dc_116_10

Subconvex Bounds for Automorphic L-functions and Applications

Gergely Harcos

A Dissertation Presented to the
Hungarian Academy of Sciences
in Candidacy for the Title
of MTA Doktora

Budapest, 2011

dc_116_10

Abstract

This work presents subconvex bounds in the q-aspect for automorphic L-functions of $\mathrm{GL}_2 \times \mathrm{GL}_1$, GL_2, $\mathrm{GL}_2 \times \mathrm{GL}_2$ type over $\mathbb{Q}$ and some of their consequences. The results were published earlier in [BlHa08a, BHM07b, HM06], but there are some benefits of collecting them in one place. First, the proofs are interrelated at several levels, which justifies a joint introduction and uniform notation for them. Second, subsequent developments allow for additional remarks and numerical improvements. In particular, the main application for Heegner points and closed geodesics (Corollary 1.4) appears in stronger form than before.

Acknowledgements

I express my deep gratitude to my master Peter Sarnak and my collaborators Valentin Blomer and Philippe Michel without whom this work would not exist. There is also a long list of teachers, colleagues, friends, family members, and institutions who provided valuable help and support over the years. I hope they will not get offended that I did not collect their names here, fearing that I would leave out someone by accident, but they will know that I do remember and thank them from my heart.

To Yvette, Flóra and Máté

Contents

dc_116_10

Chapter 1

Introduction

1.1 L-functions

This dissertation deals with L-functions, a key unifying concept of number theory. The distinguished role L-functions play in mathematics is reflected by the fact that they are subject of 2 among the 7 Millennium Prize Problems of the Clay Mathematics Institute. In order to exploit the information encoded in these objects it is crucial to investigate their analytic properties such as analytic continuation, functional equation, distribution of poles and zeros, or bounds for their size. According to the Langlands philosophy, all L-functions in arithmetic can be built up from (principal) automorphic L-functions. For automorphic L-functions, some of the required analytic properties are readily available, while others have been identified as particularly deep. Current research in the field is based to a large extent on the idea that L-functions are not isolated objects but occur in natural families. Even a single L-function is regarded as a family of L-values in the modern point of view.

It has been realized recently that certain plausible analytic properties of L-functions in natural families provide the key to the solution of deep Diophantine problems. As such they also provide links to diverse fields including algebraic geometry, combinatorics, representation theory, ergodic theory, dynamical systems, scattering theory, random matrix theory, and mathematical physics. Two central issues, not independent of each other, are vanishing and size of L-functions in families. The former problem arises in connection with the rank of abelian varieties (conjecture of Birch and Swinnerton-Dyer), the theta correspondence, and the deformation theory of hyperbolic surfaces. The latter problem can be applied in various equidistribution problems such as Linnik's problems (equidistribution of lattice points on ellipsoids, or Heegner points and closed geodesics on arithmetic hyperbolic surfaces), their refinements and generalizations related to the André–Oort conjecture (equidistribution of incomplete Galois orbits of special subvarieties on Shimura varieties), Hilbert's 11th problem (equidistribution of representations by quadratic forms in a given genus), and Quantum Unique Ergodicity (equidistribution of mass on arithmetic hyperbolic surfaces). Excellent descriptions of these and other exciting developments can be found in [Fr95, KS99, IS00, Sa03, MV06, Mi07].

In this dissertation we discuss subconvex bounds for classical automorphic L-functions and some of their applications.

1.2 The subconvexity problem

A completed principal automorphic L-function $\Lambda(\pi, s)$ of degree n over a number field F is associated to an irreducible cuspidal automorphic representation π of the group GL_n over F with unitary central character. It is a meromorphic function in the complex variable s (with possible simple poles on the lines $\Re s = 0$ and $\Re s = 1$ which occur if and only if $n = 1$ and $\pi = |\det|^{it}$), and by the cuspidality of π it is not a product of completed L-functions of smaller degree. The representation π itself can be realized as an irreducible subspace of the space of all cusp forms on the adelic quotient $\mathrm{GL}_n(F)\backslash\mathrm{GL}_n(\mathbb{A}_F)$, endowed with commuting right actions of $\mathrm{GL}_n(F_v)$ at non-archimedean places v of F and the Lie algebra of $\mathrm{GL}_n(F_v)$ at archimedean places v. This harmonizes with Flath's theorem that π can be written as a restricted tensor product $\otimes_v \pi_v$, where π_v is an irreducible admissible representation of

$\mathrm{GL}_n(F_v)$ for each place v of F. Accordingly, we have a product decomposition $\Lambda(\pi, s) = \prod_v L(\pi_v, s)$ which is absolutely convergent for $\Re s > 1$. The completed L-function is bounded in vertical strips (away from the possible poles) and a simple functional equation relates $\Lambda(\pi, s)$ to $\Lambda(\tilde{\pi}, 1-s)$, where $\tilde{\pi}$ is the contragradient representation of π satisfying $L(\tilde{\pi}_v, s) = \overline{L(\pi_v, \overline{s})}$.

The finer analytic behavior of $\Lambda(\pi, s)$ becomes transparent when the archimedean local factors $L(\pi_v, s)$ are detached from it. Indeed, in vertical strips the archimedean factors decay exponentially while the non-archimedean factors remain bounded away from zero. The product of non-archimedean factors is the finite L-function $L(\pi, s)$. Its size, the central theme of this dissertation, is measured relative to the analytic conductor $C(\pi, s)$ which captures the "local ramification data" at all places of F, see [IS00]. Combining the Phragmén–Lindelöf convexity principle with the functional equation for $\Lambda(\pi, s)$ one can deduce the convexity bound $L(\pi, s) \ll_{\varepsilon,n,F} C(\pi, s)^{\frac{1}{4}+\varepsilon}$ on the critical line $\Re s = \frac{1}{2}$. Here and later ε denotes an arbitrary positive number, and the symbol $\ll_{\varepsilon,n,F}$ abbreviates "in absolute value less than a constant depending on ε, n, F times". In fact these L-values can be uniformly recovered, up to arbitrary precision, by truncating the Dirichlet series for $L(\pi, s)$ and $L(\tilde{\pi}, 1-s)$ after about $C(\pi, s)^{\frac{1}{2}+\varepsilon}$ terms, see [Ha02]. The Generalized Riemann Hypothesis states that all zeros of $\Lambda(\pi, s)$ lie on the line $\Re s = \frac{1}{2}$. It would imply that the exponent $\frac{1}{4} + \varepsilon$ in the convexity bound can be replaced by ε. This dream estimate (not proven in a single instance) is the Generalized Lindelöf Hypothesis. A more realistic goal is to establish, for special families (or conjectural families) of representations π the existence of some $\delta = \delta(n, F) > 0$ such that $L(\pi, s) \ll_{\delta,n,F} C(\pi, s)^{\frac{1}{4}-\delta}$ on the line $\Re s = \frac{1}{2}$. This is the subconvexity problem for automorphic L-functions.

A serious motivation for deriving subconvex bounds for automorphic L-functions comes from the fact that in several equidistribution problems the error term can be expressed (by deep explicit formulae) from special values of these L-functions. Usually, the convexity bound just falls short of establishing equidistribution, while any nontrivial improvement $\delta > 0$ is sufficient. In other words, arithmetic becomes "visible" exactly when a subconvex bound is achieved for the family of L-functions at hand. There are situations where the quality of the subconvex exponent is critical. For example, [Hu72] needs some $\delta > \frac{1}{12}$ for $\zeta(\frac{1}{2} + it)$, while [CCU09] utilizes the range $\delta < \frac{1}{32}$ for a certain family of $\mathrm{GL}_2 \times \mathrm{GL}_1$ type.

Depending on various parameters involved in the analytic conductor $C(\pi, s)$ we can talk about the s-aspect, the ∞-aspect (or eigenvalue-aspect) and the q-aspect (or level-aspect) of the subconvexity problem. In this dissertation we focus on the q-aspect for families of $\mathrm{GL}_2 \times \mathrm{GL}_1$, GL_2, $\mathrm{GL}_2 \times \mathrm{GL}_2$ type over $\mathbb{Q}$, therefore we mention only briefly some recent developments in other directions: [Bl11, BlHa10, BR05, JM05, JM06, LLY06, Li11, MV10, Ve10].

1.3 Summary of results

An irreducible cuspidal automorphic representation of GL_2 over $\mathbb{Q}$ can be identified (modulo a simple equivalence) with a classical modular form on the upper half-plane $\mathcal{H}$: a primitive holomorphic cusp form integral weight $k \geqslant 1$, or a primitive Maass cusp form of weight $\kappa \in \{0, 1\}$. Such an automorphic form g shares three fundamental properties (appropriately defined):

- symmetric with respect to a congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$;
- square-integrable modulo Γ;
- simultaneous eigenfunction of the Laplace and Hecke operators.

We denote the Laplacian eigenvalue by $\frac{1}{4}+t_g^2$ and call $\mu_g := 1+|t_g|$ the spectral parameter of g (hence $\mu_g = \frac{k_g+1}{2}$ when g is holomorphic of weight k_g). We denote the eigenvalue of the n-th Hecke operator by $\lambda_g(n)$: these complex numbers are of central importance for us as they give rise to the various L-functions in this dissertation. The following hypothesis is very useful in analytic investigations.

Hypothesis H_θ. *If g is a primitive Maass cusp form of weight 0 or 1, then $\lambda_g(n) \ll_\varepsilon n^{\theta+\varepsilon}$. If g is a primitive Maass cusp form of weight 0, then $\frac{1}{4} + t_g^2 \geqslant \frac{1}{4} - \theta^2$.*

We note that for holomorphic cusp forms g the estimate $\lambda_g(n) \ll_\varepsilon n^\varepsilon$ was proved by Deligne [De74], while in the case of weight 1 Maass cusp forms $\frac{1}{4} + t_g^2 \geqslant \frac{1}{4}$ follows from the representation theory of $\mathrm{SL}_2(\mathbb{R})$. For $\theta = 0$ Hypothesis H_θ is the Ramanujan–Selberg conjectures, while any $\theta < \frac{1}{2}$

is nontrivial. Currently $\theta = \frac{7}{64}$ is known to be admissible by the deep work of Kim–Shahidi, Kim and Kim–Sarnak [KiSh02, Ki03, KiSa03].

The first family we consider consists of twisted forms $f \otimes \chi$ with a fixed primitive cusp form f and a primitive Dirichlet character χ that varies. The associated (finite) L-functions are essentially defined as Dirichlet series

$$L(f \otimes \chi, s) \approx \sum_{n=1}^{\infty} \frac{\lambda_f(n)\chi(n)}{n^s}, \qquad \Re s > 1, \tag{1.1}$$

where $\approx$ means that the ratio of the two sides is negligible for our analytic purposes. These L-functions have similar features as Riemann's zeta function and Dirichlet's L-functions, namely each of them

- decomposes as an infinite Euler product over the prime numbers;
- extends to an entire function which exhibits a symmetry with respect to $s \longleftrightarrow 1 - s$.

In particular, denoting by q the conductor of χ and by N the level of f, we have the following (simple) convexity bound[1] on the critical line $\Re s = \frac{1}{2}$:

$$L(f \otimes \chi, s) \ll_\varepsilon (|s|\mu_f N q)^\varepsilon \, |s|^{\frac{1}{2}} \mu_f^{\frac{1}{2}} N^{\frac{1}{4}} q^{\frac{1}{2}}. \tag{1.2}$$

The Generalized Lindelöf Hypothesis predicts that all the exponents in (1.2) can be replaced by ε, and the subconvexity problem aims at reducing (some of) these exponents. Our first result concerns the q-aspect of this problem, i.e. we are primarily interested in reducing the exponent $\frac{1}{2}$ of q in (1.2), but we also try to keep the other 3 exponents moderate. Historically, this special case was examined first (after the classical work of Burgess [Bu63] about the GL_1 analogue, see (1.3) below), and it served as the starting point of the systematic study of the general subconvexity problem.

The initial breakthrough was achieved in 1993 by Duke, Friedlander, Iwaniec [DFI93] who improved the exponent of q to $\frac{1}{2} - \delta$ with $\delta = \frac{1}{22}$ when f is a holomorphic cusp form of full level ($N = 1$). Their proof introduced many of the basic tools for the subconvexity problem, such as the amplification method (a technique based on estimating weighted second moments of the family) and the application of various summation formulae for the Hecke eigenvalues. Subsequent progress in this problem can be summarized as follows[2]: $\delta = \frac{1}{8}$ for f holomorphic of trivial nebentypus by Bykovskiĭ [By96], $\delta = \frac{1}{54}$ for f arbitrary[3] by Harcos [Ha03a, Ha03b], $\delta = \frac{1}{22}$ by Michel [Mi04], $\delta = \frac{1-2\theta}{10+4\theta}$ by Blomer [Bl04], $\delta = \frac{1-2\theta}{8}$ by Blomer–Harcos–Michel [BHM07a]. In the last two results θ is such that Hypothesis H_θ holds (hence $\theta = \frac{7}{64}$ is admissible), and the results depend on this parameter for a good reason. Namely, the papers [Bl04, BHM07a] proceed along the lines of [DFI93] where amplification is carried out over the characters χ. After the averaging the $\chi(n)$'s from (1.1) disappear, but the $\lambda_f(n)$'s survive in products of pairs. These pairs of Hecke eigenvalues are grouped in shifted convolution sums which are then analyzed by elaborate techniques of harmonic analysis. Still, some factors of type $\lambda_f(q)$ turn out to be very "robust", and this yields an unwanted factor q^θ in the relevant estimates. It is for this reason that Bykovskiĭ's method is remarkable as it produces $\delta = \frac{1}{8}$ without any θ. Note that this is the true analogue of Burgess' famous bound [Bu63]

$$L(\chi, s) \ll_\varepsilon (|s|q)^\varepsilon \, |s|^{\frac{1}{4}} q^{\frac{1}{4} - \frac{1}{16}}, \tag{1.3}$$

because $L(f \otimes \chi, s)$ is closely related to the products $L(\chi_1, s)L(\chi_2, s)$ with $\chi_1\chi_2 = \chi^2$. It is all the more interesting that [BHM07a] falls short of this result only by the presence of θ, although it imposes no restriction on the nebentypus or the type of f. Bykovskiĭ's key idea was to amplify over the forms f in the spectrum of level $[N, q]$. In this averaging the $\lambda_f(n)$'s from (1.1) disappear, and only the $\chi(n)$'s survive which are trivially bounded by 1. Of course this description is very vague, but hopefully it motivates well the overall discussion.

The first result in this dissertation is joint work with Valentin Blomer [BlHa08a] which pushes the method of Bykovskiĭ [By96] to its limit.

[1] In fact the convexity bound is a slightly stronger statement, we displayed the version in which the various parameters appear separated.

[2] We list results proved for all χ, hence we omit [CI00].

[3] In the case of Maass forms we assumed that the weight is 0 as the case of weight 1 is almost identical. The same is true of later developments.

Theorem 1.1. *Let f be a primitive (holomorphic or Maass) cusp form of level N and trivial nebentypus, and let χ be a primitive character modulo q. Then for $\Re s = \frac{1}{2}$ and for any $\varepsilon > 0$ one has*

$$L(f \otimes \chi, s) \ll_\varepsilon (|s|\mu_f N q)^\varepsilon \left(|s|^{\frac{1}{4}} \mu_f^{\frac{1}{2}} N^{\frac{1}{4}} q^{\frac{3}{8}} + |s|^{\frac{1}{2}} \mu_f N^{\frac{1}{2}} (N,q)^{\frac{1}{4}} q^{\frac{1}{4}}\right)$$

if f is holomorphic, and

$$L(f \otimes \chi, s) \ll_\varepsilon (|s|\mu_f N q)^\varepsilon \left(|s|^{\frac{1}{4}} \mu_f^{3} N^{\frac{1}{4}} q^{\frac{3}{8}} + |s|^{\frac{1}{2}} \mu_f^{\frac{7}{2}} N^{\frac{1}{2}} (N,q)^{\frac{1}{4}} q^{\frac{1}{4}}\right)$$

otherwise.

The novelty of this theorem is that it covers Maass forms and achieves good uniformity in the secondary parameters (e.g. it is as strong as the convexity bound in the s-aspect). In applications it is easier to handle a single term on the right hand side, hence we formulate

Corollary 1.1. *Let f be a primitive (holomorphic or Maass) cusp form of level N and trivial nebentypus, and let χ be a primitive character modulo q. Then for $\Re s = \frac{1}{2}$ and for any $\varepsilon > 0$ one has*

$$L(f \otimes \chi, s) \ll_\varepsilon (|s|\mu_f N q)^\varepsilon |s|^{\frac{1}{2}} \mu_f^{3} N^{\frac{1}{2}} q^{\frac{3}{8}}. \tag{1.4}$$

Moreover, for $q \geqslant (\mu_f N)^4$ one has

$$L(f \otimes \chi, s) \ll_\varepsilon (|s|\mu_f N q)^\varepsilon |s|^{\frac{1}{2}} \mu_f^{3} N^{\frac{1}{4}} q^{\frac{3}{8}}. \tag{1.5}$$

This corollary along with the ones below are deduced from the theorems in the next section. An important consequence of Theorem 1.1 is an improved bound for the Fourier coefficients of half-integral weight cusp forms (see [BlHa08a, Corollary 2] and [BM10, Theorem 1.5]), which in turn can be applied to various distribution problems on ellipsoids and hyperbolic surfaces [Du88, DuSP90], and representations by ternary quadratic forms with restricted variables [Bl08]. Another application is the following hybrid subconvexity bound on the critical line [BlHa08a, Theorem 1]:

$$L(f \otimes \chi, s) \ll_\varepsilon (N|s|q)^\varepsilon N^{\frac{4}{5}} (|s|q)^{\frac{1}{2}-\frac{1}{40}}.$$

Finally, Theorem 1.1 is an important ingredient in the proofs of Theorems 1.2 and 1.3 below.

The second family we consider consists of primitive cusp forms f of level q, for which the convexity bound reads

$$L(f, s) \ll_\varepsilon (|s|\mu_f q)^\varepsilon |s|^{\frac{1}{2}} \mu_f^{\frac{1}{2}} q^{\frac{1}{4}}.$$

The aim is to prove a similar bound with q-exponent $\frac{1}{4} - \delta$ (where $\delta > 0$ is fixed) and with an implied constant depending continuously on s and μ_f. History in brief is as follows: $\delta = \frac{1}{192}$ for f holomorphic of trivial nebentypus by Duke–Friedlander–Iwaniec [DFI94b], $\delta = \frac{1}{262144}$ for f holomorphic of squarefree level q and primitive nebentypus [DFI01], $\delta = \frac{1}{23041}$ for f of primitive nebentypus [DFI02].

The second result in this dissertation is joint work with Valentin Blomer and Philippe Michel [BHM07b] which establishes a stronger and more general subconvexity estimate for modular L-functions with a different method.

Theorem 1.2. *Let f be a primitive (holomorphic or Maass) cusp form of level q and nontrivial nebentypus. Then for $\Re s = \frac{1}{2}$ one has*

$$L(f, s) \ll (|s|\mu_f)^A q^{\frac{1}{4}-\frac{1}{1889}}, \tag{1.6}$$

where $A > 0$ is an absolute constant.

The novelty of this theorem is that it only requires the nebentypus to be nontrivial[4] instead of primitive, and the subconvexity exponent is stronger. Including non-primitive nebentypus is crucial in the following corollaries which have arithmetic applications.

[4]In fact, with slightly more work we could also have covered the trivial nebentypus case, see Remark 4.2.

Corollary 1.2. *Let K be a quadratic number field and $\mathcal{O} \subset K$ an order in K of discriminant $d_{\mathcal{O}}$. Let χ denote a primitive character of* $\mathrm{Pic}(\mathcal{O})$*. Then for $\Re s = \frac{1}{2}$ one has*

$$L(\chi, s) \ll |s|^A |d_{\mathcal{O}}|^{\frac{1}{4} - \frac{1}{1889}},$$

where $A > 0$ is an absolute constant.

Corollary 1.3. *Let K be a cubic number field of discriminant d_K. Then for $\Re s = \frac{1}{2}$ the Dedekind L-function of K satisfies*

$$\zeta_K(s) \ll |s|^A |d_K|^{\frac{1}{4} - \frac{1}{1889}}, \tag{1.7}$$

where $A > 0$ is an absolute constant.

Corollary 1.3 is an essential ingredient in the deep work of Einsiedler–Lindenstrauss–Michel–Venkatesh [ELMV11] which establishes a higher rank generalization of Duke's equidistribution theorem for closed geodesics on the modular surface [Du88, Theorem 1].

The third family we consider consists of Rankin–Selberg convolutions $f \otimes g$ with a fixed primitive cusp form g and a primitive cusp form f that varies. The associated (finite) L-functions are essentially defined as Dirichlet series

$$L(f \otimes g, s) \approx \sum_{n=1}^{\infty} \frac{\lambda_f(n)\lambda_g(n)}{n^s}, \qquad \Re s > 1,$$

where again $\approx$ means that the ratio is negligible for our analytic purposes. These L-functions have similar features as the ones already mentioned (Euler product, analytic continuation, symmetry), hence denoting by q the level of f and by D the level of g, we have the following convexity bound on the critical line $\Re s = \frac{1}{2}$:

$$L(f \otimes g, s) \ll_\varepsilon (|s|\mu_f\mu_g Dq)^\varepsilon |s|\mu_f\mu_g D^{\frac{1}{2}} q^{\frac{1}{2}}.$$

The aim is to prove a similar bound with q-exponent $\frac{1}{2} - \delta$ (where $\delta > 0$ is fixed) and with an implied constant depending continuously on the other parameters. This problem was solved by Kowalski–Michel–Vanderkam [KMV02] when f is holomorphic and the conductor of $\chi_f\chi_g$ (where χ_f and χ_g are the nebentypus characters of f and g) is at most $q^{\frac{1}{2}-\eta}$ for some $\eta > 0$, the corresponding savings δ then depending on η. The second condition (which is the more serious) was essentially removed by Michel [Mi04] under the assumptions that g is holomorphic and $\chi_f\chi_g$ is nontrivial.

The third result in this dissertation is joint work with Philippe Michel [HM06] which solves the subconvexity problem for Rankin–Selberg L-functions in even greater generality.

Theorem 1.3. *Let f and g be two primitive (holomorphic or Maass) cusp forms of level q, D and nebentypus χ_f, χ_g, respectively. Assume that $\chi_f\chi_g$ is not trivial. Then for $\Re s = \frac{1}{2}$ one has*

$$L(f \otimes g, s) \ll (|s|\mu_f\mu_g D)^A q^{\frac{1}{2} - \frac{1}{1413}}, \tag{1.8}$$

where $A > 0$ is an absolute constant.

The novelty of this theorem is that it contains no restriction on the type of the cusp forms involved, and the dependence on the secondary parameters is polynomial. To be precise, in [HM06] we proved the result with q-exponent $\frac{1}{2} - \frac{1}{2648}$, because at that time only a weaker version of Theorem 1.1 was available. Here we take the opportunity to update the exponents in [HM06], and indicate to some extent how the exponent of q in (1.8) depends on θ and the exponents in (1.4), see Proposition 5.1.

The above subconvexity results can be used to reprove and refine Duke's equidistribution theorem [Du88] which we discuss now briefly. For a fundamental discriminant $d < 0$ (resp. $d > 0$) denote by Λ_d the set of Heegner points (resp. closed geodesics) of discriminant d on the modular surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. As shown in Section 6.1, there is a natural bijection between Λ_d and the narrow ideal class group H_d of $\mathbb{Q}(\sqrt{d})$, in particular H_d acts on Λ_d in a natural fashion. The total volume of Λ_d is $|d|^{1/2+o(1)}$ by Siegel's theorem (cf. (6.9)), hence it is natural to ask if Λ_d becomes equidistributed in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ as $|d| \to \infty$. Linnik [Li68], using his pioneering ergodic method, could establish equidistribution under the condition that $\left(\frac{d}{p}\right) = 1$ for any fixed odd prime p. The congruence restriction was removed by Duke [Du88] using quite different techniques. Duke exploited a correspondence of Maass to relate the Weyl sums arising in this equidistribution problem to Fourier coefficients of half-integral weight Maass forms, and then he proved directly nontrivial bounds for them using a technique introduced by Iwaniec [Iw87]. The connection with subconvexity comes from the work of Waldspurger [Wa81] on the Shimura correspondence, which shows that nontrivial bounds for these Fourier coefficients are in fact equivalent to subconvexity bounds for the central twisted values $L\left(f\otimes(\frac{d}{\cdot}),\frac{1}{2}\right)$ as f ranges over the Hecke–Maass cusp forms and Eisenstein series on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. The necessary bounds follow from (1.3) and (1.4) above.

In combination with the special formulae of Zhang [Zh01] for $d<0$ and Popa [Po06] for $d>0$, Theorems 1.2 and 1.3 imply the equidistribution of substantially smaller subsets of Λ_d, as $|d|\to\infty$.

Corollary 1.4. *Let $d\mu(z)$ (resp. $ds(z)$) denote the hyperbolic probability measure (resp. hyperbolic arc length) on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. Let $g:\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}\to\mathbb{C}$ be a smooth function of compact support.*

- *If $d<0$ is a negative fundamental discriminant, $H\leqslant H_d$ is a subgroup of the narrow ideal class group of $\mathbb{Q}(\sqrt{d})$, and $z_0\in\Lambda_d$ is a Heegner point of discriminant d, then*

$$\frac{\sum_{\sigma\in H} g(z_0^\sigma)}{\sum_{\sigma\in H}1}=\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g(z)\,d\mu(z)+O_g\left([H_d:H]|d|^{-\frac{1}{2827}}\right). \tag{1.9}$$

- *If $d>0$ is a positive fundamental discriminant, $H\leqslant H_d$ is a subgroup of the narrow ideal class group of $\mathbb{Q}(\sqrt{d})$, and $G_0\in\Lambda_d$ is a closed geodesic of discriminant d, then*

$$\frac{\sum_{\sigma\in H}\int_{G_0^\sigma} g(z)\,ds(z)}{\sum_{\sigma\in H}\int_{G_0^\sigma}1\,ds(z)}=\int_{\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}} g(z)\,d\mu(z)+O_g\left([H_d:H]|d|^{-\frac{1}{2827}}\right). \tag{1.10}$$

In particular, every H-orbit in Λ_d becomes equidistributed on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ under $[H_d:H]\leqslant|d|^{\frac{1}{2828}}$ and $|d|\to\infty$. In the above bounds the implied constant is a Sobolev norm of g.

This corollary strengthens the numerical values in [HM06, Theorem 2] and [Po06, Theorem 6.5.1]. On the other hand, [HM06] and [Po06] discuss the analogous results on more general arithmetic hyperbolic surfaces, which we omit here for simplicity.

We conclude this summary by mentioning that the subconvex bounds (1.4), (1.6), (1.8) were successfully applied in a number of other situations, see [MV07, Sa07, FM11, KMY11, Ma11, MY11].

1.4 Proof of the corollaries

Proof of Corollary 1.1. By Theorem 1.1 we have

$$L(f\otimes\chi,s)\ll_\varepsilon(|s|\mu_f Nq)^\varepsilon\left(|s|^{\frac12}\mu_f^3N^{\frac12}q^{\frac38}+|s|^{\frac12}\mu_f^{\frac72}N^{\frac34}q^{\frac14}\right).$$

If the first term dominates inside the big parentheses, then (1.4) is clear. Else we have

$$|s|^{\frac12}\mu_f^3N^{\frac12}q^{\frac38}<|s|^{\frac12}\mu_f^{\frac72}N^{\frac34}q^{\frac14}\quad\Longrightarrow\quad q^{\frac18}<\mu_f^{\frac12}N^{\frac14}.$$

Combining this with the convexity bound (1.2) we arrive at (1.4) again:

$$L(f\otimes\chi,s)\ll_\varepsilon(|s|\mu_f Nq)^\varepsilon|s|^{\frac12}\mu_f^{\frac12}N^{\frac14}q^{\frac18}q^{\frac38}<(|s|\mu_f Nq)^\varepsilon|s|^{\frac12}\mu_f N^{\frac12}q^{\frac38}.$$

As for (1.5) we note that by Theorem 1.1 we have

$$L(f\otimes\chi,s)\ll_\varepsilon(|s|\mu_f Nq)^\varepsilon\left(|s|^{\frac12}\mu_f^3N^{\frac14}q^{\frac38}+|s|^{\frac12}\mu_f^{\frac72}N^{\frac34}q^{\frac14}\right).$$

If the first term dominates inside the big parentheses, then (1.5) is clear. Else we have

$$|s|^{\frac12}\mu_f^3N^{\frac14}q^{\frac38}<|s|^{\frac12}\mu_f^{\frac72}N^{\frac34}q^{\frac14}\quad\Longrightarrow\quad q^{\frac18}<\mu_f^{\frac12}N^{\frac12}\quad\Longrightarrow\quad q<(\mu_f N)^4.$$

□

Proof of Corollary 1.2. As in [DFI02] we only need to remark that depending on whether K is real or imaginary, $L(\chi, s)$ is the L-function of a Maass form of weight $\kappa \in \{0,1\}$, level d and nebentypus χ_K (the quadratic character associated with K). This follows from theorems of Hecke and Maass. One difference with Theorem 2.7 of [DFI02] is that we do not require the character χ to be associated with the maximal order $\mathcal{O}_K$. Now the bound follows from Theorem 1.2. $\square$

Proof of Corollary 1.3. If K is abelian, then $d_K = d^2$ is a square and $\zeta_K(s) = \zeta(s)L(\chi, s)L(\overline{\chi}, s)$, where χ is a Dirichlet character of order 3 and conductor d. In that case the bound (1.7) follows from Burgess's subconvex bound [Bu63]. If K is not abelian, let L denote the Galois closure of K (which is of degree 6 with Galois group isomorphic to S_3) and let $F/\mathbb{Q}$ denote the unique quadratic field contained in L, then $\zeta_K(s) = \zeta(s)L(\chi, s)$, where χ is a ring class character of F of order 3 and conductor $\mathfrak{d}$ satisfying $N_{F/\mathbb{Q}}(\mathfrak{d}) = |d_K|$. The bound (1.7) now follows from Corollary 1.2. $\square$

Proof of Corollary 1.4. The spectral expansion (2.1) is compatible with taking partial derivatives on both sides, therefore it suffices to prove the statement when g is a Hecke–Maass cusp form of full level with $\langle g, g\rangle = 1$ or a standard Eisenstein series $E_\infty(\cdot, \frac{1}{2}+it)$. More precisely, it suffices to show for such g that the left hand sides of (1.9)–(1.10) are $\ll [H_d : H](1+|t|)^A|d|^{-\frac{1}{2827}}$, where $A > 0$ is an absolute constant and $t = t_g$ is the spectral parameter of g as in (2.4). By (6.9) the denominators in (1.9)–(1.10) are $[H_d : H]^{-1}|d|^{\frac{1}{2}+o(1)}$, hence it suffices to show that the numerators satisfy

$$\sum_{\sigma\in H} \dots \ll (1+|t|)^A|d|^{\frac{1}{2}-\frac{1}{2826}}.$$

Using characters of the abelian group H_d we can rewrite this as

$$\frac{1}{[H_d : H]} \sum_{\substack{\psi\in\widehat{H}_d \\ \psi_{|H}\equiv 1}} \sum_{\sigma\in H_d} \overline{\psi(\sigma)} \dots \ll (1+|t|)^A|d|^{\frac{1}{2}-\frac{1}{2826}}.$$

The number of ψ's here is precisely $[H_d : H]$, hence it suffices to show that for any $\psi \in \widehat{H}_d$ and for any g as above we have

$$\begin{aligned} \sum_{\sigma\in H_d} \overline{\psi(\sigma)} g(z_0^\sigma) &\ll (1+|t|)^A\, |d|^{\frac{1}{2}-\frac{1}{2826}}, \qquad d<0, \\ \sum_{\sigma\in H_d} \overline{\psi(\sigma)} \int_{G_0^\sigma} g(z)\, ds(z) &\ll (1+|t|)^A\, |d|^{\frac{1}{2}-\frac{1}{2826}}, \qquad d>0. \end{aligned} \tag{1.11}$$

The twisted sums in (1.11) can be related to central automorphic L-values. The formula (which generalizes special cases by Dirichlet, Hecke, Maass, Gross–Kohnen–Zagier and others) is based on the deep work of Waldspurger [Wa81] and was carefully derived by Zhang [Zh01] for $d < 0$ and by Popa [Po06] for $d > 0$:

$$\left|\sum_{\sigma\in H_d} \overline{\psi(\sigma)} \dots\right|^2 = c_d\, |d|^{\frac{1}{2}}\, |\rho_g(1)|^2\, \Lambda\left(f_\psi \otimes g, \tfrac{1}{2}\right). \tag{1.12}$$

Here c_d is positive and takes only finitely many different values, $\rho_g(1)$ is the first Fourier coefficient of g as in (2.2)–(2.3), $\Lambda(\pi, s)$ denotes the completed L-function, and f_ψ is the automorphic induction of ψ from GL_1 over $\mathbb{Q}(\sqrt{d})$ to GL_2 over $\mathbb{Q}$ such that $\Lambda(f_\psi, s) = \Lambda(\psi, s)$. The modular form f_ψ was discovered by Hecke [He37] and Maass [Ma49] in this special case, it is of level $|d|$ and nebentypus $\left(\frac{d}{\cdot}\right)$. In particular, when g is an Eisenstein series $E_\infty(\cdot, \frac{1}{2}+it)$ the identity (1.12) follows from [Si80, pp. 70 and 88] and [Iw02, (3.25)].

Observe that in (1.12) we have $|\rho_g(1)|^2 \ll_\varepsilon (1+|t|)^\varepsilon e^{\pi|t|}$ by [HL94] and [Iw02, (3.25)], while the archimedean part of $\Lambda\left(f_\psi \otimes g, \frac{1}{2}\right)$ is a product of exponential and gamma factors which is $\ll (1+|t|)e^{-\pi|t|}$ by Stirling's approximation. Therefore (1.11) reduces to a subconvex bound (with a different $A > 0$)

$$L\left(f_\psi \otimes g, \tfrac{1}{2}\right) \ll (1+|t|)^A\, |d|^{\frac{1}{2}-\frac{1}{1413}}. \tag{1.13}$$

If the character $\psi : H_d \to \mathbb{C}^\times$ is real-valued, then it is one of the genus characters discovered by Gauss [Ga86]. In this case $L(f_\psi \otimes g, s) = L(g \otimes (\frac{d_1}{\cdot}), s)L(g \otimes (\frac{d_2}{\cdot}), s)$, where $d = d_1 d_2$ is a factorization of d into fundamental discriminants d_1 and d_2 (cf. [Si80, p. 62]), therefore (1.13) follows from (1.3) when g is an Eisenstein series and from (1.4) when g is a cusp form. If the character $\psi : H_d \to \mathbb{C}^\times$ is not real-valued, then f_ψ is a cusp form of level $|d|$ and nebentypus $\left(\frac{d}{\cdot}\right)$, therefore (1.13) follows from (1.6) when g is an Eisenstein series and from (1.8) when g is a cusp form. $\square$

1.5 About the proof of the main theorems

In this section we summarize briefly the main ideas in the proof of Theorems 1.1, 1.2, 1.3. The expert reader will notice that the ancestors to the proof are the papers [By96, KMV00, DFI02, Mi04]. Using the notation

- $\mathcal{L}(f) := L(f \otimes \chi, s)$ in the case of Theorem 1.1;
- $\mathcal{L}(f) := L(f, s)^2$ in the case of Theorem 1.2;
- $\mathcal{L}(f) := L(f \otimes g, s)$ in the case of Theorem 1.3;

the goal is to find a particular $\delta > 0$ such that $\mathcal{L}(f) \ll q^{\frac{1}{2}-\delta}$ with an implied constant depending polynomially on the secondary parameters. We achieve this by estimating the amplified second moment

$$\frac{1}{q}\int_\phi |\mathcal{M}(\phi)|^2 |\mathcal{L}(\phi)|^2 \, d\mu(\phi) \tag{1.14}$$

over the spectrum of the Laplacian acting on automorphic functions of level $\approx q$ (in the case of Theorem 1.1 the level equals $3[N, q]$) and given nebentypus, so that one of the terms corresponds to a cusp form $\phi \approx f$. Here $\mathcal{M}(\phi)$ is a suitable amplifier, and ϕ runs through Maass cusp forms, holomorphic cusp forms, and Eisenstein series with respect to a certain spectral measure $d\mu(\phi)$ designed for Kuznetsov's trace formula. The amplifier is given by $\mathcal{M}(\phi) := \sum_\ell x(\ell)\lambda_\phi(\ell)$, where $(x(\ell))$ is a finite sequence of complex numbers depending only on f. Opening the square and using multiplicativity of Hecke eigenvalues, we are left with bounding a normalized average

$$\mathcal{Q}(\ell) := \frac{1}{q}\int_\phi \lambda_\phi(\ell) |\mathcal{L}(\phi)|^2 \, d\mu(\phi)$$

for ℓ less than a small power of q. We win once we can show $\mathcal{Q}(\ell) \ll \ell^{-\delta}$ for a suitable $\delta > 0$.

By Kuznetsov's trace formula, the spectral sum $\mathcal{Q}(\ell)$ can be transformed into a weighted sum of (twisted) Kloosterman sums, the weights being of the form $\chi(m)\chi(n)$, $\tau(m)\tau(n)$, $\overline{\lambda_g}(m)\lambda_g(n)$ in the cases of Theorems 1.1, 1.2, 1.3, respectively. The set of weights $\chi(m)\chi(n)$ is considerably simpler which is mainly responsible for the better value of δ. Here we follow the original treatment of Bykovskiĭ [By96] which expresses the sum in terms of the Hurwitz ζ-function. By applying the functional equation for these ζ-function, the problem reduces to cancellation in certain complete character sums, which is then established by Weil's theorem. The set of weights $\tau(m)\tau(n)$ can be regarded as a special case of $\overline{\lambda_g}(m)\lambda_g(n)$ upon defining

$$g(z) := \frac{\partial}{\partial s} E_\infty(z, s)_{|s=\frac{1}{2}} = 2\sqrt{y}\log(e^\gamma y/4\pi) + 4\sqrt{y}\sum_{n\geqslant 1}\tau(n)\cos(2\pi n x)K_0(2\pi n y). \tag{1.15}$$

Note, however, that this g is not square-integrable, which causes technical complications and necessitates a separate treatment. At any rate, the next step in the proof of Theorems 1.2 and 1.3 is an application of Voronoi summation which turns the Kloosterman sums into simpler Gauss sums (plus a negligible term in the case of (1.15)). Opening the Gauss sums, we are left with sums roughly of the type

$$\frac{1}{q^{3/2}}\sum_h \chi_f\chi_g(h) \sum_{\ell_1 m - \ell_2 n = h} \overline{\lambda_g}(m)\lambda_g(n) W_{\ell_1,\ell_2}(m, n). \tag{1.16}$$

Here the sizes of h, m, n are $\approx q$, the weight function W_{ℓ_1,ℓ_2} is nice and depends mildly on ℓ_1, ℓ_2.

The innermost sum in (1.16) is a *shifted convolution sum* which at best exhibits square-root cancellation, hence we need to exploit oscillation in the h-parameter. To understand the h-dependence we analyze the shifted convolution sum by Kloosterman's refinement of the circle method. This approach is very appropriate: it worked efficiently in earlier related contexts [DFI93, DFI94a, Ju99, KMV02], and in fact a special case of Kloosterman's original application [Kl26] can be regarded as a special case of the problem at hand. More precisely, for technical reasons, we employ the variants of the circle method developed by Meurman [Me01] and Jutila [Ju92, Ju96]. As a result, the shifted convolution sum equals (up to negligible error) a main term plus a weighted c-sum of (untwisted) Kloosterman sums $S(h,h';c)$. The weights are defined in terms of the coefficients $\lambda_g(n)$, but in the end we only need that these are small in L^2-mean. The main term is present only for (1.15), we return to it later below. For the sum of Kloosterman sums we apply Kuznetsov's trace formula *in the other direction* in order to separate the h and h' variables. Now we encounter expressions of the type

$$\int_\psi \sum_h \chi(h)\rho_\psi(h)\,d\tilde{\mu}(\psi), \tag{1.17}$$

where the h-sum is smooth of length $\approx q$, and ψ runs through modular forms of levels $\approx \ell_1\ell_2$ and trivial nebentypus with respect to another spectral measure $d\tilde{\mu}(\psi)$. Cancellation in the h-sum is therefore equivalent to subconvexity of twisted automorphic L-functions for which we need Theorem 1.1. Some difficulties arise from the fact that (1.16) may be "ill-posed": if the support of W_{ℓ_1,ℓ_2} is such that m is much smaller than n, we have to solve an unbalanced shifted convolution problem which is reflected by the fact that the ψ-integral in (1.17) is "long". In this case the saving comes from the spectral large sieve inequalities of Deshouillers–Iwaniec [DI82].

