ON THE SPECTRAL LARGE SIEVE INEQUALITY FOR SYMMETRIC-SQUARES

MATTHEW P. YOUNG

ABSTRACT. We improve on the spectral large sieve inequality for symmetric-squares. We also prove a lower bound showing that the most optimistic upper bound is not true for this family.

1. INTRODUCTION

1.1. **Notation and motivation.** Suppose that $\mathcal{F}$ is a family of automorphic forms with associated L-functions $L(f, s) = \sum_n \lambda_f(n) n^{-s}$, $f \in \mathcal{F}$. We normalize $L(f, s)$ to have a functional equation under $s \to 1 - s$. We define a norm on a bilinear form associated to $\mathcal{F}$ by

$$(1.1) \qquad \mathcal{B}(\mathcal{F}, N) = \max_{\mathbf{a} \neq 0} |\mathbf{a}|^{-2} \sum_{f \in \mathcal{F}} \Big| \sum_{N \leq n \leq 2N} a_n \lambda_f(n) \Big|^2,$$

where $\mathbf{a} = (a_n)$ and $|\mathbf{a}|^2 = \sum_{N \leq n \leq 2N} |a_n|^2$. In practice, $\mathcal{F}$ will be a finite part of a larger infinite family, which we finitize by specifying the size of the conductors. It is also common to study modified versions of (1.1), which could include an arithmetical weight such as w_f^{-1}, where $w_f = \mathrm{Res}_{s=1} L(f \otimes \overline{f}, s)$.

A large sieve inequality for $\mathcal{F}$ is an upper bound on $\mathcal{B}(\mathcal{F}, N)$. A strong bound reflects orthgonality properties of the coefficients $\lambda_f(n)$ and has applications to moments of L-functions, the distribution of their zeros, etc. We also seek *lower* bounds on $\mathcal{B}(\mathcal{F}, N)$, which may be helpful for discerning the true size of $\mathcal{B}(\mathcal{F}, N)$. By general principles of bilinear forms (see [IK, Chapter 7.3]), it is well-known that $\mathcal{B}(\mathcal{F}, N) \gg (|\mathcal{F}| + N)^{1-\varepsilon}$.

Definition 1.1. *The* optimistic bound *(for $\mathcal{F}$) is the purported inequality*

$$(1.2) \qquad \mathcal{B}(\mathcal{F}, N) \ll (|\mathcal{F}| + N)^{1+\varepsilon},$$

valid for all N.

In some important cases, notably for families of Dirichlet characters, the optimistic bound is known to be true (e.g., see Theorem 2.3 and (1.18) below). For some other families, the optimistic bound does not hold. See Section 6 below for more discussion on known counterexamples.

It appears to be difficult to even conjecture the true size of $\mathcal{B}(\mathcal{F}, N)$ for general families.

The optimistic bound is true for the following GL_2 family. Let u_j run over the Hecke-Maass cusp forms on $SL_2(\mathbb{Z})$, of Laplace eigenvalue $1/4 + t_j^2$. Let $\lambda_j(n)$ denote the n-th Hecke eigenvalue of u_j, and set $w_j = \mathrm{Res}_{s=1} L(s, u_j \otimes u_j)$. Here these weights w_j satisfy $w_j = t_j^{o(1)}$,

and hence can be included or not without changing the results. We let $E(z, 1/2 + it)$ denote the Eisenstein series having Hecke eigenvalues $\tau_{it}(n) = \sum_{ab=n} (a/b)^{it}$. The spectral large sieve inequality (see [I1, J]) states

$$(1.3) \qquad \max_{\mathbf{a} \neq 0} |\mathbf{a}|^{-2} \sum_{T \leq t_j \leq T+\Delta} w_j^{-1} \left| \sum_{N \leq n \leq 2N} a_n \lambda_j(n) \right|^2 \ll (\Delta T + N)(NT)^\varepsilon,$$

for $1 \leq \Delta \leq T$. One may additionally incorporate a contribution from the Eisenstein series, without altering the bound on the right hand side of (1.3).

The present state of knowledge of large sieve inequalities for GL_3 families is much weaker than for GL_1 and GL_2. Some references in this direction include [DK, B, Y2, BB, Y3]. The symmetric-squares of $SL_2(\mathbb{Z})$ Maass forms are known to correspond to self-dual $SL_3(\mathbb{Z})$ Maass forms [GJ, S], and form a particularly interesting GL_3 family. The relation between the Dirichlet series coefficients of $L(u_j, s)$ and $L(\mathrm{sym}^2 u_j, s)$ is encapsulated in the identity

$$(1.4) \qquad \lambda_{\mathrm{sym}^2 u_j}(n) = \sum_{d^2 m = n} \lambda_j(m^2)$$

The main objective of this paper is a bound on $\mathcal{M}_0(\Delta, T, N)$ defined by

$$(1.5) \qquad \mathcal{M}_0(\Delta, T, N) = \max_{\mathbf{a} \neq 0} |\mathbf{a}|^{-2} \sum_{T \leq t_j \leq T+\Delta} w_j^{-1} \left| \sum_{N \leq n \leq 2N} a_n \lambda_j(n^2) \right|^2.$$

We also define the analogous contribution from the Eisenstein series, namely

$$(1.6) \qquad \mathcal{M}_\infty(\Delta, T, N) = \max_{\mathbf{a} \neq 0} |\mathbf{a}|^{-2} \int_{T \leq t \leq T+\Delta} w_t^{-1} \left| \sum_{N \leq n \leq 2N} a_n \tau_{it}(n^2) \right|^2,$$

where $w_t = |\zeta(1 + 2it)|^2$. Let $\mathcal{M}(\Delta, T, N) = \mathcal{M}_0(\Delta, T, N) + \mathcal{M}_\infty(\Delta, T, N)$. It follows from (1.3) (after inclusion of the Eisenstein series) that

$$(1.7) \qquad \mathcal{M}(\Delta, T, N) \ll (\Delta T + N^2)(NT)^\varepsilon,$$

which matches the optimistic bound only for small values of N. We regard (1.7) as a benchmark which does not use the arithmetical structure of the inner sum, namely that the Hecke eigenvalues are sampled at squares.

In a slightly different context, Duke and Kowalski [DK] showed a bound that is strongest for large values of N. Their proof proceeds by duality and uses analytic properties of the degree 9 Rankin-Selberg L-function $L(\mathrm{sym}^2 u_j \otimes \mathrm{sym}^2 u_k, s)$. More precisely, they use a contour-shifting argument and apply the convexity bound for these Rankin-Selberg L-functions. Strictly speaking, they study the level aspect of the problem and not the archimedean aspect which is the focus here. Nevertheless, their method leads to

$$(1.8) \qquad \mathcal{M}_0(\Delta, T, N) \ll (N + \Delta^{3/2} T^{5/2} N^{1/2})(NT)^\varepsilon.$$

This sets another useful frame of reference to complement (1.7). Junehyuk Jung and Min Lee (personal communication) have recently improved on (1.8), using the dual approach.

We also mention that [IK, Problem 7.29] asks for an improvement on (1.8) (though technically they state the problem for the level aspect).

1.2. **Statements of results.** The main result of this paper is an improvement on (1.7) and (1.8) in many ranges.

Theorem 1.2. *Let notation be as above. Then*

$$(1.9) \qquad \mathcal{M}(\Delta, T, N) \ll (NT)^\varepsilon \begin{cases} \Delta T + T^{1/2} N, & N \leq T, \\ \Delta N + N^{3/2}, & T \leq N \leq T^2, \\ \frac{N^2}{T}, & T^2 \leq N. \end{cases}$$

One may check that Theorem 1.2 improves on (or agrees with) (1.7) for all N and T. Likewise, one may check that Theorem 1.2 improves on (1.8) for $N \ll \Delta T^{7/3}$. Another way to gauge Theorem 1.2 is to ask in what ranges of parameters does it give the optimistic bound; this occurs for $N \ll \min(T, \Delta T^{1/2})$, while (1.7) only matches the optimistic bound for $N \ll \Delta^{1/2} T^{1/2}$. On the other hand, Theorem 1.5 below gives a *lower* bound on $\mathcal{M}(\Delta, T, N)$ of size ΔN, and Theorem 1.2 appears even more favorable in this light.

