J. Eur. Math. Soc. 23, 237–314 (2021) DOI 10.4171/JEMS/1011

Eren Mehmet Kıral · Matthew P. Young

The fifth moment of modular L-functions

Received October 13, 2017

Abstract. We prove a sharp bound on the fifth moment of modular L-functions of fixed small weight and large prime level.

Keywords. L-functions, moment, central value, spectral reciprocity

1. Introduction

Let q be a prime and $\kappa \in \{4, 6, 8, 10, 14\}$. Let $H_\kappa(q)$ be the set of weight κ Hecke eigenforms on $\Gamma_0(q)$. For any $f \in H_\kappa(q)$ (note that any such f is automatically a newform), let $\lambda_f(n)$ denote its n^{th} Hecke eigenvalue. Our main result is the following theorem:

Theorem 1.1. *We have*

$$\sum_{f \in H_\kappa(q)} L(1/2, f)^5 \ll q^{1+\theta+\varepsilon} \quad \text{as } q \to \infty \text{ among primes.} \tag{1.1}$$

Here θ is the best-known progress towards the Ramanujan–Petersson conjecture.

The currently best-known value $\theta = 7/64$ is given by the work of Kim and Sarnak [Kim03]. The central value $L(1/2, f)$ is nonnegative by [Koh85, Corollary 2] and [Wal81], so upon dropping all but one term, we deduce:

Corollary 1.2. *For any $\varepsilon > 0$, we have*

$$L(1/2, f) \ll_\varepsilon q^{(1+\theta)/5+\varepsilon}. \tag{1.2}$$

Previously, Duke, Friendlander, and Iwaniec [DFI94] bounded the amplified fourth moment in this family, and Kowalski, Michel, and VanderKam [KMV00] asymptotically evaluated a mollified fourth moment. Recently, Balkanova and Frolenkov [BF17] improved the error term in these fourth moment problems, and thereby deduced the so-far

E. M. Kıral, M. P. Young: Department of Mathematics, Texas A&M University,
College Station, TX 77843-3368, U.S.A.;
e-mail: ekiral@tamu.edu, erenmehmetkiral@protonmail.com, myoung@math.tamu.edu

Mathematics Subject Classification (2020): Primary 11F66; Secondary 11F72, 11L05

best-known subconvexity result of $L(1/2, f) \ll q^{\frac{27-30\theta}{112-128\theta}}$. Corollary 1.2 improves this further. Our method of proof takes a different course from these works, and we never solve a shifted convolution problem.

This paper has some common features with the study of the cubic moment by Conrey and Iwaniec [CI00]. Their work also bounds a moment that is 1 larger than what one may consider the "barrier" to subconvexity. That is, for the family of L-functions they consider, an upper bound on the second moment that is Lindelöf-on-average leads back precisely to the convexity bound on an individual L-value. Similarly, the fourth moment is the "barrier" in the family of Theorem 1.1. In a sense, going one full moment beyond the barrier is a way of amplifying with the L-function itself. As far as the authors are aware, prior to Theorem 1.1, the only known instances of a sharp upper bound on a moment that is 1 larger than the barrier moment are the cubic moment and its generalizations [CI00], [Ivi01], [Pet15], [You17], [PY19].

In Section 2, we give a simplified sketch of the argument. The main overall difficulty in the problem is that we require a significant amount of cancellation in multivariable sums with divisor functions and Kloosterman sums. The main thrust of the argument is to apply summation formulas that shorten the lengths of summation, eventually obtaining a sum of Kloosterman sums. To this, we apply the Bruggeman–Kuznetsov formula, which leads to a fourth moment of Hecke–Maass L-functions twisted by $\lambda_j(q)$, with an additional average over the level. This is another incarnation of a Kuznetsov/Motohashi-type formula where a moment problem in one family of L-functions is related to another moment in a "dual" family (see [MV10, Section 1.1.3]). Along the way, we encounter many "fake" main terms, which turn out to be surprisingly difficult to estimate. A straightforward bound on these would only lead to $O(q^{5/4+\varepsilon})$ in Theorem 1.1, which would be trivial. We expect that all the "fake" main terms calculated in this paper should essentially cancel, but doing so is a daunting prospect. Instead, we show that with an appropriate choice of weight functions in the approximate functional equations, all the fake main terms are bounded consistently with Theorem 1.1. The amplified/mollified fourth moment (cf. [DFI94], [KMV00]) also required a difficult analysis of the main terms, which arose from solving the shifted convolution problem. Therefore, it is not clear how to compare the main term calculations here with [DFI94], [KMV00]. The article [BHM07, Section 1.2] has a more thorough discussion of the main term analysis with the shifted convolution approach.

One of the practical difficulties in applying the Bruggeman–Kuznetsov formula in applications is that one needs to recognize the particular shape of sums of Kloosterman sums one encounters (with coprimality and congruence conditions, etc.) as one associated to a group Γ, pair of cusps $\mathfrak{a}, \mathfrak{b}$, and nebentypus χ. To this end, in [KY19] we have identified all the Kloosterman sums belonging to the congruence subgroup $\Gamma_0(N)$ and at general Atkin–Lehner cusps (i.e., those cusps equivalent to ∞ under an Atkin–Lehner involution) with general Dirichlet characters. It turns out that there is a "correct" choice of scaling matrix to use when computing the Fourier coefficients and Kloosterman sums, a choice that ensures the multiplicativity of Fourier coefficients at the Atkin–Lehner cusps. In Section 3, we record the special cases of these Kloosterman sums that are required in this work.

Another practical difficulty is estimating oscillatory integral transforms with weight functions depending on multiple variables, with numerous parameters. It is desirable to perform stationary phase analysis on a single variable at a time, yet still keep track of the behavior of the remaining variables in a succinct way. We have codified some properties of a family of weight functions that allows us to do this efficiently. The key stationary phase result, which is a modest generalization of work in [BKY13], is stated as Lemma 4.3 below, with a proof appearing in [KPY19].

In the spectral analysis of a sum of Kloosterman sums, it is necessary to treat the Maass forms, holomorphic forms, and continuous spectrum. In our situation, the Maass forms and holomorphic forms are rather similar, and lead to a twisted fourth moment of GL_2 cuspidal L-functions. The continuous spectrum is similar in many respects, but a key difference is that the "dual" family of L-functions is essentially a sum of a product of eight Dirichlet L-functions at shifted arguments. One naturally wishes to treat the continuous spectrum on the same footing as the discrete spectrum, which requires shifting some contour integrals past the poles of the Dirichlet L-functions (which occur only when the character is principal). There is potentially a large loss in savings from these poles on the 1-line compared to the contribution from the $1/2$-line. Luckily, it turns out that there is some extra savings in the residues of the Dirichlet series (essentially, from considering only the principal characters) that balances against this loss. This savings ultimately arises from a careful calculation of the Fourier expansion of the Eisenstein series on $\Gamma_0(N)$ with arbitrary N, attached to an arbitrary cusp, expanded around any Atkin–Lehner cusp. This calculation occurs in [KY19].

An astute reader may note that $\kappa = 2$ is not covered by Theorem 1.1. In fact, there is only a single instance where our proof requires that $\kappa > 2$, namely in the study of the continuous part of the spectrum at (11.27). Perhaps with further analysis one might incorporate the weight 2 case, by a more careful analysis of the residues of the Dirichlet L-functions. The restriction that q is a prime and that $\kappa \leq 14$, $\kappa \neq 12$, means that the cusp forms $f \in H_\kappa(\Gamma_0(q))$ are automatically newforms. It is reasonable to expect that using a more general Petersson formula for newforms (e.g., see [PY19]) could relax these assumptions, but the arithmetical complexity would be increased.

2. High-level sketch

Here we include an outline of the major steps used in the proof, intended for an expert audience. By approximate functional equations and the Petersson formula, we arrive at

$$\mathcal{S} := \sum_{m \ll q} \sum_{n \ll q^{3/2}} \sum_{c \equiv 0 \,(\mathrm{mod}\, q)} \frac{\tau(m)\tau_3(n)}{c\sqrt{mn}} S(m,n;c) J_{\kappa-1}\left(\frac{4\pi\sqrt{mn}}{c}\right), \tag{2.1}$$

and we wish to show $\mathcal{S} \ll q^{\theta+\varepsilon}$. The hardest case to consider is $m \asymp q$, $n \asymp q^{3/2}$, and $c \asymp \sqrt{mn} \asymp q^{5/4}$, in which case $J_{\kappa-1}(x) \approx 1$. In practice, one needs to treat the two ranges of the Bessel function (i.e., $x \ll 1$ and $x \gg 1$) differently. In this sketch, we focus on the transition region of the J-Bessel function where $x \asymp 1$.

The Weil bound applied to the Kloosterman sum shows $\mathcal{S} \ll q^{7/8+\varepsilon}$, far from $q^{\theta+\varepsilon}$.

The immediate problem with (2.1) is that Voronoi summation applied to m or n leads to a dual sum that is longer than before. The conventional wisdom is that this is a bad move. However, there do not seem to be any other moves available, so it may be necessary to take a loss in the first step. We may attempt to minimize the loss here by opening $\tau(m) = \sum_{m_1 m_2 = m} 1$, supposing $m_1 \leq m_2$ by symmetry, and applying Poisson summation in m_2 modulo c. This leads to

$$\mathcal{S} \approx \sum_{m_1 \ll q^{1/2}} \sum_{\substack{c \equiv 0 \pmod q \\ c \asymp q^{5/4}}} \sum_{k \ll q^{3/4}} \sum_{n \asymp q^{3/2}} \frac{\tau_3(n)}{\sqrt{m_1 k n c}} e\Big(\frac{m_1 n \overline{k}}{c}\Big).$$

Note that the trivial bound now gives $\mathcal{S} \ll q$, so we lost a factor $q^{1/8}$ from going the "wrong way" in Poisson. However, now we may gain from the structure of the arithmetical part by applying the well-known reciprocity formula

$$e\Big(\frac{m_1 n \overline{k}}{c}\Big) = e\Big(-\frac{m_1 n \overline{c}}{k}\Big) e\Big(\frac{m_1 n}{ck}\Big).$$

This effectively switches the roles of c and k, at the expense of introducing the potentially oscillatory factor $e(\frac{m_1 n}{ck})$ into the weight function. However, when all variables are near their maximial sizes, this factor is not oscillatory, so we shall ignore it in this sketch.

Side remark. If one applies Voronoi to the sum over m (which is more in line with the previous works on the amplified/mollified fourth moment), then one encounters a shifted divisor sum of the form $\sum_{m-n=h} \tau_2(m)\tau_3(n)$. Such sums have been considered by various authors, with the most advanced results being the recent work of B. Topaçoğulları [Top16].

One way to proceed next would be to convert the additive character into Dirichlet characters (modulo k), which has a nice benefit of separating the variables, a key step in [CI00]. This would lead to a fifth moment of Dirichlet L-functions twisted by Gauss sums, with an averaging over the modulus. One may check that Lindelöf applied to these L-functions only gives $\mathcal{S} \ll q^{1/4+\varepsilon}$, which in a sense gets back to the convexity bound.

Now it is beneficial to apply Voronoi summation in n modulo k (one may view this as opening $\tau_3(n) = \sum_{n_1 n_2 n_3 = n} 1$, and applying Poisson in each n_i). This leads to

$$\mathcal{S} \approx \sum_{m_1 \ll q^{1/2}} \sum_{\substack{c \equiv 0 \pmod q \\ c \asymp q^{5/4}}} \sum_{k \ll q^{3/4}} \sum_{p \ll q^{3/4}} \frac{\tau_3(p)}{k\sqrt{m_1 p c}} S(p, c\overline{m_1}, k). \tag{2.2}$$

The trivial bound now is $q^{5/8}$, consistent with saving $q^{1/8}$ in each of the n_i variables, just as we lost $q^{1/8}$ by Poisson in the initial m_2 variable. One could also apply Poisson in m_1 to save another $q^{1/8}$, but then the arithmetical sum becomes a hyper-Kloosterman sum, which increases the difficulty of the problem (N. Pitt has studied this problem [Pit95], but it seems very hard to obtain enough cancellation using this approach). Here we have a

Kloosterman sum to which we may apply the Bruggeman–Kuznetsov formula of level m_1. Using this, we obtain

$$\mathcal{S} \approx \sum_{m_1 \ll q^{1/2}} \sum_{\substack{t_j \ll q^{\varepsilon} \\ \text{level } m_1}} \sum_{\substack{c \equiv 0 \ (\mathrm{mod}\ q) \\ c \asymp q^{5/4}}} \sum_{p_1, p_2, p_3 \ll q^{1/4}} \frac{\nu_j(p_1 c) \nu_j(p_2 p_3)}{\sqrt{p_1 p_2 p_3 c}}. \tag{2.3}$$

We can essentially write this as

$$\mathcal{S} \approx \sum_{m_1 \ll q^{1/2}} \sum_{\substack{t_j \ll q^{\varepsilon} \\ \text{level } m_1}} \nu_j(1)^2 \frac{\lambda_j(q)}{\sqrt{q}} L(1/2, u_j)^4. \tag{2.4}$$

Here the scaling on the spectral data is that $\sum_{t_j \ll T} \nu_j(1)^2 \ll T^2 m_1^{\varepsilon}$. Thus we have converted to a twisted fourth moment of Maass form L-functions, and one can see how $q^{\theta+\varepsilon}$ emerges by bounding $|\lambda_j(q)| \ll q^{\theta+\varepsilon}$, and using a Lindelöf-on-average bound for the spectral fourth moment (which in turn is “easy”, following from the spectral large sieve inequality).

The above discussion implicitly assumed that the p_i are nonzero. The zero frequencies (where some or all p_i are 0) turn out to be the “fake” main terms alluded to in the introduction.

To handle these, we compute the weight function explicitly, and evaluate the sums over k, m_1, and c as zeta quotients. We later bound the integral by moving lines of integration, and apparent poles of the integrand are cancelled by a choice of the weight function in the approximate functional equation. To elaborate on this point, consider an overly simplified model with a sum of the form $S = \sum_{n \geq 1} \frac{1}{\sqrt{n}} V\big(\frac{n}{\sqrt{q}}\big)$, where $V(x) = \frac{1}{2\pi i} \int_{(1)} \frac{G(s)}{s} x^{-s} \, ds$, and $G(s)$ is analytic satisfying $G(0) = 1$, with rapid decay in the imaginary direction. The trivial bound applied to S gives $S = O(q^{1/4})$, using $V(x) \ll (1+x)^{-100}$. Alternatively, we have $S = \frac{1}{2\pi i} \int_{(1)} \zeta(1/2+s) q^{s/2} \frac{G(s)}{s} \, ds$, which by shifting contours to the line $\mathrm{Re}(s) = \varepsilon > 0$ gives $S = G(1/2) q^{1/4} + O(q^{\varepsilon})$. If $G(1/2) = 0$ (which one is free to assume in the context of the approximate functional equation), then in fact one has an improved bound of $S = O(q^{\varepsilon})$. This is the principal idea behind the estimation of the fake main terms. The main difficulty in practice is that one has a much more complicated sum with multiple variables and weight functions that arise as integral transforms, and it requires significant work to recognize instances of this basic idea. One should also observe that the above method of estimating S is highly reliant on the specific form of the weight function V; if it were multiplied by a compactly supported bump function (say one part of a dyadic partition of unity), then one could not deduce $S = O(q^{\varepsilon})$ anymore. Since we shall apply dyadic partitions of unity in the forthcoming treatment, for the purposes of estimating these fake main terms, it is crucial to re-assemble the partitions.

The role of the m_1 variable within the proof has some curious features. In the sketch above up through (2.4), the m_1 variable was hardly used. Precisely, we never applied a summation formula nor obtained any cancellation from this variable. Nor did we use any reciprocity involving m_1 to lower a modulus. However, nontrivial estimations involving m_1 do appear in other parts of the proof. In the evaluation of one type of fake main term in Section 13.7, we evaluate the m_1-sum similarly to the discussion in the previous paragraph; the lack of pole at $s = 1/2$ amounts to square-root cancellation in this variable. The other location is in estimating the continuous spectrum analog of (2.3) which so far was neglected in this sketch. One may show that the continuous spectrum analog of (2.4) is $O(q^{\varepsilon})$ using the fact that the number of cusps on $\Gamma_0(m_1)$ is at most $O(m_1^{1/2+\varepsilon})$. However, on average over m_1, the number of cusps is $O(m_1^{\varepsilon})$, which leads to a bound that saves an additional factor $q^{1/4}$. In a sense, this discussion indicates that the continuous spectrum is much smaller in measure in the level aspect than the discrete spectrum, on average over the level.

Another point worthy of mention is that at the very first step of the proof leading to (2.1), we may in practice apply approximate functional equations without an explicit presence of the root number. This arises because $L(s, f)^2$ has root number $+1$, but vanishes at $s = 1/2$ if $L(s, f)$ has root number -1. Therefore, when we apply an approximate functional equation of $L(1/2, f)^3$, we may substitute its root number as $+1$. As a consequence, our method does not readily generalize to bound shifted fifth moments of the form $\sum_{f\in\mathcal{H}_\kappa(q)}\prod_{i=1}^{5} L(1/2+\alpha_i, f)$. It is difficult to justify the interest in such a bound without nonnegativity, of course. Without extensive work, it is difficult to predict how crucial this root number trick is; there are certainly examples of moment problems where the presence of the root number is a major obstacle.

Added in the pre-publication: Blomer and Khan [BK19] have derived an *exact* spectral reciprocity formula for the twisted fourth moment, systematizing the method of proof in this paper. Summing over the twist variable may produce an additional L-value, by which one can form the fifth moment studied in this paper. The formula (2.4) can be seen as an impressionistic representation of their formula. In addition, by considering an amplified fourth moment in place of the fifth moment, they arrive at a subconvexity bound that improves on Corollary 1.2 for the present best-allowable value of $\theta = 7/64$, though curiously the fifth moment leads to a better result assuming $\theta = 0$.

3. Kloosterman sums and Bruggeman–Kuznetsov formula

3.1. Cusps, scaling matrices, and Kloosterman sums

We mostly follow the notation of [Iwa02]. Let N be a positive integer and $\Gamma = \Gamma_0(N)$. Let $\mathfrak{a}$ be a cusp and $\Gamma_{\mathfrak{a}} = \{\gamma \in \Gamma : \gamma\mathfrak{a} = \mathfrak{a}\}$ be the stabilizer of the cusp $\mathfrak{a}$ in Γ. A matrix $\sigma_{\mathfrak{a}} \in \mathrm{SL}_2(\mathbb{R})$ satisfying

$$\sigma_{\mathfrak{a}}\infty = \mathfrak{a} \quad \text{and} \quad \sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\sigma_{\mathfrak{a}} = \left\{\pm\left(\begin{smallmatrix}1 & n\\ 0 & 1\end{smallmatrix}\right) : n \in \mathbb{Z}\right\} \tag{3.1}$$

is called a *scaling matrix* for the cusp $\mathfrak{a}$.

Definition 3.1. Let f be a Maass form for the group Γ. The *Fourier coefficients* of f at a cusp $\mathfrak{a}$, denoted $\rho_{\mathfrak{a}f}(n)$, are defined by

$$f(\sigma_{\mathfrak{a}}z) = \sum_{n\neq 0}\rho_{\mathfrak{a}f}(n)e(nx)W_{0,it_j}(4\pi|n|y), \tag{3.2}$$

where W_{0,it_j} is the Whittaker function defined by

$$W_{0,it_j}(4\pi y) = 2\sqrt{y}\, K_{it_j}(2\pi y).$$

The Fourier coefficients $\rho_{\mathfrak{a} f}(n)$ depend on the choice of scaling matrix $\sigma_{\mathfrak{a}}$, and it may be more accurate to denote them $\rho_{\sigma_{\mathfrak{a}}, f}(n)$.

Definition 3.2. For $\mathfrak{a}$ and $\mathfrak{b}$ cusps for Γ, we define the *Kloosterman sum* associated to $\mathfrak{a}$, $\mathfrak{b}$ with modulus c as

$$S_{\mathfrak{a}\mathfrak{b}}(m, n; c) = \sum_{\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_\infty \backslash \sigma_{\mathfrak{a}}^{-1} \Gamma \sigma_{\mathfrak{b}} / \Gamma_\infty} e\left(\frac{am + dn}{c}\right). \tag{3.3}$$

Definition 3.3. The set of allowed moduli is

$$\mathcal{C}_{\mathfrak{a}\mathfrak{b}} = \left\{\gamma > 0 : \left(\begin{smallmatrix} * & * \\ \gamma & * \end{smallmatrix}\right) \in \sigma_{\mathfrak{a}}^{-1} \Gamma \sigma_{\mathfrak{b}}\right\}. \tag{3.4}$$

Notice that if $\gamma \notin \mathcal{C}_{\mathfrak{a}\mathfrak{b}}$ the Kloosterman sum of modulus γ is an empty sum.

3.2. *Atkin–Lehner cusps*

Assume that $N = rs$ with $(r, s) = 1$. We call a cusp of the form $\mathfrak{a} = 1/r$ (with $(r, s) = 1$) an *Atkin–Lehner cusp*. The Atkin–Lehner cusps are precisely those that are equivalent to ∞ under an Atkin–Lehner operator, justifying their name.

A newform is an eigenfunction of all the Hecke operators, as well as all the Atkin–Lehner operators. It turns out that one may choose a scaling matrix $\sigma_{1/r}$ for the Atkin–Lehner cusp $1/r$ to be an Atkin–Lehner operator (see [KY19, Section 2.2]). Therefore, for such a choice of scaling matrix, we have

$$\rho_{\frac{1}{r} f}(n) = \eta_s(f) \rho_{\infty f}(n), \tag{3.5}$$

where f is a newform and $\eta_s(f) = \pm 1$ is the eigenvalue of the Atkin–Lehner operator W_s.

Proposition 3.4. *Let $N = rs$ with $(r, s) = 1$, and choose $\sigma_{1/r}$ as above. Then the set of allowed moduli for the pair of cusps ∞, $1/r$ is*

$$\mathcal{C}_{\infty, 1/r} = \{\gamma = c\sqrt{s} > 0 : c \equiv 0 \pmod{r},\ (c, s) = 1\}, \tag{3.6}$$

and for such $\gamma = c\sqrt{s} \in \mathcal{C}_{\infty,1/r}$, the Kloosterman sum to modulus γ is given by

$$S_{\infty, 1/r}(m, n; c\sqrt{s}) = S(\bar{s}m, n; c), \tag{3.7}$$

where the S on the right denotes an ordinary Kloosterman sum. Consequently,

$$\sum_{\substack{(c,s)=1 \\ c \equiv 0 \ (\mathrm{mod}\, r)}} S(\bar{s}m, n; c) f(c) = \sum_{\gamma \in \mathcal{C}_{\infty,1/r}} S_{\infty,1/r}(m, n; \gamma) f(\gamma/\sqrt{s}), \tag{3.8}$$

where f is any function such that the sums converge.

For this computation, see [Mot07, Section 14], in particular (14.8). Note that (3.7) differs from a formula in [Iwa97, p. 58] by an additive character, which is due to a different choice of the scaling matrix. See also [KY19] for a generalization with different cusps and characters.

3.3. *Bruggeman–Kuznetsov formula*

We record the spectral expansion of a sum of Kloosterman sums in a spectral basis of the space $L^2(\Gamma_0(N))$. Let $\{u_j\}$ be a basis of cusp forms. Assume that u_j is an eigenfunction of the Laplace–Beltrami operator with eigenvalue $1/4+t_j^2$. Call t_j the *spectral parameter* of u_j. Define $\rho_{\mathfrak{a}j}(n) = \rho_{u_j}(\sigma_{\mathfrak{a}}, n)$ as in (3.2); our choice of $\sigma_{\mathfrak{a}}$, in practice, will be an Atkin–Lehner operator.

Likewise, write the Fourier expansion of the Eisenstein series as

$$E_{\mathfrak{c}}(\sigma_{\mathfrak{a}} z, u) = \delta_{\mathfrak{a}\mathfrak{c}} y^u + \rho_{\mathfrak{a}\mathfrak{c}}(0,u) y^{1-u} + \sum_{n\neq 0} \rho_{\mathfrak{a}\mathfrak{c}}(n,u) e(nx) W_{0,u-1/2}(4\pi|n|y). \tag{3.9}$$

Consulting [Iwa02, Theorem 3.4], we have

$$\rho_{\mathfrak{a}\mathfrak{c}}(n,u) = \begin{cases} \phi_{\mathfrak{a}\mathfrak{c}}(n,u)\frac{\pi^u}{\Gamma(u)}|n|^{u-1} & \text{if } n\neq 0, \\ \delta_{\mathfrak{a}\mathfrak{c}} y^u + \phi_{\mathfrak{a}\mathfrak{c}}(u) y^{1-u} & \text{if } n = 0, \end{cases} \tag{3.10}$$

where

$$\phi_{\mathfrak{a}\mathfrak{c}}(n,u) = \sum_{\substack{(\gamma,\delta) \text{ such that} \\ \rho = \left(\begin{smallmatrix} * & * \\ \gamma & \delta \end{smallmatrix}\right) \in \Gamma_\infty \backslash \sigma_{\mathfrak{c}}^{-1}\Gamma\sigma_{\mathfrak{a}}/\Gamma_\infty}} \frac{1}{\gamma^{2u}} e\left(\frac{n\delta}{\gamma}\right) = \sum_{\gamma\in\mathcal{C}_{\mathfrak{c}\mathfrak{a}}} \frac{S_{\mathfrak{c}\mathfrak{a}}(0,n;\gamma)}{\gamma^{2u}}, \tag{3.11}$$

and $\phi_{\mathfrak{a}\mathfrak{c}}(u) = \phi_{\mathfrak{a}\mathfrak{c}}(0,u)$. Note that our ordering of the cusps in the notation $\rho_{\mathfrak{a}\mathfrak{c}}, \phi_{\mathfrak{a}\mathfrak{c}}$ is reversed from that of [Iwa02], and also that [Iwa02, (3.22)] should have $\mathcal{S}_{\mathfrak{a}\mathfrak{c}}(n,0;c)$ in place of $\mathcal{S}_{\mathfrak{a}\mathfrak{c}}(0,n;c)$ to be consistent with [Iwa02, (2.23)]. We give an explicit computation of $\phi_{\mathfrak{a}\mathfrak{c}}(n,u)$ with Proposition 12.2 below.

For aesthetic purposes, define as in [Iwa02, (8.5), (8.6)]

$$\nu_{\mathfrak{a}j}(n) = \left(\frac{4\pi|n|}{\cosh(\pi t_j)}\right)^{1/2} \rho_{\mathfrak{a}j}(n), \qquad \nu_{\mathfrak{a}\mathfrak{c}}(n,u) = \left(\frac{4\pi|n|}{\cos(\pi(u-1/2))}\right)^{1/2} \rho_{\mathfrak{a}\mathfrak{c}}(n,u). \tag{3.12}$$

Let $g \in H_\ell(N)$, that is, let g be a holomorphic level N weight ℓ modular cusp form. Define the Fourier expansion of g at a cusp $\mathfrak{a}$ by

$$g|_{\sigma_{\mathfrak{a}}}(z) = \sum_{n=1}^{\infty} \rho_{\mathfrak{a}g}(n) n^{(\ell-1)/2} e(nz).$$

Also define

$$\nu_{\mathfrak{a}g}(n) = \left(\frac{\pi^{-\ell}\Gamma(\ell)}{4^{\ell-1}}\right)^{1/2} \rho_{\mathfrak{a}g}(n), \tag{3.13}$$

similarly to [Iwa02, (9.42)], but note that $m^{(\ell-1)/2}$ was already extracted in the definition of $\rho_{\mathfrak{a}g}(m)$.

With the notation as above, define for nonzero m and n,

$$\mathcal{K} = \sum_{\gamma\in\mathcal{C}_{\mathfrak{a},\mathfrak{b}}} S_{\mathfrak{a}\mathfrak{b}}(m,n;\gamma)\phi(\gamma), \tag{3.14}$$

where ϕ is smooth and compactly supported on $(0, \infty)$. We then quote the literature for a spectral formula for this sum. Many authors state the Bruggeman–Kuznetsov formula with a weight function of the form $\gamma^{-1} F(4\pi\sqrt{mn}/\gamma)$ in place of $\phi(\gamma)$, which amounts to the substitution $F(t) = (4\pi\sqrt{mn}/t)\phi(4\pi\sqrt{mn}/t)$.

Theorem 3.5 ([Iwa02, Chapter 9]). *Let $\mathcal{K}$ be as in (3.14). Assuming ϕ is smooth with compact support on $(0, \infty)$, we have*

$$\mathcal{K} = \mathcal{K}_d + \mathcal{K}_c + \mathcal{K}_h.$$

Here $\mathcal{K}_h = 0$ if $mn < 0$, and otherwise

$$\mathcal{K}_h = \sum_{\ell>0,\, even} \phi_h(\ell) i^\ell \sum_{g \in H_\ell(N)} \overline{\nu_{\mathfrak{b}g}(m) \nu_{\mathfrak{a}g}(n)}. \tag{3.15}$$

The discrete spectrum contribution is

$$\mathcal{K}_d = \sum_{t_j} \phi_\pm(t_j) \overline{\nu_{\mathfrak{b}j}(m)} \nu_{\mathfrak{a}j}(n), \tag{3.16}$$

where the summation is over the spectral parameters t_j of a chosen orthonormal basis $\{u_j\}_j$ of cusp forms. The continuous spectrum contribution is

$$\mathcal{K}_c = \sum_{\mathfrak{c}} \frac{1}{4\pi} \int_{-\infty}^{\infty} \phi_\pm(t) \overline{\nu_{\mathfrak{b}\mathfrak{c}}(m, 1/2+it)} \nu_{\mathfrak{a}\mathfrak{c}}(n, 1/2+it)\, dt, \tag{3.17}$$

where the choice ϕ_+ versus ϕ_- depends on whether $mn > 0$ or $mn < 0$.

Here the integral transform for ϕ_h is given as

$$\phi_h(\ell) = \int_0^\infty J_{\ell-1}(x) \frac{4\pi\sqrt{mn}}{x} \phi\left(\frac{4\pi\sqrt{mn}}{x}\right) \frac{dx}{x} = (J_{\ell-1} * (x \cdot \phi))(4\pi\sqrt{mn}). \tag{3.18}$$

With

$$B_{2it}^{+}(x) = \frac{i}{2\sinh(\pi t)} (J_{2it}(x) - J_{-2it}(x)),$$

we have

$$\phi_+(t) = \int_0^\infty B_{2it}^{+}(x) \frac{4\pi\sqrt{mn}}{x} \phi\left(\frac{4\pi\sqrt{mn}}{x}\right) \frac{dx}{x} = (B_{2it}^{+} * (x \cdot \phi))(4\pi\sqrt{mn}). \tag{3.19}$$

Similarly, with

$$B_{2it}^{-}(x) = \frac{2}{\pi} \cosh(\pi t) K_{2it}(x),$$

we have

$$\phi_-(t) = \int_0^\infty B_{2it}^{-}(x) \frac{4\pi\sqrt{|mn|}}{x} \phi\left(\frac{4\pi\sqrt{|mn|}}{x}\right) \frac{dx}{x} = (B_{2it}^{-} * (x \cdot \phi))(4\pi\sqrt{|mn|}). \tag{3.20}$$

Remarks. Here we have implemented some corrections of [Iwa02] noted by Blomer, Harcos, and Michel [BHM07]. Moreover, the right hand side slightly differs from the formulas in [Iwa02] in that the roles of m and n are reversed, consistent with the remark following Definition 3.2.

It is important to emphasize that the same scaling matrices must occur in both the definition of the Kloosterman sum and the definition of the Fourier coefficients.

We occasionally use the above integral representations, but predominantly prefer Mellin-type integrals, and we next state those formulas. The integral transforms ϕ_h, ϕ_+ and ϕ_- are realized as convolutions on the group $\left(\mathbb{R}^+, \frac{dx}{x}\right)$ and therefore their Mellin transforms can be easily computed. Let $\widetilde{\phi}(s) = \int_0^\infty \phi(x) x^s \frac{dx}{x}$, which on occasion we alternatively write as $\mathcal{M}(\phi, s)$.

Proposition 3.6. *The integral transforms ϕ_h and $\phi_\pm$ have the alternative formulas*

$$\phi_h(\ell) = \frac{1}{2\pi i} \int_{(1)} \frac{2^{s-1}\Gamma\left(\frac{s+\ell-1}{2}\right)}{\Gamma\left(\frac{\ell+1-s}{2}\right)} \widetilde{\phi}(s+1)(4\pi\sqrt{mn})^{-s}\, ds, \tag{3.21}$$

$$\phi_\pm(t) = \frac{1}{2\pi i} \int_{(2)} h_\pm(s,t) \widetilde{\phi}(s+1)(4\pi\sqrt{mn})^{-s}\, ds, \tag{3.22}$$

where

$$h_\pm(s,t) = \begin{cases} \frac{1}{\pi} 2^{s-1} \cos(\pi s/2)\Gamma(s/2+it)\Gamma(s/2-it), & \pm = +, \\ \frac{1}{\pi} 2^{s-1} \cosh(\pi t)\Gamma(s/2+it)\Gamma(s/2-it), & \pm = -. \end{cases}$$

Proof. By Mellin inversion, we have

$$\phi_h(\ell) = \frac{1}{2\pi i} \int_{(1)} \mathcal{M}(J_{\ell-1} * (x \cdot \phi), s)(4\pi\sqrt{mn})^{-s}\, ds. \tag{3.23}$$

The Mellin transform satisfies the property $\mathcal{M}(f * g, s) = \mathcal{M}(f, s)\mathcal{M}(g, s)$. The Mellin transform of the J-Bessel function (see [EMOT54, 6.8(1)]) is given as

$$\int_0^\infty J_\nu(x) x^s \frac{dx}{x} = \frac{2^{s-1}\Gamma\left(\frac{s+\nu}{2}\right)}{\Gamma\left(\frac{\nu-s}{2}+1\right)}. \tag{3.24}$$

Also note that $\widetilde{x\phi}(s) = \widetilde{\phi}(s+1)$. Therefore (3.23) can be recast as (3.21), as desired.

For ϕ_- we know by [EMOT54, §6.8, (26)] that

$$\int_0^\infty K_{2it}(x) x^s \frac{dx}{x} = 2^{s-2}\Gamma(s/2+it)\Gamma(s/2-it), \tag{3.25}$$

and therefore we obtain the minus case of (3.22).