In the case of (1.15), i.e. when $\lambda_g(n)=\tau(n)$, an extra term appears in the analysis of (1.16), namely the contribution of the main term of the shifted convolution sums. This extra term equals (up to admissible error) the contribution of the Eisenstein spectrum in (1.14) which is generally too large and is included only to make (1.14) spectrally complete. In [DFI02] the analogue of this observation is justified rigorously: the two large contributions are proved to be equal, so one can forget about both of them. In the proof of Theorem 1.2 we take a shortcut instead. We arrange the weight functions in the approximate functional equation and in Kuznetsov's trace formula in such a way that the extra term becomes negligible: in the analysis this manifests as destroying a certain pole by creating a zero artificially. In fact, our choice of the approximate functional equation can be explained as by forcing the Eisenstein contribution in (1.16) to be small, see Remark 4.1.

Finally we remark that there is a more direct and more powerful method resulting in a similar spectral expansion of shifted convolution sums, see [BlHa08b, BlHa10] and the references therein. This method avoids the double application of Kuznetsov's trace formula, but at the time of working on these projects it was limited to special situations such as holomorphic g or unbalanced shifted convolution sums (i.e. when the sizes of h, m, n are not approximately equal).

Chapter 2

Review of automorphic forms

2.1 Maass forms

Let k and D be positive integers, and χ be a character of modulus D such that $\chi(-1) = (-1)^k$. An automorphic function of weight k, level D and nebentypus χ is a function $g : \mathcal{H} \to \mathbb{C}$ 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

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

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

$$g(z) = \sum_{j\geqslant 0}\langle g, u_j\rangle u_j(z) + \sum_{\mathfrak{a}} \frac{1}{4\pi i}\int_{\Re s = \frac{1}{2}} \langle g, E_{\mathfrak{a}}(*,s)\rangle E_{\mathfrak{a}}(z,s)\, ds, \tag{2.1}$$

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

$$g(z) = \sum_{\substack{n=-\infty \\ n\neq 0}}^{+\infty} \rho_g(n) W_{\frac{n}{|n|}\frac{k}{2}, it}(4\pi|n|y)e(nx), \tag{2.2}$$

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

$$E_{\mathfrak{a}}\left(z,\tfrac{1}{2}+it\right)=\delta_{\mathfrak{a}=\infty}y^{\frac{1}{2}+it}+\phi_{\mathfrak{a}}\left(\tfrac{1}{2}+it\right)y^{\frac{1}{2}-it}+\sum_{\substack{n=-\infty\\ n\neq 0}}^{+\infty}\rho_{\mathfrak{a}}(n,t)W_{\frac{n}{|n|}\frac{k}{2},it}(4\pi|n|y)e(nx), \tag{2.3}$$

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 $g:\mathcal{H}\to\mathbb{C}$ 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 Laplacian 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:

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

We denote by $\mathcal{B}_k^h(D,\chi)$ an orthonormal basis of the space of holomorphic cusp forms of weight $k\geqslant 1$, level D and nebentypus χ.

In the sequel, we set

$$\mu_g:=1+|t_g|;\qquad t_g:=\begin{cases}\sqrt{\lambda_g-1/4} & \text{when } g \text{ is a Maass cusp form of eigenvalue } \lambda_g;\\ i(k_g-1)/2 & \text{when } g \text{ is a holomorphic cusp form of weight } k_g.\end{cases} \tag{2.4}$$

2.3 Hecke operators and Hecke eigenbases

We 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

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

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 we can choose $\mathcal{B}_k(D,\chi)$ and $\mathcal{B}_k^h(D,\chi)$ to consist of Hecke eigenforms. Then, by Atkin–Lehner theory, these orthogonal bases contain a unique scalar multiple of any primitive form.

The adelic reformulation of the theory of modular forms provides a natural alternate spectral expansion of the Eisenstein spectrum $\mathcal{E}_k(D,\chi)\subset\mathcal{L}_k(D,\chi)$. In this expansion, the basis is indexed by a set of parameters of the form[1]

$$\{(\chi_1,\chi_2,f)\mid\chi_1\chi_2=\chi,\ f\in\mathcal{B}_k(\chi_1,\chi_2)\},\tag{2.5}$$

[1] We suppress here the independent spectral parameters $\frac{1}{2}+it$ with $t\in\mathbb{R}$.

where (χ_1, χ_2) ranges over the pairs of characters of modulus D such that $\chi_1\chi_2 = \chi$ and $\mathcal{B}_k(\chi_1, \chi_2)$ is some finite set depending on (χ_1, χ_2). Specifically, $\mathcal{B}_k(\chi_1, \chi_2)$ corresponds to an orthonormal basis in the induced representation constructed out of the pair (χ_1, χ_2), see [GJ79] for more details. For $g \in \mathcal{E}_k(D, \chi)$ one has

$$g(z) = \sum_{\substack{\chi_1\chi_2=\chi \\ f\in\mathcal{B}_k(\chi_1,\chi_2)}}\sum \frac{1}{4\pi i}\int_{\Re s=\frac{1}{2}} \langle g, E_{\chi_1,\chi_2,f}(*, s)\rangle E_{\chi_1,\chi_2,f}(z, s)\, ds. \tag{2.6}$$

An important feature of this basis is that it consists of Hecke eigenforms for $\mathbf{T}^{(D)}$: for $(n, D) = 1$ one has

$$T_n E_{\chi_1,\chi_2,f}\big(z, \tfrac{1}{2} + it\big) = \lambda_{\chi_1,\chi_2}(n, t) E_{\chi_1,\chi_2,f}\big(z, \tfrac{1}{2} + it\big)$$

with

$$\lambda_{\chi_1,\chi_2}(n, t) = \sum_{ab=n} \chi_1(a)a^{it}\chi_2(b)b^{-it}. \tag{2.7}$$

We shall abbreviate $E_{\chi_1,\chi_2,f}\big(*, \frac{1}{2} + it\big)$ by $E_{\chi_1,\chi_2,f,t}$, and denote its Fourier coefficients by $\rho_f(n, t)$.

2.4 Hecke eigenvalues and Fourier coefficients

Let g be any Hecke eigenform with eigenvalue $\lambda_g(n)$ for T_n, then one has

$$\lambda_g(m)\lambda_g(n) = \sum_{d|(m,n)} \psi(d)\lambda_g(mn/d^2) \quad \text{for } (mn, D) = 1, \tag{2.8}$$

$$\overline{\lambda_g(n)} = \overline{\psi(n)}\lambda_g(n) \quad \text{for } (n, D) = 1.$$

In particular, it follows that

$$\lambda_g(m)\overline{\lambda_g(n)} = \overline{\psi(n)} \sum_{d|(m,n)} \psi(d)\lambda_g(mn/d^2) \quad \text{for } (mn, D) = 1. \tag{2.9}$$

There is a close relationship between the Fourier coefficients $\rho_g(n)$ and the Hecke eigenvalues $\lambda_g(n)$:

$$\sqrt{n}\rho_g(\pm n) = \rho_g(\pm 1)\lambda_g(n) \quad \text{for } (n, D) = 1, \tag{2.10}$$

$$\sqrt{m}\rho_g(m)\lambda_g(n) = \sum_{d|(m,n)} \chi(d)\rho_g\left(\frac{m}{d}\frac{n}{d}\right)\sqrt{\frac{mn}{d^2}} \quad \text{for } (n, D) = 1, \tag{2.11}$$

$$\sqrt{mn}\rho_g(mn) = \sum_{d|(m,n)} \chi(d)\mu(d)\rho_g\left(\frac{m}{d}\right)\sqrt{\frac{m}{d}}\lambda_g\left(\frac{n}{d}\right) \quad \text{for } (n, D) = 1. \tag{2.12}$$

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 (2.10) and (2.11) 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:

$$\rho_g(-n) = \varepsilon_g\rho_g(n) \quad \text{for } n \geqslant 1 \tag{2.13}$$

with

$$\varepsilon_g = \pm\frac{\Gamma\big(\frac{1}{2} + it_g + \frac{k}{2}\big)}{\Gamma\big(\frac{1}{2} + it_g - \frac{k}{2}\big)} \tag{2.14}$$

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

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

2.5 Spectral summation formulae

The following spectral summation formulae form an important tool for the analytic theory of modular forms. Let $\chi(-1) = (-1)^\kappa$ with $\kappa \in \{0,1\}$, and recall that $\mathcal{B}_k(D,\chi)$ (resp. $\mathcal{B}_k^h(D,\chi)$) denotes an orthonormal Hecke eigenbasis of the space of Maass (resp. holomorphic) cusp forms of weight $k \equiv \kappa\,(2)$, level D and nebentypus χ. The first formula is due to Petersson (cf. Theorem 9.6 in [Iw02]):

Proposition 2.1. *For any positive integers m,n, one has*

$$4\pi\Gamma(k-1)\sqrt{mn}\sum_{f\in\mathcal{B}_k^h(D,\chi)}\overline{\rho_f}(m)\rho_f(n) = \delta_{m,n} + 2\pi i^{-k}\sum_{c\equiv 0\,(D)}\frac{S_\chi(m,n;c)}{c}J_{k-1}\left(\frac{4\pi\sqrt{mn}}{c}\right). \tag{2.15}$$

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 $\mathcal{B}_k(D,\chi) = \{u_j\}_{j\geqslant 1}$ with u_j of Laplacian eigenvalue $\lambda_j = \frac{1}{4}+t_j^2$ and Fourier coefficients $\rho_j(n)$. The following result 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 2.2. *For any integer $k \geqslant 0$ and any $A > 0$, there exist functions $\mathcal{H}(t) : \mathbb{R}\cup i\mathbb{R} \to (0,\infty)$ and $\mathcal{I}(x) : (0,\infty) \to \mathbb{R}\cup i\mathbb{R}$ depending on k and A such that*

$$\mathcal{H}(t) \gg (1+|t|)^{k-16}e^{-\pi|t|}; \tag{2.16}$$

for any integer $j \geqslant 0$,

$$x^j\mathcal{I}^{(j)}(x) \ll_{A,j} \left(\frac{x}{1+x}\right)^{A+1}(1+x)^{1+j}; \tag{2.17}$$

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 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 [DI82] (to arbitrary levels) and by Proskurin [Pr05] (to arbitrary integral and half-integral weights). See [Iw02, Theorems 9.4–9.8][2], and also [CoPS90] for an illuminating discussion from the representation theoretic point of view. In order to state Kuznetsov's sum formula, we define the following Bessel transforms for $\varphi \in C^\infty(\mathbb{R}^+)$:

$$\dot{\varphi}(k) := i^k\int_0^\infty J_{k-1}(x)\varphi(x)\frac{dx}{x}; \tag{2.18}$$

$$\hat{\varphi}(t) := \frac{\pi i t^\kappa}{2\sinh(\pi t)}\int_0^\infty\left\{J_{2it}(x) - (-1)^\kappa J_{-2it}(x)\right\}\varphi(x)\frac{dx}{x}; \tag{2.19}$$

$$\check{\varphi}(t) := 2\cosh(\pi t)\int_0^\infty K_{2it}(x)\phi(x)\frac{dx}{x}. \tag{2.20}$$

[2]Note that in [Iw02] a few misprints occur: (9.15) should have the normalization factor $\frac{2}{\pi}$ instead of $\frac{4}{\pi}$, and in (B.49) a factor 4 is missing.

Theorem 2.1. *Let m, n, D be positive integers and $\varphi \in C^\infty(\mathbb{R}^+)$ such that $\varphi(0) = \varphi'(0) = 0$ and $\varphi^{(j)}(x) \ll_\varepsilon (1+x)^{-2-\varepsilon}$ for $0 \leqslant j \leqslant 3$. Then for $\kappa \in \{0,1\}$ one has*

$$\frac{1}{4\sqrt{mn}} \sum_{c \equiv 0\,(D)} \frac{S_\chi(m,n;c)}{c} \varphi\left(\frac{4\pi\sqrt{mn}}{c}\right) = \sum_{\substack{k \equiv \kappa\,(2) \\ k > \kappa}} \Gamma(k)\dot{\varphi}(k) \sum_{f \in \mathcal{B}_k^h(D,\chi)} \overline{\rho_f}(m)\rho_f(n)$$
$$+ \sum_{j \geqslant 1} \frac{\hat{\varphi}(t_j)}{\cosh(\pi t_j)} \overline{\rho_j}(m)\rho_j(n) + \frac{1}{4\pi} \sum_{\mathfrak{a}} \int_{-\infty}^{+\infty} \frac{\hat{\varphi}(t)}{\cosh(\pi t)} \overline{\rho_{\mathfrak{a}}}(m,t)\rho_{\mathfrak{a}}(n,t)\,dt. \quad (2.21)$$

In addition, for $\kappa = 0$ one has

$$\frac{1}{4\sqrt{mn}} \sum_{c \equiv 0\,(D)} \frac{S_\chi(m,-n;c)}{c} \varphi\left(\frac{4\pi\sqrt{mn}}{c}\right) =$$
$$\sum_{j \geqslant 1} \frac{\check{\varphi}(t_j)}{\cosh(\pi t_j)} \overline{\rho_j}(m)\rho_j(-n) + \frac{1}{4\pi} \sum_{\mathfrak{a}} \int_{-\infty}^{+\infty} \frac{\check{\varphi}(t)}{\cosh(\pi t)} \overline{\rho_{\mathfrak{a}}}(m,t)\rho_{\mathfrak{a}}(-n,t)\,dt. \quad (2.22)$$

In both identities the $(\mathfrak{a}, t)$-integral is over the Eisenstein spectrum $\mathcal{E}_k(D,\chi) \subset \mathcal{L}_k(D,\chi)$.

Remark 2.1. In (2.21) and (2.22) the sum over the singular cusps $\mathfrak{a}$ can be replaced by a sum over the parameters (2.5), then accordingly $\rho_{\mathfrak{a}}(*, t)$ need to be replaced by $\rho_f(*, t)$. The proof is identical, except that the sum in (2.6) plays the role of the second sum in (2.1).

It will be useful to have bounds for the Bessel transforms occurring in Theorem 2.1.

Lemma 2.1. *Let $\varphi(r)$ be a smooth function, compactly supported in $(R, 18R)$, satisfying*

$$\varphi^{(j)}(r) \ll_j (W/R)^j$$

for some $W \geqslant 1$ and for any $j \in \mathbb{N}_0$. Then, for $t \geqslant 0$ and for any $k > 1$, one has

$$\hat{\varphi}(it),\ \check{\varphi}(it) \ll \frac{1+(R/W)^{-2t}}{1+R/W} \qquad \text{for } 0 \leqslant t < \frac{1}{4}; \quad (2.23)$$
$$\dot{\varphi}(t),\ \hat{\varphi}(t),\ \check{\varphi}(t) \ll \frac{1+|\log(R/W)|}{1+R/W} \qquad \text{for } t \geqslant 0; \quad (2.24)$$
$$\dot{\varphi}(t),\ \hat{\varphi}(t),\ \check{\varphi}(t) \ll \left(\frac{W}{t}\right)\left(\frac{1}{t^{1/2}} + \frac{R}{t}\right) \qquad \text{for } t \geqslant 1; \quad (2.25)$$
$$\dot{\varphi}(t),\ \hat{\varphi}(t),\ \check{\varphi}(t) \ll_k \left(\frac{W}{t}\right)^k\left(\frac{1}{t^{1/2}} + \frac{R}{t}\right) \qquad \text{for } t \geqslant \max(10R, 1). \quad (2.26)$$

Proof. The inequalities (2.23), (2.24), (2.25) can be proved exactly as (7.1), (7.2) and (7.3) in [DI82]. The last inequality (2.26) 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 (2.25) is stronger). For this reason we include a detailed proof of (2.26) in the case of $\check{\varphi}(t)$. For $\hat{\varphi}(t)$ and $\dot{\varphi}(t)$ 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

$$\check{\varphi}(t) = (D_t\varphi)^\vee(t), \quad (2.27)$$

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 (2.27) 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 (2.25) and obtain
$$\check{\varphi}(t) \ll_j \left(\frac{W}{t}\right)^{2j+1}\left(\frac{1}{t^{1/2}} + \frac{R}{t}\right) \quad \text{for} \quad t \geqslant \max(10R, 1).$$
□

Lemma 2.2. *a) Let $\varphi(x)$ be a smooth function supported on $x \asymp X$ such that $\varphi^{(j)}(x) \ll_j X^{-j}$ for all $j \in \mathbb{N}_0$. For $t \in \mathbb{R}$ we have*
$$\dot{\varphi}(t),\ \hat{\varphi}(t),\ \check{\varphi}(t) \ll_C \frac{1 + |\log X|}{1 + X}\left(\frac{1+X}{1+|t|}\right)^C$$
for any constant $C \geqslant 0$. Here the Bessel transform $\hat{\varphi}$ is taken with respect to $\kappa = 0$.

b) Let $\varphi(x)$ be a smooth function supported on $x \asymp X$ such that $\varphi^{(j)}(x) \ll_j (X/Z)^{-j}$ for all $j \in \mathbb{N}_0$. For $t \in (-i/4, i/4)$ we have
$$\hat{\varphi}(t),\ \check{\varphi}(t) \ll \frac{1 + (X/Z)^{-2|\Im t|}}{1 + X/Z}.$$
Here the Bessel transform $\hat{\varphi}$ is taken with respect to $\kappa = 0$.

c) Assume that $\varphi(x) = e^{iax}\psi(x)$ for some constant a and some smooth function $\psi(x)$ supported on $x \asymp X$ such that $\psi^{(j)}(x) \ll X^{-j}$ for all $j \in \mathbb{N}_0$. Assume $aX \geqslant 1$, $t \in \mathbb{R}$, and assume $t \in \mathbb{N}$ in the case of $\dot{\varphi}$. Then
$$\dot{\varphi}(t),\ \hat{\varphi}(t),\ \check{\varphi}(t) \ll_{C,\varepsilon} \frac{1}{F^{1-\varepsilon}}\left(\frac{F}{1+|t|}\right)^C$$
for any $C \geqslant 0$, $\varepsilon > 0$ and some $F = F(X, a) < (a+1)X$.

Proof. Parts a) and b) are covered by Lemma 2.1. Part c) is [Ju99, pp. 43–45]. □

Using [GR07, 8.403.1] we can express the kernel $k_t(x) := J_{2it}(x) - (-1)^\kappa J_{-2it}(x)$ in (2.19) as
$$\begin{aligned} J_{2it}(x) - J_{-2it}(x) &= i\tanh(\pi t)\left\{Y_{2it}(x) + Y_{-2it}(x)\right\} \\ J_{2it}(x) + J_{-2it}(x) &= i\coth(\pi t)\left\{Y_{2it}(x) - Y_{-2it}(x)\right\}. \end{aligned}$$
For future reference we shall recast $\hat{\varphi}$ as follows. By [GR07, 6.561.14] the Mellin transform of the kernel equals
$$\begin{aligned} \widetilde{k_t}(s) &= \int_0^\infty k_t(x)x^{s-1}dx \\ &= \frac{2^{s-1}}{\pi}\Gamma\left(\frac{s}{2} + it\right)\Gamma\left(\frac{s}{2} - it\right)\left\{\sin\left(\pi\left(\frac{s}{2} - it\right)\right) - (-1)^\kappa \sin\left(\pi\left(\frac{s}{2} + it\right)\right)\right\}. \end{aligned}$$
Let
$$\varphi^*(u) := \widetilde{\varphi}(-1-2u)2^{1+2u}. \tag{2.28}$$
Then by Plancherel's formula
$$\begin{aligned} \hat{\varphi}(t) &= \frac{\pi i t^\kappa}{2\sinh(\pi t)}\frac{1}{2\pi i}\int_{(\sigma)} \varphi^*(u)\widetilde{k_t}(1+2u)2^{-2u}\,du \qquad (2.29) \\ &= \frac{1}{\pi}\left\{\begin{array}{l} 1 \\ it\coth(\pi t) \end{array}\right\}\frac{1}{2\pi i}\int_{(\sigma)} \varphi^*(u)\Gamma\left(\frac{1}{2} + u + it\right)\Gamma\left(\frac{1}{2} + u - it\right)\left\{\begin{array}{l} -\sin(\pi u) \\ +\cos(\pi u) \end{array}\right\}du, \end{aligned}$$
where $-\frac{1}{2} + |\Im t| < \sigma < 0$, and the upper (resp. lower) line refers to $\kappa = 0$ (resp. $\kappa = 1$).

2.6 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

$$D\left(g,\frac{a}{c},s\right) = i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\left\{\Psi^+_{k,it}(s)D\left(g,-\frac{\overline{a}}{c},1-s\right)+\Psi^-_{k,it}(s)D\left(Qg,\frac{\overline{a}}{c},1-s\right)\right\}, \tag{2.30}$$

where $\Psi^\pm_{k,it}(s)$ are meromorphic functions depending at most on k and it, $\frac14+t^2$ is the Laplacian eigenvalue of g, and $Q = Q_{\frac12+it,k}$ 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

$$D\left(g,\frac{a}{c},s\right) = i^k\chi(\overline{a})\left(\frac{c}{\pi}\right)^{1-2s}\left\{\Psi^+_{k,it}(s)D\left(g,-\frac{\overline{a}}{c},1-s\right)+\varepsilon\Psi^-_{k,it}(s)D\left(g,\frac{\overline{a}}{c},1-s\right)\right\}. \tag{2.31}$$

For $k=0$, $\Psi^\pm_{k,it}(s)$ are determined in Appendix A.4 of [KMV02] (see also [Me88]):

$$\Psi^\pm_{0,it}(s) = \frac{\Gamma\left(\frac{1-s+it}{2}\right)\Gamma\left(\frac{1-s-it}{2}\right)}{\Gamma\left(\frac{s-it}{2}\right)\Gamma\left(\frac{s+it}{2}\right)} \mp \frac{\Gamma\left(\frac{2-s+it}{2}\right)\Gamma\left(\frac{2-s-it}{2}\right)}{\Gamma\left(\frac{1+s-it}{2}\right)\Gamma\left(\frac{1+s+it}{2}\right)}. \tag{2.32}$$

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]:

$$\Phi^{\pm 1}_k(s,it) := \frac{\sqrt{\pi}}{4}\int_0^\infty\left\{W_{\frac{k}{2},it}(4y)\pm\frac{\Gamma\left(\frac12+it+\frac{k}{2}\right)}{\Gamma\left(\frac12+it-\frac{k}{2}\right)}W_{-\frac{k}{2},it}(4y)\right\}y^{s-\frac12}\frac{dy}{y}. \tag{2.33}$$

Our starting point for establishing the functional equation is the identity

$$\frac{\pi^s}{4}\int_0^\infty g(x+iy)y^{s-\frac12}\frac{dy}{y} = \Phi^\varepsilon_k(s,it)D^{+1}(g,x,s)+\Phi^{-\varepsilon}_k(s,it)D^{-1}(g,x,s), \tag{2.34}$$

where

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

In deriving this identity we use (2.2), (2.13), and (2.14) 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 (2.34),

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

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

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

Using that $D^{\pm 1}(g,-x,s)=\pm D^{\pm 1}(g,x,s)$, 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 (2.31) indeed holds with the following definition of $\Psi_{k,it}^{\pm}(s)$:

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

This formula works for $k\neq 0$ and complements (2.32) which corresponds to $k=0$.

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

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

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

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

Note that by (2.32) this formula is also valid for $k=0$.

We can simplify the above expressions for $\Psi_{k,it}^{\pm}(s)$ 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

$$\begin{aligned} \Psi_{k,it}^{+}(s) &= i^k\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\big\{-\cos(\pi s)\big\}; \\ \Psi_{k,it}^{-}(s) &= i^k\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\big\{\cos(\pi it)\big\}. \end{aligned} \tag{2.35}$$

For odd k, we obtain

$$\begin{aligned} \Psi_{k,it}^{+}(s) &= i^{k-1}\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\big\{\sin(\pi s)\big\}; \\ \Psi_{k,it}^{-}(s) &= i^{k-1}\pi^{-1}2^{2s}\Gamma(1-s+it)\Gamma(1-s-it)\big\{-\sin(\pi it)\big\}. \end{aligned} \tag{2.36}$$

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 [Me88, DI90, KMV02].

Proposition 2.3. *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 $F \in C^\infty(\mathbb{R}^+)$ is a Schwartz class function vanishing in a neighborhood of zero, then*

$$\sum_{n\geqslant 1} \sqrt{n}\rho_g(n) e\Big(n\frac{a}{c}\Big) F(n) = \frac{\chi(\overline{a})}{c} \sum_{\pm} \sum_{n\geqslant 1} \sqrt{n}\rho_g^{\pm}(n) e\Big(\mp n\frac{\overline{a}}{c}\Big) \mathcal{F}^{\pm}\Big(\frac{n}{c^2}\Big). \tag{2.37}$$

In this formula,

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

and

$$\mathcal{F}^{\pm}(y) := \int_0^\infty F(x) J_g^{\pm}\big(4\pi\sqrt{xy}\big)\,dx, \tag{2.38}$$

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;
- $$J_g^+(x) := \frac{-\pi}{\cosh(\pi t)} \big\{Y_{2it}(x) + Y_{-2it}(x)\big\}, \qquad J_g^-(x) := 4\cosh(\pi t) K_{2it}(x),$$
  if k is even, and g is not induced from a holomorphic form;
- $$J_g^+(x) := \frac{\pi}{\sinh(\pi t)} \big\{Y_{2it}(x) - Y_{-2it}(x)\big\}, \qquad J_g^-(x) := -4i\sinh(\pi t) K_{2it}(x),$$
  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 (2.37) as an inverse Mellin transform

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

By the functional equation (2.30), the right hand side can be rewritten as

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

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

$$\begin{aligned} &i^k \chi(\overline{a}) \frac{1}{2\pi i}\int_{(2)} \widetilde{F}\left(1-\frac{s}{2}\right) \left(\frac{c}{\pi}\right)^{s-1} \Psi_{k,it}^+\left(1-\frac{s}{2}\right) D\left(g, -\frac{\overline{a}}{c}, \frac{s}{2}\right) \frac{ds}{2} \\ &+ i^k \chi(\overline{a}) \frac{1}{2\pi i}\int_{(2)} \widetilde{F}\left(1-\frac{s}{2}\right) \left(\frac{c}{\pi}\right)^{s-1} \Psi_{k,it}^-\left(1-\frac{s}{2}\right) D\left(Qg, \frac{\overline{a}}{c}, \frac{s}{2}\right) \frac{ds}{2}. \end{aligned} \tag{2.39}$$

Using (2.35) and (2.36) it is straightforward to check that

$$i^k \Psi_{k,it}^{\pm}\left(1-\frac{s}{2}\right) = \frac{2}{\pi} \widetilde{J_g^{\pm}(4x)}(s),$$

so that

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

where $\mathcal{F}^{\pm}$ is the Hankel-type transform of F given by (2.38). In particular,

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

and this shows that (2.39) is equal to the right hand side of (2.37). But (2.39) is also equal to the left hand side of (2.37), therefore the proof is complete.

2.7 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 [Mi04].

If g is an L^2-normalized primitive Maass cusp form of level D, weight $\kappa \in \{0,1\}$ and eigenvalue $\frac{1}{4}+t_g^2$, then from [DFI02] and [HL94] we have for any $\varepsilon > 0$ (cf. (2.4)),

$$(D\mu_g)^{-\varepsilon}\left(\frac{\cosh(\pi t_g)}{D\mu_g^{\kappa}}\right)^{1/2} \ll_\varepsilon |\rho_g(1)| \ll_\varepsilon (D\mu_g)^{\varepsilon}\left(\frac{\cosh(\pi t_g)}{D\mu_g^{\kappa}}\right)^{1/2}. \tag{2.40}$$

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

$$\frac{(Dk)^{-\varepsilon}}{(D\Gamma(k))^{1/2}} \ll_\varepsilon |\rho_g(1)| \ll_\varepsilon \frac{(Dk)^{\varepsilon}}{(D\Gamma(k))^{1/2}}. \tag{2.41}$$

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

$$|\lambda_g(n)| \leqslant \tau(n)n^{\theta}. \tag{2.42}$$

Note that $\theta = \frac{7}{64}$ is admissible by the work of Kim–Shahidi, Kim and Kim–Sarnak [KiSh02, Ki03, KiSa03], and (2.42) holds even when n is divisible by ramified primes. Moreover, if g is holomorphic, it follows from Deligne's proof of the Ramanujan–Petersson conjecture that (2.42) holds with $\theta = 0$. Hence for all $n \geqslant 1$ and for any $\varepsilon > 0$ we have by (2.10)

$$\sqrt{n}\rho_g(n) \ll_\varepsilon \begin{cases} (nD\mu_g)^{\varepsilon}\left(\dfrac{\cosh(\pi t_g)}{D\mu_g^{\kappa}}\right)^{1/2} n^{\theta} & \text{for } \ g \in \mathcal{L}_\kappa(D,\chi),\ \kappa \in \{0,1\}; \\ \dfrac{(nDk)^{\varepsilon}}{(D\Gamma(k))^{1/2}} & \text{for } \ g \in \mathcal{S}_k(D,\chi). \end{cases} \tag{2.43}$$

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 $X \geqslant 1$ and all $\varepsilon > 0$,

$$\sum_{n \leqslant X} |\lambda_g(n)|^2 \ll_\varepsilon (D\mu_g X)^{\varepsilon} X. \tag{2.44}$$

In several occasions, we will need a substitute for (2.43) 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. By a straightforward generalization of [Mi04, Lemma 2.3] we have

Lemma 2.3. *Assume Hypothesis H_θ. For $k \geqslant 1$ let $\mathcal{B}_k^h(D,\chi) \subset \mathcal{S}_k(D,\chi)$ and for $\kappa \in \{0,1\}$ let $\mathcal{B}_\kappa(D,\chi) \subset \mathcal{L}_\kappa(D,\chi)$ denote orthonormal Hecke eigenbases. Then for $n, X \geqslant 1$, one has*

$$\sum_{\substack{k \equiv \kappa\,(2) \\ 2 \leqslant k \leqslant X}} \Gamma(k) \sum_{g \in \mathcal{B}_k^h(D,\chi)} n|\rho_g(n)|^2 + \sum_{\substack{g \in \mathcal{B}_\kappa(D,\chi) \\ |t_g| \leqslant X}} \frac{n|\rho_g(n)|^2}{\cosh(\pi t_j)} \ll (nDX)^{\varepsilon} X^2 n^{2\theta}, \tag{2.45}$$

where the implied constants depend at most on ε.

Finally we state the large sieve inequalities [DI82, Theorem 2] for modular forms of level D and trivial nebentypus.

Proposition 2.4. *Let $\mathcal{B}_k^h(D,1) \subset \mathcal{S}_k(D,1)$ and $\mathcal{B}_\kappa(D,1) \subset \mathcal{L}_\kappa(D,1)$ denote orthonormal bases. Let*

$N, X \geqslant 1$ and let (a_n) be an arbitrary sequence of complex numbers. Then

$$\left.\begin{aligned}&\sum_{\substack{k\equiv 0\,(2)\\ 2\leqslant k\leqslant X}}\Gamma(k)\sum_{f\in\mathcal{B}_k(D,1)}\Bigl|\sum_{N\leqslant n<2N}a_n\sqrt{n}\rho_f(n)\Bigr|^2\\ &\sum_{\substack{f\in\mathcal{B}_\kappa(D,1)\\ |t_f|\leqslant X}}\frac{1}{\cosh(\pi t_f)}\Bigl|\sum_{N\leqslant n<2N}a_n\sqrt{n}\rho_f(n)\Bigr|^2\\ &\sum_{\mathfrak{a}}\int_{-X}^{X}\frac{1}{\cosh(\pi t)}\Bigl|\sum_{N\leqslant n<2N}a_n\sqrt{n}\rho_{\mathfrak{a}}(n,t)\Bigr|^2\,dt\end{aligned}\right\}\ll\Bigl(X^2+\frac{N^{1+\varepsilon}}{D}\Bigr)\sum_{N\leqslant n<2N}|a_n|^2, \tag{2.46}$$

where the $(\mathfrak{a},t)$-integral is over the Eisenstein spectrum $\mathcal{E}_0(D,1)\subset\mathcal{L}_0(D,1)$. The implied constant depends at most on ε.

2.8 Bounds for exponential sums associated to cusp forms

In this section we prove uniform bounds for exponential sums

$$S_g(\alpha, X) := \sum_{n\leqslant X}\lambda_g(n)e(n\alpha) \tag{2.47}$$

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

Proposition 2.5. *Let g be a primitive Maass cusp form of level D, weight $\kappa\in\{0,1\}$ and Laplacian eigenvalue $\frac{1}{4}+t_g^2$. Then we have, uniformly for $X\geqslant 1$ and $\alpha\in\mathbb{R}$,*

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

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. See [BlHo10, T10, HT11] for recent developments in the case of square-free D.

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

$$g(x+iy)=\sum_{n\geqslant 1}\rho_g(n)\bigl\{W_{\frac{\kappa}{2},it}(4\pi ny)e(nx)+\varepsilon_gW_{-\frac{\kappa}{2},it}(4\pi ny)e(-nx)\bigr\}, \tag{2.48}$$

where $\varepsilon_g=\pm(it)^\kappa$ is the constant in (2.14). The Whittaker functions here can be expressed explicitly from K-Bessel functions:

$$\begin{aligned}W_{0,it}(4y)&=\frac{2y^{1/2}}{\sqrt{\pi}}K_{it}(2y);\\ W_{\frac12,it}(4y)&=\frac{2y}{\sqrt{\pi}}\bigl\{K_{\frac12+it}(2y)+K_{\frac12-it}(2y)\bigr\};\\ itW_{-\frac12,it}(4y)&=\frac{2y}{\sqrt{\pi}}\bigl\{K_{\frac12+it}(2y)-K_{\frac12-it}(2y)\bigr\}.\end{aligned} \tag{2.49}$$

By the Cauchy–Schwarz inequality, we have

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

Combining this estimate with (2.10), (2.40), (2.44), (2.49) and the uniform bounds of Proposition 6.2, we can conclude that

$$y^{\varepsilon}g(x+iy) \ll_{\varepsilon} (D\mu_g)^{2\varepsilon}D^{-1/2}\mu_g y^{-1/2}. \tag{2.50}$$

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 $\beta\in\mathrm{SL}_2(\mathbb{Z})$ 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 $\delta=\begin{pmatrix}a & *\\ c & *\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z})$ 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 [Iw02]) 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

$$\Im w \gg c^2/[c^2,D]. \tag{2.51}$$

Observe that

$$|g(z)|=|g(\delta v)|=|g(\sigma_{\mathfrak{a}}w)|=|h(w)|, \tag{2.52}$$

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 (2.10), (2.40), (2.44) we use the uniform bound

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

This bound follows exactly as Lemma 19.3[3] in [DFI02] upon noting that $c_{\mathfrak{a}}$ for the cusp $\mathfrak{a}=a/c$ (see Section 2.6 of [Iw02]) is at least $[c,D/c]\geqslant 1$ (cf. Lemma 2.4 of [DI82]). The analogue of (2.50) 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 (2.51) and (2.52), this implies that

$$g(x+iy)\ll_{\varepsilon}(D\mu_g)^{\varepsilon}D^{1/2}\mu_g^{3/2}. \tag{2.53}$$

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 (2.50).

With the uniform bounds (2.50) and (2.53) at hand we proceed to estimate the exponential sums $S_g(\alpha,X)$. By applying Fourier inversion to (2.48), we obtain, for any $\alpha\in\mathbb{R}$,

$$\rho_g(n)\left\{W_{\frac{\kappa}{2},it}(4\pi ny)+\frac{\Gamma\bigl(\frac12+it+\frac{\kappa}{2}\bigr)}{\Gamma\bigl(\frac12+it-\frac{\kappa}{2}\bigr)}W_{-\frac{\kappa}{2},it}(4\pi ny)\right\}e(n\alpha)=$$
$$\int_0^1\bigl\{g(\alpha+\beta+iy)\pm g(-\alpha-\beta+iy)\bigr\}e(-n\beta)\,d\beta,$$

[3] 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 (2.49) and Proposition 6.2).

where the $\pm$ on the right hand side matches the one in (2.14). Then we integrate both sides against $(\pi y)^\varepsilon \frac{dy}{y}$ to see that

$$\frac{\lambda_g(n)e(n\alpha)}{n^{1/2+\varepsilon}} = \int_0^1 G_\alpha(\beta)e(-n\beta)\,d\beta, \tag{2.54}$$

where

$$G_\alpha(\beta) := \frac{\pi^{1/2+\varepsilon}}{4\rho_g(1)\Phi^1_\kappa\left(\frac{1}{2}+\varepsilon, it_g\right)} \int_0^\infty \bigl\{g(\alpha+\beta+iy) \pm g(-\alpha-\beta+iy)\bigr\} y^\varepsilon \frac{dy}{y}. \tag{2.55}$$

The function $\Phi^1_\kappa(s,it)$ is defined in (2.33), 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 (2.49). At any rate,

$$\Phi^1_\kappa\left(\tfrac{1}{2}+\varepsilon, it_g\right) \asymp \widetilde{K_{\frac{\kappa}{2}+it_g}}\left(\tfrac{1+\kappa}{2}+\varepsilon\right) \asymp \mu_g^{(\kappa-1)/2+\varepsilon}\cosh^{-1/2}(\pi t_g),$$

so that by (2.40) we also have

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

The integral in (2.55) is convergent by (2.50) and (2.53). 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

$$G_\alpha(\beta) \ll_\varepsilon (D\mu_g)^\varepsilon D\mu_g^2, \quad \alpha \in \mathbb{R}. \tag{2.56}$$

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 (2.54) that

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

Combining (2.56) 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 2.5.

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

Proposition 2.6. *Let g be a primitive holomorphic cusp form of level D and weight k. Then we have, uniformly for $X \geqslant 1$ and $\alpha \in \mathbb{R}$,*

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

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 a variant of Lemma 3 of [Ju96] (see also Lemma 3.2 of [Bl04]).

