\frac{(d,\ell_1\ell_2)^{1/2}}{d^{1/2}[\ell_1,\ell_2]^{1-\theta}}q^{3/16}\frac{XY^{3/2}}{C^{1/2}}\left(X^{-\theta}Y^{\theta-1}+X^{-3/4}Y^{-1/4}+ZX^{-1/4}Y^{-3/4}\right).$$

If $Y/X\gg_{g,\varepsilon} P^\varepsilon Z^2$, then we split the m-sum into two parts,

$$\sum_{[\ell_1,\ell_2]m<C^2/Y}\cdots\quad+\sum_{C^2/Y\leqslant[\ell_1,\ell_2]m\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/Y)}\cdots\quad,$$

and infer similarly that

$$\tilde{D}^0\ll_{g,\varepsilon} P^{3\varepsilon}\frac{(d,\ell_1\ell_2)^{1/2}}{d^{1/2}[\ell_1,\ell_2]^{1-\theta}}q^{3/16}\frac{XY^{3/2}}{C^{1/2}}\left(X^{-\theta}Y^{\theta-1}+Z^{3/2-2\theta}X^{-\theta}Y^{\theta-1}\right).$$

In both cases we conclude that

$$\tilde{D}^0\ll_{g,\varepsilon} P^{4\varepsilon}Z\frac{(d,\ell_1\ell_2)^{1/2}}{d^{1/2}[\ell_1,\ell_2]^{1-\theta}}q^{3/16}\frac{X^{3/4}Y^{3/4}}{C^{1/2}}.$$

Finally, returning to our initial sum $\Sigma_\chi(\ell_1,\ell_2;\mathbf{c})$, we see by (4.3) that the contribution of the $\tilde{D}^0$ terms is bounded by (remember that we have reused the letter d in place of $q'd$)

$$\ll_{g,\varepsilon} P^{5\varepsilon}Z\frac{(\mathbf{c},\ell_1\ell_2)^{1/2}}{[\ell_1,\ell_2]^{1-\theta}}\mathbf{c}^{1/2}q^{3/16}\frac{X^{3/4}Y^{3/4}}{C^{1/2}}.\tag{5.1}$$

Remark 5.1. Notice that in the (important for us) case $q \sim \mathbf{c} \sim X \sim Y$ (remember that $C \geqslant Y^{1/2}$), Burgess' estimate is used crucially in order to improve over the bound Y^2.

5.2. **Preliminary truncation.** We perform a dyadic subdivision on the h variable. By (4.15) and (4.16), we can decompose $\mathcal{E}(m,n,h;c)$ as

$$\mathcal{E}(m,n,h;c) = \sum_{H \geqslant 1} \mathcal{E}_H(m,n,h;c),$$

where $H = 2^\nu$, $\nu \in \mathbf{N}$, and $\mathcal{E}_H(m,n,h;c)$ as a function of h is supported on $[-2H,-H/2] \cup [H/2,2H]$ and satisfies

$$h^j c^k \mathcal{E}_H^{(0,0,j,k)}(m,n,h;c) \ll_{j,k,\varepsilon} P^\varepsilon (P^\varepsilon \mu_g^3 Z)^k \Xi(m,n). \tag{5.2}$$

Accordingly, we have the decomposition $\tilde{D} = \sum_{H \geqslant 1} \tilde{D}_H$.

5.3. **Bounding $\tilde{D}_H^{\pm}(m,n)$.** We shall assume that $H \leqslant 8Y/d$ for otherwise $\tilde{D}_H = 0$. We split $\tilde{D}_H^{\pm}$ into two more sums getting a total of 4 terms, $\tilde{D}_H^{\pm,\pm}$ say, where

$$\tilde{D}_H^{\varepsilon_1,\varepsilon_2} := \frac{1}{L} \sum_{m \geqslant 1} \sum_{\substack{n \geqslant 1 \\ \varepsilon_2 h' > 0}} \overline{\lambda_g}(m) \lambda_g(n) \tilde{D}_H^{\varepsilon_1}(m,n)$$

with

$$\tilde{D}_H^{\varepsilon_1}(m,n) := \sum_{\varepsilon_1 h h' > 0} \chi(h) \sum_{c \equiv 0\,(N)} \frac{1}{c} S(dh,h';c) \mathcal{E}_H(m,n,h;c).$$

We only consider $\tilde{D}_H^{+,+}$ (the term corresponding to $h,h' > 0$), the other three terms being treated in the same way. We proceed by separating the variables h and c by means of Mellin transforms: we have

$$\mathcal{E}_H\left(m,n,h;\frac{4\pi\sqrt{dhh'}}{r}\right) = \frac{1}{2\pi i} \int_{(2)} \varphi_H(m,n;s;r) h^{-s} ds,$$

where

$$\varphi_H(m,n;s;r) := \int_0^{+\infty} \mathcal{E}_H\left(m,n,x;\frac{4\pi\sqrt{dxh'}}{r}\right) x^s \frac{dx}{x}$$

is a smooth function of r compactly supported in the interval $\left(2\frac{\sqrt{dHh'}}{C}, 36\frac{\sqrt{dHh'}}{C}\right)$. Hence taking $r = \frac{4\pi\sqrt{dhh'}}{c}$, we have

$$\tilde{D}_H^{+}(m,n) = \frac{1}{2\pi i} \int_{(2)} \sum_{h \geqslant 1} \frac{\chi(h)}{h^s} \sum_{c \equiv 0\,(N)} \frac{S(dh,h';c)}{c} \varphi_H\left(m,n;s,\frac{4\pi\sqrt{dhh'}}{c}\right) ds.$$

We are now in a position to apply the Kuznetsov trace formula (2.21) to the innermost c-sum. We obtain a sum of 3 terms,

$$\tilde{D}_H^{+}(m,n) = \frac{1}{2\pi i} \int_{(2)} T_H^{\mathrm{Holo}}(m,n;s)\, ds + \frac{1}{2\pi i} \int_{(2)} T_H^{\mathrm{Maass}}(m,n;s)\, ds + \frac{1}{2\pi i} \int_{(2)} T_H^{\mathrm{Eisen}}(m,n;s)\, ds, \tag{5.3}$$

where

$$T_H^{\mathrm{Holo}}(m,n;s) := 4\sum_{k\equiv 0\,(2)} \tilde\varphi_H(m,n;s;*)(k-1)\Gamma(k)\sum_{f\in\mathcal{B}_k^h(N)}\sqrt{h'}\rho_f(h')L(f\otimes\chi,s;d),$$

$$T_H^{\mathrm{Maass}}(m,n;s) := 4\sum_{j\geqslant 1}\frac{\hat\varphi_H(m,n;s;*)(t_j)}{\mathrm{ch}(\pi t_j)}\sqrt{h'}\rho_j(h')L(u_j\otimes\chi,s;d),$$

$$T_H^{\mathrm{Eisen}}(m,n;s) := \frac{1}{\pi}\sum_{\mathfrak{a}}\int_{-\infty}^{+\infty}\frac{\hat\varphi_H(m,n;s;*)(t)}{\mathrm{ch}(\pi t)}\sqrt{h'}\rho_{\mathfrak{a}}(h',t)L\left(E_{\mathfrak{a}}\left(\tfrac12+it\right)\otimes\chi,s;d\right)\,dt,$$

and

$$L(f\otimes\chi,s;d) := \sum_{h\geqslant 1}\frac{\chi(h)\sqrt{dh}\,\overline{\rho_f}(dh)}{h^s}.$$

Our next step will consist of shifting the contours of integration in (5.3) to the left up to $\Re s=\frac12$ and of bounding the three integrand on these contours. For this we will need to bound the various twisted L-functions $L(f\otimes\chi,s;d)$ on the line $\Re s=\frac12$ and the various Bessel transforms $\tilde\varphi_H(m,n;s;*)(k-1)$, $\hat\varphi_H(m,n;s;*)(t)$ and $\check\varphi_H(m,n;s;*)(t)$. This will be done in the next two sections.

5.4. **Bounds for twisted L-functions.** In this section we seek nontrivial bounds for the Dirichlet series $L(f\otimes\chi,s;d)$ when $f(z)$ has trivial nebentypus and is either a holomorphic Hecke cusp form (i.e., $f\in\mathcal{B}_k^h(N)$) or a Hecke–Maass cusp form (i.e., $f=u_j\in\mathcal{B}_0(N)$) or an Eisenstein series $f(z)=E_{\mathfrak{a}}\left(z,\frac12+it\right)$.

5.4.1. *The case f cuspidal.* Denoting by $\tilde f$ the primitive (arithmetically normalized) cusp form (of level $N'|N$) underlying f, we have the further factorization

$$\begin{aligned}L(f\otimes\chi,s;d) &= \left(\sum_{h|(dN)^\infty}\frac{\chi(h)\sqrt{dh}\,\overline{\rho_f}(dh)}{h^s}\right)\left(\sum_{(n,dN)=1}\frac{\chi(n)\lambda_f(n)}{n^s}\right)\\ &= \left(\sum_{h|(dN)^\infty}\frac{\chi(h)\sqrt{dh}\,\overline{\rho_f}(dh)}{h^s}\right)\left(\prod_{p|dN}\left(1-\frac{\chi(p)\lambda_{\tilde f}(p)}{p^s}+\frac{\chi_0(p)}{p^{2s+1}}\right)\right)L(\tilde f.\chi,s),\end{aligned}$$

where χ_0 denotes the trivial character modulo N' and

$$L(\tilde f.\chi,s)=\sum_{n\geqslant 1}\frac{\chi(n)\lambda_{\tilde f}(n)}{n^s}$$

is the twisted L-function of $\tilde f$ by the character χ. In particular, we see by (2.29), (2.31) and Hypothesis $H_{\frac{7}{64}}$ that $L(f\otimes\chi,s;d)$ is holomorphic for $\Re s\geqslant\frac12$, and for $\Re s=\frac12$ one has

$$L(f\otimes\chi,s;d)\ll_\varepsilon (PN)^\varepsilon\left(\sum_{h|(dN)^\infty}\frac{|\sqrt{dh}\rho_f(dh)|}{h^{1/2}}\right)\left|L(\tilde f.\chi,s)\right|. \tag{5.4}$$

For $L(\tilde f.\chi,s)$, one has the convexity bound (cf. (2.2))

$$L(\tilde f.\chi,s)\ll_\varepsilon (|s|\mu_f Nq)^\varepsilon|s|^{1/2}\mu_f^{1/2}[N,q^2]^{1/4}.$$

For large q, we obtained in [BHM05a], by combining the techniques of [DFI93] and the spectral large sieve of [DI82], the following improvement.

