ALGEBRAIC TWISTS OF $GL_3 \times GL_2$ L-FUNCTIONS

YONGXIAO LIN, PHILIPPE MICHEL, AND WILL SAWIN

ABSTRACT. We prove that the coefficients of a $GL_3 \times GL_2$ Rankin–Selberg *L*-function do not correlate with a wide class of trace functions of small conductor modulo primes, generalizing the corresponding result [FKM15a] for GL_2 and [KLMS20] for GL_3 . This result is inspired by a recent work of P. Sharma who discussed the case of a Dirichlet character of prime modulus.

Contents

1	Introduction	1
2.	Background material regarding automorphic forms	8
3.	First transformations	10
4.	The case $q \not n_1$	17
5.	Contribution of the $n = 0$ frequency	24
6.	Contribution from the $n \neq 0$ frequencies	25
7.	Bounding $S_V(K, X)$: the final steps	28
8.	Square-root cancellation for certain exponential sums	32
9.	Some special cases	39
Ret	ferences	45

1. INTRODUCTION

We begin by describing the content of a recent preprint of P. Sharma [Sha19] which is the starting point of the present work.

Let $(\lambda(r, n))_{r,n}$ be the Hecke eigenvalues of a GL₃ cusp form φ and $(\lambda_f(m))_m$ be Hecke eigenvalues of a GL₂ cusp form f (holomorphic or Maass); for simplicity we assume that both have level 1 (i.e., are SL₃(**Z**) and SL₂(**Z**)-invariant respectively). Their Rankin–Selberg *L*-function is the Dirichlet series

$$L(\varphi \times f,s) = \sum_{n,r \geqslant 1} \frac{\lambda(r,n)\lambda_f(n)}{(nr^2)^s}, \ \mathfrak{Re} \ s > 1.$$

This has an Euler product of degree 6 admitting analytic continuation to **C** and a functional equation relating $L(\varphi \times f, s)$ to $L(\overline{\varphi} \times \overline{f}, 1-s)$.

Given q a prime and $\chi : (\mathbf{Z}/q\mathbf{Z})^{\times} \to \mathbf{C}^{\times}$ a non-trivial Dirichlet character, the twisted L-function is

$$L(\varphi \times f \times \chi, s) = \sum_{n,r \geqslant 1} \frac{\lambda(r,n)\lambda_f(n)\chi(nr^2)}{(nr^2)^s}, \ \mathfrak{Re} \ s > 1.$$

This again has an Euler product of degree 6 admitting analytic continuation to **C** and a functional equation relating $L(\varphi \times f \times \chi, s)$ to $L(\overline{\varphi} \times \overline{f} \times \overline{\chi}, 1-s)$.

Y. L. and Ph. M. were partially supported by a DFG-SNF lead agency program grant (grant 200020L_175755) and the SNF (grant 200021_197045). W. S. served as a Clay Research Fellow while working on this paper. To appear in American Journal of Math. Mon 18th Apr, 2022.

The following bound, known as the *convexity bound* (in the q-aspect) is not too hard to establish: for $\Re \mathfrak{e} s = 1/2$, one has

$$L(\varphi \times f \times \chi, s) \ll_{f,\varphi,s} q^{3/2 + o(1)}$$

The subconvexity problem aims at improving the exponent 3/2. In [Sha19], P. Sharma provided a detailed description of a solution of this problem with the bound

$$L(\varphi \times f \times \chi, s) \ll_{f,\varphi,s} q^{3/2 - 1/16 + o(1)}$$

Inspired by the previous works [FKM15a, KLMS20], we follow Sharma's strategy and generalise this bound by replacing χ by a generic *trace function*

$$K: (\mathbf{Z}/q\mathbf{Z})^{\times} \to \mathbf{C},$$

that is the (restriction to \mathbf{F}_q^{\times} of the) Frobenius trace function associated to some geometrically irreducible middle extension sheaf \mathcal{F} on $\mathbf{P}_{\mathbf{F}_q}^1$ pure of weight 0 satisfying additional generic condition (we call such a sheaf "good").

Indeed (see §9), it follows from the (Mellin) expansion of $K_{|\mathbf{F}_q^{\times}}$ into Dirichlet characters that the series

$$L(\varphi \times f \times K, s) = \sum_{\substack{n,r \ge 1 \\ (nr^2,q)=1}} \frac{\lambda(r,n)\lambda_f(n)K(nr^2)}{(nr^2)^s}, \ \mathfrak{Re} \ s > 1,$$

has analytic continuation to **C** and satisfies a functional equation relating $L(\varphi \times f \times K, s)$ to $L(\overline{\varphi} \times \overline{f} \times \widetilde{K}^6, s)$ where \widetilde{K}^6 : $\mathbf{F}_q^{\times} \to \mathbf{C}$ is a suitable "GL₆" transform of K (see (9.2)). In most cases, \widetilde{K}^6 is essentially (the restriction to \mathbf{F}_q^{\times} of) a trace function. From this one can deduce a "convexity" bound

(1.1)
$$L(\varphi \times f \times K, s) \ll_{f,\varphi,s} q^{3/2+o(1)}, \ \mathfrak{Re} \ s = 1/2$$

Our goal is to improve the exponent 3/2.

Using approximate functional equation techniques, one sees that solving this "subconvexity problem" for $L(\varphi \times f \times K, s)$ is tantamount to bounding non-trivially sums of the shape

$$S_V^t(K,X) := \sum_{r,n} \lambda(r,n)\lambda_f(n)K(nr^2)V(\frac{nr^2}{X})$$

for V a (fixed) smooth function with compact support in [1,2[and $X \ge 1$ a positive parameter of size $X \approx q^3$ (the trivial bound being $|S_V^t(K,X)| \le X^{1+o(1)}$). Indeed when φ , f and s are fixed, q^3 is approximately the square-root of the conductor of the degree 6 *L*-functions $L(\varphi \times f \times \chi, s)$ for the primitive Dirichlet characters $\chi \pmod{q}$.

For this we introduce the following conditions on the sheaf \mathcal{F} both of which are generic (i.e., hold for a "typical" sheaf). Let $[\times \lambda] \colon \mathbb{A}^1 \to \mathbb{A}^1$ denote the map which multiplies the coordinate by λ .

- (MO) There is no any $\lambda \in \mathbf{F}_q^{\times}$ such that the geometric monodromy group of \mathcal{F} has some quotient which is equal, as a representation of the geometric fundamental group π_1 into an algebraic group, to the geometric monodromy group of the Kloosterman sheaf $[\times \lambda]^* \mathcal{K} \ell_2$ modulo ± 1 .
- (SL) The local monodromy representation of \mathcal{F} at ∞ has no summand with slope 1/2.

We will also need to assume that the sheaf \mathcal{F} is *Fourier*, *i.e.*, that it is not geometrically isomorphic to either the constant sheaf or any Artin–Schreier sheaf (whose trace functions are additive characters).

We will say that \mathcal{F} is *good* if \mathcal{F} is Fourier and satisfies both of (MO) and (SL) and say that \mathcal{F} is bad otherwise.

Our main result provides bounds for the sums $S_V^t(K, X)$ when K is the trace function of a good sheaf, X is not too far from q^3 and when the smooth function V is allowed to vary and oscillate mildly as q varies.

Theorem 1.1. Let q be a prime and K be the trace function modulo q associated with a good sheaf \mathfrak{F} . Let $Z \ge 1$ be some parameter and $V \in \mathfrak{C}^{\infty}_{c}(\mathbf{R})$ be a smooth function compactly supported in the interval [1, 2] and satisfying for all $i \ge 0$

(1.2)
$$V^{(i)}(x) \ll_i Z^i.$$

Let $X \ge 1$ be such that $Z^4 q^{11/4} < X < Z^4 q^{\frac{7-\theta_3}{2}}$. We have

(1.3)
$$S_V^t(K,X) \ll q^{o(1)} \left(Z X^{3/4} q^{11/16} + Z^{\frac{4(1-\theta_3)}{3-2\theta_3}} X^{\frac{2-\theta_3}{3-2\theta_3}} q^{\frac{11(1-\theta_3)}{4(3-2\theta_3)}} + X q^{-1/8} \right);$$

here the implicit constant depends on φ , f, $C(\mathcal{F})$ and the implicit constant in (1.2) and $\theta_3 = 5/14$ is the best known bound towards the Ramanujan–Petersson conjecture on GL₃. In particular for $X = q^3$ (the convexity range) one obtains

$$S_V^t(K,q^3) \ll Zq^{3-1/16+o(1)}$$

Remark 1.2. (1) This bound is non-trivial (i.e., is o(X)) as long as

$$X \gg q^{3-1/4+\eta}$$
, for some $\eta > 0$.

The assumption $X < Z^4 q^{(7-\theta_3)/2}$ in the statement is non-essential and possibly can be removed. We make this assumption only to simplify our treatment at one certain point; see (3.3).

(2) For $k \ge 2$ an integer, the hyper-Kloosterman sheaf $\mathcal{K}\ell_k$ whose attached trace function is given by the k-1-dimensional hyper-Kloosterman sums

$$\mathrm{Kl}_k(n;q) = \frac{1}{q^{\frac{k-1}{2}}} \sum_{\substack{x_1,\dots,x_k \in \mathbf{F}_q^\times\\x_1,\dots,x_k = n}} e\left(\frac{x_1 + \dots + x_k}{q}\right)$$

is good unless k = 2. In that case neither (SL) nor (MO) holds. We will explain in §9 how a duality principle which gives the analytic continuation of $L(\varphi \times f \times K, s)$ allows for partial results for such sums.

(3) As was pointed out to us by V. Blomer, when $\varphi = \text{Sym}_2(g)$ is the symmetric square lift of a GL₂-modular form g and taking f to be a suitable Eisenstein series, one can obtain –by a variation on [Blo12], using the Petrow–Young variant of the Conrey–Iwaniec method [CI00, PY20]– the stronger subconvex bound

$$|L(\operatorname{Sym}_2(g) \times \chi, s)|^2 \ll q^{3/2 - 1/4 + o(1)}, \ \mathfrak{Re} \ s = 1/2$$

for any Dirichlet character $\chi \pmod{q}$. Since this approach uses positivity of central values, it is not yet clear whether this could be extended to general trace functions.

1.1. **Principle of the proof.** To illustrate the main idea of our approach, we provide a quick sketch of the proof, focusing on just the "generic" case in various transformations. Therefore we will assume $X \simeq q^3$ and r = 1, and we will suppress the smooth test functions from our notation. We denote the GL₃ Hecke eigenvalues by $(\lambda(1, n))_n$ and the GL₂ Hecke eigenvalues by $(\lambda(n))_n$.

Let $L \ge 1$ be a parameter (a positive power of q) whose optimal size to be determined later, and let $\ell \in [L, 2L]$ be prime numbers. We use the Hecke relation to write $\lambda(1, n\ell) \approx \lambda(1, n)\lambda(1, \ell)$ and then use the Kronecker symbol to separate oscillations of $\lambda(1, n\ell)$ and $\lambda(n)K(n)$. Our starting point is to follow [Sha19] and write

$$\begin{split} & \mathcal{S} := \sum_{n \sim q^3} \lambda(1, n) \lambda(n) K(n) \\ & \approx \frac{1}{L} \sum_{\ell \sim L} \overline{\lambda(1, \ell)} \sum_{n \sim q^3} \lambda(1, n\ell) \lambda(n) K(n) \\ & = \frac{1}{L} \sum_{\ell \sim L} \overline{\lambda(1, \ell)} \sum_{n \sim q^3 \ell} \lambda(1, n) \sum_{m \sim q^3} \lambda(m) K(m) \delta(n, m\ell) . \end{split}$$

We first use a conductor-decreasing trick to write

$$\delta(n, m\ell) = \delta_{q|n-m\ell} \cdot \delta\left(\frac{n-m\ell}{q}, 0\right).$$

Expressing $\delta_{q|n-m\ell} = \frac{1}{q} \sum_{u(q)} e\left(\frac{(n-m\ell)u}{q}\right)$ and using the delta symbol by Duke–Friedlander–Iwaniec [DFI93] to detect the second one:

$$\delta\left(\frac{n-m\ell}{q},0\right) \approx \frac{1}{C} \sum_{c \sim C} \frac{1}{c} \sum_{u(c)}^{\star} e\left(\frac{u(n-m\ell)}{cq}\right),$$

we have the following approximation

$$\delta(n, m\ell) \approx \frac{1}{qC} \sum_{c \sim C} \frac{1}{c} \sum_{u(cq)}^{\star} e\left(\frac{u(n-m\ell)}{cq}\right).$$

Here C is a large parameter which we will choose as

$$C = \left(\frac{q^3L}{q}\right)^{1/2} = qL^{1/2}.$$

Now plugging this approximation for $\delta(n, m\ell)$ in, we can write our original sum as

$$\mathbb{S} \approx \frac{\sqrt{q^3 L} \sqrt{q^3}}{LCq} \sum_{c \sim C} \frac{1}{c} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{u(cq)} \sum_{n \sim q^3 L} \frac{\lambda(1,n)}{\sqrt{n}} e\left(\frac{un}{cq}\right) \sum_{m \sim q^3} \frac{\lambda(m)}{\sqrt{m}} K(m) e\left(\frac{-um\ell}{cq}\right).$$

We use Fourier inversion to separate the *m*-variable from K(m) and rewrite the *m*-sum above as

$$\frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) \sum_{m \sim q^3} \frac{\lambda(m)}{\sqrt{m}} e\left(\frac{-(bc+u\ell)m}{cq}\right).$$

We now use Voronoi summation to dualize the *n*- and *m*-variables respectively, getting

$$\sum_{n \sim q^3 L} \frac{\lambda(1, n)}{\sqrt{n}} e\left(\frac{un}{cq}\right) \approx \sum_{n \sim \frac{(cq)^3}{q^3 L} \approx q^3 L^{1/2}} \frac{\lambda(n, 1)}{\sqrt{n}} \operatorname{Kl}_2(\bar{u}n; cq);$$
$$\sum_{m \sim q^3} \frac{\lambda(m)}{\sqrt{m}} e\left(\frac{-(bc + u\ell)m}{cq}\right) \approx \sum_{\substack{m \sim \frac{(cq)^2}{q^3} \approx qL}} \frac{\overline{\lambda(m)}}{\sqrt{m}} e\left(\frac{\overline{bc + u\ell} m}{cq}\right).$$

Therefore, after applying the two Voronoi summations, we arrive at the following dual sum

$$\begin{split} \mathbb{S} \approx & \frac{\sqrt{q^3 L}}{LCq} \frac{q^2}{C} \sum_{c \sim C} \frac{1}{c} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{n \sim q^3 L^{1/2}} \frac{\lambda(n,1)}{\sqrt{n}} \sum_{m \sim qL} \overline{\lambda(m)} \\ & \frac{1}{q^{1/2}} \sum_{u(cq)}^{\star} \sum_{b \; (\mathrm{mod}\; q)} \widehat{K}(b) \operatorname{Kl}_2(\bar{u}n;cq) e\left(\frac{\overline{bc+u\ell} \; m}{cq}\right). \end{split}$$

By splitting the sum modulo cq into a product of sums modulo c and q respectively, the second line above can be written as

$$\sum_{u_1(c)}^{\star} \operatorname{Kl}_2(\bar{q}^2 \overline{u_1} n; c) e\left(\frac{\overline{u_1 \ell q m}}{c}\right) \cdot \frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) \sum_{u(q)}^{\star} \operatorname{Kl}_2(\bar{c}^2 \bar{u}n; q) e\left(\frac{\overline{bc + u\ell \bar{c}m}}{q}\right)$$
$$\approx \sqrt{c} e\left(\frac{n\bar{q}\ell \overline{m}}{c}\right) \sum_{u(q)}^{\star} \operatorname{Kl}_2(\bar{c}^3 \bar{u}n; q) L_{\bar{c}^2 m, \ell}(u; q),$$

where

$$L_{\alpha,\beta}(u;q) := \frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) e\left(\frac{\alpha \overline{b+\beta u}}{q}\right)$$

Hence

.

$$\mathbb{S} \approx \frac{q^{5/2}}{L^{1/2}C^{3/2}} \sum_{n \sim q^3 L^{1/2}} \frac{\lambda(n,1)}{\sqrt{n}} \sum_{c \sim C} \frac{1}{c} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{m \sim qL} \overline{\lambda(m)} e\left(\frac{n\bar{q}\ell\bar{m}}{c}\right) \sum_{u(q)} {}^{\star} \mathrm{Kl}_2(\bar{c}^3\bar{u}n;q) L_{\bar{c}^2m,\ell}(u;q).$$

We next apply Cauchy–Schwarz inequality to remove the GL_3 coefficients while keeping the *c*-sum "inside":

$$S \ll \frac{q^{5/2}}{L^{1/2}C^{3/2}} \left(\sum_{n \sim q^3L^{1/2}} \frac{|\lambda(n,1)|^2}{n} \right)^{1/2} \left(\sum_{n \sim q^3L^{1/2}} \left| \sum_{c \sim C} \sum_{\ell \sim L} \sum_{m \sim qL} \sum_{u(q)}^{\star} (\dots) \right|^2 \right)^{1/2} := \frac{q^{5/2}}{L^{1/2}C^{3/2}} \Omega^{1/2}.$$

Remark 1.3. The *n*-variable has size $q^3L^{1/2}$ and was originally weighted by the GL₃ coefficients $\lambda(n,1)$ times a periodic arithmetic function of modulus qc. After applying Cauchy-Schwarz the resulting periodic functions have moduli $qcc' \approx q^3 L$ which is not much bigger than the size of n which is now smooth and therefore one can expect to be able to analyse this sum further.

We continue to analyse the sum Ω . Opening the square and switching the order of summations, we arrive at

$$\Omega = \sum_{c \sim C} \frac{1}{c} \sum_{c' \sim C} \frac{1}{c'} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{\ell' \sim L} \lambda(1,\ell') \sum_{m \sim qL} \overline{\lambda(m)} \sum_{m' \sim qL} \lambda(m')$$

$$\sum_{u(q)}^{*} L_{\bar{c}^2 m,\ell}(u;q) \sum_{u'(q)}^{*} \overline{L_{\bar{c'}^2 m',\ell'}(u';q)} \sum_{n \sim q^3 L^{1/2}} \operatorname{Kl}_2(\bar{c}^3 \bar{u}n;q) \operatorname{Kl}_2(\bar{c'}^3 \bar{u'}n;q) e\left(\frac{n\bar{q}\ell \overline{m}}{c}\right) e\left(\frac{-n\bar{q}\ell' \overline{m'}}{c'}\right).$$

$$5$$

We apply Poisson summation to the *n*-sum; after computing the resulting Fourier transform and several manipulations we obtain

$$\begin{split} \Omega \approx & \sum_{c \sim C} \frac{1}{c} \sum_{c' \sim C} \frac{1}{c'} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{\ell' \sim L} \lambda(1,\ell') \sum_{m \sim qL} \overline{\lambda(m)} \sum_{m' \sim qL} \lambda(m') \\ & q^3 L^{1/2} \sum_{|n| < L^{1/2}} \sum_{v(q)} Z(v) \overline{Z'(v - n\overline{cc'})} \cdot \delta_{n + \ell \overline{m}c' - \ell' \overline{m'}c \equiv 0 \, (\mathrm{mod} \, cc')}, \end{split}$$

where

$$Z(v) = Z_{\alpha,\beta,\gamma}(v) := \frac{1}{q^{1/2}} \sum_{x \in \mathbf{F}_q^{\times}} \operatorname{Kl}_2(\beta \gamma x; q) K(xv) \operatorname{Kl}_2(\alpha xv; q)$$

with $(\alpha, \beta, \gamma) = (\overline{c}^2 m, \ell, \overline{c}^3)$ and Z'(v) is defined likewise with $(\alpha', \beta', \gamma') = (\overline{c'}^2 m', \ell', \overline{c'}^3)$. We consider the cases where n = 0 and $n \neq 0$ separately and denote their contribution to Ω by Ω_0 and Ω_{\neq} respectively.

For n = 0, one has c = c', and since the sheaf \mathcal{F} is good, we have by Proposition 4.5,

$$\sum_{v} Z(v) \overline{Z'(v)} \ll q \cdot \delta_{\ell m' \equiv \ell' m \pmod{q}} + q^{1/2};$$

this gives

$$\Omega_{0} = \sum_{c \sim C} \frac{1}{c^{2}} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{\ell' \sim L} \lambda(1,\ell') \sum_{m \sim qL} \overline{\lambda(m)} \sum_{m' \sim qL} \lambda(m') \\ \times q^{3} L^{1/2} (q \cdot \delta_{\ell m' \equiv \ell' m \pmod{q}} + q^{1/2}) \delta_{\ell m' \equiv \ell' m \pmod{q}} \\ \ll \frac{q^{5} L^{5/2}}{C} + \frac{q^{11/2} L^{9/2}}{C^{2}},$$

whose corresponding contribution to S is

$$\mathbb{S}_0 \ll rac{q^3}{L^{1/4}} + q^{11/4}L^{1/2}.$$

Remark 1.4. This bound is admissible is long as L is a (not too big) positive power of q and corresponds roughly to the contribution of the diagonal term $(c, \ell, m, u) = (c', \ell', m', u')$ in Ω :

$$\mathbb{S}_{00} \ll \frac{q^{5/2}}{L^{1/2}C^{3/2}} \left(q^3 L^{1/2} \cdot \frac{1}{C} \cdot L \cdot qL \cdot q \right)^{1/2} \ll \frac{q^3}{L^{1/4}}.$$

This saving was the whole point of keeping the c-sum inside in the application of Cauchy–Schwarz. This also shows how the additional parameter ℓ plays a role in this argument: without introducing the ℓ -sum at the very beginning we would have been failed to beat the convexity bound $O(q^3)$ for the diagonal contribution. This trick is reminiscent of the amplification technique which seems to be used first by Munshi in [Mun21] (see also [HN18, Lin21, KLMS20, Sha19]).

For the case $n \neq 0$, we use again Proposition 4.5,

$$\sum_{v(q)} Z(v) \overline{Z'(v-\delta)} \ll q^{1/2}, \ \delta \neq 0 \,(\text{mod }q),$$

to obtain

$$\begin{split} \Omega_{\neq} &= \sum_{c \sim C} \frac{1}{c} \sum_{c' \sim C} \frac{1}{c'} \sum_{\ell \sim L} \overline{\lambda(1,\ell)} \sum_{\ell' \sim L} \lambda(1,\ell') \sum_{m \sim qL} \overline{\lambda(m)} \sum_{m' \sim qL} \lambda(m') \\ &\times q^3 L^{1/2} \sum_{0 \neq |n| < L^{1/2}} q^{1/2} \cdot \delta_{n+\ell \overline{m}c'-\ell' \overline{m'}c \equiv 0 \, (\text{mod } cc')} \\ &\ll \frac{q^{11/2} L^5}{C^2} + \frac{q^{9/2} L^4}{C}. \end{split}$$

Correspondingly, such terms contribute to S

$$S_{\neq} \ll q^{11/4} L^{3/4}.$$

Therefore we have the following estimate for the sum of interest

$$S \ll \frac{q^3}{L^{1/4}} + q^{11/4}L^{1/2} + q^{11/4}L^{3/4}.$$

By choosing L appropriately, say $L = q^{1/4}$ we have

$$\mathcal{S} \ll q^{3-1/16}.$$

beating the convexity bound $S = O(q^{3+o(1)})$.