Lemma 2.4. *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, and for any $\varepsilon > 0$, one has*

$$\sum_{h\in\mathbb{Z}} \left| \sum_{\substack{m\leqslant X,\ n\leqslant Y \\ \ell_1 m \pm \ell_2 n = h}} \overline{\lambda_g}(m)\lambda_g(n) \right|^2 \ll_\varepsilon (D\mu_g XY)^\varepsilon D^2\mu_g^4 XY.$$

Proof. The estimate follows by combining Propositions 2.5–2.6 with the Parseval identity and the Rankin–Selberg bound (2.44):

$$\begin{aligned}
\sum_{h\in\mathbb{Z}}\Bigg|\sum_{\substack{m\leqslant X,\ n\leqslant Y\\ \ell_1 m\pm\ell_2 n=h}}\overline{\lambda_g}(m)\lambda_g(n)\Bigg|^2 &= \int_0^1\big|S_g(-\ell_1\alpha,X)S_g(\pm\ell_2\alpha,Y)\big|^2\,d\alpha\\
&\ll_\varepsilon (D\mu_g X)^\varepsilon D^2\mu_g^4 X\int_0^1\big|S_g(\pm\ell_2\alpha,Y)\big|^2\,d\alpha\\
&= (D\mu_g X)^\varepsilon D^2\mu_g^4 X\sum_{n\leqslant Y}|\lambda_g(n)|^2\\
&\ll_\varepsilon (D\mu_g XY)^{2\varepsilon}D^2\mu_g^4 XY.
\end{aligned}$$

□

Chapter 3

Twisted L-functions

3.1 Amplification

In the next three sections we give a proof of Theorem 1.1. The method is based on a paper by Bykovskiĭ [By96]. Let f_0 be a primitive (holomorphic or Maass) cusp form of Hecke eigenvalues $\lambda(n)$, archimedean parameter μ, level N and trivial nebentypus, and let χ be a primitive character modulo q for which we want to prove Theorem 1.1. We shall embed f_0 into the spectrum of $\Gamma_0(D)$ with trivial nebentypus, where D is an integer satisfying $[N, q] \mid D$ and $D > 2q$; we take

$$D := 3[N, q]. \tag{3.1}$$

More precisely, we shall choose the bases $\mathcal{B}_k^h(D, 1)$ and $\mathcal{B}_0(D, 1)$ described in Chapter 2 in such a way that one of them contains the L^2-normalized version of $f_0(z)$:

$$f_1(z) := \frac{f_0(z)}{\langle f_0, f_0\rangle_D} = \frac{f_0(z)}{[\Gamma_0(q) : \Gamma_0(D)]\,\langle f_0, f_0\rangle_q}.$$

Then (2.40) and (2.41)—applied for q in place of D—shows that

$$|\rho_{f_1}(1)|^2 \gg_\varepsilon \begin{cases} (\Gamma(k)D)^{-1}(kD)^{-\varepsilon}, & \text{for} \quad f_1 \in \mathcal{B}_k^h(D, 1), \\ \cosh(\pi\mu)D^{-1}(\mu D)^{-\varepsilon}, & \text{for} \quad f_1 \in \mathcal{B}_0(D, 1), \end{cases} \tag{3.2}$$

We shall consider an amplified square mean of the "fake" twisted L-functions[1]

$$\mathcal{L}(f \otimes \chi, s) := \sum_{n=1}^{\infty} \sqrt{n}\rho_f(n)\chi(n)n^{-s}$$

for f either in $\mathcal{B}_k^h(D, 1)$ or $\mathcal{B}_0(D, 1)$ and

$$\mathcal{L}(E_{\psi,\bar{\psi},f,t} \otimes \chi, s) := \sum_{n=1}^{\infty} \sqrt{n}\rho_f(n, t)\chi(n)n^{-s}$$

for ψ any character modulo D, $f \in \mathcal{B}_0(\psi, \bar{\psi})$ and $t \in \mathbb{R}$. The justification comes from (2.10): apart from invertible Euler factors at primes dividing D,

$$L(f_0 \otimes \chi, s) \approx \sum_{n=1}^{\infty} \lambda(n)\chi(n)n^{-s},$$

hence for $\Re s = \frac{1}{2}$ we have

$$|\mathcal{L}(f_1 \otimes \chi, s)| \gg_\varepsilon D^{-\varepsilon}|\rho_{f_1}(1)||L(f_0 \otimes \chi, s)|. \tag{3.3}$$

[1] [By96] considers true L-functions over the whole spectrum which is, technically speaking, incorrect as the spectrum includes old forms. Similarly, the "normalized orthonormal basis" considered at the bottom of [By96, p.925] is problematic as the first Fourier coefficient vanishes for old forms. We avoid these troubles by a more careful setup here and in Sections 2.3–2.4.

For integers $0 \leqslant b < a$ let us define

$$\varphi_{a,b}(x) := i^{b-a} J_a(x) x^{-b}. \tag{3.4}$$

In order to satisfy the decay conditions for Kuznetsov's trace formula, we assume $b \geqslant 2$. Let $\kappa \in \{0,1\}$ such that $a - b \equiv \kappa \pmod 2$. It is straightforward to verify, using [GR07, 6.574.2], that depending on κ we have

$$\begin{aligned}
\dot{\varphi}_{a,b}(k) &= \frac{b!}{2^{b+1}\pi} \prod_{j=0}^{b} \left\{ \left(\frac{(1-k)i}{2} \right)^2 + \left(\frac{a+b}{2} - j \right)^2 \right\}^{-1} \asymp_{a,b} \pm k^{-2b-2}, \\
\hat{\varphi}_{a,b}(t) &= \frac{b!}{2^{b+1}} \left\{ \begin{matrix} 1 \\ t\coth(\pi t) \end{matrix} \right\} \prod_{j=0}^{b} \left\{ t^2 + \left(\frac{a+b}{2} - j \right)^2 \right\}^{-1} \asymp_{a,b} (1+|t|)^{\kappa-2b-2}
\end{aligned} \tag{3.5}$$

with $\dot{\varphi}$ as in (2.18) and $\hat{\varphi}$ as in (2.19). In particular,

$$\begin{aligned}
&\dot{\varphi}_{a,b}(k) > 0 \quad \text{for} \quad 2 \leqslant k \leqslant a-b, \\
&\hat{\varphi}_{a,b}(t) > 0 \quad \text{for} \quad \text{all possible spectral parameters } t,
\end{aligned} \tag{3.6}$$

since $|\Im t| < \frac{1}{2}$ when $\kappa = 0$, and $t \in \mathbb{R}$ when $\kappa = 1$.

We choose

$$\varphi := \varphi_{20,2},$$

and for

$$\tau \in \mathbb{R}, \qquad u \in \mathbb{C}, \qquad k \in \{2,4,6,\dots\}, \qquad (\ell, D) = 1$$

we define the quantities

$$\mathcal{Q}_k^{\mathrm{holo}}(\ell) := 2i^k \Gamma(k-1) \sum_{f \in \mathcal{B}_k^h(D,1)} \lambda_f(\ell) \mathcal{L}(f \otimes \chi, u + i\tau) \overline{\mathcal{L}(f \otimes \chi, \overline{u} + i\tau)},$$

$$\begin{aligned}
\mathcal{Q}(\ell) := &\sum_{k \geqslant 2 \text{ even}} \dot{\varphi}(k) 2(k-1) i^{-k} \mathcal{Q}_k^{\mathrm{holo}}(\ell) \\
&+ \sum_{f \in \mathcal{B}_0(D,1)} \hat{\varphi}(t_f) \frac{4}{\cosh(\pi t_f)} \lambda_f(\ell) \mathcal{L}(f \otimes \chi, u + i\tau) \overline{\mathcal{L}(f \otimes \chi, \overline{u} + i\tau)} \\
&+ \sum_{\substack{\psi \bmod D \\ f \in \mathcal{B}_0(\psi, \bar{\psi})}} \sum \int_{-\infty}^{\infty} \hat{\varphi}(t) \frac{1}{\pi \cosh(\pi t)} \lambda_{\psi,\bar{\psi}}(\ell, t) \mathcal{L}(E_{\psi,\bar{\psi},f,t} \otimes \chi, u + i\tau) \overline{\mathcal{L}(E_{\psi,\bar{\psi},f,t} \otimes \chi, \overline{u} + i\tau)}\, dt,
\end{aligned}$$

with the notation (2.7) and (2.18)–(2.19).

For $u = \frac{1}{2} + \varepsilon$ and $k \geqslant 4$ we shall show in the next section

$$\begin{aligned}
\mathcal{Q}_k^{\mathrm{holo}}(\ell) &\ll_\varepsilon \left(\frac{1}{\sqrt{\ell}} + \left(\frac{\ell^{\frac14}(N,q)}{q^{\frac12} N^{\frac12}} + \frac{\ell^{\frac12}(N,q)^{\frac32}}{q^{\frac12} N} \right) \left(\frac{1+|\tau|}{k} + 1 \right) \right) ((1+|\tau|)D\ell)^\varepsilon, \\
\mathcal{Q}(\ell) &\ll_\varepsilon \left(\frac{1}{\sqrt{\ell}} + \left(\frac{\ell^{\frac14}(N,q)}{q^{\frac12} N^{\frac12}} + \frac{\ell^{\frac12}(N,q)^{\frac32}}{q^{\frac12} N} \right) (1+|\tau|) \right) ((1+|\tau|)D\ell)^\varepsilon,
\end{aligned} \tag{3.7}$$

with implied constants depending only on ε. Theorem 1.1 then follows by standard amplification: let us define the amplifier

$$x(\ell) := \begin{cases} \overline{\lambda(\ell)} & \text{for} \quad L \leqslant \ell \leqslant 2L, \;\; (\ell, D) = 1, \\ 0 & \text{else}, \end{cases} \tag{3.8}$$

where L is some parameter to be chosen in a moment. Let ω be a smooth cut-off function supported on $[1/2, 3]$. Then

$$\begin{aligned}
\sum_{\substack{(\ell,D)=1 \\ \ell \sim L}} |\lambda(\ell)|^2 &\gg_\omega \frac{1}{2\pi i} \int_{(2)} L^{(D)}(f_0 \otimes \overline{f_0}, s) \hat{\omega}(s) L^s ds \\
&\gg_\varepsilon L(q\mu D)^{-\varepsilon} + O_\varepsilon\left(q^\varepsilon (L\mu N)^{\frac12 + \varepsilon} \right),
\end{aligned}$$

where the superscript (D) indicates that the Euler factors of the Rankin–Selberg L-function at the primes dividing D have been omitted. The lower bound for the residue follows from [HL94], while the error term uses the standard (convexity) bounds for the symmetric square L-function on the line $\Re s = \frac{1}{2} + \varepsilon$. Therefore,

$$\sum_{\ell} x(\ell)\lambda(\ell) = \sum_{\substack{(\ell,D)=1\\ \ell\sim L}} |\lambda(\ell)|^2 \gg_\varepsilon L(LD)^{-\varepsilon}, \tag{3.9}$$

provided $L \geqslant q^\varepsilon(\mu N)^{1+\varepsilon}$. Assume first that f_0 is a Maass cusp form of weight zero or a holomorphic cusp form of weight 2. Then by (3.3), (3.2), (3.5) with $b = 2$, (3.6) and (3.9), we obtain

$$\begin{aligned}
&\frac{L^2(LD)^{-\varepsilon}}{\mu^{6+\varepsilon}D}\left|L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \ll_\varepsilon \\
&\qquad \sum_{\substack{k\geqslant 2 \text{ even}\\ f\in\mathcal{B}_k^h(D,1)}}\sum |\dot{\varphi}(k)| 4\Gamma(k)\left|\sum_{\ell} x(\ell)\lambda_f(\ell)\right|^2\left|\mathcal{L}\left(f\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \\
&\qquad + \sum_{f\in\mathcal{B}_0(D,1)} \hat{\varphi}(t_f)\frac{4}{\cosh(\pi t_f)}\left|\sum_{\ell} x(\ell)\lambda_f(\ell)\right|^2\left|\mathcal{L}\left(f\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \\
&\qquad + \sum_{\substack{\psi \bmod D\\ f\in\mathcal{B}_0(\psi,\bar{\psi})}}\sum \int_{-\infty}^{\infty}\hat{\varphi}(t)\frac{1}{\pi\cosh(\pi t)}\left|\sum_{\ell} x(\ell)\lambda_{\psi,\bar{\psi}}(\ell,t)\right|^2\left|\mathcal{L}\left(E_{\psi,\bar{\psi},f,t}\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2\,dt,
\end{aligned}$$

so that by (2.9) and (3.6)

$$\begin{aligned}
&\frac{L^2(LD)^{-\varepsilon}}{\mu^{6+\varepsilon}D}\left|L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \ll_\varepsilon \\
&\qquad \sum_{\ell_1,\ell_2}|x(\ell_1)x(\ell_2)|\sum_{d|(l_1,l_2)}\left\{\left|\mathcal{Q}\left(\frac{\ell_1\ell_2}{d^2}\right)\right| + \sum_{k\geqslant 20 \text{ even}} 4k|\dot{\varphi}_0(k)|\left|\mathcal{Q}_k^{\text{holo}}\left(\frac{\ell_1\ell_2}{d^2}\right)\right|\right\}.
\end{aligned}$$

Now we substitute (3.7). Note that the k-sum converges by (3.5). Changing the order of summation, we get the bound

$$\begin{aligned}
\ll_\varepsilon ((1+|\tau|)LD)^\varepsilon \Bigg\{ &\sum_d\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{-\frac{1}{2}}|x(d\ell_1)x(d\ell_2)| \\
&+ \frac{(1+|\tau|)(N,q)}{q^{\frac{1}{2}}N^{\frac{1}{2}}}\sum_d\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{\frac{1}{4}}|x(d\ell_1)x(d\ell_2)| \\
&+ \frac{(1+|\tau|)(N,q)^{\frac{3}{2}}}{q^{\frac{1}{2}}N}\sum_d\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{\frac{1}{2}}|x(d\ell_1)x(d\ell_2)|\Bigg\}.
\end{aligned}$$

In each term we have, by Cauchy–Schwarz ($a \in \mathbb{R}$),

$$\begin{aligned}
\sum_d\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^a|x(d\ell_1)x(d\ell_2)| &= \sum_d\left(\sum_\ell \ell^a|x(d\ell)|\right)^2 \leqslant \sum_d\left(\sum_{\ell\leqslant 2L}\ell^{2a}\right)\left(\sum_\ell|x(d\ell)|^2\right) \\
&= \left(\sum_{\ell\leqslant 2L}\ell^{2a}\right)\sum_\ell \tau(\ell)|x(\ell)|^2 \ll_a (1+L^{2a+1})\sum_\ell\tau(\ell)|x(\ell)|^2,
\end{aligned}$$

so that

$$\frac{L^2(LD)^{-\varepsilon}}{\mu^{6+\varepsilon}D}\left|L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \ll_\varepsilon$$
$$((1+|\tau|)LD)^\varepsilon\left(1+\frac{L^{\frac{3}{2}}(N,q)}{q^{\frac{1}{2}}N^{\frac{1}{2}}}(1+|\tau|)+\frac{L^2(N,q)^{\frac{3}{2}}}{q^{\frac{1}{2}}N}(1+|\tau|)\right)\sum_\ell \tau(\ell)|x(\ell)|^2,$$

This yields, by (3.1), (3.8) and (2.44),

$$\left|L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \ll_\varepsilon$$
$$\mu^6\left(\frac{qN}{L(N,q)}+L^{\frac{1}{2}}q^{\frac{1}{2}}N^{\frac{1}{2}}(1+|\tau|)+Lq^{\frac{1}{2}}(N,q)^{\frac{1}{2}}(1+|\tau|)\right)((1+|\tau|+\mu)Nq)^\varepsilon,$$

provided $L\geqslant q^\varepsilon(\mu N)^{1+\varepsilon}$. For such L, the second term in the parenthesis is dominated by the third one which motivates our choice

$$L:=\frac{q^{\frac{1}{4}}N^{\frac{1}{2}}}{(N,q)^{\frac{3}{4}}(1+|\tau|)^{\frac{1}{2}}}+q^\varepsilon(N\mu)^{1+\varepsilon}.$$

We obtain

$$L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\ll_\varepsilon$$
$$\mu^3\left((1+|\tau|)^{\frac{1}{4}}N^{\frac{1}{4}}q^{\frac{3}{8}}(N,q)^{-\frac{1}{8}}+(1+|\tau|)^{\frac{1}{2}}\mu^{\frac{1}{2}}N^{\frac{1}{2}}(N,q)^{\frac{1}{4}}q^{\frac{1}{4}}\right)((1+|\tau|+\mu)Nq)^\varepsilon.$$

By the functional equation and the Phragmén–Lindelöf convexity principle, we obtain Theorem 1.1 in the non-holomorphic case as well as in the case when f_0 is holomorphic of weight 2. Analogously, if f_0 is holomorphic of (even) weight $k\geqslant 4$, we get

$$\frac{L^2(kLD)^{-\varepsilon}}{kD}\left|L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\right|^2 \ll_\varepsilon \sum_{\ell_1,\ell_2}|x(\ell_1)x(\ell_2)|\sum_{d|(l_1,l_2)}\left|\mathcal{Q}_k^{\mathrm{holo}}\left(\frac{\ell_1\ell_2}{d^2}\right)\right|$$
$$\ll_\varepsilon ((1+|\tau|)LD)^\varepsilon\left(1+\frac{L^{\frac{3}{2}}(N,q)}{q^{\frac{1}{2}}N^{\frac{1}{2}}}\left(\frac{1+|\tau|}{k}+1\right)+\frac{L^2(N,q)^{\frac{3}{2}}}{q^{\frac{1}{2}}N}\left(\frac{1+|\tau|}{k}+1\right)\right)\sum_\ell \tau(\ell)|x(\ell)|^2,$$

provided $L\geqslant q^\varepsilon(kN)^{1+\varepsilon}$. Choosing

$$L:=\frac{q^{\frac{1}{4}}N^{\frac{1}{2}}k^{\frac{1}{2}}}{(N,q)^{\frac{3}{4}}(1+|\tau|+k)^{\frac{1}{2}}}+q^\varepsilon(kN)^{1+\varepsilon}$$

and using (3.1), (3.8) and (2.44), we obtain

$$L\left(f_0\otimes\chi,\frac{1}{2}+\varepsilon+i\tau\right)\ll_\varepsilon$$
$$k^{\frac{1}{2}}\left((|\tau|+k)^{\frac{1}{4}}k^{-\frac{1}{4}}N^{\frac{1}{4}}q^{\frac{3}{8}}(N,q)^{-\frac{1}{8}}+(|\tau|+k)^{\frac{1}{2}}N^{\frac{1}{2}}(N,q)^{\frac{1}{4}}q^{\frac{1}{4}}\right)((1+|\tau|)kNq)^\varepsilon.$$

This completes the proof of Theorem 1.1.

3.2 Variations on a theme of Bykovskiĭ

In order to show (3.7), we perform the following steps, cf. [By96, Section 5].

Step 0. For later purposes let us define, for $u, s \in \mathbb{C}$, $\tau, x \in \mathbb{R}$, $\eta_{1,2} \in \{\pm 1\}$ and $\varphi_{a,b}$ as in (3.4),

$$E_{u,\tau}^{\eta_1,\eta_2}(s) := \begin{cases} -\exp(\eta_1 \pi i(s/2+u)), & \text{for} \quad \eta_1 = \eta_2, \\ \exp(\eta_1 \pi \tau), & \text{for} \quad \eta_1 \neq \eta_2, \end{cases}$$

and

$$\Xi_{u,\tau}^{\eta_1,\eta_2}(x) := \frac{1}{2\pi i} \int_{(\sigma)} E_{u,\tau}^{\eta_1,\eta_2}(s) \Gamma\left(1 - \frac{s}{2} - u - i\tau\right) \Gamma\left(1 - \frac{s}{2} - u + i\tau\right) \widetilde{\varphi}_{a,b}(s) 2^{1-s} x^{-\frac{s}{2}} \, ds. \tag{3.10}$$

The Mellin transform of $\varphi_{a,b}$ equals [GR07, 6.561.14]

$$\widetilde{\varphi}_{a,b}(s) = i^{b-a} 2^{s-b-1} \Gamma\left(\frac{a-b+s}{2}\right) \left(\Gamma\left(\frac{2+a+b-s}{2}\right)\right)^{-1}. \tag{3.11}$$

Thus the integrand in (3.10) is holomorphic and by Stirling's formula the integral converges absolutely if

$$b - a \; < \; \sigma \; < \; 2 - 2\Re u \; < \; 1 + b. \tag{3.12}$$

Moreover, in this range we have, uniformly in a, τ, and $\Im u$,

$$\Xi_{u,\tau}^{\eta_1,\eta_2}(x) \ll_{b,\sigma,\Re u} x^{-\frac{\sigma}{2}} \int_{-\infty}^{\infty} (a + |t|)^{\sigma-1-b} \left\{ \left(1 + \left|\frac{t}{2} + \Im(u + i\tau)\right|\right) \left(1 + \left|\frac{t}{2} + \Im(u - i\tau)\right|\right) \right\}^{\frac{1}{2} - \frac{\sigma}{2} - \Re u} dt.$$

Breaking the integration into $|t| \leqslant 4(1 + |\Im u| + |\tau|)$ and $|t| > 4(1 + |\Im u| + |\tau|)$ we find, for integers $0 \leqslant b \leqslant 2 < a$ and σ satisfying (3.12),

$$\Xi_{u,\tau}^{\eta_1,\eta_2}(x) \ll_{\sigma,\Re u} x^{-\frac{\sigma}{2}} \begin{cases} a^{\sigma-1-b}(1 + |\Im u| + |\tau|)^{2-\sigma-2\Re u} + (1 + |\Im u| + |\tau|)^{1-2\Re u - b}, & \text{for} \quad \sigma < 1 + b, \\ a^{\sigma-b}(1 + |\Im u| + |\tau|)^{1-\sigma-2\Re u} + (1 + |\Im u| + |\tau|)^{1-2\Re u-b}, & \text{for} \quad \sigma < 1 - 2\Re u. \end{cases}$$

In particular, for $u = 1/2 + \varepsilon$ we obtain

$$\Xi_{u,\tau}^{\eta_1,\eta_2}(x) \ll_\varepsilon x^{-\frac{1}{2}+2\varepsilon}(1 + |\tau|)^{2\varepsilon}, \tag{3.13}$$

$$\Xi_{u,\tau}^{\eta_1,\eta_2}(x) \ll_\varepsilon x^{\frac{1}{2}+\varepsilon} \left(\frac{1 + |\tau|}{a} + 1\right), \tag{3.14}$$

upon choosing $\sigma = 1 - 4\varepsilon$ and $\sigma = -1 - 2\varepsilon$, respectively, while for $1/2 < \Re u < (a - b + 1)/2 - \varepsilon$ we have

$$\Xi_{u,\tau}^{\eta_1,\eta_2}(x) \ll_{a,\tau,\Re u,\varepsilon} x^{\frac{a-b}{2} - \varepsilon} \tag{3.15}$$

upon choosing $\sigma = b - a + 2\varepsilon$. For $\alpha \in \mathbb{R}$ let

$$\zeta_{(\alpha)}(s) := \sum_{n+\alpha>0} (n + \alpha)^{-s}$$

be the Hurwitz zeta-function. It satisfies a functional equation

$$\zeta_{(\alpha)}(s) = (2\pi)^{s-1}\Gamma(1 - s) \left\{ -ie\left(\frac{s}{4}\right) \zeta^{(\alpha)}(1 - s) + ie\left(-\frac{s}{4}\right) \zeta^{(-\alpha)}(1 - s) \right\}, \tag{3.16}$$

where

$$\zeta^{(\alpha)}(s) := \sum_{n=1}^{\infty} e(\alpha n) n^{-s}.$$

Step 1. Let us first assume $5/4 < \Re u < 3/2$. By combining (2.11) with Petersson's (resp. Kuznetsov's) trace formula (2.15) (resp. (2.21) and Remark 2.1) we obtain the following expressions for $\mathcal{Q}_k^{\mathrm{holo}}(\ell)$ (resp. $\mathcal{Q}(\ell)$), cf. (2.7) and [By96, (5.3)]:

$$\begin{aligned}&\frac{\lambda_{\chi,\bar\chi}(\ell,-\tau)}{2\pi i^{-k}\ell^{u}}\prod_{p\mid q}\left(1-\frac{1}{p^{2u}}\right)\zeta(2u)\\ &+\sum_{D\mid c}\frac{1}{c}\sum_{m_1,m_2}\frac{S(m_1,m_2,-\ell;c)}{m_1^{u+i\tau}m_2^{u-i\tau}}\chi(m_1)\overline{\chi(m_2)}\varphi\left(\frac{4\pi\sqrt{m_1m_2\ell}}{c}\right),\end{aligned}\tag{3.17}$$

where

$$S(m_1,m_2,m_3;c):=\frac{1}{c}\sum_{a_1,a_2,a_3\,(c)}e\left(\frac{a_1a_2a_3+m_1a_1+m_2a_2+m_3a_3}{c}\right)$$

and

$$\varphi:=\begin{cases}J_{k-1}=\varphi_{k-1,0} & \text{if } f \text{ is holomorphic of weight } k\geqslant 4;\\ \varphi_{20,2} & \text{otherwise.}\end{cases}\tag{3.18}$$

The diagonal term in the first line of (3.17) only appears in the holomorphic case. The sum in the second line converges absolutely once $\Re u > 5/4$. In the following we transform the off-diagonal term further.

Step 2. We open φ and write it as an inverse Mellin transform

$$\varphi\left(\frac{4\pi\sqrt{m_1m_2\ell}}{c}\right)=\frac{1}{2\pi i}\int_{(\sigma)}\widetilde{\varphi}(s)\left(\frac{c}{4\pi\sqrt{m_1m_2\ell}}\right)^{s}ds.$$

By (3.11) the integrand is holomorphic and the integral converges absolutely if $-3<\sigma=\Re s<0$ in both the holomorphic (note $k\geqslant 4$) and the non-holomorphic case; the m_1,m_2-sum converges absolutely if $\Re u+\sigma/2>1$, and the c-sum converges absolutely if $\sigma<-1/2$ (Weil's bound, cf. [By96, Lemmata 1 and 3]). If we impose $2-2\Re u<\sigma<-1/2$, we can interchange the s-integration and the m_1,m_2-sum. Now splitting into residue classes modulo c, we write the m_1,m_2-sum as a linear combination of a product of two Hurwitz ζ-functions getting

$$\begin{aligned}\sum_{D\mid c}\frac{1}{c^{2u+1}}\frac{1}{2\pi i}\int_{(\sigma)}\widetilde{\varphi}(s)(4\pi\sqrt{\ell})^{-s}&\sum_{b_1,b_2\,(c)}S(b_1,b_2,-\ell;c)\chi(b_1)\overline{\chi(b_2)}\\ &\times\zeta_{\left(\frac{b_1}{c}\right)}\left(\frac{s}{2}+u+i\tau\right)\zeta_{\left(\frac{b_2}{c}\right)}\left(\frac{s}{2}+u-i\tau\right)ds.\end{aligned}$$

By standard bounds for the Hurwitz ζ-function the s-integral and the c-sum converge absolutely if $\Re u>5/4$ and $-3<\sigma<0$.

Step 3. We shift the integration to any line $-3<\sigma<-2\Re u$. By [By96, Lemma 6] if $\tau\neq 0$ and by [By96, Lemma 2] if $\tau=0$, we pick up poles only if $\frac{c}{q}\mid\ell$. Since $(\ell,D)=1$, $D\mid c$ and $\frac{D}{q}>1$, this does not happen[2]. Now we apply the functional equation (3.16) for the two Hurwitz ζ-functions[3], and write them as Dirichlet series getting (cf. [By96, (5.8)])

$$\begin{aligned}\sum_{D\mid c}\frac{(2\pi)^{2u-2}}{2c^{2u+1}}\sum_{m_1,m_2\in\mathbb{Z}\setminus\{0\}}&|m_1|^{u-1+i\tau}|m_2|^{u-1-i\tau}\sum_{b_1,b_2(c)}S(b_1,b_2,-\ell;c)\chi(b_1)\overline{\chi(b_2)}e\left(\frac{m_1b_1+m_2b_2}{c}\right)\\ &\times\Xi_{u,\tau}^{\mathrm{sgn}(m_1),\mathrm{sgn}(m_2)}\left(\frac{\ell}{|m_1m_2|}\right),\end{aligned}$$

where $\Xi_{u,\tau}^{\mathrm{sgn}(m_1),\mathrm{sgn}(m_2)}$ with φ as in (3.18) was defined in (3.10). This expression converges absolutely if $\Re u>5/4$. Note that when we apply (3.13)–(3.15) in the following, we have $(a,b)=(k-1,0)$ with $k\geqslant 4$ or $(a,b)=(20,2)$.

[2] It can be shown [By96, (5.10)] that the residues in the case $\frac{c}{q}\mid\ell$ would be harmless.

[3] i.e., we apply Poisson summation to both m_1 and m_2 in (3.17)

Step 4. We transform the b_1, b_2-sum by [By96, Lemma 2] obtaining

$$\sum_{D|c} \frac{(2\pi)^{2u-2}}{2c^{2u-1}q} \sum_{\substack{m_1m_2\neq 0\\ m_1m_2\equiv \ell\,(c/q)}} |m_1|^{u-1+i\tau}|m_2|^{u-1-i\tau} \sum_{d\,(q)}^{*} \bar{\chi}\left(m_1+\frac{c}{q}\bar{d}\right)\chi\left(m_2+\frac{m_1m_2-\ell}{c/q}d\right)$$
$$\times\, \Xi_{u,\tau}^{\operatorname{sgn}(m_1),\operatorname{sgn}(m_2)}\left(\frac{\ell}{|m_1m_2|}\right).$$

We will see in a moment that this term can be analytically continued to $\Re u > 1/2$. Let us start with the terms $m_1m_2 \neq \ell$. Their contribution equals

$$\frac{1}{4\pi q}\left(\frac{2\pi}{q}\right)^{2u-1} \sum_{\substack{m_1m_2-n_1n_2=\ell\\ m_1m_2n_1n_2\neq 0\\ D/q|n_1}} \frac{X}{|m_1m_2|^{1/2}} \left|\frac{m_1m_2}{n_1^2}\right|^{u-1/2} \left|\frac{m_1}{m_2}\right|^{i\tau} \Xi_{u,\tau}^{\operatorname{sgn}(m_1),\operatorname{sgn}(m_2)}\left(\frac{\ell}{|m_1m_2|}\right), \tag{3.19}$$

where

$$X := \sum_{d\,(q)}^{*} \bar{\chi}(m_1+n_1\bar{d})\chi(m_2+n_2d) \ll_\varepsilon q^{1/2+\varepsilon}(m_1,m_2,q)^{1/2}(n_1n_2,q)^{1/2}. \tag{3.20}$$

This estimate strengthens [By96, Lemma 4] and follows essentially from the Riemann Hypothesis over finite fields. We provide a detailed proof in the next section, see Proposition 3.1. The condition $(\ell, q) = 1$ is crucial here and in the sequel. By (3.15), the term (3.19) is holomorphic in $1/2 < \Re u < 3/2$. Let us take $u := 1/2 + \varepsilon$. We split the sum in (3.19) into two parts: $|m_1m_2| > \ell$, $|m_1m_2| < \ell$. Notice that $m_1m_2 = -\ell$ cannot happen, since $m_1m_2 \equiv \ell \pmod{D/q}$ and (3.1) would then imply $(2\ell, D) \geqslant D/q > 2$ which contradicts $(\ell, D) = 1$.

Using (3.14), the terms $|m_1m_2| > \ell$ contribute at most

$$\ll_\varepsilon (\ell q)^\varepsilon \left(\frac{\ell}{q}\right)^{\frac12}\left(\frac{1+|\tau|}{a}+1\right) \sum_{\substack{d_1,d_2|q\\ (d_1,d_2)=1}} (d_1d_2)^{\frac12} \sum_{\substack{m>\ell\\ m\equiv 0\,(d_1^2)\\ m\equiv\pm\ell\,([d_2,D/q])}} \frac{1}{m^{1+\varepsilon}},$$

where $a := 20$ in the non-holomorphic case and $a := k-1$ in the holomorphic case. The smallest element in the arithmetic progression given by the inner sum is at least $\max(\ell, d_1^2, \frac12[d_2, D/q])$, therefore the above is at most

$$\begin{aligned} &\ll_\varepsilon (\ell q)^\varepsilon \left(\frac{\ell}{q}\right)^{\frac12}\left(\frac{1+|\tau|}{a}+1\right)\left(\sum_{d|q}\frac{d^{\frac12}}{[d,D/q]} + \sum_{d_1,d_2|q}\frac{(d_1d_2)^{\frac12}}{\ell^{\frac14}d_1^{\frac12}[d_2,D/q]^{\frac12}}\right)\\ &\ll_\varepsilon (\ell q)^\varepsilon \left(\frac{\ell}{q}\right)^{\frac12}\left(\frac{1+|\tau|}{a}+1\right)\left(\frac{(N,q)^{\frac32}}{N}+\frac{(N,q)}{\ell^{\frac14}N^{\frac12}}\right). \end{aligned} \tag{3.21}$$

In the last step we used the definition of D (cf. (3.1)).

By (3.13), the terms $|m_1m_2| < \ell$ contribute at most

$$\begin{aligned} &\ll_\varepsilon \frac{(\ell q(1+|\tau|))^\varepsilon}{(\ell q)^{\frac12}} \sum_{\substack{d_1,d_2|q\\ (d_1,d_2)=1}} (d_1d_2)^{\frac12} \sum_{\substack{0<m<\ell\\ m\equiv 0\,(d_1^2)\\ m\equiv\pm\ell\,([d_2,D/q])}} 1\\ &\ll_\varepsilon \frac{(\ell q(1+|\tau|))^\varepsilon}{(\ell q)^{\frac12}} \sum_{\substack{d_1,d_2|q\\ (d_1,d_2)=1}} (d_1d_2)^{\frac12}\left(\frac{\ell}{[d_1^2d_2,D/q]}+1\right)\\ &\ll_\varepsilon (\ell q(1+|\tau|))^\varepsilon\left(\left(\frac{\ell}{q}\right)^{\frac12}\frac{(N,q)^{\frac32}}{N}+\frac{1}{\sqrt{\ell}}\right). \end{aligned} \tag{3.22}$$

Finally the contribution of the terms $m_1m_2=\ell$ is

$$\begin{aligned}&\sum_{D|c}\frac{(2\pi)^{2u-2}}{2c^{2u-1}q\ell^{1-u}}\sum_{m_1m_2=\ell}\chi(m_2)\left|\frac{m_1}{m_2}\right|^{i\tau}\sideset{}{^*}\sum_{a\,(q)}\bar\chi\left(m_1+\frac{c}{q}\bar a\right)\left(\Xi^{1,1}_{u,\tau}(1)+\Xi^{-1,-1}_{u,\tau}(1)\right)\\ =\;&\frac{(2\pi)^{2u-2}\left(\Xi^{1,1}_{u,\tau}(1)+\Xi^{-1,-1}_{u,\tau}(1)\right)}{2D^{2u-1}q\ell^{1-u}}\sum_{m_1m_2=\ell}\chi(m_2)\left|\frac{m_1}{m_2}\right|^{i\tau}\sum_{c}\frac{1}{c^{2u-1}}\sideset{}{^*}\sum_{a\,(q)}\bar\chi\left(m_1+\frac{D}{q}\bar a c\right).\end{aligned}\tag{3.23}$$

We write $r:=(D/q,q)$. Then the c,a-sum equals

$$\left(\frac{q}{r}\right)^{1-2u}\sum_{b\,(q/r)}\sideset{}{^*}\sum_{a\,(q)}\bar\chi(m_1+r\bar ab)\zeta_{(\frac{b}{q/r})}(2u-1)$$

which is holomorphic for $\mathbb{C}\setminus\{1/2\}$. By the functional equation (3.16), this is for $\Re u>1/2$

$$\begin{aligned}&-i\left(\frac{q}{r}\right)^{1-2u}(2\pi)^{2u-2}\Gamma(2-2u)e\left(\frac{2u-1}{4}\right)\sum_n\frac{1}{n^{2-2u}}\sum_{b\,(q/r)}\sideset{}{^*}\sum_{a\,(q)}\bar\chi(m_1+r\bar ab)e(brn/q)\\ &+i\left(\frac{q}{r}\right)^{1-2u}(2\pi)^{2u-2}\Gamma(2-2u)e\left(\frac{1-2u}{4}\right)\sum_n\frac{1}{n^{2-2u}}\sum_{b\,(q/r)}\sideset{}{^*}\sum_{a\,(q)}\bar\chi(m_1+r\bar ab)e(-brn/q).\end{aligned}$$

The a,b-sum decomposes into Ramanujan sums,

$$\sum_b\sum_a\ldots=\sum_{\substack{d\,(q)\\ r|d}}\bar\chi(m_1+d)\sideset{}{^*}\sum_{a\,(q)}e\left(\pm\frac{adn}{q}\right)=\sum_{\substack{d\,(q)\\ r|d}}\bar\chi(m_1+d)\sum_{s|(dn,q)}s\mu\left(\frac{q}{s}\right),$$

showing that both n-sums equal

$$\sum_{\substack{d\,(q)\\ r|d}}\bar\chi(m_1+d)\sum_{s|q}s\mu\left(\frac{q}{s}\right)\sum_{\frac{s}{(d,s)}|n}\frac{1}{n^{2-2u}}=\zeta(2-2u)\sum_{\substack{d\,(q)\\ r|d}}\bar\chi(m_1+d)\sum_{s|q}\mu\left(\frac{q}{s}\right)\frac{(d,s)^{2-2u}}{s^{1-2u}}.$$

We substitute this back into (3.23), and obtain by (3.13) that this term for $u=1/2+\varepsilon$ is bounded by

$$\frac{(\ell q(1+|\tau|))^{\varepsilon}}{q\sqrt{\ell}}\sum_{d\,(q)}(d,q)\ll_\varepsilon\frac{(\ell q(1+|\tau|))^{\varepsilon}}{\sqrt{\ell}}.\tag{3.24}$$

Collecting the first line of (3.17), (3.21), (3.22), and (3.24), we arrive at (3.7) for $u=1/2+\varepsilon$.

3.3 A character sum estimate

In this section we state in more precise form the bound (3.20) and provide a detailed proof.

Proposition 3.1. *Let χ be a primitive character modulo q and let m_1, m_2, n_1, n_2 be arbitrary integers satisfying $(m_1m_2-n_1n_2,q)=1$. Then we have the uniform bound*[4]

$$X(m_1,m_2,n_1,n_2):=\sideset{}{^*}\sum_{a\,(q)}\bar\chi(m_1+n_1\bar a)\chi(m_2+n_2a)\ll q^{1/2}\tau(q)(m_1n_1^2,m_2n_2^2,q)^{1/2},$$

where the implied constant is absolute.

By the multiplicative nature of these sums it suffices to show that

$$|X(m_1,m_2,n_1,n_2)|\leqslant q^{1/2}(m_1n_1^2,m_2n_2^2,q)^{1/2}\times\begin{cases}2, & q=p^\beta\text{ for a prime }p>2;\\ 2^{5/2} & q=p^\beta\text{ for }p=2.\end{cases}\tag{3.25}$$

[4]Note that $(m_1m_2-n_1n_2,q)=1$ implies $(m_1n_1^2,m_2n_2^2,q)=(m_1,m_2,q)(n_1^2,n_2^2,q)\mid(m_1,m_2,q)(n_1n_2,q)$.

Case 1. First we discuss the case when $\beta = 1$ (that is, when q is prime). We apply [IK04, Theorem 11.23] with the parameters $n = 1$, $\mathbb{F} := \mathbb{F}_q$, and

$$f(x) := x(m_1x + n_1)^{d-1}(m_2 + n_2x),$$

where $d > 1$ is the order of χ. The only thing we have to check is that f is not a d-th power. If $d > 2$ then f can only be a d-th power if $n_1 = n_2 = 0$ in $\mathbb{F}$ in which case the displayed bound is trivial. If $d = 2$ then f can only be a d-th power if $n_1 = n_2 = 0$ or $m_1 = m_2 = 0$ in $\mathbb{F}$ in which case the displayed bound (3.25) is again trivial. Otherwise (3.25) follows from [IK04, Theorem 11.23].

Case 2. Now we discuss the case when $\beta > 1$ is even, say $\beta = 2\alpha$. We apply [IK04, Lemma 12.2] for the rational functions

$$f(x) := x\frac{m_2 + n_2x}{m_1x + n_1}, \qquad g(x) := 0.$$

Then

$$f'(x) = \frac{m_1n_2x^2 + 2n_1n_2x + m_2n_1}{(m_1x + n_1)^2},$$

therefore it suffices to show that the congruence

$$m_1n_2y^2 + 2n_1n_2y + m_2n_1 \equiv 0 \pmod{p^\alpha} \tag{3.26}$$

under the restriction

$$y(m_2 + n_2y)(m_1y + n_1) \not\equiv 0 \pmod{p} \tag{3.27}$$

has at most $2(n_1, n_2, p^\alpha)$ solutions when $p > 2$ and at most $4(n_1, n_2, p^\alpha)$ solutions when $p = 2$. We can clearly assume that $(n_1, n_2, p^\alpha) < p^\alpha$ for otherwise the assertion is trivial. Let us first assume that $p > 2$. If $p \mid m_1$ and $p \mid m_2$ then the condition $(m_1m_2 - n_1n_2, q) = 1$ shows that (3.26) has no solution satisfying $p \nmid y$. Therefore, without loss of generality, we can assume that $p \nmid m_1$. We multiply both sides of (3.26) by m_1 to see that the congruence is equivalent to

$$n_2(m_1y + n_1)^2 \equiv n_1(n_1n_2 - m_1m_2) \pmod{p^\alpha}.$$

By assumption, the parentheses on both sides are coprime with p, hence a solution can only exist if $p^\gamma \parallel n_1$ and $p^\gamma \parallel n_2$ for some $0 \leqslant \gamma \leqslant \alpha - 1$, and then the number of solutions of (3.26) under (3.27) is at most $2p^\gamma = 2(n_1, n_2, p^\alpha)$ by the structure of the group $(\mathbb{Z}/p^{\alpha-\gamma})^\times$. For $p = 2$ we adjust the above argument slightly. First of all, we can assume that $\alpha > 2$ for otherwise (3.26) trivially has at most 4 solutions. If $4 \mid m_1$ and $4 \mid m_2$ then the condition $(m_1m_2 - n_1n_2, q) = 1$ shows that (3.26) has no solution satisfying $2 \nmid y$. Therefore, without loss of generality, we can assume that $4 \nmid m_1$. We multiply both sides of (3.26) by m_1 to see that the congruence is equivalent to

$$n_2(m_1y + n_1)^2 \equiv n_1(n_1n_2 - m_1m_2) \pmod{2^\alpha(m_1, 2)}.$$

If $2 \mid n_1n_2$ then $2 \nmid m_1m_2$ and we conclude, similarly as in the case of $p > 2$, that the number of solutions of (3.26) under (3.27) is at most $4(n_1, n_2, 2^\alpha)$. If $2 \nmid n_1n_2$ then the number of solutions of the congruence

$$n_2x^2 \equiv n_1(n_1n_2 - m_1m_2) \pmod{2^\alpha(m_1, 2)}$$

is at most 4 while the map $\mathbb{Z}/2^\alpha \to \mathbb{Z}/2^\alpha(m_1, 2)$ given by $y \mapsto m_1y + n_1$ is injective, hence the number of solutions of (3.26) under (3.27) is also at most 4.

