Mathematische Annalen (2024) 389:2783–2855
https://doi.org/10.1007/s00208-023-02713-8

Mathematische Annalen

On length sets of subarithmetic hyperbolic manifolds

Alex Kontorovich[1,2] · Xin Zhang[3]

Received: 17 March 2023 / Revised: 16 August 2023 / Accepted: 21 August 2023 /
Published online: 2 September 2023

Abstract

In this paper, we study the set of lengths of closed geodesics (or equivalently, the set of traces of the fundamental group) of a hyperbolic manifold. By "subarithmetic," we mean a manifold whose set of traces takes values in a ring of algebraic integers. For such, we formulate the "Asymptotic Length-Saturation Conjecture", which states that, under certain natural conditions, there is an asymptotic local–global principle for the trace set. We prove the first instance of the conjecture for punctured, Zariski dense covers of the modular surface.

1 Introduction

By the length spectrum of a hyperbolic manifold M, we mean the set of lengths of closed geodesics on M, with multiplicity. As is well-known, closed geodesics on M correspond to hyperbolic conjugacy classes of its fundamental group

$$\Gamma = \pi_1(M) < \mathrm{Isom}(\mathbb{H}^n) \cong PO(n,1),$$

and lengths of the former are a simple function of traces of the latter, namely, trace $= 2\cosh(\text{length}/2)$. It is also classical to study the length set, that is, the set of lengths of closed geodesics, now counted *without* multiplicity; again, this is intimately related

Kontorovich is partially supported by NSF Grant DMS-1802119, BSF Grant 2020119, and the 2020–2021 Distinguished Visiting Professorship at the National Museum of Mathematics. Zhang is partially supported by ECS Grant 27307320 and NSFC grant 12001457.

✉ Alex Kontorovich
alex.kontorovich@rutgers.edu

Xin Zhang
xzhang@maths.hku.hk

[1] Department of Mathematics, Rutgers University, New Brunswick, NJ, USA

[2] National Museum of Mathematics (MoMath), New York, NJ, USA

[3] Department of Mathematics, The University of Hong Kong, Pokfulam, Hong Kong

to the set $\mathcal{T}(\Gamma)$ of traces (without multiplicity) of hyperbolic conjugacy classes of Γ. In this paper, we initiate a detailed study of the latter for (sub)arithmetic manifolds, from the viewpoint of local–global theory. In particular, we produce a density one set of "admissible" traces for subgroups of the modular group containing parabolic elements, see Theorem 1.11.

To motivate our main results, we begin with a few illustrative examples.

Example 1 Consider the Hecke $(2, 3, \infty)$ triangle reflection group, or rather, its orientation preserving cover, the modular group $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. The trace set $\mathcal{T}(\Gamma)$ of the latter is elementarily seen to be all of $\mathbb{Z}$, as for any desired integer t, one simply expresses t as $t = a + d$ and factors $bc = ad - 1$ to make a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ having trace t. This is because Γ is an arithmetic (or better yet, congruence) group, and hence *any* solution to $ad - bc = 1$ over $\mathbb{Z}$ gives an element. We can compare these facts with the well-known count (see, e.g., [14, §15.8]) for the number of points in Γ in an archimedean ball B_N of radius N: as with any lattice in $\mathrm{SL}_2(\mathbb{R})$, we crudely have

$$\#\Gamma \cap B_N \sim cN^2, \tag{1.1}$$

for some constant $c > 0$; this means that the average number of times that a particular integer $t \asymp N$ arises as a trace of a matrix in B_N is of order $N^2/N = N$. But this does not take into account the fact that trace is a conjugacy class invariant. For $t > 2$, let $H(t)$ denote the number of conjugacy classes of elements in Γ with trace t. As is well-known (see, e.g., [11]), $H(t)$ is equal to $h(t^2 - 4)$, where $h(D)$ is the classical class number, that is, the number of equivalence classes of binary quadratic forms of discriminant D (not necessarily primitive). By the Prime Geodesic Theorem we have (compare to (1.1)):

$$\sum_{t<N} H(t) \sim \frac{N^2}{\log(N^2)}, \tag{1.2}$$

and so a "typical" value of $H(t)$ (for $t \asymp N$) is more like $N/\log N$, rather than N. (Note that the fundamental unit ϵ_D for discriminant $D = t^2 - 4$ is about as small as possible, $\epsilon_D = (t + \sqrt{D})/2$, and hence this class number is as large as possible, of size about $\sqrt{D}$). The discrepancy in counting matrices versus counting conjugacy classes makes sense, as the archimedean size of elements under conjugation grow exponentially (the stabilizer group of a conjugacy class is a discrete subgroup of some $\mathrm{SO}(1, 1)$), so $(\log N)$-many matrices of size N are grouped together. This is a minor issue here, but will play a major role in the next example.