Theorem 5. *Assume Hypothesis H_θ. Let f be a primitive cusp form of level N (holomorphic or a weight 0 Maass form), and let χ be a primitive character of modulus q. For $\Re s=\frac12$ and for any $\varepsilon>0$, we have*

$$L(f.\chi,s)\ll(|s|\mu_f Nq)^\varepsilon|s|^E\mu_f^B N^A q^{1/2-\delta_{\mathrm{tw}}}.$$

The implied constant depends at most on ε, and an admissible choice for the exponents is

$$\delta_{\rm tw} := \frac{1-2\theta}{8}, \qquad A := \frac{13}{16}, \qquad B := \frac{79+12\theta}{16}, \qquad E := \frac{31+4\theta}{16}.$$

Remark 5.2. This result was established previously with the following smaller exponents $\delta_{\rm tw}$ (and some positive A, B, E):

$$\delta_{\rm tw} = \frac{1}{54} \ \text{[H03a, H03b]}, \qquad \delta_{\rm tw} = \frac{1}{22} \ \text{[M04a]}, \qquad \delta_{\rm tw} = \frac{1-2\theta}{10+4\theta} \ \text{[Bl04]}.$$

5.4.2. *The case f Eisenstein.* When $f(z)$ is of the form $E_{\mathfrak{a}}\big(z, \frac{1}{2}+it\big)$, the computations of [M04a] show that bounds for $L(f\otimes\chi, s; d)$ are reduced to bounds for products of Dirichlet L-functions. More precisely, we recall (see [DI82, Lemma 2.3]) that the cusps $\{\mathfrak{a}\}$ of $\Gamma_0(N)$ are uniquely represented by the rationals

$$\left\{\frac{u}{w}: \quad w|N, \quad u \in \mathcal{U}_w\right\},$$

where, for each $w|N$, $\mathcal{U}_w$ is a set of integers coprime with w representing each reduced residue class modulo $(w, N/w)$ exactly once, and in the half-plane $\Im t < 0$ we have for $h \neq 0$ (see [DI82, (1.17) and p.247]),

$$\sqrt{|dh|}\rho_{\mathfrak{a}}(dh,t) = \frac{\pi^{\frac{1}{2}+it}|dh|^{it}}{\Gamma\big(\frac{1}{2}+it\big)}\left(\frac{(w,N/w)}{wN}\right)^{\frac{1}{2}+it} \sum_{(\gamma,N/w)=1}\frac{1}{\gamma^{1+2it}} \sum_{\substack{\delta(\gamma w),\ (\delta,\gamma w)=1\\ \delta\gamma\equiv u \bmod (w,N/w)}} e\left(-dh\frac{\delta}{\gamma w}\right)$$

with analytic continuation to $\Im t = 0$. The congruence condition on δ can be analyzed by means of multiplicative characters modulo $(w, N/w)$:

$$\sum_{(\gamma,N/w)=1}\frac{1}{\gamma^{1+2it}} \sum_{\substack{\delta(\gamma w),\ (\delta,\gamma w)=1\\ \delta\gamma\equiv u \bmod (w,N/w)}} e\left(-dh\frac{\delta}{\gamma w}\right) = \frac{1}{\varphi((w,N/w))}\sum_{\psi \bmod (w,N/w)}\overline{\psi}(-u)\sum_{(\gamma,N/w)=1}\frac{\psi(\gamma)}{\gamma^{1+2it}}G_\psi(dh;\gamma w).$$

For each character ψ mod $(w, N/w)$, we denote by w^* its conductor and decompose w as

$$w = w^*w'w'', \qquad w'|(w^*)^\infty, \qquad (w'', w^*) = 1.$$

Accordingly, the Gauss sum factors as

$$G_\psi(dh;\gamma w) = \psi(\gamma w'')G_\psi(dh;w^*w')r(dh;\gamma w'') = \delta_{w'|dh}w'\psi(\gamma w'')G_\psi(dh/w';w^*)r(dh;\gamma w'').$$

Hence the inner sum on the right hand side equals

$$\sum_{(\gamma,N/w)=1}\frac{\psi(\gamma)}{\gamma^{1+2it}}G_\psi(dh;\gamma w) = \frac{\delta_{w'|dh}w'\overline{\psi}(dh/w')\psi(w'')G_\psi(1;w^*)}{L^{(N)}(\psi^2,1+2it)}\left(\sum_{\substack{\gamma|N^\infty\\(\gamma,N/w)=1}}\frac{\psi^2(\gamma)}{\gamma^{1+2it}}r(dh;\gamma w'')\right)\left(\sum_{\substack{a|dh\\(a,N)=1}}\frac{\psi^2(a)}{a^{2it}}\right),$$

where the superscript (N) indicates that the local factors at the primes dividing N have been removed. We deduce from here the inequality

$$\sqrt{|dh|}\rho_{\mathfrak{a}}(dh,t) \ll_\varepsilon (P(1+|t|))^\varepsilon \operatorname{ch}^{1/2}(\pi t)\frac{(dh,w)(w,N/w)}{(wN)^{1/2}}$$

$$\ll_\varepsilon (P(1+|t|))^\varepsilon \operatorname{ch}^{1/2}(\pi t)(dh,N)^{1/2}, \tag{5.5}$$

and also the identity

$$
\begin{aligned}
L(f\otimes\chi,s;d) &= \frac{\pi^{\frac12+it}d^{it}}{\Gamma\left(\frac12+it\right)}\left(\frac{(w,N/w)}{wN}\right)^{\frac12+it}\\
&\quad\times\frac{1}{\varphi((w,N/w))}\sum_{\psi \bmod (w,N/w)}\frac{w'G_\psi(1;w^*)\overline{\psi}(-ud/(d,w'))\psi(w'')}{(w'/(d,w'))^{s-it}L^{(N)}(\psi^2,1+2it)}\chi\left(\frac{w'}{(d,w')}\right)\\
&\qquad\times\sum_{h\geqslant 1}\frac{\chi(h)\overline{\psi}(h)}{h^{s-it}}\left(\sum_{\substack{\gamma|N^\infty\\(\gamma,N/w)=1}}\frac{\psi^2(\gamma)}{\gamma^{1+2it}}r\left(\frac{dhw'}{(d,w')};\gamma w''\right)\right)\left(\sum_{\substack{a|\frac{dh}{(d,w')}\\(a,N)=1}}\frac{\psi^2(a)}{a^{2it}}\right).
\end{aligned}
$$

Now the h-sum factors as

$$
\begin{aligned}
\left(\sum_{(h,dN)=1}\cdots\right)\left(\sum_{h|(dN)^\infty}\cdots\right) &= L^{(dN)}(\chi\overline{\psi},s-it)L^{(dN)}(\chi\psi,s+it)\\
&\quad\times\sum_{h|(dN)^\infty}\frac{\chi(h)\overline{\psi}(h)}{h^{s-it}}\left(\sum_{\substack{\gamma|N^\infty\\(\gamma,N/w)=1}}\frac{\psi^2(\gamma)}{\gamma^{1+2it}}r\left(\frac{dh}{(d,w')};\gamma w''\right)\right)\left(\sum_{\substack{a|\frac{dh}{(d,w')}\\(a,N)=1}}\frac{\psi^2(a)}{a^{2it}}\right).
\end{aligned}
$$

We can see that the second factor is holomorphic for $\Re s>0$ and is bounded, for any $\varepsilon>0$, by $\ll_\varepsilon (dN)^\varepsilon(d,w'')(w'')^{1-\Re s}$ in this domain. Hence $L(f\otimes\chi,s;d)$ has meromorphic continuation to the half-plane $\{s,\ \Re s>0\}$ with the only possible poles at $s=1\pm it$. The latter poles occur only if q divides $(w,N/w)$.

By Burgess' bound

$$L(\chi\overline{\psi},s-it)L(\chi\psi,s+it)\ll_\varepsilon(|s|+|t|)(qw^*)^{1/2-1/8+\varepsilon},$$

we infer that for $\Re s=\frac12$,

$$
\begin{aligned}
L\left(E_{\mathfrak a}\left(z,\tfrac12+it\right)\otimes\chi,s;d\right) &\ll_\varepsilon((1+|t|)Nq)^\varepsilon\operatorname{ch}^{1/2}(\pi t)(|s|+|t|)\frac{(w,N/w)^{1-1/8}(d,w)}{N^{\frac12}}q^{1/2-1/8}\\
&\ll_\varepsilon((1+|t|)Nq)^\varepsilon\operatorname{ch}^{1/2}(\pi t)(|s|+|t|)(d,N)q^{1/2-1/8}.
\end{aligned}
$$

Remark 5.3. In the special case where $q|(w,N/w)$, the residues of $L\left(E_{\mathfrak a}\left(z,\frac12+it\right)\otimes\chi,s;d\right)$ at $s=1\pm it$ ($t\neq 0$) are bounded by

$$
\begin{aligned}
\operatorname{res}_{s=1\pm it}L\left(E_{\mathfrak a}\left(z,\tfrac12+it\right)\otimes\chi,s;d\right) &\ll_\varepsilon((1+|t|)Nq)^\varepsilon\operatorname{ch}^{1/2}(\pi t)\frac{(d,w)(w,N/w)}{(wN)^{1/2}}\\
&\ll_\varepsilon((1+|t|)Nq)^\varepsilon\operatorname{ch}^{1/2}(\pi t)(d,N)^{1/2},
\end{aligned}
\tag{5.6}
$$

and the same bound holds for $\operatorname{res}_{s=1}(s-1)L\left(E_{\mathfrak a}\left(z,\frac12+it\right)\otimes\chi,s;d\right)$ if $t=0$.

