SUB-CONVEXITY PROBLEM FOR RANKIN-SELBERG
L-FUNCTIONS

CHANDRASEKHAR RAJU

Abstract. We establish a sub-convexity estimate for Rankin-Selberg L-functions in the combined level aspect, using the circle method. If \( p \) and \( q \) are distinct prime numbers, \( f \) and \( g \) are non-exceptional newforms (modular or Maass) for the congruence subgroups \( \Gamma_0(p) \) and \( \Gamma_0(q) \) (resp) with trivial nebentypus, then for all \( \epsilon > 0 \) we show that there exists an \( A > 0 \) such that

\[
L\left( \frac{1}{2} + it, f \times g \right) \ll \epsilon, \mu_f, \mu_g (1 + |t|)^A \frac{(pq)^{1/2 + \epsilon}}{\max\{p, q\}^{\theta/2}}.
\]

The dependence on \( \mu_f \) and \( \mu_g \), the parameters at infinity for \( f \) and \( g \) respectively, is polynomial. Further, if \( p \) is fixed and \( q \to \infty \), we improve this to

\[
L\left( \frac{1}{2} + it, f \times g \right) \ll \epsilon, \mu_f, \mu_g (p(1 + |t|))^A q^{\theta - \frac{1}{8} - \frac{2\epsilon}{3} + \epsilon},
\]

where \( \theta \) is the exponent towards Ramanujan-conjecture for cuspidal automorphic forms. Unconditionally, we can take \( \theta = 7/64 \). This improves all previously known sub-convexity estimates in this case.

1. Introduction

Understanding the behaviour of automorphic L-functions in the critical strip is an important problem in modern analytic number theory. A problem which has received a lot of attention in this area is the sub-convexity problem. Let \( L(f, s) \) be an automorphic L-function and \( s \in \mathbb{C} \) be such that \( \Re(s) = \frac{1}{2} \), then the following inequality is called the convexity bound:

\[
L(f, s) \ll Q(f, s)^{1/4 + \epsilon},
\]

where \( Q(f, s) \) is the analytic conductor of the L-functions at \( s \). We refer the reader to [16] for the definition of analytic conductor. Obtaining any positive saving in the exponent 1/4, is typically called the sub-convexity problem. The generalized Lindelöf-hypothesis, which is a consequence of the generalized Riemann-Hypothesis, asserts that for any \( \epsilon > 0 \)

\[
L(f, s) \ll Q(f, s)^\epsilon.
\]
The convexity bound, can be viewed as the trivial bound for the $L$-function on the half line $\Re(s) = \frac{1}{2}$. One way to view a sub-convexity estimate is as progress towards generalized Riemann-hypothesis. Perhaps more concretely, the sub-convexity problem for various families of $L$-functions connects to various equidistribution problems. One example of such a connection is the relation between Quantum Unique Ergodicity (QUE) and sub-convexity estimates for the symmetric square $L$-functions (see [31]). We refer the reader to [8], [16], [21], and [23] for a survey of the other applications.

The aim of this work is twofold. Firstly, we exhibit a sub-convexity estimate for Rankin-Selberg $L$-functions in the combined level aspect of both the automorphic forms, as long as the conductor of the $L$-function doesn’t drop. Secondly, as we use the circle method instead of an amplified second moment to establish this result, we are able to completely bypass the use of the Kuznetsov trace formula. This makes the proof considerably less technical. Moreover, we are able to improve the known results considerably (see Theorem 1.3) by avoiding the technical complications and directly cutting to the heart of the matter.

In the case that $f$ is a $GL(1)$ automorphic form, the sub-convexity problem was solved due to the work of Weyl [32] and Burgess [3]. Iwaniec introduced the amplification method, in [15], to prove sub-convexity estimates for $GL(2)$ $L$-functions in the spectral aspect. Following this, Duke, Friedlander and Iwaniec established sub-convexity bounds in the level aspect for $GL(2)$ $L$-functions in a series of papers culminating in [5]. The problem becomes significantly harder to tackle when we head to $GL(3)$ $L$-functions.

Munshi [25] proved a hybrid sub-convexity bounds for $GL(2)$ $L$-functions twisted by a Dirichlet character in the critical strip by a very different argument. If $f$ is a modular form for $PSL_2(\mathbb{Z})$ and $\chi$ is a character mod $q$, he shows that

$$L(1/2 + it, f \times \chi) \ll_{\epsilon} (q(3 + |t|))^{1/2 - 1/18 + \epsilon}. \quad (1.1)$$

The main novelty in the argument is to directly separate the oscillation of $f$ and $\chi$ in an approximate functional equation for $L(f \times \chi, 1/2 + it)$, using Jutila’s circle method. Using a set of factorable moduli in the circle method, he obtains some extra cancellation to break the convexity bound. He has advanced the use of circle method to obtain sub-convexity bounds in a series of papers. Notably, he obtained the first sub-convexity bound for the value of non self dual $GL(3)$ cusp form in $t$-aspect [27] (The self-dual case was already known due to work of Li [20]). He also came up with the “$GL(2)$ circle method” to establish a sub-convexity bound for $GL(3)$ cusp forms twisted by Dirichlet character [26]. Holowinsky and Nelson have simplified the latter result’s proof considerably in [12]. Munshi also used the $GL(2)$ circle method to re-establish a Burgess type bound for character twists of $GL(2)$ automorphic forms (including the Eisenstein series, which recovers Burgess’s original bound for Dirichlet
Before we begin, we set up some basic notation. We say that a cusp form $f$ is “non-exceptional”, if either $f$ is holomorphic or the eigenvalue $\lambda_f$ of $f$ under the Laplacian $(-\Delta)$ satisfies, $\lambda_f \geq \frac{1}{4}$. Selberg’s eigenvalue conjecture asserts that there are no exceptional forms.

Let $f$ and $g$ be primitive cuspidal newforms (not necessarily holomorphic) for $\Gamma_0(p)$ and $\Gamma_0(q)$, with nebentypus $\chi_f$ and $\chi_g$ respectively. These are eigenforms of suitably normalized Hecke operators $\{T_n\}_{n \geq 1}$ with eigenvalues $\lambda_f(n)$ and $\lambda_g(n)$ respectively. For all primes $l$, these eigenvalues for $f$ can be written as

$$\lambda_f(l) = \alpha_{f,1}(l) \alpha_{f,2}(l), \quad \alpha_{f,1}(l) \alpha_{f,2}(l) = \chi_f(l)$$

and similarly for $g$. The Rankin-Selberg $L$-function is defined by

$$L(f \times g, s) = L(\chi_f \chi_g, 2s) \sum_{n=1}^{\infty} \frac{\lambda_f(n) \lambda_g(n)}{n^s}.$$ 

If $(p, q) = 1$, then we have the following Euler product for $L(f \times g, s)$:

$$L(f \times g, s) = \prod_{l \text{ prime}} \prod_{i,j=1,2} \left( 1 - \frac{\alpha_{f,i} \alpha_{g,j}}{l^s} \right)^{-1}.$$ 

Moreover, the equality above holds in general even when $p$ and $q$ are not coprime except for finitely many Euler factors at the primes $l$ dividing $(p, q)$. The arithmetic conductor of this $L$-function, $Q(f \times g)$, satisfies [9]

$$\frac{(pq)^2}{(p, q)^4} \leq Q(f \times g) \leq \frac{(pq)^2}{(p, q)}.$$ (1.2)

Thus if $(p, q) = 1$, the convexity estimate for the $L$-function is

$$L(f \times g, 1/2 + it) \ll_{t, \epsilon} Q(f \times g)^{1/4+\epsilon} = (pq)^{1/2+\epsilon}.$$ 

In the case that $f$ is fixed and we let $q$ vary, sub-convexity estimates are known due to the work of Michel, Kowalski, Vanderkam, and Harcos ([18], [19], [22], and [9]). Using the amplification method and the Kuznetsov trace formula (assuming $\chi_f \chi_g$ is not trivial), they established that [9]

$$L(f \times g, 1/2) \ll_{f, \epsilon} q^{1/2-1/2648+\epsilon}.$$ 

Better exponents are known in particular cases. In particular, if $f$ is a fixed non-exceptional cuspidal automorphic form and $g$ has trivial central character, then
Kowalski, Michel, and Vanderkam [19] showed that
\[ L(f \times g, 1/2) \ll_{f, \epsilon} q^{1/2 - 1/80 + \epsilon}. \] (1.3)

In this work, we tackle the case when both \( p \) and \( q \) vary simultaneously such that \((p, q) = 1\) and \( \chi_f, \chi_g \) are both the trivial character. This question has been treated in the works of Michel-Ramakrishnan [24], Feigon-Whitehouse [6], Nelson [30], and Holowinsky-Templier [13] in situations where positivity of the central value is known. Holowinsky and Munshi [11] obtained a sub-convexity bound for this problem as long as \( p \leq q^\eta \), with \( \eta = \frac{2}{21} \). Hou and Zhang extended this to \( \eta = \frac{2}{15} \) [14]. Assuming that the form with the smaller level is holomorphic Zhilin Ye [33] proves a sub-convexity bound for all \( \eta \). It has been indicated that \( \delta = 1/801 \) (in the notation of Theorem 1.1 below) is admissible. We remove the holomorphicity assumption and improve the sub-convexity exponent considerably by a different method.

**Theorem 1.1.** Let \( f \) and \( g \) be primitive non-exceptional cuspidal newforms (not necessarily holomorphic) of prime levels \( p \) and \( q \) respectively, with trivial nebentypus. If \( \delta = 1/64 \), \( p \neq q \) and \( \Re s = 1/2 \), then for any \( \epsilon > 0 \)
\[ L(f \times g, s) \ll_{s, \epsilon} \frac{(pq)^{1/2 + \epsilon}}{\max\{p, q\}^4}, \]
with the implied constant depending polynomially on \( |s| \), \( \epsilon \), and spectral parameters of \( f \) and \( g \) at infinity.

**Remark 1.1.** We can also treat exceptional forms, at the cost of a smaller exponent of sub-convexity. The primality of \( p \) and \( q \), can be replaced by the condition that \( p \) and \( q \) are coprime. We have avoided carrying this out to simplify the exposition and keep the ideas clear. Our method also works in the case \((p, q) > 1\), as long as the arithmetic conductor of \( f \times g \) doesn't drop much. The hypothesis \( c(f \times g) \gg_\epsilon (pq)^{1+\epsilon} \) is sufficient. This does not include \( f \times f \), which is related to the sub-convexity of the symmetric square \( L \)-function \( L(\text{sym}^2 f, s) \).

If either one of \( p \) or \( q \) is very small (i.e \( p \) is bounded by a small power of \( q \), say \( q^{1/1000} \)), then this problem can be solved using the methods of [19]. The most interesting and hardest case of the theorem is when both \( p \) and \( q \) are both large. In this case, the crux of the proof is the solution to a shifted convolution problem (6.3), where the shifts are multiples of levels of the modular form. Munshi encounters a very similar problem in his work on the symmetric square \( L \)-function [29]. We state this below in the form of a theorem, as this might be of independent interest in connection to other problems.

**Theorem 1.2.** Let \( p \) be a prime number or \( p = 1 \). Let \( a, b, c, d \) be integers such that \( a \) and \( b \) are co-prime to \( p \). Further, let \( f, g \) be non-exceptional cuspidal newforms
(modular or Maass) of level $p$ and trivial nebentypus. For any $M_1, M_2, K_1, K_2 \geq 1$, we claim the following upper bounds for the shifted convolution sum $S_{f,g}$:

$$S_{f,g}(a, b, c, d, M_1, M_2) \ll p^\epsilon \min\{(M_1 M_2)^{1/2}, (M_1 M_2)^\theta X\}, \quad (1.4)$$

where $S_{f,g}$ and $X$ have been defined in section 6.3 and equation (6.12) respectively. Here $\theta$ is the exponent towards Ramanujan conjecture for $f$ and $g$. If the shift $ad - bc$ is non-zero, then

$$S_{f,g}(a, b, c, d, M_1, M_2) \ll (K_1 K_2)^{3/2} \sqrt{abp}^{3/4} X^{3/4}. \quad (1.5)$$

Furthermore, if the shift $ad - bc$ is a non-zero multiple of $p$, then

$$S_{f,g}(a, b, c, d, M_1, M_2) \ll (ab p M_1 M_2)^{1/2} (K_1 K_2)^{3/2} \sqrt{abp}^{1/4} X^{3/4}. \quad (1.6)$$

We believe that it is possible to improve the upper bound in (1.6) to $O_{K_1, K_2, \epsilon}(\sqrt{abp}^{1/4} X^{1/2})$ using spectral theory. If the level of $f$ is fixed we obtain a better exponent in Theorem 1.1, using such an improvement known due to the work of Blomer [2]. This improves the previously known sub-convexity bounds due to Kowalski, Michel, and Vanderkam (1.3).

**Theorem 1.3.** Let $f$ and $g$ be primitive non-exceptional cuspidal newforms (not necessarily holomorphic) of prime levels $p$ and $q$ respectively, with trivial nebentypus. If $\delta = \frac{1-2\theta}{27+2\theta}$, then for any $s \in \mathbb{C}$ with $\Re s = 1/2$ and any $\epsilon > 0$

$$L(f \times g, s) \ll_{\delta, \epsilon, p, q} q^{1/2-\delta+\epsilon},$$

with the implied constant depending polynomially on $|s|$, $\epsilon$, $p$, and parameters of $f$ and $g$ at infinity. Here $\theta$ is the exponent towards Ramanujan conjecture for cuspidal automorphic forms on $GL(2)$. Unconditionally, we can take $\theta = 7/64$. This gives $\delta = 0.02598 \ldots$.

Though Theorem 1.3 has been stated for $f$ having trivial central character, one can go through the proof and check that, with minor modifications, the proof works even if $f$ has a non-trivial central character. But as this is not possible in Theorem 1.1, we have chosen not to write this down separately. However, we are unable to handle $f$ being an Eisenstein series. Hence, our result does not recover a sub-convexity estimate for $GL(2)$ $L$-functions in the level aspect [5]. The issue here is the presence of main terms, and a similar issue demanding a delicate cancellation argument arose in [5]. It would be interesting to resolve this case using the circle method.

As a Corollary to Theorem 1.3, we improve the bounds obtained by Kowalski, Michel, and Vanderkam [19] for the problem of distinguishing modular forms based on their first Fourier coefficients.
Corollary 1.4. Let $f$ be a primitive cusp form and $\delta = \frac{1-2\theta}{27+28\theta}$ as in Theorem 1.3 of prime level $p$ and $\epsilon > 0$. There exists a constant $C = C(f, \epsilon)$ such that for any primitive cuspidal new form $g$ of prime level $q$, there exists $n \leq q^{1-2\delta+\epsilon}$ such that

$$\lambda_f(n) \neq \lambda_g(n).$$

Proof. The proof is identical to the proof of Corollary 1.3 [19]. Use Theorem 1.3 in place of [19, Theorem 1.1] in the proof. \hfill \Box

We briefly review the facts we need about $GL(2)$ automorphic forms in the next section. We do not need the $GL(2)$ trace formula. We will say a few more comments on this point in Section 4 (see (4.10)), where we also briefly sketch the outline of the proof. In Section 5, we carry out the initial transformations leading us to the shifted convolution problem. We obtain upper bounds for the shifted convolution problem in Section 6. We end the paper by combining the bounds obtained before to prove Theorem 1.1 and Theorem 1.3 in Section 7.

Acknowledgment. The author thanks Kannan Soundararajan and Valentin Blomer for a careful reading and helpful comments. The author is also grateful to Roman Holowinsky and Paul Nelson for their encouragement. The author is supported by B.C. and E.J. Eaves Stanford Graduate Fellowship.

2. Review of automorphic forms

We state the facts we need briefly in this section. We refer the reader to [9], [22] for a complete account.