Case 3. Finally we discuss the case when $\beta > 1$ is odd, say $\beta = 2\alpha + 1$. We apply [IK04, Lemma 12.3] for the rational functions

$$f(x) := x\frac{m_2 + n_2x}{m_1x + n_1}, \qquad g(x) := 0.$$

Then

$$f'(x) = \frac{m_1n_2x^2 + 2n_1n_2x + m_2n_1}{(m_1x + n_1)^2}, \qquad f''(x) = \frac{2n_1(n_1n_2 - m_1m_2)}{(m_1x + n_1)^3},$$

hence for $p \nmid 2n_1$ the bound (3.25) follows from the already proven fact that (3.26) under (3.27) has at most 2 solutions and for $p = 2$ the bound (3.25) follows from the already proven fact that (3.26) under (3.27) has at most $4(n_1, n_2, p^\alpha)$ solutions. For $p \mid n_1$ $(p > 2)$ it suffices to show that in any complete residue systems modulo p^α there are at most $2p^{-1}(n_1, n_2, p^{\alpha+1})$ solutions of the congruence

$$m_1 n_2 y^2 + 2n_1 n_2 y + m_2 n_1 \equiv 0 \pmod{p^{\alpha+1}} \tag{3.28}$$

under (3.27). We can clearly assume that $(n_1, n_2, p^{\alpha+1}) < p^{\alpha+1}$ for otherwise the assertion is trivial. By the condition $(m_1 m_2 - n_1 n_2, q) = 1$ we have $p \nmid m_1$, hence (3.28) is equivalent to

$$n_2(m_1 y + n_1)^2 \equiv n_1(n_1 n_2 - m_1 m_2) \pmod{p^{\alpha+1}}.$$

By assumption, the parentheses on both sides are coprime with p, hence a solution of (3.28) can only exist if $p^\gamma \parallel n_1$ and $p^\gamma \parallel n_2$ for some $1 \leqslant \gamma \leqslant \alpha$, and then the number of solutions of (3.28) under (3.27) is at most $2p^\gamma$ by the structure of the group $(\mathbb{Z}/p^{\alpha+1-\gamma})^\times$. In particular, n_1 and n_2 are both divisible by p and the solutions of (3.28) under (3.27) form $2p^{\gamma-1} = 2p^{-1}(n_1, n_2, p^{\alpha+1})$ complete residue classes modulo p^α. This completes the proof of Proposition 3.1.

Chapter 4

Modular L-functions

4.1 Preliminaries on divisor sums

Let τ be the divisor function. Exponential sums involving the divisor function can be handled by Voronoi summation. Let

$$L_w(x) := \log x + 2\gamma - 2\log w, \tag{4.1}$$

where γ is Euler's constant, and let

$$\mathcal{J}^-(x) := -2\pi Y_0(4\pi x), \qquad \mathcal{J}^+(x) := 4K_0(4\pi x)$$

with the usual Bessel functions. For later purposes we write $\mathcal{J}^\pm$ as inverse Mellin transforms using [GR07, 17.43.17, 17.43.18] or [KMV00, (36)]:

$$\begin{aligned} \mathcal{J}^+(\sqrt{x}) &= \frac{2}{2\pi i}\int_{(1)} (2\pi)^{-2u}\Gamma(u)^2 x^{-u}\,du, \\ \mathcal{J}^-(\sqrt{x}) &= \frac{2}{2\pi i}\int_{(*)} (2\pi)^{-2u}\Gamma(u)^2 x^{-u}\cos(\pi u)\,du, \end{aligned} \tag{4.2}$$

where $(*)$ is the path $\Re u = -1$ except when $|\Im u| < 1$ where it curves to hit the real axis at $u > 0$. Let $(d, c) = 1$ and let $F \in \mathcal{C}_0^\infty((0,\infty))$, then

$$\begin{aligned} \sum_{m=1}^\infty \tau(m) e\left(\frac{dm}{c}\right) F(m) =& \frac{1}{c}\int_0^\infty L_c(y)F(y)\,dy \\ &+ \frac{1}{c}\sum_\pm \sum_{m=1}^\infty \tau(m)\ e\left(\frac{\pm\bar{d}m}{c}\right)\int_0^\infty \mathcal{J}^\pm\left(\frac{\sqrt{my}}{c}\right)F(y)\,dy. \end{aligned} \tag{4.3}$$

In order to evaluate additive divisor sums, we use the following method, cf. [Me01, (2.1) and (2.4)]. Here and later in the proof, we will need smooth cut-off functions. Let henceforth ω denote a smooth function such that $\omega(x) = 1$ on $[0,1]$ and $\omega(x) = 0$ on $[2,\infty)$. Then we have

$$\left(1 - \omega\left(\frac{x}{\sqrt{Q}}\right)\right)\left(1 - \omega\left(\frac{y}{x\sqrt{Q}}\right)\right) = 0$$

for all $x, y, Q \geqslant 0$ such that $y \leqslant Q$. Therefore

$$\tau(n) = \sum_{\delta | n} \omega\left(\frac{\delta}{\sqrt{Q}}\right)\left(2 - \omega\left(\frac{n}{\delta\sqrt{Q}}\right)\right)$$

whenever $n \leqslant Q$. Let $g : [\frac{1}{2}, Q] \times [\frac{1}{2}, M] \to \mathbb{C}$ be a smooth function. Then

$$\begin{aligned}
\sum_{an \pm m = h} \tau(n)\tau(m)g(n,m) &= \sum_{n=1}^{\infty} \tau(n)\tau(\pm(h-an))g(n,\pm(h-an)) \\
&= \sum_{\delta=1}^{\infty} \omega\left(\frac{\delta}{\sqrt{Q}}\right) \sum_{\delta \mid n} \tau(\pm(h-an))g(n,\pm(h-an)) \left(2 - \omega\left(\frac{n}{\delta\sqrt{Q}}\right)\right) \\
&= \sum_{\delta=1}^{\infty} \omega\left(\frac{\delta}{\sqrt{Q}}\right) \sum_{m \equiv \pm h \, (a\delta)} \tau(m) g\left(\frac{h \mp m}{a}, m\right)\left(2 - \omega\left(\frac{h \mp m}{a\delta\sqrt{Q}}\right)\right).
\end{aligned}$$

Using additive characters and Voronoi summation (4.3), we get

$$\begin{aligned}
\sum_{m \equiv \mu \, (c)} \tau(m)F(m) = \frac{1}{c}\sum_{w \mid c} \frac{r_w(\mu)}{w} \int_0^\infty L_w(y)F(y)\,dy \\
+ \sum_{\pm} \frac{1}{c} \sum_{w \mid c} \frac{1}{w} \sum_{m=1}^{\infty} \tau(m) S(-\mu, \pm m; w) \int_0^\infty \mathcal{J}^{\pm}\left(\frac{\sqrt{my}}{w}\right) F(y)\,dy
\end{aligned}$$

for any compactly supported smooth function F, so that

$$\begin{aligned}
\sum_{an \pm m = h} \tau(n)\tau(m)g(n,m) &= \sum_{w=1}^{\infty} \frac{(a,w)r_w(h)}{w^2} \int_0^\infty L_w(\pm(h-ax))K_{(a,w),w}(x)g(x,\pm(h-ax))\,dx \\
&+ \sum_{w=1}^{\infty} \frac{(a,w)}{w^2} \sum_{n=1}^{\infty} \tau(n) S(\mp h, n; w) \int_0^\infty \mathcal{J}^{+}\left(\frac{\sqrt{n(\pm(h-ax))}}{w}\right) K_{(a,w),w}(x)g(x,\pm(h-ax))\,dx \\
&+ \sum_{w=1}^{\infty} \frac{(a,w)}{w^2} \sum_{n=1}^{\infty} \tau(n) S(\mp h, -n; w) \int_0^\infty \mathcal{J}^{-}\left(\frac{\sqrt{n(\pm(h-ax))}}{w}\right) K_{(a,w),w}(x)g(x,\pm(h-ax))\,dx,
\end{aligned} \tag{4.4}$$

where

$$r_w(h) := S(h,0;w) = \sum_{d \mid (h,w)} d\mu(w/d) \tag{4.5}$$

is the Ramanujan sum and

$$K_{r,w}(x) := \sum_{\delta=1}^{\infty} \frac{1}{\delta} \omega\left(\frac{w\delta}{r\sqrt{Q}}\right)\left(2 - \omega\left(\frac{rx}{\delta w\sqrt{Q}}\right)\right).$$

For future reference we state some properties of $K_{r,w}(x)$. A straightforward calculation shows

$$x^i w^j \frac{\partial^i}{\partial x^i} \frac{\partial^j}{\partial w^j} K_{r,w}(x) \ll_{i,j} \log Q \tag{4.6}$$

for any $i, j \geqslant 0$, and clearly

$$K_{r,w}(x) = 0 \quad \text{ if } w \geqslant 2r\sqrt{Q}. \tag{4.7}$$

4.2 Approximate functional equation

Let $f = f_0$ be a primitive (holomorphic or Maass) cusp form having L^2-norm 1, for which we want to prove Theorem 1.2. Let $t_0 = t_{f_0}$ denote its spectral parameter as defined in (2.4). For $\Re s > 1$ the L-function of f_0 is defined as a Dirichlet series in the Hecke eigenvalues of f_0

$$L(f_0, s) := \sum_{n=1}^{\infty} \lambda_{f_0}(n) n^{-s}.$$

The completed L-function is given by

$$\Lambda(f_0,s):=q^{s/2}L_\infty(f_0,s)L(f_0,s),\qquad L_\infty(f_0,s):=\pi^{-s}\Gamma\left(\frac{s+\mu_1}{2}\right)\Gamma\left(\frac{s+\mu_2}{2}\right),$$

where

$$\mu_1,\ \mu_2:=\begin{cases} it_0, & -it_0 \quad \text{when } f_0 \text{ is an even Maass form of even weight;}\\ it_0, & -it_0+1 \quad \text{when } f_0 \text{ is an even Maass form of odd weight;}\\ it_0+1, & -it_0+1 \quad \text{when } f_0 \text{ is an odd Maass form of even weight;}\\ it_0+1, & -it_0 \quad \text{when } f_0 \text{ is an odd Maass form of odd weight;}\\ -it_0, & -it_0+1 \quad \text{when } f_0 \text{ is a holomorphic form.}\end{cases}$$

Observe that Hypothesis H_θ implies

$$\Re\mu_1,\ \Re\mu_2\ \geqslant\ -\theta. \tag{4.8}$$

The completed L-function is entire and satisfies the functional equation [DFI02, (8.11)–(8.13), (8.17)–(8.19)]

$$\Lambda(f_0,s)=\omega\overline{\Lambda}(f_0,1-\overline{s}) \tag{4.9}$$

for some constant $\omega=\omega(f_0)$ of modulus 1. Relation (2.8) shows that

$$L(f_0,s)^2=L(2s,\chi)\sum_{n=1}^{\infty}\tau(n)\lambda_{f_0}(n)n^{-s},\qquad \Re s>1. \tag{4.10}$$

Let us fix a point s on the critical line $\Re s=\frac{1}{2}$ for which we want to prove Theorem 1.2. The above Dirichlet series no longer converges (absolutely) for s but a similar formula holds which is traditionally called an approximate functional equation. In order to achieve polynomial dependence in the spectral parameter t_0 we will closely follow the argument in [Ha02] specified for the shifted L-function $u\mapsto L\left(f_0,s-\frac{1}{2}+u\right)$. We define the analytic conductor [Ha02, (2.4) and Remark 2.7]

$$C=C(f_0,s):=\frac{q}{(2\pi)^2}|s+\mu_1||s+\mu_2| \tag{4.11}$$

and the auxiliary function [Ha02, (1) in Erratum]

$$F(f_0,s;u):=\frac{1}{2}C^{-u/2}q^u\frac{L_\infty(f_0,s+u)\overline{L_\infty}(f_0,s)}{\overline{L_\infty}(f_0,s-\overline{u})L_\infty(f_0,s)}+\frac{1}{2}C^{u/2}.$$

By (4.8) this function is holomorphic in $\Re u>-\frac{1}{4}$ (say) and satisfies the bound [Ha02, (2) in Erratum]

$$C^{-u/2}F(f_0,s;u)-\frac{1}{2}\ll_\sigma(1+|u|)^{2\Re u},\qquad -\frac{1}{4}<\Re u\leqslant\sigma \tag{4.12}$$

with an implied constant independent of s and f_0. In addition, we have $F(f_0,s;0)=1$, and from the functional equation (4.9) we can deduce [Ha02, (3.3)]

$$F(f_0,s;u)L(f_0,s+u)=\omega\lambda\overline{F}(f_0,s;-\overline{u})\overline{L}(f_0,s-\overline{u}),\qquad \lambda:=\frac{\overline{L_\infty}(f_0,s)}{L_\infty(f_0,s)}.$$

In particular,

$$\eta=\eta(f_0,s):=(\omega\lambda)^2$$

is of modulus 1 and with the notation

$$G^+(u):=F(f_0,s;\tfrac{1}{2}-s+u)^2,\qquad G^-(u):=\overline{F}(f_0,s;\tfrac{1}{2}-s+\overline{u})^2$$

we obtain the functional equation

$$G^+(u)L(f_0,\tfrac{1}{2}+u)^2=\eta G^-(-u)\overline{L}(f_0,\tfrac{1}{2}-\overline{u})^2. \tag{4.13}$$

Observe that (4.12) implies, for $0 < \varepsilon \leqslant \Re u \leqslant \sigma$,

$$G^{\pm}(u) \ll_{\varepsilon,\sigma} C^{\Re u} \left(1 + |\Im u \mp \Im s|\right)^{4\Re u}. \tag{4.14}$$

We fix an arbitrary entire function $P(u)$ which decays fast in vertical strips and satisfies $P(0) = 1$ as well as $P(u) = P(-u) = \overline{P(\overline{u})}$. The role of this factor is to make the dependence on s in Theorem 1.2 polynomial. We introduce another even function in order to create zeros that avoid the matching, as discussed in Section 1.5:

$$Q(u,t) := \left(u^2 - (\tfrac{1}{2} - it)^2\right)^2 \left(u^2 - (\tfrac{1}{2} + it)^2\right)^2 =: \sum_{\nu=0}^{4} \alpha_\nu(t) u^{2\nu} \tag{4.15}$$

for suitable real even polynomials $\alpha_\nu \in \mathbb{R}[T]$. Note that

$$Q\left(u, i\left(\tfrac{1}{2} - u\right)\right) = Q^{(1,0)}\left(u, i\left(\tfrac{1}{2} - u\right)\right) = 0. \tag{4.16}$$

Now we apply the usual contour shift technique to the integral

$$\frac{1}{2\pi i} \int_{(1)} L\left(f_0, \tfrac{1}{2} + u\right)^2 G^+(u) P(u + \tfrac{1}{2} - s) \frac{Q(u,t_0)}{Q(s-\frac{1}{2}, t_0)} \cdot \frac{du}{u + \frac{1}{2} - s}.$$

In combination with (4.10) and (4.13) we obtain

$$L(f_0, s)^2 = \sum_{n=1}^{\infty} \frac{\tau(n)\lambda_{f_0}(n) V_{t_0}^+(n/q)}{n^{1/2}} + \eta \sum_{n=1}^{\infty} \frac{\tau(n)\overline{\lambda_{f_0}(n)} V_{t_0}^-(n/q)}{n^{1/2}}, \tag{4.17}$$

where we define $V_t^{\pm}$ for any spectral parameter t through its Mellin transform

$$\begin{aligned} \widetilde{V}_t^+(u) &:= \widetilde{W}^+(u) Q(u,t) := q^{-u} G^+(u) L(1+2u, \chi) \frac{P(u + \frac{1}{2} - s)}{u + \frac{1}{2} - s} \cdot \frac{Q(u,t)}{Q\left(s - \frac{1}{2}, t_0\right)}, \\ \widetilde{V}_t^-(u) &:= \widetilde{W}^-(u) Q(u,t) := q^{-u} G^-(u) L(1+2u, \overline{\chi}) \frac{P(u + s - \frac{1}{2})}{u + s - \frac{1}{2}} \cdot \frac{Q(u,t)}{Q\left(s - \frac{1}{2}, t_0\right)}. \end{aligned} \tag{4.18}$$

Here we have suppressed the notational dependence of $\widetilde{V}_t^{\pm}$ and $\widetilde{W}^{\pm}$ on s and t_0 as these parameters are kept fixed in the rest of the paper. Since $Q\left(s - \frac{1}{2}, t_0\right)$ is real for $\Re s = \frac{1}{2}$ and the spectral parameter t_0, we have

$$\begin{array}{rclcrcl} \widetilde{W}^-(u) &=& \overline{\widetilde{W}^+(\overline{u})} & \quad\text{and}\quad & \widetilde{V}_t^-(u) &=& \overline{\widetilde{V}_t^+(\overline{u})} \\ W^-(x) &=& \overline{W^+(x)} & & V_t^-(x) &=& \overline{V_t^+(x)}. \end{array} \tag{4.19}$$

We can therefore drop the superscripts and write

$$W := W^+ \qquad \text{and} \qquad V_t := V_t^+.$$

Note that by (4.18) and (4.15),

$$V_t(x) = \sum_{\nu=0}^{4} \alpha_\nu(t) \left(x \frac{\partial}{\partial x}\right)^{2\nu} W(x). \tag{4.20}$$

By (4.14), (4.11), (4.15), (2.4), it follows, for $0 < \varepsilon \leqslant \Re u \leqslant \sigma$ and for any $A > 0$,

$$\widetilde{W}(u) \ll_{\varepsilon,\sigma,A} (|s| + |t_0|)^{2\Re u} \left(1 + |\Im u \mp \Im s|\right)^{-A}.$$

Therefore $\widetilde{W}$ is rapidly decaying on vertical lines and inverse Mellin transformation shows

$$x^i \frac{\partial^i}{\partial x^i} W(x) \ll_{\varepsilon,B,i} |s|^{i+1} (|s| + |t_0|)^{2\varepsilon} x^{-\varepsilon} \left(1 + \frac{x}{(|s| + |t_0|)^2}\right)^{-B}, \qquad B, i \in \mathbb{N}_0. \tag{4.21}$$

With these auxiliary functions we introduce the following family of "fake" L-functions for any cusp form f either in $\mathcal{B}_k^h(q,\chi)$ or in $\mathcal{B}_\kappa(q,\chi)$ and for any Eisenstein series $E_{\chi_1,\chi_2,f,t}$:

$$\begin{aligned}\mathcal{L}(f\otimes E,s) &:= \sum_{n=1}^{\infty}\frac{\tau(n)\sqrt{n}\rho_f(n)V_{t_f}(n/q)}{n^{1/2}},\\ \mathcal{L}(E_{\chi_1,\chi_2,f,t}\otimes E,s) &:= \sum_{n=1}^{\infty}\frac{\tau(n)\sqrt{n}\rho_f(n,t)V_t(n/q)}{n^{1/2}}.\end{aligned} \tag{4.22}$$

With this notation (4.17) reads for $f=f_0$ (cf. (2.11) and (4.19))

$$\rho_{f_0}(1)L(f_0,s)^2=\mathcal{L}(f_0\otimes E,s)+\eta(f_0,s)\overline{\mathcal{L}(f_0\otimes E,s)}. \tag{4.23}$$

In order to apply the trace formula, we wanted an approximate functional equation that is "as independent of t_0 as possible"; now the information on the spectral parameter is all encoded in the polynomial $Q(u,t)$. In [DFI02], however, the weight function was the same for all the f's which made the rest of the proof more complicated.

4.3 Amplification

In this section we introduce the amplified second moment whose estimation will lead to the proof of Theorem 1.2. The analysis relies on Kuznetsov's trace formula, therefore we use as spectral coefficients the Bessel transforms (2.18)–(2.19) of a convenient test function $\varphi\in C^\infty(\mathbb{R}^+)$. The construction is similar as in Section 3.1 for the proof of Theorem 1.1, but instead of working with a single function defined in (3.4) we shall use a linear combinations of such functions to gain more flexibility. Namely, it follows from (3.5) that for any fixed a, b and any even polynomial $\alpha\in\mathbb{C}[T]$ of degree $2d\leqslant 2b-4$ there is a linear combination

$$\varphi(x)=\sum_{\nu=0}^{d}\beta_\nu\varphi_{a-\nu,b-\nu}(x) \tag{4.24}$$

with β_ν depending on a, b and the coefficients of α such that

$$\dot\varphi(k)=\dot\varphi_{a,b}(k)\alpha\left(\tfrac{(1-k)i}{2}\right)\qquad\text{and}\qquad\hat\varphi(t)=\hat\varphi_{a,b}(t)\alpha(t). \tag{4.25}$$

That is, we can introduce any given polynomial factor in the spectral coefficients once b is sufficiently large. Note that for $b\geqslant 2$ the function $\varphi_{a,b}\in C^\infty(\mathbb{R}^+)$ also satisfies the decay conditions for Kuznetsov's trace formula, and by [GR07, 6.561.14] its transform (2.28) satisfies

$$\varphi^*_{a,b}(u)=i^{b-a}2^{-b-1}\frac{\Gamma\left((a-b-1-2u)/2\right)}{\Gamma\left((3+a+b+2u)/2\right)}\ll_{a,b}(1+|\Im u|)^{-b-2-2\Re u},\quad |\Re u|\leqslant\frac{a-b-2}{2}. \tag{4.26}$$

We now specify

$$\varphi_0(x):=\varphi_{A,10}(x)=i^{10-A}J_A(x)x^{-10} \tag{4.27}$$

for some very large A of parity κ, and for $(\ell,q)=1$ we define

$$\begin{aligned}\mathcal{Q}_k^{\mathrm{holo}}(\ell) :=& 2i^k\Gamma(k-1)\sum_{f\in\mathcal{B}_k^h(q,\chi)}\lambda_f(\ell)\left|\mathcal{L}(f\otimes E,s)\right|^2,\\ \mathcal{Q}(\ell) :=& \sum_{\substack{k\equiv\kappa\,(2)\\ k>\kappa}}\dot\varphi_0(k)2(k-1)i^{-k}\mathcal{Q}_k^{\mathrm{holo}}(\ell)+\sum_{f\in\mathcal{B}_\kappa(q,\chi)}\hat\varphi_0(t_f)\frac{4}{\cosh(\pi t_f)}\lambda_f(\ell)\left|\mathcal{L}(f\otimes E,s)\right|^2\\ &+\sum_{\substack{\chi_1\chi_2=\chi\\ f\in\mathcal{B}_\kappa(\chi_1,\chi_2)}}\int_{-\infty}^{\infty}\hat\varphi_0(t)\frac{1}{\pi\cosh(\pi t)}\lambda_{\chi_1,\chi_2}(\ell,t)\left|\mathcal{L}(E_{\chi_1,\chi_2,f,t}\otimes E,s)\right|^2dt.\end{aligned} \tag{4.28}$$

Remark 4.1. Let us explain the reason of our choice for the construction of $V_t^{\pm}$. Suppose for simplicity that q is prime (hence χ being nontrivial is primitive). In that case there are two Eisenstein series $E_{\chi,1,f,t}$ and $E_{1,\chi,f,t}$ (which are the Eisenstein series associated to the cusps $\mathfrak{a}=0,\infty$). Their contribution to the above sum equals

$$\int_{-\infty}^{\infty}\hat{\varphi}_0(t)\frac{|\rho(1,t)|^2}{\pi\cosh(\pi t)}\lambda_{\chi,1}(\ell,t)\left|\frac{1}{2\pi i}\int_{(1)}\frac{L^2(u+\frac12-it,\chi)\zeta^2(u+\frac12+it)}{L(2u+1,\chi)}\widetilde{V}_t(u)\,q^u\,du\right|^2dt. \tag{4.29}$$

The main contribution comes from the double pole of the inner integrand, and we designed V_t such that it kills this pole. This is reflected by the vanishing of (4.58) below.

We shall show

$$k^{-18}\left|\mathcal{Q}_k^{\mathrm{holo}}(\ell)\right|+|\mathcal{Q}(\ell)|\ll_{s,t_0,\varepsilon}q^{\varepsilon}\left(\ell^{c_1}q^{-c_2}+\ell^{-1/2}\right) \tag{4.30}$$

for certain positive absolute constants c_1 and c_2, uniformly in $k\geqslant A-10$, and with polynomial dependence on s and t_0. This implies Theorem 1.2: Let us choose the standard amplifier

$$x(\ell):=\begin{cases}\lambda(p)\overline{\chi}(p) & \text{if}\quad \ell=p,\quad p\nmid q,\quad \frac12\sqrt{L}<p\leqslant\sqrt{L};\\ -\overline{\chi}(p) & \text{if}\quad \ell=p^2,\quad p\nmid q,\quad \frac12\sqrt{L}<p\leqslant\sqrt{L};\\ 0 & \text{else};\end{cases}$$

for some parameter $\log L\asymp\log q$ to be chosen in a minute. Using (2.8) with $n=m=p$, we see

$$\sum_{\ell}x(\ell)\lambda(\ell)=\sum_{\substack{p\nmid q\\ \frac12\sqrt{L}<p\leqslant\sqrt{L}}}1\gg L^{1/2-\varepsilon}.$$

Therefore, by (4.23), (2.40)–(2.41), (4.27), (3.5)–(3.6), we obtain

$$\begin{aligned}&\frac{L}{q^{1+\varepsilon}}\left|L(f_0,s)\right|^4\ll_{t_0,\varepsilon}\\ &\qquad\sum_{\substack{k\equiv\kappa\,(2),\ k>\kappa\\ f\in\mathcal{B}_k^h(q,\chi)}}\sum\left|\dot{\varphi}_0(k)\right|4\Gamma(k)\left|\sum_{\ell}x(\ell)\lambda_f(\ell)\right|^2\left|\mathcal{L}(f\otimes E,s)\right|^2\\ &\qquad+\sum_{f\in\mathcal{B}_\kappa(q,\chi)}\hat{\varphi}_0(t_f)\frac{4}{\cosh(\pi t_f)}\left|\sum_{\ell}x(\ell)\lambda_f(\ell)\right|^2\left|\mathcal{L}(f\otimes E,s)\right|^2\\ &\qquad+\sum_{\substack{\chi_1\chi_2=\chi\\ f\in\mathcal{B}_\kappa(\chi_1,\chi_2)}}\sum\int_{-\infty}^{\infty}\hat{\varphi}_0(t)\frac{1}{\pi\cosh(\pi t)}\left|\sum_{\ell}x(\ell)\lambda_{\chi_1,\chi_2}(\ell,t)\right|^2\left|\mathcal{L}(E_{\chi_1,\chi_2,f,t}\otimes E,s)\right|^2dt,\end{aligned}$$

so that by (2.9), (3.6) and (4.28) we obtain

$$\begin{aligned}&\frac{L}{q^{1+\varepsilon}}\left|L(f_0,s)\right|^4\ll_{t_0,\varepsilon}\\ &\qquad\sum_{\ell_1,\ell_2}\left|x(\ell_1)x(\ell_2)\right|\sum_{d\mid(l_1,l_2)}\left\{\left|\mathcal{Q}\left(\frac{\ell_1\ell_2}{d^2}\right)\right|+\sum_{\substack{k\equiv\kappa\,(2)\\ k\geqslant A-10}}4k\left|\dot{\varphi}_0(k)\right|\left|\mathcal{Q}_k^{\mathrm{holo}}\left(\frac{\ell_1\ell_2}{d^2}\right)\right|\right\}.\end{aligned}\tag{4.31}$$

Substituting (4.30) (note that the k-sum converges by (3.5)) and changing the order of summation, this is

$$\begin{aligned}&\ll_{s,t_0,\varepsilon}q^{\varepsilon}\left\{q^{-c_2}\sum_{d}\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{c_1}|x(d\ell_1)x(d\ell_2)|+\sum_{d}\sum_{\ell_1,\ell_2}(\ell_1\ell_2)^{-1/2}|x(d\ell_1)x(d\ell_2)|\right\}\\ &\ll_{s,t_0,\varepsilon}q^{\varepsilon}\left(L^{2c_1+1/2}q^{-c_2}+1\right)\sum_{\ell}\tau(\ell)|x(\ell)|^2,\end{aligned}$$

where we used Cauchy–Schwarz twice. By (2.44), we obtain from the last two displays

$$|L(f_0,s)|^4 \ll_{s,t_0,\varepsilon} q^{1+\varepsilon}\left(L^{2c_1}q^{-c_2}+L^{-1/2}\right).$$

Choosing

$$L := q^{c_2/(2c_1+1/2)}, \tag{4.32}$$

this gives Theorem 1.2 with

$$L(f_0,s) \ll_{s,t_0,\varepsilon} q^{\frac{1}{4}-\frac{c_2}{4(4c_1+1)}+\varepsilon}. \tag{4.33}$$

It remains to show (4.30) and calculate the constants c_1 and c_2. This will be done in the next three sections.

4.4 Applying the summation formulae

As a first step we substitute (4.22) into the definition (4.28) of $\mathcal{Q}_k^{\text{holo}}(\ell)$ and $\mathcal{Q}(\ell)$. Then we apply (2.8) and the corresponding formula for the divisor function in order to remove the factors $\lambda_f(\ell)$ and $\lambda_{\chi_1,\chi_2}(t,\ell)$. Applying (2.10), this gives

$$\begin{aligned} k^{-18}\mathcal{Q}_k^{\text{holo}}(\ell) = &\sum_{de=\ell}\frac{\chi(d)}{\sqrt{d}}\sum_{ab=d}\frac{\mu(a)\tau(b)}{\sqrt{a}}\sum_{m,n}\frac{\tau(m)\tau(n)}{(mn)^{1/2}} \\ &\times \overline{k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{m}{q}\right)}k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{adn}{q}\right)\frac{i^k\Gamma(k-1)\sqrt{maen}}{2\pi(4\pi)^{k-1}}\sum_{f\in\mathcal{B}_k^h(q,\chi)}\overline{\rho_f(m)}\rho_f(aen) \end{aligned}$$

and

$$\begin{aligned} \mathcal{Q}(\ell) = &\sum_{de=\ell}\frac{\chi(d)}{\sqrt{d}}\sum_{ab=d}\frac{\mu(a)\tau(b)}{\sqrt{a}}\sum_{m,n}\frac{\tau(m)\tau(n)}{(mn)^{1/2}} \\ &\times\Bigg\{\sum_{f\in\mathcal{B}_\kappa(q,\chi)}\hat{\varphi}_0(t_f)\overline{V_{t_f}\left(\frac{m}{q}\right)}V_{t_f}\left(\frac{adn}{q}\right)\frac{4\sqrt{maen}}{\cosh(\pi t_f)}\overline{\rho_f(m)}\rho_f(aen) \\ &+\sum_{\substack{\chi_1\chi_2=\chi \\ f\in\mathcal{B}_\kappa(\chi_1,\chi_2)}}\sum\int_{-\infty}^{\infty}\hat{\varphi}_0(t)\overline{V_t\left(\frac{m}{q}\right)}V_t\left(\frac{adn}{q}\right)\frac{\sqrt{maen}}{\pi\cosh(\pi t)}\overline{\rho_f(m,t)}\rho_f(aen,t)\,dt \\ &+\sum_{\substack{k\equiv\kappa\,(2),\ k>\kappa \\ f\in\mathcal{B}_k^h(q,\chi)}}\sum\dot{\varphi}_0(k)\overline{V_{\frac{(1-k)i}{2}}\left(\frac{m}{q}\right)}V_{\frac{(1-k)i}{2}}\left(\frac{adn}{q}\right)\frac{\Gamma(k)\sqrt{maen}}{\pi(4\pi)^{k-1}}\overline{\rho_f(m)}\rho_f(aen)\Bigg\}. \end{aligned}$$

Substituting (4.20), we get something of the form

$$\begin{aligned} \mathcal{Q}(\ell) = \sum_{\nu,\xi=0}^{4}\cdots\Bigg\{&\sum_j\hat{\varphi}_0(t_f)\overline{\alpha_\nu(t_f)}\alpha_\xi(t_f)\overline{\left(x\frac{\partial}{\partial x}\right)^{2\nu}W\left(\frac{m}{q}\right)}\left(x\frac{\partial}{\partial x}\right)^{2\xi}W\left(\frac{adn}{q}\right)\cdots \\ &+\text{ Eisenstein contribution }+\text{ holomorphic contribution }\Bigg\}. \end{aligned} \tag{4.34}$$

Now we apply Kuznetsov's trace formula (2.21) for each term separately. Similarly, we apply Petersson's formula (2.15) for $\mathcal{Q}_k^{\text{holo}}(\ell)$. In the latter case we obtain a diagonal term which can be estimated trivially using (4.20) and (4.21):

$$\frac{i^k}{2\pi}\sum_{de=\ell}\frac{\chi(d)}{\sqrt{d}}\sum_{ab=d}\frac{\mu(a)\tau(b)}{a\sqrt{e}}\sum_n\frac{\tau(aen)\tau(n)}{n}\overline{k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{aen}{q}\right)}k^{-9}V_{\frac{(1-k)i}{2}}\left(\frac{adn}{q}\right)\ll_\varepsilon q^\varepsilon\ell^{-1/2}. \tag{4.35}$$

Here and henceforth we suppress the dependence on s and t_0 and merely make sure that it is polynomial at most. In either case the off-diagonal term is a linear combination of terms of the form

$$\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c} \sum_{m,n} \frac{\tau(m)\tau(n)}{(mn)^{1/2}} W_1\left(\frac{m}{q}\right) W_2\left(\frac{a^2bn}{q}\right) S_\chi(m, aen; c)\varphi\left(\frac{4\pi\sqrt{aemn}}{c}\right). \quad (4.36)$$

where φ is J_{k-1} or a suitable φ as in (4.24)–(4.25) (with $a := A$, $b := 10$, $\alpha := \overline{\alpha_\nu}\alpha_\xi$ and $d := 8$), cf. (4.27). In particular, by (4.21), (4.24), (3.4), (4.26),

$$W_{1,2}^{(i)}(x) \ll_{\varepsilon,B,i} x^{-i-\varepsilon}(1+x)^{-B}, \quad \varphi^{(i)}(x) \ll_{A,i} \left(\frac{x}{1+x}\right)^{A-10-i}, \quad \varphi^*(u) \ll (1+|\Im u|)^{-2-2\Re u} \quad (4.37)$$

for all i with some very large A, B and for all u in a wide vertical strip symmetric about the origin.

Let us now open the Kloosterman sum and apply Voronoi summation (4.3) to the m-variable. It is one of the main features of the Voronoi summation here that the twisted Kloosterman sum becomes a Gauss sum. Let

$$G_\chi(h;c) := \sum_{\substack{d \pmod c\\ (d,c)=1}} \chi(d)e\left(\frac{hd}{c}\right)$$

denote the Gauss sum, then the term (4.36) decomposes into the sum of a "diagonal" first term

$$\begin{aligned}\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2} \sum_{n} \frac{\tau(n)G_{\overline{\chi}}(aen;c)}{n^{1/2}} W_2\left(\frac{a^2bn}{q}\right) \\ \times \int_0^\infty L_c(y) W_1\left(\frac{y}{q}\right)\varphi\left(\frac{4\pi\sqrt{aeny}}{c}\right)\frac{dy}{y^{1/2}},\end{aligned} \quad (4.38)$$

and of an "off-diagonal" second term given by

$$\sum_{\pm}\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2} \sum_{h} G_{\overline{\chi}}(h;c) \sum_{aen\pm m=h} \tau(m)\tau(n)g^\pm(n,m;c), \quad (4.39)$$

where

$$g^\pm(n,m;c) := \frac{1}{n^{1/2}} W_2\left(\frac{a^2bn}{q}\right) \int_0^\infty \mathcal{J}^\pm\left(\frac{\sqrt{my}}{c}\right) W_1\left(\frac{y}{q}\right)\varphi\left(\frac{4\pi\sqrt{aeny}}{c}\right)\frac{dy}{y^{1/2}} \quad (4.40)$$

for $c \geqslant q$.

Using the weak bound (cf. (4.54))

$$|G_{\overline{\chi}}(h;c)| \leqslant c^{1/2}(c,h)^{1/2},$$

the fact that $(\ell, q) = 1$ and also the inequalities (4.37) (cf. (4.42)), we obtain that (4.38) is bounded by

$$\ll_\varepsilon q^\varepsilon \sum_{\substack{q|c\\ c\leqslant \ell^{1/2}q^{1+\varepsilon}}} \frac{1}{c^2} \sum_{n\leqslant q^{1+\varepsilon}} \frac{c^{1/2}(c,\ell n)^{1/2}}{n^{1/2-\varepsilon}} q^{1/2+\varepsilon} \ll_\varepsilon q^{3\varepsilon-1/2}, \quad (4.41)$$

As for the term (4.39), let us attach a smooth factor $\psi(m)$ to $g^\pm$ that is zero for $m \leqslant 1/2$ and 1 for $m \geqslant 3/4$. This does not affect the sum (4.39). We need this little technicality in order to apply (4.4) later. It is easy to see that $g^\pm(n,m;c)$ is negligible (i.e., $\ll q^{-C}$ for any constant $C > 0$) unless

$$\frac{q^{1-\varepsilon}}{ae} =: N^- \leqslant n \leqslant N^+ := \frac{q^{1+\varepsilon}}{a^2b}, \qquad c \leqslant \frac{\sqrt{e}q^{1+\varepsilon}}{\sqrt{ab}}, \qquad m \leqslant aenq^\varepsilon. \quad (4.42)$$

The upper bound on n follows directly from (4.37) by choosing A and B large enough. By (4.37) we can also assume that $cq^{-\varepsilon} \leqslant \sqrt{aeny}$ and $y \leqslant q^{1+\varepsilon}$. Combining these inequalities, we obtain $c^2q^{-3\varepsilon} \leqslant qaen$ which implies the lower bound on n and, in combination with the upper bound on

n, it implies the upper bound on c as well. Finally, the upper bound on m follows from (6.14) by choosing a large j there. As a by-product, we can see that the integral in (4.40) is essentially supported on $[q^{1-\varepsilon}e^{-1}, q^{1+\varepsilon}]$, hence by applying a crude bound for the Bessel functions in that integral (e.g. Proposition 6.2) we obtain

$$g^{\pm}(n,m;c) \ll_{\varepsilon} q^{1/2+\varepsilon}n^{-1/2} \quad \text{for} \quad n \leqslant q^{1+\varepsilon} \text{ and } c \geqslant q. \tag{4.43}$$

Let $\mathcal{S}(a,b,e,c;q)$ denote the weighted sum of shifted convolution sums

$$\mathcal{S}(a,b,e,c;q) := \sum_h G_{\overline{\chi}}(h;c) \sum_{\pm} \sum_{aen \pm m = h} \tau(m)\tau(n)g^{\pm}(n,m;c)\psi(m).$$

Thus (4.39) equals

$$\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2}\mathcal{S}(a,b,e,c;q). \tag{4.44}$$

Remark 4.2. Since we have assumed that χ is not trivial, $G_{\overline{\chi}}(0;c) = 0$, hence in $\mathcal{S}(a,b,e,c;q)$ the h-sum varies over the $h \neq 0$. When χ is trivial, the degenerate contribution corresponding to $h = 0$,

$$\mathcal{S}_0(a,b,e,c;q) := \varphi(c) \sum_{aen=m} \tau(m)\tau(n)g^{-}(n,m;c),$$

yields a main term which can be bounded by $\ll_{\varepsilon} q^{\varepsilon}\ell^{-1/2}$. We do not carry out this computation in this paper and rather refer to [KMV00, Section 3.6].