The plus case follows from using (3.24), the reflection formula for the gamma function, and the addition formulas for sine. ◻

3.4. Spectral large sieve

On $\Gamma_0(N)$, we normalize the Petersson inner product by

$$\langle g_1, g_2\rangle = \int_{\Gamma_0(N)\backslash\mathbb{H}} g_1(z)\overline{g_2}(z) y^{\kappa} \frac{dx\,dy}{y^2}. \tag{3.26}$$

Quoting from [BHM07], if u_j (g, respectively) is an L^2-normalized cuspidal Hecke–Maass (holomorphic, resp.) newform of level N with trivial nebentypus, then

$$|\nu_{\infty j}(1)|^2 = N^{-1}(N(1+|t_j|))^{o(1)} \quad \text{and} \quad |\nu_{\infty g}(1)|^2 = N^{-1}(Nk)^{o(1)}. \tag{3.27}$$

With the normalization (3.12), and assuming $\mathfrak{a}$ is an Atkin–Lehner cusp, the spectral large sieve inequalities give

$$\sum_{|t_j|\le T}\Big|\sum_{m\le M} a_m \nu_{\mathfrak{a}j}(m)\Big|^2 \ll \Big(T^2+\frac{M}{N}\Big)(MNT)^{\varepsilon}\sum_{m\le M}|a_m|^2,$$

and

$$\sum_{\mathfrak{c}}\int_{|t|\le T}\Big|\sum_{m\le M} a_m \nu_{\mathfrak{a}\mathfrak{c}}(m,1/2+it)\Big|^2 dt \ll \Big(T^2+\frac{M}{N}\Big)(MNT)^{\varepsilon}\sum_{m\le M}|a_m|^2, \tag{3.28}$$

and

$$\sum_{k\le T}\sum_{g\in H_k(N)}\Big|\sum_{m\le M} a_m \nu_{\mathfrak{a}g}(m)\Big|^2 \ll \Big(T^2+\frac{M}{N}\Big)(MNT)^{\varepsilon}\sum_{m\le M}|a_m|^2.$$

3.5. Newforms and oldforms

Atkin and Lehner showed the orthogonal decomposition

$$S_\kappa(N) = \bigoplus_{LM=N}\bigoplus_{f\in H_\kappa^*(M)} S_\kappa(L; f),$$

where $S_\kappa(L; f)$ is the span of the forms $f_{|\ell}$ with $\ell \mid L$, where

$$f_{|\ell}(z) = \ell^{\kappa/2} f(\ell z). \tag{3.29}$$

Their proof works with virtually no changes in the case of Maass forms (which have weight 0, in our context). For the rest of this section, we focus on the Maass case, but with a general weight κ (in order to most easily translate the results to the holomorphic case).

The formula (3.29) means that (let us agree to drop the subscript ∞ when working with the Fourier expansion at ∞)

$$\nu_{f|_\ell}(n) = \ell^{1/2}\nu_f(n/\ell). \tag{3.30}$$

Blomer and Milićević have shown in [BM15, Section 6] that there exists a basis of $S_\kappa(L; f)$ of the following type. Let f^* denote a newform of level $M \mid N$, L^2-normalized as a *level* N form, which implies $|\nu_{\infty f^*}(1)|^2 = N^{-1}(N(1+|t_j|))^{o(1)}$. Then there exists an orthonormal basis for $S_\kappa(L; f)$ of the form $g_m = \sum_{\ell | L} c_{\ell,m} f^*|_\ell$, where $c_{\ell,m} \ll N^\varepsilon$. For an Atkin–Lehner cusp $\mathfrak{a}$, we have $|\nu_{\mathfrak{a} f^*}(1)| = |\nu_{\infty f^*}(1)|$, by (3.5) .

We need the following information on the Fourier coefficients of $f^*|_\ell$ at Atkin–Lehner cusps:

Lemma 3.7. *Suppose $\mathfrak{a}$ is an Atkin–Lehner cusp of $\Gamma_0(N)$, and f^* is a newform of level M with $LM = N$. Then the set of lists of Fourier coefficients $\{(\nu_{\mathfrak{a} f^*|_\ell}(n))_{n\in\mathbb{N}} : \ell \mid L\}$ is, up to signs, the same as the set $\{(\nu_{\infty f^*|_\ell}(n))_{n\in\mathbb{N}} : \ell \mid L\}$.*

See [KY19, Lemma 2.5] for a proof.

It is crucial for our later purposes (specifically, in the proof of Lemma 11.1) to bound the Hecke eigenvalues of newforms at primes dividing the level. Let f^* be a newform (Maass or holomorphic) of level M, with trivial nebentypus, as above. If $p \mid M$, then

$$|\lambda_{f^*}(p)| \leq p^{-1/2} \tag{3.31}$$

(see [Li75, Theorem 3(iii)] or [AL70, Theorem 3(iii)]; the proofs carry over to Maass forms with virtually no changes).

4. Inert functions and oscillatory integrals

4.1. Basic definition

We begin with a class of functions defined by derivative bounds. Let $\mathcal{F}$ be any index set and $X : \mathcal{F} \to \mathbb{R}_{\geq 1}$ be a function of $\mathrm{T} \in \mathcal{F}$ with its value at T denoted by X_T.

Definition 4.1. A family $\{w_\mathrm{T}\}_{\mathrm{T}\in\mathcal{F}}$ of smooth functions supported on a product of dyadic intervals in $\mathbb{R}^d_{>0}$ is called X-*inert* if for each $j = (j_1, \ldots, j_d) \in \mathbb{Z}^d_{\geq 0}$ we have

$$C(j_1, \ldots, j_d) := \sup_{\mathrm{T}\in\mathcal{F}} \sup_{(x_1,\ldots,x_d)\in\mathbb{R}^d_{>0}} X_\mathrm{T}^{-j_1-\cdots-j_d} |x_1^{j_1} \ldots x_d^{j_d} w_\mathrm{T}^{(j_1,\ldots,j_d)}(x_1, \ldots, x_d)| < \infty. \tag{4.1}$$

In our desired applications, our family $\{w_\mathrm{T}\}_{\mathrm{T}\in\mathcal{F}}$ of inert functions will be indexed by tuples T of the form $\mathrm{T} = (M_1, M_2, N_1, N_2, N_3, C, a, \ldots)$, as well as some other parameters that arise as dual variables after Poisson summation. Each of these parameters is polynomially bounded in q. Furthermore, the relevant values of X will be at most $c(\varepsilon)q^\varepsilon$ for some constant $c(\varepsilon)$.

In addition, the weight functions encountered in this paper will typically be represented in the form

$$P(\mathrm{T})e^{i\phi(x_1,\ldots,x_d)} w_\mathrm{T}(x_1, \ldots, x_d), \tag{4.2}$$

where $P(\mathrm{T})$ is some simple function depending on the tuple T only, $\phi(x_1, \ldots, x_d)$ is the phase, and w_T is an inert function. We wish to understand how such a function behaves under Fourier and Mellin transformations. In Section 4.2, we analyze the Fourier and Mellin transforms in case $\phi = 0$, and in Section 4.3 we discuss the stationary phase analysis of the Fourier transform in the presence of a phase ϕ.

4.2. *Fourier and Mellin transforms*

Inert functions behave regularly under the Fourier transform. Suppose that $w_T(x_1, \dots, x_d)$ is X-inert, and let

$$\widehat{w_T}(t_1, x_2, \dots, x_d) = \int_{-\infty}^{\infty} w_T(x_1, \dots, x_d) e(-x_1 t_1)\, dx_1$$

denote its Fourier transform in the x_1 variable. Suppose that the support of w_T is such that $x_i \asymp X_i$ for each i. Now $\widehat{w_T}$ is not compactly supported in t_1, so it will not be inert. However, if we let $W_{T,Y_1}(t_1, x_2, \dots x_d) = w_{Y_1}(t_1)\widehat{w_T}(t_1, x_2, \dots x_d)$ where $\{w_{Y_1} : Y_1 > 0\}$ is a 1-inert family supported on $t_1 \asymp Y_1$ (or $-t_1 \asymp Y_1$) then $X_1^{-1} W_{T,Y_1}$ forms an X-inert family. Moreover, by repeated integration by parts we have

$$W_{T,Y_1}(t_1, x_2, \dots, x_d) \ll X_1 \left(1 + \frac{|t_1| X_1}{X}\right)^{-A} \asymp \left(1 + \frac{Y_1 X_1}{X}\right)^{-A},$$

so that in practice we may restrict our attention to $Y_1 \ll X q^{\varepsilon}/X_1$. See [KPY19] for more details.

A similar integration by parts argument also treats the Mellin transform, and we record the result as follows:

Lemma 4.2. *Let $w_T(x_1, \dots, x_d)$ be a family of X-inert functions such that x_1 is supported in the dyadic interval $[X_1, 2X_1]$. Let*

$$\widetilde{w}_T(s, x_2, \dots, x_d) = \int_0^{\infty} w_T(x, x_2, \dots, x_d) x^s \frac{dx}{x}.$$

Then we have $\widetilde{w}_T(s, x_2, \dots, x_d) = X_1^s W_T(s, x_2, \dots x_n)$ where $W_T(s, \cdot)$ is a family of X-inert functions in all the remaining x_i, which is entire in s and has rapid decay for $|\mathrm{Im}(s)| \gg X^{1+\varepsilon}$.

4.3. *Stationary phase*

Next we synthesize some results from [BKY13] and [KPY19].

Lemma 4.3. *Suppose that $w = w_T(t, t_2, \dots, t_d)$ is a family of X-inert functions supported on $t \asymp Z$, $t_i \asymp X_i$ for $i = 2, \dots, d$. Suppose that on the support of w_T, ϕ satisfies*

$$\frac{\partial^{a_1 + a_2 + \cdots + a_d}}{\partial t^{a_1} \dots \partial t_d^{a_d}} \phi(t, t_2, \dots, t_d) \ll \frac{Y}{Z^{a_1}} \frac{1}{X_2^{a_2} \dots X_d^{a_d}}, \tag{4.3}$$

i.e., $Y^{-1}\phi$ is 1-inert. Suppose that $Y/X^2 \gg q^{\delta}$ for some $\delta > 0$. Let

$$I = \int_{-\infty}^{\infty} w_T(t, t_2, \dots, t_d) e^{i\phi(t, t_2, \dots, t_d)}\, dt.$$

(1) *If $\left|\frac{\partial}{\partial t}\phi(t, t_2, \dots, t_d)\right| \gg Y/Z$ for all t in the support of w_T, then $I \ll_A q^{-A}$ for $A > 0$ arbitrarily large.*

(2) *If* $\frac{\partial^2}{\partial t^2}\phi(t, t_2, \dots, t_d) \gg Y/Z^2$ *for all* $t, t_2, \dots, t_d$ *in the support of* w*, and there exists* $t_0 \in \mathbb{R}$ *such that* $\phi'(t_0) = 0$ *(here,* ϕ' *denotes the derivative with respect to* t*, and note* t_0 *is necessarily unique), then*

$$I = \frac{Z}{\sqrt{Y}} e^{i\phi(t_0,t_2,\dots,t_d)} W_T(t_2, \dots, t_d) + O(q^{-A}) \tag{4.4}$$

for some X*-inert family of functions* W_T*.*

Part (1) above follows from [BKY13, Lemma 8.1]. The one-variable version of (2) above is contained in [BKY13, Proposition 8.2], which was improved to many variables in [KPY19].

4.4. *A convention*

We often renormalize a family of inert functions. For a simple example to illustrate this, say $w_{\mathrm{T}}(x)$ is X-inert, supported on $x \asymp N$. We can write $x^{-1/2} w_{\mathrm{T}}(x) = N^{-1/2} W_{\mathrm{T}}(x)$, where $W_{\mathrm{T}}(x) = (x/N)^{-1/2} w_{\mathrm{T}}(x)$. Then W_{T} forms a new X-inert family with a different list of constants $C(j)$. When doing this too many times it becomes difficult to find notation for all the new functions that arise, so we may on occasion replace W_{T} by w_{T}, which is supposed to represent a generic inert function.

Another useful convention is that, when focusing only a particular variable (say n), we may write $w_N(n, \cdot)$ where the $\cdot$ is a placeholder for the remaining variables. Writing all the variables is unwieldy, and the notion of inertness keeps track of the important behavior of the weight function with respect to the remaining variables.

We will also say that a family $\{w_{\mathrm{T}}(x_1, \dots, x_d)\}$ of inert functions such that each variable x_i is supported in $[X_i, 2X_i]$ is *very small* to mean a quantity which is of size $O_A((X_1 \dots X_d q)^{-A})$ for every $A > 0$, and uniformly in the family $\mathrm{T} \in \mathcal{F}$. More generally, we will use this terminology "very small" for more general quantities, not just inert functions. In practice, we will largely ignore very small error terms.

5. Preliminaries

5.1. *Petersson trace formula*

The Petersson formula reads

$$\sum_{f \in H_\kappa(q)} w_f \lambda_f(m) \lambda_f(n) = \delta_{n=m} + 2\pi i^{-\kappa} \sum_{c \equiv 0 \ (\mathrm{mod}\ q)} \frac{S(m, n; c)}{c} J_{\kappa-1}\left(\frac{4\pi\sqrt{mn}}{c}\right),$$

where $w_f = q^{-1+o(1)}$ are the Petersson weights. Define

$$\mathcal{M} = \mathcal{M}(q) = \sum_{f \in H_\kappa(q)} w_f L(1/2, f)^5.$$

Our main result, Theorem 1.1, is equivalent to

$$\mathcal{M} \ll_{\kappa,\varepsilon} q^{\theta+\varepsilon} \tag{5.1}$$

for any $\varepsilon > 0$.

5.2. The approximate functional equations

Let κ be a positive even integer, q a prime, and f a Hecke cusp form of weight κ and level q. Put

$$\gamma(s,\kappa) = \pi^{-s}\Gamma\left(\frac{s+\frac{\kappa-1}{2}}{2}\right)\Gamma\left(\frac{s+\frac{\kappa+1}{2}}{2}\right).$$

Let G_i $(i = 1, 2)$ be an even entire function decaying rapidly in vertical strips such that $G_i(0) = 1$. Define

$$V_1(x) = \frac{1}{2\pi i}\int_{(1)} \frac{G_1(u)}{u}\,\frac{\gamma(1/2+u,\kappa)}{\gamma(1/2,\kappa)}x^{-u}\,du,$$
$$V_2(x) = \frac{1}{2\pi i}\int_{(1)} \frac{G_2(u)}{u}\,\frac{\gamma(1/2+u,\kappa)^2}{\gamma(1/2,\kappa)^2}x^{-u}\,du.$$

If $x \gg q^{\varepsilon}$ then by shifting the contour of integration arbitrarily far to the right, we obtain $V_i(x) \ll_{\kappa,A} (xq)^{-A}$. Here and throughout, we view κ as fixed, and q as becoming large. For later use, it will be important to assume $G_i(1/2) = 0$.

Proposition 5.1. *With notation as above, we have*

$$L(1/2, f)^2 = 2\sum_{m=1}^{\infty} \frac{\lambda_f(m)\tau_2(m)}{\sqrt{m}} V\left(\frac{m}{q}\right),$$

where $\tau_2(m)$ is the (two-fold) divisor function, and

$$V(x) = \sum_{(e,q)=1}^{\infty} V_2(e^2x)/e = \frac{1}{2\pi i}\int_{(1)} \widetilde{V_2}(u)\zeta_q(1+2u)x^{-u}\,du,$$

where $\zeta_q(s) = (1-q^{-s})\zeta(s)$ is the Riemann zeta function with the q^{th} Euler factor missing.

Proof. By the Hecke relation, we have

$$L(s, f)^2 = \sum_{m_1,m_2=1}^{\infty} \sum_{\substack{e|(m_1,m_2)\\(e,q)=1}} \frac{\lambda_f(m_1m_2/e^2)}{(m_1m_2)^s} = \sum_{(e,q)=1}^{\infty} \frac{1}{e^{2s}} \sum_{m=1}^{\infty} \frac{\tau_2(m)\lambda_f(m)}{m^s}. \tag{5.2}$$

Then from the functional equation $L(s, f)^2\gamma(s,\kappa)^2q^s =: \Lambda(s, f)^2 = \Lambda(1-s, f)^2$ we get the formula, as in [IK04, Theorem 5.3]. □

Proposition 5.2. *Let ε_f be the sign of the functional equation for $L(s, f)$. Then*

$$L(1/2, f)^3 = (1+\varepsilon_f)^3 \sum_{\substack{a=1\\(a,q)=1}}^{\infty} \frac{\mu(a)}{a^{3/2}} \sum_{n=1}^{\infty} \frac{\lambda_f(na)}{\sqrt{n}}\tau_3(n, F_{a,\sqrt{q}}), \tag{5.3}$$

where

$$\tau_3(n, F_{a,\sqrt{q}}) = \sum_{n_1 n_2 n_3 = n} F_a\left(\frac{n_1}{\sqrt{q}}, \frac{n_2}{\sqrt{q}}, \frac{n_3}{\sqrt{q}}\right), \tag{5.4}$$

and

$$\begin{aligned} F_a(x_1, x_2, x_3) &= \sum_{\substack{e_1, e_2, e_3 \\ (e_1 e_2 e_3, q) = 1}} \frac{1}{e_1 e_2 e_3} V_1(a x_1 e_1 e_2) V_1(a x_2 e_1 e_3) V_1(a x_3 e_2 e_3) \\ &= \iiint_{(1)(1)(1)} \prod_{i=1}^{3} \frac{\gamma(1/2 + u_i, \kappa) G(u_i)}{(a x_i)^{u_i} \gamma(1/2, \kappa) u_i} \zeta_q(1 + u_1 + u_2) \zeta_q(1 + u_1 + u_3) \zeta_q(1 + u_2 + u_3) \\ &\qquad \times \frac{du_1\, du_2\, du_3}{(2\pi i)^3}. \end{aligned} \tag{5.5}$$

Remark. One may easily check that

$$x_1^{j_1} x_2^{j_2} x_3^{j_3} \frac{\partial^{j_1 + j_2 + j_3}}{\partial x_1^{j_1} \partial x_2^{j_2} \partial x_3^{j_3}} F_a(x_1, x_2, x_3) \ll_{j_1, j_2, j_3, A, \varepsilon} \prod_{i=1}^{3} (a x_i)^{-\varepsilon} (1 + a x_i)^{-A}. \tag{5.6}$$

In the terminology introduced later in Section 4, the property (5.6) means that F_a satisfies the same derivative bounds as an X-inert function with $X \ll q^{\varepsilon}$, in the region $x_i \gg q^{-1/2}$, for all i. Similar derivative bounds hold for $V(x)$.

Proof of Proposition 5.2. Using the approximate functional equation for each $L(1/2, f)$, and the Hecke relations, we obtain

$$\begin{aligned} L(1/2, f)^3 &= \sum_{\substack{e_1, e_2 \\ (e_1 e_2, q) = 1}} \frac{(1+\varepsilon_f)^3}{e_1} \sum_{\substack{n_1, n_2, n_3 \\ e_2 | (n_1 n_2, n_3)}} \frac{\lambda_f\left(\frac{n_1 n_2 n_3}{e_2^2}\right)}{\sqrt{n_1 n_2 n_3}} V_1\left(\frac{n_1 e_1}{\sqrt{q}}\right) V_1\left(\frac{n_2 e_1}{\sqrt{q}}\right) V_1\left(\frac{n_3}{\sqrt{q}}\right) \\ &= \sum_{\substack{e_1, e_2 \\ (e_1 e_2, q) = 1}} \frac{(1+\varepsilon_f)^3}{e_1 e_2} \\ &\quad \times \sum_{f_1 f_2 = e_2} \sum_{\substack{n_1, n_2, n_3 \\ (n_1, f_2) = 1}} \frac{\lambda_f(n_1 n_2 n_3)}{\sqrt{n_1 n_2 n_3}} V_1\left(\frac{n_1 e_1 f_1}{\sqrt{q}}\right) V_1\left(\frac{n_2 e_1 f_2}{\sqrt{q}}\right) V_1\left(\frac{n_3 f_1 f_2}{\sqrt{q}}\right). \end{aligned}$$

Using Möbius inversion to detect the coprimality condition with $\sum_{a | (n_1, f_2)} \mu(a)$, reordering the summations, and renaming the summation variables gives the more symmetric form

$$\begin{aligned} L(1/2, f)^3 &= (1 + \epsilon_f)^3 \sum_{\substack{a=1 \\ (a,q)=1}}^{\infty} \frac{\mu(a)}{a^{3/2}} \sum_{\substack{e_1, e_2, e_3 \\ (e_1 e_2 e_3, q) = 1}} \frac{1}{e_1 e_2 e_3} \\ &\quad \times \sum_{n_1, n_2, n_3} \frac{\lambda_f(a n_1 n_2 n_3)}{\sqrt{n_1 n_2 n_3}} V_1\left(\frac{a n_1 e_1 e_2}{\sqrt{q}}\right) V_1\left(\frac{a n_2 e_1 e_3}{\sqrt{q}}\right) V_1\left(\frac{a n_3 e_2 e_3}{\sqrt{q}}\right). \end{aligned}$$

This is seen to be equivalent to (5.3). □

Now apply Propositions 5.1 and 5.2 to $\mathcal{M}$. There is a significant simplification in Proposition 5.2, whereby we may replace $1+\epsilon_f$ by 2, because if $\epsilon_f = -1$, then $L(1/2, f)^2 = 0$ anyway. Applying the Petersson trace formula then yields

$$\tfrac{1}{16}\mathcal{M} = \mathcal{D} + 2\pi i^{-\kappa}\mathcal{S},$$

where $\mathcal{D}$ is the diagonal term, and

$$\mathcal{S} = \sum_{(a,q)=1}^{\infty} \frac{\mu(a)}{a^{3/2}} \sum_{c\equiv 0 \,(\mathrm{mod}\, q)} \sum_{n,m} \frac{\tau_2(m)\tau_3(n, F_{a,\sqrt{q}})S(m, na; c)}{c\sqrt{mn}} J_{\kappa-1}\left(\frac{4\pi\sqrt{mna}}{c}\right) V\left(\frac{m}{q}\right). \tag{5.7}$$

It is easy to bound the diagonal term.

Lemma 5.3. *For any $\varepsilon > 0$ we have $\mathcal{D} \ll_\varepsilon q^\varepsilon$.*

This follows easily from the fact that the functions $V_1(y)$ and $V_2(y)$ decay rapidly as $y \to \infty$, and using the bound $J_{\kappa-1}(x) \ll x$ for $\kappa \geq 2$.

Proving Theorem 1.1 is reduced to showing $\mathcal{S} \ll q^{\theta+\varepsilon}$.

6. First Poisson summation

We continue the analysis of $\mathcal{S}$ from (5.7). We open up the divisor functions using the formulas $\sum_m \tau_2(m) f(m) = \sum_{m_1,m_2} f(m_1 m_2)$, and the definition of $\tau_3(n, F_{a,\sqrt{q}})$ (see (5.4)).

6.1. *Dyadic partition of unity*

Throughout this paper we will apply dyadic decompositions of important variables. Call a number N *dyadic* if $N = 2^{k/2}$ for some integer k. A *dyadic partition of unity* is a partition of unity of the form $\sum_{k\in\mathbb{Z}} \omega(2^{-k/2}x) \equiv 1$ for $x > 0$, where ω is a fixed smooth function with support on the dyadic interval $[1, 2]$. The family $\omega_N(x) = \omega(x/N)$ forms a 1-inert family of functions. Applying this to $\mathcal{S}$, we have

$$\mathcal{S} = \sum_{M_1,M_2,N_1,N_2,N_3,C \text{ dyadic}} \mathcal{S}_{M_1,M_2,N_1,N_2,N_3,C}, \tag{6.1}$$

where the dyadic numbers are restricted to be $\geq 2^{-1/2}$, and where

$$\mathcal{S}_{M_1,M_2,N_1,N_2,N_3,C} = \sum_{(a,q)=1}^{\infty} \frac{\mu(a)}{a^{3/2}} \sum_{\substack{c\equiv 0 \,(\mathrm{mod}\, q)\\ c\asymp C}} \frac{1}{c} \sum_{n_1,n_2,n_3,m_1,m_2} \frac{S(m_1m_2, n_1n_2n_3a, c)}{\sqrt{m_1m_2n_1n_2n_3}}$$
$$\times J_{\kappa-1}\left(\frac{4\pi\sqrt{m_1m_2n_1n_2n_3a}}{c}\right) V\left(\frac{m_1m_2}{q}\right) F_a\left(\frac{n_1}{\sqrt{q}}, \frac{n_2}{\sqrt{q}}, \frac{n_3}{\sqrt{q}}\right) w_T(m_1, m_2, n_1, n_2, n_3, c). \tag{6.2}$$

The letter T here and throughout stands for the tuple of dyadic parameters, and we may use $\mathcal{S}_{\mathrm{T}}$ as shorthand for the left hand side of (6.2). For the main thrust of the argument, the precise form of w_{T} is not important. However, when calculating certain potential main terms, we have found it important to re-sum over the partition, in which case one should remember that w_{T} may be expressed as

$$w_{\mathrm{T}}(m_1, m_2, n_1, n_2, n_3, c) = \omega\left(\frac{m_1}{M_1}\right) \dots \omega\left(\frac{c}{C}\right).$$

Let $M = M_1 M_2$ and $N = N_1 N_2 N_3$.

Lemma 6.1. *Let $\varepsilon > 0$. Unless*

$$M \ll_\varepsilon q^{1+\varepsilon} \quad \text{and} \quad N_i \ll_\varepsilon \frac{q^{1/2+\varepsilon}}{a} \tag{6.3}$$

for all $i = 1, 2, 3$, we have

$$\mathcal{S}_{\mathrm{T}} \ll_A q^{-A}$$

for $A > 0$ arbitrarily large. Moreover, if $C > q^3$, we have

$$\mathcal{S}_{\mathrm{T}} \ll_\varepsilon q^\varepsilon.$$

We will henceforth assume (6.3) (which implies $N \ll a^{-3} q^{3/2+\varepsilon}$), and

$$C \le q^3. \tag{6.4}$$

Proof of Lemma 6.1. The bounds (6.3) follow from the rapid decay of the weight functions in the approximate functional equations. The bound for $C > q^3$ holds using the Weil bound for Kloosterman sums, and $J_{\kappa-1}(x) \ll x$. □

By symmetry (Dirichlet's hyperbola method), we shall assume

$$M_1 \le M_2. \tag{6.5}$$

Note $M_1 \ll q^{1/2+\varepsilon}$.

6.2. *Poisson summation*

Applying Poisson summation in m_2 modulo c, we obtain

$$\mathcal{S}_{\mathrm{T}} = \sum_{(a,q)=1} \frac{\mu(a)}{a^{3/2}} \sum_{m_1, n_1, n_2, n_3} \frac{1}{\sqrt{m_1 n_1 n_2 n_3}} \sum_{c \equiv 0 \,(\mathrm{mod}\, q)} \frac{1}{c^2} \sum_{k \in \mathbb{Z}} H(k) I(k), \tag{6.6}$$

where (with shorthand $n = n_1 n_2 n_3$)

$$H(k) = H(k, m_1, na; c) = \sum_{x \,(\mathrm{mod}\, c)} S(m_1 x, na; c) e\left(\frac{kx}{c}\right), \tag{6.7}$$

and

$$I(k)=I(m_1,k,n_1,n_2,n_3,a,c)=\int_0^{\infty} e\left(\frac{-kt}{c}\right)J_{\kappa-1}\left(\frac{4\pi\sqrt{m_1nat}}{c}\right)w_{M_2}(t,\cdot)\,\frac{dt}{\sqrt{t}}. \tag{6.8}$$

For the notation $w_{M_2}(t,\cdot)$, recall the convention described in Section 4.4. Furthermore note that the definition of the inert function w_{T} has been altered to include the function V and F_a from the second line of (6.2).

We now apply a dyadic partition of unity to the k-sum, and let $\omega(k/K)$ be a generic such piece. To simplify the notation, we simply add K to the long tuple of parameters already appearing in (6.2); we are already writing T as shorthand for this long tuple, and we shall continue this practice. Let $I_K = I_{(M_1,M_2,N_1,N_2,N_3,a,C,K)} = \omega(k/K)I$. Then for $k>0$, $I=\sum_{K \text{ dyadic}} I_K$. By redefining the inert function in (6.8) to incorporate $\omega(k/K)$, we may also view I_K as an instance of (6.8).

Remark. We may without loss of generality assume that $k>0$. The negative values of k give rise to terms that are complex conjugates of their positive counterparts. Secondly, $H(0)=0$ unless $c\mid m_1$, and those terms only contribute $O(q^{-A})$ since $m_1 \ll q^{1/2+\varepsilon}$ and $q\mid c$.

6.3. *The arithmetic function*

We now compute the arithmetic sum $H(k)$. Immediately from the definition, we obtain

$$\frac{1}{c}H(k)=\frac{1}{c}\sum_{u \,(\mathrm{mod}\, c)}^{*}\sum_{x \,(\mathrm{mod}\, c)} e\left(\frac{(m_1u+k)x+na\overline{u}}{c}\right)=\sum_{u\,(\mathrm{mod}\, c)}^{*}\delta_{m_1u\equiv -k \,(\mathrm{mod}\, c)}\,e\left(\frac{na\overline{u}}{c}\right). \tag{6.9}$$

One would like to simply substitute $u\equiv -k\overline{m_1} \pmod c$, but this is not possible because it is not guaranteed that m_1 (or k) is coprime to c. For this reason, we employ a factorization of c and the Chinese remainder theorem as follows.

Write

$$c=c_0c_2 \quad\text{and}\quad k=k_0k_1, \tag{6.10}$$

where the factorizations may be written locally, using the notation $\nu_p(n)=d$ for $p^d \,\|\, n$, as

$$c_0=\prod_{\nu_p(c)>\nu_p(k)} p^{\nu_p(c)},\qquad c_2=\prod_{1\le \nu_p(c)\le \nu_p(k)} p^{\nu_p(c)},$$
$$k_0=\prod_{\nu_p(k)\ge \nu_p(c)} p^{\nu_p(k)},\qquad k_1=\prod_{1\le \nu_p(k)<\nu_p(c)} p^{\nu_p(k)}.$$

Alternatively, using the notation $n^*=\prod_{p\mid n} p$ we have

$$(c_0,k_0)=1,\qquad c_2\mid k_0,\qquad k_1k_1^*\mid c_0, \tag{6.11}$$

and these conditions characterize c_0, c_2, k_0, k_1. Note that $(c_2, k_1) = 1$ automatically from the other conditions; indeed we also have $(c_2, c_0) = (k_1, k_0) = 1$.

The congruence condition $m_1 u \equiv -k \pmod{c}$ in (6.9) is solvable with $(u, c) = 1$ if and only if $(m_1, c) = (k, c)$. The conditions (6.10), (6.11) give $(k, c) = k_1 c_2$, and so we impose the condition $k_1 c_2 = (m_1, c_0 c_2) = \big(m_1, \frac{c_0}{k_1} k_1 c_2\big)$. Thus we define

$$m_1 = k_1 c_2 m_1', \tag{6.12}$$

where the new variable m_1' is only subject to the restriction

$$(m_1', c_0/k_1) = 1, \quad \text{i.e.} \quad (m_1', c_0) = 1,$$

where we have used the fact that c_0/k_1 has the same prime factors as c_0.

Remark. In $\mathcal{S}$, we have $q \mid c$. If $q \mid k_1 c_2$, this means $q \mid m_1$, but we have $m_1 \ll q^{1/2+\varepsilon}$, so the condition $q \mid c$ may be freely replaced with $q \mid c_0$, and we may assume

$$(q, k_1) = 1. \tag{6.13}$$

Proposition 6.2. *Given the notation above,*

$$\frac{1}{c} H(k, m_1, an; c) = e\left(-\frac{nam_1'\overline{k_0}}{c_0}\right) S(na, 0; c_2) k_1 \delta_{k_1 \mid na} \delta_{(m_1', c_0)=1} \delta_{c,k}, \tag{6.14}$$

where $\overline{k_0}$ indicates the multiplicative inverse modulo c_0 and where $\delta_{c,k} = 1$ if (6.10) and (6.11) hold, and $\delta_{c,k} = 0$ otherwise.

Proof. First, we note that $m_1 u \equiv -k \pmod{c}$ (that is, $m_1' k_1 c_2 u \equiv -k_0 k_1 \pmod{c_0 c_2}$) is equivalent to $m_1' u \equiv -k_0/c_2 \pmod{c_0/k_1}$. In other words,

$$\overline{u} \equiv -m_1' \, \overline{k_0/c_2} \pmod{c_0/k_1}. \tag{6.15}$$

Here $\overline{k_0/c_2}$ can be taken to be the multiplicative inverse modulo c_0, since every prime that divides c_0 also divides c_0/k_1 (via (6.11))

Now we apply the Chinese remainder theorem to the pair c_0 and c_2, which gives

$$\begin{aligned} \frac{1}{c} H(k, m_1, an; c) &= \sideset{}{^*}\sum_{u \,(\mathrm{mod}\, c)} \delta_{\overline{u} \equiv -m_1' \overline{k_0/c_2} \,(\mathrm{mod}\, c_0/k_1)} e\left(\frac{na\overline{u}(c_0\overline{c_0} + c_2\overline{c_2})}{c_0 c_2}\right) \\ &= \sideset{}{^*}\sum_{u \,(\mathrm{mod}\, c_2)} e\left(\frac{na\overline{u c_0}}{c_2}\right) \sideset{}{^*}\sum_{u \,(\mathrm{mod}\, c_0)} \delta_{\overline{u} \equiv -m_1' \overline{k_0/c_2} \,(\mathrm{mod}\, c_0/k_1)} e\left(\frac{na\overline{u c_2}}{c_0}\right). \end{aligned}$$

The sum modulo c_2 is a Ramanujan sum, and for the sum modulo c_0 we replace u by $\overline{u}$, which gives

$$\frac{1}{c} H(k, m_1, an; c) = S(na, 0; c_2) \sideset{}{^*}\sum_{u \,(\mathrm{mod}\, c_0)} \delta_{u \equiv -m_1' \overline{(k_0/c_2)} \,(\mathrm{mod}\, c_0/k_1)} e\left(\frac{nau\overline{c_2}}{c_0}\right).$$

The congruence restriction on u modulo c_0 may be expressed as

$$u \equiv -m_1' \overline{(k_0/c_2)} + v \frac{c_0}{k_1} \pmod{c_0} \quad \text{with} \quad v \pmod{k_1}. \tag{6.16}$$

Here v runs over *all* residue classes modulo k_1, because as long as u is coprime to c_0/k_1 it is also coprime to c_0. Thus

$$\begin{aligned}\frac{1}{c} H(k, m_1, an; c) &= S(na, 0; c_2) \sum_{v \pmod{k_1}} e\left(\frac{na(-m_1' \overline{(k_0/c_2)} + vc_0/k_1)\overline{c_2}}{c_0}\right) \\ &= S(na, 0; c_2) e\left(-\frac{nam_1' \overline{k_0}}{c_0}\right) k_1 \delta_{k_1 | na}. \end{aligned}$$

□

Inserting the conclusion of Proposition 6.2 into (6.6), and imposing (6.13), we get

$$\begin{aligned}\mathcal{S}_{\mathrm{T}} = \sum_{(a,q)=1} \frac{\mu(a)}{a^{3/2}} \sum_{m_1', n_1, n_2, n_3} \frac{1}{\sqrt{m_1' n_1 n_2 n_3}} \sum_{\substack{(c_0, m_1')=1 \\ c_0 \equiv 0 \pmod q}} \sum_{(k_0, c_0)=1} \sum_{\substack{k_1 | c_0/k_1^* \\ (k_1, q)=1}} k_1^{1/2} \delta_{k_1 | na} \\ \times \sum_{c_2 | k_0} \frac{1}{c_0 c_2^{3/2}} e\left(-\frac{n_1 n_2 n_3 a m_1' \overline{k_0}}{c_0}\right) S(n_1 n_2 n_3 a, 0; c_2) I(m_1' k_1 c_2, k_0 k_1, n_1, n_2, n_3, a, c_0 c_2)\end{aligned} \tag{6.17}$$

plus a small error term.

6.4. *Analysis of integral transform*

The asymptotic behavior of I_K depends on whether $\sqrt{aMN}/C \gg q^{\varepsilon}$ or not, since this dictates whether the Bessel function is oscillatory or not.