2.1. Voronoi summation. We recall the Voronoi summation formula from [19, Theorem A.4].

Lemma 2.1. Let $D$ be a positive integer, $\chi_D$ be a character of modulus $D$. Further, let $g$ be either a holomorphic form of weight $k_g \geq 2$ or a Maass form of eigenvalue $\lambda_g$, level $D$, and central character $\chi_D$. For $(a,c) = 1$, set $D_1 = (c,D)$ and $D_2 = D/D_1$ and assume that $(D_1,D_2) = 1$, so that $\chi_D = \chi_{D_1}\chi_{D_2}$ is the unique factorization of $\chi_D$ into characters of modulus $D_1$ and $D_2$. For $F \in C^\infty(\mathbb{R}^+)$, a smooth function
vanishing in a neighborhood of 0 and decreasing rapidly,

\[
\sum_{n \geq 1} \lambda_g(n)e\left(\frac{an}{c}\right)F(n) \\
= \frac{\chi_{D_2}(\pi)\chi_{D_2}(-c)\eta_g(D_2)}{c\sqrt{D_2}} \sum_{n \geq 1} \lambda_{g_{D_2}}(n)e\left(\frac{-naD_2}{c}\right) \int_0^\infty F(x)J^+_g\left(\frac{4\pi c}{\sqrt{D_2}x}\right)dx \\
+ \frac{\chi_{D_2}(\pi)\chi_{D_2}(c)\eta_g(D_2)}{c\sqrt{D_2}} \sum_{n \geq 1} \lambda_{g_{D_2}}(n)e\left(\frac{naD_2}{c}\right) \int_0^\infty F(x)J^-_g\left(\frac{4\pi c}{\sqrt{D_2}x}\right)dx.
\]

In this formula

- \( \eta_g(D_2) \) is the pseudo-eigenvalue of the Atkin-Lehner operator \( W_{D_2} \); if \( \lambda_g(D_2) \neq 0 \), it equals
  \[
  \eta_g(D_2) = \frac{G(\chi_{D_2})}{\lambda_g(D_2)\sqrt{D_2}};
  \]
- if \( g \) is holomorphic of weight \( k_g \), then
  \[
  J^+_g(x) = 2\pi i^{k_g}J_{k_g-1}(x), \quad J^-_g(x) = 0;
  \]
- if \( g \) is Maass form with \( (\Delta + \lambda)g = 0 \) and let \( r \) satisfy \( \lambda_g = (\frac{1}{2} + ir)(\frac{1}{2} - ir) \), and let \( \epsilon_g \) be the eigenvalue of \( g \) under the reflection operator. Then
  \[
  J^+_g(x) = \frac{-\pi}{\sin(\pi ir)}\left(J_{2ir}(x) - J_{-2ir}(x)\right), \quad J^-_g(x) = \epsilon_g \frac{4}{\pi} \cosh(\pi r)K_{2ir}(x);
  \]
- if \( r = 0 \),
  \[
  J^+_g(x) = -2\pi Y_0(x), \quad J^-_g(x) = \epsilon_g \frac{4}{\pi} K_0(x).
  \]

Remark 2.1. We shall use this Lemma only when \( \chi_D \) is the trivial character i.e \( g \) has trivial nebentypus. In this case \( W_{D_2} \) is an endomorphism on the space of cusp forms of level \( D_2 \), \( |\eta_g(D_2)| = 1 \) and \( g_{D_2} = g \) [19, Proposition A.1].

We need to understand the behaviour of the integral transforms defined in Lemma 2.1. We state what we need in the form of a Lemma below.

Lemma 2.2. Let \( a \geq 1 \), \( W : [1, 2] \to \mathbb{C} \) be a compactly supported smooth function satisfying

\[
W^{(l)}(x) \ll_l a^l,
\]
for all \( l \geq 0 \). Define \( W^\pm : \mathbb{R}^+ \to \mathbb{C} \) by

\[
W^\pm(x) = \int_0^\infty W(x)J^\pm_g\left(4\pi \sqrt{\xi x}\right)dx, \quad (2.1)
\]
where \( J_g \) has been defined in Lemma 2.1. Then for all \( j, l \geq 0 \) and \( \xi > 0 \)
\[
\xi^l \frac{\partial^l}{\partial \xi^l} W^\pm(\xi) \ll_{\mu_g,j,l} d^l \max\{1, \xi^{-\theta_g}\} \frac{a^j}{\xi^{l/2}},
\]  
where \( \mu_g \) is the parameter of \( g \) at infinity and \( \theta_g \) is defined as
\[
\theta_g = \begin{cases} 
0, & \text{if } g \text{ is holomorphic} \\
\Im(r), & \text{if } g \text{ is a Maass form and } \lambda_g = 1/2 + ir.
\end{cases}
\]  
Unconditionally, we have \( \theta_g \leq \frac{7}{64} [17] \). If \( g \) is not exceptional, then
\[
\xi^l \frac{\partial^l}{\partial \xi^l} W^\pm(\xi) \ll_{\mu_g,j,l} d^l \frac{a^j}{\xi^{l/2}}.
\]  
\( \Box \)

Proof. Making the substitution \( u = 2\pi \sqrt{\xi x} \) in (2.1), we get
\[
W^\pm(\xi) = \frac{1}{8\pi^2 \sqrt{\xi}} \int_0^\infty \frac{u}{\sqrt{\xi}} W\left(\frac{u^2}{(4\pi)^2 \xi}\right) J^\pm_g(u)du.
\]  
Differentiating within the integral, we get
\[
\left(\frac{\partial}{\partial \xi}\right)^l W^\pm(\xi) = \sum_{0 \leq k \leq l} \left(\frac{\partial}{\partial \xi}\right)^k \left(\frac{1}{8\pi^2 \sqrt{\xi}}\right) \int_0^\infty \left(\frac{\partial}{\partial \xi}\right)^{l-k} \left(\frac{u}{\sqrt{\xi}} W\left(\frac{u^2}{(4\pi)^2 \xi}\right)\right) J^\pm_g(u)du.
\]  
We apply Lemma 6.1 [19] (see remark below) with
\[
h(u) = \left(\frac{\xi}{a}\right)^{l-k} \left(\frac{\partial}{\partial \xi}\right)^{l-k} \left(\frac{u}{\sqrt{\xi}} W\left(\frac{u^2}{(4\pi)^2 \xi}\right)\right)
\]  
and \( M = 4\pi \sqrt{\xi} \) to bound the integral. Using our definition of \( J^\pm_g \) (see Lemma 2.1), we see that the real part of the Bessel function satisfies \( \Re \nu \geq -2\theta_g \). Thus if \( \Re \nu \leq 0 \), then
\[
M^{\Re \nu} \ll \max\{1, \xi^{-\theta}\}.
\]  
\( \Box \)

Remark 2.2. We would like to point out a typo in Lemma 6.1 of [19]. While the estimate stated in Lemma 6.1 [19] is
\[
\int_0^\infty J_{\nu}(x)h(x)dx \ll_{\nu,j} \frac{a^j(1 + |\log M|)}{M^{\nu-j+1}} \frac{M^{\Re \nu+j+1}}{(1 + M)^{\Re \nu+j+1/2}},
\]  
the estimate that has been shown is
\[
\int_0^\infty J_{\nu}(x)h(x)dx \ll_{\nu,j} \frac{a^j(1 + |\log M|)}{M^{\nu-j}} \frac{M^{\Re \nu+j}}{(1 + M)^{\Re \nu+j+1/2}}.
\]
Notice the absence of $+1$ in the exponent $M^{2\nu+j+1}$. Furthermore, the same inequality holds with the $J$-bessel function replaced by $K$-Bessel function or $Y$-Bessel function, without the $\frac{M^{2\nu}}{(1+M^{2\nu})}$ term.

The result above has been stated for $W$ having compact support contained in the interval $[1, 2]$. But it holds without any change for the support contained in any absolutely bounded interval, bounded away from zero. For example, $\text{supp}(W) \subset [1/1000, 1000]$ is sufficient for the purpose of this paper.

**Definition 2.3 ($H_\theta$).** We say that Hecke-cusp form $f$ of level $q$ and nebentypus $\chi$ satisfies the Hypothesis $H_\theta$, if for all $n \geq 1$ and any $\epsilon > 0$

$$|\lambda_f(n)| \ll_\epsilon n^{\theta + \epsilon},$$

where $\lambda_f(p)$ are the local parameters of $\pi_f$ at $p$ and $\lambda_f(n)$ satisfy Hecke relations.

The Ramanujan conjecture asserts that we can take $\theta = 0$, for all $q$ and $\chi$. We can unconditionally take $\theta = 7/64$ for Maass forms (Kim-Sarnak [17]) and $\theta = 0$ for holomorphic forms. Rankin-Selberg theory [9, 2.28], implies the Ramanujan conjecture on average unconditionally.

**Lemma 2.4.** Let $g$ cuspidal automorphic form of level $p$ and nebentypus $\chi_f$. Then for all $N \geq 1$ and $\epsilon > 0$

$$\sum_{1 \leq n \leq N} |\lambda_f(n)|^2 \ll_\epsilon (p(1 + |\mu_f|))N^\epsilon N,$$

where $\mu_f$ is the local parameter of $f$ at infinity.

We prove a Polya-Vinogradov type inequality for smooth partial sums of $\lambda_f(n)\lambda_g(n)$ below:

**Lemma 2.5.** Let $f$ and $g$ be cuspidal newforms of levels $p$ and $q$ respectively, $W : \mathbb{R}^+ \rightarrow \mathbb{C}$ be a smooth compactly supported function and $Q(f \times g)$ be the arithmetic conductor of $L(f \times g, s)$ (1.2). If $f \neq g$, so that $L(f \times g, s)$ has no poles, then

$$\sum_n \lambda_f(n)\lambda_g(n)W\left(\frac{n}{N}\right) \ll_\epsilon (pqN')\sqrt{N\sqrt{Q(f \times g)}}.$$

**Proof.** Let $F(s) = \sum_{n=1}^{\infty} \frac{\lambda_f(n)\lambda_g(n)n^s}{L(\chi_f, \lambda_g, 2s)} = \frac{L(f \times g, s)}{L(\chi_f, \lambda_g, 2s)}$. Using Mellin inversion,

$$\sum_n \lambda_f(n)\lambda_g(n)W\left(\frac{n}{N}\right) = \frac{1}{2\pi i} \int_{(2)} N^s F(s)\widetilde{W}(s)ds,$$
where
\[ \tilde{W}(s) = \int_0^\infty W(x)x^{s-1}dx \quad (2.11) \]
is the Mellin transform. Using integration by parts repeatedly, we can show that for all \( A \geq 0, \)
\[ \tilde{W}(s) \ll_A \min\{1, |s|^{-A}\}. \]
For \( \Re(s) > \frac{1}{2}, L(\chi_f \chi_g, 2s) \) doesn’t vanish. Hence, \( F(s) \) is analytic in the same region. Let \( \chi \) be any Dirichlet character. The following lower bound for \( L(\chi, s) \) to the right of critical strip is elementary and well known.
\[ L(\chi, 1 + \epsilon + it) = \prod_{l \text{ prime}} \left( 1 - \frac{\chi(l)}{l^{1+\epsilon+it}} \right)^{-1} \]
\[ \geq \prod_{l \text{ prime}} \left( 1 + \frac{1}{l^{1+\epsilon}} \right)^{-1} \]
\[ \geq \frac{c}{\zeta(1 + \epsilon)} \gg \epsilon, \]
where the constant \( c \) is independent of \( \chi, t, \) and \( \epsilon. \) Shifting the contour to \( \Re(s) = \frac{1}{2} + \epsilon \) in (2.10), we have
\[ \sum_n \lambda_f(n)\lambda_g(n)W\left(\frac{n}{N}\right) = \frac{1}{2\pi i} \int_{(1/2+\epsilon)} N^s F(s)\tilde{W}(s)ds \]
\[ = \frac{1}{2\pi i} \int_{[\text{Re}(s) \geq (pqN)^{-\epsilon}]} N^s \frac{L(f \times g, s)}{L(\chi_f \chi_g, 2s)} \tilde{W}(s)ds + O((pqN)^{-2018}) \]
\[ \ll_{\epsilon} (pqN)^{\epsilon}N^{1/2}Q(f \times g)^{1/4}. \]
In the last line, we have used the Phragmen-Lindelof convexity bound for \( L(f \times g, s) \) on \( \Re(s) = \frac{1}{2} + \epsilon \) to bound \( L(f \times g, s) \) from above. \( \square \)

As we shall make repeated use of the bound proved below, we state it in the form of a lemma.

**Lemma 2.6.** Let \( a, b, c \in \mathbb{N} \) be such that \( (ab, c) = 1 \) and \( X, Y \geq 1. \) If \( 0 \leq \alpha, \beta \leq \frac{1}{2}, \)
then
\[ S(\alpha, \beta) = \sum_{\substack{n \leq X \\atop m \leq Y \atop \text{am\equiv bm(c)}}} \frac{|\lambda_f(n)| |\lambda_g(m)|}{n^\alpha m^\beta} \ll_{\epsilon} (pqXY)^{\epsilon} \frac{X^{1-\alpha}Y^{1-\beta}}{c} \left( 1 + \sqrt[4]{c} + \sqrt[4]{X} + \sqrt[4]{XY} \right), \]
\[ (2.15) \]
In particular,

\[ S := S(0, 0) = \sum_{\substack{n \leq X \\ m \leq Y \\ an \equiv bn(c)}} |\lambda_f(n)||\lambda_g(m)| \ll \varepsilon (pqXY)^\varepsilon \frac{XY}{c} \left( 1 + \sqrt{\frac{c}{Y}} + \sqrt{\frac{c}{X}} + \frac{c}{\sqrt{XY}} \right). \]  

(2.16)

**Proof.** For \( \alpha = \beta = 0 \), using Cauchy-Schwarz inequality, we have

\[ S = S(0, 0) \ll \left( \sum_{\substack{n \leq X \\ m \leq Y \\ an \equiv bn(c)}} |\lambda_f(n)|^2 \right)^{1/2} \left( \sum_{\substack{n \leq X \\ m \leq Y \\ an \equiv bn(c)}} |\lambda_g(m)|^2 \right)^{1/2}. \]  

(2.17)

Now applying the Rankin-Selberg bound (2.9), we get

\[ S \ll (pqXY)^\varepsilon \left( X \left( \frac{Y}{c} + 1 \right) \right)^{1/2} \left( Y \left( \frac{X}{c} + 1 \right) \right)^{1/2}. \]  

(2.18)

For general \( \alpha \leq 1/2 \) and \( \beta \leq 1/2 \), the bound (2.15) follows from (2.16), after performing a dyadic sub-division of the sum over \( m, n \).

Remark 2.3. If both \( X, Y \) are greater than \( c \), then the first term on right hand side of (2.16) dominates the others. This bound is optimal upto \( (XY)^\varepsilon \).

2.2. Approximate functional equation. We refer the reader to [9, Section 3] for proofs. For \( s \) on the critical line, we set

\[ A = \prod_{i=1}^{4} |s + \mu_{f \times g,i}|^{1/2}, \]  

(2.19)

where the local parameters \( \mu_{f \times g,i} \) of \( \pi_f \times \pi_g \) can be computed in terms of the local parameters of \( \pi_f \) and \( \pi_g \) respectively. We can check that

\[ A \leq (|s| + \mu_f + \mu_g)^2. \]

We have essentially isolated the spectral part of the analytic conductor as \( A \). Let us define

\[ S_{f \times g}(N) = \sum_{n} \lambda_f(n)\lambda_g(n)U \left( \frac{n}{N} \right). \]  

(2.20)
where \( U : \mathbb{R} \to \mathbb{R} \) is a smooth function with compact support contained in \([1/2, 5/2]\).

By standard techniques, for \( \Re s = 1/2 \) and any \( K \geq 1 \) we can show

\[
L(f \times g, s) \ll A \log^2(pqA + 1) \sum_N \frac{|S_{f \times g}(N)|}{\sqrt{N}} \left( 1 + \frac{N}{AQ(f \times g)} \right)^{-K},
\]

where \( N \) runs over reals of the form \( N = 2^\nu, \nu \geq -1 \). Thus to prove Theorem 1.1, it is enough to prove the following statement.

**Proposition 2.7.** For \( \delta = \frac{1}{64} \), any \( 0 \leq \epsilon \leq 10^{-6} \), and any \( 1 \leq N \leq (Apq)^{1+\epsilon} \),

\[
S_{f \times g}(N) \ll_{A, \epsilon} \sqrt{N(pq)^{1/2+\epsilon}} \max\{p, q\}^\delta,
\]

where the dependence on \( A \) is polynomial.

If \( N \leq \frac{pq}{\max\{p, q\}^{2\delta}} \), then (2.22) follows from Rankin-Selberg bound (2.9). Hence, we may assume

\[
\frac{pq}{\max\{p, q\}^{2\delta}} \leq N \leq (Apq)^{1+\epsilon}.
\]

We shall define \( T \) by

\[
T := \frac{pq}{N},
\]

Then \((Apq)^{-\epsilon} \leq T \leq \max\{p, q\}^{2\delta}\). Similarly to prove Theorem 1.3, it is enough to prove the following statement:

**Proposition 2.8.** For \( \delta = \frac{1-2\theta}{2\theta + 28\theta} \), any \( 0 \leq \epsilon \leq 10^{-6} \), and any \( 1 \leq N \leq (Apq)^{1+\epsilon} \),

\[
S_{f \times g}(N) \ll_{A, p, \epsilon} \sqrt{Nq^{1/2-\delta+\epsilon}},
\]

where the dependence on \( A \) and \( p \) is polynomial.

We may likewise use the Rankin-Selberg bound (2.9) to show Proposition 2.8, when \( N \leq q^{1-2\delta} \). Proposition 2.7 and 2.8 are proved in Section 7. We shall treat both the propositions simultaneously till Section 6.