Applying (4.4) with

$$Q := N^{+} = \frac{q^{1+\varepsilon}}{a^2 b} \tag{4.45}$$

to the innermost sum, $\mathcal{S}(a,b,e,c;q)$ splits into a main term

$$\begin{aligned} \mathcal{S}^{M}(a,b,e,c;q) := \sum_{h\neq 0} G_{\overline{\chi}}(h;c) \sum_{\pm} \sum_{w=1}^{\infty} \frac{(ae,w)r_w(h)}{w^2} \\ \times \int_0^{\infty} L_w(\pm(h-aex))K_{(ae,w),w}(x)g^{\pm}\left(x, \pm(h-aex);c\right)\psi(\pm(h-aex))\,dx. \end{aligned} \tag{4.46}$$

and two error terms of the shape

$$\begin{aligned} \mathcal{S}^{E,\pm}(a,b,e,c;q) := \sum_{h\neq 0} G_{\overline{\chi}}(h;c) \sum_{\pm} \sum_{w=1}^{\infty} \frac{(ae,w)}{w^2} \sum_{n=1}^{\infty} \tau(n)S\left(\mp h, \pm n; w\right) \\ \times \int_0^{\infty} \mathcal{J}^{\pm}\left(\frac{\sqrt{n(\pm(h-aex))}}{w}\right) K_{(ae,w),w}(x)g^{\pm}(x,\pm(h-aex);c)\psi(\pm(h-aex))\,dx \end{aligned} \tag{4.47}$$

for various combinations of $\pm$. We postpone the estimation of (4.47) to Section 4.6, and start with the contribution of (4.46) to $\mathcal{S}(a,b,e,c;q)$. At this point, we need to remove the catalyst function $\psi(m)$ in (4.46) and define

$$\begin{aligned} \tilde{\mathcal{S}}^{M}(a,b,e,c;q) := \sum_{h\neq 0} G_{\overline{\chi}}(h;c) \sum_{\pm} \sum_{w=1}^{\infty} \frac{(ae,w)r_w(h)}{w^2} \\ \times \int_0^{\infty} L_w(\pm(h-aex))K_{(ae,w),w}(x)g^{\pm}\left(x, \pm(h-aex);c\right)\,dx. \end{aligned} \tag{4.48}$$

The integrands in the two terms $\tilde{\mathcal{S}}^{M}$ and $\mathcal{S}^{M}$ differ only for $x = h/(ae) + O(1/(ae))$. Since by (4.42) (cf. (4.60) below) the h-sum in both terms is essentially over $1 \leqslant |h| \leqslant eq^{1+\varepsilon}/(ab)$, the contribution of their difference to (4.44) is at most (cf. (4.1), (4.6), (4.42), (4.43))

$$\ll_{\varepsilon} q^{\varepsilon} \sum_{ae|\ell} \sum_{q|c} \frac{1}{c^2} \sum_{1\leqslant h \leqslant eq^{1+\varepsilon}} c^{1/2}(h,c)^{1/2} \sum_{w=1}^{\infty} \frac{(ae,w)(h,w)}{w^2}\left(\frac{q}{aeh}\right)^{1/2} \ll_{\varepsilon} q^{3\varepsilon-1/2}. \tag{4.49}$$

4.5 The main term

In this section, we will evaluate the contribution of the term (4.48) to (4.44):

$$\begin{aligned}&\sum_{abe=\ell}\frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}}\sum_{w\geqslant 1}\frac{(ae,w)}{w^2}\\ &\times\sum_{q\mid c}\frac{1}{c^2}\sum_{h\neq 0}r_w(h)G_{\overline{\chi}}(h;c)\sum_{\pm}\int_0^\infty L_w(\pm(h-aex))K_{(ae,w),w}(x)g^{\pm}\left(x,\pm(h-aex);c\right)\,dx.\end{aligned}\tag{4.50}$$

More precisely, we shall first evaluate the c- and h-sums above then average trivially over a, b, e, w.

To do so we proceed essentially as in [KMV00, pp. 117–122]. We substitute the definition (4.40) of $g^{\pm}$ and make a change of variables

$$\xi := \frac{|h|}{c^2}y, \qquad \eta := \frac{ae}{|h|}x$$

in order to remove all parameters from the oscillating functions. Secondly, we replace the negative values of h in (4.50) (which only contribute to the "$-$" case in $\sum_{\pm}$) by their absolute values. To simplify the notation, let us write (cf. (4.1))

$$\mathcal{L}(\eta) := L_w(h\eta) = \log\eta + 2\gamma + \log\left(\frac{h}{w^2}\right) =: \log\eta + \Lambda,$$

say. Then the c,h-sum in (4.50) equals

$$\begin{aligned}&\frac{1}{\sqrt{ae}}\sum_{q\mid c}\frac{1}{c}\sum_{h\geqslant 1}r_w(h)G_{\overline{\chi}}(h;c)\int_0^\infty\int_0^\infty\varphi(4\pi\sqrt{\xi\eta})\\ &\times\left\{\delta_{\eta<1}\mathcal{L}(1-\eta)\mathcal{J}^+(\sqrt{(1-\eta)\xi})+\delta_{\eta>1}\mathcal{L}(\eta-1)\mathcal{J}^-(\sqrt{(\eta-1)\xi})+\chi(-1)\mathcal{L}(\eta+1)\mathcal{J}^-(\sqrt{(\eta+1)\xi})\right\}\\ &\times K_{(ae,w),w}\left(\frac{h\eta}{ae}\right)W_1\left(\frac{c^2\xi}{hq}\right)W_2\left(\frac{abh\eta}{eq}\right)\frac{d\xi d\eta}{(\xi\eta)^{1/2}}.\end{aligned}\tag{4.51}$$

Let us also write

$$X_w(\eta) := K_{(ae,w),w}\left(\frac{q\eta}{a^2b}\right)W_2(\eta).$$

Its Mellin transform $\widetilde{X}_w$ satisfies essentially the same properties as $\widetilde{W}_2$. To see this, observe first that by (4.37), W_2 is up to a negligible error supported on $[0,q^{\varepsilon}]$, so we can replace $K_{(ae,w),w}\left(q\eta/(a^2b)\right)$ by

$$K_w^*(\eta) := K_{(ae,w),w}\left(\frac{q\eta}{a^2b}\right)\omega\left(\frac{\eta}{q^{\varepsilon}}\right),$$

where, as usual, ω is a smooth cut-off function. Then, by (4.6), (4.45), and sufficiently many integrations by parts, we find that

$$\widetilde{K}_w^*(u) = \int_0^\infty K_w^*(\eta)\eta^{u-1}\,d\eta \ll_{j,\Re u} q^{\varepsilon}|u|^{-j}$$

for $\Re u>0$ and any $j\geqslant 0$. Finally, by (4.21),

$$\widetilde{X}_w(u) = \frac{1}{2\pi i}\int_{(\frac{1}{2}\Re u)}\widetilde{K}_w^*(u-v)\widetilde{W}_2(v)\,dv \ll_{j,\varepsilon} q^{\varepsilon}|u|^{-j}$$

for $\varepsilon\leqslant\Re u\leqslant 5$, say.

Our next aim is to transform the double integral in (4.51) by several applications of Mellin's inversion formula: using (4.2) and (2.28), we write $\mathcal{J}^{\pm}$ and φ as inverse Mellin transforms. Then the

ξ, η-integral in (4.51) equals

$$\begin{aligned}
&\int_0^\infty \int_0^\infty \frac{4}{(2\pi i)^2} \int_{(0.2)} \int_{(*)} \varphi^*(u_1)(2\pi\sqrt{\xi\eta})^{1+2u_1}(2\pi)^{-2u_2}\Gamma(u_2)^2\xi^{-u_2} \\
&\times \left(\frac{\delta_{\eta<1}\mathcal{L}(1-\eta)}{(1-\eta)^{u_2}} + \frac{\delta_{\eta>1}\mathcal{L}(\eta-1)\cos\pi u_2}{(\eta-1)^{u_2}} + \frac{\chi(-1)\mathcal{L}(\eta+1)\cos\pi u_2}{(\eta+1)^{u_2}} \right) du_2\, du_1 \\
&\times W_1\left(\frac{c^2\xi}{hq}\right) X_w\left(\frac{abh\eta}{eq}\right) \frac{d\xi d\eta}{(\xi\eta)^{1/2}}.
\end{aligned} \tag{4.52}$$

Since the u_1-, u_2- and ξ-integrals are absolutely convergent (using (4.37)), we can pull the ξ-integration inside and calculate it explicitly in terms of the Mellin transform $\widetilde{W}_1$ of W_1. Then we write X_w as an inverse Mellin transform getting that (4.52) equals

$$\begin{aligned}
&\int_0^\infty \frac{4}{(2\pi i)^2} \int_{(0.2)} \int_{(0.6)} \varphi^*(u_1)(2\pi)^{1+2u_1-2u_2}\eta^{1/2+u_1}\Gamma(u_2)^2 \\
&\times \left(\frac{\delta_{\eta<1}\mathcal{L}(1-\eta)}{(1-\eta)^{u_2}} + \frac{\delta_{\eta>1}\mathcal{L}(\eta-1)\cos\pi u_2}{(\eta-1)^{u_2}} + \frac{\chi(-1)\mathcal{L}(\eta+1)\cos\pi u_2}{(\eta+1)^{u_2}} \right) \left(\frac{c^2}{hq}\right)^{-1-u_1+u_2} \\
&\times \widetilde{W}_1\left(1+u_1-u_2\right) du_2\, du_1 \,\frac{1}{2\pi i}\int_{(0.9)} \left(\frac{abh\eta}{eq}\right)^{-u_3} \widetilde{X}_w(u_3)\, du_3\, \frac{d\eta}{\eta^{1/2}}.
\end{aligned}$$

Here we shifted the u_2-integration to $\Re u_2 = 0.6$ since $\widetilde{W}_1(u)$ is rapidly decaying on the line $\Re u = 0.6$. Since again all integrals are absolutely convergent, we can pull the η-integration inside and calculate the three terms explicitly (as in [KMV00, (38)]) using [GR07, 3.191.1, 3.191.2, 3.194.3]. We find

$$\begin{aligned}
&\int_0^\infty \left(\frac{\delta_{\eta<1}\mathcal{L}(1-\eta)}{(1-\eta)^{u_2}} + \frac{\delta_{\eta>1}\mathcal{L}(\eta-1)\cos\pi u_2}{(\eta-1)^{u_2}} + \frac{\chi(-1)\mathcal{L}(\eta+1)\cos\pi u_2}{(\eta+1)^{u_2}} \right) \eta^{1+u_1-u_3}\frac{d\eta}{\eta} \\
&= \cos(\pi u_2)(-\partial_{u_2}+\Lambda)\frac{\Gamma(1+u_1-u_3)\Gamma(-1-u_1+u_3+u_2)}{\Gamma(u_2)} \left(\chi(-1) + \frac{\sin(\pi(u_3-u_1))}{\sin(\pi u_2)}\right) \\
&- (-\partial_{u_2}+\Lambda)\frac{\Gamma(1+u_1-u_3)\Gamma(-1-u_1+u_3+u_2)}{\Gamma(u_2)}\frac{\sin(\pi(-u_1+u_3+u_2))}{\sin(\pi u_2)}.
\end{aligned}$$

Introducing the new variable $u_4 := 1+u_1-u_2$ as a substitute for u_2, we see that (4.52) equals

$$\begin{aligned}
&\frac{-4}{(2\pi i)^3}\int_{(0.2)}\int_{(0.6)}\int_{(0.9)} \varphi^*(u_1)(2\pi)^{2u_4-1}\Gamma(1+u_1-u_4)^2\Gamma(1+u_1-u_3) \\
&\times \left(\frac{c^2}{hq}\right)^{-u_4} \widetilde{W}_1\left(u_4\right) \left(\frac{abh}{eq}\right)^{-u_3} \widetilde{X}_w(u_3) \\
&\times \Bigg\{ \cos(\pi(u_1-u_4))(\partial_{u_4}+\Lambda)\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\left(\chi(-1)+\frac{\sin(\pi(u_3-u_1))}{\sin(\pi(u_4-u_1))}\right) \\
&\qquad -(\partial_{u_4}+\Lambda)\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\frac{\sin(\pi(u_3-u_4))}{\sin(\pi(u_4-u_1))}\Bigg\}\, du_3\, du_4\, du_1.
\end{aligned} \tag{4.53}$$

Using the identities

$$\begin{aligned}
(\partial_{u_4}+\Lambda)\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)} &= \frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(u_3-u_4)+\Lambda\right), \\
\frac{\sin(\pi(u_3-u_4))}{\sin(\pi(u_4-u_1))} &= -\cos(\pi(u_1-u_3)) + \cos(\pi(u_1-u_4))\frac{\sin(\pi(u_3-u_1))}{\sin(\pi(u_4-u_1))},
\end{aligned}$$

it is straightforward to verify that the last two lines in (4.53) can be simplified to

$$\begin{aligned}
&\frac{\Gamma(u_3-u_4)}{\Gamma(1+u_1-u_4)}\Bigg\{-\pi\sin(\pi(u_1-u_3)) + \\
&\qquad \Big(\cos(\pi(u_1-u_3)) + \chi(-1)\cos(\pi(u_1-u_4))\Big)\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(u_3-u_4)+\Lambda\right)\Bigg\}.
\end{aligned}$$

In particular, we observe that the triple integral is absolutely convergent (since φ^*, $\widetilde{W}_1$, and $\widetilde{X}_w$ are sufficiently nice) and the integrand is holomorphic whenever $0 < \Re u_4 < \Re u_3 < 1 + \Re u_1$. Let us shift the u_4-contour to $\Re u_4 = \varepsilon\,(< 0.1)$ and the u_3-contour to $\Re u_3 = 1.1$.

We now substitute this triple integral back into (4.51) and perform the (absolutely convergent) sum over c and h. To justify this, we need to evaluate for $s = 1 + 2u_4$, $t = u_3 - u_4$ the Dirichlet series

$$D_{w,q}(\chi, s, t) := \sum_{q|c} \frac{1}{c^s} \sum_{h \geqslant 1} \frac{G_{\overline{\chi}}(h; c) r_w(h)}{h^t}.$$

First we need to compute the Gauss sum $G_{\overline{\chi}}(h; c)$: we denote by q^* the conductor of χ and, slightly abusing notation, we write χ also for the primitive character of modulus q^* underlying χ mod q. For $q \mid c$ we consider the unique factorization $c = q^* q_1 q_2 c_1 c_2$ where $q = q^* q_1 q_2$, $c_1 q_1 \mid (q^*)^\infty$ and $(c_2 q_2, q^*) = 1$. Then

$$G_{\overline{\chi}}(h; c) = \overline{\chi}(c_2 q_2) G_{\overline{\chi}}(h; q^* q_1 c_1) r_{c_2 q_2}(h)$$

(with $r_{c_2 q_2}(h)$ being the Ramanujan sum, cf. (4.5)). Moreover, $G_{\overline{\chi}}(h; q^* q_1 c_1) = 0$ unless $c_1 q_1 \mid h$ in which case

$$G_{\overline{\chi}}(h; q^* q_1 c_1) = \chi\left(\frac{h}{c_1 q_1}\right) c_1 q_1 G_{\overline{\chi}}(1; q^*).$$

Summarizing the above computation, one has

$$G_{\overline{\chi}}(h; c) = \delta_{c_1 q_1 | h} \overline{\chi}(c_2 q_2) \chi\left(\frac{h}{c_1 q_1}\right) c_1 q_1 r_{c_2 q_2}(h) G_{\overline{\chi}}(1; q^*). \tag{4.54}$$

Therefore

$$D_{w,q}(\chi, s, t) = \frac{\overline{\chi}(q_2) G_{\overline{\chi}}(1; q^*)}{(q^*)^s q_1^{s+t-1} q_2^s} \sum_{c_1 | (q^*)^\infty} \frac{1}{c_1^{s+t-1}} \sum_{(c_2, q^*) = 1} \frac{\overline{\chi}(c_2)}{c_2^s} \sum_{h \geqslant 1} \frac{\chi(h) r_w(c_1 q_1 h) r_{c_2 q_2}(h)}{h^t}.$$

For $\sigma := \Re s$, and $\tau := \Re t$ sufficiently large the c_1, c_2, h-sum factors as an Euler product over the primes:

$$\sum_{c_1 | (q^*)^\infty} \frac{1}{c_1^{s+t-1}} \sum_{(c_2, q^*) = 1} \frac{\overline{\chi}(c_2)}{c_2^s} \sum_{h \geqslant 1} \frac{\chi(h) r_w(c_1 q_1 h) r_{c_2 q_2}(h)}{h^t} = \prod_p \Pi_p(\chi, s, t), \tag{4.55}$$

say. We collected some useful properties of the Euler factors $\Pi_p(\chi, s, t)$ in Lemma 4.1 at the end of this section. These properties imply that for $\Re u_4 = \varepsilon\,(< 0.1)$ and $\Re u_3 = 1.1$ the series $D_{w,q}(\chi, s, t)$ is absolutely convergent and in the domain $\sigma > 1$, $\tau > 0$ it decomposes as

$$D_{w,q}(\chi, s, t) = \zeta(s + t - 1) L(\chi, t) H_{w,q}(\chi, s, t),$$

where $H_{w,q}(\chi, s, t)$ is a holomorphic function. Moreover, for $0 < \varepsilon < 0.1$,

$$\Re s = 1 + 2\varepsilon, \qquad \varepsilon/2 < \Re t < 3\varepsilon/2,$$

one has

$$H_{w,q}(\chi, s, t) \ll_\varepsilon q^\varepsilon (q_1, w) w^{1-\varepsilon/3} (q^*)^{-1/2}. \tag{4.56}$$

Using

$$\Lambda = 2\gamma - \log\bigl(w^2\bigr) + \log(h)$$

we obtain that (4.51) equals

$$\begin{aligned}\frac{1}{\sqrt{ae}} \frac{-4}{(2\pi i)^3} \int_{(0.2)} \int_{(\varepsilon)} \int_{(1.1)} & \varphi^*(u_1) (2\pi)^{2u_4 - 1} \Gamma(1 + u_1 - u_4) \Gamma(1 + u_1 - u_3) \Gamma(u_3 - u_4) \\ & \times q^{u_4} \widetilde{W}_1(u_4) \left(\frac{ab}{eq}\right)^{-u_3} \widetilde{X}_w(u_3) \left\{ \sum_{j=0}^{1} \partial_{u_3}^j \bigl(\zeta(u_3 + u_4) \widetilde{L}(u_3, u_4)\bigr) F_j \right\} du_3\, du_4\, du_1,\end{aligned}$$

where

$$\widetilde{L}(u_3,u_4) := L(\chi, u_3-u_4)H_{w,q}(\chi, 1+2u_4, u_3-u_4)$$

and

$$F_0(u_1,u_3,u_4) := -\pi\sin(\pi(u_1-u_3)) + \Big(\cos(\pi(u_1-u_3)) + \chi(-1)\cos(\pi(u_1-u_4))\Big) \\ \times\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(u_3-u_4) + 2\gamma - \log(w^2)\right),$$

$$F_1(u_1,u_3,u_4) := -\Big(\cos(\pi(u_1-u_3)) + \chi(-1)\cos(\pi(u_1-u_4))\Big).$$

Let us now shift the u_3-contour from $\Re u_3 = 1.1$ to $\Re u_3 = 2\varepsilon$; we will show below that there is no pole at $u_3+u_4=1$. Then $\Re(u_3-u_4)=\varepsilon$, hence

$$\partial_{u_3}^j L(\chi, u_3-u_4) \ll_{j,\varepsilon} (q^*)^{1/2}$$

by the functional equation for $L(\chi,t)$ with implied constants depending on j, ε and (polynomially) on $|\Im(u_3-u_4)|$. In addition, (4.56) combined with Cauchy's integral formula shows that

$$\partial_{u_3}^j H_{w,q}(\chi, 1+2u_4, u_3-u_4) \ll_{j,\varepsilon} q^{\varepsilon}(q_1,w)w^{1-\varepsilon/3}(q^*)^{-1/2},$$

therefore (4.51) summed over $abe=\ell$ is bounded by

$$\ll_\varepsilon q^{10\varepsilon}(q_1,w)\log(w)w^{1-\varepsilon/3}\sum_{abe=\ell}\frac{(e/ab)^{\Re u_3}}{a\sqrt{abe}} \ll_\varepsilon q^{12\varepsilon}(q_1,w)w^{1-\varepsilon/4}\ell^{-1/2}.$$

Finally, averaging over w the above bound against the weight $(ae,w)/w^2$, we obtain that the main term (4.50) is bounded by

$$\ll_\varepsilon q^{13\varepsilon}\ell^{-1/2}. \tag{4.57}$$

To conclude the analysis of the main term, it remains to show that the pole of the zeta-function at $u_3+u_4=1$ does not contribute anything. Let us only focus on the factors depending on u_3:

$$G(u_1,u_3,u_4) := \Gamma(1+u_1-u_3)\Gamma(u_3-u_4)\left(\frac{ab}{eq}\right)^{-u_3}\widetilde{X}_w(u_3)\left\{\sum_{j=0}^{1}\partial_{u_3}^j\big(\zeta(u_3+u_4)\widetilde{L}(u_3,u_4)\big)F_j\right\}.$$

If R_j denotes the contribution of the j-term to the residue of $G(u_1,u_3,u_4)$ at $u_3 = 1-u_4$, then

$$R_0 = \Gamma(u_1+u_4)\Gamma(1-2u_4)\left(\frac{ab}{eq}\right)^{u_4-1}\widetilde{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4)\times\Bigg\{+\pi\sin(\pi(u_1+u_4)) + \\ \left\{\begin{array}{l} +2\sin(\pi u_1)\sin(\pi u_4) \\ -2\cos(\pi u_1)\cos(\pi u_4)\end{array}\right\}\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(1-2u_4) + 2\gamma - \log(w^2)\right)\Bigg\},$$

$$R_1 = \Gamma(u_1+u_4)\Gamma(1-2u_4)\left(\frac{ab}{eq}\right)^{u_4-1}\widetilde{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4)\times\Bigg\{-\pi\sin(\pi(u_1+u_4)) + \\ \left\{\begin{array}{l} +2\sin(\pi u_1)\sin(\pi u_4) \\ -2\cos(\pi u_1)\cos(\pi u_4)\end{array}\right\}\left(-\frac{\Gamma'}{\Gamma}(u_1+u_4) + \frac{\Gamma'}{\Gamma}(1-2u_4) + \frac{\widetilde{X}_w'}{\widetilde{X}_w}(1-u_4) - \log\left(\frac{ab}{eq}\right)\right)\Bigg\}.$$

Here the upper line corresponds to $\kappa=0$ and the lower line to $\kappa=1$, and we have used $\chi(-1)=(-1)^\kappa$. Altogether the residual integral equals, after shifting the u_1-integration to $(-\varepsilon/2)$ and interchanging the u_1- and u_4-integration,

$$\frac{8}{(2\pi i)^2}\int_{(\varepsilon)}\int_{(-\varepsilon/2)}\varphi^*(u_1)(2\pi)^{2u_4-1}\Gamma(1+u_1-u_4)\Gamma(u_1+u_4)\Gamma(1-2u_4) \\ \times q^{u_4}\widetilde{W_1}(u_4)\left(\frac{ab}{eq}\right)^{u_4-1}\widetilde{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4) \\ \times\left\{\begin{array}{l} -\sin(\pi u_1)\sin(\pi u_4) \\ +\cos(\pi u_1)\cos(\pi u_4)\end{array}\right\}\left(\frac{\Gamma'}{\Gamma}(1+u_1-u_4) - \frac{\Gamma'}{\Gamma}(u_1+u_4) + \widetilde{\Lambda}(u_4)\right)du_1\,du_4, \tag{4.58}$$

where

$$\widetilde{\Lambda}(u_4) := \frac{\widetilde{X}'_w}{\widetilde{X}_w}(1-u_4) + 2\gamma - \log\left(\frac{abw^2}{eq}\right).$$

We recast the inner integral as

$$\frac{1}{2\pi i}\int_{(-\varepsilon/2)} \varphi^*(u_1)\left(\widetilde{\Lambda}(u_4) - \partial_{u_4}\right)\Gamma(1+u_1-u_4)\Gamma(u_1+u_4)\begin{Bmatrix} -\sin(\pi u_1) \\ +\cos(\pi u_1)\end{Bmatrix} du_1$$

and use (2.29) to see that (4.58) equals

$$\begin{aligned}&\frac{8q}{2\pi i}\int_{(\frac{1}{2})}(2\pi)^{2u_4-1}\Gamma(1-2u_4)\widetilde{W_1}(u_4)\left(\frac{ab}{e}\right)^{u_4-1}\widetilde{X}_w(1-u_4)\widetilde{L}(1-u_4,u_4)\\ &\times\begin{Bmatrix}\sin(\pi u_4)\\ \cos(\pi u_4)\end{Bmatrix}\left(\widetilde{\Lambda}(u_4)-\partial_{u_4}\right)\hat{\varphi}\left(i\left(\tfrac{1}{2}-u_4\right)\right)\begin{Bmatrix}1\\ \frac{\cot(\pi u_4)}{i(\frac{1}{2}-u_4)}\end{Bmatrix}du_4.\end{aligned}$$

If $\varphi = J_{k-1}$, $k \equiv \kappa \pmod 2$ then the integral vanishes by $\hat{\varphi} = 0$. Otherwise we shift ∂_{u_4} to the other factors by partial integration. Then we sum over ν as in (4.34) and recall that, by the definition of φ and W_1,

$$\widetilde{W_1}(u_4) = u_4^{2\nu}\widetilde{\widetilde{W}}(u_4) \qquad \text{and} \qquad \hat{\varphi}(t) = \hat{\varphi}_0(t)\overline{\alpha_\nu(t)}\alpha_\xi(t).$$

For $t \in \mathbb{R}$ we have $\alpha_\nu(t) \in \mathbb{R}$, hence the sum over ν introduces factors

$$\sum_{\nu=0}^{4}\alpha_\nu\left(i\left(\tfrac{1}{2}-u_4\right)\right)u_4^{2\nu} \qquad \text{or} \qquad \sum_{\nu=0}^{4}\alpha_\nu\left(i\left(\tfrac{1}{2}-u_4\right)\right)\frac{\partial}{\partial u_4}u_4^{2\nu}$$

to each term. By (4.15)–(4.16) these factors vanish, that is, the residual integral (4.58) is zero in all cases. This completes the analysis of the main term.

Without the additional zeros in the approximate functional equation, we might still succeed at the cost of much more work. Applying the functional equation of $L(s,\chi)$, expressing $K_{(ae,w),w}(y)$ in terms of $L_{\frac{w}{(ae,w)}}(y)$ and therefore X_w in terms of W, it should be possible to see that the polar contribution (4.58) resembles exactly the contribution of the cusps $\mathfrak{a} = 0, \infty$ of $\mathcal{Q}(\ell)$, see (4.29).

We conclude this section by stating and proving some useful properties for the Euler factors $\Pi_p(\chi,s,t)$ in (4.55).

Lemma 4.1. *Let $\sigma = \Re s > 1$ and $\tau = \Re t > 0$. For a prime p let v_p denote the p-adic valuation, and let ζ_p (resp. L_p) denote the corresponding Euler factor of the Riemann zeta function (resp. Dirichlet L-function).*

a) For $(p,qw) = 1$,

$$\Pi_p(\chi,s,t) = \zeta_p(s+t-1)\frac{L_p(\chi,t)}{L_p(\overline{\chi},s)}.$$

b) For $p \mid q^*$,

$$|\Pi_p(\chi,s,t)| \leqslant 3p^{\min(v_p(q_1),v_p(w))+(1-\tau)v_p(w)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

c) For $(p,q^*) = 1$, $p \mid qw$,

$$|\Pi_p(\chi,s,t)| \leqslant 4p^{v_p(q_2)+(1-\tau)v_p(w)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

Proof. a) For $(p,qw) = 1$ we use the notation

$$\alpha := v_p(c_2), \qquad \beta := v_p(h)$$

in the sum (4.55), then

$$\begin{aligned}
\Pi_p(\chi,s,t) &= \sum_{\alpha=0}^{\infty}\frac{\overline{\chi}(p^\alpha)}{p^{\alpha s}}\sum_{\beta=0}^{\infty}\frac{\chi(p^\beta)r_{p^\alpha}(p^\beta)}{p^{\beta t}}\\
&= \sum_{\beta=0}^{\infty}\frac{\chi(p^\beta)}{p^{\beta t}}\left(1+\sum_{\alpha=1}^{\beta}\frac{\overline{\chi}(p^\alpha)}{p^{\alpha s}}(p^\alpha-p^{\alpha-1})-\frac{\overline{\chi}(p^{\beta+1})}{p^{(\beta+1)s}}p^\beta\right)\\
&= \frac{1-\overline{\chi}(p)p^{-s}}{1-\overline{\chi}(p)p^{1-s}}\sum_{\beta=0}^{\infty}\frac{\chi(p^\beta)}{p^{\beta t}}\left(1-\frac{\overline{\chi}(p^{\beta+1})}{p^{(\beta+1)s}}p^{\beta+1}\right)\\
&= \frac{1-\overline{\chi}(p)p^{-s}}{1-\overline{\chi}(p)p^{1-s}}\left(\frac{1}{1-\chi(p)p^{-t}}-\frac{\overline{\chi}(p)p^{1-s}}{1-p^{1-s-t}}\right)\\
&= \frac{1-\overline{\chi}(p)p^{-s}}{(1-\chi(p)p^{-t})(1-p^{1-s-t})}.
\end{aligned}$$

b) For $p \mid q^*$ we use the notation

$$\alpha := v_p(c_1), \qquad \beta := v_p(h), \qquad \gamma := v_p(q_1), \qquad \delta := v_p(w)$$

in the sum (4.55), then clearly

$$|\Pi_p(\chi,s,t)| \leqslant \sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\sum_{\beta=0}^{\infty}\frac{\left|r_{p^\delta}(p^{\alpha+\beta+\gamma})\right|}{p^{\beta\tau}}.$$

We distinguish between two cases. For $\gamma \geqslant \delta$ we infer

$$|\Pi_p(\chi,s,t)| \leqslant \sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\sum_{\beta=0}^{\infty}\frac{p^\delta}{p^{\beta\tau}} = p^\delta\zeta_p(\sigma+\tau-1)\zeta_p(\tau).$$

For $\gamma < \delta$ we infer

$$\begin{aligned}
|\Pi_p(\chi,s,t)| &\leqslant \sum_{\alpha=0}^{\delta-\gamma-1}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\left(\frac{p^{\delta-1}}{p^{(\delta-\gamma-1-\alpha)\tau}}+\sum_{\beta=\delta-\gamma-\alpha}^{\infty}\frac{p^\delta}{p^{\beta\tau}}\right)+\sum_{\alpha=\delta-\gamma}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\sum_{\beta=0}^{\infty}\frac{p^\delta}{p^{\beta\tau}}\\
&= p^\gamma\sum_{\alpha=0}^{\delta-\gamma-1}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\left(\frac{p^{\delta-\gamma-1}}{p^{(\delta-\gamma-1-\alpha)\tau}}+\frac{p^{\delta-\gamma}}{p^{(\delta-\gamma-\alpha)\tau}}\zeta_p(\tau)\right)+p^\delta\zeta_p(\tau)\sum_{\alpha=\delta-\gamma}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\\
&\leqslant 2p^{\gamma+(\delta-\gamma)(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma-1)+p^{\gamma+(\delta-\gamma)(2-\sigma-\tau)}\zeta_p(\tau)\zeta_p(\sigma+\tau-1).
\end{aligned}$$

In both cases we conclude

$$|\Pi_p(\chi,s,t)| \leqslant 3p^{\min(\gamma,\delta)+\delta(1-\tau)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

c) For $(p,q^*)=1$, $p \mid qw$, we use the notation

$$\alpha := v_p(c_2), \qquad \beta := v_p(h), \qquad \gamma := v_p(q_2), \qquad \delta := v_p(w)$$

in the sum (4.55), then clearly

$$|\Pi_p(\chi,s,t)| \leqslant \sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha\sigma}}\sum_{\beta=0}^{\infty}\frac{\left|r_{p^\delta}(p^\beta)r_{p^{\alpha+\gamma}}(p^\beta)\right|}{p^{\beta\tau}}.$$

We distinguish between two cases. For $\gamma \geqslant \delta$ we infer (note that $\gamma > 0$ in this case)

$$\begin{aligned}|\Pi_p(\chi,s,t)| &\leqslant \sum_{\alpha=0}^{\infty}\frac{p^\delta}{p^{\alpha\sigma}}\left(\frac{p^{\alpha+\gamma-1}}{p^{(\alpha+\gamma-1)\tau}}+\sum_{\beta=\alpha+\gamma}^{\infty}\frac{p^{\alpha+\gamma}}{p^{\beta\tau}}\right)\\ &\leqslant p^\delta\zeta_p(\tau)\sum_{\alpha=0}^{\infty}\frac{1}{p^{\alpha\sigma}}\left(\frac{p^{\alpha+\gamma-1}}{p^{(\alpha+\gamma-1)\tau}}+\frac{p^{\alpha+\gamma}}{p^{(\alpha+\gamma)\tau}}\right)\\ &\leqslant 2p^{\delta+\gamma(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma+\tau-1)\\ &\leqslant 2p^{\gamma+\delta(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma+\tau-1).\end{aligned}$$

For $\gamma < \delta$ we infer

$$\begin{aligned}|\Pi_p(\chi,s,t)| &\leqslant \sum_{\alpha=0}^{\delta-\gamma-1}\frac{p^{\alpha+\gamma}}{p^{\alpha\sigma}}\left(\frac{p^{\delta-1}}{p^{(\delta-1)\tau}}+\sum_{\beta=\delta}^{\infty}\frac{p^\delta}{p^{\beta\tau}}\right)+\sum_{\alpha=\delta-\gamma}^{\infty}\frac{p^\delta}{p^{\alpha\sigma}}\left(\frac{p^{\alpha+\gamma-1}}{p^{(\alpha+\gamma-1)\tau}}+\sum_{\beta=\alpha+\gamma}^{\infty}\frac{p^{\alpha+\gamma}}{p^{\beta\tau}}\right)\\ &\leqslant 2p^{\gamma+\delta(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma-1)+2p^{\delta+\gamma(1-\tau)}\zeta_p(\tau)\sum_{\alpha=\delta-\gamma}^{\infty}\frac{1}{p^{\alpha(\sigma+\tau-1)}}\\ &= 2p^{\gamma+\delta(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma-1)+2p^{\delta-\sigma(\delta-\gamma)+\delta(1-\tau)}\zeta_p(\tau)\zeta_p(\sigma+\tau-1).\end{aligned}$$

In both cases we conclude

$$|\Pi_p(\chi,s,t)| \leqslant 4p^{\gamma+\delta(1-\tau)}\zeta_p(\sigma-1)\zeta_p(\tau).$$

The proof of Lemma 4.1 is complete. $\square$

4.6 The error term

Finally we estimate the contribution of (4.47) to (4.44). This time, we fix c and evaluate the h-sum nontrivially: in other words, we will bound the terms $\mathcal{S}^{E,\pm}(a,b,e,c;q)$ in (4.47) for c satisfying (cf. (4.42))

$$q \mid c, \qquad q \leqslant c \leqslant \frac{\sqrt{e}q^{1+\varepsilon}}{\sqrt{ab}}.$$

As a first step, we use the identity

$$\sum_{w\geqslant 1}F(w,(ae,w)) = \sum_{r\mid ae}\sum_{(ae,w)=r}F(w,r) = \sum_{rs\mid ae}\mu(s)\sum_{w\equiv 0\,(rs)}F(w,r)$$

and write (4.47) as

$$\begin{aligned}&\sum_{\pm}\sum_{rs\mid ae}r\mu(s)\sum_{w\equiv 0\,(rs)}\frac{1}{w^2}\sum_{h\neq 0}G_{\overline{\chi}}(h;c)\sum_{n=1}^{\infty}\tau(n)S(\mp h,\pm n;w)\\ &\times\int_0^\infty \mathcal{J}^{\pm}\left(\frac{\sqrt{n(\pm(h-aex))}}{w}\right)K_{r,w}(x)g^{\pm}(x,\pm(h-aex);c)\psi(\pm(h-aex))\,dx.\end{aligned}$$

We want to apply the trace formulae (2.21) and (2.22) to the w-sum. This needs some preparation. By (4.42) we can restrict the x-integration to

$$|h-aex| \leqslant aexq^{\varepsilon} \leqslant \frac{eq^{1+\varepsilon}}{ab} \tag{4.59}$$

and the h-summation to

$$|h| \leqslant \frac{eq^{1+\varepsilon}}{ab}, \tag{4.60}$$

up to negligible error. Let ρ be a smooth nonnegative function with bounded derivatives, supported on $[1/2,2]$ such that $\rho(y)+\rho(2y)=1$ for $y\in[1/2,1]$. Then $\sum_{\nu\in\mathbb{Z}}\rho(2^\nu y)=1$ for $y>0$. We apply this smooth partition of unity to all variables and insert (4.40); thus we will bound $O(\log^6 q)$ terms ((4.62), (4.64), (4.67) show that each of W, H, N, R, X, Y can be taken from the interval $[1/2,\ell^3q^{1+\varepsilon}]$), of the shape

$$\begin{aligned}&\sum_{rs\mid ae} r\mu(s)\sum_{w\equiv 0\,(rs)}\frac{\rho(w/W)}{w^2}\sum_h G_{\overline{\chi}}(h;c)\rho\left(\frac{|h|}{H}\right)\sum_n \rho\left(\frac{n}{N}\right)\tau(n)S(\mp h,\pm n;w)\\ &\times\int_0^\infty\int_0^\infty K_{r,w}(x)\rho\left(\frac{\pm(h-aex)}{R}\right)\rho\left(\frac{x}{X}\right)\rho\left(\frac{y}{Y}\right)W_1\left(\frac{y}{q}\right)W_2\left(\frac{a^2bx}{q}\right)\\ &\times\mathcal{J}^\pm\left(\frac{\sqrt{n(\pm(h-aex))}}{w}\right)\mathcal{J}^\pm\left(\frac{\sqrt{\pm(h-aex)y}}{c}\right)\varphi\left(\frac{4\pi\sqrt{aexy}}{c}\right)\frac{dydx}{(xy)^{1/2}}.\end{aligned}\tag{4.61}$$

(More precisely, for $1/2\leqslant R\leqslant 1$ we adjust the first ρ-factor by the function ψ as in the discussion following (4.41).) In view of (4.59), (4.42), (4.37), (4.7) and (4.45), and the remark following (4.41), we can assume

$$\frac12\leqslant R\leqslant aeXq^\varepsilon,\qquad \frac{q^{1-\varepsilon}}{ae}\leqslant X\leqslant\frac{q^{1+\varepsilon}}{a^2b},\qquad \frac{abq^{1-\varepsilon}}{e}\leqslant Y\leqslant q^{1+\varepsilon},\qquad \frac12\leqslant W\leqslant\frac{rq^{1/2+\varepsilon}}{a\sqrt{b}}.\tag{4.62}$$

Now we use (6.14) to integrate the first factor in the third line of (4.61) by parts sufficiently many times; in order to apply (6.14) we change variables $\mathfrak{r}:=\pm(h-aex)\asymp R$. By (4.6) and (4.37), the j-th derivative with respect to $\mathfrak{r}$ of the integrand without the $\mathcal{J}^\pm(\sqrt{n\mathfrak{r}}/w)$ factor is

$$\ll_{\varepsilon,j} q^\varepsilon\left(\frac1R+\frac{1}{Xae}+\frac{\sqrt{Y}}{c\sqrt{R}}+\frac{\sqrt{Y}}{c\sqrt{Xae}}\right)^j\ll_{\varepsilon,j} q^\varepsilon\left(\frac1R+\frac{\sqrt{Y}}{q\sqrt{R}}\right)^j.$$

This shows, by (6.14), that the integral in (4.61) is negligible unless

$$\frac{W}{\sqrt{N}}\left(\frac{1}{\sqrt{R}}+\frac{\sqrt{Y}}{q}\right)\geqslant q^{-\varepsilon}.\tag{4.63}$$

Note that this implies either $\sqrt{RN}/W\leqslant q^\varepsilon$ or $\sqrt{N}/W\leqslant q^{-1/2+\varepsilon}$ (since $Y\leqslant q^{1+\varepsilon}$), and so in any case

$$\frac{\sqrt{RN}}{W}\leqslant\sqrt{e}q^\varepsilon.\tag{4.64}$$

Let us now define

$$\begin{aligned}\Psi(h,n;z):=&\frac{z\rho(n/N)}{4\pi\sqrt{|h|n}}\rho\left(\frac{4\pi\sqrt{|h|n}}{zW}\right)\int_0^\infty\int_0^\infty K_{r,4\pi\sqrt{|h|n}/z}(x)\\ &\times\rho\left(\frac{\pm(h-aex)}{R}\right)\rho\left(\frac{x}{X}\right)\rho\left(\frac{y}{Y}\right)W_1\left(\frac{y}{q}\right)W_2\left(\frac{a^2bx}{q}\right)\\ &\times\mathcal{J}^\pm\left(z\frac{\sqrt{\pm(h-aex)}}{4\pi\sqrt{|h|}}\right)\mathcal{J}^\pm\left(\frac{\sqrt{\pm(h-aex)y}}{c}\right)\varphi\left(\frac{4\pi\sqrt{aexy}}{c}\right)\frac{dydx}{(xy)^{1/2}}.\end{aligned}\tag{4.65}$$

Then (4.61) equals

$$\sum_{rs\mid ae}r\mu(s)\sum_h G_{\overline{\chi}}(h;c)\rho\left(\frac{|h|}{H}\right)\sum_n\tau(n)\sum_{w\equiv 0\,(rs)}\frac1w S(\mp h,\pm n;w)\Psi\left(h,n;\frac{4\pi\sqrt{|h|n}}{w}\right).\tag{4.66}$$

We are now in a position to apply Kuznetsov's trace formula (2.21)–(2.22) for

$$\text{level } rs,\ \text{trivial nebentypus and weight } 0.$$

The innermost sum in (4.66) equals

$$\sum_{f\in\mathcal{B}_0(rs,1)} \hat{\Psi}(h,n;t_f)\frac{4\sqrt{|h|n}}{\cosh(\pi t_f)}\overline{\rho_f}(|h|)\rho_f(n) + \text{ two similar terms}$$

corresponding to holomorphic forms and Eisenstein series (or a similar expression with $\check{\Psi}$ in place of $\hat{\Psi}$). We substitute this into (4.66), and are left with bounding

$$\sum_{rs|ae} r\mu(s) \sum_{f\in\mathcal{B}_0(rs,1)} \sum_h G_{\overline{\chi}}(h;c)\rho\left(\frac{|h|}{H}\right)\sqrt{|h|}\overline{\rho_f}(|h|)\sum_n \tau(n)\sqrt{n}\rho_f(n)\frac{\hat{\Psi}(h,n;t_f)}{\cosh(\pi t_f)}$$

for

$$\frac{1}{2}\leqslant H\leqslant \frac{eq^{1+\varepsilon}}{ab}, \tag{4.67}$$

cf. (4.60). Finally we split the $f\in\mathcal{B}_0(rs,1)$-sum into dyadic pieces depending on the size of t_f: namely,

$$\sum_{f\in\mathcal{B}_0(rs,1)} = \sum_{|t_f|<1}\dots + \sum_\tau\sum_{|t_f|\asymp\tau}\cdots$$

for $\tau=2^k$, $k\geqslant 0$ an integer. Thus typically we need to bound sums of the form

$$\sum_{rs|ae} r\mu(s) \sum_{|t_f|\asymp\tau} \sum_h G_{\overline{\chi}}(h;c)\rho\left(\frac{|h|}{H}\right)\sqrt{|h|}\overline{\rho_f}(|h|)\sum_n \tau(n)\sqrt{n}\rho_f(n)\frac{\hat{\Psi}(h,n;t_f)}{\cosh(\pi t_f)}, \tag{4.68}$$

(plus one more sum with $\sum_{|t_f|\asymp\tau}$ replaced by $\sum_{|t_f|<1}$). Moreover, as we will see in Lemma 4.2 below, the contribution of the τ's greater than $q^\varepsilon\left(1+\frac{\sqrt{N}}{W}(\sqrt{H}+\sqrt{R})\right)$ is negligible.