Example 2 Now consider the Hecke $(2, 5, \infty)$ triangle group, or rather its cover, the group Γ generated by

$$\Gamma = \left\langle \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \phi \\ 0 & 1 \end{pmatrix} \right\rangle,$$

where $\phi = (1+\sqrt{5})/2$ is the golden mean. (Recall that $\mathrm{SL}_2(\mathbb{Z})$ has similar generators, except with ϕ replaced by 1.) The group Γ is nonarithmetic, but it is *subarithmetic*:

Definition 1.3 A manifold, or its fundamental group, is called **subarithmetic** if its set of traces takes values in a ring of algebraic integers.

In this case at hand, the ring is $\mathcal{O} = \mathbb{Z}[\phi]$. Note that Γ is a subgroup of the arithmetic group $\widetilde{\Gamma} := \mathrm{SL}_2(\mathbb{Z}[\phi])$. The latter does not act discretely on $\mathbb{H}$, but *is* a lattice in $\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$, where it acts by the Galois conjugate in the second factor. (The fact that Γ is a lattice in the first factor, while the second factor is non-compact, is one way to see its nonarithmeticity.) In fact, Γ is a *thin* group (see, e.g. [16]), as the set of $\mathbb{Z}$-points of its Zariski closure is exactly $\widetilde{\Gamma}$, and it has infinite index in the latter. The set of traces of $\widetilde{\Gamma}$ is again elementary to determine; it is the full order $\mathcal{O} = \mathbb{Z}[\phi]$, for the same reason as in Example 1. But now we may ask, which $t \in \mathcal{O}$ are also traces of Γ?

The asymptotic count (1.1) of *matrices* in a ball B_N (in $\mathbb{R}^4 \cong M_{2\times 2}(\mathbb{R})$) is still of order N^2, since Γ is a lattice in $\mathrm{SL}_2(\mathbb{R})$. But now $\mathcal{O} \cong \mathbb{Z} \oplus \phi\mathbb{Z}$ is a quadratic ring. In general, if $\mathcal{O}$ is an order in the ring of integers of a number field with k embeddings into $\mathbb{R}$ and ℓ embeddings into $\mathbb{C}$, let $\mathcal{O}_N$ denote the points of $\mathcal{O}$ in a fixed Euclidean norm-N ball in $\mathbb{R}^{k+2\ell}$ under the image of all the embeddings

$$\mathcal{O} \subset \mathbb{R} \times \cdots \times \mathbb{R} \times \mathbb{C} \times \cdots \times \mathbb{C}. \tag{1.4}$$

Returning to $\mathcal{O} = \mathbb{Z}[\phi]$, the number of elements in $\mathcal{O}_N$ is also roughly N^2, the same as the number of points of Γ in B_N.

Therefore the average number of matrices in B_N having a given trace $t \in \mathcal{O}$ is a positive constant.

But what happens when we group by conjugacy classes? (In Example 1, this caused the average count to drop by a factor of $\log N$, but here we don't have this factor to spare!) Let $H_\Gamma(t)$ denote the number of conjugacy classes of Γ having trace $t \in \mathcal{O}$. As Γ is a lattice in $\mathrm{SL}_2(\mathbb{R})$, we still have the Prime Geodesic Theorem (see (1.2)), that

$$\sum_{\mathbb{N}(t)<N^2} H_\Gamma(t) \sim \frac{N^2}{\log(N^2)}, \tag{1.5}$$

where $\mathbb{N} : \mathcal{O} \to \mathbb{Z}$ is the algebraic norm. Therefore there can't be more than $O(N^2/\log N)$ elements $t \in \mathcal{O}$ which actually arise as traces in $\Gamma \cap B_N$, and thus the density of those that do arise is zero! While we can't say much about the class number $H_\Gamma(t)$, in every conjugacy class that does arise, there should be about $\log N$ matrices of size N, as before. So when counting matrices, even though the "average" multiplicity is bounded, what's really going on is that 100% of the time, the multiplicity is exactly zero, and very rarely there are somewhat large (of size at least $\log N$) multiplicities. See also recent work of McMullen [21] in this direction. We remark that the number of elements of $\widetilde{\Gamma}$ in a ball B_N is roughly N^4, and $H_{\widetilde{\Gamma}}(t)$ is roughly of order t^2.

Example 3 For our last example, consider the Hecke $(2, 7, \infty)$ triangle group, or rather its cover, the group Γ generated by

$$\Gamma = \left\langle \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & \eta \\ 0 & 1 \end{pmatrix} \right\rangle,$$

where $\eta = 2\cos(\pi/7)$. The ring $\mathcal{O} = \mathbb{Z}[\eta]$ is cubic, and so the number of matrices of size N in Γ is of order N^2, while the number of possible values of the trace of size up to N is N^3, and hence it is clear that very few numbers in $\mathcal{O}$ can occur as traces.

Returning to the general setting, in light of these examples, to be able to say anything about traces of Γ, we need some conditions. First we assume that $\Gamma < \mathrm{PO}(n, 1)$ is discrete, finitely generated, and subarithmetic, so that $\mathrm{tr}(\Gamma)$ consists of algebraic integers. By the **trace ring**, $\mathcal{O}$, we mean the ring generated by the traces of Γ. Let $\mathcal{O}_N$ be as above (1.4).