Lemma 6.3 (Pre-transition). *Let $I_K(k)$ be defined via (6.8). If for any $\varepsilon > 0$,*

$$\sqrt{aMN}/C \ll_{\varepsilon} q^{\varepsilon}, \tag{6.18}$$

then

$$M_2^{1/2} I_K(k) = (\sqrt{aMN}/C)^{\kappa - 1} M_2 w_{\mathrm{T}}(m_1, k, n_1, n_2, n_3, a, c), \tag{6.19}$$

where $w_{\mathrm{T}}(\cdot)$ is a $q^{\varepsilon'}$-inert function. Furthermore, I_K is very small unless

$$KM_2/C \ll q^{\varepsilon}. \tag{6.20}$$

Lemma 6.4 (Post-transition). *If for any $\varepsilon > 0$,*

$$\sqrt{aMN}/C \gg_{\varepsilon} q^{\varepsilon}, \tag{6.21}$$

then

$$M_2^{1/2} I_K(k) = \frac{CM_2}{(aMN)^{1/2}} e\left(\frac{m_1 na}{ck}\right) w_{\mathrm{T}}(m_1, k, n_1, n_2, n_3, a, c) + O((kq)^{-A}), \tag{6.22}$$

where $w_{\mathrm{T}}(\cdot)$ is a $q^{\varepsilon'}$-inert function. Furthermore, I_K is very small unless

$$K \asymp (aMN)^{1/2}/M_2. \tag{6.23}$$

Proof of Lemma 6.3. Suppose that (6.18) holds. Then the Bessel function is not oscillatory, and $J_{\kappa-1}(x) = x^{\kappa-1}W(x)$ where $x^j \frac{d^j}{dx^j} W(x) \ll X^j$ with $X \ll q^\varepsilon$ (note that if $1 \ll x \ll q^\varepsilon$, it is still valid to factor out $x^{\kappa-1}$ though there is a small loss of efficiency by doing so). This is the same derivative bound as for an X-inert function, so it may be absorbed into the inert function w_{T}. Then by the discussion in Sections 4.2 and 4.4, we have

$$M_2^{1/2} I_K(k) = (\sqrt{aMN}/C)^{\kappa-1} M_2 w_{\mathrm{T}}(\cdot), \tag{6.24}$$

and $I_K(k) \ll (kq)^{-A}$ if $K \gg \frac{C}{M_2} q^\varepsilon$. Here w_{T} is $q^{\varepsilon'}$-inert. □

Proof of Lemma 6.4. Now suppose that (6.21) holds. Then we use the fact that for $x \gg 1$,

$$J_{\kappa-1}(x) = \sum_{\pm} x^{-1/2} e^{\pm ix} W_{\pm}(x),$$

where $W_\pm$ satisfies the same derivative bounds as a 1-inert function. Thus

$$\sqrt{M_2} I_K(k) = \sum_{\pm} \frac{C^{1/2}}{(aMN)^{1/4}} \int_{-\infty}^{\infty} w_{M_2}(t, \cdot) e\left(\frac{-kt}{c}\right) e\left(\frac{\pm 2\sqrt{tm_1na}}{c}\right) dt,$$

where $w_{\mathrm{T}}(t)$ is q^ε-inert (in all previously declared variables), and supported on $t \asymp M_2$.

Since $k > 0$, if the $\pm$ sign is $-$, then Lemma 4.3(1) shows that the integral is very small. Therefore, we focus on the case where the sign is $+$, in which case we obtain an oscillatory integral with phase

$$\phi(t) = -\frac{kt}{c} + \frac{2\sqrt{tm_1na}}{c}.$$

We have

$$\phi'(t) = -\frac{k}{c} + \frac{\sqrt{m_1na}}{c\sqrt{t}}, \qquad \phi''(t) = -\frac{\sqrt{m_1na}}{2ct^{3/2}}.$$

There is a unique point t_0 where $\phi'(t_0) = 0$, namely

$$t_0 = \frac{m_1na}{k^2}.$$

If it is not the case that $t_0 \asymp M_2$ (with large but absolute implied constants), then we have $|\phi'(t)| \gg \frac{\sqrt{aMN}}{cM_1}$ throughout the support of the weight function, and Lemma 4.3(1) again shows the integral is small. If $t_0 \asymp M_2$, then the location of t_0 is compatible with the support of ϕ, and stationary phase (Lemma 4.3(2)) shows that

$$\int_{-\infty}^{\infty} w_{\mathrm{T}}(t) e\left(\frac{-kt}{c}\right) e\left(\frac{2\sqrt{tm_1na}}{c}\right) dt \\ = \frac{C^{1/2} M_2}{(aMN)^{1/4}} e\left(\frac{m_1na}{ck}\right) w_{\mathrm{T}}\left(\frac{m_1na}{k^2}, \cdot\right) + O((kq)^{-A}),$$

where w_{T} on the right hand side is q^ε-inert, and supported on $m_1na/k^2 \asymp M_2$. □

7. Reciprocity and other arithmetical manipulations

Next we reorder the summation $\mathcal{S}_{\mathrm{T}}$ in (6.17). We bring the sum over $n = n_1n_2n_3$ to the inside, and open up the Ramanujan sum $S(na, 0; c_2) = \sum_{d|(na,c_2)} d\mu(c_2/d)$. This gives

$$\mathcal{S}_{\mathrm{T}} = \sum_{(a,q)=1} \frac{\mu(a)}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d\mu(c_2/d) \sum_{(k_1,q)=1} k_1^{1/2} \sum_{m_1'} \frac{\mathcal{S}'}{\sqrt{m_1'}} + O(q^{-A}), \tag{7.1}$$

where

$$\mathcal{S}' = \sum_{\substack{(c_0,m_1')=1\\ c_0\equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0,c_0)=1\\ k_0\equiv 0\,(\mathrm{mod}\, c_2)}} \sum_{\substack{n_1n_2n_3a\equiv 0\,(\mathrm{mod}\, k_1)\\ n_1n_2n_3a\equiv 0\,(\mathrm{mod}\, d)}} \frac{e\left(-\frac{n_1n_2n_3am_1'\overline{k_0}}{c_0}\right)}{\sqrt{n_1n_2n_3}} I_K(m_1'k_1c_2, k_0k_1, n_1, n_2, n_3, a, c_0c_2).$$

We shall not obtain any significant cancellation in the outer summation variables appearing in $\mathcal{S}$ (except for a "fake" main term, in Section 13.7) , but substantial cancellation is required in c_0, k_0, and the n_i.

Note that since $d \mid c_2$, $c_2 \mid k_0$, $k_1 \mid c_0$, and $(c_0, k_0) = 1$, we have $(d, k_1) = 1$. Then the congruences in the sum over $n = n_1n_2n_3$ are equivalent to $an \equiv 0 \pmod{dk_1}$, which in turn is equivalent to $n \equiv 0 \pmod{\delta_1}$, where

$$\delta_1 = \frac{k_1d}{(a, k_1d)}. \tag{7.2}$$

Note that $(\delta_1, q) = 1$ and $(k_0, q) = 1$.

Since $(c_0, k_0) = 1$, we have the reciprocity formula

$$-\frac{\overline{k_0}}{c_0} \equiv \frac{\overline{c_0}}{k_0} - \frac{1}{c_0k_0} \pmod 1.$$

Define

$$\begin{aligned} J(n_1, n_2, n_3, a, m_1', c_0, k_0, c_2, k_1) &= e\left(-\frac{nam_1'}{c_0k_0}\right) I_K(m_1'k_1c_2, k_0k_1, n_1, n_2, n_3, a, c_0c_2)\\ &= e\left(-\frac{nam_1}{ck}\right) I_K(m_1, k, n_1, n_2, n_3, a, c), \end{aligned} \tag{7.3}$$

where in the second line we have expressed J in terms of the earlier variable names (6.10), (6.12). This is sometimes convenient for tracking the sizes of certain quantities, for example, $\frac{nam_1'}{c_0k_0} \asymp \frac{NaM_1}{CK}$.

Our next goal is to apply Poisson summation in the n variables, and to do that we need some preparatory moves. First, consider a formal sum of the form

$$\sum_{\substack{\eta_1,\eta_2,\eta_3\geq 1\\ \eta_1\eta_2\eta_3\equiv 0\ (\mathrm{mod}\ r)}} J(\eta_1,\eta_2,\eta_3). \tag{7.4}$$

The product $\eta_1\eta_2\eta_3$ runs over integers of the form ηr with $\eta \geq 1$. Now define

$$r_1 = (\eta_1, r), \quad \eta_1 = \eta_1' r_1,$$

so $(\eta_1', r/r_1) = 1$. Then $\eta_1'\eta_2\eta_3 = (r/r_1)\eta$. Continuing this process, define $r_2 = (\eta_2, r/r_1)$, $\eta_2 = \eta_2' r_2$, so $\eta_1'\eta_2'\eta_3 = \frac{r}{r_1 r_2} n$ with $\left(\eta_2', \frac{r}{r_1r_2}\right) = 1$. Finally, let $r_3 = \left(\eta_3, \frac{r}{r_1 r_2}\right)$, and set $\eta_3 = \eta_3' r_3$, whence $\left(\eta_3', \frac{r}{r_1r_2r_3}\right) = 1$. Now $\eta_1'\eta_2'\eta_3' = \frac{r}{r_1r_2r_3}\eta$, and the coprimality conditions mean that $\left(\eta_1'\eta_2'\eta_3', \frac{r}{r_1r_2r_3}\right) = 1$, so $r_1r_2r_3 = r$.

Therefore, translating this discussion into formulas, we find that (7.4) equals

$$\sum_{r_1r_2r_3=r}\ \sum_{(\eta_1', r_2r_3)=1}\ \sum_{(\eta_2', r_3)=1}\ \sum_{\eta_3'} J(r_1\eta_1', r_2\eta_2', r_3\eta_3').$$

Using Möbius inversion, we see that (7.4) equals

$$\sum_{r_1r_2r_3=r}\ \sum_{e_1|r_2r_3}\mu(e_1)\sum_{e_2|r_3}\mu(e_2)\sum_{n_1,n_2,n_3\geq 1} J(r_1e_1n_1, r_2e_2n_2, r_3n_3). \tag{7.5}$$

Applying this formula to $\mathcal{S}'$, we obtain

$$\mathcal{S}' = \sum_{r_1r_2r_3=\delta_1}\ \sum_{e_1|r_2r_3}\ \sum_{e_2|r_3}\mu(e_1)\mu(e_2)\mathcal{S}'', \tag{7.6}$$

where

$$\mathcal{S}'' = \sum_{\substack{(c_0,m_1')=1\\ c_0\equiv 0\ (\mathrm{mod}\ qk_1k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0,c_0)=1\\ k_0\equiv 0\ (\mathrm{mod}\ c_2)}} \sum_{n_1,n_2,n_3\geq 1} \frac{e\left(\frac{e_1e_2\delta_1 a m_1' n_1n_2n_3\overline{c_0}}{k_0}\right)}{\sqrt{\delta_1 e_1e_2n_1n_2n_3}} \times J(r_1e_1n_1, r_2e_2n_2, r_3n_3, a, m_1', c_0, k_0, c_2, k_1).$$

We remark that in doing so we changed variables by

$$n_1 \mapsto r_1e_1n_1, \quad n_2 \mapsto r_2e_2n_2, \quad n_3\mapsto r_3n_3. \tag{7.7}$$

With the earlier definition of n as $n_1n_2n_3$, (7.7) is equivalent to $n \mapsto e_1e_2\delta_1 n$.

Next define $g_0 = (e_1e_2\delta_1 a m_1', k_0)$, and write

$$k_0 = g_0 k_0' \quad \text{and} \quad \delta_2 = \frac{e_1e_2\delta_1 a m_1'}{g_0}.$$

There are some implicit conditions on the variables that we wish to record explicitly. Note that since $(k_1, k_0) = 1$ and $d \mid c_2 \mid k_0$, we may write g_0 as

$$g_0 = \left(e_1e_2\frac{ad}{(a,d)}m_1', k_0\right) = d\left(e_1e_2\frac{a}{(a,d)}m_1', \frac{k_0}{d}\right),$$

and in particular $d \mid g_0$, a property that will be important in Section 11.7. Also note that since none of the factors of δ_2 is divisible by q (since q is prime, $(a, q) = 1$, and the original m_1 and n_i variables are $\ll q^{1/2+\varepsilon}$), we have

$$(\delta_2, q) = 1. \tag{7.8}$$

We may also observe that $(g_0, qk_1) = 1$ since $g_0 \mid k_0$, $(k_0, c_0) = 1$, and $c_0 \equiv 0 \pmod{qk_1k_1^*}$. From $k_1 \mid a\delta_1$, we also conclude that

$$k_1 \mid \delta_2. \tag{7.9}$$

Therefore,

$$\mathcal{S}'' = \sum_{\substack{g_0 \mid e_1e_2\delta_1 am_1' \\ g_0 \equiv 0 \,(\mathrm{mod}\, d)}} \mathcal{S}''', \tag{7.10}$$

where

$$\mathcal{S}''' = \sum_{\substack{(c_0, g_0m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \sum_{n_1,n_2,n_3 \geq 1} \frac{e\left(\frac{\delta_2n_1n_2n_3\overline{c_0}}{k_0'}\right)}{\sqrt{\delta_1e_1e_2n_1n_2n_3}}$$
$$\times J(r_1e_1n_2, r_2e_2n_2, r_3n_3, a, m_1', c_0, g_0k_0', c_2, k_1). \tag{7.11}$$

8. Triple Poisson

8.1. *Poisson summation formula*

Our next step is to apply Poisson summation in n_1, n_2, n_3 modulo k_0' to (7.11), to which end we state the following general version.

Proposition 8.1. *Let J be any smooth and compactly supported function on $(0, \infty)^3$, and suppose $(\alpha, k) = 1$. Then*

$$\sum_{n_1,n_2,n_3 \geq 1} e\left(\frac{n_1n_2n_3\alpha}{k}\right) J(n_1, n_2, n_3) = \frac{1}{k^3} \sum_{p_1,p_2,p_3 \in \mathbb{Z}} A(p_1, p_2, p_3; \alpha; k) B(p_1, p_2, p_3; k),$$

where

$$A(p_1, p_2, p_3; \alpha; k) = \sum_{x_1,x_2,x_3 \,(\mathrm{mod}\, k)} e\left(\frac{x_1x_2x_3\alpha - x_1p_1 - x_2p_2 - x_3p_3}{k}\right), \tag{8.1}$$

$$B(p_1, p_2, p_3; k) = \int_0^\infty \int_0^\infty \int_0^\infty J(t_1, t_2, t_3) e\left(\frac{p_1t_1}{k} + \frac{p_2t_2}{k} + \frac{p_3t_3}{k}\right) dt_1\, dt_2\, dt_3.$$

An evaluation for A is given in Lemma 8.2 below (see also Lemma 13.2).

In view of (7.11), we need to evaluate $A(p_1, p_2, p_3; \delta_2\overline{c_0}; k_0')$, and analyze

$$B(p_1, p_2, p_3; k_0') = \int_{(\mathbb{R}^+)^3} J(r_1e_1t_1, r_2e_2t_2, r_3t_3, a, m_1', c_0, g_0k_0', c_2, k_1) \times e\left(\frac{p_1t_1}{k_0'} + \frac{p_2t_2}{k_0'} + \frac{p_3t_3}{k_0'}\right) \frac{dt_1\, dt_2\, dt_3}{\sqrt{e_1e_2\delta_1t_1t_2t_3}}, \quad (8.2)$$

where $r_1r_2r_3 = \delta_1$ (see (7.6)).

8.2. *The evaluation of A*

Lemma 8.2. *Suppose* $(\alpha, k) = 1$. *Then*

$$A(p_1, p_2, p_3; \alpha; k) = k \sum_{f|(p_2,p_3,k)} f\, S\left(p_1\overline{\alpha}, \frac{p_2p_3}{f^2}; \frac{k}{f}\right),$$

where $S(m, n; c)$ *is the standard Kloosterman sum.*

Proof. By first evaluating the sum over x_3, we derive

$$A(p_1, p_2, p_3; \alpha; k) = k \sum_{\substack{x_1,x_2 \ (\mathrm{mod}\, k) \\ x_1x_2\alpha\equiv p_3 \ (\mathrm{mod}\, k)}} e\left(\frac{x_1p_1 + x_2p_2}{k}\right).$$

At this point we decompose the sum by letting $(x_1, k) = f$ with $f \mid k$. Say $x_1 = fy$ with y running over reduced residue classes modulo k/f. Note that necessarily $f \mid (p_3, k)$, and $x_2 \equiv \overline{\alpha y}\frac{p_3}{f} \pmod{k/f}$. Therefore, we may write $x_2 = \overline{\alpha y}\frac{p_3}{f} + v\frac{k}{f}$ where v runs modulo f. Hence,

$$A(p_1, p_2, p_3; \alpha; k) = k \sum_{f|(p_3,k)} \sideset{}{^*}\sum_{y \ (\mathrm{mod}\, k/f)} e\left(\frac{yp_1}{k/f}\right) e\left(\frac{\overline{\alpha y}\frac{p_3}{f}p_2}{k}\right) \sum_{v \ (\mathrm{mod}\, f)} e\left(\frac{p_2v}{f}\right).$$

The sum over v detects $f \mid p_2$, and so the formula follows. □

8.3. *Asymptotics of B*

Let us begin by unraveling the definition of B. First we recall its definition from (8.2), (7.3), and (6.8). Let us also pull out a factor $N^{-1/2}$ coming from $(e_1e_2\delta_1t_1t_2t_3)^{-1/2}$. Recall that $I_K(k)$ has a built-in inert function. One may change this inert function appropriately to achieve that B takes a simplified form:

$$N^{1/2}B(p_1, p_2, p_3; k_0') = \int_{(\mathbb{R}^+)^3} e\left(\frac{-e_1e_2\delta_1t_1t_2t_3am_1'}{c_0k_0}\right) I^*(m_1, k, r_1e_1t_1, r_2e_2t_2, r_3t_3a, c) \times e\left(\frac{p_1t_1}{k_0'} + \frac{p_2t_2}{k_0'} + \frac{p_3t_3}{k_0'}\right) dt_1\, dt_2\, dt_3,$$

where I^* has the same properties as I given in Lemmas 6.3 and 6.4 (since only the definition of the inert function has changed). Note that the support of the inert function is such that $t_i \asymp N_i'$, say, where

$$N_1' = \frac{N_1}{e_1 r_1}, \qquad N_2' = \frac{N_2}{e_2 r_2}, \qquad N_3' = \frac{N_3}{r_3}.$$

Define $N' = N_1' N_2' N_3'$. In the analytic aspect, it is usually most convenient to work with the original variable names (we may perform the substitutions later, after analyzing the integral transform). Let

$$h = e_1 e_2 r_1 r_2 r_3 = e_1 e_2 \delta_1,$$

and note that

$$N' h = N.$$

The terms with some $p_i = 0$ will be treated in Section 13, using a more elementary approach than the method used in the analysis of the nonzero terms with $p_1 p_2 p_3 \neq 0$. For the nonzero terms, we apply dyadic partitions of unity to each p_i variable, for the positive and negative values separately. For $\mathrm{P} = (P_1, P_2, P_3)$, let B_{P} be the same as B but multiplied by one function from this partition of unity with $|p_i| \asymp P_i$, $i = 1, 2, 3$; we also assume that the sign of each p_i is fixed by the partition, but suppress the signs in the notation of B_{P}. As a convention, we may incorporate the case $p_i = 0$ by setting $P_i = 0$.

Lemma 8.3 (Post-transition). *Suppose* (6.21) *holds. Then*

$$M_2^{1/2} N^{1/2} B_{\mathrm{P}}(p_1, p_2, p_3; k_0') = \frac{C}{(aMN)^{1/2}} M_2 N' w_{\mathrm{T}}(\cdot), \tag{8.3}$$

where w_{T} is q^{ε}-inert, $N' = N_1' N_2' N_3'$, and where $B_{\mathrm{P}}(p_1, p_2, p_3)$ is very small unless

$$P_i \ll \frac{k_0'}{N_i'} q^{\varepsilon} \quad \text{and} \quad K \asymp \frac{(aMN)^{1/2}}{M_2}. \tag{8.4}$$

Proof of Lemma 8.3. The main observation is that the exponential factor appearing in (6.22) cancels the exponential factor in the definition of J in (7.3). Therefore, B is a Fourier transform of a q^{ε}-inert function supported on $t_i \asymp N_i'$, and hence by the discussion in Section 4.2, (8.3) follows. □

Lemma 8.4 (Pre-transition, non-oscillatory). *Suppose* (6.18) *holds, and additionally*

$$\frac{NaM_1}{CK} \ll_{\varepsilon} q^{\varepsilon}. \tag{8.5}$$

Then $B_{\mathrm{P}}(p_1, p_2, p_3; k_0')$ is very small unless

$$\frac{N_i' P_i}{k_0'} \ll_{\varepsilon} q^{\varepsilon}, \quad i = 1, 2, 3, \quad \text{and} \quad \frac{KM_2}{C} \ll_{\varepsilon} q^{\varepsilon},$$

in which case

$$M_2^{1/2}N^{1/2}B_{\mathrm{P}}(p_1,p_2,p_3;k_0') = \left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1} M_2N'w_{\mathrm{T}}(\cdot), \tag{8.6}$$

where w_{T} *is* q^{ε}*-inert.*

Proof. In this case, the exponential factor in the definition of B_{P} is essentially not oscillatory, because of the condition (8.5). For this, it is again helpful to remember that

$$\frac{e_1e_2\delta_1t_1t_2t_3am_1'}{c_0k_0} \asymp \frac{NaM_1}{CK}.$$

Since (8.5) holds, the exponential factor may be included into the definition of the inert weight function, which is q^{ε}-inert. As in the case of Lemma 8.3, we again obtain the Fourier transform of a q^{ε}-inert function, and hence we obtain the claimed estimates. □

We record for later use that under the conditions of Lemmas 8.3 and 8.4, we have

$$P_1P_2P_3 \ll_{\varepsilon} q^{\varepsilon}\frac{h}{N}\left(\frac{k}{k_1g_0}\right)^3 \asymp q^{\varepsilon}\frac{K^3}{N}\frac{h}{(k_1g_0)^3}. \tag{8.7}$$

Lemma 8.5 (Pre-transition, oscillatory). *Suppose* (6.18) *holds, and additionally*

$$\frac{NaM_1}{CK} \gg q^{\varepsilon}. \tag{8.8}$$

Then $B_{\mathrm{P}}(p_1,p_2,p_3;k_0')$ *is very small unless each* $p_i > 0$ *and*

$$P_i \asymp \frac{NaM_1k_0'}{CKN_i'}, \quad i=1,2,3, \quad \text{and} \quad \frac{KM_2}{C} \ll_{\varepsilon} q^{\varepsilon}, \tag{8.9}$$

in which case

$$\begin{aligned} M_2^{1/2}N^{1/2}B_{\mathrm{P}}(p_1,p_2,p_3;k_0') &= O\Big(q^{-A}\prod_{i=1}^{3}(1+|p_i|)^{-A}\Big) \\ &+ \left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1} M_2N'\left(\frac{CK}{aM_1N}\right)^{3/2} e\left(\frac{2(p_1p_2p_3ck)^{1/2}}{(am_1hk_0'^3)^{1/2}}\right)w_{\mathrm{T}}(\cdot), \end{aligned} \tag{8.10}$$

where w_{T} *is* q^{ε}*-inert.*

For later use, it is convenient to observe the identity

$$\frac{2(p_1p_2p_3ck)^{1/2}}{(am_1hk_0'^3)^{1/2}} = \frac{2(p_1p_2p_3c_0)^{1/2}}{k_0'(ahm_1'/g_0)^{1/2}} = \frac{2(p_1p_2p_3c_0)^{1/2}}{k_0'\delta_2^{1/2}}. \tag{8.11}$$

The explicit oscillatory rate of B_{P} seen in (8.10) is exactly that of a Whittaker function $W\big(\frac{4\pi\sqrt{|p_1p_2p_3c_0|}}{k_0'\delta_2^{1/2}}\big)$. This suggests that at this point of the proof the situation is ripe for the application of the Bruggeman–Kuznetsov formula. From this observation we can heuristically infer that the variables p_1, p_2, p_3 and c_0 are playing similar roles, that sums over each variable will reconstitute one of the four copies of the L-functions mentioned in (2.4), and k_0' is the modulus variable in the geometric side of the Bruggeman–Kuznetsov formula.

Proof of Lemma 8.5. In this case, the phase arising from reciprocity is oscillatory, and is not cancelled by a corresponding phase from the kernel function I_K. By (6.19) and (7.3), we have

$$M_2^{1/2}N^{1/2}B_{\mathrm{P}} = \left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1} M_2 \times \int_{\mathbb{R}^3} w_{\mathrm{T}}(t_1,t_2,t_3,\cdot)\, e\left(\frac{-t_1t_2t_3am_1h}{ck}\right) e\left(\frac{t_1p_1+t_2p_2+t_3p_3}{k_0'}\right) dt_1\, dt_2\, dt_3.$$

The behavior of this oscillatory integral is derived as an example in [KPY19]. □

8.4. Mellin transform of B

For many of our later purposes, we prefer to work with the Mellin transform of B_{P} instead of B_{P} itself. Of course, B_{P} depends on a number of variables, and what is meant here is the Mellin transform *in terms of* k_0'. Define

$$\widetilde{B_{\mathrm{P}}}(s) := \int_0^\infty B_{\mathrm{P}}(p_1,p_2,p_3;x)x^s \frac{dx}{x}, \tag{8.12}$$

which is the Mellin transform of B_{P} in k_0'. Recalling $k = g_0k_1k_0'$, note that

$$x \asymp \frac{K}{g_0k_1}.$$

Let us combine the results from Lemmas 8.3 and 8.4. In these two cases, we have

$$M_2^{1/2}N^{1/2}B_{\mathrm{P}}(p_1,p_2,p_3;k_0') = \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} M_2N' w_{\mathrm{T}}(\cdot), \tag{8.13}$$

where $\delta = -1$ in Lemma 8.3, and $\delta = \kappa - 1$ in Lemma 8.4. In both cases, p_i are supported on $|p_i| \asymp P_i \ll (k_0'/N_i')q^\varepsilon$, but there are different constraints on the parameters. In any event, in terms of k_0', B_{P} is q^ε-inert, so we group these two cases together under the heading of "non-oscillatory". Lemma 4.2 then leads to the following.

Lemma 8.6 (Non-oscillatory). *Suppose the conditions of Lemma 8.3 or Lemma 8.4 hold, and put $\delta = -1$ or $\delta = \kappa - 1$ in the respective cases. Then*

$$M_2^{1/2}N^{1/2}\widetilde{B_{\mathrm{P}}}(s) = \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} M_2N' \left(\frac{K}{g_0k_1}\right)^s w_{\mathrm{T}}(s,\cdot),$$

where w_{T} is q^ε-inert in all the variables except for s, and entire in terms of s. Moreover, $w_{\mathrm{T}}(\cdot,\sigma+it)$ is very small unless $|t| \ll_\sigma q^\varepsilon$.

In the case where B_{P} is oscillatory, it turns out to be easier to use the Bessel integral representations (9.8), (9.10) in the Bruggeman–Kuznetsov formula, and so we may avoid the Mellin transform analysis of B_{P}. See the introductory paragraphs of Section 10.2 for more explanation.

9. Application of Bruggeman–Kuznetsov

Write $\mathcal{T}_{\mathrm{P}}$ for the terms from $\mathcal{S}'''$ with B replaced by B_{P} (in particular, $p_1 p_2 p_3 \neq 0$). Therefore,

$$\mathcal{T}_{\mathrm{P}} = \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \pmod{q k_1 k_1^*}}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \pmod{\frac{c_2}{(g_0, c_2)}}}} \frac{1}{k_0'^3} \times \sum_{p_1, p_2, p_3 \neq 0} A(p_1, p_2, p_3; \delta_2 \overline{c_0}; k_0') B_{\mathrm{P}}(p_1, p_2, p_3; k_0').$$

Applying Lemma 8.2, and moving the sum over k_0' to the inside, we obtain

$$\mathcal{T}_{\mathrm{P}} = \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \pmod{q k_1 k_1^*}}} \frac{1}{c_0} \sum_{p_1, p_2, p_3 \neq 0} \sum_{f \mid (p_2, p_3)} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \pmod{\frac{c_2}{(g_0, c_2)}} \\ k_0' \equiv 0 \pmod{f}}} \frac{f}{k_0'^2} S\left(p_1 c_0 \overline{\delta_2}, \frac{p_2 p_3}{f^2}; \frac{k_0'}{f}\right) B_{\mathrm{P}}(p_1, p_2, p_3; k_0').$$

We absorb $k_0'^{-2} \asymp (g_0 k_1)^2 / K^2$ into the inert function which changes the definition of B_{P} (call the new function $B_{\mathrm{P},*}$), but not any of the analytic properties it satisfies (cf. Section 4.4), giving

$$\mathcal{T}_{\mathrm{P}} = \frac{(g_0 k_1)^2}{K^2} \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \pmod{q k_1 k_1^*}}} \frac{1}{c_0} \sum_{p_1, p_2, p_3 \neq 0} \sum_{f \mid (p_2, p_3)} f \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \pmod{\frac{c_2}{(g_0, c_2)}} \\ k_0' \equiv 0 \pmod{f}}} S\left(p_1 c_0 \overline{\delta_2}, \frac{p_2 p_3}{f^2}; \frac{k_0'}{f}\right) B_{\mathrm{P},*}(p_1, p_2, p_3; k_0').$$

Let $k_0' = f k_0''$, so that the inner sum over k_0' becomes

$$\delta_{(f, \delta_2 c_0)=1} \sum_{\substack{(k_0'', \delta_2 c_0)=1 \\ k_0'' \equiv 0 \pmod{\delta_3}}} S\left(p_1 c_0 \overline{\delta_2}, \frac{p_2 p_3}{f^2}; k_0''\right) B_{\mathrm{P},*}(p_1, p_2, p_3, f k_0''),$$

where we have defined

$$\delta_3 = \frac{c_2/(g_0, c_2)}{(f, c_2/(g_0, c_2))}. \tag{9.1}$$

Finally, to ease a later summation over c_0, we detect the condition $(k_0'', c_0) = 1$ with Möbius inversion, say over the variable δ_4. Then we reverse the order of summations, and define

$$c_0 = \delta_4 c_0', \qquad \delta_5 = [\delta_3, \delta_4]. \tag{9.2}$$

We record that the summation conditions in the sum over k_0'' are empty unless

$$(\delta_2, \delta_5) = 1, \quad \text{i.e.} \quad (\delta_2, \delta_3\delta_4) = 1.$$

For later use, we also record that

$$(\delta_4, k_1) = 1, \tag{9.3}$$

since $k_1 \mid \delta_2$, and $(\delta_2, \delta_4) = 1$. Using this, and moving the sum over f to the outside, with the definitions

$$p_2 = f p_2', \qquad p_3 = f p_3',$$

we obtain

$$\mathcal{T}_{\mathrm{P}} = \frac{(g_0 k_1)^2}{K^2} \sum_{\delta_3 \mid \frac{c_2}{(g_0, c_2)}} \sum_{(\delta_4, \delta_2 g_0 m_1') = 1} \frac{\mu(\delta_4)}{\delta_4} \sum_{\substack{(f, \delta_2\delta_4) = 1 \\ (9.1) \text{ is true}}} f \sum_{\substack{(c_0', f g_0 m_1') = 1 \\ \delta_4 c_0' \equiv 0 \, (\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0'} \sum_{p_1, p_2', p_3' \neq 0} \mathcal{K}_{\mathrm{P}}, \tag{9.4}$$

where

$$\mathcal{K}_{\mathrm{P}} = \sum_{\substack{(k_0'', \delta_2) = 1 \\ k_0'' \equiv 0 \, (\mathrm{mod}\, \delta_5)}} S(p_1 \delta_4 c_0' \overline{\delta_2}, p_2' p_3'; k_0'') B_{\mathrm{P},*}(p_1, f p_2', f p_3'; f k_0''). \tag{9.5}$$

Consulting Proposition 3.4, we may now realize the Kloosterman sum in question as one belonging to the group $\Gamma = \Gamma_0(\delta_2\delta_5)$ with the pair of cusps ∞, $1/\delta_5$ (note that these are Atkin–Lehner cusps, since $(\delta_2, \delta_5) = 1$). Hence

$$\mathcal{K}_{\mathrm{P}} = \sum_{k_0''\sqrt{\delta_2} \in \mathcal{C}_{\infty, 1/\delta_5}} S_{\infty, 1/\delta_5}(p_1 \delta_4 c_0', p_2' p_3'; k_0''\sqrt{\delta_2}) B_{\mathrm{P},*}(p_1, f p_2', f p_3'; f k_0'').$$

According to Theorem 3.5, write $\mathcal{K}_{\mathrm{P}} = \mathcal{K}_d + \mathcal{K}_c + \mathcal{K}_h$, and accordingly write $\mathcal{T}_{\mathrm{P}} = \mathcal{T}_d + \mathcal{T}_c + \mathcal{T}_h$. We furthermore decompose $\mathcal{K}_{\mathrm{P}} = \sum_{\epsilon_1, \epsilon_2, \epsilon_3 \in \{-1, 1\}} \mathcal{K}_{\mathrm{P}}^{\epsilon_1, \epsilon_2, \epsilon_3}$, where the meaning is $\epsilon_i p_i \geq 1$ for $i = 1, 2, 3$, and likewise decompose $\mathcal{K}_d$, etc. To simplify the notation and reduce the number of cases to investigate, let $\mathcal{K}^+$ denote the terms with $p_i \geq 1$ for all i, and $\mathcal{K}^-$ the terms with $p_1 \leq -1$ and $p_2, p_3 \geq 1$. The other sign combinations may be easily treated by symmetry, since the Kloosterman sum is unchanged when reversing the signs on two of the p_i, and since Lemmas 8.3 and 8.4 are insensitive to the signs of the p_i, while Lemma 8.5 requires all $p_i > 0$, which is covered by $\mathcal{K}^+$.

We have

$$\mathcal{K}_d^{\pm} = \sum_{t_j \text{ level } \delta_2\delta_5} \nu_{\infty,j}(p_1\delta_4 c_0')\overline{\nu}_{1/\delta_5,j}(p_2'p_3')W_{\pm}(p_1, fp_2', fp_3'; t_j), \tag{9.6}$$

where

$$W_{\pm}(p_1, fp_2', fp_3'; t_j) = \int_{(2\theta+\varepsilon)} h_{\pm}(s,t_j)\big(4\pi\sqrt{\delta_4 c_0'|p_1|p_2'p_3'}\big)^{-s}\widetilde{\widetilde{B}}_{\mathrm{P},*}(p_1, fp_2', fp_3'; s+1)\, ds,$$

(recall $h_{\pm}$ was defined by (3.6)), and where

$$\widetilde{\widetilde{B}}_{\mathrm{P},*}(p_1, fp_2', fp_3'; s+1) := \int_0^\infty B_{\mathrm{P},*}(p_1, fp_2', fp_3'; fy/\sqrt{\delta_2})y^{s+1}\frac{dy}{y}.$$

Here the "double tilde" notation for B is meant to indicate the Mellin transform of B with respect to $\gamma = k_0''\sqrt{\delta_2}$ (where $\gamma \in \mathcal{C}_{\mathfrak{ab}}$ as in (3.14)), because we have already reserved $\widetilde{B}$ for the Mellin transform in the k_0' variable (as in Section 8.4). The relationship between these two transforms is

$$\widetilde{\widetilde{B}}_{\mathrm{P},*}(s+1) = (\sqrt{\delta_2}/f)^{s+1}\widetilde{B}_{\mathrm{P},*}(s+1).$$

Simplifying, we obtain

$$W_{\pm}(p_1,p_2,p_3;t_j) = \frac{\sqrt{\delta_2}}{f}\int_{(2\theta+\varepsilon)} h_{\pm}(s,t_j)\left(\frac{\sqrt{\delta_2}}{\sqrt{\delta_4 c_0'|p_1|p_2 p_3}}\right)^s \widetilde{B}_{\mathrm{P},*}(s+1)\, ds. \tag{9.7}$$

The holomorphic case is similar, but with a different integral kernel than $h_{\pm}(s,t_j)$.