Key to this argument is the bound for the correlation sums

$$\sum_{v} Z(v) \overline{Z'(v-\delta)} \ll \delta_{\delta=0} \delta_{\alpha/\alpha'\equiv\beta\gamma/(\beta'\gamma') \, (\mathrm{mod} \, q)} q + q^{1/2}$$

which follows from Proposition 4.5. This proposition is proven by interpreting the functions

$$v \mapsto Z(v), Z'(v)$$

as trace functions of ℓ -adic sheaves $\mathcal{Z}, \mathcal{Z}'$ on the affine line $\mathbf{A}^1_{\mathbf{F}_q}$ and by using methods from ℓ -adic cohomology. As the expression for Z suggests, the underlying sheaf \mathcal{Z} is the geometric convolution of the tensor product

$$\mathcal{K} = (\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2)$$
 with $\mathcal{L} = [y \to \beta \gamma / y]^* \mathcal{K} \ell_2$

(and likewise for \mathcal{Z}'). First we verify that if (MO) is satisfied, then \mathcal{K} and hence \mathcal{Z} are geometrically irreducible. It remains to show that if (SL) is also satisfied, \mathcal{Z} and $[-\delta]\mathcal{Z}'$ are not geometrically isomorphic if either $\delta \neq 0$, or $\alpha/\alpha' \neq \beta\gamma/(\beta'\gamma') \pmod{q}$. To do this, we analyse carefully the possible singularities of \mathcal{Z} (at most at ∞ , 0 and $\beta\gamma/\alpha$) and determine their nature (unipotent or not) at $0, \beta\gamma/\alpha$ using Deligne's semi-continuity theorem and the theory of local convolution due to Rojas-León. This last point is the most delicate part of the geometric argument.

This general approach is entirely different from the path taken by Sharma in the special case $K = \chi$ a multiplicative character [Sha19]: Sharma exploits the multiplicativity of χ and reduces the problem to checking the non-degeneracy criterion of Adolphson–Sperber for an explicit exponential sum in 9 variables. For this special case, a third route is possible: the sheaf \mathcal{Z} is an hypergeometric sheaf whose local and global properties were studied in depth by Katz in [Kat90]; the estimation of the correlation sums above can then be deduced from these properties and the general bounds for correlations sums of trace functions like [FKMS19].

Acknowledgements. We are greatful to V. Blomer for his comments and suggestions on the paper and to the anonymous referee for his/her careful reading, numerous helpful comments and for pointing out some errors and inaccuracies. A good amount of this work was carried out during the Second Symposium on Analytic Number Theory which took place in Cetraro in July 2019 and we would like to thank D. Bazzanella, S. Bettin and A. Perelli for their excellent organisation and their choice of the beautiful location. Finally we would like to thank heartily E. Kowalski for decisive discussions and numerous encouragements during the various phases of this project.

2. Background material regarding automorphic forms

In the sequel we will denote by V or W some smooth functions $V, W \in \mathcal{C}^{\infty}_{c}([1,2])$ satisfying for all $i \ge 0$

$$(2.1) V^{(i)}(x) \ll Z^i$$

for some parameter Z satisfying $1 \leq Z \leq q$.

2.1. Bounds for Hecke eigenvalues. We recall the Hecke relation for GL_3 Hecke eigenvalues implies that

$$\lambda(r,n) = \sum_{d \mid (r,n)} \mu(d) \lambda(r/d,1) \lambda(1,n/d)$$

and

(2.2)
$$\lambda(1,m)\lambda(r,n) = \sum_{\substack{d_0d_1d_2 = m \\ d_1|r,d_2|n}} \lambda(rd_2/d_1,nd_0/d_2)$$

and we have the following individual bounds

(2.3)
$$\lambda(r,n) \ll (rn)^{\theta_3 + o(1)}$$

where $\theta_3 = 5/14$ according to Kim–Sarnak (see [KS03]).

We also have the following well-known Rankin–Selberg estimates for GL₂ and GL₃ Hecke eigenvalues (see [Gol06, Sec. 12.1])

(2.4)
$$\sum_{1 \leqslant n \leqslant X} |\lambda_f(n)|^2 \ll_f X,$$

and

$$\sum_{1\leqslant m^2n\leqslant X}|\lambda(n,m)|^2\ll_{\varphi}X,$$

from which and the Hecke relation, one can also derive the bound (see [Mun21, Lemma 2])

(2.5)
$$\sum_{1 \leqslant m^2 n \leqslant X} |\lambda(n,m)|^2 m \ll_{\varphi,\varepsilon} X^{1+\varepsilon},$$

for any $\varepsilon > 0$.

2.2. Summation formulas. Let f be a GL_2 cuspform of level 1 and of weight k (if f is holomorphic) or of Laplace eigenvalue $1/4 + r_f^2$ (if f is Maass). The GL_2 Voronoi summation formula states:

Proposition 2.1. Let V be a test function as in (2.1). Let X > 0, $c, u \ge 1$ be integers such that (u,c) = 1. We have

$$\sum_{m=1}^{\infty} \lambda_f(m) e\left(\frac{um}{c}\right) V\left(\frac{m}{X}\right) = \frac{X}{c} \sum_{\substack{\pm \\ 8}} \sum_{m=1}^{\infty} \lambda_f(m) e\left(\mp \frac{\bar{u}m}{c}\right) \mathcal{V}^{\pm}\left(\frac{mX}{c^2}\right),$$

where

$$\mathcal{V}^{\pm}(y) = \int_0^\infty V(x) \mathcal{J}_f^{\pm}(4\pi\sqrt{xy}) \mathrm{d}x$$

with

$$\mathcal{J}_f^+(x) = 2\pi i^k J_{k-1}(x), \ \mathcal{J}_f^-(x) = 0 \ if \ f \ is \ holomorphic;$$

and

$$\mathcal{J}_f^+(x) = \frac{-\pi}{\sin \pi i r_f} (J_{2ir_f}(x) - J_{-2ir_f}(x)), \quad \mathcal{J}_f^-(x) = 4\varepsilon_f \cosh(\pi r_f) K_{2ir_f}(x) \quad if \ f \ is \ Maass.$$

We also recall the GL_3 Voronoi summation formula first proven by Miller and Schmid ([MS06]):

Proposition 2.2. Let W be a test function as in (2.1). Let X > 0, $c, u, r \ge 1$ be integers such that (u, c) = 1. We have

$$\sum_{n=1}^{\infty} \lambda(r,n) e\left(\frac{un}{c}\right) W\left(\frac{n}{N}\right) = c^{3/2} \sum_{\pm} \sum_{n_1 \mid rc} \sum_{n=1}^{\infty} \frac{\lambda(n,n_1)}{n_1 n} \frac{S\left(r\bar{u}, \pm n; \frac{rc}{n_1}\right)}{c^{1/2}} \mathcal{W}_{\pm}\left(\frac{nn_1^2}{c^3 r/N}\right),$$

where

$$S(m,n;c) = \sum_{x(c)}^{\star} e\left(\frac{mx+n\bar{x}}{c}\right)$$

is the Kloosterman sum and where

(2.6)
$$\mathcal{W}_{\sigma}(x) = \frac{1}{2\pi i} \int_{(1)} x^{-s} \mathcal{G}^{\sigma}(s+1) \left(\int_{0}^{+\infty} W(y) y^{-s} \frac{dy}{y} \right) ds$$

For the precise definition of the kernel function $\mathcal{G}^{\sigma}(s)$ we refer the reader to [BK19, (4.6)]. We recall the following lemma (see [KLMS20, Lemma 4.3])

Lemma 2.3. Let $\sigma \in \{-1, 1\}$. For any $j \ge 0$, any $A \ge 1$ and any $\varepsilon > 0$, we have

$$x^{j}\mathcal{W}_{\sigma}^{(j)}(x) \ll \min\left(Z^{j+1}x^{1-\theta_{3}-\varepsilon}, Z^{j+5/2+\varepsilon}\left(\frac{Z^{3}}{x}\right)^{A}\right)$$

for x > 0, where the implied constant depends on (j, A, ε) . Moreover, for $x \ge 1$, we have

$$x^{j} \mathcal{W}_{\sigma}^{(j)}(x) \ll x^{2/3} \min(Z^{j}, x^{j/3})$$

where the implied constant depends on j.

The integral transform W_{\pm} admits another representation as Hankel transform of some GL₃-Bessel functions as can be seen by switching the y and s integrals in (2.6) which we record here (see [Lin21, Lemma 3.2])

(2.7)
$$\mathcal{W}_{\pm}(x) = x \int_0^\infty W(y) J_{\varphi,\pm}(xy) dy.$$

Here $J_{\varphi,\pm}(x)$ is some GL₃-Bessel function satisfying the following properties.

Lemma 2.4. (1). Let $\rho > \max\{-\Re \mathfrak{e} \,\mu_{\varphi,1}, -\Re \mathfrak{e} \,\mu_{\varphi,2}, -\Re \mathfrak{e} \,\mu_{\varphi,3}\}$. For $x \ll 1$, we have

(2.8)
$$x^{j} J_{\varphi,\pm}^{(j)}(x) \ll_{\mu_{\varphi,1},\mu_{\varphi,2},\mu_{\varphi,3},\varrho,j} x^{-\varrho}.$$

(2). Let $K \ge 0$ be a fixed nonnegative integer. For x > 0, we may write

(2.9)
$$J_{\varphi,\pm}(x^3) = \frac{e(\pm 3x)}{x} W_{\varphi}^{\pm}(x) + E_{\varphi}^{\pm}(x),$$

where $W_{\varphi}^{\pm}(x)$ and $E_{\varphi}^{\pm}(x)$ are real-analytic functions on $(0,\infty)$ satisfying

$$W_{\varphi}^{\pm}(x) = \sum_{m=0}^{K-1} B_{m}^{\pm}(\varphi) x^{-m} + O_{K,\mu_{\varphi,1},\mu_{\varphi,2},\mu_{\varphi,3}} \left(x^{-K} \right),$$

and

$$E_{\varphi}^{\pm,(j)}(x) \ll_{\mu_{\varphi,1},\mu_{\varphi,2},\mu_{\varphi,3},j} \frac{\exp(-3\sqrt{3}\pi x)}{x},$$

for $x \gg_{\mu_{\varphi,1},\mu_{\varphi,2},\mu_{\varphi,3}} 1$, where $B_m^{\pm}(\varphi)$ are constants depending on $\mu_{\varphi,1}, \mu_{\varphi,2}$ and $\mu_{\varphi,3}$. Proof. See [Qi20, Theorem 14.1].

Next we recall the Duke–Friedlander–Iwaniec delta symbol method [DF193] in a version given by Heath-Brown [HB96].

Lemma 2.5. For any C > 1, there is a positive constant η_C and a smooth function h(x, y) defined on $(0,\infty) \times (-\infty,\infty)$ such that

$$\delta_{n=0} = \frac{\eta_C}{C} \sum_{c=1}^{\infty} \frac{1}{c} \sum_{a(c)}^{\star} e\left(\frac{an}{c}\right) h\left(\frac{c}{C}, \frac{n}{C^2}\right).$$

Here the constant η_C satisfies

$$\eta_C = 1 + O_A(C^{-A})$$

for any A > 0, and h(x,y) is a smooth function vanishing unless $x \leq \max(1,2|y|)$ and whose derivatives satisfy

$$x^i \frac{\partial^i}{\partial^i x} h(x,y) \ll_i 1 \text{ and } \frac{\partial}{\partial y} h(x,y) = 0,$$

for $x \leq 1$ and $|y| \leq x/2$, and

(2.10)
$$x^{i}y^{j}\frac{\partial^{i+j}}{\partial^{i}x\partial^{j}y}h(x,y) \ll_{i,j} 1$$

for |y| > x/2.

3. First transformations

From now on we will assume that Z satisfies

$$1 \leqslant Z \leqslant q$$

(otherwise the trivial bound is stronger that the bound claimed in Theorem 1.1). We have

$$S_{V}^{t}(K,X) := \sum_{r,n} \lambda(r,n)\lambda_{f}(n)K(nr^{2})V(\frac{nr^{2}}{X}) = \sum_{r\geq 1} S_{V,r}(K,X/r^{2}).$$

where

(3.1)
$$S_{V,r}(K,X) := \sum_{n=1}^{\infty} \lambda(r,n)\lambda_f(n)K(nr^2)V(\frac{n}{X}).$$

By applying the Rankin–Selberg estimates (2.4), (2.5) and the bound (2.3), we have

(3.2)
$$\sum_{r \ge R} S_{V,r}(K, X/r^2) \ll q^{o(1)} R^{\theta_3} X/R$$

where $\theta_3 = 5/14$; see, say [LS21, (4.2)] for the details. It will suffice to bound $S_{V,r}(K,X)$ for $r \leq R = q^{\varrho}$ for some $\varrho > 0$ to be chosen; see (7.7). In the sequel we assume that

$$r \leqslant R;$$

moreover, as we will see, R < q, so that (r, q) = 1.

Let $L \ge 1$ be some parameter and let L be the set of primes in the interval [L, 2L[. We first write

$$S_{V,r}(K,X) = \frac{1}{L^{\star}} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \sum_{n} \lambda(1,\ell) \lambda(r,n) \lambda_f(n) K(nr^2) V(\frac{n}{X})$$

where

$$L^{\star} := \sum_{\ell \in \mathcal{L}} |\lambda(1,\ell)|^2 \simeq_{\varphi} \frac{L}{\log L}$$

by the prime number theorem for automorphic forms ([LWY05, Lemma 5.1]).

In the sequel we will assume that L satisfies

$$(3.3) R < L$$

so that it is guaranteed that any $\ell \in [L, 2L]$ is coprime with r.

By the Hecke relation (2.2) we have (since $(\ell, r) = 1$ by the assumption L > R)

$$\lambda(1,\ell)\lambda(r,n) = \lambda(r,\ell n) + \delta_{\ell|n}\lambda(r\ell,n/\ell).$$

The contribution to $S_{V,r}(K, X)$ of the second term is trivially bounded by

$$(3.4) \qquad \frac{1}{L^{\star}} \sum_{\ell \in \mathcal{L}} |\lambda(1,\ell)|^2 \sum_{n \sim X/\ell} |\lambda(r,n)| \left(|\lambda_f(n)\lambda_f(\ell)| + \delta_{\ell|n} |\lambda_f(n/\ell)| \right) \|K\|_{\infty} \leqslant q^{o(1)} \frac{r^{\theta_3} L^{\theta_2} X}{L}$$

where $\theta_3 = 5/14$ and $\theta_2 = 7/64$. Therefore we obtain

$$S_{V,r}(K,X) = \frac{1}{L^{\star}} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \sum_{m=1}^{\infty} \lambda(r,m\ell) \lambda_f(m) K(mr^2) V\left(\frac{m}{X}\right) + O\left(q^{o(1)} \frac{r^{\theta_3} L^{\theta_2} X}{L}\right)$$

Now we use the delta method to separate the coefficients $\lambda(r, m\ell)$ and $\lambda_f(m)K(mr^2)$. To prepare for later manipulations, we first introduce U, a smooth function supported in (1/100, 100) and satisfying $U^{(i)}(x) \ll 1$ for $i \ge 0$ and U(x) = 1 for $x \in [1, 2]$.

Let $v = q^{\varepsilon} Z$ where $\varepsilon > 0$ is fixed but to be chosen as small as we need. Then we can rewrite $S_{V,r}(K, X)$ as

$$S_{V,r}(K,X) = \frac{1}{L^{\star}} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \sum_{n=1}^{\infty} \lambda(r,n) \sum_{m=1}^{\infty} \lambda_f(m) K(mr^2) \left(\frac{n}{m\ell}\right)^{iv} \delta_{n=m\ell} U\left(\frac{n}{X\ell}\right) V\left(\frac{m}{X}\right) + O\left(q^{o(1)} \frac{r^{\theta_3} L^{\theta_2} X}{L}\right).$$

Using a Mellin transform on U we can replace – up to a factor $q^{o(1)}$ and up to changing the definition of U– the expression $U\left(\frac{n}{X\ell}\right)$ by $U\left(\frac{n}{XL}\right)$.

Therefore from now on we will consider the sum

$$(3.5) \quad S'_{V,r}(K,X) = \frac{1}{L^{\star}} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \sum_{n=1}^{\infty} \lambda(r,n) \sum_{m=1}^{\infty} \lambda_f(m) K(mr^2) \left(\frac{n}{m\ell}\right)^{iv} \delta_{n=m\ell} U\left(\frac{n}{XL}\right) V\left(\frac{m}{X}\right).$$

Remark 3.1. The introduction of the parameter $v = q^{\varepsilon}Z$ will be useful later, when we try to localize the range of some variables; see Lemma 3.3 and (4.5). See also [AHLS20] (and Remark 4.2 there) for a similar trick.

We proceed as in [Sha19] and follow Holowinsky, Munshi and Qi [HMQ16] to rewrite the $\delta_{n-\ell m}$ in a more analytic form using the delta symbol method in Lemma 2.5: for any $n \in \mathbb{Z}$ such that $|n| \leq 4XL$ we have

$$\begin{split} \delta_{n=0} &= \delta_{q|n} \delta_{n/q=0} &= \frac{1}{q} \sum_{u(q)} e(\frac{un}{q}) \frac{1}{C} \sum_{c \leqslant 2C} \frac{1}{c} \sum_{a(c)}^{\star} e\left(\frac{an/q}{c}\right) h\left(\frac{c}{C}, \frac{n}{C^2 q}\right) + O_A(C^{-A}) \\ &= \frac{1}{C} \sum_{c \leqslant 2C} \frac{1}{cq} \sum_{u(q), a(c)}^{\star} e\left(n\frac{a+uc}{cq}\right) h\left(\frac{c}{C}, \frac{n}{C^2 q}\right) + O_A(C^{-A}) \\ &= \frac{1}{C} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{cq} \sum_{u(q), a(c)}^{\star} e\left(n\frac{aq+uc}{cq}\right) h\left(\frac{c}{C}, \frac{n}{C^2 q}\right) \\ &+ \frac{1}{C} \sum_{c \leqslant 2C/q} \frac{1}{cq^2} \sum_{u(q), a(cq)}^{\star} e\left(n\frac{a+ucq}{cq^2}\right) h\left(\frac{cq}{C}, \frac{n}{C^2 q}\right) + O_A(C^{-A}). \end{split}$$

Here we choose

$$(3.6) 2C = \left(\frac{XL}{q}\right)^{\frac{1}{2}}.$$

Observe that in the first sum of the last expression, as a varies over a set of representatives of the residue classes modulo c (prime to c) and u varies over a set of representatives of the residue classes modulo q, aq + uc varies over a set of representatives of the residue classes modulo cq prime to c.

Similarly in the second sum, the modulus c is $\leq 2(XL/q^3)^{1/2} < q$ and is therefore coprime with q. It follows that as a varies over a set of representatives of the residue classes modulo cq (prime to cq) and u varies over a set of representatives of the residue classes modulo q, a + ucq varies over a set of representatives of the residue classes modulo q.

We can therefore rewrite

$$\delta_{n=0} = \frac{1}{C} \sum_{\substack{c \leq 2C \\ (c,q)=1}} \frac{1}{cq} \sum_{u(cq)}^{\star} e\left(n\frac{u}{cq}\right) h\left(\frac{c}{C}, \frac{n}{C^2q}\right)$$

$$+ \frac{1}{C} \sum_{\substack{c \leq 2C \\ (c,q)=1}} \frac{1}{cq} \sum_{a(c)}^{\star} e\left(n\frac{a}{c}\right) h\left(\frac{c}{C}, \frac{n}{C^2q}\right)$$

$$+ \frac{1}{C} \sum_{c \leq 2C/q} \frac{1}{cq^2} \sum_{u(cq^2)}^{\star} e\left(n\frac{u}{cq^2}\right) h\left(\frac{cq}{C}, \frac{n}{C^2q}\right) + O_A(C^{-A}).$$

Remark 3.2. One reason for detecting the condition $n - m\ell$ in two such steps is that since we already have a trace function of period q, introducing an additive character modulo q does not significantly increase the complexity. Therefore, we can detect the condition $n = m\ell \pmod{q}$ at a reduced cost. After this, we can then detect $(n - m\ell)/q = 0$ using the delta symbol. So we are choosing a delta symbol that is in harmony with the modulus of our trace function (we thank the referee for pointing this out).

The outcome of such an operation is that the parameter C in (3.6) is reduced by a factor $q^{1/2}$ from the most natural choice $(XL)^{1/2}$. A similar reduction trick was used by Munshi in [Mun15], and more recently in [Mun22].

We apply (3.7) to the difference $n - m\ell$ in (3.5) and obtain

(3.8)
$$S'_{V,r}(K,X) = \text{Main} + \text{Err}_1 + \text{Err}_2 + O_A(X^{-A})$$

where

$$(3.9) \qquad \text{Main} = \frac{1}{L^{\star}Cq} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\ell \in \mathbf{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{u(cq)}^{\star} \sum_{n=1}^{\infty} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\sum_{m=1}^{\infty} \lambda_f(m) K(mr^2) e\left(\frac{-um\ell}{cq}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n-m\ell}{C^2q}\right),$$
$$\text{Err}_1 = \frac{1}{L^{\star}Cq} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\ell \in \mathbf{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{a(c)}^{\star} \sum_{n=1}^{\infty} \lambda(r,n) e\left(\frac{an}{c}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\sum_{m=1}^{\infty} \lambda_f(m) K(mr^2) e\left(\frac{-am\ell}{c}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n-m\ell}{C^2q}\right),$$

and

$$(3.11) \qquad \operatorname{Err}_{2} = \frac{1}{L^{\star}Cq^{2}} \sum_{\substack{c \leqslant 2C/q \\ (c,q)=1}} \frac{1}{c} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{a(cq^{2})} \sum_{n=1}^{\star} \lambda(r,n) e\left(\frac{an}{cq^{2}}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\sum_{m=1}^{\infty} \lambda_{f}(m) K(mr^{2}) e\left(\frac{-am\ell}{cq^{2}}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{cq}{C}, \frac{n-m\ell}{C^{2}q}\right).$$

As for the first term Main, we may further restrict to the subsum satisfying $(c, \ell) = 1$ and will bound the complementary subsum (corresponding to $\ell|c$) as an error term. That is, we further write

$$Main = Main_0 + Err_3,$$

where

(3.12)
$$\operatorname{Main}_{0} = \frac{1}{L^{\star}Cq} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\substack{\ell \in L \\ (\ell,c)=1}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{u(cq)}^{\star} \sum_{n=1}^{\infty} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\sum_{m=1}^{\infty} \lambda_{f}(m) K(mr^{2}) e\left(\frac{-um\ell}{cq}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n-m\ell}{C^{2}q}\right),$$

and

(3.13)
$$\operatorname{Err}_{3} = \frac{1}{L^{*}Cq} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{\substack{c \leqslant 2C \\ \ell \mid c, (c,q) = 1}} \frac{1}{c} \sum_{u(cq)}^{*} \sum_{n=1}^{\infty} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\sum_{m=1}^{\infty} \lambda_{f}(m) K(mr^{2}) e\left(\frac{-um\ell}{cq}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n-m\ell}{C^{2}q}\right).$$

In the sequel, we focus our analysis on the term $Main_0$ which is the hardest and is responsible for the final bound. The other three terms Err_1 , Err_2 and Err_3 are discussed briefly in §7.1.

3.1. Bounding Main₀. To prepare for the application of Voronoi summation formula to the *m*-sum we write

$$K(mr^2) = \frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) e(\frac{-bmr^2}{q}) = \frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) e(\frac{-bcr^2m}{cq}).$$

where

(3.14)
$$\widehat{K}(b) = \frac{1}{q^{1/2}} \sum_{a \in \mathbf{F}_q} K(a) e(\frac{ab}{q})$$

denote the normalized Fourier transform of K.