As an aside, (1.9) is concisely expressed by

$$(1.10) \qquad \mathcal{M}(\Delta, T, N) \ll (NT)^\varepsilon \left(\Delta(T + N) + N(T + N)^{1/2} + \frac{N^2}{T} \right).$$

Theorem 1.2 treats the Hecke eigenvalues sampled at squares, which is not the same as the Dirichlet series coefficients of the symmetric-square L-function (recall (1.4)). We next discuss the connections between these objects. Define

$$(1.11) \qquad \mathcal{M}_0^{(2)}(\Delta, T, N) = \max_{\mathbf{a} \neq 0} |\mathbf{a}|^{-2} \sum_{T \leq t_j \leq T + \Delta} w_j^{-1} \left| \sum_{N \leq n \leq 2N} a_n \lambda_{\mathrm{sym}^2 u_j}(n) \right|^2.$$

Suppose for convenience that a_n is supported on $[N, 2N]$. Then note

$$(1.12) \qquad \sum_n a_n \lambda_{\mathrm{sym}^2 u_j}(n) = \sum_d b_d \lambda_j(d^2), \qquad \text{where} \qquad b_d = \sum_k a_{dk^2}.$$

Note that if a_n is restricted to square-free integers, then $\lambda_{\mathrm{sym}^2 u_j}(n) = \lambda_j(n^2)$, so a bound on $\mathcal{M}_0(\Delta, T, N)$ may be applied under this assumption. Without this square-free restriction, we may deduce the following (see Section 5.2 for details):

Corollary 1.3. *For any* $\mathbf{a}$*, we have*

$$(1.13) \qquad \sum_{T \leq t_j \leq T + \Delta} w_j^{-1} \left| \sum_{N \leq n \leq 2N} a_n \lambda_{\mathrm{sym}^2 u_j}(n) \right|^2 \ll \mathcal{D} + \left(\Delta N + N(T + N)^{1/2} + \frac{N^2}{T} \right)(NT)^\varepsilon |\mathbf{a}|^2,$$

where

$$(1.14) \qquad \mathcal{D} = \Delta T \sum_n \left| \sum_k a_{nk^2} \right|^2 \ll \Delta T N^{1/2} |\mathbf{a}|^2.$$

In particular,

$$(1.15) \qquad \mathcal{M}_0^{(2)}(\Delta, T, N) \ll (NT)^\varepsilon \left(\Delta T N^{1/2} + \Delta N + N(T + N)^{1/2} + \frac{N^2}{T} \right).$$

Similarly to (1.9), the right hand side of (1.13) can alternatively be stated as

$$(1.16) \qquad \mathcal{D} + (NT)^\varepsilon \times \begin{cases} \Delta N + T^{1/2} N, & N \leq T, \\ \Delta N + N^{3/2}, & T \leq N \leq T^2, \\ \frac{N^2}{T}, & T^2 \leq N. \end{cases}$$

Thorner and Zaman [TZ] have proved a complementary bound on $\mathcal{M}_0^{(2)}$ by the dual approach. By following their method, they implicitly show

$$(1.17) \qquad \mathcal{M}_0^{(2)}(\Delta, T, N) \ll (NT)^\varepsilon (N + \Delta^2 T^4).$$

The shape of $\mathcal{D}$ in (1.13) is reminiscent of the family of quadratic Dirichlet characters, as we now elaborate. Heath-Brown [H-B] showed

$$(1.18) \qquad \sideset{}{^*}\sum_{d \leq X} \Big| \sideset{}{^*}\sum_{n \leq N} a_n \chi_d(n) \Big|^2 \ll (X + N)^{1+\varepsilon} |\mathbf{a}|^2,$$

where the sums restrict d, n to odd, square-free integers. The square-free restriction is vital: if d is allowed to run over squares, then one can produce a large term of size $\sqrt{X} N |\mathbf{a}|^2$ by taking $a_n = 1$ for all n. A term of this size would contradict (1.18). By the duality principle, one may produce a large term of size $\sqrt{N} X |\mathbf{a}|^2$ if n is not restricted to square-free integers (see also discussion surrounding (1.23) below). As a heuristic, one might expect

$$\frac{1}{X^*} \sideset{}{^*}\sum_{d \leq X} \chi_d(mn) \approx \delta_{mn=\square}, \qquad \text{where} \qquad X^* = \sideset{}{^*}\sum_{d \leq X} 1,$$

and consequently,

$$\sideset{}{^*}\sum_{d \leq X} \Big| \sum_{n \leq N} a_n \chi_d(n) \Big|^2 \approx X^* \sum_n \Big| \sum_k a_{k^2 n} \Big|^2.$$

The point is that the shape of this "diagonal" term matches that in Corollary 1.3. See Proposition 1.6 below for more discussion.

As a simple corollary, we record

Corollary 1.4. *For $T^{1/2} \leq \Delta \leq T$ we have*

$$(1.19) \qquad \sum_{T \leq t_j \leq T + \Delta} |L(\mathrm{sym}^2 u_j, 1/2)|^2 \ll \Delta T^{1+\varepsilon}.$$

Corollary 1.4 is not close to the state of the art, as Lam [La] previously obtained this quality of bound with $\Delta = T^{1/3}$, which was recently improved further to $\Delta = T^{1/5}$ in [KY]. On the other hand, (1.7) gives a Lindelöf-on-average bound only for $\Delta = T$. In addition, it is likely that some of the intermediate steps used in the proof of Theorem 1.2 could be useful for some applications. For instance, (5.1) or (5.3), which are formulas valid for arbitrary coefficients a_n, could be followed up with tools that use special properties of the coefficients, such as a Poisson/Voronoi summation formula. This strategy was used in [Y1].

We briefly sketch a proof of Corollary 1.4. The conductor of $L(\mathrm{sym}^2 u_j, 1/2)$, for $T \leq t_j \leq T + \Delta$, is approximately T^2. Hence, by an approximate functional equation, we reduce the second moment to an expression closely resembling

$$\sum_{T \leq t_j \leq T + \Delta} \Big| \sum_{n \ll N} \frac{\lambda_{\mathrm{sym}^2 u_j}(n)}{\sqrt{n}} \Big|^2, \qquad \text{where} \quad N = T^{1+\varepsilon}.$$

Since $a_n \ll n^{-\frac{1}{2}}$, we have $\mathcal{D} \ll \Delta T^{1+\varepsilon}$, and so (1.13) leads quickly to (1.19).

1.3. **Lower bounds.** It is unclear what to conjecture for the true size of $\mathcal{M}(\Delta, T, N)$. In other contexts, notably the family of cusp forms on $\Gamma_1(q)$ studied by Iwaniec and Li [IL], the optimistic bound is not true. Blomer and Buttcane [BB] showed that for spectral families on $SL_n(\mathbb{Z})$, with $n \geq 3$, the Eisenstein series component contributes a large term, implying that the optimistic bound is not true for these families. The recent work [Y3] showed that, at least for $n = 3$, the Blomer-Buttcane bound can be improved by restricting to the cusp forms (that is, omitting the Eisenstein series). This raises the question on what are the sizes of $\mathcal{M}_0(\Delta, T, N)$ and $\mathcal{M}_\infty(\Delta, T, N)$ separately. Towards this end, we have

Theorem 1.5. *If $\Delta \gg (NT)^\delta$ for some $\delta > 0$, then*

$$(1.20) \qquad \mathcal{M}_\infty(\Delta, T, N) \gg \Delta N (NT)^{-\varepsilon}.$$

Since the method of proof of Theorem 1.2 treats $\mathcal{M}_0 + \mathcal{M}_\infty$, Theorem 1.5 shows that we cannot remove the term ΔN in (1.10) (also seen heuristically in (1.31) below). However, since this lower bound comes from $\mathcal{M}_\infty$, this has no direct implication about $\mathcal{M}_0$ itself. See Section 1.5 for further discussions. A short proof of Theorem 1.5 appears in Section 6.

Next we discuss the true size of $\mathcal{M}_0^{(2)}$. On first inspection, (1.15) appears to be defective in that the term $\Delta T N^{1/2}$ is much larger than the size of the family. However, we have the following lower bound showing this term cannot be improved:

Proposition 1.6. *If $N \ll \Delta T^{1-\delta}$ for some $\delta > 0$, then*

$$(1.21) \qquad \mathcal{M}_0^{(2)}(\Delta, T, N) \gg \Delta T N^{1/2} (NT)^{-\varepsilon}.$$

Note that this lower bound does *not* come from Eisenstein series, so this is a different phenomenon than that observed by Blomer and Buttcane [BB]. Nevertheless, both Theorems 1.5 and Proposition 1.6 have similar proofs, as we next discuss.