We may also prefer to use the Bessel integral representation for W, which we do in case $B_{\mathrm{P},*}$ is oscillatory. For instance, we have

$$W_h(p_1, fp_2', fp_3'; \ell) = \int_0^\infty J_{\ell-1}\left(\frac{4\pi\sqrt{p_1\delta_4 c_0' p_2' p_3'}}{y}\right)B_{\mathrm{P},*}(p_1, fp_2', fp_3'; fy/\sqrt{\delta_2})\, dy.$$

Changing variables, we obtain

$$W_h(p_1, fp_2', fp_3'; \ell) = \frac{\sqrt{\delta_2}}{f}\int_0^\infty J_{\ell-1}\left(\frac{4\pi\sqrt{f^2 p_1 p_2' p_3' \delta_4 c_0'}}{\sqrt{\delta_2}\, y}\right)B_{\mathrm{P},*}(p_1, fp_2', fp_3'; y)\, dy. \tag{9.8}$$

Note that, in terms of the older variable names, we have

$$\frac{f^2 p_1 p_2' p_3' \delta_4 c_0'}{\delta_2} = \frac{p_1 p_2 p_3 c_0}{ahm_1'/g_0}. \tag{9.9}$$

Similarly, in the $+$ Maass case, we have

$$W_+(p_1, fp_2', fp_3'; t_j) = \frac{\sqrt{\delta_2}}{f} \int_0^\infty B^+_{2it_j}\left(\frac{4\pi\sqrt{f^2 p_1 p_2' p_3' \delta_4 c_0'}}{\sqrt{\delta_2}\, y}\right) B_{\mathrm{P},*}(p_1, fp_2', fp_3'; y)\, dy. \tag{9.10}$$

10. Asymptotics of W

Here we analyze the various W-functions appearing in the Bruggeman–Kuznetsov formula.

10.1. Non-oscillatory cases

First suppose the conditions of Lemma 8.3 or Lemma 8.4 hold, so that Lemma 8.6 gives the behavior of $\widetilde{B}$. Continuing from (9.7), we have

$$W_\pm(p_1, p_2, p_3; t_j) = \frac{(\sqrt{aMN}/C)^\delta M_2 N' K}{M_2^{1/2} N^{1/2}} \frac{\sqrt{\delta_2}}{f g_0 k_1} \int_{(2\theta+\varepsilon)} h_\pm(s, t_j) \left(\frac{\sqrt{\delta_2}\, K}{g_0 k_1 \sqrt{\delta_4 c_0' |p_1| p_2 p_3}}\right)^s w_{\mathrm{T}}(s, \cdot\,)\, ds. \tag{10.1}$$

Here w_{T} is q^ε-inert in all variables except s. It is entire in s, with rapid decay for $|\mathrm{Im}(s)| \gg q^\varepsilon$.

As shorthand, let

$$Y = \frac{g_0 k_1 \sqrt{C P_1 P_2 P_3}}{\sqrt{\delta_2 c_2}\, K} \asymp \left(\frac{\sqrt{\delta_2}\, K}{g_0 k_1 \sqrt{\delta_4 c_0' |p_1| p_2 p_3}}\right)^{-1}. \tag{10.2}$$

Our goal now is to show

Lemma 10.1 (Non-oscillatory). *Suppose the conditions of Lemma 8.3 or Lemma 8.4 hold. If $|t_j| \gg (1+Y)q^\varepsilon$, then $W_\pm(t_j, \cdot\,)$ is very small. Similarly, if $\ell \gg (1+Y)q^\varepsilon$, then $W_h(\ell, \cdot\,)$ is very small.*

Proof. If $s = \sigma + it$ and $|t| \gg (|t_j| q)^\varepsilon$, then by the rapid decay of w_{T}, we conclude that this part of the integral is bounded in a satisfactory manner. In the complementary region, we know from Stirling that

$$h_\pm(\sigma + it, t_j) \ll_\sigma q^\varepsilon (q^\varepsilon + |t_j|)^{\sigma - 1}.$$

Side remark: The exponential factor implicitly appearing in Stirling's bound on $h_\pm(s, t_j)$ is $\ll 1$, and one cannot do better in general, because in one of the two cases of $\pm$ sign, the exponential factor is exactly 1.

Now if $|t_j| \gg (1+Y)q^{\varepsilon}$, we shift the contour far to the left and bound it trivially. In doing so, one encounters poles at $s/2 \pm it_j = 0, -1, -2, \dots$. However, these all have large imaginary part and w_{T} is very small here, so these residues are bounded in a satisfactory manner. The integral on the new line is very small since $|t_j|/Y \gg q^{\varepsilon}$.

Next consider $W_h(\ell)$. The analysis is similar, except one replaces $h_{\pm}(s, t_j)$ by

$$h(s, \ell) := 2^{s-1} \frac{\Gamma\left(\frac{s+\ell-1}{2}\right)}{\Gamma\left(\frac{-s+\ell+1}{2}\right)}.$$

Stirling's formula gives, for $\sigma \ll \sqrt{|\ell + it|}$,

$$|h(\sigma + it, \ell)| \ll (\max(\ell, |t|))^{\sigma - 1}.$$

As before, if $\ell \gg (1+Y)q^{\varepsilon}$, then we may move the contour far to the left (some large constant not growing with q). Then $W_h(\ell)$ is small, by the same type of reasoning as in the Maass case. □

Now we reap the reward of the language of inert functions. Since w_{T} is inert in all variables, we may apply the Mellin inversion formula together with Lemma 4.2, which gives

$$W_{\pm}(p_1, p_2, p_3; t_j) = \left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1} (M_2 N)^{1/2} K \frac{\sqrt{\delta_2}}{hfg_0k_1} \int_{(2\theta+\varepsilon)} h_{\pm}(s, t_j)$$
$$\times \int \left(\frac{\sqrt{\delta_2}\, K}{g_0 k_1 \sqrt{\delta_4 c_0' |p_1| p_2 p_3}}\right)^{s} \widetilde{w_{\mathrm{T}}}(s, \mathbf{u}, \cdot) \left(\frac{P_1}{|p_1|}\right)^{u_1} \left(\frac{P_2}{p_2}\right)^{u_2} \left(\frac{P_3}{p_3}\right)^{u_3} \left(\frac{C}{\delta_4 c_0' c_2}\right)^{u_4} d\mathbf{u}\, ds, \tag{10.3}$$

plus a small error term, where $\mathbf{u} = (u_1, u_2, u_3)$ and is integrated over an arbitrary product of lines with $\mathrm{Re}(u_i)$ fixed, for $i = 1, 2, 3$. Here $\widetilde{w_{\mathrm{T}}}$ is very small except if the imaginary parts of all the variables are $\ll q^{\varepsilon}$.

10.2. Oscillatory case

Now we consider $W_{\pm}$ and W_h when B is given by Lemma 8.5. The first significant point is that W_- is small, because this corresponds to the case where $p_1 p_2 p_3 < 0$, which means $p_i < 0$ for some i, in which case B is small. Indeed, B is small unless $p_i > 0$ for all i, and so the only relevant functions are W_+ and W_h.

It is inconvenient to use (9.7) in the oscillatory case. The problem is that the oscillatory nature of B means that we may no longer restrict $|\mathrm{Im}(s)|$ to be $O(q^{\varepsilon})$, which in turn has an effect on the behavior of $h_+(s, r)$ and $h(s, \ell)$. Namely, it is no longer true that $h_+(s, r)$ and $h(s, \ell)$ satisfy analogous asymptotic formulas (due to the use of Stirling with ir large vs. ℓ large), and so it appears difficult to unify these two cases. In addition, one is forced to confront some tricky oscillatory integrals. To sidestep these problems entirely, we shall use the Bessel integral formula for W instead. The oscillatory behavior of B is actually beneficial and causes W to be essentially inert (in both the Maass and holomorphic cases).

Let us begin with W_h. Using (9.8), (9.9), and (8.11) to match phases, we have

$$W_h(p_1, p_2, p_3; \ell) = \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} \sqrt{M_2 N}\left(\frac{CK}{aM_1N}\right)^{3/2} \frac{\sqrt{\delta_2}}{fh} Z$$

plus a small error term, where

$$Z = \int_0^\infty J_{\ell-1}\left(\frac{4\pi\sqrt{p_1p_2p_3\delta_4c_0'}}{\sqrt{\delta_2}y}\right) e\left(\frac{2\sqrt{p_1p_2p_3\delta_4c_0'}}{\sqrt{\delta_2}y}\right) w_{\mathrm{T}}(y, \cdot)\, dy. \tag{10.4}$$

Here we recall that w_{T} has support on $y \asymp \frac{K}{g_0k_1}$. The fact that the phases match is pleasant.

Recall the integral representation

$$J_{\ell-1}(x) = \sum_{\pm} c_{\ell,\pm} \int_0^{\pi/2} \cos((\ell-1)\theta) e^{\pm ix\cos\theta}\, d\theta, \qquad c_{\ell,\pm} = \frac{e^{\mp i(\ell-1)\pi/2}}{\pi}. \tag{10.5}$$

This gives

$$Z = \sum_{\pm} \int_0^{\pi/2} c_{\ell,\pm} \cos((\ell-1)\theta) \int_0^\infty e\left(\frac{z(1\pm\cos\theta)}{y}\right) w_{\mathrm{T}}(y, \cdot)\, dy\, d\theta$$

with

$$z = \frac{2\sqrt{p_1p_2p_3\delta_4c_0'}}{\sqrt{\delta_2}}.$$

Changing variables $y = \frac{K}{g_0k_1x}$ now gives $x \asymp 1$, and the inner integral is a Fourier transform of an inert function. Hence

$$Z = \frac{K}{g_0k_1} \sum_{\pm} \int_0^{\pi/2} c_{\ell,\pm} \cos((\ell-1)\theta) \widehat{w_{\mathrm{T}}}\left(\frac{zg_0k_1}{K}(1\pm\cos\theta)\right) d\theta,$$

where we have redefined w_{T} (see Section 4.4). Using $\delta_4c_0' = c_0 = c/c_2$, $\delta_2 = ham_1'/g_0$, $m_1' = \frac{m_1}{k_1c_2}$, (8.9), and $k_0' = \frac{k}{k_1g_0}$, we check the size of

$$\frac{zg_0k_1}{K} \asymp \frac{\sqrt{P_1P_2P_3Ck_1^3g_0^3}}{K\sqrt{haM_1}} \asymp \frac{NaM_1}{CK},$$

which is $\gg q^{\varepsilon}$ because we are operating under the conditions of Lemma 8.5.

Now we observe that the integrand is very small unless

$$\frac{NaM_1}{CK}|1\pm\cos\theta| \ll q^{\varepsilon}.$$

Hence, the sign must be $-$, and we must have

$$\theta \ll \left(\frac{CK}{NaM_1}\right)^{1/2} q^{\varepsilon}$$

(which is $O(q^{-\delta})$ for some $\delta > 0$). This means that by using a Taylor expansion, we may develop the $\widehat{w_{\mathrm{T}}}$ part into an asymptotic expansion with leading term given by the substitution $1 - \cos\theta \mapsto \theta^2/2$. Therefore,

$$Z = \frac{K}{g_0 k_1} \int_{-\infty}^{\infty} \cos((\ell - 1)\theta) \left(\widehat{w_{\mathrm{T}}}\left(\frac{z g_0 k_1}{K} \theta^2 \right) + \cdots \right) d\theta,$$

where we were able to extend the integral to $+\infty$ since $\widehat{w_{\mathrm{T}}}$ is small otherwise, and also extend to $-\infty$ by symmetry (we have also redefined the inert function to absorb constants).

As another shorthand, let

$$Q = \frac{z g_0 k_1}{K} \asymp \frac{N a M_1}{CK}.$$

Then Z takes the form

$$Z = \frac{K}{g_0 k_1 \sqrt{Q}} \int_{-\infty}^{\infty} \exp\left(i \frac{(\ell - 1)}{\sqrt{Q}} \theta \right) \widehat{w_{\mathrm{T}}}(\theta^2)\, d\theta + \cdots .$$

If we let $g(\theta) = \widehat{w_{\mathrm{T}}}(\theta^2)$, then $g^{(j)}(\theta) \ll_{j,A} X^j (1 + \theta)^{-A}$ for arbitrary j, A, where $X \ll q^{\varepsilon}$. Therefore, this is another Fourier transform of a function with controlled derivatives, and so by the discussion in Section 4.2, it takes the form

$$\frac{K}{g_0 k_1 \sqrt{Q}} G\left(\frac{\ell - 1}{\sqrt{Q}}, \cdot \right)$$

plus a very small error term, where G would be q^{ε}-inert (in ℓ) if it had dyadic support. It is q^{ε}-inert in all the other variables, however.

Regrouping, we have

$$W_h = \left(\frac{\sqrt{aMN}}{C} \right)^{\kappa - 1} \sqrt{M_2 N} \left(\frac{CK}{aM_1 N} \right)^2 K \frac{\sqrt{\delta_2}}{f g_0 k_1 h} G\left(\frac{\ell - 1}{\sqrt{Q}}, \cdot \right)$$

plus a very small error term, where G is very small unless

$$\ell \ll \left(\frac{M_1 a N}{CK} \right)^{1/2} q^{\varepsilon}.$$

Then we may take the Mellin transform in p_1, p_2, p_3, c_0, which gives

$$W_h(p_1, p_2, p_3; c_0'; \ell) = \left(\frac{\sqrt{aMN}}{C} \right)^{\kappa - 1} \sqrt{M_2 N} \left(\frac{CK}{aM_1 N} \right)^2 K \frac{\sqrt{\delta_2}}{f g_0 k_1 h}$$
$$\times \int_{(\sigma)} \widetilde{w_{\mathrm{T}}}(\mathbf{u}, \ell, \cdot) \left(\frac{P_1}{|p_1|} \right)^{u_1} \left(\frac{P_2}{p_2} \right)^{u_2} \left(\frac{P_3}{p_3} \right)^{u_3} \left(\frac{C}{\delta_4 c_0' c_2} \right)^{u_4} d\mathbf{u} \qquad (10.6)$$

plus a small error term, where $\int_{(\sigma)}$ is integration over the line $\mathrm{Re}(u_i) = \sigma$, $i = 1, 2, 3$, and

$\sigma \in \mathbb{R}$ is arbitrary to be chosen later. Now we turn to W_+. Since the details are similar to the previous case, the exposition is brief. We follow through the steps used for W_h, where the alteration in the first step is replacing $J_{\ell-1}$ by $B_{2ir}^+(x)$. In place of (10.5), we have instead

$$\frac{J_{2ir}(x) - J_{-2ir}(x)}{\sinh(\pi r)} = \frac{2}{\pi i} \int_{-\infty}^{\infty} \cos(x \cosh v) e\Big(\frac{rv}{\pi}\Big) \, dv.$$

We shall use this for real values of r. Although this integral does not converge absolutely, we have

$$\Big| \int_{|v| \geq V} \cos(x \cosh v) e\Big(\frac{rv}{\pi}\Big) \, dv \Big| \ll \frac{1 + |r|}{x \sinh V}, \tag{10.7}$$

from integration by parts.

Forming the analog of Z from W_h, and keeping the same definition of z, we have (absorbing the absolute constant into the inert function)

$$Z = \int_{-\infty}^{\infty} e\Big(\frac{rv}{\pi}\Big) \int_{-\infty}^{\infty} e\Big(\frac{z}{y}\Big) \cos\Big(2\pi \frac{z}{y} \cosh(v)\Big) w_{\mathrm{T}}(y, \cdot) \, dy \, dv,$$

using (10.7) to reverse the order of integrations. Next write $\cos u = \frac{1}{2} e^{iu} + \frac{1}{2} e^{-iu}$; the part with e^{iu} is very small as in the W_h case. From this point on, the analysis is nearly identical to that of W_h, and the conclusion is that W_+ is very small unless

$$|t_j| \ll \Big(\frac{M_1 a N}{CK}\Big)^{1/2} q^{\varepsilon},$$

and W_+ satisfies a formula identical to that in (10.6).

In the exceptional eigenvalue case where $ir \in \mathbb{R}$, the final shape of the formula for W_+ is the same as (10.6), but the above arguments would need modification since $e(rv/\pi)$ is no longer bounded. There is a more direct route, however. We have the asymptotic expansion (see [GR00, (8.451.1)])

$$\frac{J_{2ir}(x) - J_{-2ir}(x)}{\sinh(\pi r)} \sim \sum_{\pm} e^{\pm ix} \sum_{\nu} \frac{P_{\pm}(r, \nu)}{x^{1/2+\nu}},$$

where $P_{\pm}(r, \nu)$ is a polynomial in r and ν. This is certainly valid for $r = O(1)$ and $x \gg 1$ (in the present context, $x \gg q^{\varepsilon}$). With this, it is easy to estimate Z directly, showing that it is of the form $\frac{K}{g_0 k_1 \sqrt{Q}}$ times a q^{ε}-inert function, plus a small error term. Therefore, applying Mellin inversion in the appropriate variables, we obtain an expression of the same form as (10.6).

11. Regrouping after Bruggeman–Kuznetsov

11.1. Non-oscillatory Maass cases

Here we consider the contribution to $\mathcal{T}_d$ from the parameters where B is non-oscillatory. By (9.4), (9.6), and (10.3), we obtain

$$\mathcal{T}_d^{\pm} = \frac{g_0 k_1}{K} \sum_{\delta_3 | \frac{c_2}{(g_0, c_2)}} \sum_{(\delta_4, \delta_2 g_0 m_1')=1} \frac{\mu(\delta_4)}{\delta_4} \sum_{\substack{(f, \delta_2\delta_4)=1 \\ (9.1) \text{ is true}}} \sum_{\substack{(c_0', f g_0 m_1')=1 \\ \delta_4 c_0' \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0'} \sum_{\pm p_1, p_2', p_3' \geq 1}$$
$$\sum_{t_j \text{ level } \delta_2\delta_5} \nu_{\infty,j}(p_1 \delta_4 c_0') \overline{\nu}_{1/\delta_5, j}(p_2' p_3') \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \int_{(2\theta+\varepsilon)} h_{\pm}(s, t_j)$$
$$\times \int_{(\sigma)} \left(\frac{\sqrt{\delta_2} K}{f g_0 k_1 \sqrt{\delta_4 c_0' |p_1| p_2' p_3'}} \right)^s \widetilde{w_{\mathrm{T}}}(s, \mathbf{u}, \cdot\,)$$
$$\times \left(\frac{P_1}{|p_1|} \right)^{u_1} \left(\frac{P_2}{f p_2'} \right)^{u_2} \left(\frac{P_3}{f p_3'} \right)^{u_3} \left(\frac{C}{\delta_4 c_0' c_2} \right)^{u_4} d\mathbf{u}\, ds \qquad (11.1)$$

plus a very small error term. In the above expression, we could take $\mathrm{Re}(s) > 2\theta$ without crossing any poles coming from exceptional Laplace eigenvalues (recall (3.6) for the definition of $h_{\pm}$), and $\sigma \in \mathbb{R}$ is arbitrary to be chosen later. By Lemma 10.1, we may truncate at $|t_j| \ll (1+Y)q^{\varepsilon}$ with a small error term. Now we move the sums over p_1, p_2', p_3', and c_0 to the inside, change variables $u_i \mapsto u_i - s/2$, and bound everything at that point with absolute values. In this way, we obtain

$$\mathcal{T}_d^{\pm} \ll q^{\varepsilon} \frac{g_0 k_1 c_2}{KC} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \sum_{\delta_3 | \frac{c_2}{(g_0, c_2)}} \sum_{(\delta_4, \delta_2 g_0 m_1')=1} \sum_{\substack{(f, \delta_2\delta_4)=1 \\ (9.1) \text{ is true}}} \sum_{\substack{t_j \text{ level } \delta_2\delta_5 \\ |t_j| \ll (1+Y)q^{\varepsilon}}}$$
$$\frac{1}{1+|t_j|} \int_{(2\theta+\varepsilon)} \int_{(\sigma)} \left(\frac{t_j}{Y} \right)^{2\theta+\varepsilon} |\widetilde{w_{\mathrm{T}}}(s, \mathbf{u} - s/2, \cdot\,)| \left| P_1^{u_1} \left(\frac{P_2}{f} \right)^{u_2} \left(\frac{P_3}{f} \right)^{u_3} \left(\frac{C}{\delta_4 c_2} \right)^{u_4} \right|$$
$$\times |Z_j(\mathbf{u})|\, d\mathbf{u}\, ds \qquad (11.2)$$

plus a very small error term, where

$$Z_j(\mathbf{u}) = \sum_{\substack{(c_0', f g_0 m_1')=1 \\ \delta_4 c_0' \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \sum_{p_1, p_2', p_3' \geq 1} \frac{\nu_{\infty,j}(p_1 \delta_4 c_0') \overline{\nu}_{1/\delta_5, j}(p_2' p_3')}{p_1^{u_1} p_2'^{u_2} p_3'^{u_3} c_0'^{u_4}}. \qquad (11.3)$$

Our plan is to relate $Z_j(\mathbf{u})$ to L-functions, and use a large sieve inequality to bound it on average over t_j.

11.2. Non-oscillatory holomorphic cases

These cases are nearly identical to those in Section 11.1, but the bounds will turn out to be even better due to the applicability of Deligne's bound. The key point is that for $\ell \ll (1+Y)q^{\varepsilon}$, we may claim the bound

$$|h(s,\ell)| \ll \ell^{\sigma-1},$$

which is entirely analogous to $|h(s,t_j)| \ll t_j^{\sigma-1}$. We omit the details for brevity.

11.3. Oscillatory Maass cases

As in Section 11.1, we use (9.4) and (9.6), but instead of (10.3) we use (10.6) (which, as discussed in Section 10.2, holds also for W_+ in place of W_h). Also recall that only the $+$ sign enters the picture in the oscillatory case. Thus we obtain

$$\mathcal{T}_d^+ \ll \frac{g_0 k_1 c_2}{KC}\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}\sum_{\delta_3 | \frac{c_2}{(g_0,c_2)}}\ \sum_{(\delta_4,\delta_2 g_0 m_1')=1}\ \sum_{\substack{(f,\delta_2\delta_4 c_0')=1\\ (9.1)\text{ is true}}}$$
$$\sum_{\substack{t_j \text{ level } \delta_2\delta_5\\ |t_j|\ll Y' q^{\varepsilon}}}\left(\frac{CK}{aM_1N}\right)^2\int_{(\sigma)}|\widetilde{w_{\mathrm{T}}}(t_j,\mathbf{u},\cdot)|\left|P_1^{u_1}\left(\frac{P_2}{f}\right)^{u_2}\left(\frac{P_3}{f}\right)^{u_3}\left(\frac{C}{\delta_4 c_2}\right)^{u_4}\right||Z_j(\mathbf{u})|\,d\mathbf{u}, \tag{11.4}$$

where

$$Y' = \left(\frac{M_1 aN}{CK}\right)^{1/2} q^{\varepsilon}.$$

11.4. Oscillatory holomorphic cases

These are similar to (but easier than) the oscillatory Maass cases, and so we omit them.

11.5. Continuous spectrum

First consider the non-oscillatory cases. Then analogously to (11.1), we have

$$\mathcal{T}_c^{\pm} = \frac{g_0k_1}{K}\sum_{\delta_3|\frac{c_2}{(g_0,c_2)}}\ \sum_{(\delta_4,\delta_2 g_0 m_1')=1}\frac{\mu(\delta_4)}{\delta_4}\sum_{\substack{(f,\delta_2\delta_4)=1\\ (9.1)\text{ is true}}}\ \sum_{\substack{(c_0', f g_0 m_1')=1\\ \delta_4 c_0' \equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}}\frac{1}{c_0'}\sum_{\pm p_1,p_2',p_3'\geq 1}$$
$$\sum_{\mathfrak{c}}\int_{t=-\infty}^{\infty}\nu_{\infty,\mathfrak{c}}(p_1\delta_4c_0', 1/2+it)\overline{\nu}_{1/\delta_5,\mathfrak{c}}(p_2'p_3', 1/2+it)\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}$$
$$\times\int_{(\varepsilon)}h_{\pm}(s,t)\int_{(1+\varepsilon)}\left(\frac{\sqrt{\delta_2}K}{fg_0k_1\sqrt{\delta_4c_0'|p_1|p_2'p_3'}}\right)^{s}\widetilde{w_{\mathrm{T}}}(s,\mathbf{u},\cdot)$$
$$\times\left(\frac{P_1}{|p_1|}\right)^{u_1}\left(\frac{P_2}{fp_2'}\right)^{u_2}\left(\frac{P_3}{fp_3'}\right)^{u_3}\left(\frac{C}{\delta_4c_0'c_2}\right)^{u_4}d\mathbf{u}\,ds\,dt. \tag{11.5}$$

Now we move the sums to the inside, getting

$$\mathcal{T}_c^{\pm} \ll q^{\varepsilon} \frac{g_0 k_1 c_2}{KC} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \sum_{\delta_3 | \frac{c_2}{(g_0, c_2)}} \sum_{(\delta_4, \delta_2 g_0 m_1')=1} \sum_{\substack{(f, \delta_2 \delta_4)=1 \\ (9.1) \text{ is true}}} \int_{|t| \ll (1+Y) q^{\varepsilon}}$$
$$\frac{1}{1+|t|} \int_{(\varepsilon)} \int_{(1+\varepsilon)} |\widetilde{w_{\mathrm{T}}}(s, \mathbf{u} - s/2, \cdot)| \left| P_1^{u_1} \left(\frac{P_2}{f} \right)^{u_2} \left(\frac{P_3}{f} \right)^{u_3} \left(\frac{C}{\delta_4 c_2} \right)^{u_4} \right|$$
$$\times \sum_{\mathfrak{c}} |Z_{\mathfrak{c},t}(\mathbf{u})| \, d\mathbf{u} \, ds, \qquad (11.6)$$

where

$$Z_{\mathfrak{c},t}(\mathbf{u}) = \sum_{\substack{(c_0', f g_0 m_1')=1 \\ \delta_4 c_0' \equiv 0 \, (\mathrm{mod}\, q k_1 k_1^*)}} \sum_{p_1, p_2', p_3' \geq 1} \frac{\nu_{\infty, \mathfrak{c}}(p_1 \delta_4 c_0', 1/2 + it) \overline{\nu}_{1/\delta_5, \mathfrak{c}}(p_2' p_3', 1/2 + it)}{p_1^{u_1} p_2'^{u_2} p_3'^{u_3} c_0'^{u_4}}. \qquad (11.7)$$

The **oscillatory case** is similar, leading to

$$\mathcal{T}_c^{+} \ll \frac{g_0 k_1 c_2}{KC} \left(\frac{\sqrt{aMN}}{C} \right)^{\kappa - 1} (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \sum_{\delta_3 | \frac{c_2}{(g_0, c_2)}} \sum_{(\delta_4, \delta_2 g_0 m_1')=1} \sum_{\substack{(f, \delta_2 \delta_4 c_0')=1 \\ (9.1) \text{ is true}}} \int_{|t| \ll Y' q^{\varepsilon}}$$
$$\left(\frac{CK}{aM_1 N} \right)^2 \int_{(1+\varepsilon)} |\widetilde{w_{\mathrm{T}}}(t, \mathbf{u}, \cdot)| \left| P_1^{u_1} \left(\frac{P_2}{f} \right)^{u_2} \left(\frac{P_3}{f} \right)^{u_3} \left(\frac{C}{\delta_4 c_2} \right)^{u_4} \right| \sum_{\mathfrak{c}} |Z_{\mathfrak{c},t}(\mathbf{u})| \, d\mathbf{u} \, dt. \qquad (11.8)$$

11.6. Claiming bounds on Z_j, and estimating $\mathcal{T}$

In Section 12, we will show

Lemma 11.1. *The function $Z_j(\mathbf{u})$ defined by (11.3) has analytic continuation to $\mathrm{Re}(\mathbf{u}) \geq \sigma > 1/2$. In this region it satisfies the bound*

$$\sum_{\substack{t_j \text{ level } \delta_2 \delta_5 \\ |t_j| \leq T}} |Z_j(\mathbf{u})| \ll_{\sigma, \varepsilon} q^{\theta - 1/2} \frac{(\delta_4, q)^{1/2}}{(k_1 k_1^*)^{1/2} \delta_4^{1/2}} T^{2+\varepsilon} q^{\varepsilon} \, \mathrm{Poly}(|\mathbf{u}|), \qquad (11.9)$$

where $\mathrm{Poly}(|\mathbf{u}|)$ is some fixed polynomial in the absolute values of the coordinates of $\mathbf{u}$.

The key feature is that this bound saves a factor $\delta_4^{1/2}$. This saving comes from the fact that one of the Fourier coefficients in (11.3) houses a δ_4 in the Fourier coefficient ν_j. For newforms, the fact that δ_4 divides the level gives us the $\delta_4^{-1/2}$ savings, since $\delta_5 = [\delta_3, \delta_4]$ and (3.31) is a bound for such Hecke eigenvalues. This side of the Bruggeman–Kuznetsov formula is a spectrally complete sum over all Maass forms of level $\delta_2 \delta_5$, including oldforms. The savings from bounding the Fourier coefficients of an oldform is even better, as the normalized Fourier coefficients of oldforms decrease proportionally with the increase in volume of the fundamental domain. We make this oldform-newform interplay precise in Section 12.1 below with the $AB = \delta_2\delta_5$ consideration.

Now we use Lemma 11.1 to estimate $\mathcal{T}_d^{\pm}$, and eventually $\mathcal{S}$. We do not require the factor $(k_1k_1^*)^{-1/2}$ appearing in (11.9), and in order to unify the treatment with the continuous spectrum, we shall only use a weaker bound with this factor omitted.

First consider the **non-oscillatory Maass cases**. Inserting the bound from Lemma 11.1 into (11.2) (taking $\sigma = 1/2 + \varepsilon$) gives

$$\mathcal{T}_d^{\pm} \ll q^{\varepsilon}\frac{g_0k_1c_2}{KC}\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}\sum_{\delta_3|\frac{c_2}{(g_0,c_2)}}\sum_{(\delta_4,\delta_2g_0m_1')=1}\sum_{\substack{(f,\delta_2\delta_4)=1\\ (9.1)\text{ is true}}}$$
$$\int_{(2\theta+\varepsilon)}\int_{(\sigma)}|\widetilde{w_{\mathrm{T}}}(s,\mathbf{u}-s/2,\cdot)|\frac{(P_1P_2P_3C)^{1/2}}{f\delta_4^{1/2}c_2^{1/2}}q^{\theta-1/2}\frac{(\delta_4,q)^{1/2}}{(k_1k_1^*)^{1/2}\delta_4^{1/2}}(Y^{-2\theta}+Y)$$
$$\times\,\mathrm{Poly}(|\mathbf{u}|)\,d\mathbf{u}\,ds, \qquad (11.10)$$

by considering the two cases $Y \ll 1$ and $Y \gg 1$ separately (recall Y was defined in (10.2)). Summing over δ_4 (here is where the savings of $\delta_4^{1/2}$ is important, as the δ_4-sum is now essentially a harmonic series with a benign extra factor $(\delta_4, q)^{1/2}$), f, and δ_3, and integrating over s and $\mathbf{u}$, we obtain

$$\mathcal{T}_d^{\pm} \ll q^{\varepsilon}\frac{g_0k_1c_2}{KC}\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}(Y^{-2\theta}+Y)(P_1P_2P_3)^{1/2}\left(\frac{C}{c_2}\right)^{1/2}q^{\theta-1/2}. \qquad (11.11)$$

Let us write $\mathcal{S}_d^{\pm}$ for the contribution to $\mathcal{S}$ from this part. Applying the additional summations that led from $\mathcal{S}$ to $\mathcal{S}'''$ (see (7.10), (7.6), (7.1)), we obtain

$$\mathcal{S}_d^{\pm} \ll q^{\varepsilon}\sum_a\frac{1}{a^{3/2}}\sum_{c_2}\frac{1}{c_2^{3/2}}\sum_{d|c_2}d\sum_{k_1}k_1^{1/2}\sum_{m_1'}\frac{1}{\sqrt{m_1'}}\sum_{r_1r_2r_3=\delta_1}\sum_{e_1|r_2r_3}\sum_{e_2|r_3}\sum_{\substack{g_0|e_1e_2\delta_1am_1'\\ g_0\equiv 0\ (\mathrm{mod}\ d)}}$$
$$\frac{g_0k_1c_2}{KC}\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}(Y^{-2\theta}+Y)(P_1P_2P_3)^{1/2}\left(\frac{C}{c_2}\right)^{1/2}q^{\theta-1/2}.$$

Convention. Here and below, we have not written the truncation points for these outer summation variables. In almost all cases, all that is necessary is to recall that all the variables may be bounded by some fixed power of q. The only exception is that for some estimates we need to use $m_1' \ll M_1/(k_1c_2)$.