We find that the term in (3.12) can be rewritten as

$$(3.15) \qquad \operatorname{Main}_{0} = \frac{1}{L^{\star}Cq} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\substack{\ell \in \mathbf{L} \\ (\ell,c)=1}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{u(cq)} \sum_{n=1}^{\star} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) \sum_{m=1}^{\infty} \lambda_{f}(m) e\left(\frac{-(bcr^{2}+u\ell)m}{cq}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n-m\ell}{C^{2}q}\right).$$

We can further assume that $(bcr^2 + u\ell, cq) = 1$, as otherwise we would have $(bcr^2 + u\ell, q) = q$ (since we have already assumed $(c, \ell) = 1$ and then by applying Proposition 2.1 to such terms above, we have

$$\sum_{m=1}^{\infty} \lambda_f(m) e\left(\frac{-\frac{bcr^2 + u\ell}{q}m}{c}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n - m\ell}{C^2 q}\right)$$
$$= \frac{X}{c} \sum_{m=1}^{\infty} \overline{\lambda_f(m)} e\left(\frac{\pm \frac{\overline{bcr^2 + u\ell}}{q}m}{c}\right) \widehat{\mathcal{V}}^{\pm}\left(n, \frac{mX}{c^2}\right),$$

where

(3.16)
$$\widehat{\mathcal{V}}^{\pm}(n,y) = \int_{\mathbb{R}} V(x) (Xx)^{-iv} h\left(\frac{c}{C}, \frac{n - Xx\ell}{C^2 q}\right) \mathcal{J}_f^{\pm}(4\pi\sqrt{xy}) \mathrm{d}x.$$

By integration by parts, one easily sees that $\widehat{\mathcal{V}}^{\pm}(n,y) \ll_A \left(\frac{Z+v}{\sqrt{y}}\right)^A$, for any $A \ge 0$. On the other hand with our choice of the C in (3.6), we have

$$\frac{mX}{c^2} > \frac{mX}{C^2} = \frac{mq}{L} \gg q^{2\eta} Z^2$$

for some $\eta > 0$. Therefore $\widehat{\mathcal{V}}^{\pm}\left(n, \frac{mX}{c^2}\right) \ll (q^{-\eta})^A$, of arbitrarily small size. That is, the contribution from the terms with $(bcr^2 + u\ell, cq) > 1$ in (3.15) is bounded above by $O_A(q^{-A})$. Now assuming $(bcr^2 + u\ell, cq) = 1$ and applying Proposition 2.1 to the *m*-sum in (3.15), we have

(3.17)

$$\begin{aligned} \operatorname{Main}_{0} &= \frac{X}{L^{\star}Cq^{2}} \sum_{\pm} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c^{2}} \sum_{\substack{\ell \in L \\ (\ell,c)=1}} \overline{\lambda(1,\ell)} \ell^{-iv} \frac{1}{q^{1/2}} \sum_{\substack{b(q),u(cq) \\ (bcr^{2}+u\ell,cq)=1}} \widehat{K}(b) \\ &\times \sum_{m=1}^{\infty} \overline{\lambda_{f}(m)} e(\frac{\pm \overline{bcr^{2}+u\ell}m}{cq}) \sum_{n=1}^{\infty} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right) \widehat{\mathcal{V}}^{\pm}\left(n,\frac{mX}{c^{2}q^{2}}\right) + O(q^{-A}). \end{aligned}$$

We further apply Proposition 2.2 to the *n*-sum above to obtain a sum of the form

$$(3.18) \quad \frac{X}{L^{\star}Cq} \sum_{\pm \pm} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\substack{\ell \in L \\ (\ell,c)=1}} \overline{\lambda(1,\ell)} \ell^{-iv} \\ \times \sum_{n_1|rcq} \sum_{m,n} \overline{\lambda_f(m)} \frac{\lambda(n,n_1)}{nn_1} \mathcal{C}(m,n;\frac{rcq}{n_1}) \mathcal{W}_{\pm \pm} \left(\frac{m}{c^2q^2/X}, \frac{n_1^2n}{c^3q^3r/XL}\right)$$

where

$$\mathbb{C}(m,n;\frac{rcq}{n_1}) = \frac{1}{q^{1/2}} \sum_{\substack{b(q),u(cq)\\(bcr^2+u\ell,cq)=1}}^{\star} \widehat{K}(b)e\left(\frac{\pm \overline{bcr^2+u\ell}m}{cq}\right) S\left(r\overline{u},\pm n;\frac{rcq}{n_1}\right)$$

and

(3.19)
$$\mathcal{W}_{\pm\pm}(y,z) = \int_{\mathbb{R}} V(x)(Xx)^{-iv} \mathcal{W}_{x,\pm}(z) \mathcal{J}_{f}^{\pm}(4\pi\sqrt{xy}) \mathrm{d}x \\ = z \int_{\mathbb{R}} V(x)(Xx)^{-iv} \left(\int_{0}^{\infty} W_{x}(\xi) J_{\varphi,\pm}(z\xi) \mathrm{d}\xi\right) \mathcal{J}_{f}^{\pm}(4\pi\sqrt{xy}) \mathrm{d}x$$

by (2.7). Here

$$W_x(\xi) := (XL\xi)^{iv} U(\xi) h\left(\frac{c}{C}, \frac{XL\xi - Xx\ell}{C^2 q}\right)$$

Lemma 3.3. Let $v = q^{\varepsilon}Z$, $\varepsilon > 0$ be the parameter introduced above Remark 3.1. For any $B \ge 1$ the function

$$(y,z)\mapsto \mathcal{W}_{\pm\pm}(y,z)$$

is negligible for $y, z \ge q^{-B}$ unless

$$(3.20) y \asymp v^2, \ z \asymp v^3.$$

Here negligible means that for any $A \ge 1$, and for $y, z \ge q^{-B}$ not satisfying (3.20), one has

$$\mathcal{W}_{\pm\pm}(y,z) \ll_{A,B,\varepsilon} q^{-A}$$

Proof. We recall from (3.16) that

$$\widehat{\mathcal{V}}^{\pm}(n,y) = \int_{\mathbb{R}} V(x) (Xx)^{-iv} h\left(\frac{c}{C}, \frac{n - Xx\ell}{C^2 q}\right) \mathcal{J}_f^{\pm}(4\pi\sqrt{xy}) \mathrm{d}x$$

for n a real variable satisfying $n \asymp XL$ and that

$$\mathcal{W}_{\pm\pm}(y,z) = z \int_0^\infty (XL\xi)^{iv} U(\xi) \widehat{\mathcal{V}}^{\pm} (XL\xi,y) J_{\varphi,\pm}(z\xi) \mathrm{d}\xi$$

We consider the cases where f is holomorphic (the case f is a Maass cusp form would be similar). Then $\mathcal{J}_{f}^{-}(x) = 0$, and $\mathcal{J}_{f}^{+}(x) = J_{k-1}(x)$ satisfies

(3.21)
$$x^{i}J_{k-1}^{(i)}(x) \ll x^{k-1},$$

for $x \ll 1$, and while for $x \gg 1$ we can write

(3.22)
$$J_{k-1}(x) = \sum_{\pm} \frac{e^{\pm ix}}{\sqrt{x}} W_{k-1,\pm}(x),$$

for some $W_{k-1,\pm}(x)$ satisfying $x^i W_{k-1,\pm}^{(i)}(x) \ll 1$. We also have

$$x^{i} \frac{\partial^{i}}{\partial x^{i}} h\left(\frac{c}{C}, \frac{n - Xx\ell}{C^{2}q}\right) \ll 1,$$

which follows from (2.10). This together with (3.21) implies that if $y \ll 1$ then $x^i V_{1,y}^{(i)}(x) \ll Z^i$, where

$$V_{1,y}(x) := V(x)X^{-iv}h\left(\frac{c}{C}, \frac{n - Xx\ell}{C^2q}\right)J_{k-1}(4\pi\sqrt{xy}).$$

Therefore integration by parts implies that: if $y \ll 1$ we have

$$\widehat{\mathcal{V}}^+(n,y) = \int_{\mathbb{R}} V_{1,y}(x) x^{-iv} \mathrm{d}x \ll \left(\frac{Z}{v}\right)^A \ll \left(\frac{1}{q^{\varepsilon}}\right)^A,$$

for any $A \ge 0$. That is, for $y \ll 1$, $\widehat{\mathcal{V}}^+(n, y)$ is always negligibly small.

We now assume that $y \gg 1$. We use (3.22) to write

$$\begin{aligned} \widehat{\mathcal{V}}^+\left(n,y\right) &= \sum_{\pm} \int_{\mathbb{R}} X^{-iv} V(x) h\left(\frac{c}{C}, \frac{n - Xx\ell}{C^2 q}\right) \frac{1}{(4\pi\sqrt{xy})^{1/2}} W_{k-1,\pm}(4\pi\sqrt{xy}) e\left(-\frac{v\log x}{2\pi} \pm 2\sqrt{xy}\right) \mathrm{d}x\\ &:= \sum_{\pm} \frac{1}{y^{1/4}} \int_{\mathbb{R}} V_{2,y}(x) e\left(-\frac{v\log x}{2\pi} \pm 2\sqrt{xy}\right) \mathrm{d}x, \end{aligned}$$

where $V_{2,y}(x)$ satisfies $x^i V_{2,y}^{(i)}(x) \ll Z^i$. Again, integration by parts implies that

$$\widehat{\mathcal{V}}^+(n,y) \ll \frac{1}{y^{1/4}} \left(\frac{Z}{|-v/2\pi \pm \sqrt{y}|} \right)^A,$$

for any $A \ge 0$, which shows that $\widehat{\mathcal{V}}^+(n,y)$ is negligibly small unless $|-v/2\pi \pm \sqrt{y}| \le v$. Therefore we may assume that the y-variable satisfies

$$y \simeq v^2$$

(otherwise $\widehat{\mathcal{V}}^+(n, y)$ and $\mathcal{W}_{\pm\pm}(y, z)$ have negligible size for $n \asymp XL$ and $y, z \ge q^{-B}$.) In this remaining range we evaluate $\mathcal{W}_{\pm\pm}(y, z)$. Recall from (3.19), we have

(3.23)
$$\mathcal{W}_{\pm\pm}(y,z) = \int_{\mathbb{R}} V(x)(Xx)^{-iv} \mathcal{J}_{f}^{\pm}(4\pi\sqrt{xy}) \times z \int_{0}^{\infty} (XL\xi)^{iv} U(\xi) h\left(\frac{c}{C}, \frac{XL\xi - Xx\ell}{C^{2}q}\right) J_{\varphi,\pm}(z\xi) \mathrm{d}\xi \,\mathrm{d}x$$

We consider the inner integral above. Recall here $U(\xi)$ is a test function satisfying (2.1) with $Z_U = 1$. If $z \ll 1$, then by using (2.8) and (2.10) we have $\xi^i W_{1,z}^{(i)}(\xi) \ll z^{-\varrho}$, where

$$W_{1,z}(\xi) := (XL)^{iv} U(\xi) h\left(\frac{c}{C}, \frac{XL\xi - Xx\ell}{C^2 q}\right) J_{\varphi,\pm}(z\xi).$$

Hence using integration by parts, the inner integral satisfies

$$\int_{\mathbb{R}} W_{1,z}(\xi) \xi^{iv} \mathrm{d}\xi \ll z^{-\varrho} v^{-A} \ll q^{-A\varepsilon/2}$$

by taking A sufficiently large. This implies that $\mathcal{W}_{\pm\pm}(y,z)$ is negligibly small when $q^{-B} \leq z \ll 1$.

Now we assume that $z \gg 1$ and use (2.9) to rewrite the inner integral over ξ in (3.23) as

$$\sum_{\pm} \int_0^\infty (XL)^{iv} U(\xi) h\left(\frac{c}{C}, \frac{XL\xi - Xx\ell}{C^2 q}\right) \frac{W_{\varphi}^{\pm}((z\xi)^{1/3})}{(z\xi)^{1/3}} e\left(\frac{v\log\xi}{2\pi} \pm 3(z\xi)^{1/3}\right) \mathrm{d}\xi$$
$$:= \sum_{\pm} \frac{1}{z^{1/3}} \int_0^\infty W_{2,z}(\xi) e\left(\frac{v\log\xi}{2\pi} \pm 3(z\xi)^{1/3}\right) \mathrm{d}\xi$$

up to a negligible error term. Here

$$W_{2,z}(\xi) = (XL)^{iv} \xi^{-1/3} U(\xi) h\left(\frac{c}{C}, \frac{XL\xi - Xx\ell}{C^2 q}\right) W_{\varphi}^{\pm}((z\xi)^{1/3})$$

satisfies $\xi^i W_{2,z}^{(i)}(\xi) \ll 1$. Applying integration by parts, the integral above is negligibly small unless

$$z \simeq v^3$$
.

		٦

Remark 3.4. The above argument shows in fact that for $z \ge q^{-B}$ and y > 0 one has

(3.24)
$$\mathcal{W}_{\pm\pm}(y,z) = z^{2/3} \int_{\mathbb{R}} V(x) (Xx)^{-iv} \mathcal{J}_{f}^{\pm} (4\pi \sqrt{xy}) \times \int_{0}^{\infty} W_{2,z}(\xi) e\left(\frac{v \log \xi}{2\pi} - 3(z\xi)^{1/3}\right) d\xi dx + \text{``negligible error term''}.$$

where the main term is also negligible unless possibly, when (3.20) is satisfied.

4. The case $q \not| n_1$

For the sum in (3.18), we further split it into two subsums according to $(n_1, q) = 1$ or not, and write

$$Main_0 = Main_{00} + Err_4 + O(q^{-A}),$$

where

$$(4.1) \quad \operatorname{Main}_{00} = \frac{X}{L^{\star}Cq} \sum_{\pm \pm} \sum_{\substack{c \leq 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\substack{\ell \in \mathbf{L} \\ (\ell,c)=1}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{\substack{n_1 \mid rc \\ (n_1,q)=1}} \sum_{\substack{m,n \\ (n_1,q)=1}} \overline{\lambda_f(m)} \\ \times \frac{\lambda(n,n_1)}{nn_1} \mathcal{C}(m,n;\frac{rcq}{n_1}) \mathcal{W}_{\pm \pm} \left(\frac{m}{c^2q^2/X},\frac{n_1^2n}{c^3q^3r/XL}\right)$$

and Err_4 corresponds to the complementary sum where $q|n_1$. In §6.3 we briefly analyse the contribution from Err_4 (see (6.4) and (6.5)).

We have

$$e(\frac{\pm \overline{bcr^2 + u\ell}m}{cq}) = e(\frac{\pm \overline{bcr^2 + u\ell}\overline{c}m}{q})e(\frac{\pm \overline{u\ell}\overline{q}m}{c})$$
$$S(r\overline{u}, \pm n; \frac{rcq}{n_1}) = S(\overline{c}n_1\overline{u}, \pm \overline{rc}n_1n; q)S(\overline{q}r\overline{u}, \pm \overline{q}n; rc/n_1).$$

Therefore, the (b, u) sum in $q^{-1/2} \mathcal{C}(m, n; \frac{rcq}{n_1})$ splits into a product of two sums of respective moduli rc/n_1 and q.

The modulus rc sum is denoted by

(4.2)
$$M_{n_1,r}(m,n,\ell;rc) := \sum_{u(c)}^{\star} e(\frac{\pm \overline{u\ell}\overline{q}m}{c}) S(\overline{q}r\overline{u},\pm\overline{q}n;rc/n_1).$$

The modulus q sum is given by

$$\begin{split} N_{\overline{cr}}(m,n,\ell;q) &:= \frac{1}{q} \sum_{\substack{b(q),u(q)\\(bcr^2+u\ell,q)=1}}^{\star} \widehat{K}(b)e(\frac{\pm \overline{bcr^2+u\ell}\overline{c}m}{q})S(\overline{c}n_1\overline{u},\pm\overline{rc}n_1n;q) \\ &= \frac{1}{q^{1/2}} \sum_{\substack{b(q),u(q)\\(b+u\ell,q)=1}}^{\star} \widehat{K}(b)e(\frac{\pm \overline{c}^2\overline{r}^2\overline{b+u\ell}m}{q})\operatorname{Kl}_2(\pm \overline{c}^3\overline{r}^3n_1^2n\overline{u};q) = \sum_{u(q)}^{\star} L_{\pm \overline{c}^2\overline{r}^2m,\ell}(u;q)\operatorname{Kl}_2(\pm \overline{c}^3\overline{r}^3n_1^2n\overline{u};q) \end{split}$$

where

(4.3)
$$L_{\alpha,\beta}(u;q) := \frac{1}{q^{1/2}} \sum_{\substack{b(q)\\(b+\beta u,q)=1}} \widehat{K}(b)e\left(\frac{\alpha b + \beta u}{q}\right)$$

We will sometime suppress the parameters α and β , and abbreviate $L_{\alpha,\beta}(u;q)$ as L(u;q).

From these notations we find that the sum $Main_{00}$ in (4.1) becomes

(4.4)
$$\operatorname{Main}_{00} = \frac{X}{L^{*}Cq^{1/2}} \sum_{\pm \pm} \sum_{\substack{c \leq 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\substack{\ell \in \mathbf{L} \\ (\ell,c)=1}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{\substack{n_{1}|rc \\ (n_{1},q)=1}} \sum_{\substack{m,n \\ (n_{1},q)=1}} \overline{\lambda_{f}(m)} \frac{\lambda(n,n_{1})}{nn_{1}} \times M_{n_{1},r}(m,n,\ell;rc) N_{\overline{cr}}(m,n,\ell;q) \mathcal{W}_{\pm \pm}(\frac{m}{c^{2}q^{2}/X}, \frac{n_{1}^{2}n}{c^{3}q^{3}r/XL}).$$

We break the c-sum into $O(\log q)$ dyadic intervals and for $C' \leq 2C$ we evaluate the truncated version of $S_V(K)$ where $c \sim C'$. Here C' satisfies

$$X^{1/2-\eta}/q \leq C' \leq 2C = (XL/q)^{1/2}$$

We set

(4.5)
$$M = Z^2 \frac{{C'}^2 q^{2+2\varepsilon}}{X}, \ N = Z^3 \frac{{C'}^3 q^{3+3\varepsilon} r}{XL}.$$

By Lemma 3.3 we have

$$m \simeq M, nn_1^2 \simeq N$$

which we abbreviate by

$$m \approx M, \ nn_1^2 \approx N.$$

Remark 4.1. From this discussion we see that C' cannot be too small: we have $m/M \leq 1$ and since $m \geq 1$, then

$$C' \geqslant X^{1/2}/Zq^{1+\varepsilon},$$

and this also implies that N is not too small

$$N \geqslant \frac{X^{1/2}r}{L}.$$

4.1. Cauchy–Schwarz. We will now apply Cauchy–Schwarz inequality with the n, n_1 variables outside (to get rid of the factor $\lambda(n, n_1)$) but we need some preparation.

We factor $c = c_1 c_2$ with

$$c_1 \leq C', n_1 | rc_1, c_1 | (n_1 r)^{\infty}$$
 and $(c_2, n_1 r) = 1$.

Then we apply Cauchy–Schwarz and (2.5) to remove the GL₃ coefficients in (4.4). Using Lemma 3.3 the sum Main₀₀ is bounded by four terms (for the various choices of \pm, \pm) of the shape

$$\frac{q^{o(1)}X}{LCq^{1/2}}\frac{1}{C'N}A^{1/2}B^{1/2}$$

with

$$A = \sum_{nn_1^2 \approx N} \sum_{n=1}^{\infty} |\lambda(n, n_1)|^2 n_1 \ll q^{o(1)} N = q^{o(1)} Z^3 C'^3 q^3 r / XL$$

$$B = \sum_{\substack{c_1, nn_1^2 \approx N \\ (n_1, q) = 1}} n_1 \bigg| \sum_{\substack{\ell \in L \\ (\ell, c_1) = 1}} \overline{\lambda(1, \ell)} \ell^{-iv} \sum_{m \leqslant M} \overline{\lambda_f(m)} \\ \times \sum_{\substack{c_2 \sim C'/c_1 \\ (c_2, q\ell) = 1}} M_{n_1, r}(m, n, \ell; rc_1 c_2) N_{\overline{c_1 c_2 r}}(m, n, \ell; q) \mathcal{W}_{\pm \pm} (\frac{m}{c_1^2 c_2^2 q^2 / X}, \frac{n_1^2 n}{c_1^3 c_2^3 q^3 r / XL}) \bigg|^2 U\left(\frac{n}{N/n_1^2}\right).$$

Here U is a smooth function with compact support contained in $(0, \infty)$ satisfying (2.1) with $Z_U = 1$.

Hence we obtain that

After opening the square, the second factor B equals

$$(4.7) B = \sum_{c_1, n_1} \sum_{n_1} \sum_{\substack{\ell, \ell' \\ m, m'}} \sum_{c_2, c'_2} \sum_{c_2, c'_2} \times$$

$$\sum_{n \ge 1} M_{n_1,r}(m,n,\ell;rc_1c_2) \overline{M_{n_1,r}(m',n,\ell';rc_1c'_2)} N_{\overline{c_1c_2r}}(m,n,\ell;q) \overline{N_{\overline{c_1c'_2r}}(m',n,\ell';q)} \mathcal{W}\left(\frac{n}{N/n_1^2}\right),$$

where

$$(4.8) \quad \mathcal{W}\left(\frac{n}{N/n_1^2}\right) = U\left(\frac{n}{N/n_1^2}\right) \mathcal{W}_{\pm\pm}\left(\frac{m}{c_1^2 c_2^2 q^2/X}, \frac{n_1^2 n}{c_1^3 c_2^3 q^3 r/XL}\right) \overline{\mathcal{W}_{\pm\pm}}\left(\frac{m'}{c_1^2 c_2'^2 q^2/X}, \frac{n_1^2 n}{c_1^3 c_2'^3 q^3 r/XL}\right).$$

Remark 4.2. At this point it may be useful to recall the typical size of the various quantities involved: we should imagine that $c_1 = n_1 = r = 1$ and

$$X = q^3, \ L = q^{\eta}$$
 with $\eta > 0$ as small as need be, $C = qL^{1/2}$
 $q^{1/2} \leq C' \leq qL^{1/2}.$
 $M = C'^2 q^{-1} \in [1, qL], \ N = C'^3 q^3 / q^3 L = C'^3 / L \in [q^{3/2} / L, q^3 L^{1/2}]$

and

$$c_2 c'_2 q = C'^2 q \in [q^2, q^3 L^{1/2}], \ c_2 c'_2 \approx Q/q \in [q, q^2 L^{1/2}].$$

In particular

$$N = C'^3 / L \ge (c_2 c'_2 q)^{1/2} = C' q^{1/2};$$

therefore we are in the Polya-Vinogradov range where applying the Poisson summation formula to (4.7) is beneficial.

We apply Poisson formula to the n-variable keeping in mind that

$$(4.9) n \mapsto M_{n_1,r}(m,n,\ell;rc_1c_2)\overline{M_{n_1,r}(m',n,\ell';rc_1c'_2)}N_{\overline{c_1c_2r}}(m,n,\ell;q)\overline{N_{\overline{c_1c'_2r}}(m',n,\ell';q)}$$

is periodic of period $qk := qrc_1c_2c'_2/n_1$, and see that (4.7) equals

(4.10)
$$\sum_{c_1,n_1} \sum_{n_1} \sum_{\substack{\ell,\ell' \\ m,m'}} \sum_{c_2,c_2'} \sum_{n_1^2(qk)^{1/2}} \sum_{n \in \mathbf{Z}} \operatorname{FT}(n,m,m',\ell,\ell';qk) \widehat{\mathcal{W}}(n/N^*),$$

where

$$\operatorname{FT}(n,m,m',\ell,\ell';qk) = \sum_{u,u' \pmod{q}} \sum_{L_{\pm \overline{c}^2 \overline{r}^2 m,\ell}(u;q) \overline{L_{\pm \overline{c'}^2 \overline{r}^2 m',\ell'}(u';q)} \times \frac{1}{\sqrt{qk}} \sum_{v \pmod{qk}} \operatorname{Kl}_2(\pm \overline{c}^3 \overline{r}^3 n_1^2 v \overline{u};q) \operatorname{Kl}_2(\pm \overline{c'}^3 \overline{r}^3 n_1^2 v \overline{u'};q) M_{n_1,r}(m,v,\ell;rc_1c_2) \overline{M_{n_1,r}(m',v,\ell';rc_1c'_2)} e\left(\frac{nv}{qk}\right)$$

and

(4.11)
$$N^* := qkn_1^2/N \asymp \frac{n_1}{c_1} \frac{qrC'^2}{N} (\asymp \frac{n_1XL}{Z^3c_1C'q^{2+3\varepsilon}}).$$

Here we can truncate the dual *n*-sum in (4.10) at $|n| \ll q^{2\varepsilon} ZN^*$.