Let $\mathcal{N} \subseteq [N, 2N] \cap \mathbb{Z}$. We say that the set $\mathcal{N}$ is *biased* for the family $\mathcal{F}$ if

$$(1.22) \qquad \sum_{f \in \mathcal{F}} \left| \sum_{n \in \mathcal{N}} \lambda_f(n) \right|^2 \gg (|\mathcal{F}| \cdot |\mathcal{N}|^2)^{1-\varepsilon}.$$

By taking **a** to be the characteristic function of $\mathcal{N}$, we see that if $\mathcal{F}$ has a biased set $\mathcal{N}$, then

$$(1.23) \qquad \mathcal{B}(\mathcal{F}, N) \gg (|\mathcal{F}| \cdot |\mathcal{N}|)^{1-\varepsilon},$$

which contradicts the optimistic bound for $\mathcal{F}$ provided $|\mathcal{N}| \gg (|\mathcal{F}| \cdot N)^\delta$ for some $\delta > 0$. As a simple example of a biased set, let $\mathcal{F}$ consist of the quadratic Dirichlet characters of odd conductor d with $d \asymp X$, and let $\mathcal{N}$ consist of the squares in $[N, 2N]$, so $|\mathcal{N}| \asymp \sqrt{N}$. It is easy to see that $\mathcal{N}$ is biased for this family. Of course, this does not contradict (1.18) since there n is restricted to square-free integers.

Next we explain the source of bias in the two families in Theorems 1.5 and Proposition 1.6. For the family in Theorem 1.5, let $\mathcal{N}$ consist of primes in $[N, 2N]$, so $|\mathcal{N}| \asymp \frac{N}{\log N}$. The bias arises from $\tau_{it}(p^2) = 1 + p^{2it} + p^{-2it}$, which is approximately 1 on average over $t \in [T, T + \Delta]$. Similarly, for the symmetric-square family in Proposition 1.6, we let $\mathcal{N}$ consist of p^2 with p prime and $p^2 \in [N, 2N]$, so $|\mathcal{N}| \asymp \frac{N^{1/2}}{\log N}$. The bias here may be seen from $\lambda_{\mathrm{sym}^2 u_j}(p^2) = 1 + \lambda_j(p^2)$, which again is approximately 1 on average over $t_j \in [T, T + \Delta]$. More details of the proof are presented in Section 6.

This is some evidence that the definition (1.1) might require some modification, depending on some arithmetical features of the family $\mathcal{F}$. More examples of families with biased sets would be welcome.

We also mention that Dunn and Radziwiłł [DR] have shown that the optimstic bound does not hold for the family of cubic residue symbols.

1.4. **Sketch of proof of Theorem 1.2.** The Kuznetsov formula leads to a diagonal term of size ΔT as well as a sum of Kloosterman sums roughly of the shape

$$(1.24) \qquad \frac{\Delta T}{N\sqrt{C}} \sum_{c \sim C} \sum_{m,n} a_m \overline{a_n} S(m^2, n^2; c) e_c(2mn) e^{i\frac{T^2 c}{mn}},$$

where C runs over dyadic segments with $1 \ll C \ll \frac{N^2}{\Delta T}$. Here $e_c(2mn)$ comes from the main part of the phase of the Bessel-type transform from the Kuznetsov formula, and the factor $\exp(i\frac{T^2 c}{mn})$ is a lower-order term in the phase. Since $e_c(2mn)$ is periodic in m, n modulo c, it can be joined with the Kloosterman sum $S(m^2, n^2; c)$. If $(mn, c) = 1$ then with $w = mn$ we can expand $F(w) = S(w^2, 1; c) e_c(2w)$ into Dirichlet characters, via

$$(1.25) \qquad F(w) = \sum_{\chi \,(\mathrm{mod}\ c)} \widehat{F}(\chi) \chi(w), \quad \text{where} \quad \widehat{F}(\chi) = \frac{1}{\varphi(c)} \sum_{u \,(\mathrm{mod}\ c)} \overline{\chi}(u) S(u^2, 1; c) e_c(2u).$$

We similarly have an archimedean separation of variables by the Mellin transform, via

$$(1.26) \qquad e^{i\frac{T^2 c}{mn}} \approx \frac{1}{\sqrt{P}} \int_{t \asymp P} f(t) \Big(\frac{T^2 c}{mn}\Big)^{-it} dt, \qquad P = 1 + \frac{T^2 C}{N^2},$$

where f is a smooth function satisfying $f(t) \ll 1$. Inserting these into (1.24), we obtain

$$(1.27) \qquad \frac{\Delta T}{N\sqrt{CP}} \int_{t \asymp P} f(t) T^{-2it} \sum_{c \sim C} c^{-it} \sum_{\chi \,(\mathrm{mod}\ c)} \widehat{F}(\chi) \sum_{m,n} a_m \overline{a_n} \chi(mn) (mn)^{it} dt.$$

As a heuristic, consider the contribution to (1.27) from χ primitive modulo c. For such χ, $|\widehat{F}(\chi)| \ll c^\varepsilon$. Then the classical hybrid large sieve (see Theorem 2.3) gives a bound

$$(1.28) \qquad \frac{\Delta T}{N\sqrt{CP}} (C^2 P + N) |\mathbf{a}|^2.$$

Note the latter term is $\frac{\Delta T}{\sqrt{CP}} \ll \Delta T$, which is already bounded by the diagonal term. The former term is bounded by $\frac{\Delta T}{N} \frac{T}{N} \frac{N^4}{\Delta^2 T^2} = \frac{N^2}{\Delta}$.

On the opposite extreme from χ primitive is χ trivial. It turns out (see Lemma 4.3) that the most significant contribution from trivial χ comes from c ranging over squares, in which case $\widehat{F}(\chi_0) = c^{1/2}$. These terms contribute to (1.27) an expression bounded by

$$(1.29) \qquad \frac{\Delta T}{N\sqrt{CP}} \int_{t \asymp P} \sum_{\substack{c \sim C \\ c = \square}} \sqrt{c} \Big| \sum_m a_m m^{it} \Big|^2 dt.$$

The hybrid large sieve inequality in this case reduces to the mean value theorem for Dirichlet polynomials (see [IK, Theorem 9.1]), and produces a bound of the form

$$(1.30) \qquad \frac{\Delta T}{N\sqrt{CP}} (CP + CN) |\mathbf{a}|^2.$$

Comparing this with (1.28), we see that the term CP in (1.30) is dwarfed by the term $C^2 P$ in (1.28). However, the term CN is now larger than the term N. This term with CN

contributes

$$(1.31) \qquad \frac{\Delta T}{N\sqrt{CP}}CN = \Delta N.$$

Now taking into account the bounds from the diagonal, the primitive characters, and the trivial characters, we obtain a bound of size

$$(1.32) \qquad \Delta T + \Delta N + \frac{N^2}{\Delta}.$$

Note that (1.32) has the feature that one part is increasing in Δ, and another part is decreasing in Δ. This type of structure is commonly seen in harmonic analysis in association with the duality principle. It should thus not be surprising that (1.32) may be improvable by enlarging Δ. When $N \leq T$ then we replace Δ by $\Delta + \frac{N}{\sqrt{T}}$, which leads to the first bound in Theorem 1.2. When $T \leq N \leq T^2$ then we replace Δ by $\Delta + \sqrt{N}$, leading to the second bound in Theorem 1.2. Finally, when $N \geq T^2$ we replace Δ by T, giving the third bound in Theorem 1.2.

This discussion indicates that the rigorous bound from Theorem 1.2 matches the bounds from considering the primitive characters and the trivial characters separately, which is encouraging. Most of the work in this paper boils down to treating all the cases in a uniform manner, including dealing with m, n that are not coprime to c, and characters χ that are neither primitive nor trivial.

1.5. **Remarks on possible improvements.** Since the term of size ΔN in Theorem 1.2 matches in rough order of magnitude the lower bound on $\mathcal{M}_\infty$, it is natural to ask if the term ΔN is required in bounding $\mathcal{M}_0$, or if an improvement is possible. A possible method to improve on the bound on $\mathcal{M}_0$ would be to continue to use the Kuznetsov formula, but to cancel (perhaps only partially) the contribution of the Eisenstein series from some part of the sum of Kloosterman sums. Luo [Lu] achieved this type of cancellation in a different spectral large sieve problem.