3. Initial steps: Amplification and Circle method

Let us assume without loss of generality that \( p < q \). We shall consider

\[
S(N) = S_{f \times g}(N) = \sum_{n \in \mathbb{Z}} \lambda_f(n)\lambda_g(n)U\left(\frac{n}{N}\right),
\]

12
We begin by “amplifying” the sum. We shall use the idea of Duke, Friedlander, and Iwaniec [5] to amplify the sum using the \( GL(2) \) Hecke-relations. If \((l, q) = 1\) and \(l\) is prime, then

\[
\lambda_g(l)^2 - \lambda_g(l^2) = 1.
\]

Let

\[
\mathcal{L} = \{l \in [L/2, L] : l \text{ is prime }, (l, pq) = 1\},
\]

where \(L(\leq q^{1/2})\) is a parameter to be chosen. \(L\) should be thought of as a small power of \(q\). Note that \(|\mathcal{L}| \gg \frac{L}{\log L}\). Define the amplifier \(\alpha\) by

\[
\alpha_r = \begin{cases} 
\lambda_g(l), & \text{if } r = l \in \mathcal{L}, \\
-1, & \text{if } r = l^2 \text{ and } l \in \mathcal{L}, \\
0 & \text{otherwise}. 
\end{cases}
\]

In what follows, we shall be able to save at most \(\sqrt{L}\) over the trivial bound for \(S(N)\) (see the diagonal contribution 7.1). As we need to prove Proposition 2.7, i.e show that

\[
S_{f \times g}(N) \ll_{A, \epsilon} \frac{\sqrt{N}(pq)^{1/2+\epsilon}}{\max\{p, q\}^\delta},
\]

we can make the assumption that

\[
\max\{p, q\}^\delta \leq \sqrt{L},
\]

where \(\delta\) is as in Proposition 2.7 and Proposition 2.8. This implies that (see (2.24))

\[
T \leq \max\{p, q\}^{2\delta} \leq L.
\]

Lemma 3.1. Let

\[
S_1(N) = \frac{1}{|\mathcal{L}|} \sum_{r \leq L^2} \alpha_r \sum_n \lambda_f(n)\lambda_g(nr)U(n/N).
\]

Then, for all \(L \geq 100\),

\[
S(N) = S_1(N) + O((pqN)^\epsilon N/\sqrt{L}).
\]

Proof. Using multiplicativity of the Fourier coefficients \(\lambda_g\), \(S_1(N)\) can be rewritten as

\[
\frac{1}{|\mathcal{L}|} \sum_r \alpha_r \lambda_g(r) \sum_{(n,l)=1} \lambda_f(n)\lambda_g(n)U(n/N) + \frac{1}{|\mathcal{L}|} \sum_r \alpha_r \sum_{l|n} \lambda_f(n)\lambda_g(nr)U(n/N),
\]
where \( l \) has been defined in (3.3). Adding and subtracting the \( n \) which are divisible by \( l \) to the first term, we get

\[
S_1(N) = \frac{1}{|\mathcal{L}|} \sum_r \alpha_r \lambda_g(r) \sum_n \lambda_f(n) \lambda_g(n) U(n/N) - \frac{1}{|\mathcal{L}|} \sum_r \alpha_r \lambda_g(r) \sum_{l|n} \lambda_f(n) \lambda_g(n) U(n/N) \\
+ \frac{1}{|\mathcal{L}|} \sum_r \alpha_r \sum_{l|n} \lambda_f(n) \lambda_g(nr) U(n/N).
\]

(3.8)

The second two terms, by an application of Rankin-Selberg bound (2.9) and \( H_{1/4} \) (see 2.3), are seen to be bounded above by \( O((pqN)^{\epsilon} NL^{-1/2}) \). We have chosen the amplifier \( \alpha \) such that

\[
\sum_r \alpha_r \lambda_g(r) = |\mathcal{L}|. \tag{3.9}
\]

Combining (3.8) and (3.9), we get

\[
S(N) = S_1(N) + O \left( (pqN)^{\epsilon} \frac{N}{\sqrt{L}} \right).
\]

\( \square \)

We separate the oscillation of \( f \) and \( g \) in (3.5) using circle method. The equality of integers

\[ nr = m, \]

can be rewritten as a congruence

\[ nr \equiv m \pmod{M}, \]

if the moduli \( M \) is greater than \( |nr - m| \). The main idea is to choose the moduli \( M \) as multiples of the product of the levels of \( f \) and \( g \). Let \( C = 10L^2 \) be a parameter defined by \( L \). (It is sufficient to choose \( C = 10L^2 \), but at the cost of more delicate analysis. We get a minor improvement in the result when this is done). Define the set of moduli \( \mathcal{C} \) by

\[
\mathcal{C} = \{ c \in [C, 2C] : c \text{ is prime }, (c, pq) = 1 \}. \tag{3.10}
\]

We note that for all \( c \in \mathcal{C} \) and \( l \in \mathcal{L} \), \( c \) and \( l \) are coprime. Let us define the weight function \( W(x, y) \) to be the product \( U(x)V(y) \), with \( U \) being the weight function in (3.5) and \( V : [1/3, 2] \to \mathbb{R} \) being a smooth bump function which is identically 1 on \([1/2, 5/2]\). Since the support of \( U \) is contained in \([1/2, 5/2]\), if \( U(x) \neq 0 \), then \( W(x, x) = 1 \). We would like to note the following bound on derivatives of \( W \):

\[
\frac{\partial^j}{\partial y^j} \frac{\partial^i}{\partial x^i} W(x, y) \ll_{i,j} 1. \tag{3.11}
\]
Separating the oscillation of $\lambda_f$ and $\lambda_g$ by circle method in (3.5), we get
\[
S_1(N) = \frac{1}{|C|} \sum_r \alpha_r \sum_{n,m} \lambda_f(n)\lambda_g(m) \delta(rn = m) W\left(\frac{n}{N}, \frac{m}{rN}\right)
\]
\[
= \frac{1}{|C|} \sum_r \frac{1}{|C|} \sum_{c \in C} \alpha_r \sum_{n,m} \lambda_f(n)\lambda_g(m) \delta(rn \equiv m \pmod{pqc}) W\left(\frac{n}{N}, \frac{m}{rN}\right)
\]
\[
= \frac{1}{pq|C||C|} \sum_r \sum_{c \in C} \frac{1}{c} \sum_{a(pqc)} \alpha_r \sum_{n,m} \lambda_f(n)\lambda_g(m) e\left(\frac{a(nr - m)}{pqc}\right) W\left(\frac{n}{N}, \frac{m}{rN}\right)
\]
\[
= \frac{1}{pq|C||C|} \sum_r \sum_{c \in C} \alpha_r \sum_{a(pqc)} \sum_{n} \lambda_f(n) e\left(\frac{anr}{pqc}\right) \sum_{m} \lambda_g(m) e\left(\frac{-am}{pqc}\right) W\left(\frac{n}{N}, \frac{m}{rN}\right).
\]
Thus
\[
S_1(N) = \frac{1}{pq|C||C|} \sum_r \sum_{c \in C} \alpha_r S_1^{r,c}(N),
\]
where
\[
S_1^{r,c}(N) = \sum_{a(pqc)} \sum_{n} \lambda_f(n) e\left(\frac{anr}{pqc}\right) \sum_{m} \lambda_g(m) e\left(\frac{-am}{pqc}\right) W\left(\frac{n}{N}, \frac{m}{rN}\right).
\]
The expression (3.17) will be our starting point. We shall outline our argument in the next section.

4. OUTLINE OF PROOF

In this outline, we shall assume the Ramanujan conjecture. We indicate the proof in the case $N = pq$ as this is the most important case. We shall apply circle method to the following “amplified” sum.
\[
S = \frac{1}{|C|} \sum_{r \in C} \sum_{n \sim pq} \lambda_f(n)\lambda_g(nr).
\]
The notation $n \sim pq$ means that $n$ runs over natural numbers between $[pq, 2pq]$, weighted by a smooth function. The symbol $A \sim B$, in this outline, means that $A$ transforms into $B$ after a series of steps.
Applying Voronoi summation in both the $m$ and $n$ variables to (3.15) yields

$$S \sim \frac{1}{pqC^3 |\mathcal{L}|} \sum_{r \leq C} \sum_{c \in \mathcal{C} \ a(pqc)} \alpha_r \sum_{n \sim pqC^2 \ m \sim pqC} \lambda_f(n) \lambda_g(m) e \left( \frac{\bar{a}(-n\bar{r} + m)}{pqc} \right)$$

(4.2)

$$\sim \frac{1}{C^2 |\mathcal{L}|} \sum_{|s| \ll C} \sum_{r \leq C} \sum_{c \in \mathcal{C}} \lambda_r \sum_{m \sim pqC} \lambda_f(mr + pqcs) \lambda_g(m).$$

(4.3)

We have pretended above that $(a, cpq) = 1$ for all residue classes $a \ (\text{mod} \ cpq)$. While this is obviously false, this assumption captures the essence of the proof. We get Ramanujan sums instead of the complete sum over all additive frequencies, in the actual proof. This seems to be an annoying technical issue when writing down a complete proof (see Lemma 5.1 and 5.2). Note that the trivial bound for the right hand side of (4.3) is $O((pqC)^{1+\epsilon})$ i.e we have gained a $C$ over the trivial bound. We separate the right hand side above into diagonal ($s = 0$) + non-diagonal part. For the $s = 0$ part: trivially bounding the $r$ and $c$ sum we get

$$S(s = 0) \ll \frac{1}{C} \left| \sum_{m \sim pqC} \lambda_f(m) \lambda_g(m) \right| .$$

Since $pqC$ is greater than the square root of the conductor of $f \times g$, we can get a saving in this sum by using the functional equation. Using Lemma 2.5 we get

$$\sum_{m \sim pqC} \lambda_f(m) \lambda_g(m) \ll (pq)^{1+\epsilon} \sqrt{C} .$$

Thus the diagonal part can be bounded by $O((pq)^{1+\epsilon}/\sqrt{C})$. For the off-diagonal part i.e $s \neq 0$, we combine the variables $sc = t \ll C^2$ and use Cauchy-Schwarz to eliminate the oscillation due to $\lambda_g(m)$. This leads us to our shifted convolution problem for $f$. We get

$$S^2 \ll \frac{pqC}{(C^2 |\mathcal{L}|)^2} \sum_{r_1, r_2, t_1, t_2} \alpha_{r_1} \alpha_{r_2} \sum_{m \sim pqC} \lambda_f(r_1 m + pq t_1) \lambda_f(r_2 m + pq t_2).$$

(4.4)

As we have squared the expression, we need to save $C^2$ on the right hand side. In the diagonal-terms $r_1 t_2 = r_2 t_1$, we save $|\mathcal{L}|/C^2$ which is greater than $C^2$. For the off diagonal terms we note that the shift $pq(r_1 t_2 - r_2 t_1)$ (defined in (6.3)) is a multiple of $p$. Munshi [29] encounters a similar problem of bounding

$$S(X, h) = \sum_{n \sim pX} \lambda_f(n) \lambda_f(n + ph).$$

He shows a power saving for $S(X, h)$, as long as $X \geq p^\delta$ for some $\delta > 0$ (see (1.6)). In our scenario $X = q \geq p$. We briefly sketch an outline of this argument here. We
can rewrite $S$ as
\begin{align}
S &= \sum_{n,m \sim pX} \lambda_f(n) \lambda_f(m) \delta(m = n + ph) \\
&= \sum_{n,m \sim pX \atop n + ph \equiv m(p)} \lambda_f(n) \lambda_f(m) \delta \left( \frac{n - m}{p} + h = 0 \right)
\end{align}
(4.5)
(4.6)

We pick the congruence mod $p$ using additive characters and use the Duke, Friedlander, and Iwaniec circle method (6.1) to rewrite the $\delta$ symbol. This leads us to a sum of the form
\begin{align}
S \sim \frac{1}{pX} \sum_{n,m \sim pX} \sum_{q \sim \sqrt{X}} \sum_{b(q)\ast} \lambda_f(n) \lambda_f(m) e \left( \frac{a(n - m - ph)}{pq} \right).
\end{align}
(4.7)

Applying Voronoi summation (2.1) to both $n$ and $m$ sum, we are led to
\begin{align}
S \sim \frac{1}{p} \sum_{n,m \sim p} \sum_{q \sim \sqrt{X}} \lambda_f(n) \lambda_f(m) S(m - n, ph, pq).
\end{align}
(4.8)

The important point to note here is that the Kloosterman sum modulo $pq$, factors as a Ramanujan sum modulo $p$ times a Kloosterman sum modulo $q$. The Ramanujan sum, being very small, allows us to gain the additional saving. Bounding the right hand side using the Weil bound for Kloosterman sum and $H_0$ (see 2.3), we get
\begin{align}
S \ll (pX^{3/4})^{1+\epsilon} = \frac{(pX)^{1/4}}{X^{1/4}} (pX)^\epsilon.
\end{align}
(4.9)

Using spectral theory we should be able to improve this bound to
\begin{align}
S \ll (pX)^\epsilon p \sqrt{X}.
\end{align}

We get a satisfactory bound for our problem in Section 6, imitating Munshi’s ideas.

Although our version of circle method (3.15) looks trivial, the set of moduli we chose to capture the congruence in the circle method have inbuilt into them the levels of $f$ and $g$. This feature can be noticed in the work of Aggarwal, Holowinsky, Lin, and Sun in their simplification of Munshi’s proof of Burgess bound. We would like to point out here that one could instead solve this problem by considering the following amplified second moment
\begin{align}
\sum_{h \in \mathcal{A}(pq)} |M_g(h)|^2 |L(f \times h, s)|^2,
\end{align}
(4.10)

where $\mathcal{A}(pq)$ runs over a weighted Hecke basis for automorphic forms of level $pq$ and
\begin{align}
M_g(h) = \sum_r \alpha_r \lambda_h(r),
\end{align}
is the Duke, Friedlander, and Iwaniec amplifier. Using the Kuznetsov trace formula to rewrite the spectral sum, we roughly get
\[
\sum_{r_1, r_2} \alpha_{r_1} \alpha_{r_2} \sum_{c \geq 1 \text{ coprime}} \frac{1}{c pq} \sum_{n, m \sim pq} \lambda_f(n) \overline{\lambda_f(m)} S(r_2 n, r_1 m, c pq) \varphi \left( \frac{4\pi \sqrt{nm}}{c pq} \right).
\]
(4.11)

If we apply a Voronoi transformation to either the \(n\) or \(m\) sum, we end up with a shifted convolution problem of the following shape:
\[
\sum_{r_1, r_2, s} \ldots \sum_{m} \lambda_f(m) \overline{\lambda_f}(r_1 r_2 m + pqs).
\]
(4.12)

Writing the proof using circle method cleans up the proof considerably. But it is worth noting here that behind the scenes, we are implicitly computing the spectral second moment in (4.10).

In [19], the authors compute such an amplified second moment (4.10), when \(p\) is fixed. But they use \(A(q)\) instead of \(A(pq)\). Let us consider the Petersson trace formula:
\[
\sum_{f \in \mathcal{H}_k(q)} \omega_f^{-1} \lambda_f(n) \overline{\lambda_f}(m) = \delta(m, n) + \sum_{c \geq 1} \frac{S(m, n, c)}{cq} J_{k-1} \left( \frac{4\pi \sqrt{mn}}{cq} \right).
\]

We are forced to consider all values of \(c\) on the right hand side. They treat the large values of \(c\) using a large sieve type inequality. The smaller \(c\)'s are treated as explained above. In the optimal treatment using this method we are forced to consider \(c\) as large as \(q^{1/6}\) (see below Equation 7.13 in [19]). Using circle method we are easily able to restrict our attention to only the small \(c\)'s (\(c \leq \sqrt{L}\)). This is the principal reason for our improvement in the exponent in Theorem (1.3). This is a technical problem which can be overcome by using a clever test function in the Kuznetsov trace formula (see Section 3.2, [7]). When this is done, the trace formula yields a better exponent. The limit of this method, under Ramanujan conjecture, is saving \(q^{\frac{1}{20}}\) over the trivial bound for \(L(f \times g, s)\).