Obstruction 1: Let $\alpha > 0$ be the "growth exponent" of Γ, in the sense that

$$\#\Gamma \cap B_N = N^{\alpha + o(1)}.$$

(When $\Gamma < \mathrm{PO}(n-1, 1)$ is geometrically finite and δ is the Hausdorff dimension of its limit set, then $\alpha = \delta$ [18, 25, 26].) As in Example 3 (and Example 2), to be able to study the length set, we require that α exceeds the rank of $\mathcal{O}$. One can think of this as an "archimedean local obstruction."

Obstruction 2: There are also potentially other local obstructions. Already in the case of a classical congruence group $\Gamma(q) := \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(q))$, only the numbers that are $2 (\mathrm{mod}\, q)$ can arise as traces in $\Gamma(q)$.

Definition 1.6 We say that $t \in \mathcal{O}$ is **admissible** if, for every ideal $\mathcal{I} \subset \mathcal{O}$, $t \in \mathcal{T}(\Gamma) \bmod \mathcal{I}$.

(We remark that Strong Approximation for Zariski dense groups [27] implies that it suffices to check a finite list of ideals to determine admissibility; determining this finite list in practice is often not difficult.)

Obstruction 3: There is one final archimedean local obstruction. Given any manifold, we can take a cover that destroys the systole; that is, what was the shortest closed geodesic need not remain closed under a cover, making the shortest length (that is, smallest trace) of such a moving target. So we should allow for some "small" values of $\mathcal{O}$ to not arise as traces.

We may now formulate our main conjecture.

Definition 1.7 With notation as above, we say that Γ **length-saturates** if: every admissible $t \in \mathcal{O}$ with sufficiently large norm arises in the trace set of Γ.

Definition 1.8 We say that Γ **asymptotically length-saturates** if

$$\frac{\#\mathcal{T}(\Gamma) \cap B_N}{\#\{t \in \mathcal{O}_N \quad t \text{ is admissible}\}} \to 1, \tag{1.9}$$

as $N \to \infty$.

Thus the modular group length-saturates, as in Example 1 (see also work of Marklof [19] studying distinct length sets of arithmetic 3-folds), while the Hecke $(2, 5, \infty)$ group does not even asymptotically length-saturate (it fails Obstruction 1).

Conjecture 1.10 (Asymptotic Length-Saturation). *Let $\Gamma < \mathrm{PO}(n, 1)$ be discrete, finitely generated, and subarithmetic, with growth exponent α exceeding the rank of the trace ring $\mathcal{O}$. Then Γ asymptotically length-saturates.*

The stronger statement that with the same assumptions, Γ length-saturates is *false*. Indeed, already for certain cocompact arithmetic 2-folds corresponding to the norm-one elements of a quaternionic division algebra, the trace equation can cut out a ternary indefinite inhomogeneous quadric, which can exhibit infinitely many Brauer-Manin-type obstructions.

In this paper, we make the first progress towards Conjecture 1.10, by proving asymptotic length-saturation for punctured, geometrically finite, Zariski-dense covers of the modular surface. Equivalently, the fundamental group of such a cover contains parabolic elements, is finitely-generated, and is non-elementary.

Theorem 1.11 *Let Γ be a finitely generated, non-elementary subgroup of $\mathrm{SL}_2(\mathbb{Z})$ containing a parabolic element; then Γ is asymptotically length-saturating. In fact, it is effectively so, in that the right hand side of (1.9) is $1 + O(N^{-\varepsilon})$ for some $\varepsilon > 0$, as $N \to \infty$.*

Here is an explicit family of finitely-generated groups with $\delta(\Gamma) \to 1/2$ and with no local obstructions, to which the theorem applies. For m large, consider the group $\Gamma_0 < \mathrm{SL}_2(\mathbb{Z})$ generated by

$$\begin{pmatrix}0&-1\\1&0\end{pmatrix}, \ \begin{pmatrix}1&m\\0&1\end{pmatrix}\begin{pmatrix}0&-1\\1&0\end{pmatrix}\begin{pmatrix}1&-m\\0&1\end{pmatrix}, \text{ and } \begin{pmatrix}1&-m\\0&1\end{pmatrix}\begin{pmatrix}0&-1\\1&0\end{pmatrix}\begin{pmatrix}1&m\\0&1\end{pmatrix}.$$

A fundamental domain for the action of Γ_0 is shown in Fig. 1. By strong approximation, there is some $q_0 = q_0(m)$ such that the reduction of Γ_0 mod any prime $p \nmid q_0$ is onto. Let $P = P(m)$ be a very large prime coprime to q_0 and let Γ be the group generated by Γ_0 and the translation $\begin{pmatrix}1&P\\0&1\end{pmatrix}$. Then since P is a unit mod q_0, the reduction of Γ mod q is easily seen to be all of $\mathrm{SL}_2(q)$ (by which we mean $\mathrm{SL}_2(\mathbb{Z}/q\mathbb{Z})$), for *all* q. Thus all numbers are admissible; that is, there are no local obstructions. For P and m large, the Hausdorff dimension of the limit set of Γ can be made any number exceeding $1/2$.