As a possible clue in this direction, consider the contribution from $c = p^2$ with p prime in (1.24). The Kloosterman sum of modulus p^2 can be calculated in closed form (see [I2, Section 4.3]), showing $S(m^2, n^2; p^2) = p\sum_{\pm} e_p(\pm 2mn)$ for $(p, mn) = 1$. The term with $e_p(-2mn)$ cancels the phase from $e_p(2mn)$ coming from the Bessel transform, and leads to an expression resembling

$$(1.33) \qquad \frac{\Delta T}{N}\sum_p \sum_{m,n} a_m \overline{a_n} e^{i\frac{T^2 p^2}{mn}}.$$

When $p \ll \frac{N}{T}$, the exponential is not oscillatory, so one sees a sum roughly of the form $\Delta|\sum_n a_n|^2$. This now closely resembles the term (6.2) responsible for the lower bound in Theorem 1.5.

As a possible argument *against* an improvement in this problem, one can consider the holomorphic weight k cusp form analog of $\mathcal{M}(\Delta, T, N)$. With only minor modifications of the proof, one could derive the analog of Theorem 1.2 for this family. A simplification here is that there is no Eisenstein contribution, so there would be no need to cancel a contribution from the Eisenstein spectrum. This *might* be a hint that the term of size ΔN is not removable for the cusp forms. At the least, it indicates that the Maass form and holomorphic form

analogs would have some differences. Barring any further evidence, the author makes no conjecture either for or against a possible improvement on the term ΔN.

2. Preliminaries

In this section we collect some tools needed for the proof of Theorem 1.2.

2.1. Around the Kuznetsov formula.

Proposition 2.1 (Kuznetsov formula). *Suppose that $h(t)$ is even, holomorphic in $|Im(t)| \leq \frac{1}{2} + \delta$, and satisfying $h(t) \ll (1 + |t|)^{-2-\delta}$, for some $\delta > 0$. Then for certain weights ω_j, ω_t proportional to w_j and w_t, respectively, we have*

$$(2.1) \quad \sum_j \omega_j^{-1} \lambda_j(m) \lambda_j(n) h(t_j) + \int_{-\infty}^{\infty} \omega_t^{-1} \tau_{it}(m) \tau_{it}(n) h(t) dt$$

$$= \delta_{m=n} \frac{1}{\pi} \int_{-\infty}^{\infty} h(r) r \tanh(\pi r) dr + \sum_{c=1}^{\infty} c^{-1} S(m,n;c) H_\infty\left(\frac{4\pi\sqrt{mn}}{c}\right),$$

where

$$(2.2) \qquad\qquad H_\infty(x) = 2i \int_{-\infty}^{\infty} \frac{J_{2ir}(x)}{\cosh(\pi r)} r h(r) dr.$$

Lemma 2.2 (Properties of H_∞). *Let $T^\delta \leq \Delta \leq T^{1-\delta}$ for some $\delta > 0$, and let*

$$h(t) = \frac{t^2 + \frac{1}{4}}{T^2}\left[\exp(-(t-T)^2/\Delta^2) + \exp(-(t+T)^2/\Delta^2)\right].$$

Then $H_\infty(x)$ is very small unless

$$(2.3) \qquad\qquad\qquad\qquad x \gg \Delta T^{1-\varepsilon}.$$

Moreover, for x satisfying (2.3), $H_\infty(x)$ has an asymptotic formula of the form

$$(2.4) \qquad\qquad H_\infty(x) = \sum_{\pm} \frac{\Delta T}{\sqrt{x}} e^{\pm ix} e^{i\phi(x)} I(x) + O(T^{-A}),$$

where $I = I_{\pm,A}$ is a smooth function satisfying $I^{(j)}(x) \ll_{j,A} x^{-j}$ for $x \gg \Delta T^{1-\varepsilon}$, and $\phi = \phi_\pm$ is a smooth function with an asymptotic expansion of the form $\phi(x) = c_1 \frac{T^2}{x} + c_2 \frac{T^4}{x^3} + \dots$.

This lemma can be extracted from the work of Jutila and Motohashi [JM, pp.75–76].

2.2. The GL_1 large sieve.

Theorem 2.3 (Gallagher [G]). *Let $Q, T \geq 1$. Then for any vector $\mathbf{a} = (a_n)$, we have*

$$(2.5) \qquad \int_0^T \sum_{q \leq Q} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{n \leq N} a_n \chi(n) n^{it}\right|^2 dt \ll (Q^2 T + N)|\mathbf{a}|^2.$$

We also need the following variant.

Corollary 2.4. *Let $Q, T, d, g \geq 1$, with g square-free. Define d' to be the smallest integer so that $d|(d')^2$ (so if $d = \prod_{p|d} p^{d_p}$, then $d' = \prod_p p^{\lceil d_p/2 \rceil}$). Then for any vector $\mathbf{a} = (a_n)$, we have*

$$(2.6) \qquad \int_0^T \sum_{q \leq Q} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{\substack{m,n \leq N \\ d|(m^2,n^2) \\ g|mn/d}} a_m \overline{a_n} \chi(mn)(mn)^{it} \Big| dt \ll N^\varepsilon \Big(Q^2 T + \frac{N}{d'} \Big) \sum_n |a_{nd'}|^2.$$

Proof. First set $h = (m, n)$, and change variables $m \to hm$ and $n \to hn$. After this change, the condition $d|(m^2, n^2)$ becomes $d|h^2$. Likewise, the condition $g|\frac{mn}{d}$ becomes $mn \equiv 0$ $(\mathrm{mod}\ \frac{g}{(g,h^2/d)})$. Since $(m, n) = 1$, we can parameterize the solutions to this latter congruence by writing $g_1 g_2 = \frac{g}{(g,h^2/d)}$ and imposing $m \equiv 0 \pmod{g_1}$ and $n \equiv 0 \pmod{g_2}$. Then we have

$$(2.7) \qquad \Big| \sum_{\substack{m,n \leq N \\ d|(m^2,n^2) \\ g|mn/d}} a_m \overline{a_n} \chi(mn)(mn)^{it} \Big| \leq \sum_{d|h^2} \sum_{g_1 g_2 = \frac{g}{(g,h^2/d)}} \Big| \sum_{(mg_1,ng_2)=1} a_{mg_1 h} \overline{a_{ng_2 h}} \chi(mn)(mn)^{it} \Big|.$$

We then apply Möbius inversion to detect the condition $(m, n) = 1$ and thereby separate the variables. After this step, we then apply Theorem 2.3. This gives that the left hand side of (2.6) is bounded by

$$(2.8) \qquad \sum_{d|h^2} \sum_{g_1 g_2 = \frac{g}{(g,h^2/d)}} \sum_\ell |\mu(\ell)| \Big(Q^2 T + \frac{N}{hg_1 \ell} \Big) \sum_n |a_{nhg_1 \ell}|^2.$$

The condition $h^2 \equiv 0 \pmod{d}$ is equivalent to $h \equiv 0 \pmod{d'}$. Simplifying the above expression using a divisor function bound leads quickly to (2.6). $\qquad\square$

3. STRUCTURAL STEPS

3.1. **Reduction.** We now embark on the proof of Theorem 1.2. Our first stage parallels the Fourier/Mellin decomposition indicated in Section 1.4. We insert the weight function $h(t)$ from Lemma 2.2, extend the spectral sum/integral to all t_j, t, open the square, and apply the Kuznetsov formula. The diagonal term is of size ΔT. We then obtain $\mathcal{M}(\Delta, T, N) \ll \Delta T + \max_{\mathbf{a} \neq 0} |\mathbf{a}|^{-2} |\mathcal{K}(\Delta, T, N)|$, where

$$(3.1) \qquad \mathcal{K}(\Delta, T, N) = \sum_{m,n} a_m \overline{a_n} \sum_{c=1}^\infty \frac{S(m^2, n^2; c)}{c} H_\infty \Big(\frac{4\pi mn}{c} \Big).$$

According to Lemma 2.2, we write $H_\infty(x) = e^{ix} H_+(x) + e^{-ix} H_-(x) + O(T^{-A})$, and correspondingly write $\mathcal{K} = \sum_\pm \mathcal{K}_\pm + O(T^{-A}|\mathbf{a}|^2)$. We have

$$(3.2) \qquad \mathcal{K}_\pm(\Delta, T, N) = \sum_\pm \sum_{m,n} a_m \overline{a_n} D_\pm(m, n),$$

where

$$(3.3) \qquad D_\pm(m, n) = \sum_{c=1}^\infty \frac{S(m^2, n^2; c) e_c(\pm 2mn)}{c} H_\pm \Big(\frac{4\pi mn}{c} \Big).$$

Both terms $\mathcal{K}_\pm$ may be bounded in the same way, as they are essentially complex conjugates of each other. We then proceed with developing D_+, and drop the subscript $+$ from the notation.