Remark 4.3. The truncation of the dual *n*-sum follows from the fact that $\mathcal{W}(n/N^*)$ is negligibly small when $|n| \gg q^{2\varepsilon} ZN^*$, as can be seen from the expression in (3.24). Indeed, from (4.8) the Fourier transform $\widehat{\mathcal{W}}(w)$ after Poisson summation equals

(4.12)
$$\widehat{\mathcal{W}}(w) = \int_{\mathbb{R}} U(y) \, \mathcal{W}_{\pm\pm}(-, \frac{yN}{c_1^3 c_2^3 q^3 r/XL}) \overline{\mathcal{W}_{\pm\pm}}(-, \frac{yN}{c_1^3 c_2'^3 q^3 r/XL}) e(-wy) \mathrm{d}y.$$

From (3.24), we see that the factor in the integral that depends on y is of the form

$$\int_{\mathbb{R}} U(y) e\left(-3\left(\frac{yNXL}{c_1^3 c_2^3 q^3 r}\xi_1\right)^{1/3} + 3\left(\frac{yNXL}{c_1^3 {c_2'}^3 q^3 r}\xi_2\right)^{1/3} - wy\right) \mathrm{d}y$$

where $\xi_1, \xi_2 \simeq 1$, and for $w \neq 0$ by applying integration by parts, this is

$$\ll \left(\frac{q^{\varepsilon}Z}{|w|}\right)^A$$

upon noting that $\frac{NXL}{C'^3q^3r} \approx q^{3\varepsilon}Z^3$ by (4.5). Hence $\widehat{\mathcal{W}}(w)$ is negligibly small when $|w| \gg q^{2\varepsilon}Z$.

4.2. Computation of FT. Recall $k = rc_1c_2c'_2/n_1$. We have (q, k) = 1 and we split the above sum $FT(n, m, m', \ell, \ell'; qk)$ as a product of sums FT(n; q) and FT(n; k) of respective moduli q and k (to simplify notations we do not display the dependency in m, m', ℓ, ℓ' in these expressions).

4.2.1. The k-sum. The k-sum equals

$$\operatorname{FT}(n;k) := \frac{1}{\sqrt{k}} \sum_{v(k)} M_{n_1,r}(m,v,\ell;rc_1c_2) \overline{M_{n_1,r}(m',v,\ell';rc_1c_2')} e\left(\frac{nv\bar{q}}{k}\right)$$

Lemma 4.4. We have the following bounds

$$FT(0;k) \ll \sqrt{k}rc_1c_2 \sum_{\substack{d|c_1c_2 \\ (\ell'd,\ell d')|(m\ell'-m'\ell)}} (d,d'),$$

and

$$FT(n;k) \ll \sqrt{k} \sum_{d_1|c_1} d_1 \sum_{d'_1|c_1} d'_1 \sum_{\substack{x_1(rc_1/n_1)\\\ell n_1 x_1 \equiv \mp m \pmod{d_1}}}^{\star} \sum_{\substack{d_2|(c_2,\ell n_1c'_2+nm)\\d_2|(c'_2,\ell' n_1c_2+nm')}} d_2 d'_2.$$

Proof. To see this, we recall

(4.13)
$$M_{n_1,r}(m,n,\ell;rc) = \sum_{d|c} d\mu\left(\frac{c}{d}\right) \sum_{\substack{x(rc/n_1)\\\ell n_1 x \equiv \mp m \pmod{d}}}^{\star} e\left(\frac{\pm \bar{q}n\bar{x}}{rc/n_1}\right).$$

Then

(4.14)
$$\begin{aligned} \operatorname{FT}(n;k) &= \frac{1}{\sqrt{k}} \sum_{d \mid c_1 c_2} d\mu \left(\frac{c_1 c_2}{d} \right) \sum_{d' \mid c_1 c'_2} d' \mu \left(\frac{c_1 c'_2}{d'} \right) \times \\ & \sum_{\substack{x(rc_1 c_2/n_1) \\ \ell n_1 x \equiv \mp m \pmod{d} \ell' n_1 x' \equiv \mp m' \pmod{d'}}^{\star} \sum_{\substack{v(k) \\ e \neq \overline{x}(rc_1 c_2/n_1) \\ \ell n_1 x \equiv \mp m \pmod{d} \ell' n_1 x' \equiv \mp m' \pmod{d'}}^{\star} e \left(\frac{(\pm \bar{q} \bar{x} c'_2 \mp \bar{q} \overline{x'} c_2 + n \bar{q}) v}{k} \right) \\ &= \sqrt{k} \sum_{d \mid c_1 c_2} d\mu \left(\frac{c_1 c_2}{d} \right) \sum_{d' \mid c_1 c'_2} d' \mu \left(\frac{c_1 c'_2}{d'} \right) \sum_{\substack{x(rc_1 c_2/n_1) \\ \ell n_1 x \equiv \mp m \pmod{d} \ell' n_1 x' \equiv \mp m' \pmod{d'}}^{\star} \frac{1}{\bar{x} c'_2 - \bar{x'} c_2 = \mp n \pmod{d'}} \\ &= \sqrt{k} \sum_{d \mid c_1 c_2} d\mu \left(\frac{c_1 c_2}{d} \right) \sum_{d' \mid c_1 c'_2} d' \mu \left(\frac{c_1 c'_2}{d'} \right) \sum_{\substack{x(rc_1 c_2/n_1) \\ \ell n_1 x \equiv \mp m \pmod{d} \ell' n_1 x' \equiv \mp m' \pmod{d'}}}^{\star} \frac{1}{\bar{x} c'_2 - \bar{x'} c_2 \equiv \mp n \pmod{d'}} \end{aligned}$$

We calculate the case where $n \equiv 0 \pmod{k}$ first. For $n \equiv 0 \pmod{k}$, the congruence $\overline{x}c'_2 - \overline{x'}c_2 \equiv \mp n \pmod{k}$ forces

(4.15)
$$c_2 = c'_2$$

and $x \equiv x' \pmod{rc_1c_2/n_1}$. Therefore

$$FT(0;k) = \sqrt{k} \sum_{d|c_1c_2} d\mu \left(\frac{c_1c_2}{d}\right) \sum_{d'|c_1c_2} d'\mu \left(\frac{c_1c_2}{d'}\right) \sum_{\substack{x(rc_1c_2/n_1)\\\ell n_1x \equiv \mp m \pmod{d}\\\ell' n_1x \equiv \mp m' \pmod{d'}} 1.$$

Notice that since we have $(\ell, c) = 1$ (see (3.12)), for d|c we have $(d, \ell) = 1$. The system of equations $\ell n_1 x \equiv \mp m \pmod{d}$ and $\ell' n_1 x \equiv \mp m' \pmod{d'}$ has a unique solution modulo $\left[\frac{d}{(n_1,d)}, \frac{d'}{(n_1,d')}\right]$; more-over it implies that $(\ell'd, \ell d')|(m\ell' - m'\ell)$. The number of solutions for $x \pmod{\frac{rc_1c_2}{n_1}}$ is therefore given by

$$\frac{rc_1c_2/n_1}{[\frac{d}{(n_1,d)},\frac{d'}{(n_1,d')}]} = \frac{rc_1c_2}{[n_1,d,d']}.$$

Hence

$$\mathrm{FT}(0;k) = \sqrt{k} \sum_{\substack{d \mid c_1 c_2 \\ (\ell' d, \ell d') \mid (m\ell' - m'\ell)}} d\mu \left(\frac{c_1 c_2}{d}\right) d' \mu \left(\frac{c_1 c_2}{d'}\right) \frac{r c_1 c_2}{[n_1, d, d']}.$$

In particular,

(4.16)
$$\operatorname{FT}(0;k) \ll \sqrt{k} r c_1 c_2 \sum_{\substack{d \mid c_1 c_2 \\ (\ell' d, \ell d') \mid (m\ell' - m'\ell) \\ 21}} \sum_{\substack{d \mid c_1 c_2 \\ (\ell' d, \ell d') \mid (m\ell' - m'\ell) \\ 21}} (d, d').$$

Next we consider the case where $n \neq 0$ in FT(n; k). In (4.14) we write $d = d_1d_2$ where $d_1|c_1$, $d_2|c_2$ and $d' = d'_1d'_2$ where $d'_1|c_1, d'_2|c'_2$. Then the sum splits as

$$(4.17) FT(n;k) = \sqrt{k} \sum_{d_1|c_1} d_1 \mu\left(\frac{c_1}{d_1}\right) \sum_{d_1'|c_1} d_1' \mu\left(\frac{c_1}{d_1'}\right) \sum_{d_2|c_2} d_2 \mu\left(\frac{c_2}{d_2}\right) \sum_{d_2'|c_2'} d_2' \mu\left(\frac{c_2'}{d_2'}\right) \\ \sum_{\substack{x_1(rc_1/n_1)\\\ell n_1 x_1 \equiv \mp m \pmod{d_1}\ell' n_1 x_1' \equiv \mp m' \pmod{d_1'}} \sum_{\substack{x_1'(rc_1/n_1)\\\ell n_1 x_2 \equiv \mp m \pmod{d_2}\ell' n_1 x_2' \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2(c_2)\\k_1 x_2' \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2'(c_2')\\k_1 x_2' \equiv \mp m' \pmod{d_1'}} \sum_{\substack{x_1(rc_1/n_1)\\\ell n_1 x_2 \equiv \mp m \pmod{d_2'}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \pmod{d_2'}} \sum_{\substack{x_2'(c_2)\\k_2 \equiv \mp m' \binom{x_2'(c_2)\\k_2 \equiv \mp m' \binom{x_2'(c_2)\\k_2 \equiv m' \binom{x_2'(c_2)\\k_2 \equiv \mp m' \binom{x_2'(c_2)\\k_2 \equiv \mp m' \binom{x_2'(c_2)\\k_2 \equiv m' \binom{x_2'(c_2)}$$

where

$$FT_{1}(n;k) = \sum_{d_{1}|c_{1}} d_{1}\mu\left(\frac{c_{1}}{d_{1}}\right) \sum_{d_{1}'|c_{1}} d_{1}'\mu\left(\frac{c_{1}}{d_{1}'}\right) \sum_{\substack{x_{1}(rc_{1}/n_{1})\\\ell n_{1}x_{1} \equiv \mp m \pmod{d_{1}}\ell' n_{1}x_{1}' \equiv \mp m' \pmod{d_{1}}}_{\substack{x_{1}(rc_{1}/n_{1})\\\ell n_{1}x_{1}' \equiv \mp m' \pmod{d_{1}}} 1,$$

and

$$\operatorname{FT}_{2}(n;k) = \sum_{d_{2}|c_{2}} d_{1}\mu\left(\frac{c_{2}}{d_{2}}\right) \sum_{d_{2}'|c_{2}'} d_{2}'\mu\left(\frac{c_{2}'}{d_{2}'}\right) \sum_{\substack{x_{2}(c_{2})\\\ell n_{1}x_{2} \equiv \mp m \pmod{d_{2}\ell'} n_{1}x_{2}' \equiv \mp m' \pmod{d_{2}}}_{\substack{x_{2}'(c_{2}')\\ \overline{x_{2}c_{2}'} - \overline{x_{2}'}c_{2} \equiv \mp n \pmod{d_{2}c_{2}}} 1.$$

In FT₁(n; k), the term $x'_1 \pmod{rc_1/n_1}$ is completely determined by $x_1 \pmod{rc_1/n_1}$. Therefore estimating trivially, one sees that

$$\mathrm{FT}_{1}(n;k) \ll \sum_{d_{1}|c_{1}} d_{1} \sum_{d_{1}'|c_{1}} d_{1}' \sum_{\substack{x_{1}(rc_{1}/n_{1})\\\ell n_{1}x \equiv \mp m \, (\mathrm{mod} \, d_{1})}^{\star} 1.$$

For $FT_2(n;k)$, the congruence conditions there imply that $d_2|\ell n_1c_2 + nm$ and $d'_2|\ell' n_1c_2 + nm'$. Therefore

$$FT_2(n;k) \ll \sum_{\substack{d_2 \mid (c_2,\ell n_1 c'_2 + nm) \\ d'_2 \mid (c'_2,\ell' n_1 c_2 + nm')}} d_2 d'_2.$$

4.2.2. The q-sum. The q-sum equals q

$$(4.18) \quad \operatorname{FT}(n;q) = \frac{1}{\sqrt{q}} \sum_{u,u' \pmod{q}} L_{\pm \overline{c}^2 \overline{r}^2 m,\ell}(u;q) \overline{L_{\pm \overline{c'}^2 \overline{r}^2 m',\ell'}(u';q)} \times \sum_{v \pmod{q}} \operatorname{Kl}_2(\pm \overline{c}^3 \overline{r}^3 n_1^2 v \overline{u};q) \operatorname{Kl}_2(\pm \overline{c'}^3 \overline{r}^3 n_1^2 v \overline{u'};q) e\left(\frac{\overline{k} v n}{q}\right).$$

It will be useful to transform it to make it amenable to a sheaf-theoretic treatment.

For $\alpha, \beta, \gamma, \alpha', \beta', \gamma' \in \mathbf{F}_q^{\times}$ we recall (see (4.3))

$$(4.19) \quad L(u;q) = L_{\alpha,\beta}(u;q) := \frac{1}{\sqrt{q}} \sum_{b(q)} \widehat{K}(b) e\left(\frac{\alpha \overline{(b+\beta u)}}{q}\right) = \frac{1}{\sqrt{q}} \sum_{a \in \mathbf{F}_q} K(a) \operatorname{Kl}_2(\alpha a;q) e(-\frac{\beta a u}{q}),$$

the latter identity following from the expression of the Fourier transform (3.14). We also define

$$M(u) := \operatorname{Kl}_2(\gamma u; q)$$

and define likewise L'(u;q), M'(u') with α', β', γ' . The following choices of values of the parameters correspond to our initial problem:

(4.20)
$$\alpha = \pm \overline{c}^2 \overline{r}^2 m, \ \beta = \ell, \ \alpha' = \pm \overline{c'}^2 \overline{r}^2 m', \ \beta' = \ell'$$
$$\gamma = \pm \overline{c}^3 \overline{r}^3 n_1^2, \ \gamma' = \pm \overline{c'}^3 \overline{r}^3 n_1^2, \ \delta = \overline{k} n,$$

we see (by switching the u, u' sums and the v sum) that (4.18) equals

$$\sqrt{q} \sum_{v \, (\mathrm{mod} \, q)} M \star L(v) \overline{M' \star L'(v)} e(\frac{\delta v}{q})$$

where $M \star L$ denotes the normalized multiplicative convolution

$$M \star L(v) = \frac{1}{q^{1/2}} \sum_{u \in \mathbf{F}_q^{\times}} M(u) L(v/u) = \frac{1}{q^{1/2}} \sum_{u \in \mathbf{F}_q^{\times}} M(vu) L(1/u).$$

To evaluate such sums further we will need to make a few elementary transformations. By Plancherel formula we have

(4.21)
$$\sum_{v} M \star L(v) \overline{M' \star L'(v)} e(\frac{\delta v}{q}) = \sum_{v} \widehat{M \star L}(v) \overline{\widehat{M' \star L'(v-\delta)}} = \sum_{v} Z(v) \overline{Z'(v-\delta)}$$

say. Let us now compute Z(v):

$$Z(v) = \frac{1}{q^{1/2}} \frac{1}{q^{1/2}} \sum_{x \in \mathbf{F}_q^{\times}} \sum_{u} M(x) L(u/x) e(\frac{vu}{q}) = \frac{1}{q^{1/2}} \sum_{x \in \mathbf{F}_q^{\times}} M(x) \widehat{L}(xv) + \frac{1}{q^{1/2}} \sum_{v \in \mathbf{F}_q^{\times}} M(v) \widehat{L}(vv) + \frac{1}{q^{1/2}} \sum_{v \in \mathbf{F}_q^{\times}} M(v) + \frac{1}{q^{1/2}$$

By (8.1) we have

$$\widehat{L}(x) = K(x/\beta) \operatorname{Kl}_2(\alpha/\beta x; q)$$

and

(4.22)
$$Z(v) = \frac{1}{q^{1/2}} \sum_{x \in \mathbf{F}_q^{\times}} \operatorname{Kl}_2(\beta \gamma x; q) K(xv) \operatorname{Kl}_2(\alpha xv; q).$$

In $\S{8}$ we will prove the following

Proposition 4.5. Let $T_{\mathcal{F}}(\mathbf{F}_q)$ be the subgroup of \mathbf{F}_q^{\times} defined by

$$T_{\mathfrak{F}}(\mathbf{F}_q) = \{\lambda \in \mathbf{F}_q^{\times}, \ [\times \lambda]^* \mathfrak{F} \text{ is geometrically isomorphic to } \mathfrak{F} \}.$$

Assuming that the sheaf \mathfrak{F} is good, then for any $\alpha, \beta, \alpha', \beta', \gamma, \gamma', \delta \in \mathbf{F}_q^{\times}$, we have

$$\sum_{v} Z(v) \overline{Z'(v-\delta)} = O(q^{1/2}).$$

If $\delta = 0$ the above bound holds unless

$$\alpha/\alpha' = \beta\gamma/\beta'\gamma' \in T_{\mathcal{F}}(\mathbf{F}_q)$$

in which case

$$\sum_{v} Z(v) \overline{Z'(v)} = c_{\mathcal{F}}(\alpha/\alpha')q + O(q^{1/2})$$

for $c_{\mathfrak{F}}(\alpha/\alpha')$ some complex number of modulus 1. Here the implicit constants depend only on $C(\mathfrak{F})$.

Returning to our original sum, we see that (4.18) is O(q) unless (observe that the r variable is gone) $l^2 \cdots l^2 m l' = c l^3 \ell l c^3 \ell l' \subset T_{\mathbf{T}}(\mathbf{F})$

$$c m/c^2 m' = c \ell/c^3 \ell \in T_F(\mathbf{F}_q)$$

in which case (4.18) equals $C(c'^2 m/c^2 m')q^{3/2} + O(q)$ with $|C(c'^2 m/c^2 m')| = 1$.

5. Contribution of the n = 0 frequency

In this section we bound the contribution to (4.10) of the frequency n = 0. By (4.15), we then have

(5.1)
$$c_2 = c'_2, \ c = c', \ k = rc_1c_2^2/n_1$$

We use the case $\delta = 0$ of Proposition 4.5: by (5.1) and (4.20) we have that (4.18) is O(q) unless we have the congruence modulo q

$$m/m' = \ell/\ell' \in T_{\mathcal{F}}(\mathbf{F}_q)$$

in which case (4.18) equals $C(m/m')q^{3/2} + O(q)$ with |C(m/m')| = 1. The contribution of the n = 0 frequency to (4.10) is bounded by

$$\ll \frac{N}{q^{1/2}} \sum_{n_1, c=c_1 c_2} \frac{1}{n_1 k^{1/2}} \sum_{\ell, \ell'} \sum_{\ell, \ell'} |\lambda(1, \ell)| |\lambda(1, \ell')| \\ \times \sum_{m, m' \leqslant M} \sum_{m, m' \leqslant M} |\lambda_f(m)| |\lambda_f(m')| (q^{3/2} \delta_{m/m' = \ell/\ell' \in T_{\mathcal{F}}(\mathbf{F}_q)} + q) |\mathrm{FT}(0; k)|$$

$$\ll \frac{N}{q^{1/2}} \sum_{n_1, c=c_1 c_2} \frac{1}{n_1 k^{1/2}} \sum_{\ell, \ell'} \sum_{l, \ell'} |\lambda(1, \ell')|^2 \times \sum_{m, m' \leq M} \sum_{m, m' \leq M} |\lambda_f(m)|^2 (q^{3/2} \delta_{m/m' = \ell/\ell' \in T_{\mathcal{F}}(\mathbf{F}_q)} + q) |\mathrm{FT}(0; k)|$$

$$\ll q^{o(1)} \frac{rNC'}{q^{1/2}} \sum_{n_1,c} \frac{1}{n_1} \sum_{\ell,\ell'} |\lambda(1,\ell')|^2 \times \sum_{m,m' \leqslant M} \sum_{m,m' \leqslant M} |\lambda_f(m)|^2 (q^{3/2} \delta_{m\ell' = \ell m' \pmod{q}} + q) \sum_{\substack{d,d' \mid c \\ (\ell'd,\elld') \mid (m\ell' - m'\ell)}} (d,d').$$

Writing b = (d, d') the sum is bounded by

$$\ll q^{o(1)} \frac{rNC'}{q^{1/2}} \sum_{n_1, b \mid c \sim C'} \frac{b}{n_1} \sum_{\substack{\ell, \ell' \sim L, m, m' \leq M \\ b \mid \ell'm - \ell m'}} |\lambda(1, \ell')|^2 |\lambda_f(m)|^2 (q^{3/2} \delta_{m\ell' = \ell m' \pmod{q}} + q)$$
$$\ll q^{o(1)} \frac{rNC'}{q^{1/2}} LM \sum_{b \mid c \sim C'} b \left(q^{3/2} (\frac{LM}{qb} + 1) + q(\frac{LM}{b} + 1) \right)$$
$$\ll q^{o(1)} \frac{rNC'}{q^{1/2}} LMC' (q^{1/2}LM + C'q^{3/2} + qLM + C'q) \ll q^{o(1)} \frac{rNC'^2 LM}{q^{1/2}} (C'q^{3/2} + qLM).$$

Taking the squareroot of this term and multiplying by $\frac{X^{3/2}}{r^{1/2}Z^{3/2}L^{1/2}CC'^{5/2}q^2}$, we see that the contribution of these terms to (4.6) and then to (4.1) is bounded by

(5.2)
$$q^{o(1)} \frac{r^{1/2} X M^{1/2}}{CL^{1/2} q^{3/4}} \left(C'^{1/2} q^{3/4} + q^{1/2} L^{1/2} M^{1/2} \right)$$
$$\ll q^{o(1)} r^{1/2} \left(Z \frac{X^{3/4} q^{3/4}}{L^{1/4}} + Z^2 L^{1/2} X^{1/2} q^{5/4} \right).$$

In particular for $X = q^3$ we obtain

$$q^{o(1)}r^{1/2}(Z\frac{q^3}{L^{1/4}}+Z^2L^{1/2}q^{3-1/4})$$

which when r = 1 is non-trivial for $1 < L < q^{1/2}$.

6. Contribution from the $n \neq 0$ frequencies

Recall from (4.10) that

(6.1)

$$B_{n\neq0} = \frac{N}{q^{1/2}} \sum_{c_1,n_1} \sum_{n_1} \frac{1}{n_1} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{\ell' \in \mathcal{L}} \lambda(1,\ell') \ell'^{iv} \sum_{m \leqslant M} \overline{\lambda_f(m)} \sum_{m' \leqslant M} \lambda_f(m') \sum_{c_2 \sim C'/c_1} \sum_{c_2 \sim C'/c_1} \frac{1}{c_2' \sim C'/c_1} \sum_{m' \leqslant M} \frac{1}{k^{1/2}} \sum_{n \neq 0} \operatorname{FT}(n;q) \operatorname{FT}(n;k) \widehat{\mathcal{W}}(n/N^*).$$

We consider two cases: $n \neq 0 \pmod{q}$ and $n \equiv 0 \pmod{q}$.