Remark 1.12 This family also gives examples of groups Γ for which it should be *difficult* to produce many traces! Indeed, these Γ have no small traces at all: the systole of the groups Γ_0 grows with m, and taking P large enough does not create a shorter closed geodesic. In particular, this is an example of a family of Γ's having *arbitrarily many* local–global failures.

Remark 1.13 The question of length-saturation is closely related to the Local-Global Conjecture for Apollonian packings, Zaremba's Conjecture, and McMullen's Classical Arithmetic Chaos Conjecture (see, e.g. [15, 17, 20] for discussions of these). Each of these problems amounts to understanding the image of a linear form (which in the

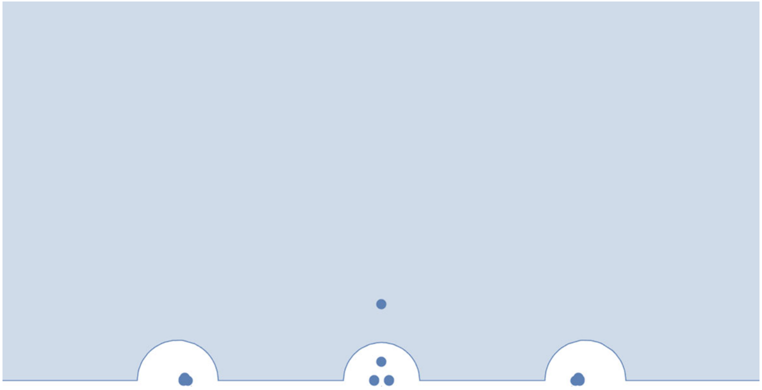

Fig. 1 The fundamental domain of Γ_0, and a typical orbit

setting of this paper is the trace) of a Zariski dense subgroup (or sub-semigroup). In each of the previous cases, the expected multiplicity in a ball of size N was some fixed positive power of N.

Note that in our setting here, there is no restriction on the growth exponent α of Γ; indeed, the Hausdorff dimension δ of the limit set of Γ, which can be as small as $\delta > 1/2$ (due to the puncture, see [1]). Counting with multiplicity, we have that:

$$\#\{\gamma \in \Gamma \cap B_N \quad \operatorname{tr}(\gamma) < N\} \;=\; N^{2\delta+o(1)}. \quad (N \to \infty)$$

So the multiplicity of a typical trace $t \asymp N$ in the trace set $t \in \mathcal{T}(\Gamma)$ may be extremely small,

$$\#\{\gamma \in \Gamma \cap B_N \quad \operatorname{tr}(\gamma) = t\} \overset{?}{=} N^{2\delta-1+o(1)}, \tag{1.14}$$

where $2\delta - 1$ can be any quantity just above 0, and yet our methods produce a density-one set of traces in this setting.

Also note that the methods introduced in [3] and applied to both the Zaremba [4] and Apollonian [5, 12, 28] settings required the linear form to have a bilinear structure. That is, the linear form was of type:

$$\gamma \mapsto \langle v, \gamma w\rangle \tag{1.15}$$

for some fixed vectors v, w. The trace is *not* of this form, and so the best one can currently say towards McMullen's conjecture is a strong level of distribution, see [7]. It is not even currently known that a positive proportion of numbers arises in the set of traces of a Zaremba-type semigroup (see [17, §3]). For related work in a somewhat different direction, see also the recent PhD thesis of Brooke Ogrodnik [23, 24].

Remark 1.16 Returning to the setting of this paper, here are some further remarks:

(1) It is sometimes possible to completely determine the trace set of Γ, even if the latter is thin. For example, take the "Lubotzky 1-2-3" group, $\Gamma = \left\langle \left(\begin{smallmatrix}1&3\\0&1\end{smallmatrix}\right), \left(\begin{smallmatrix}1&0\\3&1\end{smallmatrix}\right)\right\rangle$. It

is easy to see that every trace is $\equiv 2 (\bmod 9)$, and indeed the element $\left(\begin{smallmatrix}1&0\\3&1\end{smallmatrix}\right)\left(\begin{smallmatrix}1&3\\0&1\end{smallmatrix}\right)^n$ has trace $2+9n$, so all admissible traces are represented by this one arithmetic progression.

(2) Since we have assumed that Γ does contain a parabolic element, it is immediate that its trace set $\mathcal{T}(\Gamma)$ comprises a positive proportion of integers, since, as above, $\mathcal{T}(\Gamma)$ contains entire arithmetic progressions. Without assuming that $\Gamma\backslash\mathbb{H}$ is punctured, current technology cannot not even produce a positive proportion of traces!

(3) On the other hand, an argument based on Furstenberg's topology on the integers shows that, if there is even a single local–global failure (that is, an admissible t not in $\mathcal{T}(\Gamma)$), then finitely many such arithmetic progressions cannot possibly cover even a density-1 subset of $\mathcal{T}(\Gamma)$. A sketch is the following: declare arithmetic progressions to be open, and generate a topology from this basis; then give the admissible numbers in $\mathcal{T}(\Gamma)$ the subspace topology. If a finite number of arithmetic progressions cover a density-1 set of admissible numbers, then their complement, assumed to be non-empty, must be open; hence it contains an arithmetic progression, so the cover is not density-1.