Lemma 3.1. *We have*

$$(3.4) \qquad D(m,n) = \sum_{d|(m^2,n^2)} \sum_{g|\frac{mn}{d}} \frac{\mu(g)}{g} \sum_{(c,mn/d)=1} \frac{S((\frac{mn}{dg})^2, 1; c)e_c(\frac{2mn}{dg})}{c} H\Big(\frac{4\pi mn}{dgc}\Big).$$

Proof. We begin with the Selberg identity [I2, (4.10)]

$$S(m^2, n^2; c) = \sum_{d|(m^2,n^2,c)} dS(m^2n^2/d^2, 1; c/d),$$

which quickly leads to

$$(3.5) \qquad D(m,n) = \sum_{d|(m^2,n^2)} \sum_{c=1}^{\infty} \frac{S((mn/d)^2, 1; c)e_c(2mn/d)}{c} H\Big(\frac{4\pi mn/d}{c}\Big).$$

As shorthand, let $w = mn/d$, and parameterize the sum over c by the value of $g = \gcd(w, c)$. Then

$$(3.6) \quad \sum_{c=1}^{\infty} \frac{S(w^2, 1; c)e_c(2w)}{c} H\Big(\frac{4\pi w}{c}\Big) = \sum_{g|w} g^{-1} \sum_{(c,w/g)=1} \frac{S(w^2, 1; gc)e_c(2w/g)}{c} H\Big(\frac{4\pi w/g}{c}\Big).$$

To simplify this, we claim that if $(c, g) \neq 1$ then $S(w^2, 1; gc) = 0$. In turn this follows from the fact that $S(p^k, 1; p^c) = 0$ if $c \geq 2$ and $k \geq 1$, which can be proved elementarily. With the condition $(g, c) = 1$ in effect, we then have $S(w^2, 1; gc) = S(w^2\bar{c}^2, 1; g)S(w^2\bar{g}^2, 1; c) = \mu(g)S((w/g)^2, 1; c)$. At this point, direct substitutions complete the proof. $\square$

3.2. **Finite Fourier analysis.**

Our next stage of development is multiplicative Fourier analysis, mirroring (1.25) from the sketch. As in the sketch, define $F(w) = S(w^2, 1; c)e_c(2w)$, where $(w, c) = 1$. Then (1.25) holds, and with a change of variables we have

$$(3.7) \qquad \widehat{F}(\chi) = \frac{1}{\varphi(c)} \sum_{u,t \,(\mathrm{mod}\, c)} \overline{\chi}(ut)e_c(t(u+1)^2).$$

Applying this to (3.4), with $w = \frac{mn}{dg}$ we obtain

$$(3.8) \qquad D(m,n) = \sum_{d|(m^2,n^2)} \sum_{g|\frac{mn}{d}} \frac{\mu(g)}{g} \sum_{c=1}^{\infty} \sum_{\chi \,(\mathrm{mod}\, c)} \frac{\widehat{F}(\chi)\chi(mn\overline{dg})}{c} H\Big(\frac{4\pi mn}{dgc}\Big).$$

Section 4 is devoted to understanding $\widehat{F}(\chi)$ in greater detail.

3.3. **Archimedean Fourier/Mellin analysis.**

The representation (3.8) neatly separates the variables m, n with the multiplicative character χ. However, they remain joined inside $H(x)$, so we will use Mellin inversion as the analogous tool at the archimedean place. Suppose that $1 = \sum_X W(x/X)$ is a smooth dyadic partition of unity, where X runs over numbers of the form $2^{j/2}$ with $j \in \mathbb{Z}$, and W has support on $[1, 2]$. Define

$$(3.9) \qquad G(x) = G_X(x) = W(x/X)H(4\pi x), \quad \text{and} \quad \widetilde{G}(-s) = \int_0^{\infty} G(x)x^{-s}\frac{dx}{x}.$$

Lemma 3.2. *Suppose $X \gg \Delta T^{1-\varepsilon}$, and let*

$$(3.10) \qquad P = 1 + \frac{T^2}{X}.$$

Then $\widetilde{G}(-s)$ is analytic on $\mathbb{C}$ and satisfies

$$(3.11) \qquad |\widetilde{G}(-\sigma - it)| \ll_{\sigma, A, \varepsilon} \Delta T \frac{X^{-\sigma}}{\sqrt{XP}} \left(1 + \frac{|t|}{PT^\varepsilon}\right)^{-A}.$$

Proof. Since G is smooth and compactly supported on the positive reals, its Mellin transform is entire. By Lemma 2.2, we have

$$(3.12) \qquad \widetilde{G}(-s) = \Delta T \int_0^\infty x^{-1/2-s} W(x/X) I(x) e^{ic_1 \frac{T^2}{x} + \cdots} \frac{dx}{x}.$$

If $X \gg T^{2-\varepsilon}$, i.e., $P \ll T^\varepsilon$, then the bound follows by standard integration by parts. For $X \ll T^{2-\varepsilon}$, i.e., $P \gg T^\varepsilon$, the bound results from robust stationary phase analysis (see [KPY], for instance). $\qquad\square$

Applying (3.9) to (3.8), and interchanging the orders of summation and integration (valid for $\sigma > 2$, using only a trivial bound $|\widehat{F}(\chi)| \le c$), we deduce the following, which is similar in spirit to [J, (3.1)].

Lemma 3.3. *For $\sigma > 2$, we have*
$$(3.13)$$
$$D(m,n) = \sum_{X \text{ dyadic}} \sum_{d|(m^2,n^2)} \sum_{g|\frac{mn}{d}} \frac{\mu(g)}{g} \int_{(\sigma)} \widetilde{G}(-s) \left(\frac{mn}{dg}\right)^s \sum_{c=1}^\infty \sum_{\chi \pmod c} \frac{\widehat{F}(\chi)\chi(mn\overline{dg})}{c^{1+s}} \frac{ds}{2\pi i}.$$

4. PROPERTIES OF $\widehat{F}$

Lemma 4.1. *Suppose $c = c_1 c_2$ with $(c_1, c_2) = 1$, and $\chi = \chi_1 \chi_2$ with χ_j modulo c_j for $j = 1, 2$. Then $\widehat{F}(\chi)$ satisfies the twisted-multiplicativity relation*

$$(4.1) \qquad \widehat{F}(\chi) = \overline{\chi_1}(c_2)\overline{\chi_2}(c_1)\widehat{F}(\chi_1)\widehat{F}(\chi_2),$$

and so $|\widehat{F}(\chi)|$ is multiplicative.

Proof. Standard with Chinese remainder theorem. $\qquad\square$

In light of Lemma 4.1, it suffices to understand $\widehat{F}(\chi)$ when $c = p^k$.