6.1. $n \neq 0 \pmod{q}$. By Proposition 4.5, we have

$$FT(n;q) \ll q.$$

Combining this with Lemma 4.4, gives

$$B_{n\neq 0 \,(\mathrm{mod}\,q)} \ll q^{o(1)} N q^{1/2} \sum_{c_1,n_1} \frac{1}{n_1} \sum_{\ell \in \mathrm{L}} |\overline{\lambda(1,\ell)}| \sum_{\ell' \in \mathrm{L}} |\lambda(1,\ell')| \sum_{m} |\overline{\lambda_f(m)}| \sum_{m'} |\lambda_f(m')| \sum_{c_2} \sum_{c'_2} \sum_{m' \in \mathrm{ZN}^*} \sum_{\substack{n \ll \mathbb{ZN}^* \\ n \neq 0 \,(\mathrm{mod}\,q)}} \sum_{d_1,d'_1|c_1} d_1 d'_1 \sum_{\substack{x_1(rc_1/n_1) \\ \ell n_1 x_1 \equiv \mp m \,(\mathrm{mod}\,d_1) \, d'_2|(c'_2,\ell'n_1c'_2+nm')}} \sum_{d_2d'_2} |\widehat{\mathcal{W}}(n/N^*)|.$$

By using $|\lambda(1,\ell)||\lambda(1,\ell')||\lambda_f(m)||\lambda_f(m')| \ll |\lambda(1,\ell')\lambda_f(m)|^2 + |\lambda(1,\ell)\lambda_f(m')|^2$, we have

$$\begin{split} B_{n \neq 0 \, (\mathrm{mod} \, q)} \ll & q^{o(1)} N q^{1/2} \sum_{c_1, n_1} \sum_{l \in \mathrm{L}} \sum_{\ell' \in \mathrm{L}} |\lambda(1, \ell)|^2 \sum_{\substack{n \ll ZN^* \\ n \neq 0 \, (\mathrm{mod} \, q)}} \sum_{d_1 \mid c_1} d_1 \\ & \sum_{m \leqslant M} \sum_{m' \leqslant M} |\lambda_f(m')|^2 \sum_{c_2 \sim C'/c_1} \sum_{\substack{c_2' \sim C'/c_1 \\ \ell_1 \sim C'}} \sum_{\substack{n \ll ZN^* \\ n \neq 0 \, (\mathrm{mod} \, q)}} \sum_{d_1 \mid c_1 \atop d_1 \leq d_2 \mid (c_2, \ell_1 n c_2' + nm)} d_2 d_2'. \end{split}$$

The right hand side, when replacing c_2 by c_2d_2 and c'_2 by $c'_2d'_2$, is

$$q^{o(1)}Nq^{1/2}\sum_{c_1,n_1}\sum_{\substack{n \ll \mathbb{Z}N^* \\ n \neq 0 \,(\text{mod}\,q)}}\sum_{d_1|c_1}d_1\sum_{\substack{d_2 \ll C'/c_1}}d_2\sum_{\substack{d'_2 \ll C'/c_1}}d'_2\sum_{c_2 \sim C'/c_1d_2}$$

$$\sum_{\ell' \in \mathcal{L}}\sum_{\substack{m' \leqslant M \\ \ell'n_1c_2d_2+nm' \equiv 0 \,(\text{mod}\,d'_2)}}|\lambda_f(m')|^2 \times \sum_{\substack{c'_2 \sim C'/c_1d'_2}}\sum_{\ell \in \mathcal{L}}|\overline{\lambda(1,\ell)}|^2\sum_{\substack{x_1(rc_1/n_1) \\ x_1(rc_1/n_1)}}\sum_{\substack{m \leqslant M \\ \ell n_1c'_2d'_2+nm \equiv 0 \,(\text{mod}\,d_2)}}1.$$

Bounding the number of solutions in m of the congruence equations $\ell n_1 c'_2 d'_2 + nm \equiv 0 \pmod{d_2}$ and $\ell n_1 x_1 \equiv \mp m \pmod{d_1}$ we see that

$$\sum_{\substack{c_2' \sim C'/c_1 d_2' \ \ell \in \mathcal{L}}} \sum_{\substack{x_1(rc_1/n_1) \\ \ell n_1 c_2' d_2' + nm \equiv 0 \ (\text{mod } d_2) \\ \ell n_1 x_1 \equiv \mp m \ (\text{mod } d_1)}} 1 \qquad \ll q^{o(1)} (d_2, n) \frac{rc_1}{n_1} \frac{C'L}{c_1 d_2'} \left(1 + \frac{M}{d_1 d_2}\right) \\ \ll q^{o(1)} (d_2, n) \frac{r}{n_1} \frac{C'L}{d_2'} \left(1 + \frac{M}{d_1 d_2}\right).$$

Averaging this bound over the remaining variables we obtain

$$B_{n \neq 0 \,(\text{mod}\,q)} \ll q^{o(1)} r N q^{1/2} C' L \sum_{c_1, n_1} \sum_{\substack{n_1^2 \\ n \neq 0 \,(\text{mod}\,q)}}^{r \ll ZN^*} \sum_{\substack{n \ll ZN^* \\ n \neq 0 \,(\text{mod}\,q)}} \sum_{d_2 \ll C'/c_1} (d_2, n) \, (c_1 d_2 + M) \sum_{d'_2 \ll C'/c_1} \sum_{c_2 \sim C'/c_1 d_2} \sum_{\substack{\ell' \in \mathcal{L} \\ \ell' n_1 c_2 d_2 + nm' \equiv 0 \,(\text{mod}\,d'_2)}}^{r \ll ZN^*} |\lambda_f(m')|^2.$$

- If $\ell' n_1 c_2 d_2 + nm' \neq 0$ we interpret the congruence $\ell' n_1 c_2 d_2 + nm' \pmod{d'_2}$ as d'_2 being a divisor of that integer so that the contribution of such terms is bounded by

$$\ll q^{o(1)} r N q^{1/2} C' L \sum_{c_1, n_1} \frac{c_1}{n_1^2} Z N^* \sum_{d_2 \leqslant C'/c_1} (c_1 d_2 + M) \frac{C'}{c_1 d_2} L M$$

= $q^{o(1)} r Z N q^{1/2} C'^2 L^2 M \sum_{c_1, n_1} \sum_{n_1^2} \frac{c_1}{n_1^2} \frac{n_1}{c_1} \frac{q r C'^2}{N} \sum_{d_2 \leqslant C'/c_1} \left(1 + \frac{M}{c_1 d_2}\right)$
= $q^{o(1)} r^2 Z C'^4 q^{3/2} L^2 M \sum_{c_1, n_1} \sum_{n_1} \frac{1}{n_1} \sum_{d_2 \leqslant C'/c_1} \left(1 + \frac{M}{c_1 d_2}\right)$
= $q^{o(1)} r^2 Z C'^5 q^{3/2} L^2 M \left(1 + \frac{M}{C'}\right).$

– On the other hand, the contribution of the terms satisfying $\ell' n_1 c_2 d_2 + nm' = 0$ is bounded by (this follows from bounding the number of representations of m'n' as a product of four factors, where $n' = n/(d_2, n)$ and using the Rankin–Selberg bound (2.4))

$$q^{o(1)}rNq^{1/2}C'^{3}L\sum_{c_{1},n_{1}}\sum_{\substack{n_{1}^{2}\\n\neq 0 \pmod{q}}}\sum_{\substack{n\ll ZN^{*}\\n\neq 0 \pmod{q}}}\sum_{d_{2}\ll C'/c_{1}}(d_{2},n)\left(1+\frac{M}{C'}\right)\sum_{c_{2}\sim C'/c_{1}d_{2}}\sum_{\ell'\in \mathcal{L}}\sum_{\substack{m'\leqslant M\\\ell'n_{1}c_{2}d_{2}=-nm'}}|\lambda_{f}(m')|^{2}$$

$$\ll q^{o(1)}rNq^{1/2}C'^{3}L\left(1+\frac{M}{C'}\right)\sum_{c_{1},n_{1}}\sum_{\substack{n_{1}^{2}\\n_{1}^{2}}ZN^{*}M=q^{o(1)}r^{2}ZC'^{5}q^{3/2}LM\left(1+\frac{M}{C'}\right),$$

which is smaller than the previous term.

6.2. $q|n, n \neq 0$. In that case we write n = qn' with $n' \leq q^{o(1)}N^*/q$ and have

$$\mathrm{FT}(n'q;q) \ll q^{3/2}.$$

We have there lost a factor $q^{1/2}$ by comparison with the previous section. On the other hand we can run exactly the same argument as above with N^* replaced by N^*/q and all in all we find that

$$B_{q|n,n\neq 0} \ll q^{o(1)} r^2 Z C'^5 q^{3/2} L^2 M\left(1 + \frac{M}{C'}\right) \frac{1}{q^{1/2}}$$

The non-zero frequencies contribution to (4.6) and hence to (4.1) is bounded by

(6.2)
$$q^{o(1)} \frac{X^{3/2}}{r^{1/2} Z^{3/2} L^{1/2} C C'^{5/2} q^2} \left(B_{n \neq 0 \pmod{q}} + B_{q|n,n \neq 0} \right)^{1/2} \\ \leqslant q^{o(1)} \frac{X^{3/2}}{r^{1/2} Z^{3/2} L^{1/2} C C'^{5/2} q^2} \left(r^2 Z C'^5 q^{3/2} L^2 M \left(1 + \frac{M}{C'} \right) \right)^{1/2} \\ \leqslant q^{o(1)} r^{1/2} \left(\frac{L^{1/2} X}{q^{1/4}} + Z L^{3/4} X^{3/4} q^{1/2} \right).$$

6.3. Bounding Main₀: Conclusion. Let us recall that the sum Main₀ in (3.18) was split into two subsums depending on whether $(n_1, q) = 1$ or not.

By (5.2) and (6.2) the first subsum (4.1) is bounded by

(6.3)
$$\ll q^{o(1)} r^{1/2} \Big(Z \frac{X^{3/4} q^{3/4}}{L^{1/4}} + Z^2 L^{1/2} X^{1/2} q^{5/4} + \frac{L^{1/2} X}{q^{1/4}} + Z L^{3/4} X^{3/4} q^{1/2} \Big).$$

The complement sum (when $q|n_1$) is given (rewriting n_1 into qn_1) by

(6.4)
$$\operatorname{Err}_{4} = \frac{X}{L^{\star}Cq^{2}} \sum_{\pm \pm} \sum_{\substack{c \leqslant 2C \\ (c,q)=1}} \frac{1}{c} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{n_{1}|rc} \sum_{m,n} \overline{\lambda_{f}(m)} \times \frac{\lambda(n,n_{1}q)}{nn_{1}} \mathcal{C}(m,n;\frac{rc}{n_{1}}) \mathcal{W}_{\pm \pm} \left(\frac{m}{c^{2}q^{2}/X}, \frac{n_{1}^{2}n}{c^{3}qr/XL}\right)$$

where

$$\begin{split} \mathcal{C}(m,n;\frac{rc}{n_1}) &= \frac{1}{q^{1/2}} \sum_{b(q),u(cq)}^{} \widehat{K}(b)e\left(\frac{\pm \overline{bcr^2 + u\ell}m}{cq}\right) S\left(r\overline{u},\pm n;\frac{rc}{n_1}\right) \\ &= \frac{1}{q^{1/2}} \sum_{b(q),u(q)}^{} \widehat{K}(b)e\left(\frac{\pm \overline{bcr^2 + u\ell}m\overline{c}}{q}\right) \times \sum_{u\,(\mathrm{mod}\,c)}^{*} e\left(\frac{\pm \overline{u\ell}m\overline{q}}{c}\right) S\left(r\overline{u},\pm n;\frac{rc}{n_1}\right) \\ &= -K(0) \sum_{u\,(\mathrm{mod}\,c)}^{*} e\left(\frac{\pm \overline{u\ell}m\overline{q}}{c}\right) S\left(r\overline{u},\pm n;\frac{rc}{n_1}\right). \end{split}$$

This last *u*-sum is very similar to the sum (4.2) discussed previously (the Kloosterman sum $S(\bar{q}r\bar{u},\pm\bar{q}n;rc/n_1)$ has just been replaced by $S(r\bar{u},\pm n;rc/n_1)$) and is bounded by $(rc)^{1+o(1)}$ (cf. (4.13)). Using this bound we obtain that (6.4) is bounded by

(6.5)
$$\ll q^{o(1)+\theta_3} \frac{X}{LCq^2} L \frac{Z^2 C^2 q^2}{X} rC = q^{o(1)+\theta_3} rZ^2 \frac{XL}{q}$$

Combining this bound with (6.3) we obtain that

(6.6)
$$\begin{aligned} \operatorname{Main}_{0} \ll q^{o(1)+\theta_{3}} r Z^{2} \frac{XL}{q} + q^{o(1)} r^{1/2} \Big(Z \frac{X^{3/4} q^{3/4}}{L^{1/4}} \\ &+ Z^{2} L^{1/2} X^{1/2} q^{5/4} + \frac{L^{1/2} X}{q^{1/4}} + Z L^{3/4} X^{3/4} q^{1/2} \Big) \end{aligned}$$

7. Bounding $S_V(K, X)$: The final steps

7.1. Bounding Err_1 , Err_2 and Err_3 . Our treatment of the error terms Err_1 , Err_2 and Err_2 is similar to that of Main_0 and in the end these terms yield smaller contributions. We will again start by using Voronoi summation on the *m* and the *n* variables but, from this point on, the argument becomes much simpler than that of §4.1. For instance for Err_1 it is sufficient to bound the resulting sums trivially (no need to apply Cauchy–Schwarz to smooth the *n* variable). Since these arguments have already been presented in full details and in a more complex form in §4.1, we will be brief and pass over various technical points (for instance we will often assume the coprimality of different variables).

7.1.1. Err₁. We refer to (3.10) for the shape of this term. As before we apply Voronoi summation formulas to the *m* and *n* sums. The chief difference is that the exponential

$$n\mapsto e(\frac{an}{c})$$

has modulus $c \leq C$ instead of cq. The dual sum has length

$$(cZ)^3/XL \leqslant Z^3 X^{3/2} L^{3/2}/(Xq^{3/2}L) = Z^3 (XL)^{1/2}/q^{3/2}$$

which is rather small ($\leq L^{1/2}$ for $X = q^3$) and the exponential is (essentially) transformed into Kloosterman sums typically of the shape

(7.1)
$$n \mapsto \operatorname{Kl}_2(r\overline{a}n; rc).$$

For the m-sum, the function

$$m\mapsto K(mr^2)e(\frac{-am\ell}{c})$$

has period cq, therefore the dual m-sum (after Voronoi) has length

$$(cqZ)^2/X \leqslant Z^2qL$$

and is weighted (essentially) by the function

(7.2)
$$m \mapsto \widecheck{K}(\pm \overline{c}^2 \overline{r}^2 m) e(\pm \frac{\overline{a\ell q^2 m}}{c})$$

where

$$\widecheck{K}(m) = \frac{1}{\sqrt{q}} \sum_{b \in \mathbf{F}_q^{\times}} \widehat{K}(b) e(\frac{\overline{b}m}{q}) = \left(\widehat{K} \star e(\frac{\cdot}{q})\right)(m).$$

The convolution \check{K} is a trace function unless \mathcal{F} is geometrically isomorphic to $[x \mapsto \alpha x]^* \mathcal{K} \ell_2$, $\alpha \in \mathbf{F}_q^{\times}$ (whose Fourier transform is $[x \mapsto x^{-1}]^* \mathcal{L}_{\psi}$ for \mathcal{L}_{ψ} the Artin-Schreier sheaf attached to an additive character depending on α), but such situation is excluded by both the MO and the SL assumptions. Summing the product of (7.1) and (7.2) over $a \pmod{c}$ the resulting sum is bounded by

$$\ll \|\widecheck{K}\| c^{1/2+o(1)}$$
28

and therefore we obtain

(7.3)

$$\operatorname{Err}_{1} \ll \frac{q^{o(1)}}{LCq}L\sum_{c\leqslant C}\frac{1}{c}\frac{XL}{(cZ)^{3/2}}\frac{(cZ)^{3}}{XL}\frac{X}{cqZ}\frac{(cqZ)^{2}}{X}c^{1/2}$$

$$= \frac{q^{o(1)}Z^{5/2}}{Cq}\sum_{c\leqslant C}c^{2}q = q^{o(1)}Z^{5/2}C^{2} = q^{o(1)}Z^{5/2}\frac{XL}{q}$$

7.1.2. Err₂. We refer to (3.11) for notations and the shape of this term; in particular $c \leq C/q$. This time the two functions

$$n \mapsto e(\frac{an}{cq^2}), \ m \mapsto K(mr^2)e(\frac{-a\ell m}{cq^2})$$

have modulus $cq^2 \leq Cq$. Under Voronoi, for (a, cq) = 1 the first function transforms (essentially) into a Kloosterman sum of the shape

(7.4)
$$\operatorname{Kl}_2(\pm \overline{a}n; cq^2)$$

while the second sum transforms into

(7.5)
$$\frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) e(\pm m \frac{\overline{a\ell}(1 - \overline{a\ell}br^2 cq)}{cq^2}) = K(\pm (\overline{a\ell})^2 r^2 m) e(\pm \frac{\overline{a\ell}m}{cq^2}).$$

The sum over $a \pmod{cq^2}$, (a, cq) = 1 of the product of (7.4) and (7.5) equals

$$\sum_{\substack{a \,(\mathrm{mod}\,cq^2)}}^{\star} K(\pm(\overline{a\ell})^2 r^2 m) e(\pm \frac{a\ell m}{cq^2}) \,\mathrm{Kl}_2(\pm \overline{a}n; cq^2)$$
$$= \left(\sum_{\substack{a \,(\mathrm{mod}\,c)}}^{\star} e(\pm \frac{\overline{a\ell q^2}m}{c}) \,\mathrm{Kl}_2(\pm \overline{aq^4}n; c)\right) \left(\sum_{\substack{a \,(\mathrm{mod}\,q^2)}}^{\star} K(\pm \overline{a\ell}^2 r^2 m) e(\pm \frac{\overline{a\ell c}m}{q^2}) \,\mathrm{Kl}_2(\pm \overline{ac^2}n; q^2)\right).$$

The first factor is (essentially)

$$c^{1/2}e(-\frac{\overline{q}^2n\ell\overline{m}}{c})$$

while the second factor equals

$$\begin{split} \sum_{a \pmod{q^2}}^{\star} K(\pm a^2 r^2 m) e(\pm \frac{a\overline{c}m}{q^2}) \operatorname{Kl}_2(\pm a\overline{c^2}\ell n; q^2) = \\ \sum_{a \leqslant q}^{\star} K(\pm a^2 r^2 m) e(\pm \frac{a\overline{c}m}{q^2}) \sum_{b \pmod{q}} e(\pm \frac{b\overline{c}m}{q}) \frac{1}{q} \sum_{x \pmod{q^2}}^{\star} e(\frac{\pm (a + bq)\overline{c^2}\ell nx + \overline{x}}{q^2}) \\ &= \sum_{a \leqslant q}^{\star} K(\pm a^2 r^2 m) e(\pm \frac{a\overline{c}m}{q^2}) \sum_{\substack{x \pmod{q^2}\\x \equiv -cm\overline{\ell}n \pmod{q}}} e(\frac{\pm a\overline{c^2}\ell nx + \overline{x}}{q^2}) \\ &= \sum_{a \leqslant q}^{\star} K(\pm a^2 r^2 m) e(\pm \frac{a\overline{c}m}{q^2}) e(\mp \frac{\overline{a}\overline{c}m}{q^2}) e(-\frac{\overline{c}m\ell n}{q^2}) \sum_{y \pmod{q}} e(\frac{\pm a\overline{c^2}\ell ny - \overline{cm^2}(\ell n)^2 y}{q}) \\ &= qK(\overline{m}^3 \ell^2 r^2 n^2) e(\frac{-\overline{c}n\ell\overline{m}}{q^2}). \end{split}$$

Hence the sum over $a \pmod{cq^2}$, (a, cq) = 1 of the product of (7.4) and (7.5) equals (essentially) to

$$c^{1/2}q \cdot K(\overline{m}^{3}\ell^{2}r^{2}n^{2})e(\frac{-n\ell\overline{m}}{cq^{2}}),$$

and Err_2 is (essentially) transformed into

$$\frac{q^{o(1)}}{LCq} \sum_{c \leqslant C/q} \frac{1}{c^{1/2}} \sum_{\ell \in \mathcal{L}} \lambda(\ell, 1) \frac{XL}{(cq^2 Z)^{3/2}} \sum_{n \sim \frac{(cq^2 Z)^3}{XL}} \lambda(n, 1) \frac{X}{cq^2 Z} \sum_{m \ll \frac{(cq^2 Z)^2}{X}} \lambda_f(m) K(\overline{m}^3 \ell^2 r^2 n^2) e(\frac{-n\ell\overline{m}}{cq^2}) \\ \ll \frac{q^{o(1)} X^2}{Z^{5/2} Cq^6} \bigg| \sum_{c \leqslant C/q} \frac{1}{c^3} \sum_{n \sim Z^3 (XLq^3)^{1/2}} \lambda(n, 1) \sum_{\ell \in \mathcal{L}} \lambda(\ell, 1) \sum_{m \ll Z^2 qL} \lambda_f(m) K(\overline{m}^3 \ell^2 r^2 n^2) e(\frac{-n\ell\overline{m}}{cq^2}) \bigg|.$$

We just need to save a bit more than $O(q^{\delta}L)$ from the trivial estimate $\operatorname{Err}_2 \ll q^{o(1)}XL$.

As in the treatment of the generic term $Main_0$, we can then apply the Cauchy–Schwarz inequality to smooth the *n*-sum, but now we can put the *c*-sum outside the square and the (m, ℓ) -sums inside the square. After this, applying Poisson summation to the *n*-variable leads to the following

$$\sum_{\substack{n\sim Z^3(XLq^3)^{1/2}\\cq^2}} K(\overline{m}^3\ell^2r^2n^2)\overline{K(\overline{m'}^3\ell'^2r^2n^2)}e(\frac{-n\ell\overline{m}}{cq^2})e(\frac{n\ell'\overline{m'}}{cq^2}) = O_A(q^{-A}) + \frac{Z^3(XLq^3)^{1/2}}{cq^2}\sum_{\beta \pmod{q^2}} K(\overline{m}^3\ell^2r^2\beta^2)\overline{K(\overline{m'}^3\ell'^2r^2\beta^2)}e(\frac{-(\ell\overline{m}-\ell'\overline{m'})\beta\overline{c}}{q^2})\sum_{\gamma \pmod{c}} e(\frac{-(\ell\overline{m}-\ell'\overline{m'})\overline{q^2}\gamma}{c}).$$

Here we have noticed that in the dual sum only the n = 0 zero-frequency contributes, since

$$cq^2 \leqslant q^{-\eta} (XLq^3)^{1/2}$$

for some $\eta > 0$. Writing $\beta = \beta' + qy$, $\beta' \leq q$, $y \pmod{q}$ we see that this is further equal to

$$\begin{split} \frac{Z^3(XLq^3)^{1/2}}{q^2} &\sum_{\beta \leqslant q} K(\overline{m}^3 \ell^2 r^2 \beta^2) \overline{K(\overline{m'}^3 \ell'^2 r^2 \beta^2)} e(\frac{-(\ell \overline{m} - \ell' \overline{m'}) \beta \overline{c}}{q^2}) \\ & \times \sum_{y \, (\mathrm{mod} \, q)} e(\frac{-(\ell \overline{m} - \ell' \overline{m'}) y \overline{c}}{q}) \cdot \delta_{m\ell' \equiv m'\ell \, (\mathrm{mod} \, c)} \\ &= \frac{Z^3(XLq^3)^{1/2}}{q} \sum_{\beta \leqslant q} K(\overline{m}^3 \ell^2 r^2 \beta^2) \overline{K(\overline{m'}^3 \ell'^2 r^2 \beta^2)} e(\frac{-(\ell \overline{m} - \ell' \overline{m'}) \beta \overline{c}}{q^2}) \cdot \delta_{m\ell' \equiv m'\ell \, (\mathrm{mod} \, cq)}. \end{split}$$

Given (m, m', ℓ, ℓ') , we consider two cases.