1.1 Methods

We use the (orbital) circle method to access the trace set $\mathcal{T}(\Gamma)$. In fact, our methods apply not just to the trace function $\mathrm{tr} : \mathrm{SL}_2(\mathbb{Z}) \to \mathbb{Z}$ but to any linear form, $\mathscr{L}$, say, on SL_2 (and hence we do not group traces by conjugacy class). It turns out (see (2.1)) that the trace function is the "generic" linear form. The main theorem, from which Theorem 1.11 follows immediately, is the following.

Theorem 1.17 *Let Γ be a geometrically finite, punctured, Zariski dense subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and let $\mathscr{L} : \Gamma \to \mathbb{Z}$ be any linear form. Let $\mathcal{A}$ denote the admissible values of $\mathscr{L}$; that is, $n \in \mathcal{A}$ if and only if $n \in \mathscr{L}(\Gamma) \ (\bmod q)$, for all q. Then there is some $\Theta > 0$, so that:*

$$\frac{\#\{n \in \mathscr{L}(\Gamma) \cap [1, N]\}}{\#\{n \in \mathcal{A} \cap [1, N]\}} = 1 + O(N^{-\Theta}),$$

as $N \to \infty$. The implied constant is effective.

In the special case that the linear form is bilinear (as in (1.15)) and the critical exponent of Γ is sufficiently close to 1, the above theorem is proved by the second-named author in [29], but there are major differences between that setting (and indeed all previous work on the orbital circle method) and the present paper. For one thing, we are able to, for the first time, handle the case of $\mathscr{L}$ being the trace, which is *not* bilinear; see Remark 1.16(2). This introduces great difficulties even in the major arc analysis, as described below, requiring delicate arguments with the Burgess bound and Siegel zeros. There are also a number of key innovations in our handling of the minor arcs, among other things, requiring a "third Kloosterman refinement" to get the application and allow δ to be as small as possible, any amount exceeding $1/2$.

The starting point of our attack is to use the parabolic element in *two* ways to produce not only arithmetic progressions, but values of binary quadratic polynomials in the set of values of the linear form $\mathscr{L}$. By this we mean the following: given a fixed element $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and a parabolic, say, $\left(\begin{smallmatrix} 1 & P \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$, we can compute that

$$\operatorname{tr}\left(\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & P \\ 0 & 1 \end{smallmatrix}\right)^x\right) = a + d + cxP,$$

which is a linear form in x, whereas, say,

$$\operatorname{tr}\left(\left(\begin{smallmatrix} 1 & P \\ 0 & 1 \end{smallmatrix}\right)^x\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 & P \\ 0 & 1 \end{smallmatrix}\right)^y\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\right) = a^2 + 2bc + d^2 + (a+d)cP(x+y) + c^2P^2xy$$

is quadratic as a function of the pair (x, y). (Using three or more copies of the parabolic produces cubic or higher forms; the added cost of increasingly larger coefficient sizes seems not to be advantageous for this problem.)

Then varying x and y, and letting $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ run over certain regions of Γ in a ball of size N, we study the "representation number" $\mathcal{R}_N(n)$ of the number of times that $n \asymp N$ occurs as a value of $\mathscr{L}$. In fact our construction of $\mathcal{R}_N$ is more complicated, as we need to create multilinear forms at several scales for later estimates, see Sect. 2.1.

Following the (orbital) circle method, we decompose $\mathcal{R}_N(n)$ into a "main" term $\mathcal{M}_N(n)$ and an "error", $\mathcal{E}_N(n)$, where we integrate over the "major" and "minor" arcs, resp. We note again that the main term is expected to be of size a singular series times $N^{2\delta-1}$ (see (1.14)), which may be an arbitrarily small (but fixed) positive power of N. So we do not have much room to get an error off of the main term!

As in some other applications of the orbital circle method, we are only able to control the error terms in L^2, and hence produce only Asymptotic Length-Saturation, and not full Length-Saturation (which perhaps could be expected in this setting). Even this involves several novel techniques; there is a more standard analysis of cancellation in certain exponential sums and averages thereof, and there is also an appeal at some point to Hilbert's Nullstellensatz in effective form (see Sect. 3.2).

In the major arcs, we use the work of Bourgain-Varju [10] and Bourgain-Gamburd-Sarnak [2] for an archimedean spectral gap, together with infinite volume counting methods of [8], to obtain an estimate for the main term $\mathcal{M}_N(n)$. But there are several surprises here as well! It turns out that the singular series is a very short sum (of length N^{ε}) which is trying to approximate a quadratic Dirichlet L-function at 1, see Sect. 3.6. On GRH, this L-value can indeed be approximated by such a short sum, but our statement is unconditional! Since our error term estimate is anyway only as an "average" over n, we also average on the main term; that is, we show (see Theorem 3.38) that, for all but very few n's, the approximation is valid. But then we have another problem: we need Siegel's bound to know that the singular series, which is now one such L-value, is not too small. Again, because we are stating only an average result, we show (see Theorem 3.41) that we can bound these L-values from below for all but an exceptional set of values of n, with *effective* constants, leading to the effective constants in Theorem 1.17.