For convenience, we gather some of the previous definitions:

$$h = e_1e_2r_1r_2r_3, \quad \delta_1 = r_1r_2r_3 = \frac{k_1d}{(a,k_1d)}, \quad \delta_2 = \frac{e_1e_2\delta_1am_1'}{g_0} = \frac{ham_1'}{g_0},$$
$$N'h = N, \quad m_1 = k_1c_2m_1', \quad Y = \frac{g_0k_1\sqrt{CP_1P_2P_3}}{\sqrt{\delta_2c_2}\,K}. \qquad (11.12)$$

With these substitutions, we obtain

$$\sqrt{\delta_2}\,(Y^{-2\theta}+Y)(P_1P_2P_3)^{1/2}\left(\frac{C}{c_2}\right)^{1/2} = \frac{\delta_2 K}{g_0k_1}(Y^{1-2\theta}+Y^2),$$

and hence

$$\mathcal{S}_d^{\pm} \ll q^{\varepsilon}\sum_a \frac{1}{a^{3/2}}\sum_{c_2}\frac{1}{c_2^{3/2}}\sum_{d|c_2} d\sum_{k_1} k_1^{1/2}\sum_{m_1'}\frac{1}{\sqrt{m_1'}}\sum_{r_1r_2r_3=\delta_1}\sum_{e_1|r_2r_3}\sum_{e_2|r_3}\sum_{\substack{g_0|e_1e_2\delta_1am_1'\\ g_0\equiv 0 \ (\mathrm{mod}\ d)}} \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} M_2^{1/2}N^{1/2}\frac{c_2}{C}\frac{am_1'}{g_0}(Y^{1-2\theta}+Y^2)q^{\theta-1/2}. \quad (11.13)$$

Now we note that in this non-oscillatory case, we see from (8.7) that

$$\frac{P_1P_2P_3(g_0k_1)^2}{\delta_2c_2} \ll_{\varepsilon} q^{\varepsilon}\frac{K^3}{NaM_1},$$

which in particular means that $Y \ll_{\varepsilon} \left(\frac{CK}{NaM_1}\right)^{1/2}q^{\varepsilon}$, which is independent of g_0, k_1, c_2, etc. Now it is evident that the sums over $g_0, e_1, e_2, r_1, r_2, r_3$ contribute at most $O(q^{\varepsilon})$, and the fact that $d \mid g_0$ will cancel the other visible factor of d in (11.13). With this observation, and performing minor simplifications, we have

$$\mathcal{S}_d^{\pm} \ll q^{\varepsilon}\sum_a \frac{1}{a^{1/2}}\sum_{c_2}\frac{1}{c_2^{1/2}}\sum_{d|c_2}\sum_{k_1} k_1^{1/2}\sum_{m_1'}\sqrt{m_1'} \times \left(\frac{\sqrt{aMN}}{C}\right)^{\delta}\frac{M_2^{1/2}N^{1/2}}{C}q^{\theta-1/2}\left(\left(\frac{\sqrt{CK}}{\sqrt{NaM_1}}\right)^{1-2\theta}+\left(\frac{\sqrt{CK}}{\sqrt{NaM_1}}\right)^2\right).$$

Trivially summing over m_1' (recall $m_1' \ll M_1/(k_1c_2)$), k_1, d, c_2, and finally a, we derive

$$\mathcal{S}_d^{\pm} \ll q^{\varepsilon}\max_a M_1^{3/2}\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}\frac{M_2^{1/2}N^{1/2}a^{1/2}}{C}q^{\theta-1/2}\left(\left(\frac{\sqrt{CK}}{\sqrt{NaM_1}}\right)^{1-2\theta}+\left(\frac{\sqrt{CK}}{\sqrt{NaM_1}}\right)^2\right).$$

Now we split into the cases of Lemmas 8.3 and 8.4. In the **case of Lemma 8.3**, we have $\delta = -1$, and

$$M_1^{3/2}\left(\frac{\sqrt{aMN}}{C}\right)^{\delta}\frac{M_2^{1/2}N^{1/2}a^{1/2}}{C}q^{\theta-1/2} = \frac{M_1}{q^{1/2-\theta}}.$$

Meanwhile, using $K \asymp M_2^{-1}\sqrt{aMN}$ (see (6.23)), we have

$$\frac{CK}{NaM_1} \asymp \frac{C}{\sqrt{aMN}} \ll q^{\varepsilon}$$

via (6.21). Therefore, in this case

$$\mathcal{S}_d^{\pm} \ll_\varepsilon \frac{M_1}{q^{1/2}} q^{\theta+\varepsilon} \ll_\varepsilon q^{\theta+\varepsilon}. \tag{11.14}$$

In the **case of Lemma 8.4**, we have $\frac{CK}{NaM_1} \gg q^{-\varepsilon}$, and $\delta = \kappa - 1 \geq 1$, so with easy simplifications, we derive

$$\mathcal{S}_d^{\pm} \ll_\varepsilon q^\varepsilon \max_a \frac{M_2 K}{C} \frac{M_1 q^\theta}{q^{1/2}}.$$

Since $KM_2/C \ll q^\varepsilon$ in this case, we obtain the same bound as in (11.14).

The non-oscillatory holomorphic cases are nearly identical, so we omit the proofs.

Now consider the **oscillatory Maass case**, where we treat (11.4). Following the same steps as in the non-oscillatory cases, we obtain

$$\mathcal{T}_d^{+} \ll \frac{g_0 k_1}{K} \left(\frac{\sqrt{aMN}}{C}\right)^\delta (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \left(\frac{CK}{aM_1 N}\right) (P_1 P_2 P_3)^{1/2} \left(\frac{c_2}{C}\right)^{1/2} \frac{q^\theta}{q^{1/2}}.$$

After some simplifications, we have

$$\mathcal{S}_d^{+} \ll q^\varepsilon \sum_a \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d \sum_{k_1} k_1^{1/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \sum_{r_1 r_2 r_3 = \delta_1} \sum_{e_1 | r_2 r_3} \sum_{e_2 | r_3} \sum_{\substack{g_0 | e_1 e_2 \delta_1 a m_1' \\ g_0 \equiv 0 \, (\mathrm{mod}\, d)}}$$
$$\frac{g_0 k_1}{K} \left(\frac{\sqrt{aMN}}{C}\right)^\delta (M_2 N)^{1/2} \left(\frac{c_2}{C}\right)^{1/2} \frac{\sqrt{\delta_2}}{h} \left(\frac{CK}{aM_1 N}\right) (P_1 P_2 P_3)^{1/2} \frac{q^\theta}{q^{1/2}}. \tag{11.15}$$

We need to remember the origins of these variables. We have

$$P_1 P_2 P_3 \asymp \frac{(NaM_1)^3 k_0'^3}{C^3 K^3 N'} \asymp \left(\frac{NaM_1}{CK}\right)^3 \frac{K^3}{N} \frac{h}{(g_0 k_1)^3}. \tag{11.16}$$

Thus the bound becomes

$$\mathcal{S}_d^{+} \ll q^\varepsilon \sum_a \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d \sum_{k_1} k_1^{1/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \sum_{r_1 r_2 r_3 = \delta_1} \sum_{e_1 | r_2 r_3} \sum_{e_2 | r_3} \sum_{\substack{g_0 | e_1 e_2 \delta_1 a m_1' \\ g_0 \equiv 0 \, (\mathrm{mod}\, d)}}$$
$$\frac{g_0 k_1}{K} \left(\frac{\sqrt{aMN}}{C}\right)^\delta (M_2 N)^{1/2} \left(\frac{c_2}{C}\right)^{1/2} \sqrt{\frac{ham_1'}{g_0}} \left(\frac{NaM_1}{CK}\right)^{1/2} \left(\frac{K^3}{N} \frac{h}{(g_0 k_1)^3}\right)^{1/2} \frac{q^\theta}{hq^{1/2}}.$$

We see that the sum over g_0 gives $O(d^{-1} q^\varepsilon)$, and the h-dependence cancels out entirely, so that the δ_1-dependence is also essentially gone. Thus, we obtain

$$\mathcal{S}_d^{+} \ll q^\varepsilon \frac{q^\theta}{q^{1/2}} \sum_a \frac{1}{a} \sum_{c_2} \frac{1}{c_2} \sum_{d|c_2} \sum_{k_1} (k_1 k_1^*)^{-1/2}$$
$$\times \sum_{m_1'} \frac{(M_2 N)^{1/2}}{K C^{1/2}} \left(\frac{\sqrt{aMN}}{C}\right)^\delta \left(\frac{NaM_1}{CK}\right)^{1/2} \left(\frac{K^3}{N}\right)^{1/2}.$$

Now we sum over all the remaining variables, which gives in all

$$\mathcal{S}_d^+ \ll q^{\varepsilon} \frac{M_1 q^{\theta}}{q^{1/2}} \frac{(M_2 N)^{1/2}}{K C^{1/2}} \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} \left(\frac{NaM_1}{CK}\right)^{1/2} \left(\frac{K^3}{N}\right)^{1/2}.$$

Simplifying (in particular, $\delta = \kappa - 1$ here), we obtain

$$\mathcal{S}_d^+ \ll_{\varepsilon} q^{\varepsilon} \frac{M_1 q^{\theta}}{q^{1/2}} \frac{MaN}{C^2}.$$

Since $\sqrt{MaN} \ll_{\varepsilon} Cq^{\varepsilon}$ (see (6.18)), we obtain the same bound as in (11.14). The **oscillatory holomorphic case** is similar, but even simpler.

In summary, this shows the desired bound for the Maass forms and holomorphic forms.

11.7. Claiming bounds on $Z_{\mathfrak{c},t}$, and estimating $\mathcal{T}_c$

Recall the definition (11.5). Define flrt to be the multiplicative function defined on prime powers by

$$\mathrm{flrt}(p^{\alpha}) = p^{\lfloor \alpha/2 \rfloor}. \tag{11.17}$$

Lemma 11.2. *The function $Z_{\mathfrak{c},t}(\mathbf{u})$ has a decomposition*

$$Z_{\mathfrak{c},t}(u_1, u_2, u_3, u_4) = (Z_1^0(u_2, u_3) + Z_1^*(u_2, u_3))(Z_2^0(u_1, u_4) + Z_2^*(u_1, u_4)),$$

where for $i = 1, 2$, $Z_i^*(\alpha, \beta)$ has analytic continuation to $\mathrm{Re}(\alpha, \beta) \geq \sigma > 1/2$, and $Z_i^0(\alpha, \beta)$ is analytic for $\mathrm{Re}(\alpha, \beta) \geq \sigma > 1$. For $\mathrm{Re}(\mathbf{u}) \geq \sigma > 1/2$, we have

$$\int_{|t| \leq T} \sum_{\mathfrak{c}} |Z_1^*(u_2, u_3) Z_2^*(u_1, u_4)| \, dt \ll_{\mathbf{u}, \varepsilon} q^{\varepsilon} T^{2+\varepsilon} \left(\frac{(\delta_4, q)}{q}\right)^{1/2} \frac{\mathrm{flrt}(\delta_2)\, \mathrm{flrt}(\delta_3)^{3/2}}{\sqrt{\delta_2 \delta_5}}. \tag{11.18}$$

For $\mathrm{Re}(\mathbf{u}) \geq \sigma > 1$, we have

$$\sum_{\mathfrak{c}} |Z_1^0(u_2, u_3) Z_2^0(u_1, u_4)| \ll_{\mathbf{u}, \varepsilon} (q(1+|t|))^{\varepsilon} \frac{(\delta_4, q)}{q\sqrt{k_1 k_1^*}} \frac{1}{\delta_2 \delta_5}. \tag{11.19}$$

For $\mathrm{Re}(u_1, u_4) \geq \sigma > 1$ and $\mathrm{Re}(u_2, u_3) \geq \sigma' > 1/2$, we have

$$\int_{|t| \leq T} \sum_{\mathfrak{c}} |Z_1^*(u_2, u_3) Z_2^0(u_1, u_4)| \, dt \ll_{\mathbf{u}, \varepsilon} q^{\varepsilon} T^{1+\varepsilon} \frac{(\delta_4, q)}{q\sqrt{k_1 k_1^*}} \frac{\mathrm{flrt}(\delta_2) \sqrt{\mathrm{flrt}(\delta_3)}}{\delta_2 \sqrt{\delta_5}}. \tag{11.20}$$

For $\mathrm{Re}(u_2, u_3) \geq \sigma > 1$ and $\mathrm{Re}(u_1, u_4) \geq \sigma' > 1/2$, we have

$$\int_{|t| \leq T} \sum_{\mathfrak{c}} |Z_1^0(u_2, u_3) Z_2^*(u_1, u_4)| \, dt \ll_{\mathbf{u}, \varepsilon} q^{\varepsilon} T^{1+\varepsilon} \left(\frac{(\delta_4, q)}{q}\right)^{1/2} \frac{\mathrm{flrt}(\delta_2)\, \mathrm{flrt}(\delta_3)}{\delta_2 \delta_5}. \tag{11.21}$$

The implied dependence on $\mathbf{u}$ is at most polynomial, as in Lemma 11.1.

We comment on some important features of the above bounds. In (11.18) and (11.21) we require a factor $\delta_5^{-1/2}$ (or better) to secure convergence of the sum over δ_4. The overall power of k_1 is also important for securing convergence in each case. In terms of the final power of q that occurs in our bound on $\mathcal{S}_c$, the most important feature is the power of δ_2. This is because δ_2 contains the m_1' variable which can be as large as $q^{1/2+\varepsilon}$. Note that although $\operatorname{flrt}(n)$ may occasionally be as large as $\sqrt{n}$, it is small on average, indeed $\sum_{n\le x} \operatorname{flrt}(n) \ll x \log x$.

Using Lemma 11.2, we bound $\mathcal{T}_c$. For the non-oscillatory cases, we return to (11.6). Technically, we should return to (11.5), decompose $\mathcal{T}_c^{\pm}$ according to $Z_{\mathfrak{c},t} = (Z_1^* + Z_1^0)(Z_2^* + Z_2^0)$ into four pieces, shift contours to the lines allowed by Lemma 11.2, and only then apply the absolute values. We found it a bit easier to bound Z_1^0 and Z_2^0 slightly to the right of the 1-line instead of bounding the residues of Z_1 and Z_2, but this is more or less equivalent.

Then (note that the sum over f converges absolutely, and the t-integral is easily estimated, so we may simplify a bit in these aspects)

$$\mathcal{T}_c^{\pm} \ll q^{\varepsilon} \frac{g_0 k_1 c_2}{KC} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \sum_{\delta_3 \mid \frac{c_2}{(g_0, c_2)}} \sum_{(\delta_4, \delta_2 g_0 m_1')=1} \operatorname{flrt}(\delta_2) \operatorname{flrt}(\delta_3)$$
$$\times \Bigg(\frac{(\delta_4, q)^{1/2}}{q^{1/2}} (1+Y) \frac{(P_1 P_2 P_3 C)^{1/2} \operatorname{flrt}(\delta_3)^{1/2}}{\sqrt{\delta_2 \delta_4 \delta_5 c_2}} + \frac{(\delta_4, q)}{q} \frac{P_1 P_2 P_3 C}{\delta_2 \delta_4 \delta_5 c_2 \sqrt{k_1 k_1^*}}$$
$$+ \frac{P_1 (P_2 P_3)^{1/2} C}{\delta_2 \delta_4 c_2 \sqrt{\delta_5}} \frac{(\delta_4, q)}{q \sqrt{k_1 k_1^*}} + \frac{P_1^{1/2} P_2 P_3 C^{1/2}}{\delta_4^{1/2} c_2^{1/2} \delta_2 \delta_5} \frac{(\delta_4, q)^{1/2}}{q^{1/2}} \Bigg).$$

Using $\delta_5 \ge \sqrt{\delta_3 \delta_4}$ (recall that $\delta_5 = [\delta_3, \delta_4]$), the sums over δ_3 and δ_4 are easily evaluated, and lead to a factor of size at most $O(q^{\varepsilon})$; the only slightly tricky case uses instead

$$\operatorname{flrt}(\delta_3)^{1/2} \sum_{\delta_4} \frac{(\delta_4, q)^{1/2}}{\sqrt{\delta_4 \delta_5}} = \operatorname{flrt}(\delta_3)^{1/2} \sum_{\delta_4} \frac{(\delta_4, \delta_3 q)^{1/2}}{\delta_4 \sqrt{\delta_3}} \ll \operatorname{flrt}(c_2) \frac{(\delta_3 q)^{\varepsilon}}{\sqrt{\delta_3}}. \tag{11.22}$$

It is helpful to observe the following nice simplification. At this point we can see that the first term within the parentheses which occurred from $Z_1^* Z_2^*$ will lead to the same bound we obtained on $\mathcal{T}_d^{\pm}$, by comparison to (11.11). The only difference is the benign factor of $\operatorname{flrt}(c_2)$, which does not make the sum over c_2 appreciably larger, since $\sum_{c_2} c_2^{-1} \operatorname{flrt}(c_2) \ll q^{\varepsilon}$. Actually, apart from $\operatorname{flrt}(c_2)$, the bound is better in two ways: firstly, the factor q^{θ} may be omitted, and secondly, instead of using $\operatorname{flrt}(\delta_2)/\sqrt{\delta_2} \le 1$, we could use that $\operatorname{flrt}(n)$ is $O(n^{\varepsilon})$ on average, which could lead to a saving of the factor $M_1^{1/2}$. Instead of carrying through the calculations, we will simply abbreviate this term by $(**)$ in the forthcoming displays.

Next we wish to sum over the outer variables that make $\mathcal{S}_c$ from $\mathcal{T}_c$. To this end, we need to write the P_i, Y, and δ_2 variables in terms of these outer ones. Let

$$P_i^* = K/N_i, \qquad i = 1, 2, 3, \tag{11.23}$$

so that $P_i \ll q^{\varepsilon} P_i^* \frac{h_i}{g_0 k_1}$ where $h_1 = e_1 r_1$, $h_2 = e_2 r_2$, and $h_3 = r_3$ (so $h = h_1 h_2 h_3$). With this, we obtain

$$\mathcal{T}_c^{\pm} \ll q^{\varepsilon} \frac{g_0 k_1 c_2}{KC} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} (M_2 N)^{1/2} \frac{\sqrt{\delta_2}}{h} \operatorname{flrt}(\delta_2) \times \Bigg[(**) + \frac{h}{(g_0 k_1)^3} \frac{P_1^* P_2^* P_3^* C}{q \delta_2 c_2 \sqrt{k_1 k_1^*}} + \frac{h_1 (h_2 h_3)^{1/2}}{(g_0 k_1)^2} \frac{P_1^* (P_2^* P_3^*)^{1/2} C}{q \delta_2 c_2 \sqrt{k_1 k_1^*}} + \frac{h_1^{1/2} h_2 h_3}{(g_0 k_1)^{5/2}} \frac{(P_1^*)^{1/2} P_2^* P_3^* C^{1/2}}{q^{1/2} c_2^{1/2} \delta_2} \Bigg].$$

Recall that

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \sum_a \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d | c_2} d \sum_{k_1} k_1^{1/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \sum_{r_1 r_2 r_3 = \delta_1} \sum_{e_1 | r_2 r_3} \sum_{e_2 | r_3} \sum_{\substack{g_0 | e_1 e_2 \delta_1 a m_1' \\ g_0 \equiv 0 \ (\mathrm{mod}\, d)}} \mathcal{T}_c^{\pm},$$

and that $\delta_2 = ham_1'/g_0$ and $h = e_1 e_2 \delta_1 = e_1 e_2 r_1 r_2 r_3$.

Next we analyze the sum over g_0 in all four terms. For the flrt part, we use $\operatorname{flrt}(\delta_2) = \operatorname{flrt}(e_1 e_2 \delta_1 a m_1'/g_0) \leq \operatorname{flrt}(e_1 e_2 \delta_1 a m_1') = \operatorname{flrt}(ham_1')$, and otherwise we see that the overall power of g_0 is negative in all terms, and so the smallest value of g_0, namely d, leads to the dominant part.

Putting this together, and simplifying, we obtain

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \sum_a \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d | c_2} d \sum_{k_1} k_1^{3/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \times \sum_{r_1 r_2 r_3 = \delta_1} \sum_{e_1 | r_2 r_3} \sum_{e_2 | r_3} \frac{(M_2 N)^{1/2}}{CK} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} \frac{\operatorname{flrt}(ham_1')}{\sqrt{ham_1'}} \times \Bigg[(**) + \frac{P_1^* P_2^* P_3^* C}{q d^{3/2} c_2 k_1^3 \sqrt{k_1 k_1^*}} + \frac{P_1^* (P_2^* P_3^*)^{1/2} C}{q k_1^2 c_2 \sqrt{d h_2 h_3 k_1 k_1^*}} + \frac{(P_1^*)^{1/2} P_2^* P_3^* C^{1/2}}{d k_1^{5/2} \sqrt{q h_1 c_2}} \Bigg].$$

Our next goal is to estimate the sum over m_1'. Since m_1' is independent of δ_1 (and hence e_1, e_2), we may move the sum over m_1' to the inside. We shall use the following estimate:

$$\sum_{n \leq X} \frac{\operatorname{flrt}(nN)}{n} \ll \operatorname{flrt}(N) (XN)^{\varepsilon},$$

which can be proved by elementary methods. Applying this to the sums over m_1', we obtain with easy simplifications

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \sum_a \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d \sum_{k_1} k_1^{3/2} \sum_{r_1r_2r_3=\delta_1} \sum_{e_1|r_2r_3} \sum_{e_2|r_3} \frac{(M_2N)^{1/2}}{CK} \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} \times \frac{\mathrm{flrt}(ha)}{\sqrt{ha}} \left[(**) + \frac{P_1^*P_2^*P_3^*C}{qd^{3/2}c_2k_1^3\sqrt{k_1k_1^*}} + \frac{P_1^*(P_2^*P_3^*)^{1/2}C}{qc_2k_1^2\sqrt{dh_2h_3k_1k_1^*}} + \frac{(P_1^*)^{1/2}P_2^*P_3^*C^{1/2}}{dk_1^{5/2}\sqrt{qh_1c_2}}\right]. \tag{11.24}$$

Using $\mathrm{flrt}(ha) \le \sqrt{ha}$, we can easily see that the outer variables sum to give no significant contribution. Therefore, we have

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \max_a \frac{(M_2N)^{1/2}}{CK} \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} \times \left[(**) + q^{-1}P_1^*P_2^*P_3^*C + q^{-1}P_1^*(P_2^*P_3^*)^{1/2}C + q^{-1/2}(P_1^*)^{1/2}P_2^*P_3^*C^{1/2}\right].$$

Substituting for P_i^* and simplifying, we obtain

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \max_a \frac{(M_2N)^{1/2}}{CK} \left(\frac{\sqrt{aMN}}{C}\right)^{\delta} \left[(**) + \frac{K^3C}{Nq} + \frac{K^2C}{qN_1\sqrt{N_2N_3}} + \frac{K^{5/2}C^{1/2}}{q^{1/2}N_1^{1/2}N_2N_3}\right]. \tag{11.25}$$

Now we split once more into the two types of non-oscillatory behavior. The **post-transition** case from Lemma 8.3 has $\sqrt{aMN}/C \gg q^{\varepsilon}$, which leads to $\delta = -1$ and $K \asymp (aMN)^{1/2}/M_2$. Therefore,

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \max_a (M_2N)^{1/2} \left[(**) + \frac{aMN}{M_2^2Nq} + \frac{\sqrt{aMN}}{qM_2N_1\sqrt{N_2N_3}} + \frac{\sqrt{aMN}}{q^{1/2}M_2^{3/2}N_1^{1/2}N_2N_3}\right].$$

This simplifies to give

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \left[1 + \frac{\sqrt{M_1}}{\sqrt{M_2}} \frac{\sqrt{M_1a^2N}}{q} + \frac{M_1^{1/2}\sqrt{aN_2N_3}}{q} + \frac{\sqrt{M_1}}{\sqrt{M_2}} \frac{\sqrt{aN_1}}{\sqrt{q}}\right]. \tag{11.26}$$

Using $M_1 \ll M_2$, $M_1 \ll q^{1/2+\varepsilon}$, and $aN_i \ll q^{1/2+\varepsilon}$, we deduce that $\mathcal{S}_c^{\pm} \ll q^{\varepsilon}$. One may observe that the part of $\mathcal{S}_c^{\pm}$ arising from $Z_1^*Z_2^0$ and $Z_1^0Z_2^*$ contributes at most $O(q^{-1/4+\varepsilon})$. With some additional work, one could show the contribution from $Z_1^*Z_2^*$ is also at most $O(q^{-1/4+\varepsilon})$.

For the **non-oscillatory pre-transition** case from Lemma 8.4 with $\delta = \kappa - 1 \ge 2$ (here is the only place where the choice of $\kappa = 2$ does not work), we have $K \ll q^{\varepsilon}C/M_2$ and $C \gg q^{\varepsilon}\sqrt{aMN}$, so we obtain

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \max_a (M_2N)^{1/2} \left[(**) + \frac{aMN}{M_2^2Nq} + \frac{\sqrt{aMN}}{qM_2N_1\sqrt{N_2N_3}} + \frac{\sqrt{aMN}}{M_2^{3/2}q^{1/2}N_1^{1/2}N_2N_3}\right]. \tag{11.27}$$

This is precisely the same bound as in (11.26), and so $\mathcal{S}_c^{\pm} \ll q^{\varepsilon}$. Actually, we only need $\kappa - 1 \geq 2$ for the term arising from $Z_1^0 Z_2^0$.

Finally, we consider the **oscillatory** case (where recall $\delta = \kappa - 1$ and only the $+$ sign enters). For this, we return to (11.8), that is,

$$\mathcal{T}_c^{+} \ll \frac{g_0 k_1 c_2}{KC}\left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}\sum_{\delta_3 | \frac{c_2}{(g_0,c_2)}}\sum_{(\delta_4,\delta_2 g_0 m_1')=1}\sum_{\substack{(f,\delta_2\delta_4 c_0')=1\\ (9.1)\text{ is true}}}\int_{|t|\ll Y' q^{\varepsilon}}$$
$$\left(\frac{CK}{aM_1N}\right)^2\int_{(1+\varepsilon)}|\widetilde{w_{\mathrm{T}}}(t,\mathbf{u},\cdot)|\left|P_1^{u_1}\left(\frac{P_2}{f}\right)^{u_2}\left(\frac{P_3}{f}\right)^{u_3}\left(\frac{C}{\delta_4 c_2}\right)^{u_4}\right|\sum_{\mathfrak{c}}|Z_{\mathfrak{c},t}(\mathbf{u})|\,d\mathbf{u}\,dt,$$

where again we should technically move the contours before applying the absolute values. Then we obtain

$$\mathcal{T}_c^{+} \ll q^{\varepsilon}\frac{g_0 k_1 c_2}{KC}\left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1}(M_2N)^{1/2}\frac{\sqrt{\delta_2}}{h}\sum_{\delta_3 | \frac{c_2}{(g_0,c_2)}}\sum_{(\delta_4,\delta_2 g_0 m_1')=1}\left(\frac{CK}{aM_1N}\right)^2$$
$$\times \operatorname{flrt}(\delta_2)\operatorname{flrt}(\delta_3)\Bigg[\frac{(\delta_4,q)^{1/2}}{q^{1/2}}Y'^2\frac{(P_1P_2P_3C)^{1/2}\operatorname{flrt}(\delta_3)^{1/2}}{\sqrt{\delta_2\delta_4\delta_5 c_2}}+Y'\frac{(\delta_4,q)}{q}\frac{P_1P_2P_3C}{\delta_2\delta_4\delta_5 c_2\sqrt{k_1k_1^*}}$$
$$+Y'\frac{P_1(P_2P_3)^{1/2}C}{\delta_2\delta_4 c_2\sqrt{\delta_5}}\frac{(\delta_4,q)}{q\sqrt{k_1k_1^*}}+Y'\frac{P_1^{1/2}P_2P_3C^{1/2}}{\delta_4^{1/2}c_2^{1/2}\delta_2\delta_5}\frac{(\delta_4,q)^{1/2}}{q^{1/2}}\Bigg].$$

Luckily, we may reuse some of the previous analysis in the non-oscillatory cases. We wish to sum over all the outer variables. Note that Y' is independent of them, and we have $P_i \asymp \frac{NaM_1}{CK}\frac{k_0'}{N_i'}$; previously we had $P_i \ll k_0'/N_i'$, so the only difference here is the extra factor $\frac{NaM_1}{CK}$ (which happens to be Y'^2). Therefore, the previous method of bounding the outer variables works identically as in this case. This time the term arising from $Z_1^* Z_2^*$ is identical to (11.15), save for $\operatorname{flrt}(c_2)$. We again denote it by $(**)$. Therefore, by altering (11.25) with the appropriate factors of Y' we have

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon}\max_a\frac{(M_2N)^{1/2}}{CKY'^4}$$
$$\times\left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1}\left[(**)+\frac{Y'^7K^3C}{Nq}+\frac{Y'^5K^2C}{qN_1\sqrt{N_2N_3}}+\frac{Y'^6K^{5/2}C^{1/2}}{q^{1/2}N_1^{1/2}N_2N_3}\right].$$

Substituting for Y', simplifying, and using $K \ll q^{\varepsilon}C/M_2$, this becomes

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon}\max_a (M_2N)^{1/2}\left(\frac{\sqrt{aMN}}{C}\right)^{\kappa-1}$$
$$\times\left[(**)+\frac{(M_1aN)^{3/2}}{CM_2^{1/2}Nq}+\frac{(M_1aN)^{1/2}}{qM_2^{1/2}N_1\sqrt{N_2N_3}}+\frac{M_1aN}{q^{1/2}CM_2^{1/2}N_1^{1/2}N_2N_3}\right].$$

Using $C \gg q^{-\varepsilon}\sqrt{MaN}$ and $\kappa - 1 \geq 1$, this gives

$$\mathcal{S}_c^{\pm} \ll q^{\varepsilon} \max_a \left[1 + \frac{M_1}{M_2} \frac{(M_2 a^2 N)^{1/2}}{q} + \frac{(M_1 a N_2 N_3)^{1/2}}{q} + \frac{\sqrt{M_1}}{\sqrt{M_2}} \frac{\sqrt{aN_1}}{\sqrt{q}} \right].$$

Again as in the non-oscillatory case, we see that $\mathcal{S}_c^{\pm} \ll q^{\varepsilon}$.

12. Bounding the Dirichlet series

12.1. Discrete spectrum

In this section we prove Lemma 11.1. Towards this, we develop some properties of an auxiliary Dirichlet series.

Lemma 12.1. *Suppose $N = LM$, f^* is a newform of level M, and d, Q are nonzero integers. For $\ell \mid L$, let $f = f^*|_\ell$, write $d/\ell = d_1/\ell_1$ in lowest terms (so $d = (d,\ell)d_1$, $\ell = (d,\ell)\ell_1$), and let $d_1 = d_M d_0$ with $d_M \mid M^\infty$ and $(d_0, M) = 1$. Define the Dirichlet series*

$$Z_{d,\ell,Q}(s,u) := \sum_{\substack{m,n \geq 1 \\ (n,Q)=1}} \frac{\nu_{f^*|_\ell}(dmn)}{m^s n^u},$$

initially for $\mathrm{Re}(s), \mathrm{Re}(u)$ *large. Then $Z_{d,\ell,Q}$ has analytic continuation to* $\mathrm{Re}(s), \mathrm{Re}(u) \geq \sigma > 1/2$*, wherein it satisfies the bound*

$$|Z_{d,\ell,Q}(s,u)| \ll_{\sigma,\varepsilon} |\nu_{f^*}(1)| (d,\ell)^{1/2} d_M^{-1/2} d_0^{\theta} (dNQ)^{\varepsilon} |L(f^*,s) L(f^*,u)|. \tag{12.1}$$

For the proof of Lemma 11.1, in a short while we will apply (12.1) twice. Within the confines of this section, N refers to the level of $f^*|_\ell$ and not to a dyadic variable. We will substitute $N = \delta_2\delta_5$, $Q = fg_0m_1'$. In the two cases this lemma will be used, we have $d = 1$ and also $d = D\delta_4/(D,\delta_4)$ where $D = qk_1k_1^*$.

Proof of Lemma 12.1. Firstly, from (3.30), we have

$$Z_{d,\ell,Q}(s,u) = \sum_{\substack{m,n \geq 1 \\ (n,Q)=1}} \frac{\ell^{1/2} \nu_{f^*}(dmn/\ell)}{m^s n^u}.$$

We have $\nu_{f^*}(dmn/\ell) = \nu_{f^*}(d_1 mn/\ell_1) = \lambda_{f^*}(d_M) \nu_{f^*}(d_0 mn/\ell_1)$, and so

$$Z_{d,\ell,Q}(s,u) = \ell^{1/2} \lambda_{f^*}(d_M) \nu_{f^*}(1) \sum_{\substack{m,n \geq 1 \\ (n,Q)=1}} \frac{\lambda_{f^*}(d_0 mn/\ell_1)}{m^s n^u}.$$

Using (3.31), and complete multiplicativity of Hecke eigenvalues for primes dividing M, we get $|\lambda_{f^*}(d_M)| \leq d_M^{-1/2}$.

By an exercise in the Hecke relations (somewhat in the spirit of (5.2)), one may derive the analytic continuation and the bound

$$\sum_{\substack{m,n\geq 1\\(n,Q)=1}} \frac{\lambda_{f^*}(d_0 mn/\ell_1)}{m^s n^u} \ll_\sigma \frac{(dNQ)^\varepsilon}{\ell_1^{1/2}} d_0^\theta |L(f^*,s)L(f^*,u)|,$$

where recall that $\mathrm{Re}(s), \mathrm{Re}(w) \geq \sigma > 1/2$. □

Now we proceed to prove Lemma 11.1. Recall the definition (11.3), which factors as

$$Z_j(\mathbf{u}) = \Bigg(\sum_{\substack{(c_0', fg_0 m_1')=1\\ \delta_4 c_0' \equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}} \sum_{p_1\geq 1} \frac{\nu_{\infty,j}(p_1\delta_4 c_0')}{p_1^{u_1} c_0'^{u_4}} \Bigg) \Bigg(\sum_{p_2,p_3\geq 0} \frac{\nu_{1/\delta_5,j}(p_2p_3)}{p_2^{u_2}p_3^{u_3}} \Bigg). \tag{12.2}$$

We begin by decomposing into newforms. By the choice of basis from Section 3.5, we have

$$\sum_{|t_j|\leq T} |Z_j(\mathbf{u})| \ll \sum_{AB=\delta_2\delta_5} \sum_{\substack{f^* \text{ new, level } B\\ |t_{f^*}|\leq T}} \sum_{\substack{\ell|A\\ \ell'|A}} \Bigg|\Bigg(\sum_{\substack{(c_0', fg_0 m_1')=1\\ \delta_4 c_0' \equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}} \sum_{p_1\geq 1} \frac{\nu_{\infty,f^*|_\ell}(p_1\delta_4 c_0')}{p_1^{u_1} c_0'^{u_4}} \Bigg) \times \Bigg(\sum_{p_2,p_3\geq 0} \frac{\nu_{1/\delta_5,f^*|_{\ell'}}(p_2p_3)}{p_2^{u_2}p_3^{u_3}} \Bigg)\Bigg|. \tag{12.3}$$

By Lemmas 3.7 and 12.1 with the choice $d = 1$, we see that the sum over p_2, p_3 has analytic continuation to the desired region, and satisfies

$$\sum_{p_2,p_3\geq 1} \frac{\nu_{1/\delta_5,f^*|_{\ell'}}(p_2p_3)}{p_2^{\alpha}p_3^{\beta}} \ll |\nu_{f^*}(1)|\,|L(f^*,\alpha)L(f^*,\beta)|,$$

uniformly in ℓ' and δ_5.

The first product in (12.3) is a bit trickier. Recall from (6.13) that $(q, k_1) = 1$ and from (7.9) that $k_1 \mid \delta_2$. Let $D = qk_1k_1^*$ and note that the divisibility condition in the sum can be rewritten as $D/(\delta_4, D) \mid c_0'$. Pulling the factor $D/(D, \delta_4)$ from the c_0' term, we apply (12.1) with $d = D\delta_4/(D, \delta_4)$. This gives

$$\sum_{\substack{(c_0', fg_0 m_1')=1\\ \delta_4 c_0' \equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}} \sum_{p_1\geq 1} \frac{\nu_{f^*|_\ell}(p_1\delta_4 c_0')}{p_1^{u_1} c_0'^{u_4}} \ll q^\varepsilon |\nu_{f^*}(1)| \left(\frac{(D,\delta_4)}{D}\right)^{1/2} \frac{(d,\ell)^{1/2} d_0^\theta}{d_B^{1/2}} |L(f^*,u_1)L(f^*,u_4)|,$$

where $d/\ell = d_1/\ell_1$ is in lowest terms, and then we factor $d_1 = d_B d_0$ where $d_B \mid B^\infty$ and $(d_0, B) = 1$.