5. Voronoi transformations

We shall apply Voronoi transformations (2.1) to the \(n\) and \(m\) sums in (3.17). The modulus in Voronoi summation clearly depends on the greatest common divisor \((ar, cpq)\). By our choice \(c, p, q,\) and \(r\) are pairwise co-prime. Let us assume that \((a, cpq) = c_1 p_1 q_1, c = c_1 c_2, p = p_1 p_2',\) and \(q = q_1 q_2'\). Since we assumed \(pq\) is
square-free, using the Chinese remainder theorem, we rewrite (3.17) as

\[ S^{r,c}_1(N) = \sum_{p_1p_2=q} \sum_{\substack{c_1c_2=c \atop q_1q_2=q}} \left( \sum_n \lambda_f(n) e \left( \frac{anr}{p_2q_2c_2} \right) \right) \left( \sum_m \lambda_g(m) e \left( \frac{-am}{p_2q_2c_2} \right) \right) W \left( \frac{n}{N}, \frac{m}{m} \right). \]  

Due to our assumption that \( q \) is prime, \( q_1 \) is either 1 or \( q \) (similarly for \( p \) and \( c \)). In Lemma 5.1, we show that the contribution of \( q_1 = q \) and \( c_1 = c \) to \( S_1(N) \) are negligible. We can make a similar statement about the contribution of \( p_1 = p \), if the value of \( p \) is large. On a first reading, we encourage the reader to assume \( p_1 = 1 \), \( q_1 = 1 \), and \( c_1 = 1 \), in order to avoid unnecessary complications in the notation. Applying Voronoi summation (2.1) to the \( n \) and \( m \) sum modulo \( p_2q_2c_2 \), we get

\[ S^{r,c}_1(N) = \sum_{\pm, \pm} S^{r,c,(\pm, \pm)}_1(N), \]  

where

\[ S^{r,c,(\pm, \pm)}_1(N) = \sum_{p_1p_2=q} \sum_{\substack{c_1c_2=c \atop q_1q_2=q}} \left( \sum_n \lambda_f(n) e \left( \frac{+aq_1r}{p_2q_2c_2} \right) \right) \left( \sum_m \lambda_g(m) e \left( \frac{aq_1r}{p_2q_2c_2} \right) \right) W^{\pm, \pm} \left( \frac{n}{X}, \frac{m}{Y} \right). \]  

\[ W^{\pm, \pm} (\xi, \eta) = \frac{r N^2}{(p_1q_1)^2 \sqrt{p_1 q_1}} \int_0^\infty \int_0^\infty W(x, y) J^+_f \left( 4\pi \sqrt{\xi x} \right) J^+_g \left( 4\pi \sqrt{\eta y} \right) dxdy, \]  

and

\[ X = \frac{p_1(p_2q_2c_2)^2}{N}, \quad Y = \frac{q_1(p_2q_2c_2)^2}{r N}. \]  

Using Lemma 2.2 and the assumption that \( f \) and \( g \) are not exceptional, we see that for all \( j \geq 0 \),

\[ W^{\pm, \pm} \left( \frac{n}{X}, \frac{m}{Y} \right) \ll_j \frac{r N^2}{(p_1q_1)^2} \min \left\{ 1, \left( \frac{X}{n} \right)^{-j}, \left( \frac{Y}{m} \right)^{-j}, \left( \frac{XY}{mn} \right)^{-j} \right\}. \]  

Among the four possible choices in \{\pm, \pm\}, we shall restrict our attention to \{+, +\}, as this is prototypical. Let us denote \( S^{r,c,(+, +)}_1(N) \) by \( S^{r,c}_2(N) \). We denote

\[ S^{\text{dual}}_1(N) := \frac{1}{pq|\mathcal{L}|c} \sum_r \sum_{c \in \mathcal{C}^*} \alpha_r c S^{r,c}_2(N). \]
By repeating the proof in a very straightforward way for all the other choices of \(\{\pm, \pm\}\) i.e \(\{+,-\}, \{-, +\}\), and \(\{-,-\}\), we can check that the bound obtained for \(S_1^{dual}(N)\) holds for \(S_1(N)\) also. We dyadically divide the sum over \(n\) and \(m\) in (5.3), using a smooth partition of unity. As this is a standard technique, we refer to Lemma 1 of [26] for details. This gives

\[
S_2^{r,c}(N) = \sum_{(A, \rho_1)} \sum_{p_1} \sum_{c_1} \sum_{(B, \rho_2)} \sum_{q_2} \left( \sum_n \lambda_f(n) e\left( \frac{\pm ap_1 q_1 n}{p_2 q_2 c_2} \right) \right) \left( \sum_m \lambda_g(m) e\left( \frac{\pm aq_1 m}{p_2 q_2 c_2} \right) \right) \left( \sum_{i,j} \frac{\partial^i}{\partial \eta^i} \frac{\partial^j}{\partial \xi^j} W_2(\xi, \eta) \right) \rho_1(\frac{n}{A}) \rho_2(\frac{m}{B}),
\]

where the pairs \((\rho_1, A)\) and \((\rho_2, B)\) are locally finite smooth partitions of unity (see [26] for more details). The important point is that the support of \(\rho_j\) is contained in \([1/2, 1]\), for all \(\rho_j\) appearing in the partition of unity. Hence the sum over \(n\) and \(m\) runs smoothly over integers in \([A/2, A]\) and \([B/2, B]\) respectively. It is also convenient to re-normalize the weight function \(W^{+, +}\), so that it is absolutely bounded. To this end, we define

\[
W_2(\xi, \eta) = \frac{(p_2 q_2 c_2)^2 \sqrt{p_1 q_1}}{r N^2} W^{+, +}\left( \frac{A\xi}{X}, \frac{B\eta}{Y} \right) \rho_1(\xi) \rho_2(\eta).
\]

Then, \(W_2 : \mathbb{R}^2 \to \mathbb{C}\) is a smooth function with compact support contained in \([1/2, 1] \times [1/2, 1]\). Lemma 2.2 implies that for all \(i, j \geq 0\)

\[
\frac{\partial^i}{\partial \eta^i} \frac{\partial^j}{\partial \xi^j} W_2(\xi, \eta) \ll_{i,j} 1.
\]

If \(n \geq (pq)^{\epsilon} X\) or \(m \geq (pq)^{\epsilon} Y\), choosing \(j = 2018 \epsilon^{-1}\) in (5.6), we get \(W^{+, +} \ll \epsilon \frac{1}{(pq)^{1000}}\).

Hence the contribution of such terms to \(S_1(N)\) is negligible. Thus

\[
S_2^{r,c}(N) = \sum_{p_1 p_2 = p, c_1 c_2 = c} \sum_{q_1 q_2 = q} S_2^{r,c}(A, B),
\]

where \(A\) and \(B\) runs over powers of 2, \(\rho_1\) and \(\rho_2\) are the smooth functions arising from the partitions of unity, and

\[
S_2^{r,c}(A, B) = \frac{r N^2}{(p_2 q_2 c_2)^2 \sqrt{p_1 q_1}} \sum_{a(p_2 q_2 c_2)^*} \left( \sum_n \lambda_f(n) e\left( \frac{\pm ap_1 q_1 n}{p_2 q_2 c_2} \right) \right) \left( \sum_m \lambda_g(m) e\left( \frac{\pm aq_1 m}{p_2 q_2 c_2} \right) \right) W_2(\frac{n}{A}, \frac{m}{B}).
\]
Summing up over the reduced classes \( a \) modulo \( p' \), \( q' \), \( c' \), we get Ramanujan sums. This gives

\[
S_{r,c}^2(A, B) = \sum_{n,m} \frac{rN^2}{(p_2q_2c'_2)^2 \sqrt{pq_1}} \left( \sum_{\lambda_f(n) \lambda_g(m) \tau_{p_2q'_2c'_2} (p_1rm - q_1n)W_2 \left( \frac{n}{A}, \frac{m}{B} \right) } \right),
\]

where

\[
\tau_q(n) = \sum_{a(n)} e \left( \frac{aq}{n} \right) = \sum_{d|\gcd(q,n)} \mu \left( \frac{q}{d} \right) d,
\]

is the Ramanujan sum. We shall use the identity (5.14) to expand the Ramanujan sum. This gives

\[
S_{r,c}^2(N) = \sum_{p_1p_2p_3 = 1} \sum_{q_1q_2q_3 = 1} \sum_{(p_1, A \leq \langle pq \rangle X)} \sum_{(p_2, B \leq \langle pq \rangle Y)} S_{r,c}^3(A, B),
\]

where

\[
S_{r,c}^3(A, B) = \mu(p_2) \mu(q_2) \mu(c_2) \frac{rN^2}{(p_2q_2c_2)^2 \sqrt{pq_1}} \times \left( \sum_{\lambda_f(n) \lambda_g(m) \delta[p_1rm = q_1n(p_3q_3c_3)]W_2 \left( \frac{n}{A}, \frac{m}{B} \right) } \right),
\]

\[
X = \frac{P_1(p_2p_3q_2q_3c_2c_3)^2}{N}, \quad Y = \frac{q_1(p_2p_3q_2q_3c_2c_3)^2}{rN}.
\]

We have made the substitution \( p'_2 = p_2p_3, \ q'_2 = q_2q_3, \) and \( c'_2 = c_2c_3. \) Since \( q \) is prime, \( q_2 \) is either 1 or \( q \) (similarly for \( p \) and \( c \)). Lemma 5.2 shows that the contribution of \( q_2 = q \) and \( c_2 = c \) to \( S_2(A, B) \) is negligible (The contribution of \( p_2 = p \) is also be shown to be negligible, if \( p \) is large). On a first reading, we encourage the readers to assume \( p_3 = p, \ q_3 = q, \) and \( c_3 = c. \)

Before we proceed further we would like to get rid of the boundary cases which make a smaller contribution. If we do not assume the primality of \( p \) and \( q \), we would have to consider the various factorizations possible and give a separate argument for each of these, depending on the sizes of the factors. Though straightforward, it is messy.

**Lemma 5.1.** Let \( S_{r,c,q_1=q}^2(N) \) be the contribution of the terms with \( q_1 = q \) to (5.15). Let \( S_{1,q_1=q}^{\text{dual}} \) be the contribution of such terms to \( S_{1,q_1=q}^{\text{dual}}(N) \) (5.7) i.e

\[
S_{1,q_1=q}^{\text{dual}}(N) = \frac{1}{pq|C||C|} \sum_r \sum_{c \in C} \alpha_r S_{r,c,q_1=q}^2(N).
\]
Then
\[
S_{1}^{\text{dual,q}=q}(N) \ll (pq)^{\epsilon} \sqrt{Npq} \left( \frac{C \sqrt{L}}{q^{3/2}} \frac{\sqrt{C}}{q} \right).
\]  
(5.19)

Similarly, let \( S_{1}^{\text{dual,p}=p}(N) \) and \( S_{1}^{\text{dual,c}=c}(N) \) be the contribution of the terms with \( p_1 = p \) and \( c_1 = c \) respectively to (5.7). Then
\[
S_{1}^{\text{dual,p}=p}(N) \ll (pq)^{\epsilon} \sqrt{Npq} \left( \frac{C \sqrt{L}}{p^{3/2}} \frac{\sqrt{C}}{p} \right)
\]  
(5.20)

and
\[
S_{1}^{\text{dual,c}=c}(N) \ll (pq)^{\epsilon} \sqrt{Npq} \frac{L}{C}.
\]  
(5.21)

Proof. If \( q_1 = q \), then \( q_2 = q_3 = 1 \). Using Lemma 2.6 to bound the right hand side of (5.16), we have
\[
S_3^{r,c}(A, B) \ll \frac{rN^2}{(pc)^2} \frac{AB}{\sqrt{p} \sqrt{q} \sqrt{p_3c_3}} \left( 1 + \sqrt{\frac{pc}{A}} + \sqrt{\frac{pc}{B}} + \sqrt{\frac{pc}{AB}} \right).
\]  
(5.22)

The function on right hand side of the inequality is increasing in \( A \) and \( B \). Thus \( S_3(A, B) \) is bounded by \( S_3((pq)^{\epsilon}X, (pq)^{\epsilon}Y) \), for all \( A \leq (pq)^{\epsilon}X \) and \( B \leq (pq)^{\epsilon}Y \). Moreover, among all factorizations of \( p_1p_2p_3 = p \) and \( c_1c_2c_3 = c \), the maximum value is attained at \( p_3 = p \) and \( c_3 = c \). In this case, \( X = \frac{(pc)^2}{N} \) and \( Y = \frac{q(pc)^2}{rN} \) (5.17). Putting this in (5.15), we get
\[
S_2^{r,c,q_1=q} \ll (pq)^{\epsilon} \frac{rN^2}{(pc)^2} \left( \frac{pc}{q} \right)^{\epsilon} \left( 1 + \frac{\sqrt{N}}{pc} + \frac{rN}{pc} + \frac{\sqrt{rN^2}}{q(pc)^2} \right)
\]  
(5.23)

\[
\ll (pq)^{\epsilon} \left( \frac{(pc)^2q}{\sqrt{pc}} + (pc)^2 \sqrt{q} \right).
\]  
(5.24)

Finally, using this bound in (5.18),
\[
S_1^{\text{dual,q}=q}(N) \ll (pq)^{\epsilon} \frac{1}{pqL||C||} \sum_r \sum_{c \in C} \frac{\alpha_r}{c} \left( \frac{(pc)^2}{\sqrt{q}} + \frac{(pc)^2q}{\sqrt{cT}} \right)
\]  
(5.25)

\[
\ll (pq)^{\epsilon} \left( \frac{pqC}{q^{3/2}} + \frac{\sqrt{NpqC}}{q} \right)
\]  
(5.26)

\[
\ll (pq)^{\epsilon} \sqrt{Npq} \left( \frac{C \sqrt{L}}{q^{3/2}} + \frac{\sqrt{C}}{q} \right)
\]  
(5.27)
We can establish the bound for $S_{dual,p_1} = p_1$ similarly. If we follow the same method for $S_{dual,c_1} = c_1(N)$, we obtain

$$S_{dual,c_1} = c_1(N) \ll (pq)^e \frac{\sqrt{NpqL}}{C}.$$  \hspace{1cm} \text{(5.28)}

We can prove slightly better bounds for these sums, but it is unnecessary for our purpose.

We exhibit a satisfactory bound for the contribution of terms with $q_2 = q$ to (5.15) in the next lemma.

**Lemma 5.2.** Let $S_{dual,q_2} = q$ be the contribution of terms with $q_2 = q$ to $S_{dual}(N)$ (5.7) i.e

$$S_{dual,q_2}(N) = \frac{1}{pq|C||C|} \sum_r \sum_{c_2} \alpha_r \sum_{c_2} S_{r,c,q_2}(N).$$  \hspace{1cm} \text{(5.29)}

Then

$$S_{dual,q_2}(N) \ll (pq)^e \frac{NL}{q}.$$  \hspace{1cm} \text{(5.30)}

Similarly, let $S_{dual,p_2} = p$ and $S_{dual,c_2} = c$ be the contribution of terms with $p_2 = p$ and $c_2 = c$ to $S_{dual}(N)$ respectively. Then

$$S_{dual,p_2}(N) \ll (pq)^e \frac{NL}{p}.$$  \hspace{1cm} \text{(5.31)}

$$S_{dual,c_2}(N) \ll (pq)^e \frac{NL}{C}.$$  \hspace{1cm} \text{(5.32)}

**Proof.** If $q_2 = q$, then $q_1 = q_2 = 1$. Then

$$S_{dual,q_2}(A, B) = \mu(p_2) \mu(c_2) \frac{rN^2}{(p_2 q c_2)^2 p_3 c_3 \sqrt{p_1}} \times \left( \sum_{n,m} \lambda_f(n) \lambda_g(m) \delta[p_1 rm \equiv n(p_3 c_3)] W_2 \left( \frac{n \cdot m}{A \cdot B} \right) \right).$$  \hspace{1cm} \text{(5.33)}

We shall indicate the argument in the case $p_3 = p$ and $c_3 = c$. ($p_3 = 1$ or $c_3 = 1$ works the same way and we get a better bound). In this case (see (5.17))

$$X = pqc^2 T \quad \text{and} \quad Y = \frac{pqc^2 T}{r}.$$  \hspace{1cm}

We shall apply Voronoi summation to the sum over $n, m$. For this purpose we have to rewrite the congruence condition in terms of additive characters. Without getting
into complete details (the steps are identical to the ones carried out in this section till (5.3)), we get the following “inequality”:

\[
\left(\sum_{n,m} \lambda_f(n)\lambda_g(m)\delta[rm \equiv n(pc)]W_2\left(\frac{n}{A}, \frac{m}{B}\right)\right) \ll \sum_{\pm,\pm} \left(\sum_{n,m} \lambda_f(n)\lambda_g(m)\delta[m \equiv qrn(pc)]\right) W_2^{\pm,\pm} \left(\frac{n}{A'}, \frac{m}{B'}\right), \tag{5.34}
\]

where

\[
W_2^{\pm,\pm}(\xi, \eta) = \frac{AB}{\sqrt{q(pc)^2}} \int_0^\infty \int_0^\infty W_2(x, y)J_f^\pm(4\pi \sqrt{\xi x})J_g^\pm(4\pi \sqrt{\eta y}) \, dx \, dy, \tag{5.35}
\]

\[
A' = \frac{(pc)^2}{A}, \text{ and } B' = \frac{q(pc)^2}{B}.
\]

Using the bounds for derivatives of \(W_2\) (5.10) and Lemma 2.2, we conclude that for all \(j \geq 0\),

\[
W_2^{\pm,\pm} \left(\frac{n}{A'}, \frac{m}{B'}\right) \ll_j \frac{AB}{\sqrt{q(pc)^2}} \min\left\{ 1, \left(\frac{A'}{n}\right)^{-j}, \left(\frac{B'}{m}\right)^{-j}, \left(\frac{A'B'}{mn}\right)^{-j} \right\}. \tag{5.36}
\]

Thus

\[
\left(\sum_{n,m} \lambda_f(n)\lambda_g(m)\delta[rm \equiv n(pc)]W_2\left(\frac{n}{A}, \frac{m}{B}\right)\right) \ll \frac{AB}{\sqrt{q(pc)^2}} \sum_{\substack{n \leq (pq)^d A' \\text{ or } m \leq (pq)^d B' \\text{ or } m \equiv qrn(pc)}} |\lambda_f(n)||\lambda_g(m)|\delta[m \equiv qrn(pc)]. \tag{5.37}
\]

We use Lemma 2.6 to bound the right hand side. Putting this back into (5.33), we get

\[
S_3^{r,c,q2=q}(A, B) \ll (pq)^{\frac{r^2}{cT^2}} \sqrt{qpc} \left( 1 + \sqrt{\frac{A}{pc}} + \sqrt{\frac{B}{qpc}} + \sqrt{\frac{AB}{pc\sqrt{q}}} \right). \tag{5.38}
\]

Using this bound in (5.15) and the upper bounds \(A \leq (pq)^c X\) and \(B \leq (pq)^c Y\), we get

\[
S_2^{r,c,q2=q}(N) \ll (pq)^{\frac{r^2qpc\sqrt{T}}{T}}. \tag{5.39}
\]
Putting this back into (5.29), we get
\[ S^{\text{dual}, q_2 = q}_1(N) \ll \frac{1}{pqL||c||} \sum_r \sum_{c \in C} |\alpha_r|_c \left| S^{\text{dual}, q_2 = q}_2(N) \right| \]
(5.40)
\[ \ll (pq)^{NL} q^{-1}. \]  
(5.41)

The bound for \( S^{\text{dual}, p_2 = p}_1 \) and \( S^{\text{dual}, c_2 = c}_2 \) can be shown along the same lines.