Outline

We begin in Sect. 2 with the setup of the circle method, introducing the main representation function, and its decomposition into a main term and error, corresponding to major and minor arcs. The next two sections provide preparatory lemmata for the main arguments. We record in Sect. 3 various infinite volume counting theorems in congruence towers, the savings off of such counts in progressions for *arbitrarily large* modulus (this is where Nullstellensatz is used), as well as the analysis of the singular series, which involves Weil and Burgess bounds. In Sect. 4, we prepare various exponential sum estimates used in the minor arcs analysis, using more standard analytic techniques such as estimates for Kloosterman-type sums. These allow us to complete the major arcs analysis in Sect. 5 and then the minor arc analysis in Sect. 6.

Notation

We use the standard notation $e(x)$ we mean $e^{2\pi i x}$, and $e_q(x) = e(x/q)$. The notation $\sum'_{r(q)}$ means summing over $r \pmod q$ with $(r, q) = 1$. We use the symbols $X = O(Y)$ and $X \ll Y$ interchangeably, and by $X \asymp Y$, we mean $X \ll Y \ll X$. All implied constants, unless specified otherwise, may depend at most on Γ and the linear form $\mathscr{L}$, which are treated as fixed.

2 Preliminaries

We henceforth take $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ to be a given finitely-generated, Zariski dense group containing parabolic elements. We consider a linear form $\mathscr{L} : \mathrm{SL}_2(\mathbb{Z}) \to \mathbb{Z}$ which is not everywhere vanishing; explicitly, this means that

$$\mathscr{L} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto Aa + Bb + Cc + Dd = \mathrm{tr}\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} A & C \\ B & D \end{pmatrix}\right], \tag{2.1}$$

and we assume that at least one coefficient A, B, C, D is non-zero. Note that $\mathscr{L}$ is of bilinear type (see (1.15)) exactly when its "discriminant",

$$\Delta = \Delta_{\mathscr{L}} = AD - BC$$

vanishes. After conjugation, we may assume that Γ contains the fixed parabolic element

$$\begin{pmatrix} 1 & P \\ 0 & 1 \end{pmatrix} \in \Gamma.$$

We make a few further simplifying assumptions.

- We may assume that $\gcd(A, B, C, D) = 1$, since otherwise we can pull out a common factor.

- By applying fixed elements of Γ inside $\mathscr{L}$, we may assume that the coefficients $A,\ B,\ C$ and D are all non-zero. Indeed, one has that:

$$\mathscr{L}\left(\gamma_1\begin{pmatrix} a & b\\ c & d\end{pmatrix}\gamma_0\right)=\operatorname{tr}\left[\begin{pmatrix} a & b\\ c & d\end{pmatrix}\gamma_0\begin{pmatrix} A & C\\ B & D\end{pmatrix}\gamma_1\right],$$

  and by the Zariski density of Γ, there exist elements $\gamma_0,\gamma_1\in\Gamma$ so that $\gamma_0\left(\begin{smallmatrix} A & C\\ B & D\end{smallmatrix}\right)\gamma_1$ has every entry non-zero.
- By Strong Approximation and passing to a finite index subgroup of Γ if necessary, we may assume that for all $(q_1,q_2)=1$, we have:

$$\Gamma/\Gamma(q_1q_2)\cong\Gamma/\Gamma(q_1)\times\Gamma/\Gamma(q_2), \tag{2.2}$$

  where

$$\Gamma(q)\ =\ \{\gamma\in\Gamma:\gamma\equiv I(\operatorname{mod}q)\} \tag{2.3}$$

  is the "principal congruence" subgroup of (the possibly thin group) Γ. Moreover, for all "good" primes p, we have that for $q=p^\ell$, the mod q reduction is onto

$$\Gamma(q)\backslash\Gamma=\mathrm{SL}_2(q).$$

- For a finite list of "bad" primes p (including $p=2$), we have an exponent ("saturation level") $k=k_p$ so that

$$\Gamma(p^k)\backslash\Gamma=\{I\}, \tag{2.4}$$

  and for $\ell>k$, $\Gamma(p^\ell)\backslash\Gamma$ is the full lift of the identity from $\mathrm{SL}_2(p^k)$ to $\mathrm{SL}_2(p^\ell)$. In particular, the parabolic element $\left(\begin{smallmatrix} 1 & P\\ 0 & 1\end{smallmatrix}\right)\in\Gamma$ satisfies

$$P\equiv 0(p^k) \tag{2.5}$$

  for all bad primes p.
- We may assume, by increasing the saturation densities k_p if necessary, that

$$k_p>L_{p,B}, \tag{2.6}$$

  where $L=L_{p,B}$ is determined by $p^L\|B$. (Here, as below, B is the coefficient of b in $\mathscr{L}$, as defined in (2.1).)