Using $|\nu_{f^*}(1)|^2 = (AB)^{-1}q^{o(1)}$ via (3.27), we then have

$$\sum_{\substack{t_j \text{ level } \delta_2\delta_5 \\ |t_j|\leq T}} |Z_j(\mathbf{u})| \ll q^{\varepsilon}\left(\frac{(D,\delta_4)}{D}\right)^{1/2} \sum_{AB=\delta_2\delta_5} \sum_{\ell|A} (d,\ell)^{1/2} d_B^{-1/2} d_0^{\theta}$$
$$\times \frac{1}{AB} \sum_{\substack{f^* \text{ new, level } B \\ |t_{f^*}|\leq T}} |L(f^*,u_1)L(f^*,u_2)L(f^*,u_3)L(f^*,u_4)|.$$

A standard argument with the spectral large sieve (e.g., see [Mot97, Theorem 3.4] for the level 1 case) implies

$$\sum_{\substack{t_j \text{ level } \delta_2\delta_5 \\ |t_j|\leq T}} |Z_j(\mathbf{u})| \ll_{\mathbf{u},\varepsilon} T^{2+\varepsilon} q^{\varepsilon}\left(\frac{(D,\delta_4)}{D}\right)^{1/2} \sum_{AB=\delta_2\delta_5} \frac{1}{A} \sum_{\ell|A} (d,\ell)^{1/2} d_B^{-1/2} d_0^{\theta}; \qquad (12.4)$$

here and throughout the implied dependence on $\mathbf{u}$ is at most polynomial.

At this point, the proof of Lemma 12.1 has reduced to elementary estimates with arithmetic functions. The factorization $d = (d,\ell)d_B d_0$ depends on ℓ, so it takes some work to estimate the sum over ℓ. To this end, we also factor d in an alternative way, independently of ℓ, by $d = d' fgh$ where $(d', AB) = 1$, $f \mid A^{\infty}$, $(f, B) = 1$, $g \mid B^{\infty}$, $(g, A) = 1$, and $h \mid A^{\infty}$ and $h \mid B^{\infty}$. Note that $(f, g) = (f, h) = (g, h) = 1$. Then writing the old variables in terms of these, we have

$$d = \underbrace{(f,\ell)(h,\ell)}_{(d,\ell)}\, \underbrace{g\frac{h}{(h,\ell)}}_{d_B}\, \underbrace{d'\frac{f}{(f,\ell)}}_{d_0}.$$

Inserting this into (12.4), and summing over $\ell \mid L$, we obtain

$$\sum_{\substack{t_j \text{ level } \delta_2\delta_5 \\ |t_j|\leq T}} |Z_j(\mathbf{u})| \ll_{\mathbf{u},\varepsilon} T^{2+\varepsilon} q^{\varepsilon} d'^{\theta}\left(\frac{(D,\delta_4)}{D}\right)^{1/2} \sum_{AB=\delta_2\delta_5} \frac{1}{A} \frac{(f,A)^{1/2-\theta} f^{\theta}(h,A)}{\sqrt{gh}}.$$

Writing $d' f = \frac{D}{(D,\delta_4)}\delta_4\frac{1}{gh}$, we obtain

$$\sum_{\substack{t_j \text{ level } \delta_2\delta_5 \\ |t_j|\leq T}} |Z_j(\mathbf{u})| \ll_{\mathbf{u},\varepsilon} T^{2+\varepsilon} q^{\varepsilon}\left(\frac{(D,\delta_4)}{D}\right)^{1/2-\theta} \delta_4^{\theta} \sum_{AB=\delta_2\delta_5} \frac{1}{A} \frac{(f,A)^{1/2-\theta}(h,A)}{(gh)^{1/2+\theta}}.$$

Now let us pause to gauge our progress towards (11.9). The inner sum over A easily gives $O(q^{\varepsilon})$, and so if we trivially bound this part, and use $D = qk_1k_1^*$ (also recall $(k_1,\delta_4) = 1$ from (9.3)), we get the bound

$$\sum_{t_j} |Z_j(\mathbf{u})| \ll_{\mathbf{u},\varepsilon} q^{\theta-1/2}\frac{1}{(k_1k_1^*)^{1/2}}(q,\delta_4)^{1/2-\theta}(k_1k_1^*)^{\theta}\delta_4^{\theta}T^2q^{\varepsilon}.$$

so we need to save $\delta_4^{1/2+\theta}(k_1k_1^*)^{\theta}$, which will come from better estimating the sum over A.

This inner sum over A, B may be factored into prime powers. For the primes $p \nmid k_1\delta_4$, all we use is that the local factor is ≤ 1 (leading to an $O(q^\varepsilon)$ bound from these primes, by the observation in the previous paragraph). Recall $\delta_4 \mid \delta_2\delta_5$ since $\delta_5 = [\delta_3, \delta_4]$ (see (9.2)), and $\delta_4 \mid d$, so $\delta_4 \mid fgh$. For $p \mid \delta_4$, say $p^\nu \parallel \delta_4$, $p^f \parallel f$, and so on for g, h, A, and B, by an abuse of notation. Now the variables in the exponents are written additively. Since $\delta_4 \mid fgh$, in additive notation we have $f + g + h \geq \nu$. Also, $A + B \geq \nu$, since $\delta_4 \mid \delta_2\delta_5$.

In the case $B = 0$, we have $g = h = 0$ and the local factor is

$$\frac{1}{p^A}(p^f, p^A)^{1/2-\theta} \leq p^{\nu(-1/2-\theta)},$$

which is the local factor of $\delta_4^{-1/2-\theta}$. In the case $A = 0$, the local factor is no larger than the local factor of $\delta_4^{-1/2-\theta}$, as can be seen easily. Finally, if $A, B > 0$, then $f = g = 0$, $h \geq \nu$, and so the local factor equals

$$\frac{1}{p^A}\min(p^h, p^A)p^{-h(1/2+\theta)} \leq p^{-\nu(1/2+\theta)}.$$

Now suppose that $p \mid k_1$. Then since $k_1 \mid \delta_2$ and $k_1 \mid d$ (whence $k_1 \mid fgh$), essentially the same proof used for δ_4 shows the local factors for primes dividing k_1 give $O(k_1^{-1/2})$.

In summary, this shows

$$\sum_{\substack{t_j \text{ level } \delta_2\delta_5 \\ |t_j| \leq T}} |Z_j(\mathbf{u})| \ll_{\mathbf{u},\varepsilon} T^2 q^\varepsilon \left(\frac{(q, \delta_4)}{qk_1k_1^*}\right)^{1/2-\theta} (k_1\delta_4)^{-1/2}.$$

Using $(k_1k_1^*)^\theta \leq k_1^{2\theta} \leq k_1^{1/2}$ gives (11.9), as desired.

12.2. *Continuous spectrum*

In this section we prove Lemma 11.2.

12.2.1. Fourier coefficients of Eisenstein series. Here we quote an explicit evaluation of the Fourier coefficients of $\phi_{\mathfrak{a}\mathfrak{c}}(n, u)$, where $\mathfrak{a} = 1/r$ is an Atkin–Lehner cusp, and $\mathfrak{c}$ is an arbitrary cusp of $\Gamma_0(N)$. The proof appears in [KY19]. Let $\mathfrak{c} = v/f$ where $f \mid N$, $(v, f) = 1$, and v runs modulo $(f, N/f)$; by [Iwa97, Proposition 2.6], every cusp $\mathfrak{c}$ may be represented in this form. Let

$$N' = N/f, \qquad N'' = N'/(f, N'), \tag{12.5}$$

and write

$$f_r = (f, r), \qquad f_s = (f, s), \qquad r = f_r r', \qquad s = f_s s'.$$

In addition, write

$$f_r = f_r' f_0, \quad \text{where} \quad (f_0, r') = 1 \text{ and } f_r' \mid (r')^\infty,$$

and similarly

$$s' = s_f' s_0, \quad \text{where} \quad (s_0, f_s) = 1 \text{ and } s_f' \mid f_s^\infty.$$

Proposition 12.2. *Let notation be as above. Then* $\phi_{\mathfrak{a}\mathfrak{c}}(n,u)=0$ *unless*

$$n=\frac{f_r'}{(f_r',r')}\frac{s_f'}{(s_f',f_s)}k$$

for some integer k. *In this case, write* $k=k_rk_s\ell$, *where*

$$k_r=(k,(f_r',r')),\qquad k_s=(k,(s_f',f_s)).$$

Then

$$\phi_{\mathfrak{a}\mathfrak{c}}(n,u)=\frac{S(\ell,0;s_0f_0)}{(N''sf_r^2)^u}\frac{f_r'}{(f_r',r')}\frac{s_f'}{(s_f',f_s)}\sum_{\substack{d\mid k\\(d,f_sr')=1}}d^{1-2u}\frac{1}{\varphi\big(\frac{(f_r',r')}{k_r}\big)}\frac{1}{\varphi\big(\frac{(s_f',f_s)}{k_s}\big)}\times$$
$$\sum_{\chi\ \left(\mathrm{mod}\ \frac{(f_r',r')}{k_r}\right)}\ \sum_{\psi\ \left(\mathrm{mod}\ \frac{(s_f',f_s)}{k_s}\right)}\frac{(\chi\psi)(\ell)\tau(\overline{\chi})\tau(\overline{\psi})}{L(2u,\overline{\chi^2\psi^2}\chi_0)}(\chi\psi)(\overline{s_0f_0d^2v})\chi(-k_s\overline{(s_f',f_s)})\psi(k_r\overline{(f_r',r')}),\tag{12.6}$$

where χ_0 *is the principal character modulo* f_sr'.

For later calculations, it will be useful to notice that the condition $(k/k_r,(f_r',r')/k_r)=1$ (and similarly in the s-aspect) is automatic from the presence of $(\chi\psi)(\ell)$. Moreover, we have $(f_r',r')=(f_r,r/f_r)$, and similarly $(s_f',f_s)=(f_s,s/f_s)$, and so $(f_r',r')(s_f',f_s)=(f,N/f)$. Note the condition $d\mid k$ together with $(d,f_sr')=1$ implies $d\mid\ell$.

12.2.2. Proof of Lemma 11.2. By (3.10) and (3.12), we have

$$\nu_{\mathfrak{a}\mathfrak{b}}(n,u)=\alpha(u)\phi_{\mathfrak{a}\mathfrak{b}}(n,u)|n|^{u-1/2},\quad\text{where}\quad\alpha(u)=\frac{2\pi^{u+1/2}}{\Gamma(u)(\cos(\pi(u-1/2)))^{1/2}}.$$

Note that $|\alpha(1/2+it)|$ is independent of $t\in\mathbb{R}$. Define

$$Z_1=Z_1(\alpha,\beta)=\sum_{m,n\geq1}\frac{\nu_{1/r,\mathfrak{c}}(mn,1/2+it)}{m^\alpha n^\beta},$$
$$Z_2=Z_2(\alpha,\beta)=\sum_{\substack{m,n\geq1\\\delta m\equiv0\ (\mathrm{mod}\ D)\\(m,Q)=1}}\frac{\nu_{\infty,\mathfrak{c}}(\delta mn,1/2+it)}{m^\alpha n^\beta},$$

where we assume the level is N as in Section 12.2.1. This meaning of N is valid only within the confines of this subsection, and hence should not be confused with the dyadic variable N in the rest of the article. For Lemma 11.2, we shall need $N=\delta_2\delta_5$, $Q=fg_0m_1'$, $D=qk_1k_1^*$, $\delta=\delta_4$, and $r=\delta_5$, but we do not need to make these specifications yet. With this notation, we have $Z_{\mathfrak{c},t}(\mathbf{u})=Z_1(u_2,u_3)Z_2(u_1,u_4)$.

The plan of the proof is to first derive bounds on Z_1^0,Z_1^*,Z_2^0,Z_2^* individually, and follow this with estimates for the sums over $\mathfrak{c}$.

Using Proposition 12.2, we have (with $u = 1/2 + it$)

$$Z_1 = \frac{f_r'}{(f_r', r')} \frac{s_f'}{(s_f', f_s)} \frac{\alpha(u)}{(N'' s f_r^2)^u} \sum_{k_r | (f_r', r')} \sum_{k_s | (s_f', f_s)} \frac{1}{\varphi\big(\frac{(f_r', r')}{k_r}\big)} \frac{1}{\varphi\big(\frac{(s_f', f_s)}{k_s}\big)}$$
$$\times \sum_{\chi \left(\mathrm{mod}\ \frac{(f_r', r')}{k_r}\right)} \sum_{\psi \left(\mathrm{mod}\ \frac{(s_f', f_s)}{k_s}\right)} \frac{\tau(\overline{\chi})\tau(\overline{\psi})}{L(2u, \overline{\chi^2\psi^2}\chi_0)} (\chi\psi)(\overline{s_0 f_0 w'}) \chi(-k_s \overline{(s_f', f_s)}) \psi(k_r \overline{(f_r', r')})$$
$$\times \sum_{(d, f_s r')=1} (\overline{\chi\psi})(d^2) d^{1-2u} \sum_{\substack{m,n \geq 1 \\ (*)}} \frac{(\chi\psi)(\ell) S(\ell, 0; s_0 f_0)}{m^\alpha n^\beta} (mn)^{u-1/2}.$$

Here $(*)$ stands for the following conditions: $mn = \frac{f_r'}{(f_r', r')} \frac{s_f'}{(s_f', f_s)} k_r k_s \ell$ and $\ell \equiv 0 \pmod{d}$. Write $Z_1 = Z_1^0 + Z_1^*$ where Z_1^0 corresponds to the part with both χ and ψ principal.

By a trivial bound, we have for $\mathrm{Re}(\alpha), \mathrm{Re}(\beta) \geq \sigma > 1$,

$$|Z_1^0| \ll_\sigma \frac{N^\varepsilon}{f_r \sqrt{sN''}\, (f_r', r')(s_f', f_s)} \frac{1}{|\zeta(1+2it)|} \ll_\sigma \frac{(N(1+|t|))^\varepsilon}{(f, N/f) f_r \sqrt{sN''}}. \tag{12.7}$$

Meanwhile, we have

$$|Z_1^*| \ll \frac{f_r'}{(f_r', r')} \frac{s_f'}{(s_f', f_s)} \frac{N^\varepsilon}{f_r \sqrt{sN''}} \sum_{k_r | (f_r', r')} \sum_{k_s | (s_f', f_s)} \frac{1}{\varphi\big(\frac{(f_r', r')}{k_r}\big)} \frac{1}{\varphi\big(\frac{(s_f', f_s)}{k_s}\big)}$$
$$\times \sideset{}{'}\sum_{\chi \left(\mathrm{mod}\ \frac{(f_r', r')}{k_r}\right)} \sideset{}{'}\sum_{\psi \left(\mathrm{mod}\ \frac{(s_f', f_s)}{k_s}\right)} \frac{|\tau(\overline{\chi})\tau(\overline{\psi})|}{|L(1+2it, \overline{\chi^2\psi^2}\chi_0)|} |Y_1|,$$

where the notation $\sum'$ means the principal character is omitted, and with

$$V = \frac{f_r'}{(f_r', r')} \frac{s_f'}{(s_f', f_s)} k_r k_s$$

we let

$$Y_1 = \sum_{(d, f_s r')=1} (\overline{\chi\psi})(d^2) d^{1-2u} \sum_{mn \equiv 0 \ (\mathrm{mod}\ dV)} \frac{(\chi\psi)\big(\frac{mn}{V}\big) S\big(\frac{mn}{V}, 0; s_0 f_0\big)}{m^\alpha n^\beta} (mn)^{u-1/2}.$$

Moreover, the analytic continuation of Z_1^* will be inherited from that of Y_1.

Similarly to (7.4) and (7.5), one can show the formal identity

$$\sum_{mn \equiv 0 \ (\mathrm{mod}\ D)} f(m,n) = \sum_{CAB=D} \mu(C) \sum_{m,n} f(CAm, CBn). \tag{12.8}$$

Applying this to Y_1 with $D = dV$, we obtain

$$Y_1 = V^{u-1/2} \sum_{(d, f_s r')=1} (\overline{\chi\psi})(d) d^{1/2-u} \sum_{CAB=dV} \frac{\mu(C)(\chi\psi)(C) C^{u-1/2}}{A^\alpha B^\beta C^{\alpha+\beta}} \times \sum_{m,n} \frac{(\chi\psi)(mn) S(Cdmn, 0; s_0 f_0)}{m^\alpha n^\beta (mn)^{1/2-u}}.$$

One can readily observe that $(d, V) = 1$ and $(V, s_0 f_0) = 1$. In the factorization $CAB = dV$, one may split each of C, A, B uniquely into its part dividing d and dividing V separately, and thereby factor the sum. In this way, we obtain (with $\mathrm{Re}(u) = 1/2$)

$$|Y_1| \ll \frac{N^\varepsilon}{V^\sigma} \left| \sum_{(CAB, f_s r')=1} \frac{\mu(C)(\overline{\chi\psi})(AB)}{A^\alpha B^\beta C^{\alpha+\beta} (AB)^{u-1/2}} \sum_{m,n} \frac{(\chi\psi)(mn) S(C^2 ABmn, 0; s_0 f_0)}{m^\alpha n^\beta (mn)^{1/2-u}} \right|.$$

Next we open the Ramanujan sum as a divisor sum, which gives

$$|Y_1| \ll \frac{N^\varepsilon}{V^\sigma} \sum_{g | s_0 f_0} g \left| \sum_{(CAB, f_s r')=1} \sum_{\substack{m,n \\ C^2 ABmn \equiv 0 \pmod g}} \frac{\mu(C)(\overline{\chi\psi})(AB)}{A^\alpha B^\beta C^{\alpha+\beta} (AB)^{u-1/2}} \frac{(\chi\psi)(mn)}{m^\alpha n^\beta (mn)^{1/2-u}} \right|.$$

Now it is not difficult to see the analytic continuation of Y_1 to $\mathrm{Re}(\alpha, \beta) \geq \sigma = 1/2 + \varepsilon$, and therein we obtain the bound

$$|Y_1| \ll N^\varepsilon \left(\frac{(f_r', r')}{f_r'} \frac{(s_f', f_s)}{s_f'} \right)^{1/2} \frac{(s_0 f_0)^{1/2}}{(k_r k_s)^{1/2}} \times |L(\alpha - it, \chi\psi) L(\alpha + it, \overline{\chi\psi}) L(\beta - it, \chi\psi) L(\beta + it, \overline{\chi\psi})|.$$

We recall the well-known bound on the fourth moment of Dirichlet L-functions:

$$\sum_{\chi \pmod N} |L(\sigma + it, \chi)|^4 \ll_{\sigma,\varepsilon} (1 + |t|)^{1+\varepsilon} N^{1+\varepsilon} \tag{12.9}$$

for $\sigma \geq 1/2$. Moreover, we have the hybrid version

$$\int_{|t| \leq T} \sum_{\chi \pmod N} |L(\sigma + it, \chi)|^4 \ll_{\sigma,\varepsilon} (1 + T)^{1+\varepsilon} N^{1+\varepsilon}. \tag{12.10}$$

For references, consult [Mon71, Chapter 10] or [Gal70, Theorem 2]; the statements in these sources do not precisely claim these bounds, but the methods can be easily modified. In addition, we have

$$\frac{1}{|L(1 + 2it, \chi)|} \ll (1 + |t|)^\varepsilon N^\varepsilon,$$

for which see [MV07, Theorem 11.4]. Thus applying Hölder's inequality and the bound (12.9) we obtain

$$|Z_1^*| \ll_{\alpha,\beta,\varepsilon} \frac{N^\varepsilon(1+|t|)^{1+\varepsilon}}{f_r\sqrt{sN''}}(f_r s')^{1/2} = \frac{N^\varepsilon(1+|t|)^{1+\varepsilon}}{\sqrt{fN''}}. \tag{12.11}$$

Using (12.10), we alternatively have

$$\int_{|t|\le T} |Z_1^*| \ll_{\alpha,\beta,\varepsilon} \frac{q^\varepsilon T^{1+\varepsilon}}{\sqrt{fN''}}. \tag{12.12}$$

Next we study Z_2, which is more difficult than Z_1. We have $\mathfrak{a} = \infty \sim 1/N$, so $r = N$ and $s = 1$, and also $f_r = f$, $r' = N'$. First we perform a minor simplification by writing the congruence $\delta m \equiv 0 \pmod{D}$ as $m \equiv 0 \pmod{\frac{D}{(\delta,D)}}$ (so necessarily $(Q, \frac{D}{(\delta,D)}) = 1$ as otherwise the sum is empty). Then we have

$$Z_2 = \left(\frac{(D,\delta)}{D}\right)^\alpha Y_2, \quad \text{where} \quad Y_2 := \sum_{\substack{m,n\ge 1\\(m,Q)=1}} \frac{\nu_{\infty,\mathfrak{c}}(amn, 1/2+it)}{m^\alpha n^\beta},$$

and where

$$a = \frac{\delta D}{(\delta, D)} = [\delta, D].$$

Applying Proposition 12.2, we obtain

$$Y_2 = \frac{f_N'}{(f_N', N')}\frac{\alpha(u)}{(N''f^2)^u} \sum_{k_N|(f_N',N')} \frac{1}{\varphi((f_N',N')/k_N)} \sum_{\chi \pmod{(f_N',N')/k_N}} \frac{\tau(\overline{\chi})\chi(-\overline{f_0}w')}{L(2u, \overline{\chi^2}\chi_0)}$$
$$\times \sum_{(d,N')=1} d^{1-2u}\overline{\chi}(d^2) \sum_{(*)} \frac{\chi(\ell)S(\ell,0;f_0)}{m^\alpha n^\beta}(amn)^{u-1/2}.$$

Now $(*)$ stands for the following conditions: $amn = \frac{f_N'}{(f_N',N')}k_N\ell$, $amn \equiv 0 \pmod{d}$, and $(m, Q) = 1$. From the condition $(d, N') = 1$, we equivalently obtain $\ell \equiv 0 \pmod{d}$. Define

$$b = \frac{f_N'}{(f_N', N')}k_N = \frac{f_N'}{(f_N', N')/k_N}, \tag{12.13}$$

and write

$$a = (a,b)a' \quad \text{and} \quad b = (a,b)b', \tag{12.14}$$

so that the condition $b \mid amn$ is equivalent to $b' \mid mn$. Then $\ell = a'\frac{mn}{b'}$, and we have

$$Y_2 = \frac{f_N'}{(f_N', N')}\frac{\alpha(u)a^{u-1/2}}{(N''f^2)^u}$$
$$\times \sum_{k_N|(f_N',N')} \frac{1}{\varphi((f_N',N')/k_N)} \sum_{\chi \pmod{(f_N',N')/k_N}} \frac{\tau(\overline{\chi})\chi(\overline{f_0}w'a')}{L(2u, \overline{\chi^2}\chi_0)} X_2,$$

where

$$X_2 := \sum_{(d,N')=1} d^{1-2u}\overline{\chi}(d^2) \sum_{\substack{mn\equiv 0\,(\mathrm{mod}\, b')\\ amn\equiv 0\,(\mathrm{mod}\, d)}} \frac{\chi(mn/b')S(a'mn/b',0;\,f_0)}{m^\alpha n^\beta}(mn)^{u-1/2}.$$

Here $(b', f_0 d) = 1$, since $b \mid (N')^\infty$. Opening the Ramanujan sum, we have

$$X_2 = \sum_{e|f_0} e\mu(f_0/e) \sum_{(d,N')=1} d^{1-2u}\overline{\chi}(d^2) \sum_{\substack{mn\equiv 0\,(\mathrm{mod}\, b')\\ amn\equiv 0\,(\mathrm{mod}\, d)\\ a'mn\equiv 0\,(\mathrm{mod}\, e)}} \frac{\chi(mn/b')}{m^\alpha n^\beta}(mn)^{u-1/2}.$$

Let $g = (a, d)$, so that

$$X_2 = \sum_{\substack{g|a\\ (g,N')=1}} g^{1-2u}\overline{\chi}(g^2) \sum_{e|f_0} e\mu(f_0/e) \sum_{(d,N'a/g)=1} d^{1-2u}\overline{\chi}(d^2)$$
$$\times \sum_{\substack{mn\equiv 0\,(\mathrm{mod}\, b')\\ mn\equiv 0\,(\mathrm{mod}\, d)\\ a'mn\equiv 0\,(\mathrm{mod}\, e)}} \frac{\chi(mn/b')}{m^\alpha n^\beta}(mn)^{u-1/2}.$$

Applying (12.8) to X_2 with the modulus d, we obtain

$$X_2 = \sum_{\substack{g|a\\ (g,N')=1}} g^{1-2u}\overline{\chi}(g^2) \sum_{e|f_0} e\mu(f_0/e) \sum_{(d,N'a/g)=1} d^{1-2u}\overline{\chi}(d^2) \sum_{CAB=d} \mu(C)$$
$$\times \sum_{\substack{Cdmn\equiv 0\,(\mathrm{mod}\, b')\\ a'Cdmn\equiv 0\,(\mathrm{mod}\, e)}} \frac{\chi(Cdmn/b')}{(CAm)^\alpha (CBn)^\beta}(Cdmn)^{u-1/2}.$$

Since $C \mid d$, $(d, N') = 1$, and $b' \mid (N')^\infty$, the congruence $Cdmn \equiv 0 \pmod{b'}$ is equivalent to $mn \equiv 0 \pmod{b'}$. We can then write

$$X_2 = \sum_{\substack{g|a\\ (g,N')=1}} g^{1-2u}\overline{\chi}(g^2) \sum_{e|f_0} e\mu(f_0/e) \sum_{(CAB,N'a/g)=1} \frac{\mu(C)\overline{\chi}(AB)(AB)^{1/2-u}}{C^{\alpha+\beta}A^\alpha B^\beta}$$
$$\times \sum_{\substack{mn\equiv 0\,(\mathrm{mod}\, b')\\ a'C^2ABmn\equiv 0\,(\mathrm{mod}\, e)}} \frac{\chi(mn/b')}{m^\alpha n^\beta}(mn)^{u-1/2}.$$

Next we use (12.8) again, this time on the congruence modulo b', which gives

$$X_2 = (b')^{u-1/2} \sum_{\substack{g|a\\(g,N')=1}} g^{1-2u}\overline{\chi}(g^2) \sum_{e|f_0} e\mu(f_0/e) \sum_{xyz=b'} \frac{\mu(x)\chi(x)x^{u-1/2}}{x^{\alpha+\beta}y^{\alpha}z^{\beta}}$$
$$\times \sum_{(CAB,N'a/g)=1} \frac{\mu(C)\overline{\chi}(AB)(AB)^{1/2-u}}{C^{\alpha+\beta}A^{\alpha}B^{\beta}} \sum_{a'C^2ABxb'mn\equiv 0 \,(\mathrm{mod}\, e)} \frac{\chi(mn)}{m^{\alpha}n^{\beta}}(mn)^{u-1/2}.$$

Similarly to the Z_1 case, one can see the meromorphic continuation with a pole only in case χ is principal. In addition, we have the bound (with $u = 1/2 + it$)

$$|X_2| \ll_{\sigma} \frac{N^{\varepsilon}}{(b')^{\sigma}} |L(\alpha - it, \chi)L(\alpha + it, \overline{\chi})L(\beta - it, \chi)L(\beta + it, \overline{\chi})| \sum_{e|f_0} e\left(\frac{(a', e)}{e}\right)^{\sigma}.$$

Note

$$\sum_{e|f_0} e\left(\frac{(a', e)}{e}\right)^{\sigma} \ll N^{\varepsilon}(a', f_0)^{\sigma}(1 + f_0)^{1-\sigma}.$$

Now write $Z_2 = Z_2^0 + Z_2^*$ where Z_2^0 corresponds to the principal characters, and similarly write $Y_2 = Y_2^0 + Y_2^*$. For $\mathrm{Re}(\alpha, \beta) \geq \sigma = 1 + \varepsilon$, we have trivially

$$X_2 \ll N^{\varepsilon}\frac{(a', f_0)}{b'} = N^{\varepsilon}\frac{(a, f_0 b)}{b},$$

recalling (12.14). Thus

$$Y_2^0 \ll \frac{f'_N}{(f'_N, N')} \frac{N^{\varepsilon}}{\sqrt{N'' f^2}} \sum_{k_N|(f'_N, N')} \frac{k_N}{(f'_N, N')} \frac{(f'_N, N')(a, f_0 b)}{f'_N k_N}.$$

Here b is a function of k_N (cf. (12.13)), and is maximal when $k_N = (f'_N, N')$, in which case $b = f'_N$. Recalling $f_0 f'_N = f$, which implies $(a, f_0 b) \leq (a, f)$, and using $(f'_N, N') = (f, N/f)$, in all we obtain

$$Y_2^0 \ll \frac{N^{\varepsilon}(a, f)}{(f, N/f)\sqrt{N'' f^2}} \frac{1}{|\zeta(1 + 2it)|}.$$

Finally,

$$Z_2^0 \ll \frac{N^{\varepsilon}(1 + |t|)^{\varepsilon}}{(f, N/f) f \sqrt{N''}} ([\delta, D], f)\frac{(\delta, D)}{D}. \tag{12.15}$$

Using (12.9) and Hölder's inequality, we have, with $\sigma = 1/2 + \varepsilon$,

$$|Y_2^*| \ll_{\alpha,\beta,\varepsilon} \frac{f'_N}{(f'_N, N')} \frac{N^{\varepsilon}(1 + |t|)^{1+\varepsilon}}{(N'' f^2)^{1/2}} \sum_{k_N|(f'_N, N')} \left(\frac{(f'_N, N')}{k_N}\right)^{1/2} \frac{(a', f_0)^{1/2} f_0^{1/2}}{b'^{1/2}}.$$

Note the simplification

$$\frac{f'_N}{(f'_N, N')}\left(\frac{(f'_N, N')}{k_N}\right)^{1/2}\frac{f_0^{1/2}}{b'^{1/2}} = f^{1/2}\frac{(a,b)^{1/2}}{k_N}.$$

We have $(a', f_0) = (a, f_0)$ since $(b, f_0) = 1$, and $(a,b)^{1/2}/k_N \le (a, f'_N)^{1/2}$. Thus

$$|Y_2^*| \ll_{\alpha,\beta,\varepsilon} \frac{N^\varepsilon(1+|t|)^{1+\varepsilon}}{(N''f)^{1/2}}(a, f)^{1/2}.$$

Hence

$$|Z_2^*| \ll_{\alpha,\beta,\varepsilon} \left(\frac{(\delta, D)}{D}\right)^{1/2}\frac{N^\varepsilon(1+|t|)^{1+\varepsilon}}{(N''f)^{1/2}}\left(\frac{\delta D}{(\delta, D)}, f\right)^{1/2}. \tag{12.16}$$

Also, in a similar way to the Z_1^* case,

$$\int_{|t|\le T}|Z_2^*|\,dt \ll_{\alpha,\beta,\varepsilon} \left(\frac{(\delta, D)}{D}\right)^{1/2}\frac{N^\varepsilon T^{1+\varepsilon}}{(N''f)^{1/2}}\left(\frac{\delta D}{(\delta, D)}, f\right)^{1/2}. \tag{12.17}$$

Now we proceed to prove the desired bounds in Lemma 11.2. The cusps may be parameterized by v/f with $f \mid N$ and $v \pmod{(f, N/f)}$ (with v coprime to f). During the course of the proof, it will be helpful to refer to the following divisor-sum bounds:

$$\sum_{d|N}\frac{(d, N/d)^2}{N} \ll N^\varepsilon\frac{\mathrm{flrt}(N)}{\sqrt{N}} \quad\text{and}\quad \sum_{d|N}\frac{(d, N/d)^2 d^{1/2}}{N} \ll N^\varepsilon\frac{\mathrm{flrt}(N)^{3/2}}{\sqrt{N}}, \tag{12.18}$$

each of which can be checked prime-by-prime by multiplicativity. If desired, the former inequality could be bounded by $N^\varepsilon \mathrm{flrt}(N)^2/N$. Along the same lines, we note

$$\sum_{d|N}\frac{(d, N/d)^2}{d^{1/2}N} \ll N^\varepsilon\frac{\mathrm{flrt}(N)^{3/2}}{N}, \tag{12.19}$$

as well as

$$\sum_{d|N}\frac{(d, N/d)d^{1/2}}{N} \ll N^\varepsilon\frac{\sqrt{\mathrm{flrt}(N)}}{\sqrt{N}} \quad\text{and}\quad \sum_{d|N}\frac{(d, N/d)}{N} \ll N^\varepsilon\frac{\mathrm{flrt}(N)}{N}. \tag{12.20}$$

Combining (12.12) and (12.16), we obtain

$$\int_{|t|\le T}\sum_{\mathfrak{c}}|Z_1^*Z_2^*|\,dt \ll_{\mathbf{u},\varepsilon} N^\varepsilon T^{2+\varepsilon}\left(\frac{(\delta, D)}{D}\right)^{1/2}\sum_{f|N}\frac{(f, N/f)^2}{N}\left(\frac{\delta D}{(\delta, D)}, f\right)^{1/2}.$$

We have $N = \delta_2\delta_5$, $D = qk_1k_1^*$, $\delta = \delta_4 \mid \delta_5$. With these substitutions, and recalling $(\delta_4, k_1) = 1$, we obtain

$$\int_{|t|\le T}\sum_{\mathfrak{c}}|Z_1^*Z_2^*|\,dt \ll_{\mathbf{u},\varepsilon} q^\varepsilon T^{2+\varepsilon}\left(\frac{(\delta_4, q)}{qk_1k_1^*}\right)^{1/2}\sum_{f|\delta_2\delta_5}\frac{(f, \delta_2\delta_5/f)^2}{\delta_2\delta_5}\left(k_1k_1^*\frac{\delta_4 q}{(\delta_4, q)}, f\right)^{1/2}.$$

By multiplicativity, and using $(\delta_2, \delta_5) = 1 = (\delta_2, q)$, and $k_1 \mid \delta_2$, the inner sum over f factors and simplifies as

$$\sum_{f|\delta_2\delta_5} \frac{(f, \delta_2\delta_5/f)^2}{\delta_2\delta_5} \left(k_1 k_1^* \frac{\delta_4 q}{(\delta_4, q)}, f\right)^{1/2} = \left(\sum_{g|\delta_2} \frac{(g, \delta_2/g)^2 (g, k_1k_1^*)^{1/2}}{\delta_2}\right)\left(\sum_{h|\delta_5} \frac{(h, \delta_5/h)^2}{\delta_5}\left(\frac{\delta_4 q}{(\delta_4, q)}, h\right)^{1/2}\right).$$

Using $(g, k_1k_1^*)^{1/2} \leq \sqrt{k_1k_1^*}$, $\left(\frac{\delta_4 q}{(\delta_4, q)}, h\right)^{1/2} \leq h^{1/2}$, and (12.18), we obtain

$$\int_{|t|\leq T} \sum_{\mathfrak{c}} |Z_1^* Z_2^*|\, dt \ll_{\mathbf{u},\varepsilon} N^\varepsilon T^{2+\varepsilon} \left(\frac{(\delta_4, q)}{qk_1k_1^*}\right)^{1/2} \frac{\operatorname{flrt}(\delta_2)\operatorname{flrt}(\delta_5)^{3/2}}{\sqrt{\delta_2\delta_5}} \sqrt{k_1k_1^*}. \tag{12.21}$$

Recall that $\delta_5 = [\delta_3, \delta_4]$, and that δ_4 is square-free. Therefore, $[\delta_3, \delta_4] = \delta_3 \frac{\delta_4}{(\delta_3,\delta_4)}$ where $\left(\delta_3, \frac{\delta_4}{(\delta_3,\delta_4)}\right) = 1$. Since flrt is multiplicative, and trivial on square-free numbers, this implies

$$\operatorname{flrt}(\delta_5) = \operatorname{flrt}(\delta_3)\operatorname{flrt}\left(\frac{\delta_4}{(\delta_3, \delta_4)}\right) = \operatorname{flrt}(\delta_3), \tag{12.22}$$

a simplification we will make repeatedly below. Applying (12.22) to (12.21) gives (11.18).