Since \( C = 10L^2 \) and \( L \leq q^{1/2} \), the bounds in Lemma 5.1 and Lemma 5.2 imply that
\[ S^{\text{dual}, q_1 = q}_1(N), S^{\text{dual}, q_2 = q}_1, S^{\text{dual}, c_1 = c}_2, S^{\text{dual}, c_2 = c}_2 \ll (pq)^{\frac{\sqrt{Npq}}{\sqrt{L}}}. \]  
(5.42)

This allows us to get rid of these boundary terms. If the size of \( p \) was large, say \( p \approx q \), then we could have gotten rid of \( p_1 = p \) and \( p_2 = p \) also. But since \( p \) could be very small, this is not possible. Restricting to \( q_3 = q \) and \( c_3 = c \), we rewrite (5.15) as
\[ S_{r,c}^2(N) = \sum_{p_1p_2p_3 = p} \sum_{(\rho_1, A) \leq (pq)^Y X} \mu(p_2) \frac{rN^2}{(p_2)^2p_3q\sqrt{p_1}} \left( \sum_{n,m} \lambda_f(n)\lambda_g(m)\delta[p_1rm \equiv qn(p_3q)W_2 \left( \frac{n}{A}, \frac{m}{B} \right) \right) \].
(5.43)

At this point we rewrite the congruence \( p_1rm \equiv qn(p_3q) \) in (5.43) as an equality,
\[ n = p_1rm + sp_3qc. \]
The value of \( s \) can be negative. Let \( \tilde{p} = (p_1, p_2, p_3) \) be any factorization of \( p \). Since \( n \leq A \) and \( m \leq B \), we get
\[ |s| \leq S_{\tilde{p}} := \max \left\{ \frac{A}{p_3qc}, \frac{p_1rB}{p_3qc} \right\}. \]  
(5.44)

Thus
\[ S_{r,c}^2(N) = \sum_{p_1p_2p_3 = p} \sum_{(\rho_1, A) \leq (pq)^Y X} \mu(p_2) \frac{rN^2}{\sqrt{p_1p_2^2p_3q}} \left( \sum_{|s| \leq S_{\tilde{p}}} \sum_{m} \tilde{\lambda}_f(p_1rm + sp_3qc)\tilde{\lambda}_g(m)W_2 \left( \frac{p_1rm + sp_3qc}{A}, \frac{m}{B} \right) \right). \]  
(5.45)
When \( s = 0 \) in the equation above the sum over \( r, c \) collapses. Hence, we need a separate treatment of the terms with \( s = 0 \). When \( s = 0 \), the inner sum is of the form
\[
S(B) = \sum_{m \leq B} \overline{\lambda_f(m)}\lambda_f(m),
\]
which appears to be the conjugate of the sum we that we started with in (3.1). But, since the length of the \( m \)-sum, \( B = Y = \frac{(pq)^2}{N} \gg pqC^2 \), is greater than the square root of the arithmetic conductor of \( L(f \times g, s) \), we can get some cancellation in this sum. We exhibit a satisfactory bound for the contribution of the diagonal terms i.e \( s = 0 \), in the next Lemma.

**Lemma 5.3.** Let the contribution of the terms with \( s = 0 \) in (5.45) to (5.7) be \( S_{1}^{dual,s=0}(N) \) i.e
\[
S_{1}^{dual,s=0}(N) = \frac{1}{pq|L||C|} \sum_{r} \sum_{c \in C} \alpha_r c S_{2}^{r,c,s=0}(N)
\]
\[
= \frac{1}{pq|L||C|} \sum_{r} \sum_{c \in C} \alpha_r c \sum_{p_1p_2p_3=p} \sum_{\rho_1, A \leq (pq)^{r} X} \mu(p_2) \frac{rN^2}{\sqrt{p_1p_2p_3qc}}
\]
\[
\times \left( \sum_{m} \overline{\lambda_f(p_1rm)}\lambda_g(m)W_2 \left( \frac{p_1rm}{A}, \frac{m}{B} \right) \right).
\]

Then
\[
S_{1}^{dual,s=0}(N) \ll (pq)^{r} \frac{\sqrt{Npq}}{\sqrt{L}}.
\]

**Proof.** Among the three choices i.e \( p_1 = p, p_2 = p, \) or \( p_3 = p \) in (5.47), \( p_3 = p \) makes the largest contribution. We shall exhibit the bound (5.49) in this case. It is straightforward to handle the other two cases using the same method.

\[
S_{1}^{dual,p_3=p,s=0}(N) \ll \frac{N^2}{(pqC)^2|L|} \sum_{r} |\alpha_r| \sum_{\rho_1, A \leq (pq)^{r} X} \left( \sum_{m} \overline{\lambda_f(rm)}\lambda_g(m)W_2 \left( \frac{rm}{A}, \frac{m}{B} \right) \right).
\]

(5.50)

Using the Rankin-Selberg bound (2.9), the sum over \( m \) can be rewritten as follows:
\[
\sum_{m} \overline{\lambda_f(rm)}\lambda_g(m) = \overline{\lambda_f(r)} \sum_{\ell \mid m} \overline{\lambda_f(m)}\lambda_g(m) + \sum_{\ell \mid m} \overline{\lambda_f(rm)}\lambda_g(m)
\]
\[
= \overline{\lambda_f(r)} \sum_{m} \overline{\lambda_f(m)}\lambda_g(m) + O \left( (pq)^{r} |\sigma_f(r\ell)||\lambda_g(\ell)| \frac{B}{L} \right),
\]
\[
(5.51)
\]
\[
(5.52)
\]
\[ \sigma_f(n) = \sum_{d|n} |\lambda_f(d)|, \tag{5.53} \]

and \( l \) has been defined in (3.3). Noting that \( B \leq (pq)^{\epsilon} Y = (pq)^{\epsilon} \frac{(pq)^2}{rN} \), we use Lemma 2.5 to bound the \( m \)-sum. This gives

\[ \sum_m \bar{\lambda}_f(rm)\bar{\lambda}_g(m)W(m/Y) \ll (pq)^\epsilon \left( |\lambda_f(r)\lambda_g(l)| \frac{(pq)^2}{rNL} \right). \]

Substituting this bound for the \( m \)-sum in (5.50) and using \( H_{1/4} \) (2.3) to bound the second term, we get

\[ S_{\text{dual}, p_3 = p, s = 0}^\epsilon (N) \ll (pq)^\epsilon \frac{N}{\sqrt{L}} \ll (pq)^\epsilon \frac{\sqrt{Npq}}{\sqrt{L}}. \tag{5.54} \]

\[ \square \]

As the structure of the sum (5.45) is different depending on whether \( p_1 = p \), \( p_2 = p \), and \( p_3 = p \), we have to treat them separately. Since the size of \( Y \) (see (5.17)) depends on \( r \), it is convenient to dyadically divide the sum over \( r \) in (5.7). Exchanging the sum over \( \{r, c\} \) and \( m \) in (5.7), (5.45), we have

\[ S_{\text{dual}}^\epsilon (N) = \frac{1}{pq|\mathcal{L}||\mathcal{C}|} \sum_{R = 2^\nu \leq L^2} \sum_{r \in [R, 2R]} \sum_{c \in \mathcal{C}} \sum_{\alpha} S_{\nu, c}^{\epsilon} (N) \]

\[ = \sum_{p_1 p_2 p_3 = p} S_{(p_1, p_2, p_3)} (N) + O \left( (pq)^\epsilon \frac{\sqrt{Npq}}{\sqrt{L}} \right), \tag{5.56} \]

where

\[ S_{\bar{p}} = \frac{1}{pq|\mathcal{L}||\mathcal{C}|} \frac{N^2}{\sqrt{p_1^2 p_2^2 p_3 q}} \sum_{R = 2^\nu \leq L^2} \sum_{(p_1, A, S)} \sum_{m \leq B} \sum_{(p_2, B, \leq (pq)^\epsilon Y_{\bar{p}})} \]

\[ \left( \sum_{r \in [R, 2R]} \sum_{c \in \mathcal{C}} \sum_{|s| \neq 0, \leq S_{\bar{p}}} \frac{r\alpha r}{c^2} \bar{\lambda}_f(p_1 rm + p_3 qcs)W_2 \left( \frac{p_1 rm + p_3 qcs}{A}, \frac{m}{B} \right) \right), \tag{5.57} \]

\[ X_{\bar{p}} = \frac{p_1 (p_2 p_3 q^2)^2}{N}, Y_{\bar{p}} = \frac{(p_2 p_3 q^2)^2}{RN}, \text{ and } S_{\bar{p}} = \max \left\{ \frac{A}{p_3 q^2}, \frac{p_1 r B}{p_3 q^2} \right\}. \tag{5.60} \]
For example: if $\vec{p} = (1, 1, p)$, then

$$S_{(1,1,p)} = \frac{1}{pq|\mathcal{L}||\mathcal{C}|} \frac{N^2}{pq} \sum_{R=2^\nu \leq L^2} \sum_{(\rho_1, A \leq (pq)^r X_3)} \sum_{m \leq B} \lambda_g(m) \times$$

$$\left( \sum_{r \leq [R,2R]} \sum_{c \in \mathcal{C}} \sum_{|s| \neq 0, \leq S_3} \frac{r \alpha}{c^2} \lambda_f(r m + pq cs) W_2 \left( \frac{r m + pq cs}{A}, \frac{m}{B} \right) \right),$$

$$X_3 = \frac{(pqc)^2}{N}, Y_3 = \frac{(pqc)^2}{RN}, \text{ and } S_3 = \max \left\{ \frac{A}{pqc}, \frac{RB}{pqc} \right\}.$$  

(5.63)

We plan to eliminate the oscillation due to $\lambda_g(m)$ in (5.57) using Cauchy-Schwarz inequality, leading us to a shifted convolution problem. Our bounds for the shifted convolution problem are not optimal if the weight functions don’t have small Sobolev norms (see the dependence on $K_1$ and $K_2$ in (1.5)). If $A$ and $p_1 r B$ are not of the same size, then the weight functions appearing in the shifted convolution problem have large derivatives. In order to circumvent this issue, we separate the weight function $W_2(x, y)$ as a product of a function in $x$ and $y$. There are many standard ways of doing this. Since $W_2$ is compactly supported smooth function with support contained in $[1/2, 1] \times [1/2, 1]$, using Fourier inversion formula, we can see that

$$W_2(x, y) = \int_{z \in \mathbb{R}} \hat{W}_2(x, z) e^{-zy} dz,$$

where $\hat{W}_2(x, z)$ is the Fourier transform of $W_2$ with respect to the second variable i.e

$$\hat{W}_2(x, z) = \frac{1}{2\pi} \int_{\eta \in \mathbb{R}} W_2(x, \eta) e^{-\eta z} d\eta.$$

Using the estimate (5.10) for the derivatives of $W_2$ and integration by parts, we see that for all $N, j \geq 0$

$$\left( \frac{d}{dx} \right)^j \hat{W}_2(x, z) \ll_{N,j} \min \left\{ 1, \frac{1}{|z|^N} \right\}.$$  

(5.64)

Choosing $N = 2019 \epsilon^{-1}$ in (5.64), we have

$$W_2(x, y) = \int_{|z| \leq (pq)^r} \hat{W}_2(x, z) e^{-zy} dz + O \left( \frac{1}{(pq)^{2018}} \right).$$  

28
Substituting this expression in (5.57), we get

\[
S_p = \frac{1}{pq|\mathcal{L}||\mathcal{C}|\sqrt{p_1p_2p_3q}} \int_{|z| \leq (pq)^x} \sum_{R=2^\nu \leq L^2} \sum_{(p_1,A \leq (pq)^x \mathcal{X}_p)} \sum_{m \leq B} \lambda_g(m)e\left(-\frac{zm}{B}\right) \times \tag{5.57}
\]

\[
\left(\sum_{r \in [R,2R]} \sum_{c \in \mathcal{C}} \sum_{|s| \neq 0} \frac{r \alpha_r c^2 \lambda_f(p_1rm + p_3qcs)}{\lambda_f(p_1'r'm + p_3qcs)} W_{2,z}\left(\frac{p_1'r'm + p_3qcs}{A}\right)\right) + O((pq)^{-1000}),
\]

where

\[
W_{2,z}(x) := \widehat{W}_2(x, z).
\]  

(5.66)

\[ W_{2,\nu}(x) := \widehat{W}_2(x, z). \]  

(5.67)

\[ W_{2,\nu}(x) \] is a smooth function with compact support contained in \([1/2, 1]\) with derivatives satisfying (5.64). Applying Cauchy-Schwarz inequality on the sum over \(m\) and taking the supremum over \(z\) (using the Rankin-Selberg bound (2.9)), we have

\[ (S_p)^2 \ll (pq)^x \left(\frac{N}{pq}\frac{p_1}{(p_2|\mathcal{L}||\mathcal{C}|)^2}\right)^4 \sup_z \sum_{R=2^\nu \leq L^2} \sum_{(p_1,A \leq (pq)^x \mathcal{X}_p)} \sum_{(p_2,B \leq (pq)^x \mathcal{Y}_p)} R^2 B \sum_{r,r' \in [R,2R]} \sum_{|t||t'| \neq 0} \beta_t \beta_{t'} \]

\[
\left(\sum_{m} \lambda_f(p_1rm + p_3qts) \lambda_f(p_1'r'm + p_3qts) W_{2,z}\left(\frac{p_1'r'm + p_3qts}{A}\right) W_{2,z}\left(\frac{p_1'r'm + p_3qts}{A}\right)\right),
\]

where

\[
\gamma_r := \frac{r}{R} \alpha_r \leq 2|\alpha_r|, \tag{5.69}
\]

\[
\beta_t := \sum_{c \in \mathcal{C}, c|t} \frac{C^2}{c^2} \leq 4d(t), \tag{5.70}
\]

and

\[
T_p = 2CS_p = \max \left\{ \frac{A}{p_3q}, \frac{p_1rB}{p_3q} \right\}. \tag{5.71}
\]

The sum over \(m\) leads us to our shifted convolution problem. The important feature to observe above is that when \(p_3 = p\) (this is the important case for large \(p\)), the shifts (defined in 6.3) are multiples of \(p\), which is the level of the automorphic form \(f\). On a first reading, the reader can skip the next section assuming the contents of Theorem 1.2, without losing continuity.

### 6. Shifted convolution problem

We recall the \(\delta\) method of Duke, Friedlander, and Iwaniec [4]. We use the version due to Heath-Brown in [10]. The following Lemma is from Munshi’s paper [29, Lemma...
In this section we shall use the letter $q$ to denote the moduli in the circle method. This should not be confused with the level of the modular form $g$ in the previous sections.