- If $m\ell' = m'\ell$, the β -sum above is $\ll q$. Such terms contribute to Err_2

$$\frac{q^{o(1)}X^2}{Z^{5/2}Cq^6} \sum_{c \leqslant C/q} \frac{1}{c^3} (XLq^3)^{1/4} \left(\sum_{\ell \in \mathcal{L}} |\lambda(\ell, 1)|^2 \sum_{m \ll Z^2 qL} |\lambda_f(m)|^2 \frac{Z^3 (XLq^3)^{1/2}}{q} \cdot q \right)^{1/2} \ll q^{o(1)}X/q^{1/2}.$$

- If $m\ell' \neq m'\ell$, we can write $m\ell' = m'\ell + cq\delta$ with $\delta \neq 0$, then the β -sum after Poisson summation is

$$\sum_{\beta \pmod{q}} K(m'\ell\ell'r^2\beta^2) \overline{K(m'\ell^2r^2\beta^2)} e(\frac{\beta\delta}{q}).$$

We distinguish two further cases.

– If \mathcal{F} is "non-exceptional" in the sense of [FKM14, p. 1686] (i.e., \mathcal{F} is not geometrically isomorphic to the product of a Kummer and an Artin–Schreier sheaf or in other terms K is not proportional to the product of an additive and a multiplicative character $(\mod q)$) then for any $\alpha \in \mathbf{F}_q^{\times}$ the pull-back sheaf $[x \mapsto \alpha x^2]^* \mathcal{F}$ whose trace function is $x \mapsto K(\alpha x^2)$ is non-exceptional

and by [FKM14, Theorem 6.3] the above sum is $\ll q^{1/2}$ for every $\delta \pmod{q}$ unless $\ell/\ell' \pmod{q}$ belongs to a subset $B \subset \mathbf{F}_q^{\times}$ satisfying |B| = O(1). Then such terms contribute to Err_2

$$q^{o(1)} \frac{X^2}{Z^{5/2} C q^6} \sum_{c \leqslant C/q} \frac{1}{c^3} (XLq^3)^{1/4} \left(\sum_{\substack{\ell,\ell',m,m'\\m\ell' \equiv m'\ell \pmod{cq}}} \frac{Z^3 (XLq^3)^{1/2}}{q} \cdot (q^{1/2} + q\delta_{\ell/\ell' \in B}) \right)^{1/2} \\ \ll q^{o(1)} Z (XL)^{3/4} + q^{o(1)} Z X^{3/4} L^{1/4} q^{1/4}.$$

- If \mathcal{F} is exceptional then we may assume that $K(x) = \chi(x)e(\frac{kx}{q})$ for $k \in \mathbf{F}_q$ and χ a (non-trivial) Dirichlet character and the β -sum equals

$$\chi(m'\ell\ell')\overline{\chi(m'\ell^2)}\sum_{\beta \pmod{q}} \sum_{\substack{k \pmod{q}}} e(\frac{km'\ell(\ell'-\ell)r^2\beta^2+\beta\delta}{q}) \ll q^{1/2} + q\delta_{\substack{\ell \equiv \ell' \pmod{q}\\\delta \equiv 0 \pmod{q}}}.$$

In either cases we obtain

(7.6)
$$\operatorname{Err}_2 \ll q^{o(1)} X/q^{1/2} + q^{o(1)} Z(XL)^{3/4} + q^{o(1)} ZX^{3/4} L^{1/4} q^{1/4}$$

7.1.3. Err₃. Recall from (3.13) Err₃ takes the following shape

$$\operatorname{Err}_{3} = \frac{1}{L^{\star}Cq} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{\substack{c \leqslant 2C \\ \ell \mid c, (c,q) = 1}} \frac{1}{c} \sum_{u(cq)} \sum_{n=1}^{\star} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right)$$
$$\frac{1}{q^{1/2}} \sum_{b \pmod{q}} \widehat{K}(b) \sum_{m=1}^{\infty} \lambda_{f}(m) e\left(\frac{-(bcr^{2}/\ell + u)m}{cq/\ell}\right) m^{-iv} V\left(\frac{m}{X}\right) h\left(\frac{c}{C}, \frac{n - m\ell}{C^{2}q}\right).$$

We treat exactly the same manner as what we did for the main term $Main_0$ in (3.15). The chief difference occurs when we apply Proposition 2.1 to the *m*-sum, as now the modulus of the additive character is cq/ℓ rather than cq. Instead of getting (3.17), we obtain

$$\operatorname{Err}_{3} = \frac{X}{L^{\star}Cq^{2}} \sum_{\pm} \sum_{\ell \in \mathcal{L}} \overline{\lambda(1,\ell)} \ell^{-iv} \sum_{\substack{c \leq 2C\\\ell \mid c, (c,q) = 1}} \frac{\ell}{c^{2}} \frac{1}{q^{1/2}} \sum_{b(q), u(cq)} \widehat{K}(b)$$
$$\times \sum_{m=1}^{\infty} \overline{\lambda_{f}(m)} e(\frac{\pm \overline{bcr^{2}/\ell + um}}{cq/\ell}) \sum_{n=1}^{\infty} \lambda(r,n) e\left(n\frac{u}{cq}\right) n^{iv} U\left(\frac{n}{XL}\right) \widehat{\mathcal{V}}_{n}^{\pm} \left(\frac{mX}{(c/\ell)^{2}q^{2}}\right).$$

The extra factor ℓ in the summand of the *c*-sum will compensate with the congruence condition $\ell | c$, as compared to Main₀, but now the effective length of the dual *m*-sum is $m \simeq M/L^2$, reduced by a factor L^2 as compared to $m \simeq M$ for the case of Main₀; cf. (4.5).

Then we proceed as what we did for Main₀, and the sum $\mathcal{C}(m, n; \frac{rcq}{n_1})$ in (3.18) will be changed into $\mathcal{C}(m\ell^2, n; \frac{rcq}{n_1})$. Correspondingly, the parameter m in §4.2 will have to be changed into $m\ell^2$ (and $m' \to m'\ell'^2$). The change of the parameter $m \to m\ell^2$ in the character sum has the effect of swapping the parameters ℓ and ℓ' in the congruence conditions in §5, and everything else essentially remains the same. Since the length of the m-sum is reduced from M to M/L^2 , by book-keeping eventually we will save some factor $O(L^{\eta}), \eta > 0$ for Err₃, compared to the bound that we obtained for Main₀. 7.2. Conclusion. Collecting the bounds (6.6), (7.3) and (7.6) and (3.4) into (3.8) we obtain that

$$S_{V,r}(K,X) \ll q^{o(1)} \frac{X}{L} r^{\theta_3} L^{\theta_2} + q^{o(1)+\theta_3} r Z^2 \frac{XL}{q} + q^{o(1)} r^{1/2} \left(Z \frac{X^{3/4} q^{3/4}}{L^{1/4}} + Z^2 L^{1/2} X^{1/2} q^{5/4} + L^{1/2} X q^{-1/4} + Z L^{3/4} X^{3/4} q^{1/2} \right)$$

and on taking

 $L = q^{1/4}$

to equate the first and fourth terms inside the parentheses we get

$$S_{V,r}(K,X) \ll q^{o(1)} r^{1/2} \left(Z X^{3/4} q^{11/16} + Z^2 X^{1/2} q^{11/8} + X q^{-1/8} + q^{\theta_3} r^{1/2} Z^2 X q^{-3/4} \right).$$

Replacing X by X/r^2 and averaging this bound over $r \leq R$ we obtain

$$\sum_{r \leqslant R} S_{V,r}(K, X/r^2) \ll q^{o(1)} \left(ZX^{3/4} q^{11/16} + Z^2 R^{1/2} X^{1/2} q^{11/8} + Xq^{-1/8} + Z^2 Xq^{\theta_3 - 3/4} \right)$$

Combining this with (3.2), we have

$$\sum_{r \leqslant R} S_{V,r}(K, X/r^2) + \sum_{r \geqslant R} S_{V,r}(K, X/r^2) \\ \ll q^{o(1)} \left(ZX^{3/4}q^{11/16} + Z^2R^{1/2}X^{1/2}q^{11/8} + Xq^{-1/8} + Z^2Xq^{\theta_3 - 3/4} + XR^{\theta_3 - 1} \right),$$

which upon choosing

(7.7)
$$R = \left(\frac{X}{Z^4 q^{11/4}}\right)^{\frac{1}{3-2\theta_3}}$$

to equate the second and the fifth terms gives

$$S_V^t(K,X) \ll q^{o(1)} \left(Z X^{3/4} q^{11/16} + Z^{\frac{4(1-\theta_3)}{3-2\theta_3}} X^{\frac{2-\theta_3}{3-2\theta_3}} q^{\frac{11(1-\theta_3)}{4(3-2\theta_3)}} + X q^{-1/8} \right).$$

This completes the proof of Theorem 1.1.

Plugging (7.7) into our previous assumption $R < L = q^{1/4}$ in (3.3) we see the restriction

$$Z^4 q^{11/4} < X < Z^4 q^{\frac{7-\theta_3}{2}}.$$

8. Square-root cancellation for certain exponential sums

In this section we establish Proposition 4.5. Let K be the trace function of an irreducible middle extension sheaf \mathcal{F} on $\mathbb{P}^1_{\mathbf{F}_a}$ of conductor $C(\mathcal{F})$. For $\alpha, \beta, \alpha', \beta' \in \mathbf{F}_q^{\times}$ we define

(8.1)
$$L(u;q) := \frac{1}{\sqrt{q}} \sum_{b(q)} \widehat{K}(b) e\left(\frac{\alpha \overline{(b+\beta u)}}{q}\right) = \frac{1}{\sqrt{q}} \sum_{a} K(a) \operatorname{Kl}_2(\alpha a;q) e(-\frac{\beta a u}{q}),$$

and define L'(u;q) in the same way with α', β' instead. We also set

(8.2)
$$Z(v) = \frac{1}{q^{1/2}} \sum_{x \in \mathbf{F}_q^{\times}} K(xv) \operatorname{Kl}_2(\alpha xv; q) \operatorname{Kl}_2(\beta x; q)$$

and define Z'(v) likewise using α', β' . In this section we provide bounds for the sums

(8.3)
$$\sum_{v} Z(v) \overline{Z'(v-\delta)}$$

for $\delta \in \mathbf{F}_q$ with different methods depending on whether $\delta = 0$ or not.

We recall the key assumptions:

- (MO) There is no $\lambda \in \mathbf{F}_q^{\times}$ such that the geometric monodromy group of \mathcal{F} has some quotient which is equal, as a representation of π_1 into an algebraic group, to the geometric monodromy representation of $[\times \lambda]^* \mathcal{K} \ell_2$ modulo ± 1 .
- (SL) The local monodromy representation of \mathcal{F} at ∞ has no summand with slope 1/2.

8.1. The case $\delta = 0$.

Lemma 8.1. Assume \mathcal{F} satisfies (MO). Then

$$L(u;q) = O(c(\mathcal{F})^{O(1)}).$$

Proof. By Deligne's theorem, because \mathcal{F} and $\mathcal{K}\ell_2(\alpha a) \otimes \mathcal{L}_{\psi}(\beta a)$ are irreducible, this cancellation holds unless \mathcal{F} is geometrically isomorphic to $\mathcal{K}\ell_2(\alpha a) \otimes \mathcal{L}_{\psi}(\beta a)$, which clearly violates assumption (MO).

Lemma 8.2. We have

$$\left|\sum_{u} L(u;q)\right| \leqslant c(\mathcal{F})q^{1/2}$$

Proof.

$$\sum_{u} L(u;q) = q^{-1/2} \sum_{u} \sum_{a} K(a) \operatorname{Kl}_2(\alpha a;q) e(-\frac{\beta a u}{q}) = q^{1/2} K(0) \operatorname{Kl}_2(0) = K(0) q^{1/2}$$

 $|K(0)| \leqslant c(\mathcal{F}).$

and

For the third Lemma we need to define the following group

(8.4)
$$T_{\mathcal{F}}(\mathbf{F}_q) = \{\lambda \in \mathbf{F}_q^{\times}, \ [\times \lambda]^* \mathcal{F} \text{ is geometrically isomorphic to } \mathcal{F}\}.$$

Lemma 8.3. Suppose \mathcal{F} satisfies (MO). Then

$$\sum_{u(q)} L(u;q)\overline{L'(u;q)} = O(c(\mathfrak{F})^{O(1)}q^{1/2})$$

unless

$$\alpha/\alpha' = \beta/\beta' \in T_{\mathcal{F}}(\mathbf{F}_q).$$

In that later case we have

$$\sum_{u(q)} L(u;q)\overline{L'(u;q)} = c_{\mathcal{F}}(\alpha/\alpha')q + O(c(\mathcal{F})^{O(1)}q^{1/2})$$

for $c_{\mathcal{F}}(\alpha/\alpha')$ a complex number of modulus 1. In particular if $T_{\mathcal{F}}(\mathbf{F}_q) = \{1\}$ we have

$$\sum_{u(q)} L(u;q) \overline{L'(u;q)} = \delta_{\substack{\alpha = \alpha' \\ \beta = \beta'}} q + O(c(\mathfrak{F})^{O(1)} q^{1/2}).$$

Proof. By Plancherel formula the sum equals

(8.5)
$$\sum_{a \pmod{q}} K(\beta^{-1}a) \operatorname{Kl}_2((\alpha/\beta)a;q) \overline{K(\beta'^{-1}a) \operatorname{Kl}_2((\alpha'/\beta')a;q)}$$

Square-root cancellation now follows unless there is a nontrival map from $[\times \beta^{-1}]^* \mathcal{F} \otimes [\times \beta'^{-1}]^* \mathcal{F}^{\vee}$ to $[\times \alpha/\beta]^* \mathcal{K} \ell_2 \otimes [\times \alpha'/\beta'] \mathcal{K} \ell_2$. If a nontrivial map exists, then an irreducible component on the left side must be isomorphic to an irreducible component on the right.

If the irreducible component on the right is trivial, then there is a nontrivial map from $[\times\beta^{-1}]^*\mathcal{F}\otimes$ $[\times\beta'^{-1}]^*\mathcal{F}^{\vee}$ to the constant sheaf \mathbf{Q}_{ℓ} , hence a nontrivial map from $[\times\beta^{-1}]^*\mathcal{F}$ to $[\times\beta'^{-1}]^*\mathcal{F}$, which must be an isomorphism because both sides are irreducible, giving $\beta/\beta' = \lambda$. Furthermore, because \mathbf{Q}_{ℓ} appears on the other side and $\mathcal{K}\ell_2$ is not geometrically isomorphic to any non-trivial multiplicative twist of itself [Kat88], we must have $\alpha/\beta = \alpha'/\beta'$ so $\alpha/\alpha' = \beta/\beta' = \lambda$.

If the irreducible component on the right is nontrivial, then its monodromy must be a nontrivial quotient of the monodromy of the product of two $\mathcal{K}\ell_2$, which is either SL_2 or $SL_2 \times SL_2$ depending on whether the two copies are isomorphic or not (by the Goursat-Kolchin-Ribet criterion in the later case). In either case, it has a further quotient equal to PGL_2 , with π_1 acting by its geometric monodromy action on one of the Kloosterman sheaves, modulo ± 1 . This quotient must also appear, as a π_1 -representation, as a quotient of the monodromy on the left side. Because it is a simple group, it must appears as a quotient of the monodromy of either $[\times \beta^{-1}]^*\mathcal{F}$ or $[\times \beta'^{-1}]^*\mathcal{F}$. This implies that assumption (MO) is violated.

8.2. The case $\delta \neq 0$.

Proposition 8.4. Assume that \mathcal{F} satisfies both (MO) and (SL). Then for any $\alpha, \beta, \alpha', \beta', \delta \in \mathbf{F}_q^{\times}$ we have

(8.6)
$$\sum_{v} Z(v) \overline{Z'(v-\delta)} = O(q^{1/2})$$

Here the implicit constants depend only on $C(\mathfrak{K})$.

We will prove this in several steps, using a series of lemmas. Let us first observe that by (8.2), Z(v) is the trace function of the sheaf

(8.7)
$$\mathcal{Z} := (\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2) \star [y \to \beta/y]^* \mathcal{K} \ell_2$$

where \star is multiplicative convolution. Our lemmas will focus mostly on understanding the geometry of this sheaf convolution and to compare with the sheaf \mathcal{Z}' defined using the parameters α', β' . We denote

$$\mathcal{K} = \mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2, \ \mathcal{L} = [y \to \beta/y]^* \mathcal{K} \ell_2$$

and

$$\mathcal{K}' = \mathfrak{F} \otimes [imes lpha']^* \mathcal{K} \ell_2, \ \mathcal{L}' = [y o eta'/y]^* \mathcal{K} \ell_2.$$

Lemma 8.5. If \mathcal{F} satisfies (MO), then \mathcal{Z} is geometrically irreducible

Proof. It follows from Goursat's lemma that, if \mathcal{F} satisfies (MO), the monodromy of $\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2$ is the product of the monodromy group of \mathcal{F} and the monodromy group of $[\times \alpha]^* \mathcal{K} \ell_2$. Because it is a tensor product of irreducible representations, it is an irreducible representation of the product group. Then because $[y \to \beta/y]^* \mathcal{K} \ell_2$ is an object of dimension 1 in the sheaf convolution Tannakian category, convolving with it preserves irreducibility. (Alternately, we can view the convolution as a Fourier transform, change of variables $y \to y^{-1}$, then another Fourier transform, and each of these steps preserve irreducibility.)

It follows immediately from Lemma 8.5 that, if \mathcal{F} satisfies (MO), then the bound of (8.6) holds unless \mathcal{Z}' is geometrically isomorphic to $[+\delta]^*\mathcal{Z}$. So this case is what we will focus on eliminating.

Lemma 8.6. If \mathcal{F} satisfies (SL), then \mathcal{Z} is lisse away from $\{0, \beta/\alpha, \infty\}$

Proof. The middle convolution $\mathcal{K} \star \mathcal{L}$ is equal to the compactly supported convolution $\mathcal{K} \star_{!} \mathcal{L}$ up to a lisse sheaf, so it suffices to prove this for the compactly supported convolution. The compactly supported convolution is

$$R\pi_!((\mathfrak{F}\otimes[\times\alpha]^*\mathfrak{K}\ell_2)\otimes[(x,v)\to\beta x/v]^*\mathfrak{K}\ell_2)$$
₃₄

where $\pi : \mathbb{G}_m \times \mathbb{G}_m \to \mathbb{G}_m$ is the projection from the torus with coordinates (x, v) to the torus with coordinate v. To prove lisseness, we apply Deligne's semicontinuity theorem and examine the variation with v of the Euler characteristic of the sheaf (in the *x*-variable) $(\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2) \otimes [x \to \beta x/v]^* \mathcal{K} \ell_2$

Let V be an indecomposable summand of the local monodromy representation of \mathcal{F} at ∞ . By assumption (SL), the slope of V is not 1/2.

If the slope of V is > 1/2, then $V \otimes [\times \alpha]^* \mathcal{K}\ell_2$ has the same slope, and thus for any given v, the sheaf in the x-variable, $(V \otimes [\times \alpha]^* \mathcal{K}\ell_2) \otimes [x \to \beta x/v]^* \mathcal{K}\ell_2$ has the same slope, which in particular is independent of v, so the contribution of this representation to the Swan conductor is constant.

If the slope of V is < 1/2, then the slope of $V[\times \alpha]^* \mathcal{K}\ell_2$ is exactly 1/2, and the slope of $(V \otimes [\times \alpha]^* \mathcal{K}\ell_2) \otimes [x \to \beta x/v]^* \mathcal{K}\ell_2$ is at most 1/2. The slope of $(V \otimes [\times \alpha]^* \mathcal{K}\ell_2) \otimes [x \to \beta x/v]^* \mathcal{K}\ell_2$ is less than 1/2 if and only if $[\times \alpha]^* \mathcal{K}\ell_2 \otimes [x \to \beta x/v]^* \mathcal{K}\ell_2$ has a summand of slope < 1/2. By known properties of the Kloosterman sheaf [Kat88, 10.4.5], this happens if and only if $\alpha = \beta v^{-1}$.

Because the Swan conductor at ∞ of $(\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2) \otimes [x \to \beta x/v]^* \mathcal{K} \ell_2$ is constant away from $v = \beta/\alpha$, and its Swan conductor elsewhere is constant, by Deligne's semicontinuity theorem its cohomology is lisse (in the v variable) away from $v = \beta/\alpha$.

Lemma 8.7. If \mathcal{F} satisfies (SL), then $\mathcal{K} \star \mathcal{L}$ has a nontrivial singularity at zero.

Proof. Let us first check that, if \mathcal{F} satisfies (SL), at least one of the following four conditions must be satisfied:

- (1) \mathcal{F} is singular at some point on \mathbb{G}_m .
- (2) The local monodromy representation of \mathcal{F} at zero is not unipotent.
- (3) The local monodromy representation of \mathcal{F} at zero has a Jordan block of size ≥ 3 .
- (4) The local monodromy representation of \mathcal{F} at ∞ has a summand with slopes > 1/2.

This follows from an Euler characteristic calculation. Assume none of these happen. Because \mathcal{F} is a nontrivial irreducible middle extension sheaf on \mathbb{A}^1 , its Euler characteristic is nonpositive. Because \mathcal{F} is unipotent at 0, it is tamely ramified at 0. Because \mathcal{F} is lisse on \mathbb{G}_m and tamely ramified at 0, its Euler characteristic is its rank, minus its drop at 0, minus its Swan conductor at ∞ . Because all the local monodromy at 0 is unipotent, with unipotent blocks of size at most 2, its drop at 0 is at most half the rank. Because its slopes at ∞ are < 1/2, its Swan conductor at ∞ is less than half the rank. So the Euler characteristic is positive, giving a contradiction.

So at least one of (1) to (4) must happen. This imply corresponding conditions for $\mathcal{K} = \mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2$. We must have either

- (1) \mathcal{K} is singular at some point on \mathbb{G}_m .
- (2) The local monodromy representation of \mathcal{K} at zero is not unipotent.
- (3) The local monodromy representation of \mathcal{K} at zero has a Jordan block of size ≥ 4 .
- (4) The local monodromy representation of \mathcal{K} at ∞ has a summand with slopes > 1/2.

These follow by straightforward arguments. The most subtle are that, in case (2), we must check that the tensor product of a non-unipotent representation with a nonzero unipotent representation is non-unipotent, and in case (3), we must check that a Jordan block of size ≥ 3 tensored with a Jordan block of size ≥ 2 produces a Jordan block of size ≥ 4 . Both follow from standard representation theory.

We will now show, in each of these cases, that $\mathcal{K} \star \mathcal{L}$ has nontrivial local monodromy at zero. In fact, we will break case (2) into the cases (2w) where the local monodromy at 0 is wild and (2t) where the local monodromy at 0 is tame but not unipotent. Rojas-León has shown [RL13, Theorem 16] that the wild part of the local monodromy at 0 of a sheaf convolution $\mathcal{K} \star \mathcal{L}$ is given by a sum of the values of certain functors applied to the local monodromy representations of \mathcal{K} and \mathcal{L} at different points. We will show, in cases (1), (2w), and (4), that one of those functors produces a

nontrivial value on \mathcal{K} and \mathcal{L} and thus the local monodromy representation of $\mathcal{K} \star \mathcal{L}$ at 0 is wild. In cases (2t) and (3) we will show, using a different result of Rojas-León (encapsulating earlier work of Katz), that the local monodromy at 0 of $\mathcal{K} \star \mathcal{L}$ contains a nontrivial tame component. In every case we will deduce it is nontrivial.

The functors appearing in [RL13, Theorem 16] are defined by swapping 0 and ∞ in the functors defined in [RL13, Theorem 9], which is harmless as \mathbb{G}_m has a symmetry switching 0 and ∞ . Thus we will need to swap 0 and ∞ when citing results from [RL13].