Combining (12.7) and (12.15) and specializing the variables, we obtain

$$\sum_{\mathfrak{c}} |Z_1^0 Z_2^0| \ll (q(1+|t|))^\varepsilon \left(\frac{(\delta_4, q)}{qk_1k_1^*}\right) \sum_{\mathfrak{c}} \frac{(f, [\delta_4, qk_1k_1^*])}{(f, N/f)^2 N'' s^{1/2} f_r f}.$$

Using (12.5) and summing over $u \pmod{(f, N/f)}$, we obtain

$$\sum_{\mathfrak{c}} \frac{(f, [\delta_4, qk_1k_1^*])}{(f, N/f)^2 N'' s^{1/2} f_r f} \leq \frac{1}{N} \sum_{f|N} \frac{(f, [\delta_4, qk_1k_1^*])}{s^{1/2} f_r}.$$

Using the coprimality conditions, we have $[\delta_4, qk_1k_1^*] = k_1k_1^*[\delta_4, q]$, which in turn divides $k_1k_1^*[\delta_5, q]$. Now k_1 is in the s-part of the level (since $s = \delta_2$ and $k_1 \mid \delta_2$), while we also have $(q, \delta_2) = 1$ by (7.8) so that $[\delta_4, q]$ is coprime to s, and hence to f_s. Now we may see that the sum above factors as

$$\frac{1}{N} \sum_{f|N} \frac{(f, [\delta_4, qk_1k_1^*])}{s^{1/2} f_r} \leq \left(\sum_{f_r|r} \frac{(f_r, [\delta_4, q])}{r f_r}\right)\left(\sum_{f_s|s} \frac{(f_s, k_1k_1^*)}{s^{3/2}}\right).$$

Using $(f_r, [\delta_4, q]) \leq f_r$ and $(f_s, k_1k_1^*) \leq f_s^{1/2}\sqrt{k_1k_1^*} \leq \sqrt{sk_1k_1^*}$ leads immediately to (11.19).

Finally, we examine the two cross terms. From (12.12) and (12.15), and using simplifications as in the above cases, we have

$$\int_{|t|\leq T} \sum_{\mathfrak{c}} |Z_1^* Z_2^0|\, dt \ll_{\mathbf{u},\varepsilon} \left(\frac{(\delta_4, q)}{qk_1k_1^*}\right) q^\varepsilon T^{1+\varepsilon} \sum_{f|\delta_2\delta_5} \frac{(f, \delta_2\delta_5/f)}{\delta_2\delta_5\sqrt{f}} (f, k_1k_1^*[\delta_4, q]).$$

The inner sum factors as

$$\Big(\sum_{g|\delta_2} \frac{(g,\delta_2/g)}{\delta_2\sqrt{g}}(g,k_1k_1^*)\Big)\Big(\sum_{h|\delta_5} \frac{(h,\delta_5/h)}{\delta_5\sqrt{h}}(h,[\delta_4,q])\Big).$$

Using $(g,k_1k_1^*) \le \sqrt{gk_1k_1^*}$, $(h,[\delta_4,q]) \le h$, and (12.20), we have

$$\sum_{f|\delta_2\delta_5} \frac{(f,\delta_2\delta_5/f)}{\delta_2\delta_5\sqrt{f}}(f,k_1k_1^*[\delta_4,q]) \ll q^{\varepsilon} \frac{\sqrt{\mathrm{flrt}(\delta_5)}\,\mathrm{flrt}(\delta_2)}{\delta_2\sqrt{\delta_5}}\sqrt{k_1k_1^*}.$$

On account of (12.22), (11.20) follows.

Similarly, combining (12.7) and (12.16), we have

$$\int_{|t|\le T} \sum_{\mathbf{c}} |Z_1^0 Z_2^*|\,dt \ll_{\mathbf{u},\varepsilon} \Big(\frac{(\delta_4,q)}{qk_1k_1^*}\Big)^{1/2} q^{\varepsilon}T^{1+\varepsilon} \sum_{f|N} \frac{f(f,N/f)\big(k_1k_1^*\frac{\delta_4 q}{(\delta_4,q)},f\big)^{1/2}}{f_r N\sqrt{sf}}.$$

Following the discussion of the $Z_1^0Z_2^0$ case (recall $r=\delta_5$ and $s=\delta_2$), the inner sum over f factors as

$$\Big(\sum_{g|\delta_2} \frac{(g,\delta_2/g)(k_1k_1^*,g)^{1/2}g^{1/2}}{\delta_2^{3/2}}\Big)\Big(\sum_{h|\delta_5} \frac{(h,\delta_5/h)\big(\frac{\delta_4 q}{(\delta_4,q)},h\big)^{1/2}}{\delta_5\sqrt{h}}\Big).$$

Using (12.20) and (12.22), we find that this is

$$\ll q^{\varepsilon}\sqrt{k_1k_1^*}\,\frac{\sqrt{\mathrm{flrt}(\delta_2)}\,\mathrm{flrt}(\delta_3)}{\delta_2\delta_5}.$$

Hence we obtain (11.21) (in fact, with a slightly better power of $\mathrm{flrt}(\delta_2)$).

13. Zero terms

13.1. Overview

In this section, we analyze the contribution to $\mathcal{S}$ from the terms with some p_i zero. Recall the original expression for $\mathcal{S}'''$ from (7.11), and Proposition 8.1.

Let us write

$$\mathcal{S}''' = \sum_{\mathrm{P}} \mathcal{T}_{\mathrm{P}} + \mathcal{S}'''_{0,0,0} + \mathcal{S}'''_{0,0} + \mathcal{S}'''_{0},$$

where $\sum_{\mathrm{P}} \mathcal{T}_{\mathrm{P}}$ corresponds to the terms with all p_i nonzero, $\mathcal{S}'''_{0,0,0}$ corresponds to the terms with all three p_i zero, $\mathcal{S}'''_{0,0}$ corresponds to the terms with exactly two p_i zero, and finally $\mathcal{S}'''_{0}$ has the terms with exactly one p_i zero. Recall that the sum over P is the sum over the dyadic partitions of unity. The partition is mainly beneficial for estimating $\mathcal{T}_{\mathrm{P}}$, and we usually wish to remove the partition as much as possible when estimating the zero terms.

Applying the additional summations that led from $\mathcal{S}$ to $\mathcal{S}'''$ (see (7.10), (7.6), (7.1) or alternatively (13.8) and (13.9) below), we likewise define $\mathcal{S}_{0,0,0}$, $\mathcal{S}_{0,0}$, and $\mathcal{S}_0$. Implicit in the definition of these quantities is that prior to the definition of $\mathcal{S}'''$, we applied a partition of unity. When it is necessary to emphasize this, we may write $\mathcal{S}_{0,0,0}^{(\mathrm{T})}$ where T stands for the tuple $(M_1, M_2, C, N_1, N_2, N_3, K)$, and likewise for $\mathcal{S}_{0,0}$ and $\mathcal{S}_0$. Then $\sum_{\mathrm{T}} \mathcal{S}_{0,0,0}^{(\mathrm{T})}$ represents the quantity after reassembling the partition. For simplicity of notation we may on occasion drop the superscript T.

Our primary goal is to show

Theorem 13.1. *With an appropriate choice of $G_i(s)$ in the approximate functional equations, we have*

$$\sum_{\mathrm{T}} \mathcal{S}_{0,0,0}^{(\mathrm{T})} \ll q^{\varepsilon}.$$

We will show the same bounds for $\mathcal{S}_{0,0}$ and $\mathcal{S}_0$. We make extensive use of the assumption

$$G_i(1/2) = 0. \tag{13.1}$$

Next we specialize Lemma 8.2 to degenerate p_i.

Lemma 13.2. *Let $(\alpha, k) = 1$. If some p_i is zero, then $A(p_1, p_2, p_3; \alpha; k)$ does not depend on α. Furthermore,*

$$\frac{1}{k} A(0, 0, 0; \alpha; k) = (\mathrm{Id} * \varphi)(k), \tag{13.2}$$

where $*$ indicates Dirichlet convolution, $\mathrm{Id}(n) = n$, and φ is Euler's totient function.

This is a short calculation, so we omit the proof. If some p_i is zero then since A does not depend on α, by abuse of notation we may drop α from the notation.

13.2. The case with all p_i zero

The case with $p_1 = p_2 = p_3 = 0$ is surprisingly delicate. It turns out that trivially bounding these terms leads only to $\mathcal{S}_{0,0,0} \ll q^{1/4+\varepsilon}$. Therefore, we have to make use of some further cancellation.

For notational simplicity, let us write $A(0, 0, 0; *; k_0') = A(k_0')$ and $B(0, 0, 0; k_0') = B(k_0')$. We then have

$$\mathcal{S}_{0,0,0}''' = \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \pmod{q k_1 k_1^*}}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \pmod{c_2/(g_0, c_2)}}} \frac{1}{k_0'^3} A(k_0') B(k_0'). \tag{13.3}$$

The function B depends on a choice of a partition of unity in the n_1, n_2, n_3 variables (as well as c, k, m_1, m_2, but here the focus is on the n_i). Our next goal is to recombine the partitions of unity in the dyadic numbers N_1, N_2, N_3.

13.3. Recombining partitions of unity

We write the weight function explicitly. Say

$$J(n_1, n_2, n_3) = J_*(n_1 n_2 n_3, a, m_1', c_0, g_0 k_0', c_2, k_1) F_a\left(\frac{n_1}{\sqrt{q}}, \frac{n_2}{\sqrt{q}}, \frac{n_3}{\sqrt{q}}\right) \frac{\omega\left(\frac{n_1}{N_1}, \frac{n_2}{N_2}, \frac{n_3}{N_3}\right)}{\sqrt{n_1 n_2 n_3}},$$

where $\omega(t_1, t_2, t_3) = \omega(t_1)\omega(t_2)\omega(t_3)$ (recall ω gave rise to the dyadic partition of unity), and

$$J_*(n, \cdot) = e\left(-\frac{nam_1}{ck}\right) \int_0^\infty e\left(\frac{-kt}{c}\right) J_{\kappa-1}\left(\frac{4\pi\sqrt{m_1 nat}}{c}\right) w_{M_2}(t, \cdot)\, \frac{dt}{\sqrt{t}}.$$

Here the weight function w_{M_2} is a piece of a dyadic partition of unity in the m_1, m_2, c, and k variables times $V(m_1 m_2/q)$. The function J_* has $n = n_1 n_2 n_3$ appearing as a block.

By (8.2), we have (introducing subscripts on B now to emphasize the choice of the partition of unity)

$$B_{N_1,N_2,N_3}(k_0') = \iiint_{(\mathbb{R}^+)^3} \frac{J_*(e_1 e_2 \delta_1 t_1 t_2 t_3, a, m_1', c_0, g_0 k_0', c_2, k_1)}{\sqrt{\delta_1 e_1 e_2}} \\ \times F_a\left(\frac{t_1 r_1 e_1}{\sqrt{q}}, \frac{t_2 r_2 e_2}{\sqrt{q}}, \frac{t_3 r_3}{\sqrt{q}}\right) \omega\left(\frac{t_1 e_1 r_1}{N_1}, \frac{t_2 e_2 r_2}{N_2}, \frac{t_3 r_3}{N_3}\right) \frac{dt_1\, dt_2\, dt_3}{\sqrt{t_1 t_2 t_3}}. \tag{13.4}$$

The c, k, m_1, m_2 partitions are implicit in the definition of J_*.

Next we sum over all dyadic numbers $N_1, N_2, N_3 \geq 2^{-1/2}$, and thereby partially reconsititute the partition of unity originally applied in Section 6.1. Note that any error term obtained from the restriction (6.3) is very small. We obtain

$$\sum_{2^{-1/2} \leq N_1, N_2, N_3 \text{ dyadic}} B_{N_1,N_2,N_3}(k_0') \\ = \iiint_{(\mathbb{R}^+)^3} \frac{J_*(t_1 t_2 t_3, a, m_1', c_0, g_0 k_0', c_2, k_1)}{\delta_1 e_1 e_2} \\ \times F_a\left(\frac{t_1}{\sqrt{q}}, \frac{t_2}{\sqrt{q}}, \frac{t_3}{\sqrt{q}}\right) W(t_1, t_2, t_3) \frac{dt_1\, dt_2\, dt_3}{\sqrt{t_1 t_2 t_3}}, \tag{13.5}$$

where $W(t_1, t_2, t_3) = \sum_{2^{-1/2} \leq N_1, N_2, N_3 \text{ dyadic}} \omega(t_1/N_1, t_2/N_2, t_3/N_3)$. Note that the function $1 - W(t_1, t_2, t_3)$ is 0 if $t_i \geq 1$ for all i. It is a slightly subtle point that it is not true that $W(t_1, t_2, t_3) = 1$ for all $t_i > 0$.

Our immediate goal is to replace the W function by 1, and estimate the error. The basic idea is that $1 - W(t_1, t_2, t_3)$ should save a factor $q^{1/4}$ from the fact that at least one of the t_i is ≤ 1, in place of $q^{1/2+\varepsilon}$. Here this numerology comes from $\int_1^{q^{1/2}} t^{-1/2}\, dt \asymp q^{1/4}$, but $\int_0^1 t^{-1/2}\, dt \ll 1$. In light of the claim that the trivial bound on $\mathcal{S}_{0,0,0}$ leads to $O(q^{1/4+\varepsilon})$, one naturally expects that this reasoning should lead to an acceptable final bound. Our next order of business is to confirm this expectation.

Lemma 13.3. *Let*

$$B_\Delta(k_0') = \iiint_{(\mathbb{R}^+)^3} \frac{J_*(t_1t_2t_3, a, m_1', c_0, g_0k_0', c_2, k_1)}{\delta_1 e_1 e_2} \times F_a\left(\frac{t_1}{\sqrt{q}}, \frac{t_2}{\sqrt{q}}, \frac{t_3}{\sqrt{q}}\right)(1 - W(t_1, t_2, t_3)) \frac{dt_1\, dt_2\, dt_3}{\sqrt{t_1t_2t_3}}.$$

Let $\mathcal{S}_\Delta'''$ *be as in* (13.3) *but with* B *replaced with* B_Δ. *Then*

$$\mathcal{S}_\Delta''' \ll q^\varepsilon \frac{(g_0, c_2)}{Cc_2k_1k_1^*} \frac{m_1^{1/2}M_2}{a^{3/2}\delta_1e_1e_2}. \tag{13.6}$$

Proof. Notice that the support of $1 - W(t_1, t_2, t_3)$ is essentially included in the union of domains where one of the variables is in $[0, 1]$ and the other two are restricted to $[0, q^{1/2+\varepsilon}/a]$. The bound on the other two variables comes from the dropoff due to the function $F_{a,\sqrt{q}}(t_1, t_2, t_3)$.

Using only the trivial bound $J_{\kappa-1}(x) \ll x$ and $|I| = |J_*|$, we derive from (6.8) (which we bound trivially) that

$$|J_*(t_1t_2t_3, am_1', c_0, g_0k_0', c_2, k_1)| \ll M_2\sqrt{m_1at_1t_2t_3}/C. \tag{13.7}$$

Therefore, using the above restrictions on the size of the t_i, and (5.6), we derive

$$B_\Delta(k_0') \ll q^\varepsilon \frac{qm_1^{1/2}M_2}{a^{3/2}\delta_1e_1e_2C}.$$

For the arithmetical part, we have

$$\frac{1}{k_0'^3} A(k_0') \ll \frac{\tau(k_0')}{k_0'}.$$

Hence

$$\mathcal{S}_\Delta''' \ll q^\varepsilon \sum_{\substack{(c_0, g_0m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, qk_1k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \frac{qm_1^{1/2}M_2}{a^{3/2}\delta_1e_1e_2C} \frac{\tau(k_0')}{k_0'},$$

which quickly leads to (13.6). □

Recall that

$$\mathcal{S}_{0,0,0}'' = \sum_{\substack{g_0 | e_1e_2\delta_1am_1' \\ g_0 \equiv 0 \,(\mathrm{mod}\, d)}} \mathcal{S}_{0,0,0}''', \qquad \mathcal{S}_{0,0,0}' = \sum_{r_1r_2r_3=\delta_1} \sum_{e_1|r_2r_3} \sum_{e_2|r_3} \mu(e_1)\mu(e_2)\mathcal{S}_{0,0,0}'', \tag{13.8}$$

$$\mathcal{S}_{0,0,0} = \sum_{(a,q)=1} \frac{\mu(a)}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d\mu(c_2/d) \sum_{k_1} k_1^{1/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \mathcal{S}_0'. \tag{13.9}$$

Let $\mathcal{S}''_{\Delta}$, $\mathcal{S}'_{\Delta}$ and $\mathcal{S}_{\Delta}$ be defined similarly. Using Lemma 13.3 and $(g_0, c_2) \le c_2$ then implies that

$$\mathcal{S}_{\Delta} \ll q^{\varepsilon} \sum_{a} \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d \sum_{k_1} k_1^{1/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \frac{(m_1' k_1 c_2)^{1/2} M_2}{C k_1 k_1^* a^{3/2}}.$$

Using $m_1 = m_1' k_1 c_2 \ll q^{1+\varepsilon}/M_2$ and $C \gg q$, we obtain

$$\mathcal{S}_{\Delta} \ll q^{1+\varepsilon}/C \ll q^{\varepsilon}.$$

Define $\overline{\mathcal{S}}'''_{0,0,0}$ to be the same as $\mathcal{S}'''_{0,0,0}$ but with W replaced by 1, so that

$$\mathcal{S}'''_{0,0,0} = \overline{\mathcal{S}}'''_{0,0,0} + \mathcal{S}'''_{\Delta},$$

and similarly for $\overline{\mathcal{S}}''_{0,0,0}$, etc. To show Theorem 13.1, we therefore need to show $\sum_{\mathrm{T}} \overline{\mathcal{S}}_{0,0,0} \ll_{\varepsilon} q^{\varepsilon}$.

13.4. The function $B(k_0')$

From now on, we let $\overline{B}(k_0')$ be the function obtained from the right hand side of (13.5) after replacing W by 1, and summing over the dyadic variables C and K. This has the shape

$$\overline{B}(k_0') = \iiint_{(\mathbb{R}^+)^3} \frac{J_*(t_1t_2t_3, a, m_1', c_0, g_0k_0', c_2, k_1)}{\delta_1 e_1 e_2} F_a\left(\frac{t_1}{\sqrt{q}}, \frac{t_2}{\sqrt{q}}, \frac{t_3}{\sqrt{q}}\right) \frac{dt_1\, dt_2\, dt_3}{\sqrt{t_1t_2t_3}}, \tag{13.10}$$

where we did not give a new name to J_* after summing over C and K. This is the relevant function for evaluating $\overline{\mathcal{S}}_{0,0,0}$.

Proposition 13.4. *Denote*

$$\mathcal{H}(s, w, u, \kappa) = (-1)^{\kappa/2} \frac{(2\pi)^{s+w+u-1}\Gamma(s+w+u)\Gamma(\kappa/2-w-u)\Gamma(\kappa/2-s)}{\Gamma(\kappa/2+s)\Gamma(\kappa/2+w+u)}. \tag{13.11}$$

Here $\mathcal{H}$ is holomorphic in the region

$$\mathrm{Re}(s), \mathrm{Re}(w+u) < \kappa/2 \quad \textit{and} \quad \mathrm{Re}(s+w+u) > 0,$$

with polynomial growth in $\mathrm{Im}(s)$, $\mathrm{Im}(w)$, *and* $\mathrm{Im}(u)$ *in vertical strips. With this notation,*

$$\overline{B}(k_0') = \frac{1}{\delta_1 e_1 e_2} \int_{(1-\varepsilon)} \frac{\gamma(1/2+s,\kappa)^3 G(s)^3}{\gamma(1/2,\kappa)^3 s^3} \zeta_q(1+2s)^3 q^{3s/2} a^{-3s}$$
$$\times \int_{(1-2\varepsilon)} \widetilde{V}(w) \left(\frac{m_1}{q}\right)^{-w} \int_{(-2\varepsilon)} M_2^u \widetilde{\omega}(u,\cdot) \frac{1}{k^{s-w-u}} \frac{(am_1)^{s-1/2}}{c^{s+w+u-1}}$$
$$\times \mathcal{H}(s,w,u,\kappa) \frac{du\, dw\, ds}{(2\pi i)^3},$$

where $k = g_0 k_0' k_1$, $m_1 = m_1' k_1 c_2$, $c = c_0 c_2$, *and* $\widetilde{\omega}(u,\cdot) = \widetilde{\omega}(u)\omega(m_1/M_1)$.

Proof. Note that in (13.10), the factor $t_1t_2t_3$ shows up as a block in both J and the denominator. Letting $y = t_1t_2t_3$ (viewing t_2 and t_3 as fixed), we have

$$\overline{B}(k_0') = \int_0^\infty \frac{J_*(y, a, m_1', c_0, g_0k_0', c_2, k_1)}{\delta_1 e_1 e_2} \int_0^\infty \int_0^\infty F_a\Big(\frac{y/(t_2t_3)}{\sqrt{q}}, \frac{t_2}{\sqrt{q}}, \frac{t_3}{\sqrt{q}}\Big) \frac{dt_2}{t_2} \frac{dt_3}{t_3} \frac{dy}{\sqrt{y}}.$$

We first claim that

$$\int_0^\infty \int_0^\infty F_a\Big(\frac{y/(t_2t_3)}{\sqrt{q}}, \frac{t_2}{\sqrt{q}}, \frac{t_3}{\sqrt{q}}\Big) \frac{dt_2}{t_2} \frac{dt_3}{t_3} = \int_{(2)} \frac{q^{3s/2}}{(a^3y)^s} \frac{\gamma(1/2+s,\kappa)^3 G(s)^3}{\gamma(1/2,\kappa)^3 s^3} \zeta_q(1+2s)^3 \frac{ds}{2\pi i}.$$

This is an exercise in Mellin inversion, directly using the definition (5.5). Secondly, we claim

$$\int_0^\infty J_*(y, a, m_1', c_0, g_0k_0', c_2, k_1) y^{-1/2-s} dy = \int_{(1-2\varepsilon)} \widetilde{V}(w) \Big(\frac{m_1}{q}\Big)^{-w} \int_{(-2\varepsilon)} M_2^u \widetilde{\omega}(u, \cdot) \frac{1}{k^{s-w-u}} \frac{(am_1)^{s-1/2}}{c^{s+w+u-1}} \mathcal{H}(s, w, u, \kappa) \frac{du\, dw}{(2\pi i)^2}. \tag{13.12}$$

Putting these two claims together then completes the proof.

Now we show (13.12). From (7.3) and (6.8), and summing over the C and K partitions, we have

$$J_*(y, a, m_1', c_0, g_0k_0', c_2, k_1) = e\Big(-\frac{yam_1'}{c_0g_0k_0'}\Big) I(m_1'k_1c_2, g_0k_0'k_1, ya, c_0c_2)$$
$$= \int_0^\infty e\Big(-\frac{yam_1'}{c_0g_0k_0'}\Big) e\Big(\frac{-g_0k_0'k_1t}{c_0c_2}\Big) J_{\kappa-1}\Big(\frac{4\pi\sqrt{m_1'k_1c_2yat}}{c_0c_2}\Big) V_1\Big(\frac{m_1'k_1c_2t}{q}\Big) \omega_{M_2}(t, \cdot) \frac{dt}{\sqrt{t}}$$
$$= \int_0^\infty e\Big(-\frac{yam_1}{ck}\Big) e\Big(\frac{-kt}{c}\Big) J_{\kappa-1}\Big(\frac{4\pi\sqrt{m_1yat}}{c}\Big) V_1\Big(\frac{m_1t}{q}\Big) \omega_{M_2}(t, \cdot) \frac{dt}{\sqrt{t}},$$

where for simplicity in the final line above we have written the expression in terms of the earlier variable names, and where $\omega_{M_2}(t, \cdot) = \omega(t/M_2)\omega(m_1/M_1)$ (since we have summed over C and K, as well as the N_i). Therefore, (13.12) equals

$$\int_0^\infty \int_0^\infty e\Big(-\frac{yam_1}{ck}\Big) e\Big(\frac{-kt}{c}\Big) J_{\kappa-1}\Big(\frac{4\pi\sqrt{m_1yat}}{c}\Big) V_1\Big(\frac{m_1t}{q}\Big) \omega_{M_2}(t, \cdot) \frac{dt}{\sqrt{t}} y^{-s} \frac{dy}{\sqrt{y}}.$$

Changing variables by $y = z/t$ (after interchanging the order of integration) shows that (13.12) equals

$$\int_0^\infty \int_0^\infty e\Big(-\frac{zam_1}{ckt}\Big) e\Big(\frac{-kt}{c}\Big) J_{\kappa-1}\Big(\frac{4\pi\sqrt{m_1za}}{c}\Big) V_1\Big(\frac{m_1t}{q}\Big) \omega_{M_2}(t, \cdot) t^s \frac{dt}{t} z^{-s} \frac{dz}{\sqrt{z}}.$$

Rewriting V_1 and ω_{M_2} in terms of their Mellin transforms, we find that (13.12) is

$$\int_{(1-2\varepsilon)} \widetilde{V}(w)\left(\frac{m_1}{q}\right)^{-w} \int_{(-\varepsilon)} M_2^u \widetilde{\omega}(u,\cdot)\mathcal{I}\,\frac{du\,dw}{(2\pi i)^2}, \tag{13.13}$$

where

$$\mathcal{I}=\int_0^\infty J_{\kappa-1}\left(\frac{4\pi\sqrt{m_1az}}{c}\right)z^{-s}\left(\int_0^\infty e\left(-\frac{m_1az}{ckt}\right)e\left(\frac{-kt}{c}\right)t^{s-w-u}\,\frac{dt}{t}\right)\frac{dz}{\sqrt{z}}. \tag{13.14}$$

We will derive an explicit formula for $\mathcal{I}$ by consulting tables of integrals.

Lemma 13.5. *For* $|\mathrm{Re}(s-w-u)|<1$, *we have*

$$\int_0^\infty e\left(-\frac{m_1az}{ckt}\right)e\left(\frac{-kt}{c}\right)t^{s-w-u}\,\frac{dt}{t}$$
$$= -i\pi\left(\frac{\sqrt{m_1az}}{k}\right)^{s-w-u} e^{-\pi i\frac{s-w-u}{2}} H^{(2)}_{s-w-u}\left(\frac{4\pi\sqrt{m_1az}}{c}\right).$$

Here $H_\nu^{(2)}(z)=J_\nu(z)-iY_\nu(z)$ *is the Hankel function of the second kind.*

Proof. This follows from [GR00, (3.871.1), (3.871.2)], or formulas (17) and (36) in [EMOT54, Section 6.5]. □

Even though the original calculation requires $|\mathrm{Re}(s-w-u)|<1$ for convergence, note that the Hankel function $H_\nu^{(2)}(x)$, for any $x>0$, is an entire function of ν [DLMF, Section 10.2] and hence we may move our lines of integration in s, w and u to any location without encountering any poles from the Hankel function.

Inserting this evaluation into (13.14), we have

$$\mathcal{I}=-i\pi e^{-\pi i\frac{s-w-u}{2}}\int_0^\infty\left(\frac{\sqrt{m_1az}}{k}\right)^{s-w-u} H^{(2)}_{s-w-u}\left(\frac{4\pi\sqrt{m_1az}}{c}\right)$$
$$\times J_{\kappa-1}\left(\frac{4\pi\sqrt{m_1az}}{c}\right)z^{1/2-s}\,\frac{dz}{z}.$$

The Bessel and Hankel functions have the same argument, which is quite pleasant. By changing variables, we have

$$\mathcal{I}=\frac{-ie^{-\pi i\frac{s-w-u}{2}}}{2k^{s-w-u}}\,\frac{(am_1)^{s-1/2}(4\pi)^{s+w+u}}{c^{s+w+u-1}}\int_0^\infty H^{(2)}_{s-w-u}(z)J_{\kappa-1}(z)z^{1-s-w-u}\,\frac{dz}{z}. \tag{13.15}$$

The z-integral may be evaluated in closed form.

Lemma 13.6. *For* $\mathrm{Re}(\pm\nu-\mu)<\mathrm{Re}(\lambda)<1$, *we have*

$$\int_0^\infty H_\nu^{(2)}(x)J_\mu(x)x^\lambda\,\frac{dx}{x}$$
$$=\frac{i2^{\lambda-1}\Gamma(1-\lambda)\Gamma\left(\frac{\nu+\mu+\lambda}{2}\right)\Gamma\left(\frac{\mu-\nu+\lambda}{2}\right)}{\pi\Gamma\left(\frac{\nu+\mu-\lambda}{2}+1\right)\Gamma\left(\frac{\mu-\nu-\lambda}{2}+1\right)}e^{-\frac{\pi}{2}i(\mu-\nu+\lambda)}. \tag{13.16}$$

Proof. We may use formulas (33) and (36) of [EMOT54, Section 6.8] (but note (36) is missing a $\Gamma(1-\lambda)$ term), and simplifying using gamma function identities. □

Substituting

$$\lambda = 1 - s - w - u, \qquad \nu = s - w - u, \qquad \mu = \kappa - 1,$$

we find that the region of convergence corresponds to

$$\mathrm{Re}(s), \mathrm{Re}(w+u) < \kappa/2 \quad \text{and} \quad \mathrm{Re}(s+w+u) > 0, \tag{13.17}$$

which are satisfied by the lines of integration given in (13.13). Furthermore,

$$\mathcal{I} = \frac{(-i)^{\kappa} e^{\pi i \frac{s+w+u}{2}}}{k^{s-w-u}} \frac{(am_1)^{s-1/2}}{c^{s+w+u-1}} \frac{(2\pi)^{s+w+u-1}\Gamma(s+w+u)\Gamma\big(\frac{\kappa}{2}-w-u\big)\Gamma\big(\frac{\kappa}{2}-s\big)}{\Gamma\big(\frac{\kappa}{2}+s\big)\Gamma\big(\frac{\kappa}{2}+w+u\big)},$$

giving

$$\mathcal{I} = \frac{\mathcal{H}(s,w,u,\kappa)}{k^{s-w-u}} \frac{(am_1)^{s-1/2}}{c^{s+w+u-1}}. \tag{13.18}$$

An application of Stirling's approximation shows the growth in $\mathrm{Im}(s)$, $\mathrm{Im}(w)$ and $\mathrm{Im}(u)$ is bounded by a polynomial. Inserting the formula for $\mathcal{I}$ into (13.13) completes the proof of Proposition 13.4. □

13.5. Bounding the zero term

Now let us recall that $k = g_0 k_0' k_1$, $m_1 = m_1' k_1 c_2$ and $c = c_0 c_2$. We will substitute the evaluation of $\overline{B}$ into $\overline{\mathcal{S}}'''_{0,0,0}$ which was defined as (13.3) (with the partition of unity removed). This gives

$$\overline{\mathcal{S}}'''_{0,0,0} = \frac{1}{\delta_1 e_1 e_2} \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \frac{A(k_0')}{k_0'^3} \int_{(1-\varepsilon)} \frac{\gamma(1/2+s,\kappa)^3 G(s)^3}{\gamma(1/2,\kappa)^3 s^3}$$
$$\times \zeta_q(1+2s)^3 q^{3s/2} a^{-3s} \int_{(1-2\varepsilon)} \widetilde{V}(w) \left(\frac{q}{m_1' k_1 c_2}\right)^{w} \int_{(-2\varepsilon)} M_2^u \widetilde{\omega}(u,\cdot)$$
$$\times \frac{\mathcal{H}(s,w,u,\kappa)}{(g_0 k_0' k_1)^{s-w-u}} \frac{(a m_1' k_1 c_2)^{s-1/2}}{(c_0 c_2)^{s+w+u-1}} \frac{du\, dw\, ds}{(2\pi i)^3}. \tag{13.19}$$

Next examine the Dirichlet series

$$\mathcal{Z}(s,w,u) = \zeta_q(1+2s)^3 \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0^{s+w+u}} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \frac{\frac{1}{k_0'^3} A(k_0')}{k_0'^{s-w-u}}.$$

Using the evaluation $A(k_0') = k_0' \sum_{d|k_0'} d\varphi(k_0'/d)$ (from Lemma 13.2) and Möbius inversion to remove the condition $(k_0', c_0) = 1$, one may derive

$$\mathcal{Z}(s, w, u) = \frac{\zeta(1+s-w-u)^2 \zeta(s+w+u)}{c_2^{s-w-u+1}(qk_1k_1^*)^{s+w+u}} (g_0, c_2)^{s-w-u+1} \Delta(s, w, u), \quad (13.20)$$

where $\Delta(s, w, u)$ is analytic for

$$\mathrm{Re}(s) > 0, \quad \mathrm{Re}(u+w) < 1 + \mathrm{Re}(s), \quad \mathrm{Re}(s+w+u) > 0, \quad (13.21)$$

and bounded by $O(q^{\varepsilon})$ in that region. The sum defining $\mathcal{Z}(s, w, u)$ converges absolutely for $\mathrm{Re}(s+w+u) > 1$ and $\mathrm{Re}(w+u) < \mathrm{Re}(s)$.

Inserting (13.20) into (13.19), we obtain

$$\begin{aligned}\overline{\mathcal{S}}_{0,0,0}''' = \int_{(1-\varepsilon)} \frac{\gamma(1/2+s,\kappa)^3 G(s)^3}{\delta_1 e_1 e_2 \gamma(1/2,\kappa)^3 s^3} \frac{q^{3s/2}}{a^{3s}} \int_{(1-2\varepsilon)} \widetilde{V}(w) \left(\frac{q}{m_1' k_1 c_2}\right)^w \int_{(-2\varepsilon)} M_2^u \widetilde{\omega}(u, \cdot) \\ \times \frac{\zeta(1+s-w-u)^2 \zeta(s+w+u)}{c_2^{s-w-u+1}(qk_1k_1^*)^{s+w+u}} (g_0, c_2)^{s-w-u+1} \\ \times \frac{\Delta \mathcal{H}(s, w, u, \kappa)}{(g_0 k_1)^{s-w-u}} \frac{(am_1' k_1 c_2)^{s-1/2}}{c_2^{s+w+u-1}} \frac{du\, dw\, ds}{(2\pi i)^3}. \end{aligned} \quad (13.22)$$

Now move the contour of integration in w to the line $\mathrm{Re}(w) = 4\varepsilon$. In doing that note that we still have $\mathrm{Re}(s+w+u) = 1-\varepsilon+4\varepsilon-\varepsilon = 1+2\varepsilon > 1$ and $\mathrm{Re}(s-w-u+1) = 1-\varepsilon-4\varepsilon+2\varepsilon+1 = 2-3\varepsilon > 1$, so we do not pass over any poles. Now move the line of integration in s to $\mathrm{Re}(s) = 3\varepsilon$. By doing so, we pick up the residue from the simple pole at $s = 1-w-u$.