**Lemma 6.1.** For any $Q \geq 1$, there exists a positive constant $c_0$ and a smooth function $h(x, y)$ defined on $(0, \infty) \times \mathbb{R}$, such that

$$
\delta(n, 0) = c_0 \sum_{q=1}^{\infty} \frac{1}{q} \sum_{\gamma(q)} e\left(\frac{\gamma n}{q}\right) h\left(\frac{q}{Q}, \frac{n}{Q^2}\right).
$$

(6.1)

The constant $c_0$ satisfies $c_0 = 1 + O_A(Q^{-A})$ for any $A > 0$. Moreover $h(x, y) \ll x^{-1}$ for all $y$, and $h(x, y)$ is non-zero only for $x \leq \max\{1, 2y\}$. If $|y| \leq \frac{x}{2}$ and $x \leq 1$, then

$$
\frac{\partial}{\partial y} h(x, y) = 0.
$$

Furthermore for all $N, j \geq 0$ and $x \leq 1$, $h$ satisfies

$$
y^j \frac{\partial^j}{\partial y^j} h(x, y) \ll_N x^N + \min\{1, (x/|y|)^N\} \ll 1.
$$

(6.2)

**Remark 6.1.** We have normalized $h(x, y)$ differently from [29]. We have multiplied the $h$ in Munshi’s paper by $x$ (see Lemma 23, [29]).

**Lemma 6.2.** For any $0 \leq x \leq 1$ and $h$ as in (6.1),

$$
\int_{\mathbb{R}} |h(x, y)| dy \ll x.
$$

(6.3)

**Proof.** Using (6.2)

$$
\int_{\mathbb{R}} |h(x, y)| dy = \int_{|y| \leq x} |h(x, y)| dy + \int_{|y| \geq x} |h(x, y)| dy
$$

(6.4)

$$
\ll x + \int_{|y| \geq x} \frac{x^2}{|y|^2} dy
$$

(6.5)

$$
\ll x.
$$

(6.6)

□

In practice to detect to $n = 0$ for a sequence of integers in the range $[-X, X]$, it is logical to choose $Q = 2\sqrt{X}$, so that in the generic range for $q$ there is no oscillation of the weight function $h(x, y)$.

**Definition 6.3.** Let $K_1, K_2, M_1, M_2 \geq 1$, $W : [1/2, 3] \times [1/2, 3] \to \mathbb{C}$ be a compactly supported smooth function satisfying

$$
W^{(i,j)}(x, y) \ll K_1^i K_2^j,
$$

(6.7)
for all $i, j \geq 0$. Let $a, b, c, d$ be integers and $a, b \neq 0$. If $f$ and $g$ are cuspidal automorphic forms (modular or Maass) of arbitrary level and nebentypus we define the shifted convolution sum $S$ as,

$$S_{f,g}(a, b, c, d, M_1, M_2) = \sum_{m=1}^{\infty} \lambda_f(am + c)\lambda_g(bm + d)W_1\left(\frac{am + c}{M_1}, \frac{bm + d}{M_2}\right). \quad (6.8)$$

We define $(ad - bc)$ to be the “shift” of the sum $S_{f,g}$.

We first establish a bound for $S_{f,g}(a, b, c, d, M_1, M_2)$ in the case $(ab, p) = 1$.

**Remark 6.2.** We have assumed the primality of $p$, as this is the only case we shall need in this paper. The proof can be easily extended to include any $p \geq 1$. We can improve the bounds in this Lemma by using Spectral theory for $GL(2)$ to bound a sum of Kloosterman sums (see equation 6.27).

**Proof of Theorem 1.2.** The first bound in (1.4) follows from the Rankin-Selberg bound (2.9). The second bound follows from hypothesis $H_\theta$ (2.3).

We separate the oscillation of $f$ and $g$ in (6.8) by introducing a $\delta$ symbol i.e

$$S = \sum_{n_1 \equiv c(a), n_2 \equiv d(b)} \lambda_f(n_1)\lambda_g(n_2)\delta\left(\frac{n_1 - c}{a} = \frac{n_2 - d}{b}\right)W_1\left(\frac{n_1}{M_1}, \frac{n_2}{M_2}\right). \quad (6.9)$$

We imitate Munshi’s ideas (see [29]) to factor the $\delta$ symbol as a congruence mod $p$,

$$\frac{n_1 - c}{a} \equiv \frac{n_2 - d}{b} (p), \quad (6.10)$$

followed by the equality

$$\frac{n_1 - c}{ap} - \frac{n_2 - d}{bp} = 0. \quad (6.11)$$

If $|ad - bc| > 3(|b|M_1 + |a|M_2)$ then the shifted convolution sum $S_{f,g}$ is 0, as the summation is empty. Thus the difference $|\frac{n_1 - c}{ap} - \frac{n_2 - d}{bp}|$ is bounded by $\frac{X}{2}$, where

$$X = 12\left(\frac{M_1}{|a|} + \frac{M_2}{|b|}\right). \quad (6.12)$$

Let us define

$$Q = \sqrt{\frac{X}{p}}. \quad (6.13)$$

We assume $Q \geq 1$, otherwise (1.4) implies (1.5) and (1.6). We pick up the congruence (6.10) using additive characters modulo $p$ and the equality (6.11) using the Duke,
Friedlander, and Iwaniec circle method (6.1), with \(Q\) as above. Thus

\[
\delta \left( \frac{n_1 - c}{a} = \frac{n_2 - d}{b} \right) = \frac{c_0}{pQ} \sum_{q=1}^{\infty} \frac{1}{q} \sum_{\alpha(p)} \sum_{\gamma(q)^*} e \left( \frac{\alpha(n_1 - c)}{ap} - \frac{\alpha(n_2 - d)}{bp} \right) e \left( \frac{\gamma(n_1 - c)}{apq} - \frac{\gamma(n_2 - d)}{bpq} \right) h \left( \frac{q}{Q} \frac{n_1 - c - n_2 - d}{ap - bp} Q^2 \right).
\]

Substituting this into (6.9), we get

\[
S = \frac{c_0}{abpQ} \sum_{q=1}^{\infty} \frac{1}{q} \sum_{\alpha(ap)} \sum_{\beta(bp)} \sum_{\alpha \equiv \beta(p)} \sum_{\gamma(q)^*} \lambda_f(n_1) \lambda_g(n_2) e \left( \frac{\alpha(n_1 - c)}{ap} - \frac{\beta(n_2 - d)}{bp} \right) e \left( \frac{\gamma(n_1 - c)}{apq} - \frac{\gamma(n_2 - d)}{bpq} \right) h \left( \frac{q}{Q} \frac{n_1 - c - n_2 - d}{ap - bp} Q^2 \right) W \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right).
\]

We have expanded the congruence \(n_1 \equiv c(a)\) and \(n_2 \equiv d(b)\) using additive characters and combined the frequency mod \(p\) using the Chinese remainder theorem. Combining the frequency \(\gamma\) modulo \(q\) with \(\alpha(ap)\) and \(\beta(bp)\), we get

\[
S = \frac{c_0}{abpQ} \sum_{q=1}^{\infty} \frac{1}{q} \sum_{\alpha(ap)} \sum_{\beta(bp)} \sum_{\alpha \equiv \beta(p)} \sum_{\gamma(q)^*} \lambda_f(n_1) \lambda_g(n_2) e \left( \frac{\alpha(n_1 - c)}{apq} - \frac{\beta(n_2 - d)}{bpq} \right) W \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right),
\]

(6.16)

where

\[
W_1 \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right) = h \left( \frac{q}{Q} \frac{n_1 - c - n_2 - d}{ap - bp} Q^2 \right) W \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right).
\]

(6.17)

Using (6.2) and (6.7), we see that for all \(i, j \geq 0\)

\[
W_1^{(i,j)}(x, y) \ll K'_1 K''_2 \left( \frac{Q}{q} \right)^{i+j}.
\]

(6.18)

We plan to apply the Voronoi summation formula (2.1) to the \(n_1\) and \(n_2\) sum in (6.16). The modulus in Voronoi summation depends on gcd(\(\alpha, apq\)) and gcd(\(\beta, bpq\)). Since \(\alpha\) and \(\beta\) are coprime to \(q\), the gcd(\(\alpha, apq\)) and gcd(\(\beta, bpq\)) are gcd(\(\alpha, ap\)) and gcd(\(\beta, bp\)) respectively. We will assume that (\(\alpha, ap\) = (\(\beta, bp\) = 1 and (\(q, \text{Rad}(ab)) = 1, to simplify notation. (\text{Rad}(a) is the radical of a, defined to be \text{Rad}(a) = \prod_{p|a} p^\infty.) We shall indicate the changes required to handle the general case towards the end of the
proof. The same bound established below holds below for the other cases too. Let us define

\[ S_{1,1,1} = c_0 \frac{1}{abpQ} \sum_{(q,ab)=1} \sum_{\alpha \divides \beta} \frac{\lambda_f(n_1)\lambda_g(n_2)e\left(\frac{\alpha(n_1-c)}{apq} - \frac{\beta(n_2-d)}{bpq}\right)}{n_1 n_2} \mathcal{W}_1\left(\frac{n_1}{M_1}, \frac{n_2}{M_2}\right). \]

(6.19)

The \((1,1,1)\), in the superscript of \(S\), refers to the greatest common divisors \((\alpha, ap\)), \((\beta, bp\)), and \((q, \text{Rad}(ab))\) respectively. Applying the Voronoi summation formula to \(n_1\) and \(n_2\) in (6.19), we get

\[ S_{1,1,1} = c_0 \frac{1}{abpQ} \sum_{(q,ab)=1} \sum_{\alpha \divides \beta} \frac{\lambda_f(n_1)\lambda_g(n_2)e\left(\frac{\alpha n_1}{apq} - \frac{\beta n_2}{bpq}\right)}{n_1 n_2} e\left(-\frac{\alpha c}{apq} + \frac{\beta d}{bpq}\right) \mathcal{W}_2\left(\frac{M_1 n_1}{(apq)^2}, \frac{M_2 n_2}{(bpq)^2}\right) + \sum \ldots, \]

(6.20)

where

\[ \mathcal{W}_2(\xi, \eta) = \frac{M_1 M_2}{abc(\sqrt{\eta})^2} \int_0^\infty \int_0^\infty \mathcal{W}_1(x, y) J^+_f\left(4\pi \sqrt{\xi x}\right) J^+_g\left(4\pi \sqrt{\eta y}\right) dx dy. \]

(6.21)

The summation \(\sum_{\ldots}\), refers to the three other terms coming from the \(+\) terms on the right hand side of the Voronoi formula (2.1). They have the same structure and can be handled similarly. Lemma 2.2 and the bound (6.18) implies that for all \(N, M \geq 0\)

\[ W_2(\xi, \eta) \ll \left(\frac{(K_1 Q)^2}{q^2\xi}\right)^{-N} \left(\frac{(K_2 Q)^2}{q^2\eta}\right)^{-M}. \]

Thus, choosing \(N\) and \(M\) sufficiently large, we can restrict the summation over \(n_1\) and \(n_2\) in (6.20) to

\[ n_1 \ll (pX)^e \frac{(K_1 apQ)^2}{M_1} = N_1 \quad \text{and} \quad n_2 \ll (pX)^e \frac{(K_2 bpQ)^2}{M_2} = N_2. \]

(6.22)

In the complementary range for \(n_1\) and \(n_2\) we shall use standard estimates for Bessel functions [19, Lemma C.2] to bound \(W_2\) as follows. We can assume \(bM_1 \leq aM_2\), without loss of generality. Making the substitution \(x \mapsto x/M_1\) and \(y \mapsto y/M_2\) in the
integral (6.21), we have

\[
W_2(\xi, \eta) = \frac{1}{ab(pq)^2} \int_{M_1/2}^{3M_1} \int_{M_2/2}^{3M_2} W \left( \frac{x}{M_1}, \frac{y}{M_2} \right) h \left( \frac{q}{Q}, \frac{x-y-ad-bc}{abX} \right) J_f^+ \left( 4\pi \sqrt{\frac{\xi x}{M_1}} \right) J_g^+ \left( 4\pi \sqrt{\frac{\eta y}{M_2}} \right) dx dy.
\]

\[
W_2(\xi, \eta) \ll \frac{1}{(\xi \eta)^{1/4}(abpq)^2} \int_{bM_1/2}^{3bM_1} \int_{aM_2/2}^{3aM_2} \left| W \left( \frac{x}{bM_1}, \frac{y}{aM_2} \right) \right| \left| h \left( \frac{q}{Q}, \frac{x-y+ad-bc}{abX} \right) \right| dx dy
\]

\[
\ll \frac{1}{(\xi \eta)^{1/4}(abpq)^2} \int_{bM_1/2}^{3bM_1} \int_{aM_2/2}^{3aM_2} \left| W \left( \frac{x}{bM_1}, \frac{x-u+h}{aM_2} \right) \right| \left| h \left( \frac{q}{Q}, \frac{u}{abX} \right) \right| du dx,
\]

where \( h = ad - bc \) is the shift. Using estimate (6.3) to bound the integral over \( u \), we get the bound

\[
W_2(\xi, \eta) \ll \frac{M_1 M_2 q}{(\xi \eta)^{1/4} ab(pq)^2 Q}.
\]

Using the Chinese remainder theorem to split the exponential sum modulo \( pq \) and \( ab \), equation (6.20) can be rewritten as

\[
S_{1,1,1} = \frac{c_0}{abpq} \sum_{(q, ab) = 1} \frac{1}{q} \sum_{n_1 \leq N_1} \sum_{n_2 \leq N_2} \frac{\Lambda_f(n_1) \Lambda_g(n_2)}{n_1 n_2} S(\bar{ab}(bn_1 - an_2); ab(ad - bc); pq)
\]

\[
S(pqn_1; -pqc; a) S(-pqn_2; pqd; b) W_2 \left( \frac{M_1 n_1}{(apq)^2}, \frac{M_2 n_2}{(bpq)^2} \right),
\]

where \( S(m, n, c) = \sum_{c(c)} e \left( \frac{am+cn}{c} \right) \) is the Kloosterman sum. We shall use the Weil bound

\[
S(m, n, c) \ll (m, n, c)^{1/2} \epsilon^{1/2}.\]

The crucial point here is that the shift \( (ad - bc) \) will be a multiple of \( p \) in our application. In this case, the Kloosterman sum modulo \( p \) becomes a Ramanujan sum (5.14). If \( n \equiv 0(c) \), then

\[
S(m, n, c) = S(m, 0, c) = r_c(m) \ll (m, c)^{\epsilon}.
\]

This allows us to save an additional \( \sqrt{p} \) in the Kloosterman sum modulo \( p \). We first consider the case \( ad - bc \equiv 0(p) \). Using the estimate (6.26) to bound \( W_2 \) along with
the Weil bound in (6.27), we get
\[
S^{1.1,1} \ll (abpM_1M_2)^{(M_1M_2)^{3/4}/abp^2Q^2} \sum_{q \leq Q} \frac{1}{\sqrt{q}} \sum_{n_1 \leq N_1} \sum_{n_2 \leq N_2} \frac{|\lambda_f(n_1)|}{n_1^{1/4}} \frac{|\lambda_g(n_2)|}{n_2^{1/4}} \times
\]
\[(n_1, a)^{1/2}(n_2, b)^{1/2}(ad - bc, q)^{1/2}(bn_1 - an_2, p) + O((pM_1M_2)^{-2018}).
\]

Isolating the biggest common divisor of \((\text{Rad}(a), n_1)\) and \((\text{Rad}(b), n_2)\), we get
\[
S^{1.1,1} \ll (abpX)^{(M_1M_2)^{3/4}/abp^2Q^2} \sum_{d_3|(ad - bc)} \frac{1}{\sqrt{q}} \sum_{d_1|a, d_2|b} (d_1d_2)^{1/2} \sum_{d_1|d_1', d_2|d_2', (\text{Rad}(d_1), d_2|\text{Rad}(d_2))} \frac{|\lambda_f(d_1')|}{d_1^{1/4}} \frac{|\lambda_f(d_2')|}{d_2^{1/4}}
\]
\[
\left\{ \begin{array}{l}
\sum_{n_1 \leq \frac{N_1}{d_1}, n_2 \leq \frac{N_2}{d_2}} \frac{|\lambda_f(n_1)|}{n_1^{1/4}} \frac{|\lambda_g(n_2)|}{n_2^{1/4}} + \sum_{n_1 \leq \frac{N_1}{d_1}, n_2 \leq \frac{N_2}{d_2}} \frac{|\lambda_f(n_1)|}{n_1^{1/4}} \frac{|\lambda_g(n_2)|}{n_2^{1/4}}
\end{array} \right\}.
\] (6.30)

We use Lemma 2.6 with \(X = \frac{N_1}{d_1}\), \(Y = \frac{N_2}{d_2}\) and \(c = p\), to bound the first term inside the bracket (see equation (6.22) for the definition of \(N_1\) and \(N_2\)). This gives,
\[
S^{1.1,1} \ll (abpX)^{(M_1M_2)^{3/4}/abp^2Q^2} \times (N_1N_2)^{3/4} \sum_{q \leq Q} \frac{1}{\sqrt{q}}
\]
\[
\ll (abpX)^{(K_1K_2)^{3/2}/abp^1/4X^{3/4}}.
\] (6.31)

The right hand side of the inequality above also bounds the second term in the bracket of equation (6.30). This proves the bound claimed in equation (1.6).