5.5. **Bounds for $\varphi_H(m,n;s;r)$ and its Bessel transforms.** We also need to exhibit bounds for $\tilde\varphi_H(m,n;s;*)(k-1)$, $\hat\varphi_H(m,n;s;*)(t)$ and $\check\varphi_H(m,n;s;*)(t)$. For this purpose, we first record an estimate for φ_H and its partial derivatives. Using (5.2) and several integrations by parts, we see that for any $j,k\geqslant 0$ and $\Re s\geqslant-\frac12$,

$$r^k\frac{\partial^k}{\partial r^k}\varphi_H(m,n;s;r)\ll_{j,k,\varepsilon}P^\varepsilon(P^\varepsilon\mu_g^3Z)^{j+k}|s|^{-j}\Xi(m,n)H^{\Re s},\tag{5.7}$$

where $\Xi(m,n)$ is defined in (4.13); moreover, as a function of r, $\varphi_H(m,n;s;r)$ is supported on

$$\left(2\frac{\sqrt{dHh'}}{C}, 36\frac{\sqrt{dHh'}}{C}\right) = (R, 18R),$$

say. We will apply these bounds in conjunction with the following lemma; as this lemma is an immediate generalization of Lemma 7.1 of [DI82], we do not reproduce the proof.

Lemma 5.1. *Let $\varphi(r)$ be a smooth function, compactly supported in $(R, 18R)$, satisfying*

$$\varphi^{(i)}(r) \ll (W/R)^i$$

for some $W \geqslant 1$ and for any integer $i \geqslant 0$, the implied constant depending on i. Then, for $t \geqslant 0$ and for any $k > 1$, one has

$$\hat{\varphi}(it),\ \check{\varphi}(it) \ll \frac{1 + (R/W)^{-2t}}{1 + R/W} \qquad \text{for } 0 \leqslant t < \frac{1}{4}; \tag{5.8}$$

$$\hat{\varphi}(t),\ \check{\varphi}(t),\ \tilde{\varphi}(t) \ll \frac{1 + |\log(R/W)|}{1 + R/W} \qquad \text{for } t \geqslant 0; \tag{5.9}$$

$$\hat{\varphi}(t),\ \check{\varphi}(t),\ \tilde{\varphi}(t) \ll \left(\frac{W}{t}\right)\left(\frac{1}{t^{1/2}} + \frac{R}{t}\right) \qquad \text{for } t \geqslant 1; \tag{5.10}$$

$$\hat{\varphi}(t),\ \check{\varphi}(t),\ \tilde{\varphi}(t) \ll_k \left(\frac{W}{t}\right)^k \left(\frac{1}{t^{1/2}} + \frac{R}{t}\right) \qquad \text{for } t \geqslant \max(10R, 1). \tag{5.11}$$

Proof. The inequalities (5.8), (5.9), (5.10) can be proved exactly as (7.1), (7.2) and (7.3) in [DI82]. The last inequality (5.11) is an extension of (7.4) in [DI82], but we only claim it in the restricted range $t \geqslant \max(10R, 1)$. On the one hand, we were unable to reconstruct the proof of (7.4) in [DI82] for the entire range $t \geqslant 1$; on the other hand, [DI82] only utilizes this inequality for $t \gg \max(R, W)$ (cf. page 268 there, and note also that for $t \ll W$ the bound (5.10) is stronger). For this reason we include a detailed proof of (5.11) in the case of $\check{\varphi}(t)$. For $\hat{\varphi}(t)$ and $\tilde{\varphi}(t)$ the proof is very similar.

We may assume that $k = 2j + 1$ is a positive odd integer. The Bessel differential equation

$$r^2 K_{2it}''(r) + r K_{2it}'(r) = (r^2 - 4t^2) K_{2it}(r)$$

gives an identity

$$\check{\varphi}(t) = (D_t \varphi)^\vee(t), \tag{5.12}$$

where

$$D_t \varphi(r) := r \left(\frac{r\varphi(r)}{r^2 - 4t^2}\right)'' + r\left(\frac{\varphi(r)}{r^2 - 4t^2}\right)'.$$

This transform $D_t\varphi$ is smooth and compactly supported in $(R, 18R)$, and it is straightforward to check that

$$\|(D_t\varphi)^{(i)}\|_\infty \ll_i (W/t)^2 (W/R)^i \quad \text{for} \quad t \geqslant \max(10R, 1).$$

By iterating (5.12) it follows that

$$\check{\varphi}(t) = (D_t^j \varphi)^\vee(t),$$

where $D_t^j\varphi$ is a smooth function, compactly supported in $(R, 18R)$, satisfying

$$\|(D_t^j\varphi)^{(i)}\|_\infty \ll_{j,i} (W/t)^{2j} (W/R)^i \quad \text{for} \quad t \geqslant \max(10R, 1).$$

We bound $(D_t^j\varphi)^\vee(t)$ by (5.10) and obtain

$$\check{\varphi}(t) \ll_j \left(\frac{W}{t}\right)^{2j+1} \left(\frac{1}{t^{1/2}} + \frac{R}{t}\right) \quad \text{for} \quad t \geqslant \max(10R, 1).$$

$\square$

5.6. **Putting it all together.** We now use the results of the preceding two sections to conclude the proof of Theorem 4. We start by estimating the contribution of the Maass spectrum to $\tilde{D}_H^{+,+}$:

$$\frac{1}{L}\sum_{\substack{m,n\geqslant 1\\ h'>0}}\overline{\lambda_g}(m)\lambda_g(n)\frac{1}{2\pi i}\int_{(2)}T_H^{\mathrm{Maass}}(m,n;s)\,ds=\frac{1}{2\pi i}\int_{(1/2)}\frac{1}{L}\sum_{\substack{m,n\geqslant 1\\ h'>0}}\overline{\lambda_g}(m)\lambda_g(n)T_H^{\mathrm{Maass}}(m,n;s)\,ds.$$

With some $T_0\geqslant\max(10R,1)$ to be determined later, we further decompose $T_H^{\mathrm{Maass}}(m,n;s)$ as

$$T_H^{\mathrm{Maass}}(m,n;s)=T_{H,\leqslant T_0}^{\mathrm{Maass}}(m,n;s)+T_{H,>T_0}^{\mathrm{Maass}}(m,n;s),$$

corresponding to the contributions of the eigenforms $u_j\in\mathcal{B}_0(N)$ such that $|t_j|\leqslant T_0$ and $|t_j|>T_0$, respectively (observe that the first portion contains the exceptional spectrum whenever it exists).

Setting $W:=P^\varepsilon\mu_g^3Z$, we can apply (5.8) and (5.9) to $\varphi=\varphi_H(m,n;s;*)$ in the light of (5.7). Using also (5.4) and Theorem 5, we obtain, for any $j\geqslant 0$,

$$\begin{aligned}T_{H,\leqslant T_0}^{\mathrm{Maass}}(m,n;s)\ll_{j,\varepsilon}(PT_0)^\varepsilon\frac{W^j}{|s|^{j-E-\varepsilon}}\Xi(m,n)(\ell_1\ell_2)^AH^{1/2}q^{1/2-\delta_{\mathrm{tw}}}&\\ \times\sum_{|t_i|\leqslant T_0}\frac{|\sqrt{h'}\rho_i(h')|}{\mathrm{ch}(\pi t_i)}\left(\sum_{h|(dN)^\infty}\frac{|\sqrt{dh}\rho_i(dh)|}{h^{1/2}}\right)\frac{1+|\log(R/W)|+(R/W)^{-2|\Im t_i|}}{1+R/W}T_0^B&.\end{aligned}$$

By several applications of the Cauchy–Schwarz inequality and the bound (2.29), we can see that

$$\sum_{|t_i|\leqslant T_0}\frac{|\sqrt{h'}\rho_i(h')|}{\mathrm{ch}(\pi t_i)}\left(\sum_{h|(dN)^\infty}\frac{|\sqrt{dh}\rho_i(dh)|}{h^{1/2}}\right)\ll_\varepsilon(mnPT_0)^\varepsilon(h'd)^\theta T_0^2. \tag{5.13}$$

In addition, since $H\leqslant 8Y/d$ and $R=2\sqrt{dHh'}/C$, we have

$$\frac{1+|\log(R/W)|+(R/W)^{-2|\Im t_i|}}{1+R/W}H^{1/2}\ll_\varepsilon P^\varepsilon\left(\frac{W^2C^2}{h'Y}\right)^\theta\left(\frac{Y}{d}\right)^{1/2}.$$

Hence by summing over m,n and using (2.28) and (4.13), we find that

$$\begin{aligned}\sum_{\substack{\ell_1m\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/X)\\ \ell_2n\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/Y)}}\left|\overline{\lambda_g}(m)\lambda_g(n)T_{H,\leqslant T_0}^{\mathrm{Maass}}(m,n;s)\right|\ll_{j,g,\varepsilon}&\\ (PT_0)^{5\varepsilon}Z^{3+2\theta}\frac{W^j}{|s|^{j-E-\varepsilon}}(\ell_1\ell_2)^Ad^\theta\frac{C^3}{\ell_1\ell_2}\left(\frac{C^2}{Y}\right)^\theta\left(\frac{Y}{d}\right)^{1/2}q^{1/2-\delta_{\mathrm{tw}}}T_0^{B+2}&.\end{aligned}$$

For $T_{H,>T_0}^{\mathrm{Maass}}(m,n;s)$, we use (5.11), (5.4) and Theorem 5: we obtain, for any $j\geqslant 0$ and any $k>1$,

$$\begin{aligned}T_{H,>T_0}^{\mathrm{Maass}}(m,n;s)\ll_{j,\varepsilon}P^\varepsilon\frac{W^j}{|s|^{j-E-\varepsilon}}\Xi(m,n)(\ell_1\ell_2)^AH^{1/2}q^{1/2-\delta_{\mathrm{tw}}}&\\ \times\sum_{|t_i|>T_0}\frac{|\sqrt{h'}\rho_i(h')|}{\mathrm{ch}(\pi t_i)}\left(\sum_{h|(dN)^\infty}\frac{|\sqrt{dh}\rho_i(dh)|}{h^{1/2}}\right)|t_i|^{B+\varepsilon}\left(\frac{W}{t_i}\right)^k\left(\frac{1}{t_i^{1/2}}+\frac{R}{t_i}\right)&.\end{aligned}$$