- (1) Fix s a singularity of \mathcal{K} in \mathbb{G}_m . We must show that the function $\overline{\varrho}_{(s,0)}$ applied to $(\mathcal{K}^w_{(s)}, \mathcal{L}^w_{(0)})$ is non-trivial. This follows from [RL13, Proposition 12], using the assumption that \mathcal{K} has a singularity at s, and the fact that \mathcal{L} has non-trivial wild local monodromy at 0.
- (2w) In this case we must show that the functor $\overline{\varrho}_{(0,0)}$ applied to $(\mathcal{K}^w_{(0)}, \mathcal{L}^w_{(0)})$ (the wild parts of the local monodromy representations of these two sheaves at 0) is non-trivial. This follows from [RL13, Proposition 11] which shows that this functor applied to any two nontrivial wild representations is nontrivial, our assumption that \mathcal{K} has nontrivial wild local monodromy at 0, and the fact that \mathcal{L} has nontrivial wild local monodromy at 0.
- (2t) For each character χ of the tame fundamental group of \mathbb{G}_m , and for U_k a unipotent Jordan block of size k, Rojas-León defines [RL13, §6] a polynomial $P_{\mathcal{K},\chi}(T)$ associated to a middle extension sheaf \mathcal{K} , whose coefficient of T^k is the multiplicity of $\mathcal{L}_{\chi} \otimes U_k$ as a direct summand of the local monodromy of \mathcal{K} at ∞ if k > 0 and the multiplicity of $\mathcal{L}_{\chi} \otimes U_{-k}$ as a direct summand of the local monodromy of \mathcal{K} at 0 if k < 0, and such that $P_{\mathcal{K},\chi}(1)$ is minus the Euler characteristic of \mathcal{K} . He proves that $P_{\mathcal{K}\star\mathcal{L},\chi}(T) = P_{\mathcal{K},\chi}(T)P_{\mathcal{L},\chi}(T)$ [RL13, Proposition 28].

If the local monodromy representation of \mathcal{K} at 0 is tame but not unipotent, then it contains some $\mathcal{L}_{\chi} \otimes U_k$ for some k > 0 and nontrivial character χ . Thus $P_{\mathcal{K},\chi}(T)$ contains some term T^{-k} . Because χ is nontrivial, \mathcal{L}_{χ} does not appear in the local monodromy of \mathcal{L} at 0 or ∞ , and so $P_{\mathcal{L},\chi}(T) = -\chi(\mathbb{G}_m, \mathcal{L}) = 1$. Thus $P_{\mathcal{K},\chi}(T)P_{\mathcal{L},\chi}(T)$ contains some term T^{-k} , so the local monodromy representation of $\mathcal{K} \star \mathcal{L}$ contains $\mathcal{L}_{\chi} \otimes U_k$ and is nontrivial.

- (3) Now taking χ to be trivial, we see that $P_{\mathcal{K},1}(T)$ contains some term of the form T^{-k} for $k \ge 4$. On the other hand, because the local monodromy of \mathcal{L} at ∞ is a unipotent block of size 2, $P_{\mathcal{L},1}(T) = T^2$. Thus $P_{\mathcal{K},\chi}(T)P_{\mathcal{L},\chi}(T)$ contains some term of the form T^{2-k} for $k \ge 4$, and thus the local monodromy representation of $\mathcal{K} \star \mathcal{L}$ contains U_{k-2} and is nontrivial.
- (4) In this case we must show that the functor $\overline{\varrho}_{(\infty,0)}$ applied to $(\mathcal{K}^w_{(\infty)}, \mathcal{L}^w_{(0)})$, the wild parts of the local monodromy representations of these two sheaves at ∞ and 0 respectively, is nontrivial. This follow from [RL13, Proposition 13] which implies (swapping 0 and ∞) that $\overline{\varrho}_{(\infty,0)}$ is nontrivial as soon as some slope *a* of the first input is greater than some slope *b* of the first input, our assumption that \mathcal{K} has some slope > 1/2 at ∞ , and the fact that \mathcal{L} has slope 1/2 at 0.

Let us adopt some notation. Let $U_k(\mathcal{F})$ be the number of unipotent blocks of size k in the local monodromy representations of \mathcal{F} at 0, let $\chi_x(\mathcal{F})$ be the drop plus the Swan conductor at the point x, and let $N_s(\mathcal{F})$ be the rank of the local monodromy representation of \mathcal{F} at ∞ with slope s.

Lemma 8.8. The dimension of the monodromy invariant subspace of $\mathcal{K} \star \mathcal{L}$ at 0 is

$$\sum_{k \ge 2} U_k(\mathcal{F}) + \sum_{k \ge 4} U_k(\mathcal{F})$$

Proof. The dimension of the monodromy invariants at 0 is the number of unipotent blocks in the local monodromy representation at 0, which is the sum of all the coefficients of negative powers of

T in $P_{\mathcal{K}\star\mathcal{L},1}(T)$, which is the sum of all coefficients of powers < -2 of T in $P_{\mathcal{K},1}(T)$, which is the number of unipotent blocks of size at least 3 of \mathcal{K} . A unipotent block of \mathcal{F} of size k produces a unipotent block of $\mathcal{F}\otimes[\times\alpha]^*\mathcal{K}\ell_2$ of size k+1 and one of size k-1, so it produces one unipotent block of size 3 if $k \ge 2$ and another if $k \ge 4$.

Lemma 8.9. Assume \mathcal{F} satisfies (SL). The dimension of the monodromy invariant subspace of $\mathcal{K} \star \mathcal{L}$ at β/α is

$$4\operatorname{Swan}_{0}(\mathcal{F}) + 4\sum_{x \in \mathbb{G}_{m}} \chi_{x}(\mathcal{F}) + \sum_{s} N_{s}(\mathcal{F})(2s + 2\max(s, 1/2)) - \left(2\sum_{k \ge 1} U_{k}(\mathcal{F}) + \sum_{k \ge 2} U_{k}(\mathcal{F}) + \sum_{k \ge 4} U_{k}(\mathcal{F})\right)$$

Proof. The dimension of the monodromy invariants at β/α is minus the Euler characteristic of

$$\mathcal{K} \otimes [y \to \beta/(\alpha y)]^* \mathcal{L} = \mathcal{K} \otimes [\times \alpha]^* \mathcal{K} \ell_2 = \mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2 \otimes [\times \alpha]^* \mathcal{K} \ell_2$$

on \mathbb{G}_m , minus the dimensions of the spaces of local monodromy invariants at 0 and ∞ of

$$\mathfrak{K} \otimes [y \to v/y]^* \mathcal{L} = \mathfrak{F} \otimes [\times \alpha]^* \mathfrak{K} \ell_2 \otimes [\times \beta/v]^* \mathfrak{K} \ell_2$$

for generic v. These extra terms come from taking the middle convolution and not the compactlysupported convolution.

By combining the formula for the Euler characteristic of a middle extension sheaf with a calculation of the invariants and Swan conductor of a tensor product monodromy representation at each point, the Euler characteristic term is 4 times the sum of the drop plus Swan at all the finite singularities of \mathcal{F} , plus the sum over all local monodromy representations at ∞ of rank r and slope s of $r(2s + 2\max(s, 1/2))$, plus 4 times the Swan conductor of \mathcal{F} at 0.

Because $[\times \alpha]^* \mathcal{K}\ell_2$ and $[\times \beta/v]^* \mathcal{K}\ell_2$ both have local monodromy representations at 0 unipotent of rank 2, the dimension of the local monodromy invariants of $\mathcal{F} \otimes [\times \alpha]^* \mathcal{K}\ell_2 \otimes [\times \alpha]^* \mathcal{K}\ell_2$ at 0 is 2 for each unipotent block of size 1 in the local monodromy of \mathcal{F} at 0, 3 for each unipotent block of size 2, and 4 for each unipotent block of greater size.

Because the local monodromy of $[\times \beta/v]^* \mathcal{K}\ell_2$ at ∞ is irreducible and distinct for distinct v, for only finitely many $v \operatorname{can} \mathcal{F} \otimes [\times \alpha]^* \mathcal{K}\ell_2 \otimes [\times \beta/v]^* \mathcal{K}\ell_2$ have nonzero local monodromy invariants at ∞ , meaning that for generic v there are no local monodromy invariants. \Box

Lemma 8.10. Assume \mathfrak{F} satisfies (SL). Suppose that the dimension of the monodromy invariants subspace of $\mathfrak{K} \star \mathfrak{L}$ at 0 is equal to the dimension of the monodromy invariant subspace of $\mathfrak{K}' \star \mathfrak{L}'$ at β'/α' . Then \mathfrak{F} is lisse on \mathbb{G}_m , with the local monodromy at 0 unipotent with all blocks of size 2, 3, and 4, all slopes of the local monodromy representation at ∞ at most 1/2, and $\chi(\mathbb{A}^1, \mathfrak{F}) = 0$.

Proof. By Lemmas 8.8 and 8.9 our assumption implies that

$$\sum_{k \ge 2} U_k(\mathcal{F}) + \sum_{k \ge 4} U_k(\mathcal{F})$$

$$= 4 \operatorname{Swan}_{0}(\mathcal{F}) + 4 \sum_{x \in \mathbb{G}_{m}} \chi_{x}(\mathcal{F}) + \sum_{s} N_{s}(\mathcal{F})(2s + 2\max(s, 1/2)) - \left(2 \sum_{k \ge 1} U_{k}(\mathcal{F}) + \sum_{k \ge 2} U_{k}(\mathcal{F}) + \sum_{k \ge 4} U_{k}(\mathcal{F})\right)$$

This gives

$$2\sum_{k\geq 1} U_k(\mathcal{F}) + 2\sum_{k\geq 2} U_k(\mathcal{F}) + \sum_{k\geq 3} U_k(\mathcal{F}) + \sum_{k\geq 4} U_k(\mathcal{F})$$

= 4 Swan₀(\mathcal{F}) + 4 $\sum_{x\in\mathbb{G}_m} \chi_x(\mathcal{F}) + \sum_s N_s(\mathbb{F})(2s + 2\max(s, 1/2))$

On the other hand, we have

$$0 \leqslant -\chi(\mathcal{F}) = \operatorname{Swan}_{0}(\mathcal{F}) + \sum_{x \in \mathbb{G}_{m}} \chi_{x}(\mathcal{F}) + \sum_{s} N_{s}(\mathcal{F})s - \sum_{k \ge 1} U_{k}(\mathcal{F})$$

and

$$\sum_{k \ge 1} k U_k(\mathcal{F}) \leqslant \operatorname{rank}(\mathcal{F}) = \sum_s N_s(\mathcal{F}) \leqslant \sum_s N_s 2 \max(s, 1/2)$$

Subtracting twice the first inequality plus the second inequality from our identity, we must have

$$-U_1(\mathcal{F}) - \sum_{k \ge 5} (k-4)U_k(\mathcal{F}) \ge 2\operatorname{Swan}_0(\mathcal{F}) + 2\sum_{x \in \mathbb{G}_m} \chi_x(\mathcal{F}) \ge 0.$$

Thus, for the identity to hold, every inequality must be sharp. It follows that we must have $\chi(\mathfrak{F}) = 0, \sum_{k \ge 1} k U_k(\mathfrak{F}) = \operatorname{rank}(\mathfrak{F})$ (meaning the local monodromy at 0 is unipotent), $U_k(\mathfrak{F}) = 0$ for $k \neq 2, 3, 4, \chi_x(\mathcal{F}) = 0$ for $x \in \mathbb{G}_m$ (meaning \mathcal{F} is lisse on \mathbb{G}_m , and $2 \max(s, 1/2) = 0$ for all s with $s \leq 0$ (meaning all slopes at $\infty \leq 1/2$).

Lemma 8.11. Under the assumptions of Lemma 8.10, the local monodromy of $\mathcal{K} \star \mathcal{L}$ at 0 is unipotent.

Proof. We will calculate the wild part and the χ -tensor unipotent part of the local monodromy at 0 for each nontrivial character χ using [RL13]'s local convolution theory, and show that they are trivial, deducing that the local monodromy is unipotent.

By Lemma 8.10, \mathcal{F} is lisse on \mathbb{G}_m and tame at 0, so $\mathcal{F}_{(s)}$ vanishes for all $s \in \mathbb{G}_m$ and $\mathcal{F}_{(0)}^w = 0$. Similarly, \mathcal{L} is lisse on \mathbb{G}_m and tame at ∞ , so $\mathcal{L}_{(s)}$ vanishes for all $s \in \mathbb{G}_m$ and $\mathcal{L}_{(\infty)}^w = 0$. So the only term in [RL13, Theorem 16] that could be nonvanishing is

$$\overline{\varrho}_{\infty,0}(\mathcal{F}^w_{(\infty)},\mathcal{L}^w_{(0)}).$$

But by Lemma 8.10, $\mathcal{F}^w_{(\infty)}$ has all slopes < 1/2, and $\mathcal{L}^w_{(0)}$ has slope 1/2, so by [RL13, Proposition [13] we have

$$\overline{\varrho}_{\infty,0}(\mathcal{F}^w_{(\infty)},\mathcal{L}^w_{(0)})=0$$

as well.

For χ , because \mathcal{F} is unipotent at $0, \mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2$ is unipotent at 0, and because \mathcal{F} has no slope exactly equal to 1/2 at ∞ , $\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2$ is totally wild at infinity, so $\mathcal{L}_{\chi} \otimes U_k$ does not appear in the local monodromy of $\mathcal{F} \otimes [\times \alpha]^* \mathcal{K} \ell_2$ at 0 or ∞ for χ nontrivial. Thus $P_{\mathcal{K},\chi}(T)$ is a constant, and by the same reasoning $P_{\mathcal{L},\chi}(T)$ is a constant, so their product is a constant, and thus by [RL13, Proposition 28] $\mathcal{L}_{\chi} \otimes U_k$ does not appear in the local monodromy of $\mathcal{K} \star \mathcal{L}$ at zero or infinity.

Lemma 8.12. Under the assumptions of Lemma 8.10, the local monodromy of $\mathcal{K}' \star \mathcal{L}'$ at β'/α' is not unipotent.

Proof. Assume for contradiction that this representation is unipotent. We have

$$\mathcal{K}' = \mathcal{K}' \star ([y \to \beta'/y]^* \mathcal{K}\ell_2 \star [y \to \beta'y]^* \mathcal{K}\ell_2) = \mathcal{K}' \star \mathcal{L}' \star [\times \beta']^* \mathcal{K}\ell_2$$

Using this identity, we apply [RL13]'s local convolution theory to calculate the slope 1/2 part of the local monodromy of \mathcal{K}' at ∞ . Because $[\times \beta']^* \mathcal{K} \ell_2$ is lisse away from 0 and ∞ and tame at 0, the only functors that contribute to the wild part of \mathcal{K}' at ∞ are $\varrho_{(0,\infty)}, \varrho_{(\beta'/\alpha',\infty)}$, and $\varrho_{(\infty,\infty)}$. Because the unique slope at ∞ of $[\times \beta']^* \mathcal{K}\ell_2$ is 1/2, by [RL13] the image of $\varrho_{(0,\infty)}$ has slopes < 1/2and by [RL13, Proposition 11], the image of $\rho_{(\infty,\infty)}$ has slopes < 1/2. So the slope 1/2 part only arises from

$$\varrho_{(\beta'/\alpha',\infty)}\left((\mathcal{K}'\star\mathcal{L}')_{(\beta'/\alpha')},[\times\beta']^*\mathcal{K}\ell_{2(\infty)}^{w}\right).$$

Because $\varrho_{(\beta'/\alpha',\infty)}$ is exact and $(\mathcal{K}'\star\mathcal{L}')_{(\beta'/\alpha')}$ is unipotent, this is an iterated extension of

$$\varrho_{(\beta'/\alpha',\infty)}\left(1,[\times\beta']^*\mathcal{K}\ell_{2(\infty)}^w\right) = [\times\alpha']^*\mathcal{K}\ell_{2(\infty)}.$$

(This identity can be checked by applying local convolution functors to calculate the local monodromy at ∞ of $\delta_{\beta'/\alpha'} \star [\times\beta']^* \mathcal{K}\ell_2$, say.)

Then the slope 1/2 part of the local monodromy of $(\mathcal{F} \otimes [\times \alpha']^* \mathcal{K} \ell_2)$ at ∞ is an iterated extension of $[\times \alpha']^* \mathcal{K} \ell_2$, hence $[\times \alpha']^* \mathcal{K} \ell_2$ tensor a unipotent representation.

Let V be the local monodromy representation of \mathcal{F} at ∞ . By Lemma 8.10 and assumption (SL), V has all slopes < 1/2, so $V \otimes [\times \alpha']^* \mathcal{K} \ell_2$ has all slopes exactly 1/2. Thus $V \otimes [\times \alpha']^* \mathcal{K} \ell_2$ is $[\times \alpha']^* \mathcal{K} \ell_2$ tensored with a unipotent representation. Because V is a summand of

$$V \otimes ([\times \alpha']^* \mathcal{K}\ell_2 \otimes [\times \alpha']^* \mathcal{K}\ell_2) = (V \otimes [\times \alpha']^* \mathcal{K}\ell_2) \otimes [\times \alpha']^* \mathcal{K}\ell_2,$$

it is a summand of

$$[\times \alpha']^* \mathcal{K}\ell_2 \otimes [\times \alpha']^* \mathcal{K}\ell_2$$

tensored with a unipotent representation, and because it has slope < 1/2, it is a summand of the slope < 1/2 part of $[\times \alpha']^* \mathcal{K} \ell_2 \otimes [\times \alpha']^* \mathcal{K} \ell_2$ tensored with a unipotent representation. The slope < 1/2 part of $[\times \alpha']^* \mathcal{K} \ell_2 \otimes [\times \alpha']^* \mathcal{K} \ell_2$ is the sum of a trivial representation and quadratic character, so V is a summand of the sum of a unipotent representation and another unipotent representation tensored with a quadratic character.

Hence V is a sum of unipotent representations and unipotent representations tensored with the quadratic character. In particular, \mathcal{F} is tamely ramified at ∞ . So because it is irreducible and lisse on \mathbb{G}_m , tamely ramified at ∞ , and unipotent at 0, it must be the trivial sheaf, which contradicts our assumption that \mathcal{F} is nontrivial.

Proof of Proposition 8.4. It is clear from their constructions that the conductors of \mathcal{Z} and \mathcal{Z}' are $O(c(\mathcal{F})^{O(1)})$. By Lemma 8.5, \mathcal{Z}' is irreducible. So we get the bound of Proposition 8.4 unless \mathcal{Z}' is geometrically isomorphic to $[+\delta]^*\mathcal{Z}$. Suppose this is the case.

By Lemma 8.7, $[+\delta]^*\mathcal{Z}$ has a singularity at $-\delta$ which should be a singularity of \mathcal{Z}' . Since $\delta \neq 0, \infty$, Lemma 8.6 (applied to \mathcal{Z}') imply that $-\delta = \beta'/\alpha'$ and the dimension of the monodromy invariant of \mathcal{Z} and \mathcal{Z}' respectively at 0 and β'/α' are equal. By Lemma 8.11, the local monodromy of $[+\delta]^*\mathcal{Z}$ at $-\delta$ is unipotent while by Lemma 8.12, the local monodromy of \mathcal{Z}' at β'/α' is not unipotent, a contradiction.

9. Some special cases

In this section we discuss three cases of functions K for which Theorem 1.1 does not apply directly either because K is not a trace function or is a trace function associated to a "bad" sheaf \mathcal{F} .

9.1. A general duality principle. We start with the general duality principle hinted in the introduction. Let $K : \mathbf{F}_q^{\times} \to \mathbf{C}$ be some function; we define the Dirichlet series

$$L(\varphi \times f \times K, s) := \sum_{\substack{n, r \geqslant 1 \\ (nr^2, q) = 1}} \frac{\lambda(r, n)\lambda_f(n)K(nr^2)}{(nr^2)^s}, \ \mathfrak{Re} s > 1.$$

We will show that $L(\varphi \times f \times K, s)$ admits analytic continuation to **C** and satisfies a functional equation. We then use this functional equation to obtain non-trivial bounds for $S_V^t(\text{Kl}_2, X)$ when the sheaf \mathcal{F} is $\mathcal{K}\ell_2$.

9.1.1. The functional equation for standard L-functions. Let φ and f be GL₃ and GL₂ cusp forms of level one; to simplify the exposition and the shape of the functional equation we assume that f is holomorphic of some weight $k \ge 2$. The Rankin–Selberg L-function $L(\varphi \times f, s)$ has analytic continuation to **C** and satisfies the functional equation

$$\Lambda(\varphi \times f, s) = \varepsilon(\varphi \times f)\Lambda(\overline{\varphi} \times f, 1 - s)$$

where

$$\varepsilon(\varphi \times f) = i^{3k} = i^k = \pm 1$$

is the root number;

$$\Lambda(\varphi \times f, s) = L_{\infty}(\varphi \times f, s)L(\varphi \times f, s)$$

is the *L*-function completed by the Archimedean local factor

$$L_{\infty}(\varphi \times f, s) = \prod_{\substack{i=1,2\\j=1,2,3}} \Gamma_{\mathbf{R}}(s - \mu_{f,i} - \mu_{\varphi,j}), \ \Gamma_{\mathbf{R}}(s) = \pi^{-s/2} \Gamma(s/2)$$

with

$$\mu_{f,1} = -\frac{k-1}{2}, \ \mu_{f,2} = -\frac{k}{2}$$

and $(\mu_{\varphi,j})_{j=1,2,3}$ being the Langlands parameters (of the principal series representation attached to φ) which satisfy

$$\mu_{\varphi,1} + \mu_{\varphi,2} + \mu_{\varphi,3} = 0$$

(and belong to $(i\mathbf{R})^3$ under the Ramanujan–Petersson conjecture); $\overline{\varphi}$ denotes the automorphic form dual to φ (attached to the contragredient automorphic representation of φ) with parameters $(\overline{\mu}_{\varphi,j})_{j=1,2,3}$ and Hecke eigenvalues $(\overline{\lambda(r,n)})_{(r,n)}$. Under our assumption, $L_{\infty}(\varphi \times f,s)$ and $L_{\infty}(\overline{\varphi} \times f,s)$ have no poles for $\Re \mathfrak{e} s > -1/2 + \theta_3 = -\frac{1}{7}$.

For $\chi \pmod{q}$ a non-trivial Dirichlet character, the twisted *L*-function $L(\varphi \times f \times \chi, s)$ has analytic continuation to **C** and satisfies the functional equation

(9.1)
$$\Lambda(\varphi \times f \times \chi, s) = \varepsilon(\varphi \times f \times \chi)\Lambda(\overline{\varphi} \times f \times \overline{\chi}, 1-s)$$

where

$$\Lambda(\varphi \times f \times \chi, s) = q^{3s} L_{\infty}(\varphi \times f, s) L(\varphi \times f \times \chi, s),$$

and

$$\varepsilon(\varphi \times f \times \chi) = \varepsilon_{\chi}^6 \varepsilon(\varphi \times f)$$

with

$$\varepsilon_{\chi} = q^{-1/2} \sum_{x \in \mathbf{F}_q^{\times}} \chi(x) e(\frac{x}{q})$$

being the normalized Gauss sum (notice that the Archimedean local factor does not depend on (the parity of) χ because f is holomorphic; this is the main reason why we have made the simplifying assumption).