If \((ad - bc, p) = 1\), then the proof works out identically till equation (6.27). But we have a Kloosterman modulo \(p\) instead of the Ramanujan sum now. Thus, using the Weil bound, we won’t have the first term in the bracket of (6.30) and the second term would be multiplied by \(\sqrt{p}\). Hence, the upper bound of equation (1.5) is multiplied by an additional \(\sqrt{p}\).
We had restricted our attention to the terms satisfying $\text{gcd}(\alpha, apq) = (\beta, bpq) = (q, ab) = 1$ in (6.16):

$$
S = \frac{c_0}{abpQ} \sum_{q=1}^{\infty} \frac{1}{q} \sum_{n_1 \equiv \beta(qp) (\alpha, q) = 1} \lambda_f(n_1) \lambda_g(n_2) e \left( \frac{\alpha(n_1 - c)}{apq} - \frac{\beta(n_2 - d)}{bpq} \right) W_1 \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right).
$$

(6.33)

We now indicate the modifications necessary to handle the general case. Let us first consider the contribution of the frequencies $\alpha$ which are divisible by $p$, denoted by $S^{\alpha \equiv 0(p)}$. Since $\beta \equiv \alpha(p)$, this implies $\beta \equiv 0(p)$. Furthermore, since $(\alpha, q) = 1$ and $p$ divides $\alpha$, $(q, p) = 1$. We decompose $a = a_1 a_2$, $b = b_1 b_2$, and $q = q_1 q_2$ such that $(\alpha, a_1 p) = a_1 p$, $(\beta, b_1 p) = b_1 p$, and $(q, \text{Rad}(ab)) = q_1$ respectively. Then

$$
S^{\alpha \equiv 0(p)} = \sum_{a_1 | a} \sum_{b_1 | b} \sum_{q_1 | \text{Rad}(ab)} \sum_{q_1 \leq Q} S^{a_1, b_1, q_1},
$$

(6.34)

where

$$
S^{a_1, b_1, q_1} = \frac{c_0}{abpQ} \sum_{(q_2, ab) = 1} \frac{1}{q_1 q_2} \sum_{\alpha \equiv a_2 q_1 q_2 (\alpha, q_1 q_2)} \sum_{\beta \equiv b_2 q_1 q_2 (\beta, q_1 q_2)} \sum_{n_1 \equiv n_1 (\alpha, q_1 q_2)} \lambda_f(n_1) \lambda_g(n_2) e \left( \frac{\alpha(n_1 - c)}{a_2 q_1 q_2} - \frac{\beta(n_2 - d)}{b_2 q_1 q_2} \right) W_1 \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right),
$$

(6.35)

$$
W_1 \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right) = h \left( \frac{q_1 q_2}{Q^2}, \frac{n_1 - c}{ap} - \frac{n_2 - d}{bp} \right) W \left( \frac{n_1}{M_1}, \frac{n_2}{M_2} \right).
$$

(6.36)

We apply Voronoi summation (2.1) formula to the $n_1$ and $n_2$ sum. The only difference in this case is that moduli $(a_2 q_1 q_2$ and $b_2 q_1 q_2$) are not divisible by $p$. The
Voronoi transformed sum is
\[ S_{a_1 b_1 q_1} = \frac{c_0}{abpq_1} \sum_{(q_2, a_2 b_2) = 1} \frac{1}{q_2} \sum_{\alpha(a_2 q_1, q_2) \alpha(b_2 q_1, q_2) \alpha_1 \beta_1 \beta_2} \sum_{n_1 n_2} \lambda_f(n_1) \lambda_g(n_2) e \left( -\frac{p\alpha n_1}{a_2 q_1} + \frac{p\beta n_2}{b_2 q_1} \right) e \left( -\frac{\alpha c}{a_2 q_1} + \frac{\beta d}{b_2 q_1} \right) W_2 \left( \frac{M_1 n_1}{p(a_2 q_1)^2}, \frac{M_2 n_2}{p(b_2 q_1)^2} \right) \]
\[ + \sum_{(-, +)} \sum_{(+, -)} \sum_{(-, -)} \ldots, \] (6.37)

where
\[ W_2 (\xi, \eta) = \frac{M_1 M_2}{a_2 b_2 p q_1} \int_0^\infty \int_0^\infty W_1(x, y) J_f^+ \left( 4\pi \sqrt{\xi x} \right) J_g^+ \left( 4\pi \sqrt{\eta y} \right) d\xi d\eta, \] (6.38)

and the summation \( \sum_{(-, +)} \ldots \), refers to the three other terms coming from the \( \pm \) terms on the right hand side of the Voronoi formula (2.1). Like before, they have the same structure and can be handled similarly. We use the Chinese remainder theorem to rewrite the exponential sums as Kloosterman sums. Then the sum over \( \alpha(a_2 q_1, q_2) \) and \( \beta(b_2 q_1, q_2) \) satisfying \( a_1 \alpha \equiv b_1 \beta(q_1, q_2) \), can be rewritten as
\[ \sum_{\alpha(a_2 q_1, q_2) \beta(b_2 q_1, q_2) a_1 \alpha \equiv b_1 \beta(q_1, q_2)} e \left( -\frac{p\alpha n_1}{a_2 q_1} + \frac{p\beta n_2}{b_2 q_1} \right) e \left( -\frac{\alpha c}{a_2 q_1} + \frac{\beta d}{b_2 q_1} \right) = \]
\[ S(abq_1(ad - bc); b_2 a_2 q_1 p(a_2 b_1 n_2 - a_1 b_2 n_1); q_2) \]
\[ \frac{1}{q_1} \sum_{\delta(n_1)} S(\delta a_1 - \frac{\delta}{q_2}; -p\delta n_1; a_2 q_1) S(\delta b_1; \delta a_2 q_2; b_2 q_1). \] (6.39)

Using the Weil-bound for Kloosterman sum (6.28), this is bounded by
\[ (a_2 b_2 q_2 q_1)^{1/2 + \epsilon} (ad - bc, q_2)^{1/2} (n_1, a_2 q_1)^{1/2} (n_2, b_2 q_1)^{1/2}. \]
Substituting this bound into (6.37), using the Rankin-Selberg bound (2.9), we get
\[ S_{a_1 b_1 q_1} \ll (abpX)^\epsilon (K_1 K_2)^{3/2} \sqrt{abq_1}^{-1/2} p^{-3/4} X^{3/4}. \]
This implies
\[ S^{a_1 b_1 q_1} \ll (abpX)^\epsilon (K_1 K_2)^{3/2} \sqrt{abp^{-3/4}} X^{3/4}, \]
which is stronger than the one claimed in (1.6).

We are left with the contribution of frequencies \( \alpha \) such that \( (\alpha, p) = 1 \) to (6.27), denoted by \( S^{(\alpha, p)} = 1 \). If \( \alpha \neq 0(p) \), then \( \beta \neq 0(p) \), as \( \alpha \equiv \beta(pq) \). Let us decompose \( a =$
\(a_1a_2, b = b_1b_2, \) and \(q = q_1q_2\) such that \((\alpha, apq) = a_1, (\beta, bpq) = b_2,\) and \((q, \Rad(ab)) = q_1\) respectively. Then

\[
S^{(a,p)=1} = \frac{c_0}{abpq} \sum_{q_1|\Rad(ab)} \sum_{(q_2,ab)=1} q_1q_2 \sum_{a_1|a} \sum_{b_1|b} \frac{1}{a_2pq_1q_2} \alpha(a_2pq_1q_2) \beta(b_2pq_1q_2) \sum_{\alpha \equiv \beta \mod b_1} \beta(pq_1q_2). \\
\sum_{n_1} \lambda_f(n_1) \lambda_g(n_2)e\left(\frac{\alpha(n_1-c)}{a_2pq_1q_2} - \frac{\beta(n_2-d)}{b_2pq_1q_2}\right) W_1\left(n_1 M_1 n_2 M_2\right). \\
(6.40)
\]

We apply Voronoi summation to the \(n_1\) and \(n_2\) sum modulo \(a_2pq_1q_2\) and \(b_2pq_1q_2\) respectively, and proceed exactly like the proof above for \(S^{1,1,1}.\

\]

We now consider the case \(p|(a,b).\)

**Theorem 6.4.** Let \(p\) be a prime number or \(p = 1.\) Let \(a, b, c, d\) be integers such that \(a\) and \(b\) are non-zero and \(p\) divides \(\gcd(a,b).\) Let \(f, g\) be non-exceptional cuspidal newforms (modular or Maass) of level \(p\) and trivial nebentypus. For any \(M_1, M_2, K_1, K_2 \geq 1,\) we claim the following upper bounds for the shifted convolution sum \(S_{f,g}:\\
S_{f,g}(a,b,c,d, M_1, M_2) \ll p^e \min\{(M_1M_2)^{1/2}, (M_1M_2)^\theta X\},
\]
where \(X\) has been defined in (6.12) and \(\theta\) is the exponent towards Ramanujan-conjecture for \(f\) and \(g.\) Further, if the shift \(ad - bc\) is non-zero, then

\[
S_{f,g}(a,b,c,d, M_1, M_2) \ll p^e (K_1K_2)^{3/2} \sqrt{ab} X^{3/4}.
\]

**Proof.** The proof works exactly like the previous one. We just skip the step of factoring the delta symbol through a congruence mod \(p\) i.e (6.10). Let \(Q = \sqrt{X}.\) We use the following expression

\[
\delta\left(\frac{n_1-c}{a} = \frac{n_2-d}{b}\right) = \frac{c_0}{Q} \sum_{q=1}^\infty \frac{1}{q} \sum_{\gamma(q)=1} e\left(\gamma(n_1-c) \frac{aq}{\gamma(q)} - \gamma(n_2-d) \frac{bq}{\gamma(q)}\right) h\left(\frac{q}{Q}, \frac{n_1-c - n_2-d}{Q}\right).
\]

instead of (6.14) and proceed identically from here.

\]

**Theorem 6.5.** Let \(p\) be fixed. Let \(a, b, c, d\) be integers such that \(a\) and \(b\) are positive. Further, let \(f, g\) be non-exceptional newforms of level \(p\) and any nebentypus.

For any \(M_1, M_2, K_1, K_2 \geq 1,\) we claim the following upper bounds for the shifted convolution sum \(S_{f,g}:\\
S_{f,g}(a,b,c,d, M_1, M_2) \ll_p \min\{(M_1M_2)^{1/2}, (M_1M_2)^\theta X\},
\]

(6.44)
where \( X \) has been defined in (6.12) and \( \theta \) is the exponent towards Ramanujan-conjecture for cusp forms on congruence subgroups of \( SL_2(\mathbb{Z}) \). If the shift \( ad - bc \) is non-zero then,

\[
S_{f,g}(a, b, c, d, M_1, M_2) \ll_{p,K_1,K_2} (abX)^{1/2+\theta}
\]  

(6.45)

where the dependence on \( p \) is polynomial.

**Proof.** Theorem 1.3 of \cite{2} shows that if the shift \( ad - bc \neq 0 \) and \( a, b \) are coprime, then

\[
S_{f,g}(a, b, c, d, M_1, M_2) \ll_{p,K_1,K_2} (bM_1 + aM_2)^{1/2+\theta} \ll (abX)^{1/2+\theta},
\]

where \( \theta \) is the exponent towards Ramanujan conjecture for cusp forms on congruence subgroups and \( X \) is as in (6.12). Using a minor modification we can handle the case \( (a, b) > 1 \). The basic idea is to handle the sum over \( q \) in (6.27) using spectral theory for \( GL(2) \). (Blomer \cite{2} carries this out using Jutila’s circle method.) \( \square \)

---

7. **Proof of Theorem 1.1 and Theorem 1.3**

Let us split (5.68) into the Diagonal part \( D \) (7.2) (terms with zero shift) and the Off-Diagonal part \( O \) (7.13) (terms with non-zero shift):

\[
(S_{\vec{p}})^2 \ll D_{\vec{p}} + O_{\vec{p}} + O((pq)^{-1000})
\]  

(7.1)

Note that the shift is \( p_1p_3q(r't' - r't) \). We first consider the contribution of terms with 0 shift (see 6.3 for definition of shift) in (5.68).

\[
D_{\vec{p}} = (pq)^\epsilon \left( \frac{N}{pq} \right)^4 \frac{p_1}{(p_2|\mathcal{L}||\mathcal{C}|)^2} \sup_{z} \sum_{R=2^v \leq L^2} \sum_{(p_1,A \leq (pq)^\epsilon X_{\vec{p}})} \frac{R^2 B}{C^4} \sum_{r,r' \in [R,2R]} \gamma_r \gamma_{r'} \sum_{\substack{\beta_t \beta_{t'} \beta_t \beta_{t'} \\ |t|,|t'| \neq 0}} \sum_{r' = r't} \lambda_f(p_1rm + p_3qt) \lambda_f(p_1rm' + p_3qt') W_{2,z} \left( \frac{p_1rm + p_3qt}{A} \right) W_{2,z} \left( \frac{p_1rm' + p_3qt'}{A} \right).
\]  

(7.2)

If we assume the Ramanujan conjecture for Fourier coefficients of \( f \), then the bound claimed in the lemma below is straightforward. We have to be a little careful as we want to treat this using Rankin-Selberg bounds alone.

**Lemma 7.1.** For all factorizations \( \vec{p} = (p_1, p_2, p_3) \) of \( p \), if \( D_{\vec{p}} \) is defined as in (7.2), then

\[
D_{\vec{p}} \ll (pq)^\epsilon \frac{N(pq)}{L}
\]  

(7.3)

39
**Proof.** Recall (5.60) and (5.71):

\[
X_{\vec{p}} = \frac{p_1(p_2 p_3 q c)^2}{N}, \quad Y_{\vec{p}} = \frac{(p_2 p_3 q c)^2}{r N}, \quad \text{and} \quad T_{\vec{p}} = \max \left\{ \frac{A}{p_3 q^2}, \frac{p_1 r B}{p_3 q} \right\}.
\]  

(7.4)

\(\gamma_r\) is non-zero only when \(r\)'s are primes or squares of primes and we have dyadically divided the \(r\) sum. Thus if \(r \neq r'\), then \((r,r') = 1\). Hence \(rt' = r't\) implies that one of the following holds:

\[
r = r', \ t = t'\quad \text{(7.5)}
\]

\[
r \neq r', \ t = rk, t' = r'k\quad \text{(7.6)}
\]

In the first case (we shall call this \(D_1\)),

\[
D_{\vec{p}}^1 \ll (pq)^{\epsilon} \left(\frac{N}{pq}\right)^4 \frac{p_1}{(p_2|\mathcal{L}|)|\mathcal{C}|} \sup_R \sum_{R=2^v \leq L^2} \sum_{(p_1 A \leq (pq)^{\epsilon} X_{\vec{p}})} \sum_{(p_2 B \leq (pq)^{\epsilon} Y_{\vec{p}})} \frac{R^2 B}{C^4} \sum_{r \in [R, 2R]} |\gamma_r|^2 \sum_{|t| \leq T_{\vec{p}}} |\beta_t|^2 \left(\sum_m |\lambda_f(p_1 r m + p_3 q t)|^2 \right) W_{2,z} \left(\frac{p_1 r m + p_3 q t}{A}\right) W_{2,z} \left(\frac{p_1 r m + p_3 q t}{A}\right).
\]  

(7.7)

As \(m\) and \(t\) vary in the inner sum, \(n = rp_1 m + tp_3 q\) runs over numbers smaller than \(A\). Since \((rp_1, p_3 q) = 1\)

\[rp_1 m_1 + t_1 p_3 q = rp_1 m_2 + t_2 p_3 q\]

then \(m_1 = m_2 + up_3 q\) and \(t_1 = t_2 - urp_1\). Thus the multiplicity of any \(n\) is at most \(O\left(1 + \frac{B}{p_3 q} + \frac{A}{p_1 p_3 r q^{\epsilon}}\right)\). We recall from (5.69) that \(\beta_t \ll d(t)\) and \(\gamma_r \leq 2|\alpha_r|\). Using the Rankin-Selberg bound (2.9) to bound the sum over \(m\) and \(t\), we get

\[
D_{\vec{p}}^1 \ll (pq)^{\epsilon} \left(\frac{N}{pq}\right)^4 \frac{p_1}{(p_2|\mathcal{L}|)|\mathcal{C}|} \sup_R \sum_{R=2^v \leq L^2} \sum_{(p_1 A \leq (pq)^{\epsilon} X_{\vec{p}})} \sum_{(p_2 B \leq (pq)^{\epsilon} Y_{\vec{p}})} \frac{R^2 B}{C^4} \sum_{r \in [R, 2R]} |\gamma_r|^2 X_{\vec{p}} \left(1 + \frac{Y_{\vec{p}}}{p_3 q}\right)
\]  