We take $k>3/2+B+\varepsilon$ to ensure the convergence of the sum over the $\{u_j\}$, and then we sum over m,n using (2.28) and (4.13). As before, we may restrict ourselves to some range

$$\ell_1m\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/X)\qquad\text{and}\qquad\ell_2n\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/Y),$$

the remaining contribution being negligible. In this range

$$h'\ll_{g,\varepsilon}P^\varepsilon Z^2(C^2/X)\qquad\text{and}\qquad R\ll_{g,\varepsilon}P^\varepsilon Z(Y/X)^{1/2},$$

therefore we obtain, using also (5.13),

$$\sum_{\substack{\ell_1 m\ll_{g,\varepsilon} P^\varepsilon Z^2(C^2/X)\\ \ell_2 n\ll_{g,\varepsilon} P^\varepsilon Z^2(C^2/Y)}} \left|\overline{\lambda_g}(m)\lambda_g(n)T^{\mathrm{Maass}}_{H,>T_0}(m,n;s)\right| \ll_{j,k,g,\varepsilon} (PT_0)^{5\varepsilon}Z^{3+2\theta}$$
$$\times \frac{W^j}{|s|^{j-E-\varepsilon}}(\ell_1\ell_2)^A d^\theta \frac{C^3}{\ell_1\ell_2}\left(\frac{C^2}{X}\right)^\theta\left(\frac{Y}{d}\right)^{1/2} q^{1/2-\delta_{\mathrm{tw}}}\left(\frac{W}{T_0}\right)^k T_0^{B+2}\left(\frac{1}{T_0^{1/2}}+\frac{Z(Y/X)^{1/2}}{T_0}\right).$$

Summing up and using also (4.5), we infer that

$$\frac{1}{L}\sum_{\substack{m,n\geqslant 1\\ h'>0}} \overline{\lambda_g}(m)\lambda_g(n)T^{\mathrm{Maass}}_H(m,n;s) \ll_{j,k,g,\varepsilon} (PT_0)^{6\varepsilon}Z^{3+2\theta}\frac{W^j}{|s|^{j-E-\varepsilon}}(\ell_1\ell_2)^A d^\theta$$
$$\times C\left(\frac{C^2}{Y}\right)^\theta\left(\frac{Y}{d}\right)^{1/2} q^{1/2-\delta_{\mathrm{tw}}}T_0^{B+2}\left\{1+\left(\frac{W}{T_0}\right)^k\left(\frac{(Y/X)^\theta}{T_0^{1/2}}+\frac{Z(Y/X)^{1/2+\theta}}{T_0}\right)\right\}.$$

Upon choosing

$$T_0 := \max(10R, WY^{1/k}) \ll_{g,\varepsilon} WY^{1/k}(Y/X)^{1/2}$$

and taking k very large (in terms of ϵ), the above becomes

$$\ll_{j,g,\varepsilon} P^{13\varepsilon}Z^{3+2\theta}\frac{W^{j+B+2}}{|s|^{j-E-\varepsilon}}(\ell_1\ell_2)^A d^{\theta-1/2}q^{1/2-\delta_{\mathrm{tw}}}(Y/X)^{(B+2)/2}Y^{1/2-\theta}C^{1+2\theta}.$$

We use this bound with $j>1+E+\varepsilon$ (to ensure convergence in the s-integral), and integrate over s. In this way we obtain that the contribution of the Maass spectrum to $\tilde{D}^{+,+}$ is bounded by

$$\ll_{g,\varepsilon} P^{14\varepsilon}Z^{3+2\theta}W^{E+1+B+2}(\ell_1\ell_2)^A d^{\theta-1/2}q^{1/2-\delta_{\mathrm{tw}}}(Y/X)^{(B+2)/2}Y^{1/2-\theta}C^{1+2\theta},$$

hence by (4.3) the global contribution of the Maass spectrum to $\Sigma_\chi(\ell_1,\ell_2;\mathbf{c})$ is bounded by (remember that we have reused the letter d in place of $q'd$)

$$\ll_{g,\varepsilon} P^{24\varepsilon}Z^{13+3\theta}(\ell_1\ell_2)\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\mathrm{tw}}}(Y/X)^4Y^{1/2-\theta}C^{1+2\theta}. \tag{5.14}$$

Similar arguments (using also (2.31) and (5.11) for $\tilde{\varphi}$) show that the same bound holds for the holomorphic and the Eisenstein spectrum (in fact in a stronger form). For the Eisenstein portion, however, an additional term might occur, coming from the poles of $L\left(E_{\mathfrak{a}}(z,\frac12+it)\otimes\chi,s\right)$ at $s=1\pm it$. This additional term occurs only if $q|(w,N/w)$ for some $w|N$ (in particular $q\leqslant N^{1/2}=(D\ell_1\ell_2)^{1/2}$) and (by (5.5), (5.6), (5.9), and (5.7) with $j=1+\delta$ for $\delta>0$ small) contributes to $\tilde{D}^{+,+}_{>H}(m,n)$ at most

$$\ll_{g,\varepsilon} P^{2\varepsilon}W\Xi(m,n)(d,\ell_1\ell_2)^{1/2}(h',\ell_1\ell_2)^{1/2}\frac{Y}{d},$$

and the contribution of these residues to $\Sigma_\chi(\ell_1,\ell_2;\mathbf{c})$ is bounded by

$$\ll_{g,\varepsilon} P^{3\varepsilon}WZ^3(\mathbf{c},\ell_1\ell_2)^{1/2}q^{1/2}YC \ll_g P^{4\varepsilon}Z^4(\mathbf{c},\ell_1\ell_2)^{1/2}(\ell_1\ell_2)^{1/2}YC. \tag{5.15}$$

Collecting all the previous estimates, we obtain that $\Sigma_\chi(\ell_1,\ell_2;\mathbf{c})$ is bounded by the sum of the terms in (4.7), (5.1), (5.14), plus the additional term (5.15) if $q|(w,N/w)$ for some $w|N$. To conclude, we discuss now the choice of the parameter C.

A comparison of (5.14) with (4.7) suggests the choice

$$C_{\mathrm{opt}} := Z^{-\frac{11+3\theta}{2(1+\theta)}}(\ell_1\ell_2)^{-\frac{1}{4(1+\theta)}}\mathbf{c}^{-\frac{\theta}{2(1+\theta)}}q^{-\frac{1-2\theta-2\delta_{\mathrm{tw}}}{4(1+\theta)}}(X/Y)^{\frac{2}{1+\theta}}X^{\frac{1}{4(1+\theta)}}Y^{1/2} =: D_{\mathrm{opt}}Y^{1/2},$$

say. Clearly, $C_{\mathrm{opt}}\leqslant Y$ and the condition $C_{\mathrm{opt}}\geqslant Y^{1/2}$ is fulfilled if and only if

$$X \geqslant X_{\mathrm{opt}} := Z^{\frac{22+6\theta}{9}}(l_1l_2)^{\frac19}\mathbf{c}^{\frac{2\theta}{9}}q^{\frac{1-2\theta-2\delta_{\mathrm{tw}}}{9}}Y^{\frac89}. \tag{5.16}$$

Under this condition it follows from $Y \geqslant X$, $\mathbf{c} \geqslant q$ and $\delta_{\rm tw} \leqslant \frac{1}{8}$ that

$$q^{3/16}\frac{X^{3/4}Y^{3/4}}{C_{\rm opt}^{1/2}} \leqslant \frac{X^{1/2}Y^{3/2}}{C_{\rm opt}},$$

so that (5.1) is bounded by (4.7) (when $P^{2\varepsilon}$ is replaced by $P^{5\varepsilon}$). Therefore, we obtain Theorem 4 when (5.16) is satisfied (cf. (5.14)):

$$\Sigma_\chi(\ell_1,\ell_2;\mathbf{c}) \ll_{g,\varepsilon} P^{24\varepsilon}Z^{13+3\theta}(\ell_1\ell_2)\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\rm tw}}(Y/X)^4 Y D_{\rm opt}^{1+2\theta},$$

plus the additional term (5.15), if $q|(w,N/w)$ for some $w|N$, which equals

$$P^{4\varepsilon}Z^4(\mathbf{c},\ell_1\ell_2)^{1/2}(\ell_1\ell_2)^{1/2}YC_{\rm opt} = P^{4\varepsilon}Z^4(\mathbf{c},\ell_1\ell_2)^{1/2}(\ell_1\ell_2)^{1/2}Y^{3/2}D_{\rm opt}.$$

If (5.16) is not satisfied (i.e., $X \leqslant X_{\rm opt}$, hence $D_{\rm opt} \leqslant 1$), we choose $C = Y^{1/2} = Y^{1/2}\max(1, D_{\rm opt})$. We see that (4.7) is bounded by (5.14) whose value is given by

$$\ll_{g,\varepsilon} P^{24\varepsilon}Z^{13+3\theta}(\ell_1\ell_2)\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\rm tw}}(Y/X)^4Y.$$

The diagonal contribution (5.1) is bounded by

$$\ll_{g,\varepsilon} P^{5\varepsilon}Z\frac{(\mathbf{c},\ell_1\ell_2)^{1/2}}{[\ell_1,\ell_2]^{1-\theta}}\mathbf{c}^{1/2}q^{3/16}X^{3/4}Y^{1/2} \leqslant P^{5\varepsilon}Z(\ell_1\ell_2)^{\theta}\mathbf{c}^{1/2}q^{3/16}X^{1/4}(X/Y)^{1/2}Y.$$

Translating $X \leqslant X_{\rm opt}$ into

$$X(X/Y)^8 \leqslant Z^{22+6\theta}(\ell_1\ell_2)\mathbf{c}^{2\theta}q^{1-2\theta-2\delta_{\rm tw}},$$

and using also $\mathbf{c} \geqslant q$ and $\delta_{\rm tw} \leqslant \frac{1}{8}$, we can see that

$$q^{3/16}X^{1/4}(X/Y)^2 \leqslant Z^{(11+3\theta)/2}(\ell_1\ell_2)^{1/4}\mathbf{c}^{\theta}q^{1/2-\theta-\delta_{\rm tw}}.$$

It follows that (5.1) is bounded by

$$\ll_{g,\varepsilon} P^{5\varepsilon}Z^{(13+3\theta)/2}(\ell_1\ell_2)^{1/4+\theta}\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\rm tw}}(Y/X)^{3/2}Y.$$

In particular, if (5.16) is not satisfied, then (4.7), (5.1) and (5.14) are all bounded by

$$P^{24\varepsilon}Z^{13+3\theta}(\ell_1\ell_2)\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\rm tw}}(Y/X)^4Y.$$

Finally, if $q|(w,N/w)$ for some $w|N$, the additional term (5.15) equals

$$P^{4\varepsilon}Z^4(\mathbf{c},\ell_1\ell_2)^{1/2}(\ell_1\ell_2)^{1/2}Y^{3/2}.$$

This concludes the proof of Theorem 4.