It will be useful to separate the h, n, t_f variables; we proceed by partial summation: for $j\in\mathbb{N}_0$ let

$$\Xi_j(h,n;z) := \frac{\partial^j}{\partial h^j}\frac{\partial}{\partial n}\rho\left(\frac{|h|}{H}\right)\Psi(h,n;z); \tag{4.69}$$

note that differentiation commutes with taking Bessel transforms. Then by partial summation (4.68) equals a sum of two expressions (corresponding to the signs $\pm$)

$$\sum_{rs|ae} r\mu(s)\int_0^\infty\int_0^\infty \sum_{|t_f|\asymp\tau}\frac{\hat{\Xi}_1(\pm\mathfrak{h},\mathfrak{n};t_f)}{\cosh(\pi t_f)}\sum_{h\leqslant\mathfrak{h}} G_{\overline{\chi}}(\pm h;c)\sqrt{h}\overline{\rho_f}(h)\sum_{n\leqslant\mathfrak{n}}\tau(n)\sqrt{n}\rho_f(n)\,d\mathfrak{h}\,d\mathfrak{n}, \tag{4.70}$$

but we can also suppress the partial summation with respect to h getting two expressions (corresponding to the signs $\pm$)

$$-\sum_{rs|ae} r\mu(s)\int_0^\infty \sum_{|t_f|\asymp\tau}\sum_{h\geqslant 1} G_{\overline{\chi}}(\pm h;c)\sqrt{h}\overline{\rho_f}(h)\frac{\hat{\Xi}_0(\pm h,\mathfrak{n};t_f)}{\cosh(\pi t_f)}\sum_{n\leqslant\mathfrak{n}}\tau(n)\sqrt{n}\rho_f(n)\,d\mathfrak{n}. \tag{4.71}$$

We summarize the properties of $\hat{\Xi}_j(t)=\hat{\Xi}_j(h,n;t)$ in the following lemma.

Lemma 4.2. *Let*

$$Z := \frac{q^\varepsilon R\sqrt{Y}}{NWae\sqrt{X}}\left(1+\frac{\sqrt{RN}}{W}\right)^{-1/2} \tag{4.72}$$

and

$$\tilde{Z} := \min\left(1,\frac{\sqrt{HN}}{W},\frac{\sqrt{H}}{\sqrt{R}}\right). \tag{4.73}$$

Then for $n\asymp N$, $|h|\asymp H$ and for any $j\in\mathbb{N}_0$ we have

$$\dot{\Xi}_j(t),\ \hat{\Xi}_j(t),\ \check{\Xi}_j(t)\ll_{j,\varepsilon}\frac{Z}{1+|t|}\left(\frac{e}{H}\right)^j\tilde{Z}^{-2|\Im t|},$$

assuming $|\Im t| < 1/2$ *and* $t \in \mathbb{N}$ *in the case of* $\dot{\Xi}_j(t)$. *Moreover, all three functions are negligible unless*

$$|t| \leqslant q^{\varepsilon}\left(1+\frac{\sqrt{N}}{W}(\sqrt{H}+\sqrt{R})\right). \tag{4.74}$$

Proof. Let us first show that the function Ξ_j defined by (4.69) and (4.65) and supported on $z \asymp \sqrt{HN}/W$ satisfies

$$z^i\frac{\partial^i}{\partial z^i}\Xi_j(h,n;z) \ll_{i,j,\varepsilon} Z\left(\frac{e}{H}\right)^j\left(1+\frac{\sqrt{RN}}{W}\right)^i \tag{4.75}$$

for all $i,j \in \mathbb{N}_0$. To verify this we fix i and the sign of h and observe that, by the Leibniz rule for the operator $z^i(\partial^i/\partial z^i)$, the left hand side is a finite linear combination of integrals of the form (cf. (4.69) and (4.65))

$$\frac{\partial^j}{\partial h^j}\int_0^\infty\int_0^\infty A(h-aex,y)\ B\left(\frac{h-aex}{h}\right)C(h,x,y)\,dx\,dy, \tag{4.76}$$

where we have used an obvious abstract notation and suppressed the dependence on n, z for simplicity. In particular, $A:\mathbb{R}\setminus\{0\}\times(0,\infty)\to\mathbb{C}$ is a smooth function supported on a product of compact intervals $t \asymp \pm R$, $y \asymp Y$ satisfying

$$A(t,y) \ll_{\varepsilon} q^{\varepsilon}\left(1+\frac{\sqrt{RY}}{c}\right)^{-1/2},$$

$B(t) := z^k(\partial^k/\partial z^k)\mathcal{J}^{\pm}\big(z\sqrt{|t|}/(4\pi)\big)$ for some $0 \leqslant k \leqslant i$ satisfying in the relevant range (cf. (4.64))

$$t^s\frac{\partial^s}{\partial t^s}B(t) \ll_{s,i,\varepsilon} q^{\varepsilon}\left(1+\frac{\sqrt{RN}}{W}\right)^{s+i-1/2}, \qquad z\sqrt{|t|} \asymp \frac{\sqrt{RN}}{W} \leqslant \sqrt{e}q^{\varepsilon}, \tag{4.77}$$

and $C:(\mathbb{R}\setminus\{0\})\times(0,\infty)\times(0,\infty)\to\mathbb{C}$ is a smooth function supported on a product of compact intervals $h \asymp H$, $x \asymp X$, $y \asymp Y$ satisfying

$$H^rX^s\frac{\partial^r}{\partial h^r}\frac{\partial^s}{\partial x^s}C(h,x,y) \ll_{r,s,i,\varepsilon} \frac{q^{\varepsilon}}{NW\sqrt{XY}}\left(1+\frac{\sqrt{aeXY}}{c}\right)^s. \tag{4.78}$$

Now for $j \geqslant 1$ we rewrite (4.76) as

$$\begin{aligned}&\frac{\partial^{j-1}}{\partial h^{j-1}}\int_0^\infty\int_0^\infty\frac{\partial}{\partial h}\left\{A(h-aex,y)B\left(\frac{h-aex}{h}\right)\right\}C(h,x,y)\,dx\,dy\\ +&\frac{\partial^{j-1}}{\partial h^{j-1}}\int_0^\infty\int_0^\infty A(h-aex,y)\ B\left(\frac{h-aex}{h}\right)\frac{\partial}{\partial h}C(h,x,y)\,dx\,dy.\end{aligned}$$

The inner integral in the first term equals

$$\begin{aligned}&-\frac{1}{h}\int_0^\infty A(h-aex,y)\ B_0\left(\frac{h-aex}{h}\right)C(h,x,y)\,dx\\ &+\frac{1}{ae}\int_0^\infty A(h-aex,y)\ B\left(\frac{h-aex}{h}\right)C_0(h,x,y)\,dx,\end{aligned} \tag{4.79}$$

where

$$B_0(t) := t\frac{\partial}{\partial t}B(t), \qquad C_0(h,x,y) := \frac{\partial}{\partial x}C(h,x,y).$$

This decomposition is not obvious but follows easily by using the identities

$$\frac{\partial}{\partial h}A(h-aex,y) = -\frac{1}{ae}\frac{\partial}{\partial x}A(h-aex,y), \qquad \frac{\partial}{\partial h}B\left(\frac{h-aex}{h}\right) = -\frac{x}{h}\frac{\partial}{\partial x}B\left(\frac{h-aex}{h}\right),$$

and then integrating by parts in

$$\int_0^\infty \frac{\partial}{\partial x}\left\{A(h-aex,y)B\left(\frac{h-aex}{h}\right)\right\}C(h,x,y)\,dx.$$

From (4.77)–(4.79) we can see that (4.76) is a linear combination of 3 integrals of the form

$$\left(\frac{\sqrt{e}}{H}+\frac{1}{aeX}+\frac{\sqrt{Y}}{c\sqrt{aeX}}\right)\frac{\partial^{j-1}}{\partial h^{j-1}}\int_0^\infty\int_0^\infty A(h-aex,y)\;B_1\left(\frac{h-aex}{h}\right)C_1(h,x,y)\,dx\,dy,$$

where A is as before; B_1 and C_1 have the same support as B and C and satisfy the same bound as in (4.77) and (4.78), respectively. By (4.62) and (4.67) we see that

$$\frac{\sqrt{e}}{H}+\frac{1}{aeX}+\frac{\sqrt{Y}}{c\sqrt{aeX}}\ll_\varepsilon q^\varepsilon\frac{e}{H}.$$

By iterating this process we can finally decompose (4.76) as a linear combination of 3^j integrals of the form

$$\left(\frac{e}{H}\right)^j\int_0^\infty\int_0^\infty A(h-aex,y)\;B_j\left(\frac{h-aex}{h}\right)C_j(h,x,y)\,dx\,dy, \tag{4.80}$$

where A is as before; B_j and C_j have the same support as B and C and satisfy the same bound as in (4.77) and (4.78), respectively. By estimating the integral pointwise we obtain (4.75) immediately.

The lemma follows now from part a) of Lemma 2.2, if t is real and $\sqrt{RN}/W\leqslant q^\varepsilon$. If $\sqrt{RN}/W\geqslant q^\varepsilon$ then we look more closely at the first factor in the third line of (4.65). In the $\mathcal{J}^+$ case we are done by the rapid decay of the Bessel K-function. In the $\mathcal{J}^-$ case we use the asymptotic expansion of the Bessel Y-function to see that for large x,

$$\mathcal{J}^-(x)=\frac{1}{\sqrt{x}}e(2x)J_1(x)+\frac{1}{\sqrt{x}}e(-2x)J_2(x)$$

with smooth functions $J_{1,2}$ satisfying $J_{1,2}^{(j)}(x)\ll_j x^{-j}$. Now a similar argument as above together with part c) of Lemma 2.2 yields the proof of Lemma 4.2. A technical point to note here is that in this case we develop the above decomposition for $i=0$ only and then estimate the z-derivatives and the Bessel transforms inside the resulting integrals (4.80) individually. In our exposition we did not follow this path as we wanted to suppress the z-dependence for simplicity. Finally, if t is imaginary, part b) of Lemma 2.2 completes the proof of Lemma 4.2. □

We will bound separately the contribution of the τ's not exceeding a specific parameter $\mathcal{T}$ and of the τ's larger than this parameter. In the former case we shall use (4.71), in the latter (4.70).

4.7 The case of large spectral parameter

Using (4.54), Lemma 4.2 and Cauchy–Schwarz, (4.70) can be estimated from above by

$$(q^*)^{1/2}c_1q_1\sum_{rs\mid ae}r\sum_{d\mid c_2q_2}d\int_{\mathfrak{n}\asymp N}\int_{\mathfrak{h}\asymp H}\frac{Ze}{\tau H}\left(\sum_{|t_f|\asymp\tau}\frac{1}{\cosh(\pi t_f)}\left|\sum_{\substack{h\leqslant\mathfrak{h}/c_1q_1d\\(h,\frac{c_2q_2}{d})=1}}\chi(h)\sqrt{c_1q_1dh}\,\overline{\rho_f}(c_1q_1dh)\right|^2\right)^{\frac12}$$
$$\times\left(\sum_{|t_f|\asymp\tau}\frac{1}{\cosh(\pi t_f)}\left|\sum_{n\leqslant\mathfrak{n}}\tau(n)\sqrt{n}\rho_f(n)\right|^2\right)^{\frac12}d\mathfrak{h}\;d\mathfrak{n}.$$

Decompose d into d_2d_2' such that $d_2\mid q_2^\infty$ and $(d_2',q_2)=1$; then for f a Hecke eigenform one has (since $(rs,q)=1$)

$$\sqrt{c_1q_1dh}\,\overline{\rho_f}(c_1q_1dh)=\lambda_f(c_1q_1d_2)\sqrt{d_2'h}\,\overline{\rho_f}(d_2'h),$$

so that by the large sieve inequalities (2.46) one obtains that (4.70) is bounded by

$$\ll_\varepsilon q^\varepsilon (q^*)^{1/2}(c_1q_1)^{1+\theta}\sum_{rs|ae} r\sum_{d|c_2q_2} d_2'd_2^{1+\theta}\frac{Ze}{\tau H}HN \times\left(\tau+\left(\frac{H}{c_1q_1d_2rs}\right)^{1/2}\right)\left(\frac{H}{c_1q_1d}\right)^{1/2}\left(\tau+\left(\frac{N}{rs}\right)^{1/2}\right)N^{1/2}.$$

Here we clearly have the inequalities $r\leqslant rs\leqslant ae\leqslant \ell$,

$$\begin{aligned} d_2'd_2^{1+\theta}\left(\tau+\left(\frac{H}{c_1q_1d_2rs}\right)^{1/2}\right)\left(\frac{H}{c_1q_1d}\right)^{1/2} &= {d_2'}^{1/2}d_2^{\theta}\left(d_2^{1/2}\tau+\left(\frac{H}{c_1q_1rs}\right)^{1/2}\right)\left(\frac{H}{c_1q_1}\right)^{1/2}\\ &\leqslant (c_2q_2)^{1/2+\theta}\left(\tau+\left(\frac{H}{c_1q_1q_2r}\right)^{1/2}\right)\left(\frac{H}{c_1q_1}\right)^{1/2}, \end{aligned}$$

and

$$r\left(\tau+\left(\frac{H}{c_1q_1q_2r}\right)^{1/2}\right)\left(\tau+\left(\frac{N}{r}\right)^{1/2}\right)\leqslant \ell\left(\tau+\left(\frac{H}{c_1q_1q_2\ell}\right)^{1/2}\right)\left(\tau+\left(\frac{N}{\ell}\right)^{1/2}\right).$$

Using these and the definition (4.72) of Z, we obtain, according to (4.62), (4.63), (4.67), (4.74), that (4.70) is bounded by

$$\begin{aligned} &\ll_\varepsilon q^\varepsilon(q^*)^{1/2}(c_1c_2q_1q_2)^{1/2+\theta}\frac{\ell^{3/2}R}{\tau W}\frac{Y^{1/2}}{X^{1/2}}q^{1/2}N^{1/2}\left(\tau+\frac{q^{1/2}}{(c_1q_1q_2)^{1/2}}\right)\left(\tau+\frac{N^{1/2}}{\ell^{1/2}}\right)\\ &\ll_\varepsilon q^\varepsilon c^{1/2}\left(\frac{c}{q^*}\right)^{\theta}\ell^2\frac{R}{\tau W}q^{1/2}N^{1/2}\left(\tau+(q^*)^{1/2}\right)\left(\tau+\frac{N^{1/2}}{\ell^{1/2}}\right). \end{aligned}\tag{4.81}$$

Let us recall that (by (4.62), (4.63), (4.67), (4.74))

$$1\leqslant q^\varepsilon\frac{W}{N^{1/2}}\left(\frac{1}{R^{1/2}}+\frac{1}{q^{1/2}}\right),\qquad \tau\leqslant q^\varepsilon\left(1+\frac{\ell^{1/2}q^{1/2}N^{1/2}}{W}\right),\qquad W\leqslant q^\varepsilon \ell q^{1/2};$$

we observe that the first two conditions imply that

$$\tau\leqslant q^\varepsilon \ell^{1/2}\left(1+\left(\frac{q}{R}\right)^{1/2}\right).$$

If we assume that

$$\tau\geqslant \ell^{1/2}\mathcal{T}\qquad\text{for some}\qquad \mathcal{T}\gg q^\varepsilon,$$

then

$$R^{1/2}\ll q^{1/2+\varepsilon}\mathcal{T}^{-1}\ll q^{1/2},$$

and in particular, $(RN)^{1/2}\ll q^\varepsilon W$. Now we bound the four terms of the product

$$\frac{R}{\tau W}q^{1/2}N^{1/2}\left(\tau+(q^*)^{1/2}\right)\left(\tau+\frac{N^{1/2}}{\ell^{1/2}}\right)$$

in (4.81):

$$\begin{gathered} \tau\frac{RN^{1/2}}{W}q^{1/2}\ll_\varepsilon \tau q^\varepsilon R^{1/2}q^{1/2}\ll_\varepsilon \ell^{1/2}q^{1+\varepsilon},\\ \frac{RN^{1/2}}{W}(q^*)^{1/2}q^{1/2}\ll_\varepsilon q^\varepsilon R^{1/2}(q^*)^{1/2}q^{1/2}\ll_\varepsilon \frac{(q^*)^{1/2}q^{1+\varepsilon}}{\mathcal{T}},\\ \frac{RN}{W}\frac{q^{1/2}}{\ell^{1/2}}\ll_\varepsilon q^\varepsilon W\frac{q^{1/2}}{\ell^{1/2}}\ll_\varepsilon \ell^{1/2}q^{1+\varepsilon},\\ \frac{1}{\tau}\frac{RN}{W\ell^{1/2}}(q^*)^{1/2}q^{1/2}\ll_\varepsilon q^\varepsilon\frac{W}{\tau\ell^{1/2}}(q^*)^{1/2}q^{1/2}\ll_\varepsilon \frac{(q^*)^{1/2}q^{1+\varepsilon}}{\mathcal{T}}. \end{gathered}$$

The same argument works for holomorphic forms and Eisenstein series and gives the same estimates. Therefore the total contribution of large eigenvalues to the sum (cf. (4.44))

$$\sum_{abe=\ell} \frac{\chi(ab)\mu(a)\tau(b)}{a\sqrt{b}} \sum_{q|c} \frac{1}{c^2} \mathcal{S}^{E,\pm}(a,b,e,c;q) \tag{4.82}$$

is bounded by

$$\ll_\varepsilon q^\varepsilon \left(\frac{\ell^2}{\mathcal{T}} \left(\frac{q^*}{q}\right)^{1/2-\theta} + \frac{\ell^{5/2}}{q^{1/2}} \left(\frac{q^*}{q}\right)^{-\theta} \right). \tag{4.83}$$

4.8 The case of small spectral parameter

The estimate (4.83) is useful if τ is not too small, that is, if $\mathcal{T}$ is at least some small power of q. In fact we shall later specify $\mathcal{T}$ so that $\log \mathcal{T} \asymp \log q$. In view of the preceding section, we suppose that

$$0 \leqslant \tau \leqslant \ell^{1/2}\mathcal{T}. \tag{4.84}$$

For such small τ we use (4.71) which can be bounded by

$$\ll_\varepsilon q^\varepsilon \sum_{rs|ae} r \int_{\mathfrak{n}\asymp N} \left(\sum_{|t_f|\asymp\tau} \left| \sum_{h\geqslant 1} G_{\overline{\chi}}(h;c)\sqrt{h}\overline{\rho_f}(h) \frac{\hat{\Xi}_0(\pm h,\mathfrak{n};t_f)}{\sqrt{\cosh(\pi t_f)}} \right|^2 \right)^{1/2} \left(\tau\sqrt{N} + \frac{N}{\sqrt{rs}}\right) d\mathfrak{n}, \tag{4.85}$$

using Cauchy–Schwarz and the large sieve (2.46).

For $f \in \mathcal{B}_0(rs,1)$ (which we recall is a Hecke eigenform), let $\mathcal{L}(f,u)$ denote the Dirichlet series

$$\mathcal{L}(f,u) := \sum_{h\geqslant 1} \frac{G_{\overline{\chi}}(h;c)\sqrt{h}\rho_f(h)}{h^u}.$$

In the following we study this Dirichlet series in order to estimate the h-sum in (4.85). The Dirichlet series is absolutely convergent for $\Re u \gg 1$; by (4.54), one has

$$\begin{aligned}
\mathcal{L}(f,u) &= \sum_{h_1|(rsc)^\infty} \sum_{(h_2,rsc)=1} \frac{G_{\overline{\chi}}(h_1;c)\chi(h_2)\sqrt{h_1h_2}\rho_f(h_1h_2)}{h_1^u h_2^u} \\
&= L^{(rsc)}(f\otimes\chi,u) \times \overline{\chi}(c_2q_2)G_{\overline{\chi}}(1;q^*)(c_1q_1)^{1-u} \sum_{h|(rsc)^\infty} \frac{r_{c_2q_2}(h)\sqrt{c_1q_1h}\rho_f(c_1q_1h)\chi(h)}{h^u} \\
&= L^{(rsc)}(f\otimes\chi,u) \times \mathcal{H}(f,u),
\end{aligned}$$

say, with

$$L^{(rsc)}(f\otimes\chi,u) := \sum_{(h,rsc)=1} \frac{\lambda_f(h)\chi(h)}{h^u}$$

and

$$\mathcal{H}(f,u) := \overline{\chi}(c_2q_2)G_{\overline{\chi}}(1;q^*)(c_1q_1)^{1-u} \sum_{d|c_2q_2} d^{1-u}\chi(d)\mu\left(\frac{c_2q_2}{d}\right) \sum_{h|(rsc)^\infty} \frac{\sqrt{c_1q_1dh}\rho_f(c_1q_1dh)\chi(h)}{h^u}.$$

On the one hand,

$$L^{(rsc)}(f\otimes\chi,u) = \prod_{p|rsc} \left(1 - \frac{\lambda_{\tilde{f}}(p)\chi(p)}{p^s} + \frac{\chi(p^2)}{p^{2s}}\right) \times L(\tilde{f}\otimes\chi,u)$$

where $\tilde{f}$ is the newform (of level dividing rs) underlying the Hecke eigenform f (and with the same spectral parameter t_f). Applying a subconvex bound of the form

$$L(f\otimes\chi,s) \ll_\varepsilon (|s|\mu_f Nq)^\varepsilon |s|^\alpha \mu_f^\beta N^\gamma q^{\frac{1}{2}-\delta} \tag{4.86}$$

one has

$$L^{(rsc)}(f\otimes\chi,u)\ll_\varepsilon(|u|(1+|t_f|)rsc)^\varepsilon|u|^\alpha(\tau)^\beta(rs)^\gamma(q^*)^{1/2-\delta};\tag{4.87}$$

in particular, we remark that (1.5) is applicable if (cf. (4.84))

$$q^*\geqslant(\ell^{3/2}\mathcal{T})^4\geqslant(rs\tau)^4.\tag{4.88}$$

On the other hand, $\mathcal{H}(f,u)$ is holomorphic for $\Re u\geqslant 1/2$ and satisfies in this domain the uniform bound

$$\mathcal{H}(f,u)\ll(q^*c_1q_1)^{1/2}\sum_{d|c_2q_2}d^{1/2}\sum_{h|(rsc)^\infty}\frac{\sqrt{c_1q_1dh}|\rho_f(c_1q_1dh)|}{h^{1/2}},\tag{4.89}$$

cf. (5.33). By Mellin inversion, the h-sum in (4.85) equals, without the factor $\sqrt{\cosh(\pi t_f)}$ and after replacing $f(z)$ by $\overline{f(-\bar z)}$,

$$\frac{1}{2\pi i}\int_{(1/2)}\mathcal{L}(f,u)\left(\int_0^\infty\hat{\Xi}_0(\pm x,\mathfrak{n};t_f)\,x^{u-1}\,dx\right)du.$$

By partial integration and Lemma 4.2 we see

$$\int_0^\infty\hat{\Xi}_0(x,\mathfrak{n};t_f)\,x^{u-1}\,dx\ll_\varepsilon\frac{Z\sqrt{H}}{1+\tau}\left(\frac{e}{|u|}\right)^\nu\tilde{Z}^{-2|\Im t_f|}\tag{4.90}$$

on $\Re u=1/2$, for any $\nu\geqslant 0$ (at first for integer ν, but then by convexity also for real ν). We choose $\nu:=\alpha+1+\varepsilon$ in order to ensure absolute convergence of the u-integral. Using Cauchy–Schwarz and (2.45), we see that

$$\begin{aligned}\left(\sum_{|t_f|\asymp\tau}\left|\sum_{h|(rsc)^\infty}\frac{\sqrt{c_1q_1dh}|\rho_f(c_1q_1dh)|}{h^{1/2}\sqrt{\cosh(\pi t_f)}}\right|^2\right)^{\frac12}&\ll_\varepsilon(rsc)^\varepsilon\sum_{h|(rsc)^\infty}\left(\frac{1}{h^{1-\varepsilon}}\sum_{|t_f|\asymp\tau}\frac{c_1q_1dh|\rho_f(c_1q_1dh)|^2}{\cosh(\pi t_f)}\right)^{\frac12}\\&\ll_\varepsilon c^\varepsilon(c_1q_1d)^\theta(1+\tau),\end{aligned}\tag{4.91}$$

where $\theta=7/64<1/2$ (cf. (2.42)). Collecting the estimates (4.87), (4.89), (4.90), (4.91), we can bound (4.85) by

$$\begin{aligned}&\ll_\varepsilon\ q^\varepsilon\sum_{rs|ae}re^{\alpha+1}\,(\tau)^\beta\,(rs)^\gamma(q^*)^{1-\delta}\sum_{d|c_2q_2}(c_1q_1d)^{1/2+\theta}NZ\sqrt{H}\left(\tau\sqrt{N}+\frac{N}{\sqrt{rs}}\right)\tilde{Z}^{-2\theta_0}\\&\ll_\varepsilon\ c^{2\varepsilon}\,(\tau)^\beta\,\ell^{2+\alpha+\gamma}(q^*)^{1-\delta}\left(\frac{c}{q^*}\right)^{1/2+\theta}NZ\sqrt{H}\left(\tau\sqrt{N}+\frac{N}{\sqrt{\ell}}\right)\tilde{Z}^{-2\theta_0},\end{aligned}\tag{4.92}$$

where $\theta_0=0$ if $\tau\geqslant 1$ and $\theta_0=\theta$ if $\tau\leqslant 1$.

If $\tau\geqslant 1$, $\tilde{Z}^{-2\theta_0}=1$ and we use the bound (4.84) to obtain that (4.92) is at most

$$c^{2\varepsilon}\mathcal{T}^\beta\ell^{2+\alpha+\beta/2+\gamma}(q^*)^{1-\delta}\left(\frac{c}{q^*}\right)^{1/2+\theta}NZ\sqrt{H}\left(\tau\sqrt{N}+\frac{N}{\sqrt{\ell}}\right).\tag{4.93}$$

We deal now with the sum (4.85) where the summation $\sum_{|t_f|\asymp\tau}$ is replaced by $\sum_{|t_f|<1}$: we recall that $\tilde{Z}$ depends on H according to (4.73), so that (4.92) is an increasing function of H. Thus we estimate (4.92) from above using (4.67). But then, together with (4.62), we see that $\tilde{Z}\geqslant\ell^{-1/2}q^{-\varepsilon}$ so that $\tilde{Z}^{-2\theta_0}\leqslant q^\varepsilon\ell^\theta$; in that case however, there is no factor $(\mathcal{T}\sqrt{\ell})^\beta$. Since $\beta\geqslant 3/8>2\theta$, the contribution of $|t_f|<1$ is dominated by (4.93).

Using (4.72), and the bound (cf. (4.62) and (4.67))

$$\frac{\sqrt{Y}}{ae\sqrt{X}}\sqrt{H}\ll_\varepsilon q^{1/2+\varepsilon},$$

we are left with

$$c^{2\varepsilon}\mathcal{T}^{\beta}\ell^{2+\alpha+\beta/2+\gamma}(q^*)^{1-\delta}q^{1/2}\left(\frac{c}{q^*}\right)^{1/2+\theta}\frac{RN^{1/2}}{W}\left(1+\frac{\sqrt{RN}}{W}\right)^{-1/2}\left(\mathcal{T}\sqrt{\ell}+\frac{\sqrt{N}}{\sqrt{\ell}}\right),$$

subject to

$$\frac{\sqrt{RN}}{W}\leqslant\sqrt{\ell}q^{\varepsilon},\qquad R\leqslant\ell q^{1+\varepsilon},\qquad W\leqslant\ell q^{1/2+\varepsilon}.$$

Averaging over $c\equiv 0 \pmod q$, we see that the total contribution of small eigenvalues to (4.82) is at most

$$\begin{aligned}
&\ll_\varepsilon q^{\varepsilon}\ell^{\alpha+\frac{\beta}{2}+\gamma+2}\mathcal{T}^{\beta}\frac{(q^*)^{1/2-\theta-\delta}}{q^{1-\theta}}\frac{R^{\frac{3}{4}}N^{\frac{1}{4}}}{W^{\frac{1}{2}}}\left(\mathcal{T}\sqrt{\ell}+\frac{\sqrt{N}}{\sqrt{\ell}}\right)\\
&\ll_\varepsilon q^{\varepsilon}\ell^{\alpha+\frac{\beta}{2}+\gamma+2}\mathcal{T}^{\beta}\frac{(q^*)^{1/2-\theta-\delta}}{q^{1-\theta}}\left(R^{\frac{1}{2}}\mathcal{T}\ell^{\frac{3}{4}}+W\ell^{\frac{1}{4}}\right),\\
&\ll_\varepsilon q^{\varepsilon}\ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{13}{4}}\mathcal{T}^{\beta+1}\left(\frac{q^*}{q}\right)^{1/2-\theta}(q^*)^{-\delta}.
\end{aligned}\tag{4.94}$$

The same bound holds for holomorphic cusp forms. The case of Eisenstein series is somewhat different at least when they are parametrized by the cusps for their Fourier coefficients are not multiplicative anymore. Instead we proceed as in [Mi04, HM06] and calculate the coefficients directly. Unfolding the Gauss sum leads for each cusp $\mathfrak{a}=\frac{v}{w}$, $w\mid rs$, to the normalized series

$$\sum_h\frac{\chi(h)\sqrt{gh}\overline{\rho_{\mathfrak{a}}}(1/2+it,gh)}{h^u\sqrt{\cosh(\pi t)}},\tag{4.95}$$

where $g:=c_1q_1dd'$ and $dd'\mid c_2q_2$. By the computation of Section 5.7 this series can be written in terms of products of two Dirichlet L-functions $L(\chi\overline{\varphi},u-it)L(\chi\varphi,u+it)$ for certain characters φ having conductor dividing $\left(w,\frac{rs}{w}\right)$, times a holomorphic function in $\Re u\geqslant 1/2$ that is bounded on $\Re u=1/2$ by

$$\ll_\varepsilon (grs)^{\varepsilon}(g,w)\left(w,\frac{rs}{w}\right)^{1/2}(rs)^{-1/2}.$$

Here we used that $(rs,q)=1$. In particular, the function defined by (4.95) can be holomorphically continued to $\Re u\geqslant 1/2$ and on $\Re u=1/2$ it is bounded by

$$\ll_\varepsilon (q(1+|t|)|u|)^{\varepsilon}(|u|+|t|)^{3/8}(g,rs)\left(w,\frac{rs}{w}\right)^{7/8}(rs)^{-1/2}q^{3/8},$$

according to Heath-Brown's hybrid bound [HB80] for Dirichlet L-functions. Summing over all cusps of $\Gamma_0(rs)$ and noting that

$$\sum_{w\mid rs}\varphi\left(w,\frac{rs}{w}\right)\left(w,\frac{rs}{w}\right)^{7/8}(rs)^{-1/2}\ll_\varepsilon(rs)^{7/16+\varepsilon},$$

we obtain a bound of at least the same quality as in the case of Maass cusp forms if we assume $\alpha,\beta\geqslant 3/8$, $\gamma\geqslant 7/16$, $\delta\leqslant 1/8$. Then we proceed analogously.

4.9 Putting it all together

Collecting (4.35), (4.41), (4.49), (4.57), (4.83) and (4.94), we obtain that

$$\begin{aligned}
&k^{-18}\left|\mathcal{Q}_k^{\mathrm{holo}}(\ell)\right|+|\mathcal{Q}(\ell)|\\
&\quad\ll_{s,t_0,\varepsilon} q^{\varepsilon}\left(\frac{1}{\ell^{1/2}}+\frac{\ell^2}{\mathcal{T}}\left(\frac{q^*}{q}\right)^{1/2-\theta}+\frac{\ell^{5/2}}{q^{1/2}}\left(\frac{q^*}{q}\right)^{-\theta}+\ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{13}{4}}\mathcal{T}^{\beta+1}\left(\frac{q^*}{q}\right)^{1/2-\theta}(q^*)^{-\delta}\right)\\
&\quad\ll_{s,t_0,\varepsilon} q^{\varepsilon}\left(\frac{1}{\ell^{1/2}}+\frac{\ell^2}{\mathcal{T}}\left(\frac{q^*}{q}\right)^{1/2-\theta}+\ell^{\alpha+\frac{\beta}{2}+\gamma+\frac{13}{4}}\mathcal{T}^{\beta+1}(q^*)^{-\delta}\left(\frac{q^*}{q}\right)^{1/2-\theta}\right)
\end{aligned}\tag{4.96}$$

(in the first inequality above the last term is always larger than the third one).

Set $q^* = q^\eta$ with $\eta \in [0,1]$. If η is small (to be determined in a moment) we choose $\mathcal{T} := q^\varepsilon \sqrt{\ell}$ and apply the convexity bound (1.2) (cf. (4.86)) with

$$\alpha = \frac{1}{2}, \qquad \beta = \frac{1}{2}, \qquad \gamma = \frac{1}{4}, \qquad \delta = 0,$$

and so we arrive at we arrive at (4.30) with

$$c_1 := 5 \qquad \text{and} \qquad c_2 := (1-\eta)\left(\frac{1}{2} - \theta\right).$$

Substituting the expressions for c_1 and c_2 into (4.33) we obtain

$$L(f_0, s) \ll_{s,t_0,\varepsilon} q^{\frac{1}{4} - \frac{(1-\eta)(1-2\theta)}{168} + \varepsilon}. \tag{4.97}$$

If η is large, we use the exponents provided by the subconvex bound (1.5) (cf. (4.86)),

$$\alpha := \frac{1}{2}, \qquad \beta := 3, \qquad \gamma := \frac{1}{4}, \qquad \delta := \frac{1}{8},$$

assuming that (4.88) holds. Equating the second and third terms on the right hand side of (4.96) we choose

$$\mathcal{T} := (q^*)^{\frac{\delta}{\beta+2} + \varepsilon} \ell^{-\frac{\alpha+\beta/2+\gamma+5/4}{\beta+2}},$$

provided

$$q^{\eta\delta - \varepsilon} > \ell^{\alpha + \frac{\beta}{2} + \gamma + \frac{5}{4}} \tag{4.98}$$

(so that $\log \mathcal{T} \asymp \log q$), and provided

$$q^{\eta(\beta+2-4\delta)-\varepsilon} > \ell^{-4\alpha+4\beta-4\gamma+7} \tag{4.99}$$

(in order to satisfy (4.88)). Under these assumptions we obtain a total error term of

$$\ll_\varepsilon q^\varepsilon \left(\frac{1}{\ell^{1/2}} + \frac{\ell^{\frac{\alpha+5\beta/2+\gamma+21/4}{\beta+2}}}{q^{\eta\frac{\delta}{\beta+2} + (1-\eta)\left(\frac{1}{2}-\theta\right)}} \right) \ll_\varepsilon q^\varepsilon \left(\frac{1}{\ell^{1/2}} + \frac{\ell^{\frac{\alpha+5\beta/2+\gamma+21/4}{\beta+2}}}{q^{\frac{\delta}{\beta+2}}} \right),$$

since $\frac{1}{2} - \theta \geqslant \frac{\delta}{\beta+2}$ for any $\beta \geqslant 0$ and any $\delta \in [0, 1/2]$. Hence we arrive at (4.30) with

$$c_1 := \frac{\alpha + 5\beta/2 + \gamma + 21/4}{\beta + 2} \qquad \text{and} \qquad c_2 := \frac{\delta}{\beta+2}.$$

We choose L as in (4.32):

$$L := q^{c_2/(2c_1+1/2)}.$$

In (4.31) we apply (4.30) for $\ell \leqslant L^2$, and it is easily checked that (4.98) and (4.99) are satisfied as long as $\eta \geqslant 14/59$. Substituting the expressions for c_1 and c_2 into (4.33) we obtain

$$L(f_0, s) \ll_{s,t_0,\varepsilon} q^{\frac{1}{4} - \frac{\delta}{16\alpha+44\beta+16\gamma+92} + \varepsilon} \ll_{s,t_0} q^{\frac{1}{4} - \frac{1}{1889}}$$

for $\eta \geqslant 14/59$ while for $\eta \leqslant 14/59$ the bound (4.97) is stronger. This concludes the proof of Theorem 1.2.