2.1 Setup of the circle method

For $\gamma\in\Gamma$, we construct the shifted binary quadratic form:

$$\mathfrak{f}_\gamma(x,y)\ =\ \mathscr{L}\Big(\left(\begin{smallmatrix} 1 & Px\\ 0 & 1\end{smallmatrix}\right)\gamma\left(\begin{smallmatrix} 1 & Py\\ 0 & 1\end{smallmatrix}\right)\Big), \tag{2.7}$$

so that, if $\gamma = \left(\begin{smallmatrix} a_\gamma & b_\gamma \\ c_\gamma & d_\gamma \end{smallmatrix}\right)$, then

$$\begin{aligned}\mathfrak{f}_\gamma(x, y) = (Aa_\gamma + Bb_\gamma + Cc_\gamma + Dd_\gamma) + (Ac_\gamma + Bd_\gamma)Px + (Ba_\gamma + Dc_\gamma)Py \\ +Bc_\gamma P^2 xy. \end{aligned} \tag{2.8}$$

Note that, for any integers $x, y \in \mathbb{Z}$ and any $\gamma \in \Gamma$, the value of $\mathfrak{f}_\gamma(x, y)$ arises in $\mathscr{L}(\Gamma)$. Let N be the main growing parameter, and T, X be parameters determined by:

$$T = N^{1/100}, \;\; X = N^{99/200}, \;\; \text{so that } \; TX^2 = N. \tag{2.9}$$

Decompose T further at

$$T = T_1 T_2, \;\; \text{with } \; T_2 = T_1^{\mathcal{C}}, \tag{2.10}$$

with $\mathcal{C}$ a very large constant depending only on the spectral gap for Γ, see (3.7).

We now define the main ensemble $\mathscr{F}_T$ as follows

$$\mathscr{F}_T := \left\{ \gamma_1 \cdot \gamma_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \begin{array}{c} \gamma_1, \gamma_2 \in \Gamma \\ \frac{1}{2}T_1 \le \|\gamma_1\| < T_1 \\ \frac{1}{2}T_2 \le \|\gamma_2\| < T_2 \\ \frac{1}{100}T < a,b,c,d \end{array} \right\}. \tag{2.11}$$

We show in Lemma 3.3 that $\mathscr{F}_T$ has cardinality $\asymp T^{2\delta}$. This is a sub-multi-set of Γ, as the product $\gamma_1\gamma_2$ may have multiplicity; that is, for a fixed $\xi \in \mathscr{F}_T$,

$$\sum_{\gamma \in \mathscr{F}_T} \mathbf{1}_{\{\gamma = \xi\}} \ll T_1^{2\delta}. \tag{2.12}$$

Fixing a smooth nonnegative bump function Υ with $\operatorname{supp} \Upsilon \subset [\frac{1}{2}, 1]$, define the main "representation number"

$$\mathcal{R}_N(n) \; = \; \sum_{\gamma \in \mathscr{F}_T} \sum_{x \in \mathbb{Z}} \sum_{y \in \mathbb{Z}} \Upsilon\left(\frac{x}{X}\right) \Upsilon\left(\frac{y}{X}\right) \mathbf{1}_{\{\mathfrak{f}_\gamma(x,y) = n\}}.$$

We decompose $\mathcal{R}_N$ into a "main" term and an error according to a (smoothed) major/minor arcs decomposition of the circle. To this end, let

$$\psi(t) \; = \; \max(1 - |t|, 0)$$

be the "tent" function whose Fourier transform is the Fejér-type kernel:

$$\widehat{\psi}(\xi) = \frac{\sin^2(\pi\xi)}{(\pi\xi)^2}.$$

We fix parameters Q_0, K_0 to be determined as follows. We set

$$Q_0 = N^{\alpha_0}, \quad K_0 = N^{\kappa_0}, \tag{2.13}$$

where the exponents satisfy

$$\kappa_0 = 3\alpha_0 \tag{2.14}$$

and

$$5\alpha_0 + \kappa_0 < \Theta. \tag{2.15}$$

Here Θ is the minimum of the two values in Lemmas 3.4 and 3.5. Setting

$$\alpha_0 = \Theta/10, \quad \kappa_0 = 3\Theta/10$$

will satisfy all the criteria.

With these choices, let

$$\Psi_{N,K_0}(\beta) := \sum_{m\in\mathbb{Z}} \psi((\beta+m)\tfrac{N}{K_0}),$$

and define the "major arcs" weight function as:

$$\mathfrak{M}(\theta) \;=\; \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \Psi_{N,K_0}(\theta - \tfrac{r}{q}). \tag{2.16}$$

Then the "main" term is given by:

$$\mathcal{M}_N(n) \;=\; \int_0^1 \mathfrak{M}(\theta)\widehat{\mathcal{R}_N}(\theta)e(-n\theta)d\theta, \tag{2.17}$$

and of course the error is

$$\mathcal{E}_N(n) \;=\; \int_0^1 (1-\mathfrak{M}(\theta))\widehat{\mathcal{R}_N}(\theta)e(-n\theta)d\theta,$$

so that

$$\mathcal{R}_N(n) = \mathcal{M}_N(n) + \mathcal{E}_N(n).$$

3 Major arc technical estimates

We record here a number of technical estimates needed in the analysis of the main term.