6. Heegner points on Shimura curves

In this section, we establish equidistribution results for incomplete orbits of Heegner points on Shimura curves associated with indefinite quaternion algebras over $\mathbf{Q}$. We consider only the case of compact Shimura curves (corresponding to non-split quaternion algebras) and give afterwards the necessary modification needed to deal with the case of modular curves $X_0(q)$.

6.1. **Indefinite quaternion algebras.** For more background on Shimura curves and Heegner points, we refer to [BD96, Da04, Z02, Z04].

Let $q^- > 1$ be a square-free integer with an even number of prime factors, and let B be the (indefinite) quaternion algebra over $\mathbf{Q}$ ramified exactly at the primes dividing q^-; for each place v of $\mathbf{Q}$, $B_v := B(\mathbf{Q}_v)$ is either the unique quaternion algebra over $\mathbf{Q}_v$ if v divides q^-, or is isomorphic to $M_2(\mathbf{Q}_v)$; for such v (in particular, for $v=\infty$), we fix an isomorphism $\phi_v : B(\mathbf{Q}_v) \to M_2(\mathbf{Q}_v)$. We set

$$\mathcal{O} := B \cap \prod_{v\neq\infty}\mathcal{O}_v,$$

where $\mathcal{O}_v$ is the unique maximal order of B_v if $v|q^-$ or $\mathcal{O}_v = \phi_v^{-1}(M_2(\mathbf{Z}_v))$; this is a maximal order of B. More generally, for q^+ coprime with q^- and for v finite, we consider the set of matrices

$$M_{2,v}(q^+) := \left\{ g \in M_2(\mathbf{Z}_v) :\ g \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{q^+} \right\},$$

and we denote by $\mathcal{O}_{q^+} \subset \mathcal{O}$ the sub-order of conductor q^+:

$$\mathcal{O}_{q^+} := \left\{ x \in \mathcal{O} :\ \phi_v(x) \in M_{2,v}(q^+),\ v|q^+ \right\}.$$

In the sequel, we put $q := q^+ q^-$.

The group $\Gamma_0(q^-, q^+) := \phi_\infty(\mathcal{O}_{q^+}^\times)$ is a discrete cofinite (in fact, cocompact, since $q^- > 1$) subgroup of $\mathrm{GL}_2(\mathbf{R})$, and one associates to it the complex algebraic (Shimura) curve[6] $X_0(q^-, q^+)$ with complex uniformization[7]

$$X_0(q^-, q^+)(\mathbf{C}) := \left[\mathcal{H}^+ \cup \mathcal{H}^-\right] / \Gamma_0(q^-, q^+),$$

where $\mathcal{H}^+ \cup \mathcal{H}^-$ denotes the union of the upper and lower half-planes equipped with the standard $\mathrm{GL}_2(\mathbf{R})$ action; there is a natural identification of this space with $\mathrm{Hom}(\mathbf{C}, B_\infty)$, $\mathrm{GL}_2(\mathbf{R})$ acting by conjugation, and under this identification one has

$$X_0(q^-, q^+)(\mathbf{C}) \simeq \mathrm{Hom}(\mathbf{C}, B_\infty)/\phi_\infty(\mathcal{O}_{q^+}^\times). \tag{6.1}$$

We denote by $\mathcal{L}(q^-, q^+)$ the L^2-space of functions on $X_0(q^-, q^+)(\mathbf{C})$ equipped with the standard (Petersson) inner product induced by the Poincaré measure $\frac{dxdy}{y^2}$ on $\mathcal{H}^+$). As for the case of the modular curve ($q^- = 1$), this space is generated by eigenfunctions of the automorphic Laplacian (Maass forms): it has a purely discrete spectrum. As in the case of modular curves, $X_0(q^-, q^+)$ is also endowed with an algebra of Hecke correspondences $\mathbf{T}^{(q)}$ defined over $\mathbf{Q}$ and generated by Hecke operators T_n ($(n, q) = 1$); there is a basis of $\mathcal{L}(q^-, q^+)$ consisting of Hecke–Maass eigenforms.

6.2. **Atkin–Lehner theory.** We need to describe more precisely the structure of such a Hecke–Maass eigenbasis. For $q'|q^+$, we have $\tau(q^+/q')$ degeneracy morphisms, defined over $\mathbf{Q}$, $(\nu_{q',d})_{d|q^+/q'}$ say,

$$\nu_{q',d} : X_0(q^-, q^+) \to X_0(q^-, q')$$

induced by the inclusions

$$\begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} M_{2,v}(q^+) \begin{pmatrix} d^{-1} & 0 \\ 0 & 1 \end{pmatrix} \subset M_{2,v}(q').$$

Atkin–Lehner theory can be extended to the general case of quaternion algebras: as usual, one says that an L^2-normalized Hecke eigenform on $X_0(q^-, q^+)$ is *new* (of level q^+) if it is orthogonal to all the functions of the form $\nu_{q',d}^* \psi$, where q' ranges over all the divisors of q^+ distinct from q^+, ψ ranges over the Hecke–Maass eigenforms on $X_0(q^-, q')$, and d ranges over the divisors of q^+/q'. If $q^- \neq 1$, one has the orthogonal decomposition

$$\mathcal{L}(q^-, q^+) = \bigoplus_{q'|q^+} \bigoplus_{g_{q'}} \mathbf{C}\langle \nu_{q',d}^* g_{q'},\ d|q^+/q' \rangle, \tag{6.2}$$

where $g_{q'}$ ranges over the set of L^2-normalized new forms on $X_0(q^-, q')$. Moreover, multiplicity one remains valid: by a result of Zhang [Z01a], the Hecke eigenspace of $\mathcal{L}(q^-, q^+)$ containing a new form (of level q^+) is one-dimensional, and, more generally, the dimension of $\mathbf{C}\langle \nu_{q',d}^* g_{q'},\ d|q^+/q' \rangle$ equals $\tau(q^+/q')$.

[6] In fact, by Shimura theory, $X_0(q^-, q^+)$ is defined over $\mathbf{Q}$.

[7] When $q^- = 1$, the quotient below needs to be compactified and yields the standard modular curve $X_0(q)$.

6.3. **Heegner points.** Fix an embedding of $\overline{\mathbf{Q}}$ into $\mathbf{C}$. Let $K \subset \overline{\mathbf{Q}}$ be an imaginary quadratic number field of discriminant $-D$ coprime with q, and denote by $\mathcal{O}_K$ its ring of integers. The *Heegner points* (of conductor 1) in $X_0(q^-,q^+)(\mathbf{C})$ are by definition the images under (6.1) of the conjugacy classes of optimal embeddings of $\mathcal{O}_K$ into $\mathcal{O}_{q^+}$ (i.e., the $\psi \in \mathrm{Hom}(K,B)$ such that $\psi(K)\cap\mathcal{O}_{q^+} = \psi(\mathcal{O}_K)$).

We denote by $H_{q^-,q^+}(K)$ the set of Heegner points. This set is non-empty if and only if

(Heegner) Every prime factor of q^- is inert in K and every prime factor of q^+ splits in K,

a condition which we always assume in the sequel. When non-empty, $H_{q^-,q^+}(K)$ is endowed with a natural free action of $\mathrm{Pic}(\mathcal{O}_K)$, the Picard group of $\mathcal{O}_K$; the orbits under this action are called *orientations*. There are $2^{\omega(q)}$ orbits each of size

$$\left|\mathrm{Pic}(\mathcal{O}_K)\right| \gg_\varepsilon D^{\frac{1}{2}-\varepsilon}$$

by Siegel's theorem. In fact, the Heegner points are defined over H_K, the Hilbert class field of K, and the Galois action of $G_K := \mathrm{Gal}(H_K/K)$ corresponds to that of $\mathrm{Pic}(\mathcal{O}_K) \simeq G_K$ via the Artin map.

The following theorem gives the equidistribution of orbits of Heegner points under subgroups of G_K of large index.

Theorem 6. *For any continuous function $g : X_0(q^-,q^+)(\mathbf{C}) \to \mathbf{C}$, there exists a bounded function $\varepsilon_g : \mathbf{R}^+ \to \mathbf{R}^+$, depending only on q^-,q^+,g, which satisfies*

$$\lim_{x\to 0}\varepsilon_g(x) = 0$$

such that: for any imaginary quadratic field K with discriminant $-D$ satisfying the Heegner condition, for any subgroup $G \subset G_K$. and for any Heegner point $z_0 \in H_{q^-,q^+}(K)$, one has

$$\left|\frac{1}{|G|}\sum_{\sigma\in G} g(z_0^\sigma) - \int_{X_0(q^-,q^+)} g(z)\,d\mu_{q^-,q^+}(z)\right| \leqslant \varepsilon_g\big([G_K:G]D^{-\frac{1}{5297}}\big),$$

where $d\mu_{q^-,q^+}(z)$ denotes the normalized hyperbolic measure on $X_0(q^-,q^+)(\mathbf{C})$:

$$d\mu_{q^-,q^+}(z) := \frac{1}{\mathrm{Vol}(X_0(q^-,q^+))}\frac{dxdy}{y^2} \qquad \text{with} \qquad \mathrm{Vol}(X_0(q^-,q^+)) := \int_{X_0(q^-,q^+)(\mathbf{C})}\frac{dxdy}{y^2}.$$

In particular, for any imaginary quadratic field $K \subset \mathbf{C}$ satisfying the Heegner condition and for any subgroup $G \subset G_K$ of index satisfying $[G_K : G] \leqslant |G_K|^{\frac{1}{5298}}$, the orbit $G.z_0$ becomes equidistributed on $X_0(q^-,q^+)$ relatively to $\mu_{q^-,q^+}(z)$ as $D \to +\infty$.