Chapter 5

Rankin–Selberg L-functions

5.1 Approximate functional equation

For s on the critical line $\Re s = \frac{1}{2}$, we set (cf. (2.4))

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

By standard techniques (see [Mi04] for instance), one can show that for s with $\Re s = \frac{1}{2}$ and for any $A \geqslant 1$, one has a bound of the form

$$L(f \otimes g, s) \ll_A \log^2(qDP+1) \sum_N \frac{|L_{f\otimes g}(N)|}{\sqrt{N}} \left(1 + \frac{N}{qDP}\right)^{-A}, \tag{5.1}$$

where $N \geqslant 1$ ranges over the powers of 2, 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

$$x^j W^{(j)}(x) \ll_{j,A} P^j \tag{5.2}$$

for all $j \geqslant 0$. In particular, Theorem 1.3 follows from

Proposition 5.1. *Assume Hypothesis H_θ for any $0 \leqslant \theta \leqslant \frac{1}{2}$ and that $\chi_f\chi_g$ is nontrivial. Let B and $\delta_{\rm tw}$ be as in* (5.35)*. Then for any $0 < \varepsilon \leqslant 10^{-3}$ and any integer*

$$1 \leqslant N \leqslant (qDP)^{1+\varepsilon}, \tag{5.3}$$

one has

$$\frac{L_{f\otimes g}(N)}{\sqrt{N}} \ll_\varepsilon q^{100\varepsilon} q^{\frac{1}{2} - \frac{(1-2\theta)\delta_{\rm tw}}{88+18B-28\theta-80\theta^2}}.$$

The implied constant depends on ε and polynomially on 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 (5.1) of the N's such that $N \geqslant (qDP)^{1+\varepsilon}$ is bounded by

$$\ll_\varepsilon (qDP)^{200\varepsilon}.$$

For the remaining terms, we apply Proposition 5.1, getting

$$\begin{aligned} L(f\otimes g, s) &\ll_{\varepsilon,D,P} q^{100\varepsilon}\log^3(qDP+1) q^{\frac{1}{2} - \frac{(1-2\theta)\delta_{\rm tw}}{88B+18B-28\theta-80\theta^2}} \\ &\ll_{\varepsilon,D,P} q^{200\varepsilon} q^{\frac{1}{2} - \frac{(1-2\theta)\delta_{\rm tw}}{88B+18B-28\theta-80\theta^2}}. \end{aligned}$$

5.2 Amplification

As usual, the bound for $L_{f\otimes g}(N)$ in Proposition 5.1 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 $\mathcal{B}_k\big([q,D],\chi\big) = \{u_j\}_{j\geqslant 1}$ 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 5.1. 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 (2.15) 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 5.2 below.

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

$$(\ell, qD) \neq 1 \quad \Longrightarrow \quad x_\ell = 0. \tag{5.4}$$

For $f(z) \in \mathcal{L}_k\big([q,D],\chi\big)$ 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 [Mi04], it follows from (2.10) and (2.11) that for our original primitive form $f = f_0$, $L_{f\otimes g}(\vec{x}, N)$ factors as

$$L_{f_0\otimes g}(\vec{x}, N) = \rho_{f_0}(1) \left(\sum_{\ell \leqslant L} x_\ell \lambda_{f_0}(\ell) \right) L_{f_0\otimes g}(N). \tag{5.5}$$

Thus we form the "spectrally complete" quadratic form

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

where $\mathcal{H}(t)$ is as in Proposition 2.2, 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 5.2. *Assume Hypothesis H_θ for any $0 \leqslant \theta \leqslant \frac{1}{2}$, and let B and $\delta_{\rm tw}$ be as in* (5.35). *With the above notation, suppose that $\chi\chi_g$ is nontrivial and let $q^* > 1$ denote its conductor; moreover, suppose that g satisfies*

$$w \mid D \quad \Longrightarrow \quad q^* \nmid (w, D/w). \tag{5.6}$$

Then for any $1 \leqslant L \leqslant q$, any $0 < \varepsilon \leqslant 10^{-3}$ and any N satisfying (5.3)*, there is $A = A(\varepsilon)$ as in Proposition 2.2 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

$$\delta_L := \frac{41 + 8B - 14\theta - 40\theta^2}{6 + 2B}, \qquad \delta_q := \frac{1 - 2\theta}{3 + B}\, \delta_{\rm tw}$$

and

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

If g does not satisfy (5.6), *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

$$\delta_{3L} := 9 - \frac{17\theta + 20\theta^2}{3+B}, \qquad \delta_{3q} := \frac{1}{2} - \frac{2\theta}{3+B}\delta_{\mathrm{tw}},$$
$$\delta_{4L} := 13 - \frac{17\theta + 20\theta^2}{3+B}, \qquad \delta_{4q} := \frac{1+4\theta-2\delta_{\mathrm{tw}}}{4(1+\theta)}.$$

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

Proof of Proposition 5.1. As explained in Section 4 of [Mi04], Proposition 5.1 now follows from Proposition 5.2. Indeed, by (5.5) and by positivity, in particular by (2.16), one has

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

Moreover, for a Maass cusp form f_0 of weight $k \in \{0,1\}$, we have, by (5.5), (2.16) and (2.40),

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

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 (2.16) and (2.41),

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

for some absolute positive constant $C > 0$, the implied constant depending at most on ε. We suppose first that g satisfies (5.6); by Proposition 5.2, we have

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

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 (2.44) we have

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

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

$$L = L_0 := q^{\frac{2\delta_q}{1+2\delta_L}} = q^{\frac{2(1-2\theta)\delta_{\mathrm{tw}}}{44+9B-14\theta-40\theta^2}},$$

we conclude the proof of Proposition 5.1 when g satisfies (5.6). When g does not satisfy (5.6), 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 5.1 holds in that case, too. □

Remark 5.2. The above estimates prove Proposition 5.1 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 (2.15)) exactly as in the next section, and Proposition 5.2 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 (2.41)) yields Proposition 5.1 with a polynomial dependency in k_{f_0} as well.

5.3 Analysis of the quadratic form

We compute the quadratic form $Q(\vec{x}, N)$ by applying the spectral summation formula of Proposition 2.2. $Q(\vec{x}, N)$ decomposes into a diagonal part and a non-diagonal one:

$$\begin{aligned} Q(\vec{x}, N) = &\sum_{\ell_1,\ell_2} \overline{x_{\ell_1}} x_{\ell_2} \sum_{\substack{a_1b_1e_1=\ell_1 \\ a_2b_2e_2=\ell_2}} \mu(a_1)\mu(a_2)\chi\chi_g(\overline{a_1}a_2)\chi(\overline{b_1}b_2)\overline{\lambda_g}(b_1)\lambda_g(b_2) \\ &\times \left\{ S^{\mathrm{D}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N\right) + \sum_{c\equiv 0\,([q,D])} \frac{1}{c^2} S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N; c\right)\right\} \\ = &\, c_A Q^{\mathrm{D}}(\vec{x}, N) + Q^{\mathrm{ND}}(\vec{x}, N), \end{aligned} \tag{5.7}$$

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

$$\begin{aligned} &S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2 \end{pmatrix}, N; c\right) := \\ &\qquad c\sum_{m,n} \overline{\lambda_g}(m)\lambda_g(n) S_\chi(a_1e_1m, a_2e_2n; c)\mathcal{I}\left(\frac{4\pi\sqrt{a_1a_2e_1e_2mn}}{c}\right)\overline{W}(a_1d_1m)W(a_2d_2n). \end{aligned} \tag{5.8}$$

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 [Mi04, Section 4.1.1] yield

$$Q^{\mathrm{D}}(\vec{x}, N) \ll_\varepsilon (qNP)^\varepsilon N \sum_{d,\ell_1,\ell_2} |x_{d\ell_1}||x_{d\ell_2}| \frac{\sigma_g(\ell_1)\sigma_g(\ell_2)}{\sqrt{\ell_1\ell_2}} \ll_{g,\varepsilon} (qNP)^{2\varepsilon} N \|\vec{x}\|_2^2 \tag{5.9}$$

for any $\varepsilon > 0$.

We transform (5.8) further by applying the Voronoi summation formula of Proposition 2.3 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 (5.4) we have $(e, qD) = 1$, hence $D|[q,D]|c^*$. Again, the arguments of [Mi04, 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 (2.14) (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 [Mi04, Section 4.1.2] yields

$$S^{\mathrm{ND}}\left(\begin{pmatrix} a_1 & b_1 & e_1 \\ a_2 & b_2 & e_2\end{pmatrix}, N; c\right) =$$

$$\chi(e^*)\chi\chi_g(c^\flat)e\sum_{f|c^\flat} f\mu\left(\frac{c^\flat}{f}\right)\sum_{f'|f^*}\mu(f')\overline{\chi_g(f')\lambda_g(f^*/f')}\sum_{\pm}\varepsilon_g^{\pm}\Sigma^{\pm}(a_1e_1e^*f'f^*, e), \quad (5.10)$$

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

$$\Sigma^{\pm}(a_1e_1e^*f'f^*, e) := \sum_h G_{\chi\chi_g}(h; c^\sharp)S_h^{\pm}(a_1e_1e^*f'f^*, e) \quad (5.11)$$

with

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

Since $\chi\chi_g$ is not the trivial character, $G_{\chi\chi_g}(0; c)S_0 = 0$, and we are left to evaluate (5.11) over the frequencies $h \neq 0$. This will be done in Theorem 5.1.

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 5.1. *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 (2.17). 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

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

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

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

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 (6.12) we integrate by parts $2k$ times (where we may assume that $k=0$ for $y \leqslant Y_0$). We obtain, using also Propositions 6.1 and 6.2, (2.17), (5.2) and that $u \sim N/b$,

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

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, $\nu \in \mathbb{N}$, $F_Y(x,y)$ is supported on $[\mathbf{X}/4, 4\mathbf{X}] \times [Y/4, 4Y]$ and satisfies

$$x^iy^i\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j}F_Y(x,y) \ll_{i,j,k,\varepsilon} Z^{i+j}\left(1+\frac{Y}{\mathbf{Y}}\right)^{-k}\left(\frac{\mathbf{Y}}{Y}\right)^{\theta_g+\varepsilon} \tag{5.12}$$

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 5.1 (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 (5.17) 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 $\theta_g\leqslant\theta$)

$$\begin{aligned}\Sigma^{\pm}_{Y\leqslant Y_{\min}}(\mathbf{l_1},\mathbf{l_2})&\ll_{P,g,\varepsilon} q^{10\varepsilon}N\frac{LN}{c}c^{1/2}\left(\frac{L^2N}{e^2}\right)^{1/2}\left(\frac{dNY_{\min}}{df'f^*}\right)^{1/2}\left(\frac{\mathbf{Y}}{Y_{\min}}\right)^{\theta}\\&\ll_{P,g,\varepsilon} q^{10\varepsilon}N\frac{L^{2+2\theta}N^{2+\theta}}{(f^*)^{1/2}e^{1+\theta}c^{1/2}}Y_{\min}^{1/2-\theta}.\end{aligned}\tag{5.13}$$

For $Y>Y_{\min}$, we apply Theorem 5.1 and for this we set (cf. the next section), with $B:=3$,

$$\eta_{1Z}:=\frac{13+2B+4\theta}{2},\qquad \eta_{1L}:=\frac{1}{2},\qquad \eta_{1X}:=0,\qquad \eta_{1Y}:=1,\qquad \eta_{1Y/X}:=\frac{2+B}{2},$$

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

and

$$D_Z:=-\frac{9+2B+4\theta}{4(1+\theta)},\quad D_L:=0,\quad D_X:=\frac{1}{4(1+\theta)},\quad D_Y:=0,\quad D_{Y/X}:=-\frac{2+B}{4(1+\theta)},$$

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

$$\eta_{2Z}:=\eta_{1Z}+(1+2\theta)D_Z=\frac{17+2B+4\theta}{4(1+\theta)},\quad \eta_{2L}:=\eta_{1L}+(1+2\theta)D_L=\frac{1}{2},$$

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

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

It follows from (5.4) that $\mathbf{l_1}\mathbf{l_2}$ is coprime with $\mathbf{q}=q^*\mid q$ and also with D, therefore if the cusp form g satisfies (5.6) then Theorem 5.1 yields, by (5.12),

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

i.e.,

$$\Sigma^{\pm}_{Y>Y_{\min}}(\mathbf{l_1},\mathbf{l_2}) \ll_{P,g,\varepsilon} q^{10\varepsilon} N\Bigg(L^{\frac{25+4B+8\theta}{2}} N^{\frac{19+3B+6\theta}{2}} \frac{(c^\flat)^{\frac{1-2\theta}{2}}}{e^{\frac{4+B+2\theta}{2}} c^{7+B+\theta}} (q^*)^{\eta_{1q}} Y_{\min}^{-\frac{2+B+2\theta}{2}} + L^{\frac{37+4B+36\theta+8\theta^2}{4(1+\theta)}} N^{\frac{27+3B+24\theta+4\theta^2}{4(1+\theta)}} \frac{(c^\flat)^{\frac{1}{2(1+\theta)}}}{e^{\frac{6+B+8\theta+4\theta^2}{4(1+\theta)}} c^{\frac{19+2B+12\theta}{4(1+\theta)}}} (q^*)^{\eta_{2q}} Y_{\min}^{-\frac{1+B+2\theta+4\theta^2}{4(1+\theta)}}\Bigg).$$

A comparison of the second portion of this bound with (5.13) suggests the choice

$$Y_{\min} := L^{\frac{29+4B+20\theta}{3+B}} N^{\frac{19+3B+12\theta}{3+B}} (c^\flat)^{\frac{2}{3+B}} e^{-\frac{2+B}{3+B}} (f^*)^{\frac{2(1+\theta)}{3+B}} c^{-\frac{17+2B+10\theta}{3+B}} (q^*)^{\frac{1-2\theta-2\delta_{\mathrm{tw}}}{3+B}}.$$

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

$$\Sigma^{\pm}(\mathbf{l_1},\mathbf{l_2}) \ll_{P,g,\varepsilon} q^{10\varepsilon} N (q^*)^{\delta_{q^*}} \times\Bigg(L^{\frac{17-(74+20B)\theta-40\theta^2}{6+2B}} N^{\frac{19+3B-(44+12B)\theta-24\theta^2}{6+2B}} \frac{(c^\flat)^{\frac{-1-B-(10+2B)\theta}{6+2B}}}{(f^*)^{\frac{2+B+(4+B)\theta+2\theta^2}{3+B}} e^{\frac{8+3B+2\theta}{6+2B}} c^{\frac{8-B-(48+12B)\theta-20\theta^2}{6+2B}}} + L^{\frac{41+8B-(26+4B)\theta-40\theta^2}{6+2B}} N^{\frac{31+7B-(20+4B)\theta-24\theta^2}{6+2B}} \frac{(c^\flat)^{\frac{1-2\theta}{3+B}}}{(f^*)^{\frac{1+B+2\theta+4\theta^2}{6+2B}} e^{\frac{8+3B+2\theta}{6+2B}} c^{\frac{20+3B-(24+4B)\theta-20\theta^2}{6+2B}}}\Bigg),$$

where

$$\delta_{q^*} := (1-2\theta)\frac{1-2\theta-2\delta_{\mathrm{tw}}}{6+2B}.$$

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

$$\sum_{c\equiv 0\,([q,D])} \sum_{f|c^\flat} \frac{ef(f^*)^\theta}{c^2} \frac{(c^\flat)^{\frac{-1-B-(10+2B)\theta}{6+2B}}}{(f^*)^{\frac{2+B+(4+B)\theta+2\theta^2}{3+B}} e^{\frac{8+3B+2\theta}{6+2B}} c^{\frac{8-B-(48+12B)\theta-20\theta^2}{6+2B}}} \ll_\varepsilon q^{\varepsilon-\frac{20+3B-(48+12B)\theta-20\theta^2}{6+2B}}$$

and

$$\sum_{c\equiv 0\,([q,D])} \sum_{f|c^\flat} \frac{ef(f^*)^\theta}{c^2} \frac{(c^\flat)^{\frac{1-2\theta}{3+B}}}{(f^*)^{\frac{1+B+2\theta+4\theta^2}{6+2B}} e^{\frac{8+3B+2\theta}{6+2B}} c^{\frac{20+3B-(24+4B)\theta-20\theta^2}{6+2B}}} \ll_\varepsilon q^{\varepsilon-\frac{32+7B-(24+4B)\theta-20\theta^2}{6+2B}}.$$

Collecting all the terms (see (5.7), (5.10), (5.11)) and using also $q^* \leqslant q$, we deduce that for g satisfying (5.6) 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

$$\delta_L := 2\theta + \frac{41+8B-(26+4B)\theta-40\theta^2}{6+2B} = \frac{41+8B-14\theta-40\theta^2}{6+2B},$$

$$\begin{aligned}\delta_q :=&\ -\delta_{q^*} - \frac{19+3B-(44+12B)\theta-24\theta^2}{6+2B} + \frac{20+3B-(48+12B)\theta-20\theta^2}{6+2B}\\ =&\ -\delta_{q^*} - \frac{31+7B-(20+4B)\theta-24\theta^2}{6+2B} + \frac{32+7B-(24+4B)\theta-20\theta^2}{6+2B}\\ =&\ \frac{1-2\theta}{3+B}\,\delta_{\mathrm{tw}}.\end{aligned}$$

For g not satisfying (5.6), an additional term occurs whose contribution to $Q^{\mathrm{ND}}(\vec{x},N)$ is bounded by (cf. Theorem 5.1)

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

with

$$\delta_{3L} := 9 - \frac{17\theta+20\theta^2}{3+B}, \qquad \delta_{3q} := \frac{1}{2} - \frac{2\theta}{3+B}\delta_{\mathrm{tw}},$$
$$\delta_{4L} := 13 - \frac{17\theta+20\theta^2}{3+B}, \qquad \delta_{4q} := \frac{1+4\theta-2\delta_{\mathrm{tw}}}{4(1+\theta)}.$$

The above estimates together with (5.9) conclude the proof of Proposition 5.2.

5.4 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 (2.2)) $\rho_g(1)$ equals one and consequently, by (2.10), 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

$$x^iy^j\frac{\partial^i}{\partial x^i}\frac{\partial^j}{\partial y^j}F(x,y)\ll Z^{i+j} \tag{5.14}$$

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

We consider the sum

$$\Sigma^{\pm}_{\chi}(\ell_1,\ell_2;\mathbf{c}):=\sum_{h\neq 0}G_\chi(h;\mathbf{c})S^{\pm}_h(\ell_1,\ell_2),$$

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

$$S^{\pm}_h(\ell_1,\ell_2):=\sum_{\ell_1m\mp\ell_2n=h}\overline{\lambda_g}(m)\lambda_g(n)F(\ell_1m,\ell_2n). \tag{5.15}$$

Our goal is

Theorem 5.1. *Assume Hypothesis H_θ for any $0\leqslant\theta\leqslant\frac12$, and let B and $\delta_{\rm tw}$ be as in* (5.35). *Set*

$$P:=\mathbf{c}D\mu_g\ell_1\ell_2Z(X+Y),$$

and assume (as one may by symmetry) that $Y\geqslant X$. Set also

$$D_{\rm opt}:=Z^{-\frac{9+2B+4\theta}{4(1+\theta)}}\mathbf{c}^{-\frac{\theta}{2(1+\theta)}}q^{-\frac{1-2\theta-2\delta_{\rm tw}}{4(1+\theta)}}(X/Y)^{\frac{2+B}{4(1+\theta)}}X^{\frac{1}{4(1+\theta)}}.$$

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:

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

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 (2.4)*) of g.*

Remark 5.3. 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

$$\Sigma^{\pm}_\chi(\ell_1, \ell_2; \mathbf{c}) = \chi(\mathbf{c}')q'G_\chi(1; q)\sum_{d|\mathbf{c}'} d\mu(\mathbf{c}'/d)\overline{\chi}(d)\sum_{h\neq 0}\overline{\chi}(h)S^{\pm}_{hq'd}(\ell_1, \ell_2). \tag{5.16}$$

Observe that by (5.14), (5.15) and (2.44) 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 2.4: integrating by parts and applying Cauchy–Schwarz, we obtain

$$\begin{aligned}
\Sigma^{\pm}_\chi(\ell_1, \ell_2; \mathbf{c}) &\leqslant q'q^{1/2}\sum_{d|\mathbf{c}'} d \iint_{(\mathbb{R}^+)^2} \ell_1\ell_2\left|\frac{\partial^2}{\partial x\partial y}F(\ell_1 x, \ell_2 y)\right| \sum_{h\neq 0}\left|\sum_{\substack{m\leqslant x,\ n\leqslant y\\ \ell_1 m\mp\ell_2 n = dq'h}} \overline{\lambda_g}(m)\lambda_g(n)\right| dxdy \\
&\leqslant Z^2 q'q^{1/2}\sum_{d|\mathbf{c}'} d \max_{\substack{x\ll X/\ell_1\\ y\ll Y/\ell_2}} \sum_{|dq'h|\leqslant \ell_1 x+\ell_2 y}\left|\sum_{\substack{m\leqslant x,\ n\leqslant y\\ \ell_1 m\mp\ell_2 n = h}} \overline{\lambda_g}(m)\lambda_g(n)\right| \\
&\ll_\varepsilon P^\varepsilon D\mu_g^2 Z^2 q'q^{1/2}\sum_{d|\mathbf{c}'} d\left(\frac{X+Y}{dq'}\right)^{1/2}\left(\frac{XY}{\ell_1\ell_2}\right)^{1/2} \\
&\ll_\varepsilon P^{2\varepsilon} D\mu_g^2 Z^2 \mathbf{c}^{1/2}(X+Y)^{1/2}\left(\frac{XY}{\ell_1\ell_2}\right)^{1/2}.
\end{aligned} \tag{5.17}$$

On the other hand, an application of the δ-symbol method of [DFI94b] yields (cf. [Mi04, Section 7.1], [Ha03b, 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.

5.5 Jutila's variant of the circle method

In the sequel, 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 5.1 follows from (5.17). We denote by $D(g,\ell_1,\ell_2,q'd)$ the h-sum in (5.16); 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:

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

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 [Ha03a], we apply Jutila's method of overlapping intervals [Ju92, Ju96] 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

$$L \gg_\varepsilon C^{2-\varepsilon}/N \gg_{g,\varepsilon} C^{2-\varepsilon}/(\ell_1\ell_2) \tag{5.18}$$

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 [Ju92] one has

$$\int_{[-1,2]}|I(\alpha)-\tilde{I}(\alpha)|^2 d\alpha \ll_\varepsilon \frac{C^{2+\varepsilon}}{\delta L^2} \ll_{g,\varepsilon} C^{2\varepsilon}(\ell_1\ell_2)^2\frac{Y}{C^2}. \tag{5.19}$$

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 (5.19) that

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

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. (2.47))

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

where by (5.14),

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

and by Proposition 2.5,

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

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

$$\ll_{g,\varepsilon}P^{2\varepsilon}Z^2(\ell_1\ell_2)^{1/2}\mathbf{c}^{1/2}\frac{X^{1/2}Y^{3/2}}{C}. \tag{5.20}$$

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 2.3 to the variables m,n (cf. (2.10), (2.13), (2.14)) 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 $\varepsilon_g^-=\pm1$ is the sign in (2.14) (for g not induced from a holomorphic form),

$$\tilde{D}^{\pm,\pm}:=\frac{1}{L}\sum_{m,n}\overline{\lambda_g}(m)\lambda_g(n)\sum_{c\equiv 0\,(N)}\sum_h\chi(h)\frac{S(dh,\mp\ell_1 m\pm\ell_2 n;c)}{c}\mathcal{E}^{\pm,\pm}(m,n,h;c) \tag{5.21}$$

and

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

Notice that the definition of E and the various assumptions made so far imply that

$$E(x,y,z)=0 \ \text{ unless } \ x\sim X/\ell_1,\ y\sim Y/\ell_2,\ |dz|\leqslant 4Y. \tag{5.22}$$

Moreover,

$$E^{(i,j,k)}(x,y,z)\ll_{i,j,k}\frac{Z^{i+j}\ell_1^i\ell_2^jd^k}{X^iY^{j+k}}, \tag{5.23}$$

so that for any fixed h

$$\|E^{(i,j,k)}(*,*,h)\|_1\ll_{i,j,k}\frac{Z^{i+j}\ell_1^{i-1}\ell_2^{j-1}d^kXY}{X^iY^{j+k}}, \tag{5.24}$$

and therefore

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

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

$$\frac{\ell_1\ell_2w(c)}{c}\iint_{(\mathbb{R}^+)^2}E(x,y,h)J_{1,\nu_1}\left(\frac{4\pi\ell_1\sqrt{mx}}{c}\right)J_{2,\nu_2}\left(\frac{4\pi\ell_2\sqrt{ny}}{c}\right)dxdy, \tag{5.25}$$

where

$$\big\{J_{1,\nu}(x),J_{2,\nu}(x)\big\}\subset\Big\{\frac{Y_\nu(x)}{\cosh(\pi t)},\ \cosh(\pi t)K_\nu(x)\Big\}$$

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

$$\big\{J_{1,\nu}(x),J_{2,\nu}(x)\big\}\subset\Big\{\frac{Y_\nu(x)}{\sinh(\pi t)},\ \sinh(\pi t)K_\nu(x)\Big\}$$

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 (5.25) 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 (6.12), 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

$$\begin{aligned}\frac{\ell_1\ell_2w(c)}{c}\left(\frac{\ell_1\sqrt{m}}{c}\right)^{-2i}\left(\frac{\ell_2\sqrt{n}}{c}\right)^{-2j}\iint_{(\mathbb{R}^+)^2}&\frac{\partial^{i+j}}{\partial x^i\partial y^j}\big\{E(x,y,h)W_1^{-\nu_1}W_2^{-\nu_2}\big\}\\ &\times W_1^{\nu_1+i}W_2^{\nu_2+j}J_{1,\nu_1+i}(W_1)J_{2,\nu_2+j}(W_2)\,dxdy,\end{aligned}$$

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 (5.22). Using (5.24) and Proposition 6.2 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

$$\Xi(m,n) := \frac{XY}{C}\left\{\left(1+\frac{C^2}{\ell_1 mX}\right)\left(1+\frac{C^2}{\ell_2 nY}\right)\right\}^{|\Im t_g|}\left\{\left(1+\frac{\ell_1 mX}{C^2}\right)\left(1+\frac{\ell_2 nY}{C^2}\right)\right\}^{-1/4}. \tag{5.26}$$

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

$$d|h| \leqslant 4Y, \qquad c \sim C, \qquad m \ll_\varepsilon P^\varepsilon \frac{\mu_g^4 Z^2 C^2}{\ell_1 X}, \qquad n \ll_\varepsilon P^\varepsilon \frac{\mu_g^4 Z^2 C^2}{\ell_2 Y}, \tag{5.27}$$

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

$$\mathcal{E}^{\pm,\pm}(m,n,h;c) \ll_\varepsilon P^\varepsilon \Xi(m,n). \tag{5.28}$$

The partial derivatives

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

can be estimated similarly. We shall restrict our attention to the range (5.27); the argument also yields that outside this range the partial derivatives are very small. By (6.13), the above partial derivative is a linear combination of integrals of the form

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

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

$$\begin{aligned} h^j c^k (\mathcal{E}^{\pm,\pm})^{(0,0,j,k)}(m,n,h;c) &\ll_{j,k,\varepsilon} P^\varepsilon \left(\frac{d|h|}{Y}\right)^j \mu_g^k \left(1+\frac{\sqrt{\ell_1 mX}}{C}+\frac{\sqrt{\ell_2 nY}}{C}\right)^k \Xi(m,n) \\ &\ll_{j,k,\varepsilon} P^\varepsilon (P^\varepsilon \mu_g^3 Z)^k \Xi(m,n). \end{aligned} \tag{5.29}$$

5.6 Expanding the c-sum

Our next step will be to expand spectrally the c-sum in (5.21) 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 2.1. 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 $h(\ell_1 m - \ell_2 n)$ when it is nonzero. So our first step will be to isolate the contribution of the m,n such that $\ell_1 m - \ell_2 n = 0$ (the contribution of the $h = 0$ is void since we assume that χ is nontrivial). Thus we have the splitting

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

where

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

with

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

the Ramanujan sum, and

$$\begin{aligned} \tilde{D}^\pm :=& \frac{1}{L}\sum_{\ell_1 m-\ell_2 n\neq 0}\overline{\lambda_g}(m)\lambda_g(n)\sum_{c\equiv 0\,(N)}\sum_{\substack{h\\ \pm h(\ell_1 m-\ell_2 n)>0}}\chi(h)\frac{S(dh,\ell_1 m-\ell_2 n;c)}{c}\mathcal{E}(m,n,h;c) \\ =& \frac{1}{L}\sum_{\ell_1 m-\ell_2 n\neq 0}\overline{\lambda_g}(m)\lambda_g(n)\tilde{D}^\pm(m,n) \end{aligned}$$

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 $h' := \ell_1 m - \ell_2 n \neq 0$.

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 (5.29) 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 (2.12), (2.44), (5.18) and (5.26) 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 (5.16) that the contribution of the $\tilde{D}^0$ terms is bounded by (remember that we have reused the letter d in place of $q'd$)

$$\ll_{g,\varepsilon} P^{5\varepsilon} Z \frac{(\mathbf{c},\ell_1\ell_2)^{1/2}}{[\ell_1,\ell_2]^{1-\theta}} \mathbf{c}^{1/2} q^{3/16} \frac{X^{3/4}Y^{3/4}}{C^{1/2}}. \tag{5.30}$$

Remark 5.4. 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.

We perform a dyadic subdivision on the h variable. By (5.28) and (5.29), 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 $H = 2^\nu$, $\nu \in \mathbb{N}$, 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

$$h^j c^k \mathcal{E}_H^{(0,0,j,k)}(m,n,h;c) \ll_{j,k,\varepsilon} P^\varepsilon (P^\varepsilon \mu_g^3 Z)^k \Xi(m,n). \tag{5.31}$$

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

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

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

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,

$$\tilde{D}_H^+(m,n) = \frac{1}{2\pi i}\int_{(2)} T_H^{\mathrm{Holo}}(m,n;s)\,ds + \frac{1}{2\pi i}\int_{(2)} T_H^{\mathrm{Maass}}(m,n;s)\,ds + \frac{1}{2\pi i}\int_{(2)} T_H^{\mathrm{Eisen}}(m,n;s)\,ds, \tag{5.32}$$

where

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

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

$$T_H^{\mathrm{Eisen}}(m,n;s) := \frac{1}{\pi}\sum_{\mathfrak{a}}\int_{-\infty}^{+\infty} \frac{\hat{\varphi}_H(m,n;s;*)(t)}{\cosh(\pi t)}\sqrt{h'}\rho_{\mathfrak{a}}(h',t)L\left(E_{\mathfrak{a}}\left(\tfrac{1}{2}+it\right)\otimes\chi,s;d\right)\,dt,$$

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 (5.32) to the left up to $\Re s = \frac{1}{2}$ 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 $\Re s = \frac{1}{2}$ and the various Bessel transforms $\dot{\varphi}_H(m,n;s;*)(k)$, $\hat{\varphi}_H(m,n;s;*)(t)$ and $\check{\varphi}_H(m,n;s;*)(t)$. This will be done in the next two sections.

5.7 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., $f\in\mathcal{B}_k^h(N,1)$) or a Hecke–Maass cusp form (i.e., $f=u_j\in\mathcal{B}_0(N,1)$) or an Eisenstein series $f(z)=E_{\mathfrak{a}}\left(z,\frac{1}{2}+it\right)$.

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 (2.45) and Hypothesis $H_{\frac{7}{64}}$ that $L(f\otimes\chi,s;d)$ is holomorphic for $\Re s\geqslant\frac{1}{2}$, and for $\Re s=\frac{1}{2}$ one has

$$L(f\otimes\chi,s;d) \ll_\varepsilon (PN)^\varepsilon\left(\sum_{h|(dN)^\infty}\frac{|\sqrt{dh}\rho_f(dh)|}{h^{1/2}}\right)\left|L(\tilde{f}.\chi,s)\right|. \tag{5.33}$$

By (1.4), for $L(\tilde{f}.\chi,s)$ one has the subconvexity bound

$$L(\tilde{f}.\chi,s) \ll (|s|\mu_f Nq)^\varepsilon|s|^{\frac{1}{2}}\mu_f^B N^{\frac{1}{2}}q^{\frac{1}{2}-\delta_{\mathrm{tw}}}, \tag{5.34}$$

with the parameters

$$B := 3, \qquad \delta_{\mathrm{tw}} := \frac{1}{8}. \tag{5.35}$$

When $f(z)$ is of the form $E_{\mathfrak{a}}\left(z,\frac{1}{2}+it\right)$, the computations of [Mi04] 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]),

$$\sqrt{|dh|}\rho_{\mathfrak{a}}(dh,t) = \frac{\pi^{\frac12+it}|dh|^{it}}{\Gamma\left(\frac12+it\right)}\left(\frac{(w,N/w)}{wN}\right)^{\frac12+it}\sum_{(\gamma,N/w)=1}\frac{1}{\gamma^{1+2it}}\sum_{\substack{\delta(\gamma w),\ (\delta,\gamma w)=1\\ \delta\gamma\equiv u \bmod (w,N/w)}} e\left(-dh\frac{\delta}{\gamma w}\right)$$

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

$$\begin{aligned}\sqrt{|dh|}\rho_{\mathfrak{a}}(dh,t) &\ll_\varepsilon (P(1+|t|))^\varepsilon\cosh^{1/2}(\pi t)\frac{(dh,w)(w,N/w)}{(wN)^{1/2}}\\ &\ll_\varepsilon (P(1+|t|))^\varepsilon\cosh^{1/2}(\pi t)(dh,N)^{1/2},\end{aligned}\tag{5.36}$$

and also the identity

$$\begin{aligned}L(f\otimes\chi,s;d) &= \frac{\pi^{\frac12+it}d^{it}}{\Gamma\left(\frac12+it\right)}\left(\frac{(w,N/w)}{wN}\right)^{\frac12+it}\\ &\times\frac{1}{\varphi((w,N/w))}\sum_{\psi \bmod (w,N/w)}\frac{w'G_\psi(1;w^*)\overline{\psi}(-ud/(d,w'))\psi(w'')}{(w'/(d,w'))^{s-it}L^{(N)}(\psi^2,1+2it)}\chi\left(\frac{w'}{(d,w')}\right)\\ &\times\sum_{h\geqslant1}\frac{\chi(h)\overline{\psi}(h)}{h^{s-it}}\left(\sum_{\substack{\gamma|N^\infty\\(\gamma,N/w)=1}}\frac{\psi^2(\gamma)}{\gamma^{1+2it}}r\left(\frac{dhw'}{(d,w')};\gamma w''\right)\right)\left(\sum_{\substack{a|\frac{dh}{(d,w')}\\(a,N)=1}}\frac{\psi^2(a)}{a^{2it}}\right).\end{aligned}$$

Now the h-sum factors as

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

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

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

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

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

Remark 5.5. In the special case where $q|(w,N/w)$, the residues of $L\left(E_{\mathfrak{a}}\big(z,\frac12+it\big)\otimes\chi,s;d\right)$ at $s=1\pm it$ $(t\neq0)$ are bounded by

$$\begin{aligned}\operatorname{res}_{s=1\pm it}L\left(E_{\mathfrak{a}}\big(z,\tfrac12+it\big)\otimes\chi,s;d\right)&\ll_\varepsilon((1+|t|)Nq)^\varepsilon\cosh^{1/2}(\pi t)\frac{(d,w)(w,N/w)}{(wN)^{1/2}}\\&\ll_\varepsilon((1+|t|)Nq)^\varepsilon\cosh^{1/2}(\pi t)(d,N)^{1/2},\end{aligned}\tag{5.37}$$

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

5.8 Putting it all together

We will need to bound the Bessel transforms $\dot{\varphi}_H(m,n;s;*)(k)$, $\hat{\varphi}_H(m,n;s;*)(t)$ and $\check{\varphi}_H(m,n;s;*)(t)$. For this purpose, we first record an estimate for φ_H and its partial derivatives. Using (5.31) and several integrations by parts, we see that for any $j,k\geqslant0$ and $\Re s\geqslant-\frac12$,

$$r^k\frac{\partial^k}{\partial r^k}\varphi_H(m,n;s;r)\ll_{j,k,\varepsilon}P^\varepsilon(P^\varepsilon\mu_g^3Z)^{j+k}|s|^{-j}\Xi(m,n)H^{\Re s},\tag{5.38}$$

where $\Xi(m,n)$ is defined in (5.26); 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 Lemma 2.1.

We are now ready to combine the results of the preceding sections to conclude the proof of Theorem 5.1. We start by estimating the contribution of the Maass spectrum to $\tilde{D}_H^{+,+}$:

$$\frac1L\sum_{\substack{m,n\geqslant1\\h'>0}}\overline{\lambda_g}(m)\lambda_g(n)\frac{1}{2\pi i}\int_{(2)}T_H^{\mathrm{Maass}}(m,n;s)\,ds=\frac{1}{2\pi i}\int_{(1/2)}\frac1L\sum_{\substack{m,n\geqslant1\\h'>0}}\overline{\lambda_g}(m)\lambda_g(n)T_H^{\mathrm{Maass}}(m,n;s)\,ds.$$

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 $u_j \in \mathcal{B}_0(N,1)$ 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 $W := P^\varepsilon \mu_g^3 Z$, we can apply (2.23) and (2.24) to $\varphi = \varphi_H(m,n;s;*)$ in the light of (5.38). Using also (5.33) and (5.34), we obtain, for any $j \geqslant 0$,

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

By several applications of the Cauchy–Schwarz inequality and the bound (2.45), we can see that

$$\sum_{|t_i|\leqslant T_0} \frac{|\sqrt{h'}\rho_i(h')|}{\cosh(\pi t_i)} \left(\sum_{h|(dN)^\infty} \frac{|\sqrt{dh}\rho_i(dh)|}{h^{1/2}} \right) \ll_\varepsilon (mnPT_0)^\varepsilon (h'd)^\theta T_0^2. \tag{5.39}$$

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 (2.44) and (5.26), we find that

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

For $T_{H,>T_0}^{\mathrm{Maass}}(m,n;s)$, we use (2.26), (5.33) and (5.34): we obtain, for any $j \geqslant 0$ and any $k > 1$,

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

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 (2.44) and (5.26). As before, we may restrict ourselves to some range

$$\ell_1 m \ll_{g,\varepsilon} P^\varepsilon Z^2(C^2/X) \qquad \text{and} \qquad \ell_2 n \ll_{g,\varepsilon} P^\varepsilon Z^2(C^2/Y),$$

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 (5.39),

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

Summing up and using also (5.18), we infer that

$$\frac{1}{L}\sum_{\substack{m,n\geqslant 1\\ h'>0}}\overline{\lambda_g}(m)\lambda_g(n)T_H^{\mathrm{Maass}}(m,n;s)\ll_{j,k,g,\varepsilon}(PT_0)^{6\varepsilon}Z^{3+2\theta}\frac{W^j}{|s|^{j-1/2-\varepsilon}}(\ell_1\ell_2)^{1/2}d^\theta$$
$$\times C\left(\frac{C^2}{Y}\right)^\theta\left(\frac{Y}{d}\right)^{1/2}q^{1/2-\delta_{\mathrm{tw}}}T_0^{B+2}\left\{1+\left(\frac{W}{T_0}\right)^k\left(\frac{(Y/X)^\theta}{T_0^{1/2}}+\frac{Z(Y/X)^{1/2+\theta}}{T_0}\right)\right\}.$$

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

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

We use this bound with $j>3/2+\varepsilon$ (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

$$\ll_{g,\varepsilon}P^{14\varepsilon}Z^{3+2\theta}W^{7/2+B}(\ell_1\ell_2)^{1/2}d^{\theta-1/2}q^{1/2-\delta_{\mathrm{tw}}}(Y/X)^{(B+2)/2}Y^{1/2-\theta}C^{1+2\theta},$$

hence by (5.16) 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 $q'd$)

$$\ll_{g,\varepsilon}P^{24\varepsilon}Z^{13/2+B+2\theta}(\ell_1\ell_2)^{1/2}\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\mathrm{tw}}}(Y/X)^{(B+2)/2}Y^{1/2-\theta}C^{1+2\theta}. \tag{5.40}$$

Similar arguments (using also (2.45) and (2.26) for $\dot{\varphi}$) 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 $L\left(E_{\mathfrak{a}}\left(z,\frac{1}{2}+it\right)\otimes\chi,s\right)$ 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 (5.36), (5.37), (2.24), and (5.38) 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

$$\ll_{g,\varepsilon}P^{3\varepsilon}WZ^3(\mathbf{c},\ell_1\ell_2)^{1/2}q^{1/2}YC\ll_g P^{4\varepsilon}Z^4(\mathbf{c},\ell_1\ell_2)^{1/2}(\ell_1\ell_2)^{1/2}YC. \tag{5.41}$$

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 (5.20), (5.30), (5.40), plus the additional term (5.41) if $q|(w,N/w)$ for some $w|N$. To conclude, we discuss now the choice of the parameter C.