Lemma 4.2 (Primitive case). *Suppose that χ has conductor p^k, with $k \ge 1$. If $p = 2$ then $\widehat{F}(\chi) = 0$. If p is odd and χ is not the Legendre symbol, then*

$$(4.2) \qquad |\widehat{F}(\chi)| = \frac{p}{p-1}.$$

If χ is the Legendre symbol, then

$$(4.3) \qquad |\widehat{F}(\chi)| = \frac{\sqrt{p}}{p-1}.$$

Proof. We use the formula (3.7) and evaluate the t-sum in terms of a Gauss sum, giving

$$(4.4) \qquad \widehat{F}(\chi) = \frac{\tau(\overline{\chi})}{\varphi(p^k)} \sum_{u \, (\mathrm{mod} \, p^k)} \overline{\chi}(u) \chi^2(u+1) = \chi(-1) \frac{\tau(\overline{\chi}) J(\overline{\chi}, \chi^2)}{\varphi(p^k)},$$

where $J(\chi, \psi)$ is the Jacobi sum. Note that if $p = 2$ then $\widehat{F}(\chi) = 0$ since $2 | u(u+1)$ for all u. For p odd then either χ^2 has conductor p^k, or χ is the Legendre symbol (and so $k = 1$). In case χ^2 has conductor p^k, then according to [IK, (3.18)], then $J(\overline{\chi}, \chi^2) = \frac{\tau(\chi^2)\tau(\overline{\chi})}{\tau(\chi)}$, and so the proof is complete. Finally, when χ is the Legendre symbol, then $\sum_{u(\mathrm{mod}\, p)} \overline{\chi}(u)\chi^2(u+1) = -\chi(-1)$, giving (4.3). $\qquad\qquad\qquad\square$

Lemma 4.3 (Trivial case). *Suppose that $\chi = \chi_0$ is trivial. Then we have*

$$(4.5) \qquad \widehat{F}(\chi) = \begin{cases} \frac{1}{p-1}, & k = 1, \\ p^{k/2}, & k \text{ even}, \\ 0, & k \geq 3 \text{ odd}. \end{cases}$$

Proof. We return to (3.7), and evaluate the sum over t as a Ramanujan sum, giving

$$(4.6) \qquad \widehat{F}(\chi_0) = \frac{1}{\varphi(p^k)} \sum_{d|p^k} d\mu(p^k/d) \sideset{}{^*}\sum_{\substack{u \, (\mathrm{mod} \, p^k) \\ (u+1)^2 \equiv 0 \, (\mathrm{mod} \, d)}} 1.$$

If $k = 1$ it is easy to check (4.5) by brute force. If k is even, then the condition $(u+1)^2 \equiv 0$ (mod d), for both $d = p^k$ and $d = p^{k-1}$, is equivalent to $u + 1 \equiv 0$ (mod $p^{k/2}$). The desired formula then follows easily.

Finally, consider $k \geq 3$ odd. For $d = p^j$ with $j \in \{k, k-1\}$, the condition $(u+1)^2 \equiv 0$ (mod d) is equivalent to $u + 1 \equiv 0$ (mod $p^{\lceil j/2 \rceil}$). Thus

$$(4.7) \qquad \widehat{F}(\chi) = \frac{1}{\varphi(p^k)} \Big(p^k \sideset{}{^*}\sum_{\substack{u \, (\mathrm{mod} \, p^k) \\ u+1 \equiv 0 \, (\mathrm{mod} \, p^{\frac{k+1}{2}})}} 1 - p^{k-1} \sideset{}{^*}\sum_{\substack{u \, (\mathrm{mod} \, p^k) \\ u+1 \equiv 0 \, (\mathrm{mod} \, p^{\frac{k-1}{2}})}} 1 \Big) = 0. \qquad \square$$

From Lemma 4.3 we easily deduce:

Corollary 4.4. *For each integer $c \geq 1$, let χ_0 denote the trivial character modulo c; then*

$$(4.8) \qquad \sum_{c \leq x} c^{-1} |\widehat{F}(\chi_0)| \ll x^\varepsilon.$$

Definition 4.5. *Suppose χ is a Dirichlet character of modulus c and conductor c^*. We say χ is semi-primitive if for every prime $p|c$, we have $1 \leq v_p(c^*) < v_p(c)$, where v_p is the p-adic valuation.*

Lemma 4.6 (Semi-primitive case). *Suppose that χ has conductor p^j, with $1 \leq j < k$. If $j \not\equiv k$ (mod 2) then $\widehat{F}(\chi) = 0$. If $j \equiv k$ (mod 2), then*

$$(4.9) \qquad |\widehat{F}(\chi)| \leq \begin{cases} p^{k/2}, & \chi^2 = 1, \\ 0, & \chi^2 \neq 1, \, p \text{ odd}, \\ 0, & \chi^2 \neq 1, \, p = 2, k > j + 2, \\ 2^{5/2}, & \chi^2 \neq 1, \, p = 2, k = j + 2. \end{cases}$$

Proof. We return to the formula (3.7). Suppose that ψ of conductor p^j induces χ. Changing variables $t \to t + p^j$ keeps $\overline{\chi}(t)$ invariant, showing that the inner sum over t vanishes unless $(u+1)^2 \equiv 0 \pmod{p^{k-j}}$. The sum over t is hence a Gauss sum repeated p^{k-j} times, giving

$$(4.10) \qquad \widehat{F}(\chi) = \frac{p^{k-j}\tau(\overline{\psi})}{\varphi(p^k)} \sum_{\substack{u \,(\mathrm{mod}\ p^k) \\ (u+1)^2 \equiv 0 \,(\mathrm{mod}\ p^{k-j})}} \overline{\psi}(u)\psi\Big(\frac{(u+1)^2}{p^{k-j}}\Big).$$

If $k - j$ is odd, then the congruence condition implies $(u + 1)^2 \equiv 0 \pmod{p^{k-j+1}}$, which causes each summand to vanish. This proves that $\widehat{F}(\chi) = 0$ if $j \not\equiv k \pmod 2$.

We proceed under the assumption $j \equiv k \pmod 2$. The congruence condition is equivalent to $u + 1 \equiv 0 \pmod{p^{\frac{k-j}{2}}}$. We write this as $u = -(1 + p^{\frac{k-j}{2}}y)$, where y now runs modulo $p^{\frac{k+j}{2}}$. This gives

$$(4.11) \qquad \widehat{F}(\chi) = \frac{p^{k-j}\tau(\overline{\psi})\psi(-1)}{\varphi(p^k)} \sum_{y \,(\mathrm{mod}\ p^{\frac{k+j}{2}})} \overline{\psi}(1 + p^{\frac{k-j}{2}}y)\psi^2(y).$$

Using Fourier inversion in the form $\overline{\psi}(x) = \frac{1}{\tau(\psi)}\sum_{t(\mathrm{mod}\ p^j)} \psi(t)e_{p^j}(tx)$, applied with $x = 1 + p^{\frac{k-j}{2}}y$, combined with simplifications, gives

$$(4.12) \qquad \sum_{y \,(\mathrm{mod}\ p^{\frac{k+j}{2}})} \overline{\psi}(1 + p^{\frac{k-j}{2}}y)\psi^2(y) = \frac{\tau(\overline{\psi})}{\tau(\psi)} \sum_{y \,(\mathrm{mod}\ p^{\frac{k+j}{2}})} \psi^2(y)e_{p^j}(p^{\frac{k-j}{2}}y).$$

Noting the inner sum is the same sum repeated $\frac{p^{\frac{k+j}{2}}}{p^j}$ times, we obtain

$$(4.13) \qquad \widehat{F}(\chi) = \frac{p^{k-j}\tau(\overline{\psi})^2\psi(-1)}{\varphi(p^k)\tau(\psi)}p^{\frac{k-j}{2}} \sum_{y \,(\mathrm{mod}\ p^j)} \psi^2(y)e_{p^j}(p^{\frac{k-j}{2}}y).$$

Now our work breaks into cases. First suppose that the conductor of ψ^2 is p^j. For p odd, this condition is equivalent to $\chi^2 \neq 1$. In this case, the sum vanishes, giving the second line of (4.9). If $\chi^2 = 1$ then $j = 1$, and direct evaluation gives a bound consistent with the first line of (4.9)

Finally, consider $p = 2$. If $j \in \{2, 3\}$ then $\chi^2 = 1$, and if $j \geq 4$ then the conductor of ψ^2 must equal 2^{j-1}. Also, note that there are no characters of conductor 2^1, so this covers everything. For $j = 2, 3$ (equivalently, $\chi^2 = 1$), we bound the inner sum over y trivially, giving the desired bound. Next suppose $j \geq 4$, so ψ^2 has conductor 2^{j-1}. If $k > j + 2$ then the sum vanishes. If $k = j + 2$, then the inner sum over y is simply a Gauss sum (of modulus 2^{j-1}) repeated twice. This gives the claimed bound. $\square$

Corollary 4.7. *We have*

$$(4.14) \qquad \sum_{c \leq x} \sum_{\substack{\chi_2 \,(\mathrm{mod}\ c) \\ \textit{semi-primitive}}} c^{-1}|\widehat{F}(\chi_2)| \ll x^\varepsilon.$$

Proof. By Lemma 4.1, it suffices to show the bound separately for c running over odd intgers, and for c running over powers of 2. For c a power of 2, we simply use that $|\widehat{F}(\chi)| \ll 1$.