Proof. By Weyl's equidistribution criterion and the spectral decomposition (6.2), it is sufficient to prove that for any $q'|q^+$, any $d|q^+/q'$, and any primitive Maass form g on $X_0(q^-,q')$, one has

$$\frac{1}{|G|}\sum_{\sigma\in G}\nu_{q',d}^* g(z_0^\sigma) = o_g\big([G_K:G]D^{-\frac{1}{5297}}\big) \quad \text{as} \quad D\to+\infty.$$

Here

$$\nu_{q',d}^* g(z_0^\sigma) = g(\nu_{q',d}(z_0^\sigma)) = g((\nu_{q',d}z_0)^\sigma);$$

Indeed, one can check (e.g., by using the adelic formulation of $\nu_{q',d}$ and of the Heegner points given in [BD96]) that $\nu_{q',d}z_0$ is a Heegner point on $X_0(q^-,q')$ and then, since $\nu_{q',d}$ is defined over $\mathbf{Q}$, it commutes with the Galois action. This reduces us to the case where g is an L^2-normalized Maass newform of level q^+.

By Fourier analysis, one has

$$\frac{1}{|G|}\sum_{\sigma\in G} g(z_0^\sigma) = \frac{1}{|G_K|}\sum_{\substack{\psi\in\widehat{G}_K\\ \psi_{|G}\equiv 1}}\sum_{\sigma\in G_K}\psi(\sigma)g(z_0^\sigma).$$

We invoke now a deep formula of Zhang [Z01a, Z01b, Z04] which expresses the (square of) the innermost sum in terms of central values of Rankin–Selberg L-functions. More precisely, we denote by $\tilde{g}$ the (unique) arithmetically normalized primitive Maass form for $\Gamma_0(q)$, having the same eigenvalues as g for the Laplace and the Hecke operators T_n $(n,q)=1$ (which exists by the Jacquet–Langlands correspondence), and by $f_\psi(z)$ the theta series associated to the class group character ψ; the latter is a weight one holomorphic form of level D with nebentypus the Kronecker symbol χ_K. One has the following formula of Zhang (which generalizes formulae of Hecke, Maass, Gross–Zagier and others):

$$\Bigl|\sum_{\sigma\in G_K}\psi(\sigma)g(z_0^\sigma)\Bigr|^2=\frac{\sqrt{D}}{4\langle\tilde{g},\tilde{g}\rangle}L\bigl(f_\psi\otimes\tilde{g},\tfrac12\bigr).$$

If ψ factors through the norm $N_{K/\mathbf{Q}}$ (i.e., ψ is of order 2), f_ψ is an Eisenstein series; in this case,

$$L\bigl(f_\psi\otimes\tilde{g},\tfrac12\bigr)=L\bigl(\chi_1\otimes\tilde{g},\tfrac12\bigr)L\bigl(\chi_2\otimes\tilde{g},\tfrac12\bigr),$$

where χ_1, χ_2 are Dirichlet characters such that $\chi_1\chi_2=\chi_K$. Otherwise, f_ψ is a cusp form. Hence, either by Theorem 5 (in the former case) or by Theorem 1 (in the latter), we infer that for any $\varepsilon>0$,

$$\frac{1}{|G|}\sum_{\sigma\in G}g(z_0^\sigma)\ll_g\frac{[G_K:G]}{|G_K|}D^{\frac12-\frac{1}{5296}}\ll_{g,\varepsilon}D^{\varepsilon-\frac{1}{5296}}[G_K:G]=o_g\bigl(D^{-\frac{1}{5297}}[G_K:G]\bigr),$$

since $|G_K|\gg_\varepsilon D^{\frac12-\varepsilon}$ for any $\varepsilon>0$. □

Remark 6.1. As in Remark 1.4, the polynomial control in the remaining parameters for the subconvexity bound can be used to bound the discrepancy and obtain

$$D_{z_0}(G):=\sup_{B\subset X_0(q^-,q^+)}\left|\frac{1}{|G|}\sum_{\substack{\sigma\in G\\ z_0^\sigma\in B}}1-\frac{\mathrm{Vol}(B)}{\mathrm{Vol}(X_0(q^-,q^+))}\right|\ll_{q^+,q^-}[G_K:G]D^{-\frac{1}{5297}-\eta}$$

for some absolute constant $\eta>0$ (independent of q^+q^-).

6.4. **Heegner points on modular curves.** When $q^-=1$, $X_0(1,q^+)$ is the usual modular curve $X_0(q)$. We suppose for simplicity that q is square-free. In that case, Theorem 6 continues to hold with the exponent $\frac{1}{5297}$ replaced by $\frac{1}{23042}$, the reason for this weaker exponent coming from the existence of the continuous spectrum and a weaker subconvexity bound for the corresponding L-function. As the proof is similar, we merely sketch it. The only difference with the cocompact case is that one has to deal with the extra contribution coming from the Eisenstein spectrum. Indeed, the treatment of the discrete spectrum is identical to the one given above (except that the Jacquet–Langlands correspondence is the identity). For the Eisenstein spectrum, the same reasoning reduces the problem to the verification that, for any class group character ψ, the twisted Weyl sum satisfies

$$\sum_{\sigma\in G_K}\psi(\sigma)g(z_0^\sigma)\ll_g D^{\frac12-\frac{1}{23041}}\quad\text{as}\quad D\to+\infty,\tag{6.3}$$

where g ranges over the Eisenstein series $E_\mathfrak{a}\bigl(z,\frac12+it\bigr)$ associated to the various cusps $\mathfrak{a}$ of $X_0(q)$, and t ranges over $\mathbf{R}$. Elementary computations (similar to the ones in Section 5.4.2) show that $E_\mathfrak{a}\bigl(z,\frac12+it\bigr)$ is a linear combination of Eisenstein series of the form

$$\nu_{q',d}^*E^{\chi,\overline{\chi}}\bigl(z,\tfrac12+it\bigr)=E^{\chi,\overline{\chi}}\bigl(\nu_{q',d}z,\tfrac12+it\bigr),$$

where

$$E^{\chi,\overline{\chi}}(z,s):=\sum_{\substack{(m,n)\in\mathbf{Z}^2-\{(0,0)\}\\ q'|m}}\chi(m)\overline{\chi}(n)\frac{y^s}{|mz+n|^{2s}}$$

is the Eisenstein series associated to the pair of characters $(\chi,\overline{\chi})$, χ ranging over the primitive characters of modulus q' such that $(q')^2|q$, and d ranging over the divisors of $q/(q')^2$. Since we have assumed q square-free, $q'=1$, χ is trivial, and $E^{\chi,\overline{\chi}}(z,s)$ is the full level Eisenstein series $E(z,s)$.

Since $\nu_{1,d}(z_0^\sigma) = (\nu_{1,d}z_0)^\sigma$ and since $\nu_{1,d}z_0$ is a Heegner point on $X_0(1)$, it is sufficient to show (6.3) for $E\big(z, \frac{1}{2}+it\big)$. In that case the twisted Weyl sums have a well-known expression (see [GZ86, p.248] for example)

$$\left|\sum_{\sigma\in G_K} \psi(\sigma)E\big(z_0^\sigma, \tfrac{1}{2}+it\big)\right|^2 = \frac{|\mathcal{O}_K^\times|}{2}\frac{\sqrt{D}}{2}\big|L\big(f_\psi, \tfrac{1}{2}+it\big)\big|^2,$$

hence (6.3) follows from (1.3).

7. Appendix: Bounds for Bessel functions

In this appendix we recall some basic facts concerning Bessel functions and prove uniform bounds for Bessel functions of the first kind (Proposition 7.1) and of the second and third kinds (Proposition 7.2).

For $s \in \mathbf{C}$, the Bessel functions satisfy the recurrence relations

$$\big(x^s J_s(x)\big)' = x^s J_{s-1}(x), \qquad \big(x^s Y_s(x)\big)' = x^s Y_{s-1}(x), \qquad \big(x^s K_s(x)\big)' = -x^s K_{s-1}(x).$$

In particular, if $r > 0$ and H_s denotes either J_s, Y_s or K_s, then

$$\frac{d}{dx}\big((r\sqrt{x})^{s+1}H_{s+1}(r\sqrt{x})\big) = \pm(r^2/2)(r\sqrt{x})^s H_s(r\sqrt{x}), \tag{7.1}$$

and for any $j \geqslant 0$,

$$x^j\frac{d^j}{d^jx}H_s\left(\frac{r}{x}\right) = Q_j(s)H_s\left(\frac{r}{x}\right) + Q_{j-1}(s)\left(\frac{r}{x}\right)^1 H_{s-1}\left(\frac{r}{x}\right) + \cdots + Q_0(s)\left(\frac{r}{x}\right)^j H_{s-j}\left(\frac{r}{x}\right), \tag{7.2}$$

where each Q_i is a polynomial of degree i whose coefficients depend on i and j.

Proposition 7.1. *For any integer $k \geqslant 1$, the following uniform estimate holds:*

$$J_{k-1}(x) \ll \begin{cases} \frac{x^{k-1}}{2^{k-1}\Gamma\left(k-\frac{1}{2}\right)}, & 0 < x \leqslant 1; \\ kx^{-1/2}, & 1 < x. \end{cases}$$

The implied constant is absolute.