The remaining integral. The contribution from the final integral to $\overline{\mathcal{S}}_{0,0,0}'''$ is

$$\ll_{\varepsilon} \frac{(g_0, c_2) q^{\varepsilon}}{\delta_1 e_1 e_2 \sqrt{am_1' k_1 c_2}}, \quad (13.23)$$

for any $\varepsilon > 0$. The contribution to $\overline{\mathcal{S}}_{0,0,0}$ from this part is then calculated (recall that $\delta_1 = k_1 d/(a, k_1 d)$ and $\delta_2 = e_1 e_2 \delta_1 am_1'/g_0$) to be at most

$$\begin{aligned}&\sum_a \frac{1}{a^{3/2}} \sum_{c_2} \frac{1}{c_2^{3/2}} \sum_{d|c_2} d \sum_{k_1} k_1^{1/2} \sum_{m_1'} \frac{1}{\sqrt{m_1'}} \sum_{r_1r_2r_3=\delta_1} \sum_{e_1|r_2r_3} \sum_{e_2|r_3} \sum_{g_0|e_1e_2\delta_1 am_1'} \frac{(g_0, c_2) q^{\varepsilon}}{\delta_1 e_1 e_2 \sqrt{am_1' k_1 c_2}} \\ &\ll q^{\varepsilon} \sum_a \frac{1}{a^2} \sum_{c_2} \frac{1}{c_2^2} \sum_{d|c_2} \sum_{k_1} \frac{(k_1 d, a)}{k_1} \sum_{m_1'} \frac{1}{m_1'} \sum_{e_1|r_2r_3} \sum_{e_2|r_3} \frac{1}{e_1 e_2} \sum_{g_0|e_1e_2\delta_1 am_1'} (g_0, c_2),\end{aligned}$$

where all the summations may be truncated at some fixed power of q (cf. the convention in Section 11.6). Summing over everything trivially using $(g_0, c_2) \le c_2$ shows that the integral contribution to $\overline{\mathcal{S}}_{0,0,0}$ is $O(q^{\varepsilon})$.

The $s = 1 - w - u$ residue. This residue contributes to $\overline{\mathcal{S}}'''_{0,0,0}$ the following:

$$\int_{(4\varepsilon)} \frac{\gamma(3/2-w-u,\kappa)^3 G(1-w-u)^3}{\delta_1 e_1 e_2 \gamma(1/2,\kappa)^3 (1-w-u)^3} \widetilde{V}(w) \Big(\frac{1}{m_1' k_1 c_2}\Big)^w \times \int_{(-2\varepsilon)} M_2^u \widetilde{\omega}(u,\cdot) \frac{q^{\frac{1-w-u}{2}-u}}{a^{3(1-w-u)}} \frac{\zeta(2-2w-2u)^2}{c_2^{2-2w-2u}(k_1 k_1^*)} (g_0,c_2)^{2-2w-2u} \times \frac{\Delta\mathcal{H}(s,w,u,\kappa)}{(g_0 k_1)^{1-2w-2u}} (a m_1' k_1 c_2)^{1/2-w-u} \frac{du\, dw}{(2\pi i)^2}. \tag{13.24}$$

Now move the line of integration in w to $\mathrm{Re}(w) = 1 - \varepsilon$. This will pass over an apparent double pole of $\zeta(2-2w-2u)$ but the triple zero of $G(1-w-u)^3$ cancels it. Then by a trivial bound, the residue is

$$\ll \frac{g_0 q^{\varepsilon}}{\delta_1 e_1 e_2 a^{1/2} c_2^{3/2} k_1^{3/2} k_1^* m_1'^{3/2}}. \tag{13.25}$$

The contribution coming from the residue can be bounded by

$$\sum_a \frac{1}{a^2} \sum_{c_2} \frac{1}{c_2^3} \sum_{d|c_2} d \sum_{k_1} \frac{1}{k_1 k_1^*} \sum_{m_1'} \frac{1}{m_1'^2} \sum_{r_1 r_2 r_3 = \delta_1} \frac{1}{\delta_1} \sum_{e_1 | r_2 r_3} \sum_{e_2 | r_3} \frac{1}{e_1 e_2} \sum_{g_0 | e_1 e_2 \delta_1 a m_1'} g_0.$$

Let us trivially bound $g_0 \leq e_1 e_2 \delta_1 a m_1'$. All the remaining sums are easily bounded, so this part is also $O(q^{\varepsilon})$.

This completes the proof of Theorem 13.1.

13.6. One of the p_i is zero

This case is the easiest, since (as it turns out) we may bound everything trivially and obtain the desired bound $\mathcal{S}_0 \ll q^{\varepsilon}$.

The original sum is symmetric in p_1, p_2 and p_3, so it suffices to estimate the terms with $p_3 = 0$ and $p_1, p_2 \neq 0$ (the expression for A from Lemma 8.2 may not appear symmetric in the p_i, but of course it must be due to the original definition (8.1)). We apply a dyadic partition of unity to the p_1 and p_2 variables. Let $P_1, P_2 \neq 0$, set $P_3 = 0$, let $\mathrm{P} = (P_1, P_2, 0)$ and consider

$$\mathcal{S}'''_{\mathrm{P}} = \sum_{\substack{(c_0, g_0 m_1') = 1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0) = 1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \sum_{\substack{p_1 \asymp P_1 \\ p_2 \asymp P_2}} \frac{1}{k_0'^3} A(p_1, p_2, 0; k_0') B(p_1, p_2, 0).$$

Here

$$A(p_1, p_2, 0; k_0') = k_0' \sum_{f | (p_2, k_0')} f S(p_1, 0; k_0'/f) \ll k_0'^{1+\varepsilon} (p_1 p_2, k_0').$$

Note that we only need to consider the non-oscillatory cases for B, where B is given by (8.13), since in the oscillatory case all the p_i must be nonzero or else B is very small. Then

$$\mathcal{S}_{\mathrm{P}}''' \ll \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \frac{1}{k_0'^2} \sum_{p_1, p_2 \neq 0} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} \frac{\sqrt{M_2 N}}{h} (p_1 p_2, k_0'),$$

where recall $\delta = \kappa - 1 \geq 1$ in the pre-transition non-oscillatory range, and $\delta = -1$ in the post-transition range. Recall $P_1 P_2 \ll q^{\varepsilon} \frac{k_0'^2}{N_2' N_3'} \ll q^{\varepsilon} \frac{k_0'^2 h}{N_2 N_3}$. Therefore,

$$\mathcal{S}_{\mathrm{P}}''' \ll q^{\varepsilon} \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} \frac{\sqrt{M_2 N}}{N_2 N_3} \frac{K(g_0, c_2)}{g_0 k_1 c_2} \frac{1}{q k_1 k_1^*}.$$

It is then not difficult to see that

$$\mathcal{S}_{\mathrm{P}} \ll q^{\varepsilon} \max_a \left(\frac{\sqrt{aMN}}{C} \right)^{\delta} \frac{\sqrt{M_2 N}\, K}{q N_2 N_3 \sqrt{a}} \sqrt{M_1}.$$

In the **post-transition case**, this bound becomes

$$\mathcal{S}_{\mathrm{P}} \ll q^{\varepsilon} \max_a \frac{M_1 N_1}{q} \ll q^{\varepsilon}.$$

A calculation shows the **pre-transition non-oscillatory case** leads to the same bound.

In all cases, summing over the dyadic tuples of P gives $\mathcal{S}_0 \ll_{\varepsilon} q^{\varepsilon}$, as desired.

13.7. Two of the p_i are zero

We finally consider the case where say $p_1 \neq 0$ and $p_2 = p_3 = 0$. This case leads to some new subtleties not present in the case with all p_i zero. The first step is to extend the sum to all $p_1 \in \mathbb{Z}$, and then subtract back the term with $p_1 = 0$. We already showed with Theorem 13.1 that the term with all p_i zero is bounded in an acceptable way. After this, we apply Poisson summation backwards. The net effect is precisely the same as when only applying Poisson summation in the n_2 and n_3 variables, and setting $p_2 = p_3 = 0$ (up to the term with all p_i zero).

It is perhaps easiest to return to (7.11). Define $\mathcal{Q}$ to be the term we get from this, after Poisson in n_2 and n_3, and substitution of $p_2 = p_3 = 0$, so that

$$\mathcal{Q} = \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0, c_2)})}} \sum_{n_1 \geq 1} \frac{A^*(n_1; k_0') B^*(n_1)}{\sqrt{n_1}},$$

where

$$A^*(n_1; k_0') = \frac{1}{k_0'^2} \sum_{x_2, x_3 \,(\mathrm{mod}\, k_0')} e\left(\frac{\delta_2 n_1 x_2 x_3 \overline{c_0}}{k_0'} \right)$$

and

$$B^*(n_1) = \frac{1}{\sqrt{\delta_1 e_1 e_2}} \int_0^\infty \int_0^\infty F_a\Big(\frac{r_1 e_1 n_1}{\sqrt{q}}, \frac{r_2 e_2 t_2}{\sqrt{q}}, \frac{r_3 t_3}{\sqrt{q}}\Big)\omega\Big(\frac{n_1 e_1 r_1}{N_1}, \frac{t_2 e_2 r_2}{N_2}, \frac{t_3 r_3}{N_3}\Big) \\ \times J_*(e_1 e_2 \delta_1 n_1 t_2 t_3, a, m_1', c_0, g_0 k_0', c_2, k_1) \frac{dt_2\, dt_3}{\sqrt{t_2 t_3}}.$$

Since some of the details are similar to the case where all p_i are zero (and easier), we will be more brief in such occasions. We may evaluate A^* directly from the definition, using a similar method to the proof of Lemma 8.2, which gives

$$A^*(n_1; k_0') = \frac{1}{k_0'} \sum_{f|k_0'} \varphi(k_0'/f)\delta(n_1 \equiv 0 \pmod{k_0'/f}).$$

We have $J_*(\cdots) \ll M_2^{1/2}$, which follows from bounding $J_{\kappa-1}(x) \ll 1$, and so

$$B^*(n_1) \ll \Big(\frac{M_2 N_2 N_3}{r_1 r_2 r_3 e_1 e_2 e_2 r_2 r_3}\Big)^{1/2}.$$

In turn, this leads to the estimate

$$\mathcal{Q} \ll_\varepsilon q^\varepsilon \frac{\sqrt{M_2 N}\,(g_0, c_2)}{\delta_1 e_1 e_2 q k_1 k_1^* c_2}.$$

Then the contribution to $\mathcal{S}$ from $\mathcal{Q}$ is seen to be $O(q^{-1+\varepsilon}(MN)^{1/2}) = O(q^{1/4+\varepsilon})$.

The next step is to replace the sum of the P_2, P_3 partitions of unity by 1, as in Section 13.3. Since $\mathcal{Q} \ll_\varepsilon q^{1/4+\varepsilon}$, the error in doing so is expected to be at most $O(q^\varepsilon)$, as described in the paragraph preceding Lemma 13.3. We omit the details, as this case is easier than the case with all p_i zero. Since $n_1 \geq 1$ automatically, we may easily sum over the N_1 partition (avoiding the analytic problems near the origin). We also reassemble the C and K partitions, at no cost. Define $\overline{\mathcal{Q}}$ to be the sum obtained after all these partitions are removed, and $\overline{B^*}(n_1)$ to be the new function. For simplicity of notation we will not change the name of the function J_*. Then by the change of variables $y = e_1 e_2 \delta_1 n_1 t_2 t_3$ (viewing t_3 as fixed), we have

$$\overline{B^*}(n_1) = \int_0^\infty \int_0^\infty \frac{J_*(y, a, m_1', c_0, g_0 k_0', c_2, k_1)}{\delta_1 e_1 e_2 \sqrt{n_1}} F_a\Big(\frac{r_1 e_1 n_1}{\sqrt{q}}, \frac{\frac{y}{e_1 r_1 r_3 n_1 t_3}}{\sqrt{q}}, \frac{r_3 t_3}{\sqrt{q}}\Big) \frac{dt_3}{t_3} \frac{dy}{\sqrt{y}}.$$

By an exercise in Mellin inversion, one may show

$$\int_0^\infty F_a\Big(\frac{r_1 e_1 n_1}{\sqrt{q}}, \frac{\frac{y}{e_1 r_1 n_1 r_3 t_3}}{\sqrt{q}}, \frac{r_3 t_3}{\sqrt{q}}\Big) \frac{dt_3}{t_3} \\ = \int_{(1)} \int_{(1)} \frac{\gamma(1/2 + s_1, \kappa) G(s_1)}{\gamma(1/2, \kappa) s_1} \frac{\gamma(1/2 + s, \kappa)^2 G(s)^2}{\gamma(1/2, \kappa)^2 s^2} \\ \times a^{-2s - s_1} \zeta_q(1 + s_1 + s)^2 \zeta_q(1 + 2s) \Big(\frac{\sqrt{q}}{e_1 r_1 n_1}\Big)^{s_1} \Big(\frac{q e_1 r_1 n_1}{y}\Big)^s \frac{ds_1\, ds}{(2\pi i)^2}.$$

Then using (13.12), we derive

$$\overline{B^*}(n_1) = \int_{(1-2\varepsilon)} \frac{\widetilde{V}(w)}{\delta_1 e_1 e_2 \sqrt{n_1}} \int_{(1)} \frac{\gamma(1/2+s_1,\kappa)G(s_1)}{\gamma(1/2,\kappa)s_1} \int_{(1)} \frac{\gamma(1/2+s,\kappa)^2 G(s)^2}{\gamma(1/2,\kappa)^2 s^2}$$
$$\times \int_{(0)} M_2^u \widetilde{\omega}(u,\cdot) a^{-2s-s_1} \zeta_q(1+s_1+s)^2 \zeta_q(1+2s) \left(\frac{\sqrt{q}}{e_1 r_1 n_1}\right)^{s_1} (q e_1 r_1 n_1)^s$$
$$\times \left(\frac{q}{m_1}\right)^w \frac{\mathcal{H}(s,w,u,\kappa)}{k^{s-w-u}} \frac{(a m_1)^{s-1/2}}{c^{s+w+u-1}} \frac{du\, ds\, ds_1\, dw}{(2\pi i)^4},$$

where recall that $k = g_0 k_0' k_1$, $m_1 = m_1' k_1 c_2$, and $c = c_0 c_2$. Moreover, as in Section 13.5, we have $\widetilde{\omega}(u,\cdot) = \widetilde{\omega}(u)\omega(m_1/M_1)$, since we have summed over N_1, N_2, N_3, C, and K.

Applying these changes of variables, and inserting this into the definition of $\overline{\mathcal{Q}}$, we obtain

$$\overline{\mathcal{Q}} = \frac{1}{\delta_1 e_1 e_2} \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, \frac{c_2}{(g_0,c_2)})}} \frac{1}{k_0'} \sum_{f \mid k_0'} \varphi\left(\frac{k_0'}{f}\right) \sum_{n_1 \equiv 0 \,(\mathrm{mod}\, k_0'/f)} \frac{1}{n_1}$$
$$\times \int_{(1-2\varepsilon)} \widetilde{V}(w) \int_{(1)} \frac{\gamma(1/2+s_1,\kappa)G(s_1)}{\gamma(1/2,\kappa)s_1} \int_{(1-\varepsilon)} \frac{\gamma(1/2+s,\kappa)^2 G(s)^2}{\gamma(1/2,\kappa)^2 s^2}$$
$$\times \int_{(0)} M_2^u \widetilde{\omega}(u) a^{-2s-s_1} \zeta_q(1+s_1+s)^2 \zeta_q(1+2s) \left(\frac{\sqrt{q}}{e_1 r_1 n_1}\right)^{s_1} (q e_1 r_1 n_1)^s$$
$$\times \left(\frac{q}{m_1' k_1 c_2}\right)^w \frac{\mathcal{H}(s,w,u,\kappa)}{(g_0 k_0' k_1)^{s-w-u}} \frac{(a m_1' k_1 c_2)^{s-1/2}}{(c_0 c_2)^{s+w+u-1}} \frac{du\, ds\, ds_1\, dw}{(2\pi i)^4}.$$

With the displayed lines of integration, all the outer sums converge absolutely. Indeed, we have

$$\sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0^{s+w+u}} \sum_{\substack{(k_0', \delta_2 c_0)=1 \\ k_0' \equiv 0 \,(\mathrm{mod}\, c_2/(g_0,c_2))}} \frac{1}{k_0'^{1+s-w-u}} \sum_{f \mid k_0'} \varphi\left(\frac{k_0'}{f}\right) \sum_{n_1 \equiv 0 \,(\mathrm{mod}\, k_0'/f)} \frac{1}{n_1^{1+s_1-s}}$$
$$= \zeta(1+s_1-s) \sum_{\substack{(c_0, g_0 m_1')=1 \\ c_0 \equiv 0 \,(\mathrm{mod}\, q k_1 k_1^*)}} \frac{1}{c_0^{s+w+u}} \sum_{(f, \delta_2 c_0)=1} \frac{1}{f^{1+s-w-u}} \sum_{\substack{(\ell, \delta_2 c_0)=1 \\ \ell \equiv 0 \left(\mathrm{mod}\, \frac{\frac{c_2}{(g_0,c_2)}}{(f, \frac{c_2}{(g_0,c_2)})}\right)}} \frac{\varphi(\ell)}{\ell^{2+s_1-w-u}}. \tag{13.26}$$

As long as we assume that

$$\mathrm{Re}(1+s_1-w-u) > 0, \quad \mathrm{Re}(1+s-w-u) > 0, \quad \mathrm{Re}(s+w+u) > 0,$$

the coprimality conditions are benign. Then the Dirichlet series in (13.26) is of the form

$$\zeta(1+s_1-s)\frac{\zeta(s+w+u)}{(qk_1k_1^*)^{s+w+u}}\zeta(1+s-w-u)\zeta(1+s_1-w-u)\left(\frac{(g_0,c_2)}{c_2}\right)^{1+\min(s,s_1)-w-u}\Delta,$$

where Δ is holomorphic and bounded by q^{ε}, and $\min(s, s_1)$ means the variable with the smaller real part. The factors $\zeta_q(1+s_1+s)^2\zeta_q(1+s)$ may be absorbed into the definition of Δ provided that $\mathrm{Re}(s), \mathrm{Re}(s_1) > 0$.

Moving the summations to the inside, we derive

$$\overline{\mathcal{Q}} = \int_{(1-2\varepsilon)} \frac{\widetilde{V}(w)}{\delta_1 e_1 e_2} \int_{(1)} \frac{\gamma(1/2+s_1,\kappa)G(s_1)}{\gamma(1/2,\kappa)s_1} \int_{(1-\varepsilon)} \frac{\gamma(1/2+s,\kappa)^2 G(s)^2}{\gamma(1/2,\kappa)^2 s^2} \int_{(0)} M_2^u \widetilde{\omega}(u)$$
$$\times \frac{\zeta(1+s_1-s)}{a^{2s+s_1}} \frac{\zeta(s+w+u)}{(qk_1k_1^*)^{s+w+u}} \zeta(1+s-w-u)\zeta(1+s_1-w-u)(qe_1r_1)^s\left(\frac{\sqrt{q}}{e_1r_1}\right)^{s_1}$$
$$\times \left(\frac{(g_0,c_2)}{c_2}\right)^{1+\min(s,s_1)-w-u}\left(\frac{q}{m_1'k_1c_2}\right)^w \frac{\Delta\mathcal{H}(s,w,u,\kappa)}{(g_0k_1)^{s-w-u}} \frac{(am_1'k_1c_2)^{s-1/2}}{c_2^{s+w+u-1}} \frac{du\,ds\,ds_1\,dw}{(2\pi i)^4}. \tag{13.27}$$

For ease of reference, we list all the constraints on the variables (using $\kappa \geq 2$):

$$0 < \mathrm{Re}(s+w+u), \quad \mathrm{Re}(s) < 1,$$
$$\mathrm{Re}(w+u) < 1 + \min(\mathrm{Re}(s), \mathrm{Re}(s_1)), \quad \mathrm{Re}(s), \mathrm{Re}(s_1) > 0. \tag{13.28}$$

Now we move the contours as follows. First, move w from $1-2\varepsilon$ to 4ε, which does not involve crossing any poles. Following this, move s_1 to 5ε, which crosses a pole at $s_1 = s$ only. Next, move s to 6ε, which crosses a pole at $s+w+u = 1$ only. We will deal with this pole momentarily.

The pole at $s_1 = s$. This contributes to $\overline{\mathcal{Q}}$

$$\frac{1}{\delta_1 e_1 e_2} \int_{(4\varepsilon)} \widetilde{V}(w) \int_{(1-\varepsilon)} \frac{\gamma(1/2+s,\kappa)^3 G(s)^3}{\gamma(1/2,\kappa)^3 s^3} \int_{(0)} M_2^u \widetilde{\omega}(u) \frac{\zeta(s+w+u)}{(qk_1k_1^*)^{s+w+u}} q^{3s/2}$$
$$\times \frac{\zeta(1+s-w-u)^2}{a^{3s}} \left(\frac{(g_0,c_2)}{c_2}\right)^{1+s-w-u} \left(\frac{q}{m_1'k_1c_2}\right)^w$$
$$\times \frac{\Delta\mathcal{H}(s,w,u,\kappa)}{(g_0k_1)^{s-w-u}} \frac{(am_1'k_1c_2)^{s-1/2}}{c_2^{s+w+u-1}} \frac{du\,ds\,dw}{(2\pi i)^3}.$$

A careful scrutiny of this formula shows that is is essentially identical to (13.22) (we did not check that the Δ function is literally equal in the two cases, but this would not be surprising). Here we need that we can move w to 4ε and then u to -2ε without crossing any poles; this move in w was our first step following (13.22), so this is easily checked. Therefore, by the work in the case with all p_i zero, the contribution to $\mathcal{S}_{0,0}$ from this pole is $O(q^{\varepsilon})$.

The new contour. On the new line, with all the variables having real parts at multiples of ε, the contribution to $\overline{\mathcal{Q}}$ is

$$\ll q^{\varepsilon} \frac{1}{\delta_1 e_1 e_2} \frac{c_2^{1/2}}{(am_1' k_1)^{1/2}} \frac{(g_0, c_2)}{c_2}.$$

Recalling that $\delta_1 = k_1 d/(a, k_1 d)$, it is not hard to see that inserting this bound into (7.10), (7.6), (7.1) gives a final contribution to $\mathcal{S}_{0,0}$ of size $O(q^{\varepsilon})$.

The pole at $s = 1 - u - w$. Denote this contribution to $\overline{\mathcal{Q}}$ by $\overline{\mathcal{Q}}_{\text{Res}}$. Then

$$\begin{aligned}\overline{\mathcal{Q}}_{\text{Res}} = \int_{(4\varepsilon)} \frac{\widetilde{V}(w)}{\delta_1 e_1 e_2} \int_{(5\varepsilon)} & \frac{\gamma(1/2+s_1,\kappa)G(s_1)}{\gamma(1/2,\kappa)s_1} \frac{\gamma(3/2-u-w,\kappa)^2 G(1-u-w)^2}{\gamma(1/2,\kappa)^2(1-u-w)^2} \\ & \times \int_{(0)} M_2^u \widetilde{\omega}(u) \frac{\zeta(s_1+u+w)}{(qk_1k_1^*)} \zeta(2-2w-2u)\zeta(1+s_1-w-u) \\ & \times \left(\frac{(g_0,c_2)}{c_2}\right)^{1+\min(1-u-w,s_1)-w-u} \left(\frac{\sqrt{q}}{e_1 r_1}\right)^{s_1} \frac{(qe_1r_1)^{1-u-w}}{a^{s_1+2(1-u-w)}} \\ & \times \left(\frac{q}{m_1' k_1 c_2}\right)^{w} \frac{\Delta\mathcal{H}(1-u-w,w,u,\kappa)}{(g_0k_1)^{1-2w-2u}} (am_1'k_1c_2)^{1/2-u-w} \frac{du\, ds_1\, dw}{(2\pi i)^3}.\end{aligned}$$

The constraints $\text{Re}(w+u) < 1 + \text{Re}(s)$ and $0 < \text{Re}(s) < 1$ with $s = 1-u-w$ simply become $0 < \text{Re}(u+w) < 1$.

Finally, we move w to $1 - 10\varepsilon$, crossing a pole at $w = s_1 - u$ only. On the new lines of integration, the contribution to $\overline{\mathcal{Q}}$ is

$$\ll q^{\varepsilon} \frac{g_0 k_1}{\delta_1 e_1 e_2} \frac{1}{k_1 k_1^*} \frac{1}{m_1' k_1 c_2} \frac{1}{\sqrt{am_1' k_1 c_2}} \ll \frac{a^{1/2} q^{\varepsilon}}{c_2^{3/2} \sqrt{m_1'}\, k_1^* k_1^{3/2}},$$

using only the weak bound $g_0 \le \delta_1 e_1 e_2 m_1' a$. It is easy to see that the final contribution to $\mathcal{S}_{0,0}$ from this is $O(q^{\varepsilon})$.

The pole at $w = s_1 - u$. This contributes

$$\begin{aligned}\overline{\mathcal{Q}}_{\text{Res}'} := \frac{1}{\delta_1 e_1 e_2} \int_{(5\varepsilon)} & \widetilde{V}(s_1-u) \frac{\gamma(1/2+s_1,\kappa)G(s_1)}{\gamma(1/2,\kappa)s_1} \frac{\gamma(1/2+1-s_1,\kappa)^2 G(1-s_1)^2}{\gamma(1/2,\kappa)^2(1-s_1)^2} \\ \times \int_{(0)} & M_2^u \widetilde{\omega}(u) \frac{\zeta(2s_1)\zeta(2-2s_1)}{(qk_1k_1^*)} \left(\frac{(g_0,c_2)}{c_2}\right)^{1+\min(1-s_1,s_1)-s_1} \left(\frac{\sqrt{q}}{e_1r_1}\right)^{s_1} (qe_1r_1)^{1-s_1} \\ & \times \left(\frac{q}{m_1'k_1c_2}\right)^{s_1-u} \frac{\Delta\mathcal{H}(1-s_1,s_1-u,u,\kappa)}{(g_0k_1)^{1-2s_1} a^{s_1+2(1-s_1)}} (am_1'k_1c_2)^{1/2-s_1} \frac{du\, ds_1}{(2\pi i)^2}.\end{aligned}$$

In terms of q, this part is $O(q^{\varepsilon})$, but the problem now is that the sum over m_1' will not be absolutely convergent. The way around this roadblock is to move the contour to a location where the m_1'-sum converges absolutely, and shift the contour back. Having $G(1/2) = 0$

once again is crucial. To this end, it is important to sum over the partition of unity in the M_1 and M_2 variables.

One may check that $\mathcal{H}(1-s_1, s_1-u, u, \kappa)$ is actually independent of u. Therefore, it is easy to sum $\overline{\mathcal{Q}}_{\mathrm{Res}'}$ over M_2: It is not hard to show that if $D(u)$ is a Dirichlet series absolutely convergent on the line $\mathrm{Re}(u)=0$, then

$$\sum_{M_2 \text{ dyadic}} \frac{1}{2\pi i}\int_{(0)} M_2^u \widetilde{\omega}(u) D(u)\, du = D(0).$$

Now we move the s_1-contour to $3/4$ (crossing no poles since $G(1/2)=0$), and sum over m_1' and M_1, which gives

$$\sum_{M_1}\sum_{m_1'} \frac{\omega_{M_1}(m_1')}{\sqrt{m_1'}} \sum_{M_2} \overline{\mathcal{Q}}_{\mathrm{Res}'} = \frac{1}{\delta_1 e_1 e_2}\int_{(3/4)} \widetilde{V}(s_1-u)\frac{\gamma(1/2+s_1,\kappa)G(s_1)}{\gamma(1/2,\kappa)s_1}$$
$$\times \frac{\gamma(3/2-s_1,\kappa)^2 G(1-s_1)^2}{\gamma(1/2,\kappa)^2(1-s_1)^2}\,\frac{\zeta(2s_1)^2\zeta(2-2s_1)}{qk_1k_1^*}\left(\frac{(g_0,c_2)}{c_2}\right)^{1+\min(1-s_1,s_1)-s_1}$$
$$\times \left(\frac{\sqrt{q}}{e_1r_1}\right)^{s_1}\frac{(qe_1r_1)^{1-s_1}}{a^{s_1+2(1-s_1)}}\left(\frac{q}{k_1c_2}\right)^{s_1}\frac{\Delta\mathcal{H}(1-s_1,s_1,0,\kappa)}{(g_0k_1)^{1-2s_1}}(ak_1c_2)^{1/2-s_1}\frac{ds_1}{2\pi i}.$$

Now we move the s_1-contour back to ε, which shows that this term is bounded by

$$\frac{q^{\varepsilon}}{\delta_1 e_1 e_2}\,\frac{1}{k_1k_1^*}\,\frac{(g_0,c_2)}{c_2}\,\frac{e_1r_1}{g_0k_1}\,\frac{1}{a^2}(ak_1c_2)^{1/2}.$$

Using the crude bounds $\frac{(g_0,c_2)}{g_0} \le 1$, $\frac{e_1r_1}{\delta_1e_1e_2} \le 1$, and summing trivially over k_1, d, c_2, and a shows that this part contributes $O(q^{\varepsilon})$ to $\mathcal{S}_{0,0}$.

Acknowledgments. This material is based upon work of M.Y. supported by the National Science Foundation under agreement No. DMS-1401008. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The first author was partially supported by an AMS-Simons travel grant.

References

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

[BF17] Balkanova, O., Frolenkov, D.: New error term for the fourth moment of automorphic L-functions. J. Number Theory **173**, 293–303 (2017) Zbl 1420.11087 MR 3581920

[BHM07] Blomer, V., Harcos, G., Michel, P.: Bounds for modular L-functions in the level aspect. Ann. Sci. École Norm. Sup. (4) **40**, 697–740 (2007) Zbl 1185.11034 MR 2382859

[BK19] Blomer, V., Khan, R.: Twisted moments of L-functions and spectral reciprocity. Duke Math. J. **168**, 1109–1177 (2019) Zbl 07055222 MR 3934595

[BKY13] Blomer, V., Khan, R., Young, M.: Distribution of mass of holomorphic cusp forms. Duke Math. J. **162**, 2609–2644 (2013) Zbl 1312.11028 MR 3127809

[BM15] Blomer, V., Milićević, D.: The second moment of twisted modular L-functions. Geom. Funct. Anal. **25**, 453–516 (2015) Zbl 1400.11097 MR 3334233

[CI00] Conrey, J. B., Iwaniec, H.: The cubic moment of central values of automorphic L-functions. Ann. of Math. (2) **151**, 1175–1216 (2000) Zbl 0973.11056 MR 1779567

[DFI94] Duke, W., Friedlander, J. B., Iwaniec, H.: Bounds for automorphic L-functions. II. Invent. Math. **115**, 219–239 (1994) Zbl 0812.11032 MR 1258904

[DLMF] NIST Digital Library of Mathematical Functions. http://dlmf.nist.gov/, Release 1.0.24 of 2019-09-15

[EMOT54] Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F. G.: Tables of Integral Transforms. Vol. I. McGraw-Hill, New York (1954) Zbl 0055.36401 MR 0061695

[Gal70] Gallagher, P. X.: A large sieve density estimate near $\sigma = 1$. Invent. Math. **11**, 329–339 (1970) Zbl 0219.10048 MR 0279049

[GR00] Gradshteyn, I. S., Ryzhik, I. M.: Table of Integrals, Series, and Products. 6th ed., Academic Press, San Diego, CA (2000) Zbl 0981.65001 MR 1773820

[Ivi01] Ivić, A.: On sums of Hecke series in short intervals. J. Théor. Nombres Bordeaux **13**, 453–468 (2001) Zbl 0994.11020 MR 1879668

[Iwa97] Iwaniec, H.: Topics in Classical Automorphic Forms. Grad. Stud. Math. 17, Amer. Math. Soc., Providence, RI (1997) Zbl 0905.11023 MR 1474964

[Iwa02] Iwaniec, H.: Spectral Methods of Automorphic Forms. 2nd ed., Grad. Stud. Math. 53, Amer. Math. Soc., Providence, RI (2002) Zbl 1006.11024 MR 1942691

[IK04] Iwaniec, H., Kowalski, E.: Analytic Number Theory. Amer. Math. Soc. Colloq. Publ. 53, Amer. Math. Soc., Providence, RI (2004) MR 2061214

[Kim03] Kim, H. H.: 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. H. Kim and P. Sarnak). J. Amer. Math. Soc. **16**, 139–183 (2003) Zbl 1018.11024 MR 1937203

[KPY19] Kıral, E. M., Petrow, I., Young, M. P.: Oscillatory integrals with uniformity in parameters. J. Théor. Nombres Bordeaux **31**, 145–159 (2019) MR 3994723

[KY19] Kıral, E. M., Young, M. P.: Kloosterman sums and Fourier coefficients of Eisenstein series. Ramanujan J. **49**, 391–409 (2019) Zbl 07067771 MR 3949076

[Koh85] Kohnen, W.: Fourier coefficients of modular forms of half-integral weight. Math. Ann. **271**, 237–268 (1985) Zbl 0542.10018 MR 783554

[KMV00] Kowalski, E., Michel, P., VanderKam, J.: Mollification of the fourth moment of automorphic L-functions and arithmetic applications. Invent. Math. **142**, 95–151 (2000) Zbl 1054.11026 MR 1784797

[Li75] Li, W. C. W.: Newforms and functional equations. Math. Ann. **212**, 285–315 (1975) Zbl 0278.10026 MR 0369263

[MV10] Michel, P., Venkatesh, A.: The subconvexity problem for GL_2. Publ. Math. Inst. Hautes Études Sci. **111**, 171–271 (2010) Zbl 1376.11040 MR 2653249

[Mon71] Montgomery, H. L.: Topics in Multiplicative Number Theory. Lecture Notes in Math. 227, Springer, Berlin (1971) Zbl 0216.03501 MR 0337847

[MV07] Montgomery, H. L., Vaughan, R. C.: Multiplicative Number Theory. I. Classical Theory. Cambridge Stud. Adv. Math. 97, Cambridge Univ. Press, Cambridge (2007) Zbl 1245.11002 MR 2378655

[Mot07] Motohashi, Y.: The Riemann zeta-function and Hecke congruence subgroups. II. arXiv:0709.2590 (2007)

[Mot97] Motohashi, Y.: Spectral Theory of the Riemann Zeta-Function. Cambridge Tracts in Math. 127, Cambridge Univ. Press, Cambridge (1997) Zbl 1144.11003 MR 1489236

[Pet15] Petrow, I. N.: A twisted Motohashi formula and Weyl-subconvexity for L-functions of weight two cusp forms. Math. Ann. **363**, 175–216 (2015) Zbl 1331.11035 MR 3394377

[PY19] Petrow, I., Young, M. P.: A generalized cubic moment and the Petersson formula for newforms. Math. Ann. **373**, 287–353 (2019) Zbl 07051745 MR 3968874

[Pit95] Pitt, N. J. E.: On shifted convolutions of $\zeta^3(s)$ with automorphic L-functions. Duke Math. J. **77**, 383–406 (1995) Zbl 0855.11024 MR 1321063

[Top16] Topaçoğulları, B.: The shifted convolution of divisor functions. Quart. J. Math. **67**, 331–363 (2016) Zbl 1356.11070 MR 3509996

[Wal81] Waldspurger, J.-L.: Sur les coefficients de Fourier des formes modulaires de poids demi-entier. J. Math. Pures Appl. (9) **60**, 375–484 (1981) Zbl 0431.10015 MR 646366

[You17] Young, M. P.: Weyl-type hybrid subconvexity bounds for twisted L-functions and Heegner points on shrinking sets. J. Eur. Math. Soc. **19**, 1545–1576 (2017) Zbl 1430.11067 MR 3635360