Now consider c odd. Lemma 4.6 implies that c factors in the form $b_1 b_2^2$ where b_1 is square-free, $b_1 | b_2$, and χ_2 is the character of modulus $b_1 b_2^2$ induced by the Jacobi symbol of conductor b_1. Then $|\widehat{F}(\chi_2)| \ll (b_1 b_2^2)^{1/2}$. To complete the proof, we use

$$(4.15) \qquad \sum_{b_1 b_2^2 \le x} \sum_{b_1 | b_2} \frac{(b_1 b_2^2)^{1/2}}{b_1 b_2^2} \ll x^\varepsilon. \qquad \qquad \square$$

5. Final steps

5.1. Proof of Theorem 1.2.
We apply Lemma 3.3 to (3.2), giving

$$(5.1) \quad \mathcal{K}(\Delta, T, N) = \sum_{X \text{ dyadic}} \sum_d \sum_g \frac{\mu(g)}{g} \int_{(\sigma)} \frac{\widetilde{G}_X(-s)}{(dg)^s} \sum_{c=1}^\infty \sum_{\chi \,(\mathrm{mod}\ c)} \frac{\widehat{F}(\chi)\chi(\overline{dg})}{c^{1+s}} B(\chi, s) \frac{ds}{2\pi i},$$

where $\sigma \gg 1$ is large, and where

$$(5.2) \qquad\qquad B(\chi, s) = \sum_{\substack{d | (m^2, n^2) \\ g | mn/d}} a_m \overline{a_n} \chi(mn)(mn)^s.$$

By Lemma 3.2, and shifting the contour far to the right, we may truncate the sum at $cdg \ll C_{\max} := \frac{N^2}{X}(NT)^\varepsilon$. After this truncation all the sums are finite, and we then shift the contour of integration to $\mathrm{Re}(s) = 0$. Then

$$(5.3) \quad |\mathcal{K}(\Delta, T, N)| \ll \sum_{X \text{ dyadic}} \sum_{dg \ll C_{\max}} \frac{1}{g} \int_{|t| \ll PT^\varepsilon} \frac{\Delta T}{(XP)^{1/2}} \sum_{c \ll \frac{C_{\max}}{dg}} \sum_{\chi \,(\mathrm{mod}\ c)} \frac{|\widehat{F}(\chi)|}{c} |B(\chi, it)| dt,$$

plus a small error term. Next we factor c and χ as follows. Write $\chi = \chi_0 \chi_1 \chi_2$ and $c = c_0 c_1 c_2$, with $(c_i, c_j) = 1$ for $i \ne j$, and where χ_j has modulus c_j. The factorization is characterized by the assumption that χ_0 is trivial, χ_1 is primitive modulo c_1, and that χ_2 is semi-primitive. This factorization corresponds to the three cases from Lemmas 4.2, 4.3, and 4.6. Lemma 4.1 implies $|\widehat{F}(\chi_0 \chi_1 \chi_2)| = |\widehat{F}(\chi_0)\widehat{F}(\chi_1)\widehat{F}(\chi_2)|$. Using this factorization, and arranging the expression appropriately, we have

$$(5.4) \quad |\mathcal{K}(\Delta, T, N)| \ll \sum_{X \text{ dyadic}} \sum_{dg \ll C_{\max}} \frac{1}{g} \sum_{c_0 c_2 \ll \frac{C_{\max}}{dg}} \sum_{\substack{\chi_2 \,(\mathrm{mod}\ c_2) \\ \text{semi-primitive}}} \frac{|\widehat{F}(\chi_0)\widehat{F}(\chi_2)|}{c_0 c_2}$$

$$\int_{|t| \ll PT^\varepsilon} \frac{\Delta T}{(XP)^{1/2}} \sum_{c_1 \ll \frac{C_{\max}}{dg c_0 c_2}} \sideset{}{^*}\sum_{\chi_1 \,(\mathrm{mod}\ c_1)} \frac{|\widehat{F}(\chi_1)|}{c_1} |B(\chi_1 \chi_0 \chi_2, it)| dt,$$

plus a small error term.

Lemma 4.2 implies $|\widehat{F}(\chi_1)| \ll c_1^{\varepsilon}$. Then by Corollary 2.4 (absorbing $\chi_0\chi_2$ into the definition of the vector $\mathbf{a}$), we have that the second line in (5.4) satisfies the bound

$$(5.5) \quad \int_{|t| \ll PT^{\varepsilon}} \frac{\Delta T}{(XP)^{1/2}} \sum_{c_1 \ll \frac{C_{\max}}{dgc_0c_2}} \sideset{}{^*}\sum_{\chi_1 \,(\mathrm{mod}\ c_1)} \frac{|\widehat{F}(\chi_1)|}{c_1} |B(\chi_1\chi_0\chi_2, it)| dt$$

$$\ll (NT)^{\varepsilon} \max_{1 \leq C_1 \ll \frac{C_{\max}}{dgc_0c_2}} \frac{\Delta T}{C_1(XP)^{1/2}} \left(C_1^2 P + \frac{N}{d'} \right) \sum_n |a_{nd'}|^2.$$

Simplifying and substituting into (5.4), we obtain

$$(5.6) \quad |\mathcal{K}(\Delta, T, N)| \ll \sum_{X \text{ dyadic}} \sum_{dg \ll C_{\max}} \frac{1}{g} \sum_{c_0c_2 \ll \frac{C_{\max}}{dg}} \sum_{\substack{\chi_2 \,(\mathrm{mod}\ c_2) \\ \text{semi-primitive}}} \frac{|\widehat{F}(\chi_0)\widehat{F}(\chi_2)|}{c_0c_2}$$

$$\Delta T(NT)^{\varepsilon} \left(\frac{C_{\max}\sqrt{P}}{dgc_0c_2\sqrt{X}} + \frac{N}{d'\sqrt{XP}} \right) \sum_n |a_{nd'}|^2.$$

Next we apply Corollaries 4.4 and 4.7 to treat the sums over c_0 and c_2. We also sum trivially over g, which is bounded by a truncated harmonic series. The sum over d is similar, since $\sum_{d \leq x} \frac{1}{d'} \ll x^{\varepsilon}$. In all, this gives

$$(5.7) \quad |\mathcal{K}(\Delta, T, N)| \ll \sum_{\substack{X \gg \Delta T \\ \text{dyadic}}} \Delta T(NT)^{\varepsilon} \left(\frac{C_{\max}\sqrt{P}}{\sqrt{X}} + \frac{N}{\sqrt{XP}} \right) |\mathbf{a}|^2.$$

The former term with $\frac{C_{\max}\sqrt{P}}{d\sqrt{X}}$ matches with the first term in (1.28); precisely, we have

$$(5.8) \quad \max_{X \gg \Delta T} \Delta T \frac{N^2}{X} \frac{\sqrt{1 + \frac{T^2}{X}}}{\sqrt{X}} \ll \frac{N^2}{\Delta}.$$

For the latter term, we have

$$(5.9) \quad \max_{X \gg \Delta T} \Delta T \frac{N}{\sqrt{XP}} \ll \Delta N,$$

which matches with (1.31).

To summarize, we have shown a bound

$$(5.10) \quad \mathcal{M}(\Delta, T, N) \ll \Delta T + (NT)^{\varepsilon} \left(\Delta N + \frac{N^2}{\Delta} \right)$$

which agrees with (1.32). The proof of Theorem 1.2 is completed by choosing Δ as described in the paragraph following (1.32). $\qquad\square$

5.2. Deduction of Corollary 1.3.

The left hand side of (1.13) takes the form

$$(5.11) \quad \sum_{T \leq t_j \leq T+\Delta} \left| \sum_d b_d \lambda_j(d^2) \right|^2, \qquad b_d = \sum_k a_{dk^2}.$$

By Cauchy's inequality, the sum over d can be broken down into dyadic segments. Consider a piece with $d \asymp D$. The proof of Theorem 1.2 estimated the diagonal terms and the non-diagonal terms separately. The diagonal term, represented by ΔT in (1.10), precisely leads to $\mathcal{D}$. The remaining non-diagonal terms are then bounded by

$$(5.12) \qquad (NT)^{\varepsilon} \Big(\Delta D + D(T+D)^{1/2} + \frac{D^2}{T} \Big) \sum_{d \asymp D} |b_d|^2.$$

By Cauchy's inequality, using that $k^2 \asymp \frac{N}{D}$, we have

$$(5.13) \qquad \sum_{d \asymp D} |b_d|^2 = \sum_{d \asymp D} \Big| \sum_k a_{dk^2} \Big|^2 \ll \frac{\sqrt{N}}{\sqrt{D}} \sum_{d,k} |a_{dk^2}|^2 \ll \frac{\sqrt{N}}{\sqrt{D}} N^{\varepsilon} |\mathbf{a}|^2.$$

Inserting this into (5.12), and summing over dyadic values of D then gives rise to (1.13), since for the non-diagonal terms, $D \asymp N$ is the essential range.