(7.8)

\[
\ll (pq)^{\epsilon} \frac{N(pq)}{L}.
\]  

(7.9)

For the second case of zero shift i.e (7.6):

\[
D_{\vec{p}}^2 = (pq)^{\epsilon} \left(\frac{N}{pq}\right)^4 \frac{p_1}{(p_2|\mathcal{L}|)|\mathcal{C}|} \sup_R \sum_{R=2^v \leq L^2} \sum_{(p_1 A \leq (pq)^{\epsilon} X_{\vec{p}})} \sum_{(p_2 B \leq (pq)^{\epsilon} Y_{\vec{p}})} \frac{R^2 B}{C^4} \sum_{r \neq r' \in [R, 2R]} \sum_{|k| \leq 2T_{\vec{p}}/R} \gamma_{r'} \gamma_{r'} \beta_{r' k} \beta_{r' k} \left(\sum_m \lambda_f(r(p_1 m + p_3 q k)) \lambda_f(r'(p_1 m + p_3 q k)) \right) W_{2,z} \left(\frac{r(p_1 m + p_3 q k)}{A}\right) W_{2,z} \left(\frac{r'(p_1 m + p_3 q k)}{A}\right).
\]  

(7.10)
As \( m \) and \( k \) vary in the inner sum, \( n = p_1m + kp_3q \) runs over numbers smaller than \((pq)\epsilon X_{\bar{f}}/r\). Since \((p_1, p_3q) = 1\)

\[
p_1m_1 + k_1p_3q = p_1m_2 + k_2p_3q
\]
then \( m_1 = m_2 + tp_3q \) and \( k_1 = k_2 - tp_1 \). Thus the multiplicity of any \( n \) is at most \( O(1 + (pq)^\epsilon X_{\bar{f}}/p_3q) \). So, the sum over \( m \) and \( k \) is bounded by

\[
\sum_{m,k} \cdots \ll (pq)^\epsilon \left( 1 + \frac{Y_{\bar{f}}}{p_3q} \right) \sum_{n \leq (pq)^\epsilon X_{\bar{f}}/r} |\lambda_f(rn)||\lambda_f(r'n)|
\]

\[
\ll (pq)^\epsilon \left( 1 + \frac{Y_{\bar{f}}}{p_3q} \right) |\sigma_f(r)||\sigma_f(r')|\frac{X_{\bar{f}}}{r},
\]
where \( \sigma_f \) has been defined in (5.53). It is straightforward to verify the bound claimed in the Lemma for \( D_{\bar{f}} \) now.

Having treated the zero-shift terms, we are left with the non-zero shifts \( \mathcal{O}_{\bar{f}} := (pq)^\epsilon \left( \frac{N}{pq} \right)^4 \frac{p_1}{(p_2|\ell||C|^2)^2} \sup_z \sum_{R = 2^z \leq L^2} \sum_{(p_1, A/(pq)^\epsilon Y_{\bar{f}})} \frac{R^2 B}{C^4} \sum_{r, r' \in [R,2R]} \frac{1}{|r| |-r'|} \sum_{|t|, |t'| = 0 \leq T_{\bar{f}} \neq r \neq r'} \beta_t \beta_{t'}
\[
\left( \sum_m \lambda_f(p_1rm + p_3qt)\lambda_f(p_1r'm + p_3qt')W_{2,z} \left( \frac{p_1rm + p_3qt}{A} \right) \frac{p_1r'm + p_3qt'}{A} \right)^z.
\]

We have to treat the cases \( p_1 = p, p_2 = p \) and \( p_3 = p \) separately. We plan to bound the inner sum over \( m \) by applying Theorem 6.4 if \( p_1 = p \) and Theorem 1.2 otherwise.

**Lemma 7.2.** Let \( \mathcal{O}_{(p,1,1)} \) be defined as in (7.13). Then

\[
\mathcal{O}_{(p,1,1)} \ll (pq)^\epsilon \frac{N(pq)L^{25/4}}{p^{7/4} q^{1/4}}.
\]

**Proof.** If \( \bar{f} = (p, 1, 1) \), then \( X_{\bar{f}} = \frac{pq^2}{N}, Y_{\bar{f}} = \frac{pq^2}{RN} \), and \( T_{\bar{f}} = \max \left\{ \frac{A}{q}, \frac{pB}{q} \right\} \) (see (5.60) and (5.71)). Use Theorem (6.4) with \( a = pr, b = pr', c = qt, d = qt', M_1 = M_2 = A, \) and \( K_1 = K_2 = 1 \) to bound the \( m \)-sum. Then,

\[
\sum_{m} \lambda_f(prm + qt)\lambda_f(pr'm + qt')W_{2,z} \left( \frac{prm + qt}{A} \right) \frac{pr'm + qt'}{A} \ll (pq)^\epsilon (pR)^{1/4} A^{3/4}.
\]

Recall from (5.69) that \( \beta_t \ll d(t) \) and \( \gamma_r \leq 2|\alpha_r| \). Rankin-Selberg bound implies that \( \sum_r |\gamma_r| \ll (pq)^\epsilon L \). Taking absolute values and using the bound above in
(7.13), we get
\[ O_{(p,1,1)} \ll (pq)^t \frac{N(pq)T^{3/4}L^{11/2}}{p^{7/4}q^{1/4}} \ll (pq)^t \frac{N(pq)L^{25/4}}{p^{7/4}q^{1/4}}, \]
(7.15)
where \( T \leq L \) has been defined in (2.24). (It is possible to remove the \( T^{3/4} \) in the numerator, by choosing \( C = L^2/T \) in (3.10).) \( \square \)

**Lemma 7.3.** Let \( O_{(1,p,1)} \) be defined as in (7.13). Then
\[ O_{(1,p,1)} \ll (pq)^t \frac{\sqrt{p}N(pq)L^{25/4}}{q^{1/4}}. \]
(7.16)

**Proof.** Use Theorem (1.2), with \( a = r, b = r', c = qt, d = qt', M_1 = M_2 = A, \) and \( K_1 = K_2 = 1 \) to bound the \( m \)-sum in (7.13). The shift \( (rt' - r't) \) may not be multiple of \( p \). Hence, we have
\[
\sum_m \lambda_f(rm + qt)\lambda_f(r'm + qt')W_{2,z} \left( \frac{rm + qt}{A} \right) W_{2,z} \left( \frac{r'm + qt'}{A} \right) \ll (pq)^t R^{1/4} (pA)^{3/4}.
\]
If \( \bar{p} = (1, p, 1) \), then \( X_{\bar{p}} = \frac{(pq)^2}{N}, Y_{\bar{p}} = \frac{(pq)^2}{AN}, \) and \( T_{\bar{p}} = \max \left\{ \frac{A}{q}, \frac{rB}{q} \right\} \) (see (5.60) and (5.71)). The bound claimed in the Lemma is a straightforward consequence of the bound mentioned above, proceeding along the lines of Lemma (7.2). \( \square \)

**Lemma 7.4.** Let \( O_{(1,1,p)} \) be defined as in (7.13). Then
\[ O(1,1,p) \ll (pq)^t \frac{N(pq)L^{25/4}}{q^{1/4}}. \]
(7.17)

**Proof.** Use Theorem 1.2 with \( a = r, b = r', c = qt, d = qt', M_1 = M_2 = A, \) and \( K_1 = K_2 = 1 \) to bound the \( m \)-sum in (7.13). In this case the shift is divisible by \( p \). This is the reason we save an additional \( \sqrt{p} \) as compared to Lemma 7.3. \( \square \)

**Lemma 7.5.** If \( p \) is fixed. Then for any factorization \( p_1p_2p_3 = p, \)
\[ O_{\bar{p}} \ll (pq)^t \frac{N(pq)L^{23/4}(qL)^{6}}{q^{1/2}}, \]
(7.18)
where \( \theta \) is the bound towards Ramanujan conjecture for the congruence subgroup \( \Gamma_0(ab) \).

**Proof.** We use Theorem 6.5 with \( a = p_1r, b = p_1r', c = p_3qt, d = p_3qt', M_1 = M_2 = A, \) and \( K_1 = K_2 = 1 \) to bound the \( m \)-sum in (7.13). Proceeding like before, we get the claim stated in the lemma. Note that we save \( q^{1/2} \) as opposed to \( q^{1/4} \) in the previous Lemma. \( \square \)
7.1. Proof of Theorem 1.1.

Proof. We trace our steps starting from the beginning. Lemma 3.6 shows that

\[ S(N) = S_1(N) + O((pq)^{\epsilon} \frac{N}{\sqrt{L}}). \]

Using Equation (5.2) and the observation that all four choices of \( \{\pm, \pm\} \) behave the same way, we have

\[ S(N) = S_{\text{dual}}^{1}(N) + O((pq)^{\epsilon} \frac{N}{\sqrt{L}}). \]

Eliminating the boundary terms in (5.55), led us to

\[ S(N) = \sum_{p_1p_2p_3=p} S(p_1, p_2, p_3)(N) + O \left( (pq)^{\epsilon} \frac{\sqrt{Npq}}{\sqrt{L}} \right). \quad (7.19) \]

After Cauchy-Schwarz, equation (7.1) shows that

\[ S(N) \ll \sum_{p_1p_2p_3=p} D(p_1, p_2, p_3)(N)^{1/2} + O(p_1, p_2, p_3)(N)^{1/2} + O \left( (pq)^{\epsilon} \frac{\sqrt{Npq}}{\sqrt{L}} \right). \]

We use Lemma 7.1 to control the diagonal contribution. We shall use the bound from Lemma 7.3 for \( O_{p_2=p} \) when \( p \) is small. Otherwise we shall use Lemma 5.31 to get rid of the contribution of the terms with \( p_2 = p \). Thus using Lemma 5.31, 7.2, 7.3, and 7.4, we obtain:

\[ S(N)^2 \ll (pq)^{\epsilon} N(pq) \left( \frac{1}{L} + \frac{L^{25/4}}{q^{1/4}} + \min \left\{ \frac{\sqrt{pL^{25/4}}}{q^{1/4}}, \frac{L^2}{p^2} \right\} \right). \quad (7.20) \]

If \( p \leq L^{3/2} \), then we use the first bound in \( \min\{\ldots\} \) and the second bound otherwise. Making the choice \( L = q^{1/32} \), we get

\[ S(N)^2 \ll (pq)^{\epsilon} N(pq). \quad (7.21) \]

Thus

\[ S(N) \ll \epsilon, A \ (pq)^{\epsilon} \sqrt{Npq} \frac{1}{64}. \]

This proves Proposition 2.7 and therefore, Theorem 1.1. \( \square \)

7.2. Proof of Theorem 1.3.

Proof. Proceeding as in the previous proof, we get equation (7.19). Using Lemma 7.1 and 7.5 to bound \( D \) and \( O \) respectively, we get

\[ S(N)^2 \ll_{p, \epsilon} q^{\epsilon} Nq \left( \frac{1}{L} + \frac{L^{23/4+7\theta}}{q^{1/2-\theta}} \right). \]
Equating the two terms, we get
\[ L = q^{2(1-2\theta)\frac{27+28}{27+28\theta}}. \] (7.22)

Thus
\[ S(N) \ll q^{\frac{\sqrt{Nq}}{q^{\frac{1-2\theta}{1-2\theta}}}}. \]

This proves Proposition 2.8 and therefore, Theorem 1.3.

□

REFERENCES

[1] K. Aggarwal, R. Holowinsky, Y. Lin, and Q. Sun, The Burgess bound via a trivial delta method, ArXiv e-prints (2018), available at 1803.00542.
[2] Valentin Blomer, Shifted convolution sums and subconvexity bounds for automorphic L-functions, International Mathematics Research Notices 2004 (2004), no. 73, 3905–3926.
[3] David A Burgess, On Character Sums and L-Series. II, Proceedings of the London Mathematical Society 3 (1963), no. 1, 524–536.
[4] W. Duke, J. B. Friedlander, and H. Iwaniec, A quadratic divisor problem, Inventiones mathematicae 115 (1994), no. 1, 209–217.
[5] William Duke, John B Friedlander, and Henryk Iwaniec, The subconvexity problem for Artin L-functions, Inventiones mathematicae 149 (2002), no. 3, 489–577.
[6] Brooke Feigon, David Whitehouse, et al., Averages of central L-values of Hilbert modular forms with an application to subconvexity, Duke Mathematical Journal 149 (2009), no. 2, 347–410.
[7] Étienne Fouvry, Emmanuel Kowalski, and Philippe Michel, Algebraic twists of modular forms and hecke orbits, Geometric and Functional Analysis 25 (2015), no. 2, 580–657.
[8] John B. Friedlander, Bounds for L-functions, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), 1995, pp. 363–373. MR1403937
[9] Gergely Harcos and Philippe Michel, The subconvexity problem for Rankin–Selberg L-functions and equidistribution of Heegner points. II, Inventiones mathematicae 163 (2006), no. 3, 581–655.
[10] D. R. Heath-Brown, A new form of the circle method, and its application to quadratic forms, J. reine angew. Math. 481 (1996), 149–206.
[11] Roman Holowinsky and Ritabrata Munshi, Level aspect subconvexity for Rankin-Selberg L-functions, arXiv preprint arXiv:1203.1300 (2012).
[12] Roman Holowinsky and Paul D Nelson, Subconvex bounds on GL(3) via degeneration to frequency zero, arXiv preprint arXiv:1801.08593 (2018).
[13] Roman Holowinsky and Nicolas Templier, First moment of Rankin–Selberg central L-values and subconvexity in the level aspect, The Ramanujan Journal 33 (2014), no. 1, 131–155.
[14] Fei Hou and Meng Zhang, Hybrid bounds for Rankin-Selberg L-functions, Journal of Number Theory 175 (2017), 21–41.
[15] Henryk Iwaniec, The spectral growth of automorphic L-functions, J. reine angew. Math 428 (1992), 139–159.
[16] Henryk Iwaniec and Peter Sarnak, Perspectives on the analytic theory of L-functions, Visions in mathematics, 2000, pp. 705–741.
[17] Henry Kim et al., *Functoriality for the exterior square of GL(4) and the symmetric fourth of GL(2)*, Journal of the American Mathematical Society 16 (2003), no. 1, 139–183.

[18] Emmanuel Kowalski, Philippe Michel, and Jeffrey VanderKam, *Mollification of the fourth moment of automorphic L-functions and arithmetic applications*, Inventiones mathematicae 142 (2000), no. 1, 95–151.

[19] Emmanuel Kowalski, Philippe Michel, Jeffrey VanderKam, et al., *Rankin-Selberg L-functions in the level aspect*, Duke Mathematical Journal 114 (2002), no. 1, 123–191.

[20] Xiaoqing Li, *Bounds for GL(3) × GL(2) L-functions and GL(3) L-functions*, Ann. of Math. (2) 173 (2011), no. 1, 301–336. MR2753605

[21] Philippe Michel, *Familles de fonctions L de formes automorphes et applications*, J. Théor. Nombres Bordeaux 15 (2003), no. 1, 275–307.

[22] , *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Annals of mathematics (2004), 185–236.

[23] , *Analytic Number Theory and Families of Automorphic L-functions* (2006).

[24] Philippe Michel and Dinakar Ramakrishnan, *Consequences of the Gross–Zagier formulae: Stability of average L-values, subconvexity, and non-vanishing mod p*, Number theory, analysis and geometry, 2012, pp. 437–459.

[25] Ritabrata Munshi, *The circle method and bounds for L-functions-I*, Mathematische Annalen 358 (2014), no. 1-2, 389–401.

[26] , *The circle method and bounds for L-functions -IV: subconvexity for twists of GL(3) L-functions*, Annals of Mathematics 182 (2015), 617–672.

[27] , *The circle method and bounds for L-functions-III: t-aspect subconvexity for GL(3) L-functions*, Journal of the American Mathematical Society 28 (2015), no. 4, 913–938.

[28] , *A note on Burgess bound*, arXiv preprint arXiv:1710.02354 (2017).

[29] , *Subconvexity for symmetric square L-functions*, arXiv preprint arXiv:1709.05615 (2017).

[30] Paul D Nelson, *Stable averages of central values of Rankin-Selberg L-functions: some new variants*, Journal of Number Theory 133 (2013), no. 8, 2588–2615.

[31] Peter Sarnak, *Estimates for Rankin-Selberg L-functions and quantum unique ergodicity*, J. Funct. Anal. 184 (2001), no. 2, 419–453. MR1851004

[32] Hermann Weyl, *Zur Abschatzung von \( \zeta(1+ti) \)*, Math. Zeitschr. 10 (1921), 88–101.

[33] Zhilin Ye, *The second moment of Rankin-Selberg L-function and hybrid subconvexity bound*, arXiv preprint arXiv:1404.2336 (2014).