9.2. Functional equation for algebraically twisted *L*-functions. Let $K : \mathbf{F}_q^{\times} \to \mathbf{C}$ be a function on \mathbf{F}_q^{\times} (extended by 0 to \mathbf{F}_q); we define the Mellin transform for $\chi \pmod{q}$

$$\widetilde{K}(\chi) = \frac{1}{(q-1)^{1/2}} \sum_{x \in \mathbf{F}_q^{\times}} K(x) \overline{\chi}(x)$$

and we define for (n,q) = 1 the "GL₆-transform" of K as

(9.2)
$$\widetilde{K}^{6}(n) = \frac{1}{q^{1/2}} \sum_{\substack{x \in \mathbf{F}_{q}^{\times} \\ 40}} \mathrm{Kl}_{6}(nx;q) K(x)$$

where Kl_6 is the hyper-Kloosterman sum in six variables

$$\mathrm{Kl}_{6}(x;q) = \frac{1}{q^{5/2}} \sum_{\substack{x_{1},\cdots,x_{6} \in \mathbf{F}_{q}^{\times} \\ x_{1},\cdots,x_{6} = x}} e(\frac{x_{1}+\cdots+x_{6}}{q}).$$

Proposition 9.1. The completed series

$$\Lambda(\varphi \times f \times K, s) = q^{3s} L_{\infty}(\varphi \times f, s) L(\varphi \times f \times K, s)$$

has analytic continuation to C and satisfies the functional equation

$$\begin{split} \Lambda(\varphi \times f \times K, s) &= \varepsilon(\varphi \times f) \Lambda(\overline{\varphi} \times f \times \widecheck{K}^{0}, 1-s) + \\ &+ \frac{q^{3s}}{(q-1)^{1/2}} \widetilde{K}(\chi_{0}) \Lambda^{(q)}(\varphi \times f, s) \\ &- \frac{\varepsilon(\varphi \times f) q^{-3s}}{(q-1)^{1/2}} \widetilde{K}(\chi_{0}) \Lambda^{(q)}(\overline{\varphi} \times f, 1-s) \end{split}$$

where $\chi_0 = 1_{|\mathbf{F}_q^{\times}|}$ denotes the trivial character modulo q and $\Lambda^{(q)}(\cdots)$ denotes the complete L-function with the local factor at q being removed.

Proof. By the Mellin inversion formula

$$K(nr^{2}) = \frac{1}{(q-1)^{1/2}} \sum_{\chi \pmod{q}} \widetilde{K}(\chi)\chi(nr^{2}),$$

we have

$$\Lambda(\varphi \times f \times K, s) = \frac{q^{3s}}{(q-1)^{1/2}} \widetilde{K}(\chi_0) \Lambda^{(q)}(\varphi \times f, s) + \frac{1}{(q-1)^{1/2}} \sum_{\substack{\chi \pmod{q} \\ \chi \neq \chi_0}} \widetilde{K}(\chi) \Lambda(\varphi \times f \times \chi, s)$$

and the result follows by applying the functional equations (9.1), the identity

$$\widetilde{K}^{6}(nr^{2}) = \frac{1}{(q-1)^{1/2}} \sum_{\chi \pmod{q}} \varepsilon_{\chi}^{6} \widetilde{K}(\chi) \overline{\chi}(nr^{2}),$$

and the value of the Ramanujan sum $\varepsilon_{\chi_0} = -q^{-1/2}$.

By standard contour shifts we deduce the following.

Corollary 9.2. Let V be a smooth compactly supported function satisfying (1.2); we have for any $X \ge 1$

$$\sum_{(nr,q)=1} \lambda(r,n)\lambda_f(n)K(nr^2)V(\frac{nr^2}{X}) = \varepsilon(\varphi \times f)\frac{X}{q^3} \sum_{(nr,q)=1} \overline{\lambda(r,n)}\lambda_f(n)\widecheck{K}^6(nr^2)\widecheck{V}^6(\frac{nr^2}{q^6/X}) + O(Z^B \frac{|\widetilde{K}(\chi_0)|}{q^{1/2-\theta_3}})$$

where

$$\widetilde{V}^{6}(x) = \frac{1}{2i\pi} \int_{(3/2)} \widetilde{V}(1-s) \frac{L_{\infty}(\overline{\varphi} \times f, s)}{L_{\infty}(\varphi \times f, 1-s)} x^{-s} ds$$

and $\tilde{V}(s) = \int_0^\infty V(y) y^s \frac{dy}{y}$ is the Mellin transform of V. Here $\theta_3 = 5/14$ is the best known bound towards the Ramanujan–Petersson conjecture on GL₃, $B \ge 0$ is an absolute constant (any B > 5/2will do) and the implicit constant in the $O(\cdots)$ term depends on φ , f and the implicit constants in (1.2).

Proof. We have

$$V(x) = \frac{1}{2i\pi} \int \widetilde{V}(s) x^{-s} ds$$

(the integration is along the vertical line $\Re \mathfrak{e} s = 1 + 1/14$) so that

$$\sum_{(nr,q)=1} \lambda(r,n)\lambda_f(n)K(nr^2)V(\frac{nr^2}{X}) = \frac{1}{2i\pi} \int \frac{\Lambda(\varphi \times f \times K,s)}{L_{\infty}(\varphi \times f,s)} \widetilde{V}(s)(\frac{X}{q^3})^s ds$$
$$= q^3 \varepsilon(\varphi \times f) \frac{1}{2i\pi} \int L(\overline{\varphi} \times f \times \widetilde{K}^6, 1-s) \frac{L_{\infty}(\overline{\varphi} \times f, 1-s)}{L_{\infty}(\varphi \times f,s)} \widetilde{V}(s)(\frac{X}{q^6})^s ds$$
$$+ \frac{\widetilde{K}(\chi_0)}{(q-1)^{1/2}} \frac{1}{2i\pi} \int L^{(q)}(\varphi \times f,s) \widetilde{V}(s)X^s ds$$
$$- \frac{\varepsilon(\varphi \times f)\widetilde{K}(\chi_0)}{(q-1)^{1/2}} \frac{1}{2i\pi} \int L^{(q)}(\overline{\varphi} \times f, 1-s) \frac{L_{\infty}(\overline{\varphi} \times f, 1-s)}{L_{\infty}(\varphi \times f,s)} \widetilde{V}(s)(\frac{X}{q^6})^s ds$$

upon applying Proposition 9.1. In the first integral we make the change of variable $s \leftrightarrow 1-s$ getting

$$\varepsilon(\varphi \times f) \frac{X}{q^3} \frac{1}{2i\pi} \int_{(-1/14)} L(\overline{\varphi} \times f \times \widecheck{K}^6, s) \frac{L_{\infty}(\overline{\varphi} \times f, s)}{L_{\infty}(\varphi \times f, 1-s)} \widetilde{V}(1-s) (\frac{X}{q^6})^{-s} ds$$

and shifting the contour back to $\Re \mathfrak{e} s = 3/2$ without hitting any poles we obtain the first sum. For the second and third integrals we shift the contour to $\Re e s = 0$ and apply trivial bounds; the term q^{θ_3} arise from the following bound for the inverse of the local factor

$$L_q(\varphi \times f, s)^{-1} \ll q^{\theta_3}, \ \mathfrak{Re} s = 0.$$

In particular we deduce that

(9.3)
$$\sum_{n,r} \lambda(r,n)\lambda_f(n)K(nr^2)V(\frac{nr^2}{X}) \ll_{\varphi,f} Z^B q^{3+o(1)} \left(\|\breve{K}^6\|_{\infty} + \|K\|_{\infty} q^{\theta_3-3} \right).$$

for some absolute constant $B \ge 0$ and the implicit constant depends on φ, f and the implicit constants in (1.2). This bound is stronger than the trivial bound $O(X^{1+o(1)})$ as long as

$$X \geqslant q^{3+\delta}, \ \delta > 0$$

and one can use it to prove the "convexity bound" (1.1) mentioned in the introduction.

Applying Theorem 1.1 to the sum

$$\sum_{n,r} \overline{\lambda(r,n)} \lambda_f(n) \widecheck{K}^6(nr^2) \widecheck{V}^6(\frac{nr^2}{q^6/X})$$

(after a dyadic partition of unity), one deduces that

Corollary 9.3. Notations being as in the above corollary; suppose that the transform \widecheck{K}^6 defined in (9.2) is L^1 -close to the trace function K' of a good sheaf \mathcal{F}' , then we have

$$\begin{split} S_V^t(X) \ll Z^{B'} q^{o(1)} \frac{X}{q^3} \bigg((q^6/X)^{3/4} q^{11/16} + (q^6/X)^{\frac{2-\theta_3}{3-2\theta_3}} q^{\frac{11(1-\theta_3)}{4(3-2\theta_3)}} + (q^6/X) q^{-1/8}) \\ &+ \| \widecheck{K}^6 - K' \|_1 (\frac{q^5}{X} + 1) (\frac{q^6}{X})^{\theta_3} \bigg). \end{split}$$

Here $\theta_3 = 5/14$ is the best known bound towards the Ramanujan-Petersson conjecture on GL₃; $B' \ge 0$ is an absolute constant,

$$\|\breve{K}^{6} - K'\|_{1} := \sum_{x \in \mathbf{F}_{q}^{\times}} |\breve{K}^{6}(x) - K'(x)|,$$

and the implicit constant depends on φ , f, $|\tilde{K}(\chi_0)|$, $C(\mathcal{F}')$ and the implicit constants in (1.2).

Remark 9.4. Ignoring the Z parameter, the above bound is non-trivial as long as

$$X \gg q^{3-1/12+\eta}, \ \eta > 0$$

and for $X = q^3$ we obtain

$$S_V^t(q^3) \ll q^{3-1/16+o(1)}.$$

The typical situation to apply Corollary 9.3 is when K is the trace function attached to a geometrically non-trivial and irreducible sheaf \mathcal{F} of weight 0. Unless \mathcal{F} is geometrically isomorphic to a sheaf of the shape $[x \mapsto \alpha x^{-1}]^* \mathcal{K}\ell_6$ for some $\alpha \neq 0$, the function \breve{K}^6 is close to the trace function, K' say, of a geometrically irreducible sheaf of weight 0, namely the convolution sheaf

$$\check{\mathcal{F}}^{0} := \mathcal{K}\ell_{6} \star [x \mapsto x^{-1}]^{*}\mathcal{F}$$

(alternatively $\check{\mathcal{F}}^6$ is the sheaf obtained from \mathcal{F} by applying 6 times the composition of $[x \mapsto x^{-1}]^*$ and the geometric Fourier transform). In such a case one has

$$|\widetilde{K}(\chi_0)|, \ \|\widecheck{K}^{0} - K'\|_1 = O_{C(\mathcal{F})}(1).$$

We discuss three explicit examples below.

9.3. Arithmetic progressions. One of the motivations for investigating algebraic twists of *L*-functions is the level of distribution of their coefficients in arithmetic progressions of large modulus; specifically let

$$\lambda_F(m) = \sum_{nr^2=m} \lambda(r,n)\lambda_f(n);$$

one would like to prove that for (a, q) = 1 and V satisfying (1.2), one has

(9.4)
$$\sum_{m \equiv a \pmod{q}} \lambda_F(m) V(\frac{m}{X}) = o_{\varphi, f, V}(\frac{X}{q})$$

for $q = X^{\vartheta}$ and some $\vartheta > 0$ (a level of distribution) as large as possible (since φ and f are cuspidal there should be no main term).

Taking

$$K(n) = q^{1/2} \delta_{n \equiv a \, (\mathrm{mod} \, q)}$$

we have for (m, q) = 1

$$\widecheck{K}^{0}(m) = \mathrm{Kl}_{6}(am;q);$$

therefore applying Corollary 9.2 and bounding the resulting sums trivially (using Deligne's bound $|\operatorname{Kl}_6(am;q)| \leq 6$) we obtain

$$\sum_{nr^2 \equiv a \pmod{q}} \lambda(r, n) \lambda_f(n) V(\frac{nr^2}{X}) \ll_{V, \varphi, f} q^{5/2 + o(1)}$$

which gives (9.4) as long as

$$\vartheta < \vartheta_6 = 2/7.$$
43

To go beyond $\vartheta_6 = 2/7$ we would need, at least, to be able to improve the trivial bound for the sum

$$\sum_{n,r} \overline{\lambda(r,n)} \lambda_f(n) \operatorname{Kl}_6(anr^2;q) \widecheck{V}^6(\frac{nr^2}{X'})$$

for $X' = q^6/X$ a bit less than $q^{5/2}$. While $\text{Kl}_6(an; q)$ is definitely "good" (its geometric monodromy group is SL₆ and its slopes at ∞ are equal to 1/6), Theorem 1.1 is non-trivial only for

$$X' \gg q^{5/2+1/4+\delta}, \ \delta > 0$$

This demonstrates the importance and difficulty of detecting cancellation for algebraically twisted sums of shorter length.

Remark 9.5. This level of distribution $\vartheta_6 = 2/7$ corresponds to Selberg's level of distribution $\vartheta_2 = 2/3$ for the divisor function (or the Fourier coefficients of modular forms) or $\vartheta_3 = 1/2$ for the ternary divisor function (see [FI85, HB86, FKM15b] for improvements in this last case).

Remark 9.6. An Archimedean analog of this question is to improve the *Landau's type* bound for the sharp-cut sum

$$\sum_{m \leqslant X} \lambda_F(m) \ll X^{5/7 + o(1)}$$

(cf. [FI05, Prop. 1.1] for the degree m = 6 in the notations of that paper); under the Ramanujan–Petersson conjecture, this amounts to non-trivial estimates for

$$\sum_{m \leqslant X'} \overline{\lambda_F(m)} \, e(t(m/X')^{1/6})$$

when X' is a bit less than $t^{5/2}$. This has recently been worked out by the first named author and Q. Sun in [LS21] and this might suggest that if the modulus q is a suitably factorable integer (either a smooth number or a large power of a fixed prime number) other exponential sum methods might allow to pass the 2/7 barrier for such a composite modulus.

9.4. Additive characters. We now discuss Theorem 1.1 when K is a trace function of a sheaf \mathcal{F} which is not Fourier: this implies that for q large enough (depending on $C(\mathcal{F})$) K is proportional to an additive character $\psi(\bullet) = e(\frac{a\bullet}{q})$.

In this case, a much more general bound is expected (cf. [Mil06]): for any $\alpha \in \mathbf{R}$, one should have

(9.5)
$$\sum_{n,r} \lambda(r,n)\lambda_f(n)e(\alpha nr^2)V(\frac{nr^2}{X}) \ll_{\varphi,f,V} X^{1-\eta+o(1)}$$

for some $\eta > 0$ (possibly $\eta = 1/2$) and this should hold uniformly for $\alpha \in \mathbf{R}$. Such a bound would be analogous to Wilton's bound for Fourier coefficients of GL_2 automorphic forms or Miller's bound for GL_3 automorphic forms; however for ranks greater than 3 such bounds are still unknown.

As was pointed out in [Mil06] if α is a rational number the analytic properties of $L(\varphi \times f \times \chi, s)$ for χ a Dirichlet character of modulus dividing the denominator of α yield (9.5) at least if X is large compared to q. For instance, for q a prime number, (an, q) = 1 we have

$$\widetilde{e(\frac{a\bullet}{q})}^6(n) = \mathrm{Kl}_5(-\overline{a}n;q) - (-1)^5 q^{-3}$$

and Corollary 9.2 together with Deligne's bound $|Kl_5(an;q)| \leq 5$ yield (9.5) as long as

$$X \geqslant q^{3+\delta}, \ \delta > 0.$$

Now since the sheaf $[x \mapsto -\overline{a}x]^* \mathcal{K}\ell_5$ is good (it is Fourier, its geometric monodromy group is SL_5 and all its ∞ -slopes equal 1/5) we have

$$\sum_{r,n} \lambda(r,n)\lambda_f(n)e(\frac{anr^2}{q})V(\frac{nr^2}{q^3}) \ll q^{2+\theta_3+o(1)} + q^{3-1/16+o(1)}$$

(the first term on the right is the contribution of the n's divisible by q) so that (9.5) holds in the slightly wider range

$$X \geqslant q^{3-1/16+\delta}, \ \delta > 0.$$

9.5. Kloosterman sums. Finally we observe that the Kloosterman sheaf $\mathcal{K}\ell_2$ while being Fourier is definitely not good: neither (SL) nor (MO) is satisfied. On the other hand,

$$\widetilde{\mathrm{Kl}}_{2}^{6}(n) = \mathrm{Kl}_{4}(n;q) - q^{-5/2} - q^{-7/2}$$

and the Kloosterman sheaf $\mathcal{K}\ell_4$ is good: its geometric monodromy group is SL₄ and all its ∞ -slopes are 1/4. By Corollary 9.2 we have

$$\sum_{r,n} \lambda(r,n)\lambda_f(n) \operatorname{Kl}_2(nr^2;q) V(\frac{nr^2}{q^3}) \ll_{\varphi,f,V} q^{3-1/16+o(1)}$$

which again is non-trivial as long as

$$X \geqslant q^{3-1/16+\delta}, \ \delta > 0.$$

Remark 9.7. Further examples of geometrically irreducible sheaves satisfying neither (MO) nor (SL) are the sheaves of the shape

$$\bigotimes_{i=1}^{r} \operatorname{Sym}_{k_{i}} \circ [\times \lambda_{i}]^{*} \mathcal{K} \ell_{2}$$

where Sym_k denotes the k-th symmetric power representation of SL_2 and λ_i , $i = 1, \dots, r$ are distinct elements of \mathbf{F}_q^{\times} . The underlying trace function is

$$K: x \in \mathbf{F}_q^{\times} \mapsto \prod_{i=1}^r \operatorname{sym}_{k_i}(\theta_{\lambda_i x})$$

where $2\cos(\theta_{\lambda_i x}) = \text{Kl}_2(\lambda_i x)$ and $\text{sym}(k\theta) = \sin((k+1)\theta)/\sin(\theta)$. In that case all the non-trivial slopes of the monodromy at ∞ are 1/2 and the set of λ 's for which the condition in (MO) fails are precisely the λ_i . It should be possible to check for most or all of these sheaves \mathcal{F} that $\overset{\sim}{\mathcal{F}}^6$ is good.

References

- [AHLS20] K. Aggarwal, R. Holowinsky, Y. Lin, and Q. Sun, The Burgess bound via a trivial delta method, Ramanujan J. 53 (2020), no. 1, 49–74.
 - [Blo12] V. Blomer, Subconvexity for twisted L-functions on GL(3), Amer. J. Math. 134 (2012), no. 5, 1385–1421.
 - [BK19] V. Blomer and R. Khan, Twisted moments of L-functions and spectral reciprocity, Duke Math. J. 168 (2019), no. 6, 1109–1177.
 - [CI00] J. B. Conrey and H. Iwaniec, The cubic moment of central values of automorphic L-functions, Ann. of Math. (2) 151 (2000), no. 3, 1175–1216.
 - [DFI93] W. Duke, J. Friedlander, and H. Iwaniec, Bounds for automorphic L-functions, Invent. Math. 112 (1993), no. 1, 1–8.
- [FKM14] É. Fouvry, E. Kowalski, and Ph. Michel, Algebraic trace functions over the primes, Duke Math. Journal 163 (2014), no. 9, 1683-1736.
- [FKM15a] É. Fouvry, E. Kowalski, and Ph. Michel, Algebraic twists of modular forms and Hecke orbits, GAFA 25 (2015), no. 2, 580-657.
- [FKM15b] É. Fouvry, E. Kowalski, and Ph. Michel, On the exponent of distribution of the ternary divisor function, Mathematika 61 (2015), no. 1, 121–144.

- [FKMS19] É. Fouvry, E. Kowalski, Ph. Michel, and W. Sawin, Lectures on applied l-adic cohomology, Analytic methods in arithmetic geometry, Contemp. Math., vol. 740, Amer. Math. Soc., [Providence], RI, 2019, pp. 113–195.
 - [FI85] J. B. Friedlander and H. Iwaniec, Incomplete Kloosterman sums and a divisor problem, Ann. of Math. (2) 121 (1985), no. 2, 319–350. With an appendix by Bryan J. Birch and Enrico Bombieri.
 - [FI05] J. B. Friedlander and H. Iwaniec, Summation formulae for coefficients of L-functions, Canad. J. Math. 57 (2005), no. 3, 494–505.
 - [Gol06] D. Goldfeld, Automorphic forms and L-functions for the group $GL(n, \mathbf{R})$, Cambridge Studies in Advanced Mathematics, vol. 99, Cambridge University Press, Cambridge, 2006. With an appendix by Kevin A. Broughan.
 - [HB86] D. R. Heath-Brown, The divisor function $d_3(n)$ in arithmetic progressions, Acta Arith. 47 (1986), no. 1, 29–56.
 - [HB96] _____, A new form of the circle method, and its application to quadratic forms, J. Reine Angew. Math. 481 (1996), 149–206.
- [HMQ16] R. Holowinsky, R. Munshi, and Z. Qi, Character sums of composite moduli and hybrid subconvexity, Advances in the theory of automorphic forms and their L-functions, Contemp. Math., vol. 664, Amer. Math. Soc., Providence, RI, 2016, pp. 135–148.
 - [HN18] R. Holowinsky and P. D. Nelson, Subconvex bounds on GL₃ via degeneration to frequency zero, Math. Ann. 372 (2018), no. 1-2, 299–319.
 - [Kat88] N. M. Katz, Gauss sums, Kloosterman sums, and monodromy groups, Annals of Mathematics Studies, vol. 116, Princeton University Press, Princeton, NJ, 1988.
 - [Kat90] _____, Exponential sums and differential equations, Annals of Mathematics Studies, vol. 124, Princeton University Press, Princeton, NJ, 1990.
 - [KS03] H. H. Kim and P. Sarnak, Refined estimates towards the Ramanujan and Selberg conjectures, J. Amer. Math. Soc. 16 (2003), no. 1, 139–183. Appendix to H. Kim, Functoriality for the exterior square of GL₄ and the symmetric fourth of GL₂.
- [KLMS20] E. Kowalski, Y. Lin, Ph. Michel, and W. Sawin, Periodic twists of GL₃-automorphic forms, Forum Math. Sigma 8 (2020), Paper No. e15, 39.
 - [Lin21] Y. Lin, Bounds for twists of GL(3) L-functions, J. Eur. Math. Soc. (JEMS) 23 (2021), no. 6, 1899–1924.
 - [LS21] Y. Lin and Q. Sun, Analytic twists of $GL_3 \times GL_2$ automorphic forms, Int. Math. Res. Not. IMRN 19 (2021), 15143–15208.
- [LWY05] J. Liu, Y. Wang, and Y. Ye, A proof of Selberg's orthogonality for automorphic L-functions, Manuscripta Math. 118 (2005), no. 2, 135–149.
- [Mil06] S. D. Miller, Cancellation in additively twisted sums on GL(n), Amer. J. Math. 128 (2006), no. 3, 699–729.
- [MS06] S. D. Miller and W. Schmid, Automorphic distributions, L-functions, and Voronoi summation for GL(3), Ann. of Math. (2) 164 (2006), no. 2, 423–488.
- [Mun15] R. Munshi, The circle method and bounds for L-functions—III: t-aspect subconvexity for GL(3) Lfunctions, J. Amer. Math. Soc. 28 (2015), no. 4, 913–938.
- [Mun21] _____, Twists of GL(3) L-functions, In: Müller W., Shin S.W., Templier N. (eds) Relative Trace Formulas, Simons Symposia, Springer, Cham, 2021, pp. 351–378.
- [Mun22] _____, Subconvexity for GL(3) × GL(2) L-functions in t-aspect, J. Eur. Math. Soc. (JEMS) 24 (2022), no. 5, 1543–1566.
- [PY20] I. Petrow and M. P. Young, The Weyl bound for Dirichlet L-functions of cube-free conductor, Ann. of Math. (2) 192 (2020), no. 2, 437–486.
- [Qi20] Z. Qi, Theory of fundamental Bessel functions of high rank, Mem. Amer. Math. Soc. 267 (2020), no. 1303, vii+123.
- [RL13] A. Rojas-León, Local convolution of l-adic sheaves on the torus, Math. Z. 274 (2013), no. 3-4, 1211–1230.
- [Sha19] P. Sharma, Subconvexity for $GL(3) \times GL(2)$ twists in level aspect, Preprint (2019). arXiv:1906.09493.

EPFL/MATH/TAN, STATION 8, CH-1015 LAUSANNE, SWITZERLAND *Email address*: yongxiao.lin@epfl.ch

EPFL/MATH/TAN, STATION 8, CH-1015 LAUSANNE, SWITZERLAND *Email address*: philippe.michel@epfl.ch

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, 2990 BROADWAY, NEW YORK, NY 10027, USA *Email address:* sawin@math.columbia.edu