6. Lower bounds

6.1. **Proof of Theorem 1.5.** We briefly sketch the proof of Theorem 1.5. We take $a_n = 1$ for $n = p$ prime, and $a_n = 0$, otherwise. Note that $\tau_{it}(p^2) = 1 + p^{2it} + p^{-2it}$. Let $A = \sum_p 1$ and $B(t) = \sum_p (p^{2it} + p^{-2it})$, so that

$$(6.1)$$
$$\int_T^{T+\Delta} \Big| \sum_{N \le n \le 2N} a_n \tau_{it}(n^2) \Big|^2 dt = \int_T^{T+\Delta} |A + B(t)|^2 dt = \int_T^{T+\Delta} |A|^2 + 2\mathrm{Re}(A\overline{B}(t)) + |B(t)|^2 dt.$$

Note

$$(6.2) \qquad \int_T^{T+\Delta} |A|^2 dt = \int_T^{T+\Delta} \Big| \sum_{p \sim N} a_p \Big|^2 dt = \Delta \frac{N^2}{(\log N)^2} (1 + o(1)),$$

by the prime number theorem. Meanwhile, by the mean value theorem for Dirichlet polynomials (see Theorem 2.3 with $Q = 1$), we have

$$(6.3) \qquad \int_T^{T+\Delta} |B(t)|^2 dt \ll (\Delta + N) \sum_p 1 \ll (\Delta + N) \frac{N}{\log N}.$$

Moreover, by Cauchy-Schwarz, we have

$$(6.4) \qquad \int_T^{T+\Delta} |AB(t)| dt \ll \Delta^{1/2} (\Delta + N)^{1/2} \Big(\frac{N}{\log N} \Big)^{3/2}.$$

Thus

$$(6.5) \qquad \int_T^{T+\Delta} \Big| \sum_{N \le n \le 2N} a_n \tau_{it}(n^2) \Big|^2 dt = \Delta \frac{N^2}{\log^2 N} \Big[1 + o(1) + O\Big(\frac{(\log N)^{1/2}}{\Delta^{1/2}} \Big) \Big].$$

Since $\sum_n |a_n|^2 \sim \frac{N}{\log N}$, and using $w_t^{-1} = T^{o(1)}$, this shows Theorem 1.5.

6.2. **Proof of Proposition 1.6.** The idea and details of the proof are similar to those presented in Section 6.1, so we will be brief. We take a_n defined by $a_n = 1$ for $n = p^2 \asymp N$, with p prime, and $a_n = 0$ otherwise. Note $\lambda_{\mathrm{sym}^2 u_j}(p^2) = 1 + \lambda_j(p^2)$. Let $A = \sum_p a_{p^2} \asymp \frac{\sqrt{N}}{\log N}$ and $B_j = \sum_p \lambda_j(p^2)$. Then we have

$$(6.6) \qquad \sum_{T \leq t_j \leq T+\Delta} w_j^{-1}|A + B_j|^2 = \sum_{T \leq t_j \leq T+\Delta} w_j^{-1}\Big(|A|^2 + 2\mathrm{Re}(A\overline{B_j}) + |B_j|^2\Big).$$

The rest of the proof now plays out nearly identically to that of Theorem 1.5. The conditions in place in Proposition 1.6, together with (1.7), show that $\sum_{t_j} w_j^{-1}|B_j|^2 = o(\sum_{t_j} w_j^{-1}|A|^2)$, which implies that

$$(6.7) \qquad \sum_{T \leq t_j \leq T+\Delta} w_j^{-1}|A + B_j|^2 \gg \Delta T|A|^2 \asymp \Delta T \frac{\sqrt{N}}{\log N}|\mathbf{a}|^2.$$

UPDATE REFS

References

[B] V. Blomer, *Applications of the Kuznetsov formula on GL(3)*. Invent. Math. 194 (2013), no. 3, 673–729.

[BB] V. Blomer and J. Buttcane, *Global decomposition of GL(3) Kloosterman sums and the spectral large sieve.* J. Reine Angew. Math. 757 (2019), 51–88

[DK] W. Duke, and E. Kowalski, *A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations.* With an appendix by Dinakar Ramakrishnan. Invent. Math. 139 (2000), no. 1, 1–39.

[DR] A. Dunn and M. Radziwiłł, *Bias in cubic Gauss sums: Patterson's conjecture.* arXiv:2109.07463.

[G] P. X. Gallagher, *A large sieve density estimate near $\sigma = 1$.* Invent. Math. 11 1970 329–339.

[GJ] S. Gelbart and H. Jacquet, *A relation between automorphic representations of GL(2) and GL(3).* Ann. Sci. École Norm. Sup. (4) 11 (1978), no. 4, 471–542.

[H-B] D.R. Heath-Brown, *A mean value estimate for real character sums.* Acta Arith. 72 (1995), no. 3, 235–275.

[I1] H. Iwaniec, *Fourier coefficients of cusp forms and the Riemann zeta-function.* Seminar on Number Theory, 1979–1980 (French), Exp. No. 18, 36 pp., Univ. Bordeaux I, Talence, 1980.

[I2] H. Iwaniec, *Topics in Classical Automorphic Forms*, Grad. Stud. Math., vol 17, Amer. Math. Soc., 1997.

[IK] H. Iwaniec and E. Kowalski, *Analytic number theory.* American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004.

[IL] H. Iwaniec and Xiaoqing Li, *The orthogonality of Hecke eigenvalues.* Compos. Math. 143 (2007), no. 3, 541–565.

[J] M. Jutila, *On spectral large sieve inequalities.* Funct. Approx. Comment. Math. 28 (2000), 7–18.

[JM] M. Jutila and Y. Motohashi, *Uniform bound for Hecke L-functions.* Acta Math. 195 (2005), 61–115.

[KY] R. Khan and M. Young, *Moments and hybrid subconvexity for symmetric-square L-functions.* Preprint, 2020, arXiv:2009.08419. To appear in J. Inst. Math. Jussieu.

[KPY] E. Kıral, I. Petrow, and M. Young, *Oscillatory integrals with uniformity in parameters.* J. Théor. Nombres Bordeaux 31 (2019), no. 1, 145–159.

[La] J. W. C. Lam, *The second moment of the central values of the symmetric square L-functions.* Ramanujan J. 38 (2015), no. 1, 129–145.

[Lu] W. Luo, *The spectral mean value for linear forms in twisted coefficients of cusp forms.* Acta Arith. 70 (1995), no. 4, 377–391.

[S] D. Soudry, *On Langlands functoriality from classical groups to GL_n.* Astérisque, no. 298 (2005), 335–390.

[TZ] J. Thorner and A. Zaman, *An unconditional GL_n large sieve*, Adv. Math. 378 (2021), 107529, 24 pp.

[Y1] M. Young, *The second moment of $GL(3) \times GL(2)$ L-functions at special points.* Math. Ann. 356 (2013), no. 3, 1005–1028.

[Y2] M. Young, *Bilinear forms with GL_3 Kloosterman sums and the spectral large sieve.* Int. Math. Res. Not. IMRN 2016, no. 21, 6453–6492.

[Y3] M. Young, *An improved spectral large sieve inequality for $SL_3(\mathbb{Z})$.* Preprint, 2021, arXiv:2102.02796.

DEPARTMENT OF MATHEMATICS, TEXAS A&M UNIVERSITY, COLLEGE STATION, TX 77843-3368, U.S.A.

Email address: myoung@math.tamu.edu