Proof. For $x > k^2$, the asymptotic expansion of J_{k-1} (see Section 7.13.1 of [O74]) provides the stronger estimate $J_{k-1}(x) \ll x^{-1/2}$ with an absolute implied constant.

For $1 < x \leqslant k^2$, we use Bessel's original integral representation (see Section 2.2 of [W44]),

$$J_{k-1}(x) = \frac{1}{2\pi}\int_0^{2\pi} \cos\big((k-1)\theta - x\sin\theta\big)\,d\theta,$$

to deduce that in this range

$$|J_{k-1}(x)| \leqslant 1 \leqslant kx^{-1/2}.$$

For the remaining range $0 < x \leqslant 1$, the required estimate follows from the Poisson-Lommel integral representation (see Section 3.3 of [W44])

$$J_{k-1}(x) = \frac{x^{k-1}}{2^{k-1}\Gamma\left(k-\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\int_0^\pi \cos(x\cos\theta)\sin^{2k-2}\theta\,d\theta. \quad \square$$

Proposition 7.2. *For any $\sigma > 0$ and any $\varepsilon > 0$, the following uniform estimates hold in the strip $|\Re s| \leqslant \sigma$:*

$$e^{-\pi|\Im s|/2} Y_s(x) \ll \begin{cases} \bigl(1+|\Im s|\bigr)^{\sigma+\varepsilon} x^{-\sigma-\varepsilon}, & 0 < x \leqslant 1+|\Im s|; \\ \bigl(1+|\Im s|\bigr)^{-\varepsilon} x^{\varepsilon}, & 1+|\Im s| < x \leqslant 1+|s|^2; \\ x^{-1/2}, & 1+|s|^2 < x. \end{cases}$$

$$e^{\pi|\Im s|/2} K_s(x) \ll \begin{cases} \bigl(1+|\Im s|\bigr)^{\sigma+\varepsilon} x^{-\sigma-\varepsilon}, & 0 < x \leqslant 1+\pi|\Im s|/2; \\ e^{-x+\pi|\Im s|/2} x^{-1/2}, & 1+\pi|\Im s|/2 < x. \end{cases}$$

The implied constants depend at most on σ and ε.

Proof. The last estimate for Y_s follows from its asymptotic expansion (see Section 7.13.1 of [O74]). The last estimate for K_s follows from Schläfli's integral representation (see Section 6.22 of [W44]),

$$K_s(x) = \int_0^\infty e^{-x\,\mathrm{ch}(t)}\,\mathrm{ch}(st)\,dt,$$

by noting that

$$\mathrm{ch}(t) \geqslant 1 + t^2/2 \qquad \text{and} \qquad |\,\mathrm{ch}(st)| \leqslant e^{\sigma t}.$$

We shall deduce the remaining uniform bounds from the integral representations

$$4K_s(x) = \frac{1}{2\pi i}\int_{\mathcal{C}} \Gamma\left(\frac{w-s}{2}\right)\Gamma\left(\frac{w+s}{2}\right)\left(\frac{x}{2}\right)^{-w} dw,$$

$$-2\pi Y_s(x) = \frac{1}{2\pi i}\int_{\mathcal{C}} \Gamma\left(\frac{w-s}{2}\right)\Gamma\left(\frac{w+s}{2}\right)\cos\left(\frac{\pi}{2}(w-s)\right)\left(\frac{x}{2}\right)^{-w} dw,$$

where the contour $\mathcal{C}$ is a broken line of 2 infinite and 3 finite segments joining the points

$$\begin{array}{ccc} -\varepsilon - i\infty, & -\varepsilon - i\bigl(2+2|\Im s|\bigr), & \sigma+\varepsilon - i\bigl(2+2|\Im s|\bigr), \\ \sigma+\varepsilon + i\bigl(2+2|\Im s|\bigr), & -\varepsilon + i\bigl(2+2|\Im s|\bigr), & -\varepsilon + i\infty. \end{array}$$

These formulae follow by analytic continuation from the well-known but more restrictive inverse Mellin transform representations of the K- and Y-Bessel functions, cf. formulae 6.8.17 and 6.8.26 in [E54].

If we write in the second formula

$$\cos\left(\frac{\pi}{2}(w-s)\right) = \cos\left(\frac{\pi}{2}w\right)\cos\left(\frac{\pi}{2}s\right) + \sin\left(\frac{\pi}{2}w\right)\sin\left(\frac{\pi}{2}s\right),$$

then it becomes apparent that the remaining inequalities of the lemma can be deduced from the uniform bound

$$\int_{\mathcal{C}} e^{\pi\max(|\Im s|,|\Im w|)/2}\left|\Gamma\left(\frac{w-s}{2}\right)\Gamma\left(\frac{w+s}{2}\right)\left(\frac{x}{2}\right)^{-w} dw\right| \ll_{\sigma,\varepsilon} \left(\frac{x}{1+|\Im s|}\right)^{-\sigma-\varepsilon} + \left(\frac{x}{1+|\Im s|}\right)^{\varepsilon}.$$

By introducing the notation

$$G(s) = e^{\pi|\Im s|/2}\Gamma(s),$$

$$M_s(x) = \int_{\mathcal{C}}\left|G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\left(\frac{x}{2}\right)^{-w} dw\right|,$$

the previous inequality can be rewritten as

$$M_s(x) \ll_{\sigma,\varepsilon} \left(\frac{x}{1+|\Im s|}\right)^{-\sigma-\varepsilon} + \left(\frac{x}{1+|\Im s|}\right)^{\varepsilon}. \tag{7.3}$$

Case 1. $|\Im s|\leqslant 1$.

If w lies on either horizontal segments of $\mathcal{C}$ or on the finite vertical segment joining $\sigma+\varepsilon\pm i\bigl(2+2|\Im s|\bigr)$, then $w\pm s$ varies in a fixed compact set (depending at most on σ and ε) disjoint from the negative axis $(-\infty,0]$. It follows that for these values w we have

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\ll_{\sigma,\varepsilon}1,$$

i.e.,

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\left(\frac{x}{2}\right)^{-w}\ll_{\sigma,\varepsilon}x^{-\sigma-\varepsilon},$$

and the same bound holds for the contribution of these values to $M_s(x)$.

If w lies on either infinite vertical segments of $\mathcal{C}$, then

$$|\Im(w\pm s)|\asymp|\Im w|>1,$$

whence Stirling's approximation yields

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\asymp_{\varepsilon}|\Im w|^{-\varepsilon-1}.$$

It follows that the contribution of the infinite segments to $M_s(x)$ is $\ll_{\sigma,\varepsilon}x^{\varepsilon}$.

Altogether we infer that

$$M_s(x)\ll_{\sigma,\varepsilon}x^{-\sigma-\varepsilon}+x^{\varepsilon},$$

which is equivalent to (7.3).

Case 2. $|\Im s|>1$.

If w lies on either horizontal segments of $\mathcal{C}$, then

$$|\Im(w\pm s)|\asymp|\Im s|,$$

whence Stirling's approximation yields

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\asymp_{\sigma,\varepsilon}|\Im s|^{\Re w-1},$$

i.e.,

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\left(\frac{x}{2}\right)^{-w}\asymp_{\sigma,\varepsilon}\frac{1}{|\Im s|}\left(\frac{|\Im s|}{x}\right)^{\Re w}.$$

It follows that the contribution of the horizontal segments to $M_s(x)$ is

$$\ll_{\sigma,\varepsilon}|\Im s|^{-1+\sigma+\varepsilon}x^{-\sigma-\varepsilon}+|\Im s|^{-1-\varepsilon}x^{\varepsilon}.$$

If w lies on the finite vertical segment of $\mathcal{C}$ joining $\sigma+\varepsilon\pm i\bigl(2+2|\Im s|\bigr)$, then

$$\Re(w\pm s)\geqslant\varepsilon\qquad\text{and}\qquad\max|\Im(w\pm s)|\asymp|\Im s|,$$

whence Stirling's approximation implies

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\ll_{\sigma,\varepsilon}\begin{cases}|\Im s|^{\sigma+\varepsilon/2-1/2} & \text{if }\ \min|\Im(w\pm s)|\leqslant 1;\\ |\Im s|^{\sigma+\varepsilon-1} & \text{if }\ \min|\Im(w\pm s)|>1.\end{cases}$$

It follows that the contribution of the finite vertical segment to $M_s(x)$ is

$$\ll_{\sigma,\varepsilon}|\Im s|^{\sigma+\varepsilon}x^{-\sigma-\varepsilon}.$$

If w lies on either infinite vertical segments of $\mathcal{C}$, then

$$|\Im(w\pm s)|\asymp|\Im w|>|\Im s|,$$

whence Stirling's approximation yields

$$G\left(\frac{w-s}{2}\right)G\left(\frac{w+s}{2}\right)\asymp_{\varepsilon}|\Im w|^{-\varepsilon-1}.$$

It follows that the contribution of the infinite vertical segments to $M_s(x)$ is

$$\ll_{\sigma,\varepsilon} |\Im s|^{-\varepsilon} x^{\varepsilon}.$$

Altogether we infer that

$$M_s(x) \ll_{\sigma,\varepsilon} |\Im s|^{\sigma+\varepsilon} x^{-\sigma-\varepsilon} + |\Im s|^{-\varepsilon} x^{\varepsilon},$$

which is equivalent to (7.3).

The proof of Proposition 7.2 is complete. □

References

[BD96] M. Bertolini, H. Darmon, *Heegner points of Mumford–Tate curves*, Invent. Math. 126 (1996), 413–456. 43, 45

[BR99] J. Bernstein, A. Reznikov, *Analytic continuation of representations and estimates of automorphic forms*, Ann. of Math. 150 (1999), 329–352. 5

[BR05] J. Bernstein, A. Reznikov, *Periods, subconvexity and representation theory* , J. Differential Geom. 69 (2005) (to appear) 7

[Bl04] V. Blomer, *Shifted convolution sums and subconvexity bound for automorphic L-functions*, Int. Math. Res. Not. 2004, 3905–3926. 7, 19, 38

[BHM05a] V. Blomer, G. Harcos, P. Michel, *A Burgess-like subconvex bound for twisted L-functions (with Appendix 2 by Z. Mao)*, Forum Math. (to appear) 7, 8, 37

[BHM05b] V. Blomer, G. Harcos, P. Michel, *On the subconvexity problem for L-functions of modular forms with nebentypus*, in preparation. 2

[B63] D. A. Burgess, *On character sums and L-series*, Proc. London Math. Soc. 12 (1962), 193–206.; *II*, ibid. 13 (1963), 524–536.