A comparison of (5.40) with (5.20) suggests the choice

$$C_{\mathrm{opt}}:=Z^{-\frac{9+2B+4\theta}{4(1+\theta)}}\mathbf{c}^{-\frac{\theta}{2(1+\theta)}}q^{-\frac{1-2\theta-2\delta_{\mathrm{tw}}}{4(1+\theta)}}(X/Y)^{\frac{B+2}{4(1+\theta)}}X^{\frac{1}{4(1+\theta)}}Y^{1/2}=:D_{\mathrm{opt}}Y^{1/2},$$

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

$$X\geqslant X_{\mathrm{opt}}:=Z^{\frac{9+2B+4\theta}{B+3}}\mathbf{c}^{\frac{2\theta}{B+3}}q^{\frac{1-2\theta-2\delta_{\mathrm{tw}}}{B+3}}Y^{\frac{B+2}{B+3}}. \tag{5.42}$$

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

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

so that (5.30) is bounded by (5.20) (when $P^{2\varepsilon}$ is replaced by $P^{5\varepsilon}$). Therefore, we obtain Theorem 5.1 when (5.42) is satisfied (cf. (5.40)):

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

plus the additional term (5.41), if $q|(w,N/w)$ for some $w|N$, which equals

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

If (5.42) is not satisfied (i.e., $X < X_{\mathrm{opt}}$, hence $D_{\mathrm{opt}} < 1$), we choose $C = Y^{1/2} = Y^{1/2}\max(1, D_{\mathrm{opt}})$. We see that (5.20) is bounded by (5.40) whose value is given by

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

The diagonal contribution (5.30) 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 $X < X_{\mathrm{opt}}$ into

$$X(X/Y)^{B+2} < Z^{9+2B+4\theta}\mathbf{c}^{2\theta}q^{1-2\theta-2\delta_{\mathrm{tw}}},$$

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

$$q^{3/16}X^{1/4}(X/Y)^{(B+2)/4} < Z^{9/2+B+2\theta}\mathbf{c}^{\theta}q^{1/2-\theta-\delta_{\mathrm{tw}}}.$$

It follows that (5.30) is bounded by

$$\ll_{g,\varepsilon} P^{5\varepsilon}Z^{11/2+B+2\theta}(\ell_1\ell_2)^{\theta}\mathbf{c}^{1/2+\theta}q^{1/2-\theta-\delta_{\mathrm{tw}}}(Y/X)^{B/4}Y.$$

In particular, if (5.42) is not satisfied, then (5.20), (5.30) and (5.40) are all bounded by

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

Finally, if $q|(w,N/w)$ for some $w|N$, the additional term (5.41) 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 5.1.

Chapter 6

Appendix

6.1 Heegner points, closed geodesics, and ideal classes

In this section we discuss how the narrow ideal classes in an imaginary (resp. real) quadratic number field give rise to Heegner points (resp. closed geodesics) on the modular surface $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.

Let us start the discussion with the equivalence of integral binary quadratic forms. The concept was introduced by Lagrange [La73] and studied by Gauss [Ga86] in a systematic fashion.

An *integral binary quadratic form* is a homogeneous polynomial

$$\langle a, b, c\rangle := ax^2 + bxy + cy^2 \in \mathbb{Z}[x, y]$$

with associated *discriminant*

$$d := b^2 - 4ac \in \mathbb{Z}.$$

The possible discriminants are the integers congruent to 0 or 1 mod 4. We shall assume that the form $\langle a, b, c\rangle$ is not a product of linear factors in $\mathbb{Z}[x, y]$, then d is not a square, hence $ac \neq 0$. If $d < 0$ then $ac > 0$ and we shall assume that we are in the *positive definite* case $a, c > 0$. Furthermore, we shall assume that d is a *fundamental discriminant* which means that it cannot be written as $d'e^2$ for some smaller discriminant d'. Then $\langle a, b, c\rangle$ is a *primitive* form which means that a, b, c are relatively prime. The possible fundamental discriminants are the square-free numbers congruent to 1 mod 4 and 4 times the square-free numbers congruent to 2 or 3 mod 4.

Example 1. The first few negative fundamental discriminants are: -3, -4, -7, -8, -11, -15, -19, -20, -23, -24. The first few positive fundamental discriminants are: 5, 8, 12, 13, 17, 21, 24, 28, 29, 33.

Lagrange [La73] discovered that every form $\langle a, b, c\rangle$ with a given discriminant d can be reduced by some integral unimodular substitution

$$(x, y) \mapsto (\alpha x + \beta y, \gamma x + \delta y), \qquad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),$$

to some form with the same discriminant that lies in a finite set depending only on d. Forms that are connected by such a substitution are called *equivalent*. It is easiest to understand this reduction by looking at the simple substitutions

$$(x, y) \overset{T}{\mapsto} (x - y, y) \qquad \text{and} \qquad (x, y) \overset{S}{\mapsto} (-y, x). \tag{6.1}$$

The induced actions on forms are given by

$$\langle a, b, c\rangle \overset{T}{\mapsto} \langle a, b - 2a, c + a - b\rangle \qquad \text{and} \qquad \langle a, b, c\rangle \overset{S}{\mapsto} \langle c, -b, a\rangle.$$

Now a given form $\langle a, b, c\rangle$ can always be taken to some $\langle a, b', c'\rangle$ with $|b'| \leqslant |a|$ by applying T or T^{-1} a few times. If $|a| \leqslant |c'|$ then we stop our reduction. Otherwise we apply S to get some $\langle a'', b'', c''\rangle$ with $|a''| < |a|$ and we start over with this form. In this algorithm we cannot apply S infinitely many times because $|a|$ decreases at each such step. Hence in a finite number of steps we arrive at an equivalent form $\langle a,b,c\rangle$ whose coefficients satisfy

$$|b| \leqslant |a| \leqslant |c|, \qquad b^2-4ac=d. \tag{6.2}$$

These constraints are satisfied by finitely many triples (a,b,c). Indeed, we have

$$|d| = |b^2-4ac| \geqslant 4|ac|-b^2 \geqslant 3b^2, \tag{6.3}$$

so there are only $\ll |d|^{1/2}$ choices for b and for each such choice there are only $\ll_\varepsilon d^\varepsilon$ choices for a and c since the product ac is determined by b. We have shown that the number of equivalence classes of integral binary quadratic forms of fundamental discriminant d, denoted $h(d)$, satisfies the inequality

$$h(d) \ll_\varepsilon |d|^{1/2+\varepsilon}. \tag{6.4}$$

In the case $d<0$ it is straightforward to compile a maximal list of inequivalent forms satisfying (6.2). There is an algorithm for $d>0$ as well but it is less straightforward. In fact the subsequent findings of this lecture can be turned into an algorithm for all d. Note that for $d>0$ (6.3) implies $4ac=b^2-d<0$, hence by an extra application of S we can always arrange for a reduced form $\langle a,b,c\rangle$ with $a>0$.

Example 2. The equivalence classes for $d=-23$ are represented by the forms $\langle 1,1,6\rangle$, $\langle 2,\pm1,3\rangle$. Hence $h(-23)=3$. The equivalence classes for $d=21$ are represented by the forms $\langle 1,1,-5\rangle$, $\langle -1,1,5\rangle$. Hence $h(21)=2$.

To obtain a geometric picture of equivalence classes of forms we shall think of $\mathbb{Q}(\sqrt{d})$ as embedded in $\mathbb{C}$ such that $\sqrt{d}/i>0$ for $d<0$ and $\sqrt{d}>0$ for $d>0$. For $q_1,q_2\in\mathbb{Q}$ we shall consider the conjugation

$$\overline{q_1+q_2\sqrt{d}} := q_1-q_2\sqrt{d}.$$

Each form $\langle a,b,c\rangle$ decomposes as

$$ax^2+bxy+cy^2=a(x-zy)(x-\bar{z}y),$$

where

$$z:=\frac{-b+\sqrt{d}}{2a}, \qquad \bar{z}:=\frac{-b-\sqrt{d}}{2a}.$$

Using (6.1) we can see that the action of $\mathrm{SL}_2(\mathbb{Z})$ on z and $\bar{z}$ is the usual one given by fractional linear transformations:

$$z\overset{T}{\mapsto} z+1 \qquad \text{and} \qquad z\overset{S}{\mapsto} -1/z.$$

Therefore in fact we are looking at the standard action of $\mathrm{SL}_2(\mathbb{Z})$ on certain conjugate pairs of points of $\mathbb{Q}(\sqrt{d})$ embedded in $\mathbb{C}$. For $d<0$ we consider the points $z\in\mathcal{H}$ and obtain $h(d)$ points on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$. These are the *Heegner points* of discriminant $d<0$. For $d>0$ we consider the geodesics $G_{\bar{z},z}\subset\mathcal{H}$ connecting the real points $\{\bar{z},z\}$ and obtain $h(d)$ geodesics on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$.

It is a remarkable fact that for $d>0$ any geodesic $G_{\bar{z},z}$ as above becomes closed when projected to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$, and its length is an important arithmetic quantity associated with the number field $\mathbb{Q}(\sqrt{d})$. To see this take any matrix $M\in\mathrm{GL}_2^+(\mathbb{R})$ which takes 0 to $\bar{z}$ and ∞ to z, for example[1]

$$M:=\begin{pmatrix} z & \bar{z}\\ 1 & 1\end{pmatrix},$$

then M takes the positive real axis (resp. geodesic) connecting $\{0,\infty\}$ to the real segment (resp. geodesic) connecting $\{\bar{z},z\}$. In particular, using that M is a conformal automorphism of the Riemann sphere, we see that $G_{\bar{z},z}$ is the semicircle above the real segment $[\bar{z},z]$, parametrized as

$$G_{\bar{z},z}=\{g(\lambda)i:\ \lambda>0\}, \qquad \text{where} \qquad g(\lambda):=M\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}.$$

[1] we assume here that $a>0$ which is legitimate as we have seen

Moreover, the unique isometry of $\mathcal{H}$ fixing the geodesic $G_{\bar{z},z}$ and taking $g(1)i$ to $g(\lambda)i$ is given by the matrix

$$M\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}M^{-1}\in \mathrm{SL}_2(\mathbb{R}).$$

Therefore we want to see that for some $\lambda>1$ the matrix

$$M\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}M^{-1}=\frac{1}{z-\bar{z}}\begin{pmatrix}z\lambda-\bar{z}\lambda^{-1} & z\bar{z}(\lambda^{-1}-\lambda)\\ \lambda-\lambda^{-1} & z\lambda^{-1}-\bar{z}\lambda\end{pmatrix} \tag{6.5}$$

is in $\mathrm{SL}_2(\mathbb{Z})$, and then the projection of $G_{\bar{z},z}$ to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ has length

$$\int_1^{\lambda^2}\frac{dy}{y}=2\ln(\lambda)$$

for the smallest such $\lambda>1$. A necessary condition for λ is that the sum and difference of diagonal elements of the matrix (6.5) are integers and so are the anti-diagonal elements as well. Using that

$$z-\bar{z}=\frac{\sqrt{d}}{a},\qquad z+\bar{z}=\frac{-b}{a},\qquad z\bar{z}=\frac{c}{a}$$

this is equivalent to:

$$\lambda+\lambda^{-1}\in\mathbb{Z},\qquad \{a,b,c\}\frac{\lambda-\lambda^{-1}}{\sqrt{d}}\subset\mathbb{Z}.$$

As $\gcd(a,b,c)=1$ we can simplify this to

$$\lambda+\lambda^{-1}\in\mathbb{Z},\qquad\text{and}\qquad \frac{\lambda-\lambda^{-1}}{\sqrt{d}}\in\mathbb{Z}.$$

In other words, there are integers m,n such that

$$\lambda=\frac{m+n\sqrt{d}}{2}\qquad\text{and}\qquad \lambda^{-1}=\frac{m-n\sqrt{d}}{2}. \tag{6.6}$$

As $\lambda>1$ the integers m,n are positive and they satisfy the diophantine equation

$$m^2-dn^2=4. \tag{6.7}$$

The equations (6.6)–(6.7) are not only necessary but also sufficient for (6.5) to lie in $\mathrm{SL}_2(\mathbb{Z})$. Namely, (6.5)–(6.7) imply that

$$M\begin{pmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{pmatrix}M^{-1}=\begin{pmatrix}\frac{m-bn}{2} & -nc\\ na & \frac{m+bn}{2}\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z}) \tag{6.8}$$

since

$$m\pm bn\equiv m^2-dn^2\equiv 0 \pmod 2.$$

The λ's given by (6.6)–(6.7) are exactly the totally positive[2] units in the ring of integers $\mathcal{O}_d$ of $\mathbb{Q}(\sqrt{d})$. These units form a group isomorphic to $\mathbb{Z}$ by Dirichlet's theorem, therefore there is a smallest $\lambda=\lambda_d>1$ among them (which generates the group). In other words, the sought $\lambda=\lambda_d>1$ exists and comes from the smallest positive solution of (6.7). In classical language, the matrices (6.8) are the *automorphs* of the form $\langle a,b,c\rangle$.

To summarize, the $\mathrm{SL}_2(\mathbb{Z})$-orbits of forms $\langle a,b,c\rangle$ with given fundamental discriminant d give rise to $h(d)$ Heegner points on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ for $d<0$ and $h(d)$ closed geodesics of length $2\ln(\lambda_d)$ for $d>0$ where $\lambda_d=(m+n\sqrt{d})/2$ is the smallest totally positive unit of $\mathcal{O}_d$ greater than 1. This geometric picture is even more interesting in the light of the following refinement of (6.4) which is a consequence of Dirichlet's class number formula and Siegel's theorem (see [Da00, Chapters 6 and 21]):

$$\begin{aligned}|d|^{1/2-\varepsilon}\ll_\varepsilon h(d)\ll_\varepsilon |d|^{1/2+\varepsilon},\qquad & d<0,\\ d^{1/2-\varepsilon}\ll_\varepsilon h(d)\ln(\lambda_d)\ll_\varepsilon d^{1/2+\varepsilon},\qquad & d>0.\end{aligned} \tag{6.9}$$

[2] i.e. positive under both embeddings $\mathbb{Q}(\sqrt{d})\hookrightarrow\mathbb{R}$

This shows that the set of Heegner points of discriminant $d < 0$ has cardinality about $|d|^{1/2}$, while the set of closed geodesics of discriminant $d > 0$ has total length about $d^{1/2}$.

We now prove that the equivalence classes of forms of fundamental discriminant d can be mapped bijectively to narrow ideal classes of the quadratic number field $\mathbb{Q}(\sqrt{d})$ in a natural fashion. As the latter classes form an abelian group under multiplication, we obtain a natural multiplication law on the equivalence classes of forms. This law, discovered by Gauss [Ga86], is called *composition* in the classical theory. In combination with the previous paragraphs, we obtain that the narrow ideal classes correspond bijectively to the Heegner points (if $d < 0$) or the closed geodesics (if $d > 0$) of discriminant d on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$, and the narrow ideal class group acts on these geometric objects accordingly.

Recall that a fractional ideal of $\mathbb{Q}(\sqrt{d})$ is a finitely generated $\mathcal{O}_d$-module contained in $\mathbb{Q}(\sqrt{d})$ and two nonzero fractional ideals are equivalent (in the narrow sense) if their quotient is a principal fractional ideal generated by a totally positive element of $\mathbb{Q}(\sqrt{d})$. Here "totally positive element" can clearly be changed to "element of positive norm" where the norm of $\mu \in \mathbb{Q}(\sqrt{d})$ is given by $N(\mu) = \mu\bar{\mu}$. Recall also that we can represent equivalence classes of forms of fundamental discriminant d by some

$$Q_i(x, y) = a_i x^2 + b_i xy + c_i y^2 = a_i(x - z_i y)(x - \bar{z}_i y), \qquad i = 1, \ldots, h(d),$$

with

$$a_i > 0, \qquad z_i := \frac{-b_i + \sqrt{d}}{2a_i}, \qquad \bar{z}_i := \frac{-b_i - \sqrt{d}}{2a_i}.$$

It will suffice to show that each fractional ideal I of $\mathbb{Q}(\sqrt{d})$ is equivalent to some fractional ideal

$$I_i := \mathbb{Z} + \mathbb{Z}z_i, \qquad i = 1, \ldots, h(d),$$

and that the fractional ideals I_i are pairwise inequivalent.

Any fractional ideal I can be written as

$$I = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 \qquad \text{with} \qquad \frac{\bar{\omega}_1\omega_2 - \omega_1\bar{\omega}_2}{\sqrt{d}} > 0.$$

We associate to I (and ω_1, ω_2) the binary quadratic form

$$Q_I(x, y) := \frac{(x\omega_1 - y\omega_2)(x\bar{\omega}_1 - y\bar{\omega}_2)}{N(I)},$$

where $N(I) > 0$ is the absolute norm of I, i.e. the multiplicative function that agrees with $(\mathcal{O}_d : I)$ for integral ideals I. We claim first that $Q_I(x, y)$ has integral coefficients and discriminant d. To see the claim we can assume that I is an integral ideal since $Q_I(x, y)$ does not change if we replace I by nI (and ω_i by $n\omega_i$) for some positive integer n. Then ω_1, ω_2 and their conjugates are in $\mathcal{O}_d$ and the claim amounts to:

- $N(I) \mid \omega_1\bar{\omega}_1,\ \omega_1\bar{\omega}_2 + \bar{\omega}_1\omega_2,\ \omega_2\bar{\omega}_2$;
- $(\omega_1\bar{\omega}_2 - \bar{\omega}_1\omega_2)^2 = N(I)^2 d$.

The first statement follows from the fact that $\omega_1, \omega_2, \omega_1 + \omega_2$ are elements of I, hence their norms are divisible by $N(I)$. The second statement follows by writing $\mathcal{O}_d$ as $\mathbb{Z} + \mathbb{Z}\omega$ and then noting that

$$\begin{vmatrix} \omega_1 & \bar{\omega}_1 \\ \omega_2 & \bar{\omega}_2 \end{vmatrix}^2 = (\mathcal{O}_d : I)^2 \begin{vmatrix} 1 & 1 \\ \omega & \bar{\omega} \end{vmatrix}^2 = N(I)^2 d.$$

The claim implies that there is a unique i and a unique $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$Q_I(\alpha x + \beta y, \gamma x + \delta y) = Q_i(x, y).$$

We can write this as

$$\frac{N(\alpha\omega_1 - \gamma\omega_2)}{N(I)}(x - zy)(x - \bar{z}y) = a_i(x - z_i y)(x - \bar{z}_i y),$$

where

$$z := \frac{-\beta\omega_1 + \delta\omega_2}{\alpha\omega_1 - \gamma\omega_2}. \tag{6.10}$$

This implies immediately that

$$N(\alpha\omega_1 - \gamma\omega_2) = a_i N(I) > 0. \tag{6.11}$$

Then a straightforward calculation yields

$$\frac{z - \bar{z}}{\sqrt{d}} = \frac{\alpha\delta - \beta\gamma}{N(\alpha\omega_1 - \gamma\omega_2)} \frac{\bar{\omega}_1\omega_2 - \omega_1\bar{\omega}_2}{\sqrt{d}} > 0$$

which by

$$\frac{z_i - \bar{z}_i}{\sqrt{d}} = \frac{1}{a_i} > 0$$

forces that $z = z_i$. But then (6.10)–(6.11) imply that

$$I = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \mathbb{Z}(\alpha\omega_1 - \gamma\omega_2) + \mathbb{Z}(-\beta\omega_1 + \delta\omega_2)$$

is equivalent to

$$\mathbb{Z} + \mathbb{Z}z = \mathbb{Z} + \mathbb{Z}z_i = I_i.$$

Now assume that I_i and I_j are equivalent, i.e. there is some $\mu \in \mathbb{Q}(\sqrt{d})$ such that

$$\mu(\mathbb{Z} + \mathbb{Z}z_i) = \mathbb{Z} + \mathbb{Z}z_j, \qquad N(\mu) > 0.$$

Then we certainly have some $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z})$ such that

$$\mu = \alpha + \beta z_j, \qquad \mu z_i = \gamma + \delta z_j.$$

In particular,

$$z_i = \frac{\gamma + \delta z_j}{\alpha + \beta z_j} \qquad \text{with} \qquad N(\alpha + \beta z_j) > 0.$$

By a straightforward calculation as before,

$$\frac{z_i - \bar{z}_i}{\sqrt{d}} = \frac{\alpha\delta - \beta\gamma}{N(\alpha + \beta z_j)} \frac{z_j - \bar{z}_j}{\sqrt{d}},$$

which shows that

$$\alpha\delta - \beta\gamma = 1 \qquad \text{and} \qquad N(\alpha + \beta z_j) = \frac{z_j - \bar{z}_j}{z_i - \bar{z}_i} = \frac{a_i}{a_j}.$$

Now we obtain

$$a_i(x - z_i y)(x - \bar{z}_i y) = a_j\big((\alpha + \beta z_j)x - (\gamma + \delta z_j)y\big)\big((\alpha + \beta \bar{z}_j)x - (\gamma + \delta \bar{z}_j)y\big),$$

i.e.

$$Q_i(x, y) = Q_j(\alpha x - \gamma y, -\beta x + \delta y), \qquad \begin{pmatrix} \alpha & -\gamma \\ -\beta & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

This clearly implies that $i = j$, since otherwise the forms Q_i and Q_j are inequivalent.

Incidentally, we see that the equivalence class of the associated form $Q_I(x, y)$ only depends on the narrow class of I (in particular, it is independent of the choice of ordered basis of I) and two fractional ideals I and J are in the same narrow class if and only if $Q_I(x, y)$ and $Q_J(x, y)$ are equivalent.

6.2 Bessel functions

In this section we prove some basic facts concerning Bessel functions.

For $s \in \mathbb{C}$, 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 $\alpha > 0$ and B_s denotes either J_s, Y_s or K_s, then

$$(\alpha\sqrt{x})^s B_s(\alpha\sqrt{x}) = \pm\frac{2}{\alpha^2}\frac{d}{dx}\big((\alpha\sqrt{x})^{s+1} B_{s+1}(\alpha\sqrt{x})\big). \tag{6.12}$$

and for any $j \in \mathbb{N}$,

$$x^j \frac{d^j}{d^j x} B_s\left(\frac{\alpha}{x}\right) = Q_j(s) B_s\left(\frac{\alpha}{x}\right) + Q_{j-1}(s)\left(\frac{\alpha}{x}\right)^1 B_{s-1}\left(\frac{\alpha}{x}\right) + \cdots + Q_0(s)\left(\frac{\alpha}{x}\right)^j B_{s-j}\left(\frac{\alpha}{x}\right), \tag{6.13}$$

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

Lemma 6.1. *Let $F \in C_c^\infty(\mathbb{R}^+)$ be a smooth function of compact support. For $s \in \mathbb{C}$ let B_s denote either of the Bessel functions J_s, Y_s or K_s. Then for $\alpha > 0$ and $j \in \mathbb{N}$ we have*

$$\int_0^\infty F(x) B_s(\alpha\sqrt{x})\,dx = \pm\left(\frac{2}{\alpha}\right)^j \int_0^\infty \frac{d^j}{dx^j}\big(F(x)x^{-\frac{s}{2}}\big) x^{\frac{s+j}{2}} B_{s+j}(\alpha\sqrt{x})\,dx. \tag{6.14}$$

Proof. Using (6.12) and applying integration by parts j times we obtain

$$\begin{aligned}
\int_0^\infty F(x) B_s(\alpha\sqrt{x})\,dx &= \pm\left(\frac{2}{\alpha^2}\right)^j \int_0^\infty F(x)(\alpha\sqrt{x})^{-s}\frac{d^j}{dx^j}\big((\alpha\sqrt{x})^{s+j} B_{s+j}(\alpha\sqrt{x})\big)\,dx \\
&= \pm\left(\frac{2}{\alpha^2}\right)^j \int_0^\infty \frac{d^j}{dx^j}\big(F(x)(\alpha\sqrt{x})^{-s}\big)(\alpha\sqrt{x})^{s+j} B_{s+j}(\alpha\sqrt{x})\,dx \\
&= \pm\left(\frac{2}{\alpha}\right)^j \int_0^\infty \frac{d^j}{dx^j}\big(F(x)x^{-\frac{s}{2}}\big)x^{\frac{s+j}{2}} B_{s+j}(\alpha\sqrt{x})\,dx.
\end{aligned}$$

□

Proposition 6.1. *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 [Ol74]) 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 [Wa44]),

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

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 [Wa44])

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

Proposition 6.2. *For any $\sigma > 0$ and any $\varepsilon > 0$, the following uniform estimates hold in the strip $|\Re s| \leqslant \sigma$:*

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

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

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 [Ol74]). The last estimate for K_s follows from Schläfli's integral representation (see Section 6.22 of [Wa44]),

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

by noting that

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

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

$$\begin{matrix} -\varepsilon - i\infty, & -\varepsilon - i\big(2+2|\Im s|\big), & \sigma+\varepsilon - i\big(2+2|\Im s|\big), \\ \sigma+\varepsilon + i\big(2+2|\Im s|\big), & -\varepsilon + i\big(2+2|\Im s|\big), & -\varepsilon + i\infty. \end{matrix}$$

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 [Er54].

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

$$M_s(x) \ll_{\sigma,\varepsilon} \left(\frac{x}{1+|\Im s|}\right)^{-\sigma-\varepsilon} + \left(\frac{x}{1+|\Im s|}\right)^{\varepsilon}. \tag{6.15}$$

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 (6.15).

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 (6.15).

The proof of Proposition 6.2 is complete. □

Bibliography

[BM10] E. M. Baruch, Z. Mao, A generalized Kohnen–Zagier formula for Maass forms, *J. London Math. Soc.* **82** (2010), 1–16.

[BR05] J. Bernstein, A. Reznikov, Periods, subconvexity of L-functions and representation theory, *J. Differential Geom.* **70** (2005), 129–141.

[Bl04] V. Blomer, Shifted convolution sums and subconvexity bound for automorphic L-functions, *Int. Math. Res. Not.* **2004**, 3905–3926.

[Bl08] V. Blomer, Ternary quadratic forms, and sums of three squares with restricted variables, *Anatomy of integers (J. M. de Koninck, A. Granville, F. Luca eds.)*, 1–17., CRM Proc. Lecture Notes, 46, Amer. Math. Soc., Providence, RI, 2008.

[Bl11] V. Blomer, Subconvexity for twisted L-functions on $GL(3)$, *Amer. J. Math.*, to appear

[BlHa08a] V. Blomer, G. Harcos, Hybrid bounds for twisted L-functions, *J. Reine Angew. Math.* **621** (2008), 53–79.

[BlHa08b] V. Blomer, G. Harcos, The spectral decomposition of shifted convolution sums, *Duke Math. J.* **144** (2008), 321–339.

[BlHa10] V. Blomer, G. Harcos, Twisted L-functions over number fields and Hilbert's eleventh problem, *Geom. Funct. Anal.* **20** (2010), 1–52.

[BHM07a] V. Blomer, G. Harcos, P. Michel, A Burgess-like subconvex bound for twisted L-functions (with Appendix 2 by Z. Mao), *Forum Math.* **19** (2007), 61–105.

[BHM07b] V. Blomer, G. Harcos, P. Michel, Bounds for modular L-functions in the level aspect, *Ann. Sci. École Norm. Sup.* **40** (2007), 697–740.

[BlHo10] V. Blomer, R. Holowinsky, Bounding sup-norms of cusp forms of large level, *Invent. Math.* **179** (2010), 645–681.

[Bu63] D. A. Burgess, On character sums and L-series, *Proc. London Math. Soc.* **12** (1962), 193–206.; II, *ibid.* **13** (1963), 524–536.

[By96] V. A. Bykovskiĭ, A trace formula for the scalar product of Hecke series and its applications, translation in *J. Math. Sci. (New York)* **89** (1998), 915–932.

[CCU09] F. Chamizo, E. Christóbal, A. Ubis, Lattice points in rational ellipsoids, *J. Math. Anal. Appl.* **350** (2009), 283–289.

[CoPS90] J. Cogdell, I. Piatetskii-Shapiro, The arithmetic and spectral analysis of Poincaré series, Perspectives in Mathematics 13, Academic Press, Boston, MA, 1990.

[CI00] B. Conrey, H. Iwaniec, The cubic moment of central values of automorphic L-functions, *Ann. of Math.* **151** (2000), 1175–1216.

[Da00] H. Davenport, Multiplicative number theory [revised and with a preface by H. L. Montgomery], 3rd edition, Graduate Texts in Mathematics 74, Springer-Verlag, New York, 2000.

[De74] P. Deligne, La conjecture de Weil. I, *Inst. Hautes Études Sci. Publ. Math.* **43** (1974), 274–307.

[DI82] J-M. Deshouillers, H. Iwaniec, Kloosterman sums and Fourier coefficients of cusp forms, *Invent. Math.* **70** (1982/83), 219–288.

[Du88] W. Duke, Hyperbolic distribution problems and half-integral weight Maass forms, *Invent. Math.* **92** (1988), 73–90.

[DFI93] W. Duke, J. B. Friedlander, H. Iwaniec, Bounds for automorphic L-functions, *Invent. Math.* **112** (1993), 1–8.

[DFI94a] W. Duke, J. B. Friedlander, H. Iwaniec, A quadratic divisor problem, *Invent. Math.* **115** (1994), 209–217.

[DFI94b] W. Duke, J. Friedlander, H. Iwaniec, Bounds for automorphic L-functions. II, *Invent. Math.* **115** (1994), 219–239.; Erratum, *ibid.* **140** (2000), 227–242.

[DFI01] W. Duke, J. Friedlander, H. Iwaniec, Bounds for automorphic L-functions. III, *Invent. Math.* **143** (2001), 221–248.

[DFI02] W. Duke, J. B. Friedlander, H. Iwaniec, The subconvexity problem for Artin L-functions, *Invent. Math.* **149** (2002), 489–577.

[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.

[DuSP90] 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.

[ELMV11] M. Einsiedler, E. Lindenstrauss, P. Michel, A. Venkatesh, Distribution of periodic torus orbits and Duke's theorem for cubic fields, *Annals of Math.*, to appear

[Er54] A. Erdélyi et al., Tables of integral transforms, Vol. I. [based on notes left by H. Bateman], McGraw-Hill, New York, 1954.

[FM11] A. Folsom, R. Masri, The asymptotic distribution of traces of Maass–Poincaré series, *Adv. Math.* **226** (2011), 3724–3759.

[Fr95] J. B. Friedlander, Bounds for L-functions, *Proc. Int. Congr. Math. (Zürich, 1994), Vol. I*, 363–373, Birkhäuser, Basel, 1995.

[Ga86] C. F. Gauss, Disquisitiones arithmeticae [translated by A. A. Clarke and revised by W. C. Waterhouse, C. Greither and A. W. Grootendorst], Springer-Verlag, New York, 1986.

[GJ79] S. Gelbart, H. Jacquet, Forms on GL_2 from the analytic point of view, *Automorphic forms, representations, and L-functions (A. Borel, W. Casselman eds.), Part 1*, 213–251, Proc. Sympos. Pure Math. 33, Amer. Math. Soc., Providence, RI, 1979.

[GR07] I. S. Gradshteyn, I. M. Ryzhik, Tables of integrals, series, and products, 7th edition, Academic Press, New York, 2007.

[Ha02] G. Harcos, Uniform approximate functional equation for principal L-functions, *Int. Math. Res. Not.* **2002**, 923–932.; Erratum, *ibid.* **2004**, 659–660.

[Ha03a] G. Harcos, An additive problem in the Fourier coefficients of Maass forms, *Math. Ann.* **326** (2003), 347–365.

[Ha03b] G. Harcos, New bounds for automorphic L-functions, Ph. D. thesis, Princeton University, 2003.

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

[HT11] G. Harcos, N. Templier, On the sup-norm of Maass cusp forms of large level. II, submitted

[HB80] D. R. Heath-Brown, Hybrid bounds for Dirichlet L-functions II, *Quart. J. Math. Oxford Ser. (2)*, **31** (1980), 157–167.

[He37] E. Hecke, Über Modulfunktionen und die Dirichletschen Reihen mit Eulerscher Produktentwicklung. I., *Math. Ann.* **114** (1937), 1–28; *II.*, ibid. **114** (1937), 316–351.

[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.

[Hu72] M. N. Huxley, On the difference between consecutive primes, *Invent. Math.* **15** (1972), 164–170.

[Iw87] H. Iwaniec, Fourier coefficients of modular forms of half-integral weight, *Invent. Math.* **87** (1987), 385–401.

[Iw02] H. Iwaniec, Spectral methods of automorphic forms, 2nd edition, Graduate Studies in Mathematics 53, American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002.

[IK04] H. Iwaniec, E. Kowalski, Analytic number theory, American Mathematical Society Colloquium Publications 53, American Mathematical Society, Providence, RI, 2004.

[IS00] H. Iwaniec, P. Sarnak, Perspectives on the analytic theory of L-functions, *Geom. Funct. Anal. Special Volume* (2000), 705–741.

[Ju92] M. Jutila, Transformations of exponential sums, *Proceedings of the Amalfi Conference on Analytic Number Theory (Maiori 1989)*, 263–270, Univ. Salerno, Salerno, 1992.

[Ju96] M. Jutila, A variant of the circle method, *Sieve methods, exponential sums and their applications in number theory*, 245–254, Cambridge University Press, 1996.

[Ju99] M. Jutila, Convolutions of Fourier coefficients of cusp forms, *Publ. Inst. Math. (Beograd) (N.S.)* **65(79)** (1999), 31–51.

[JM05] M. Jutila, Y. Motohashi, Uniform bounds for Hecke L-functions, *Acta Math.* **195** (2005), 61–115.

[JM06] M. Jutila, Y. Motohashi, Uniform bounds for Rankin-Selberg L-functions, *Multiple Dirichlet series, automorphic forms, and analytic number theory*, 243–256, Proc. Sympos. Pure Math., 75, Amer. Math. Soc., Providence, RI, 2006.

[KS99] N. M. Katz, P. Sarnak, Zeroes of zeta functions and symmetry, *Bull. Amer. Math. Soc. (N.S.)* **36** (1999), 1–26.

[KMY11] B. D. Kim, R. Masri, T. H. Yang, Nonvanishing of Hecke L-functions and the Bloch–Kato conjecture, *Math. Ann.* **349** (2011), 301–343.

[Ki03] 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.

[KiSa03] H. Kim, P. Sarnak, Appendix: Refined estimates towards the Ramanujan and Selberg Conjectures, *J. Amer. Math. Soc.* 16 (2003), 175–181.

[KiSh02] H. Kim, F. Shahidi, Cuspidality of symmetric powers with applications, *Duke Math. J.* **112** (2002), 177–197.

[Kl26] H. D. Kloosterman, On the representation of numbers in the form $ax^2+by^2+cz^2+dt^2$, *Acta Math.* **49** (1926), 407–464.

[KMV00] E. Kowalski, P. Michel, J. VanderKam, Mollification of the fourth moment of automorphic L-functions and arithmetic applications, *Invent. Math.* **142** (2000), 95–151.

[KMV02] E. Kowalski, P. Michel, J. VanderKam, Rankin–Selberg L-functions in the level aspect, *Duke Math. J.* **114** (2002), 123–191.

[La73] J. L. Lagrange, Recherche d'arithmétique, *Nouv. Mém. Acad. Berlin* (1773), 265–312.

[LLY06] Y.-K. Lau, J. Liu, Y. Ye, A new bound $k^{2/3+\varepsilon}$ for Rankin-Selberg L-functions for Hecke congruence subgroups, *Int. Math. Res. Pap.* **2006**, Art. ID 35090, 78 pp.

[Li68] 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.

[Li11] X. Li, Bounds for $GL(3) \times GL(2)$ L-functions and $GL(3)$ L-functions, *Annals of Math.* **173** (2011), 301–336.

[Ma49] H. Maass, Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen, *Math. Ann.* **121** (1949), 141–183.

[Ma11] R. Masri, The asymptotic distribution of traces of cycle integrals of the j-function, submitted

[MY11] R. Masri, T. H. Yang, Nonvanishing of Hecke L-functions for CM fields and ranks of abelian varieties, submitted

[Me88] T. Meurman, On exponential sums involving the Fourier coefficients of Maass wave forms, *J. Reine Angew. Math.* **384** (1988), 192–207.

[Me01] T. Meurman, On the binary additive divisor problem, *Number theory (Turku 1999)*, 223–246, de Gruyter, Berlin, 2001.

[Mi04] P. Michel, The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points, *Ann. of Math.*, **160** (2004), 185–236.

[Mi07] P. Michel, Analytic number theory and families of automorphic L-functions, *Automorphic forms and applications*, 181–295, IAS/Park City Math. Ser. 12, Amer. Math. Soc., Providence, RI, 2007.

[MV06] P. Michel, A. Venkatesh, Equidistribution, L-functions and ergodic theory: on some problems of Yu. Linnik, *Proc. Int. Congr. Math. (Madrid 2006), Vol. II*, 421–457, Eur. Math. Soc., Zürich, 2006

[MV07] P. Michel, A. Venkatesh, Heegner points and non-vanishing of Rankin/Selberg L-functions, *Analytic number theory*, 169–183, Clay Math. Proc. 7, Amer. Math. Soc., Providence, RI, 2007.

[MV10] P. Michel, A. Venkatesh, The subconvexity problem for GL_2, *Publ. Math. Inst. Hautes Études Sci.* **111** (2010), 171–271.

[Ol74] F. W. J. Olver, Asymptotics and special functions, Academic Press, New York, 1974.

[Po06] A. Popa, Central values of Rankin L-series over real quadratic fields, *Compos. Math.* **142** (2006), 811–866.

[Pr05] N. V. Proskurin, On the general Kloosterman sums, translation in *J. Math. Sci. (New York)* **129** (2005), 3874–3889.

[Sa03] P. Sarnak, Spectra of hyperbolic surfaces, *Bull. Amer. Math. Soc. (N.S.)* **40** (2003), 441–478.

[Sa07] P. Sarnak, Reciprocal geodesics, *Analytic number theory*, 217–237, Clay Math. Proc. 7, Amer. Math. Soc., Providence, RI, 2007.

[Si80] C. L. Siegel, Advanced analytic number theory, 2nd edition, Tata Institute of Fundamental Research Studies in Mathematics, 9, Tata Institute of Fundamental Research, Bombay, 1980.

[T10] N. Templier, On the sup-norm of Maass cusp forms of large level, *Selecta Math. (N.S.)* **16** (2010), 501–531.

[Ve10] A. Venkatesh, Sparse equidistribution problems, period bounds and subconvexity, *Ann. of Math.* **172** (2010), 989–1094.

[Wa81] J.-L. Waldspurger, Sur les coefficients de Fourier des formes modulaires de poids demi-entier, *J. Math. Pures et Appliquées* **60** (1981), 374–484.

[Wa44] G. N. Watson, A treatise on the theory of Bessel functions, Cambridge University Press, Cambridge, 1944.

[Zh01] S. Zhang, Gross–Zagier formula for GL_2, *Asian J. Math.* **5** (2001), 183–290.