[C04] J. Cogdell, *L-functions and Converse Theorems for* GL_n, Park City Lecture Notes (2004, to appear), available at `http://www.math.okstate.edu/~cogdell/` 1, 19

[CPS90] J. Cogdell, I.I. Piatetskii-Shapiro, *The arithmetic and spectral analysis of Poincaré series*, Perspectives in Mathematics 13, Academic Press, Boston, MA, 1990. 14

[Da04] H. Darmon, *Rational points on modular elliptic curves*, CBMS Regional Conference Series in Mathematics 101, American Mathematical Society, Providence, RI, 2004. 43

[DI82] J-M. Deshouillers, H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. Math. 70 (1982/83), 219–288. 17, 18, 37, 38, 40

[D88] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. 92 (1988), 73–90. 3, 4

[DFI93] W. Duke, J. Friedlander, H. Iwaniec, *Bounds for automorphic L-functions*, Invent. Math. 112 (1993), 1–8. 3, 5, 7, 37

[DFI94a] W. Duke, J. Friedlander, H. Iwaniec, *A quadratic divisor problem*, Invent. Math. 115 (1994), 209–217. 2, 5, 6, 7, 30

[DFI94b] W. Duke, J. Friedlander, H. Iwaniec, *Bounds for automorphic L-functions. II*, Invent. Math. 115 (1994), 219–239. 30

[DFI97a] W. Duke, J. Friedlander, H. Iwaniec, *Bilinear forms with Kloosterman fractions*, Invent. Math. 128 (1997), 23–43. 2

[DFI97b] W. Duke, J. Friedlander, H. Iwaniec, *Representations by the determinant and mean values of L-functions*, Sieve methods, exponential sums, and their applications in number theory (Cardiff, 1995), 109–115, London Math. Soc. Lecture Note Ser. 237, Cambridge Univ. Press, Cambridge, 1997. 2

[DFI01] W. Duke, J. Friedlander, H. Iwaniec, *Bounds for automorphic L-functions. III*, Invent. Math. 143 (2001), 221–248. 2

[DFI02] W. Duke, J. Friedlander, H. Iwaniec, *The subconvexity problem for Artin L-functions*, Invent. Math. 149 (2002), 489–577. 2, 4, 10, 11, 12, 13, 15, 18, 21

[DI90] W. Duke, H. Iwaniec, *Bilinear forms in the Fourier coefficients of half-integral weight cusp forms and sums over primes*, Math. Ann. 286 (1990), 783–802. 10, 12

[DS90] W. Duke, R. Schulze-Pillot, *Representation of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids*, Invent. Math. 99 (1990), 49–57. 4

[E54] A. Erdélyi et al., *Tables of integral transforms, Vol. I. [based on notes left by H. Bateman]*, McGraw-Hill, New York, 1954. 48

[GZ86] B. Gross, D. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. 84 (1986), 225–320. 5, 47

[H03a] G. Harcos, *An additive problem in the Fourier coefficients of Maass forms,* Math. Ann. 326 (2003), 347–365. 3, 5, 6, 31, 38

[H03b] G. Harcos, *New bounds for automorphic L-functions,* Ph. D. thesis, Princeton University, 2003. 5, 6, 30, 38

[HM05] G. Harcos, P. Michel, *On some shifted convolution problems,* in preparation. 6

[HL94] J. Hoffstein, P. Lockhart, *Coefficients of Maass forms and the Siegel zero (with an appendix by D. Goldfeld, J. Hoffstein and D. Lieman)*, Ann. of Math. 140 (1994), 161–181. 15

[H85] M. Huxley, *Introduction to Kloostermania*, in *Elementary and Analytic Theory of Numbers*, 217–306, Banach Center Publ. 17, Warsaw, 1985.

[I87] H. Iwaniec, *Fourier coefficients of modular forms of half-integral weight*, Invent. Math. 87 (1987), 385–401. 3, 14

[I95] H. Iwaniec, *Introduction to the spectral theory of automorphic forms*, Biblioteca de la Revista Matemática Iberoamericana, Revista Matemática Iberoamericana, Madrid, 1995. 13, 17, 18

[IS00] H. Iwaniec, P. Sarnak, *Perspectives in the analytic theory of L-functions*, Geom. Funct. Anal. 2000, Special Volume, Part II, 705–741. 3

[J72] H. Jacquet, *Automorphic forms on* $\mathrm{GL}(2)$*. Part II.*, Lecture Notes in Mathematics 278, Springer-Verlag, Berlin–New York, 1972. 1, 19

[J92] M. Jutila, *Transformations of exponential sums*, Proceedings of the Amalfi Conference on Analytic Number Theory (Maiori 1989), Univ. Salerno, Salerno, 1992, 263–270. 6, 31

[J96] M. Jutila, *A variant of the circle method*, Sieve methods, exponential sums and their applications in number theory, 245–254, Cambridge University Press, 1996. 6, 19, 31

[J99] M. Jutila, *Convolutions of Fourier coefficients of cusp forms*, Publ. Inst. Math. (Beograd) (N.S.) 65(79) (1999), 31–51. 5, 6

[K03] H. Kim, *Functoriality for the exterior square of* GL_4 *and the symmetric fourth of* GL_2 *(with Appendix 1 by D. Ramakrishnan and Appendix 2 by H. Kim and P. Sarnak)*, J. Amer. Math. Soc. 16 (2003), 139–183. 2

[KS03] H. Kim, P. Sarnak, *Appendix: Refined estimates towards the Ramanujan and Selberg Conjectures*, J. Amer. Math. Soc. 16 (2003), 175–181. 2, 5

[KS02] H. Kim, F. Shahidi, *Cuspidality of symmetric powers with applications*, Duke Math. J. 112 (2002), 177–197. 2

[KMV02] E. Kowalski, P. Michel, J. VanderKam, *Rankin–Selberg L-functions in the level aspect*, Duke Math. J. 114 (2002), 123–191. 1, 5, 7, 10, 12, 30

[KSt02] B. Krötz, R. J. Stanton, *Holomorphic extension of representations: (I) automorphic functions*, Ann. of Math. 159 (2004), 641–724. 5

[L68] Y. V. Linnik, *Ergodic properties of algebraic fields [translated from the Russian by M. S. Keane]*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 45, Springer-Verlag, New York, 1968. 3

[LS95] W. Luo, P. Sarnak, *Quantum ergodicity of eigenfunctions on* $\mathrm{PSL}_2(\mathbf{Z})\backslash\mathbf{H}$, Inst. Hautes Études Sci. Publ. Math. No. 81 (1995), 207–237. 4

[M88] T. Meurman, *On exponential sums involving the Fourier coefficients of Maass wave forms*, J. Reine Angew. Math. 384 (1988), 192–207. 10, 12

[M04a] P. Michel, *The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. 160 (2004), 185–236. 1, 2, 3, 4, 5, 6, 8, 15, 16, 20, 21, 22, 24, 30, 38

[M04b] P. Michel, *Analytic number theory and families of L-functions*, Park City Lecture Notes (2004, to appear), available at `http://www.math.univ-montp2.fr/~michel/` 3

[Mo04] Y. Motohashi, *A note on the mean value of the zeta and L-functions. XIV*, preprint, 2004, `arxiv:math.NT/0401085` 6

[O74] F. W. J. Olver, *Asymptotics and special functions*, Academic Press, New York, 1974. 47, 48

[P03] A. Popa, *Central values of Rankin L-series over real quadratic fields*, Ph.D. thesis, Harvard University, 2003. 4

[R00] D. Ramakrishnan, *Modularity of the Rankin–Selberg L-series, and multiplicity one for* $\mathrm{SL}(2)$, Ann. of Math. 152 (2000), 45–111. 19

[S94] P. Sarnak, *Integrals of products of eigenfunctions*, Int. Math. Res. Not. 1994, 251–260. 3, 5

[S01] P. Sarnak, *Estimates for Rankin–Selberg L-functions and quantum unique ergodicity*, J. Funct. Anal. 184 (2001), 419–453. 3, 5, 6

[S65] A. Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Sympos. Pure Math., Vol. VIII, 1–15, Amer. Math. Soc., Providence, RI, 1965. 5

[V05] A. Venkatesh, *Sparse equidistribution problems, period bounds, and subconvexity*, preprint, 2005, `arxiv:math.NT/0506224` 7

[W81] J.-L. Waldspurger, *Sur les coefficients de Fourier des formes modulaires de poids demi-entier*, J. Math. Pures et Appliquées 60 (1981), 374–484. 3

[W44] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge University Press, Cambridge, 1944. 47, 48

[Z01a] S. Zhang, *Heights of Heegner points on Shimura curves*, Ann. of Math. 153 (2001), 27–147. 4, 5, 44, 46

[Z01b] S. Zhang, *Gross–Zagier formula for* GL_2, Asian J. Math. 5 (2001), 183–290. 4, 5, 46

[Z02] S. Zhang, *Elliptic curves, L-functions, and CM points*, Current developments in mathematics, 2001, 179–219, Int. Press, Somerville, MA, 2002. 43

[Z04] S. Zhang, *Gross–Zagier formula for* GL_2*. II*, preprint, 2002, available at `http://www.math.columbia.edu/~szhang/` 4, 43, 46

The University of Texas at Austin, Mathematics Department, 1 University Station C1200, Austin, TX 78712-0257, USA

E-mail address: `gharcos@math.utexas.edu`

I3M, UMR CNRS 5149, Université Montpellier II CC 051, 34095 Montpellier Cedex 05, France

E-mail address: `michel@math.univ-montp2.fr`