3.1 Spectral analysis and counting

Let $\delta = \delta_\Gamma$ be the Hausdorff dimension of the limit set of Γ, and recall that $\delta > 1/2$. By Patterson-Sullivan theory [25, 26], δ is related to the bottom eigenvalue $\lambda_0 = \delta(1-\delta)$ of the hyperbolic Laplacian acting on $L^2(\Gamma\backslash\mathbb{H})$. Work of Lax-Phillips [18] shows that the spectrum of the latter below $1/4$ consists of finitely many eigenvalues. By Bourgain-Varju [10] and Bourgain-Gamburd-Sarnak [2], there is a uniform spectral gap, in the following sense.

Theorem 3.1 *([2, 10]). There exists a "spectral gap,"*

$$\Theta_0 = \Theta_0(\Gamma) > 0 \tag{3.2}$$

so that, for all $q \geq 1$, the eigenvalue λ_0 is the bottom of the spectrum of $L^2(\Gamma(q)\backslash\mathbb{H})$, and all other eigenvalues are at least $\lambda_\Theta := s_\Theta(1 - s_\Theta)$, where $s_\Theta := \delta - \Theta_0$. Here $\Gamma(q)$ is as defined in (2.3).

Recalling the construction of $\mathscr{F}_T$ from (2.11), we record the following counting results, which follow from now-standard techniques.

Lemma 3.3 *As $T \to \infty$,*

$$\#\mathscr{F}_T \asymp T^{2\delta}.$$

Proof This follows from infinite volume counting methods in Zariski dense groups with $\delta > 1/2$; see, e.g., [8]. □

Lemma 3.4 *There exists $\Theta > 0$ so that, for any $q \geq 1$, $\gamma_0 \in \Gamma(q)\backslash\Gamma$, $|\beta| < 1/X^2$, and $x, y \asymp X$, we have that*

$$\sum_{\substack{\gamma\in\mathscr{F}_T \\ \gamma\equiv\gamma_0 (\operatorname{mod} q)}} e(\beta\mathfrak{f}_\gamma(x,y)) = \frac{1}{[\Gamma:\Gamma(q)]}\sum_{\gamma\in\mathscr{F}_T} e(\beta\mathfrak{f}_\gamma(x,y)) + O(|\mathscr{F}_T|N^{-\Theta}),$$

as $T \to \infty$.

Proof The proof is the same as that of the similar statement in [8, Theorem 1.14]. □

Lemma 3.5 *There exists $\Theta > 0$ so that, for $x, y \asymp X$ and $n \asymp N$, we have:*

$$\sum_{\gamma\in\mathscr{F}_T}\int_{\mathbb{R}} \psi(\beta\tfrac{N}{K_0})e(\beta(\mathfrak{f}_\gamma(x,y)-n))d\beta \gg \frac{|\mathscr{F}_T|}{K_0} + O(|\mathscr{F}_T|N^{-\Theta}),$$

as $T \to \infty$.

Proof We first note that

$$\int_{\mathbb{R}} \psi(\beta \tfrac{N}{K_0}) e(\beta(\mathfrak{f}_\gamma(x,y)-n))d\beta = \frac{K_0}{N}\widehat{\psi}((\mathfrak{f}_\gamma(x,y)-n)\frac{K_0}{N}) \geq 0,$$

and if $|\mathfrak{f}_\gamma(x,y)-n)\frac{K_0}{N}| < \frac{1}{2}$, then $\widehat{\psi}(\cdot) > \frac{2}{5}$. So we need to show the count:

$$\sum_{\gamma\in\mathscr{F}_T} \mathbf{1}_{\{|\mathfrak{f}_\gamma(x,y)-n|<\frac{N}{2K_0}\}} \gg \frac{|\mathscr{F}_T|}{K_0} + O(|\mathscr{F}_T|N^{-\Theta}).$$

The latter follows from the same techniques as the proof of [8, Theorem 1.15]. □

3.2 Nullstellensatz

Theorem 3.6 *Let Θ_0 be the spectral gap in (3.2). Define $\mathcal{C}$ by*

$$\mathcal{C} = 3\times 10^8/\Theta_0, \tag{3.7}$$

which is needed to specify the construction of the set $\mathscr{F}_T$ in (2.11) and (2.10). There exists an $\eta_0 > 0$ depending only on the spectral gap for Γ, so that, for all $1 \leq q < N$, and all $r(\bmod q)$,

$$\sum_{\gamma\in\mathscr{F}_T} \mathbf{1}_{\{c_\gamma\equiv r(\bmod q)\}} \ll \frac{1}{q^{\eta_0}}|\mathscr{F}_T|. \tag{3.8}$$

The proof of this theorem follows a similar strategy to that of [5, Lemma 5.2]; unfortunately, that proof contains a minor gap, so we give full details here for how to overcome it.

Proof We first drop the condition $\frac{1}{100}T < a, b, c, d$ from $\mathscr{F}_T$ in (2.11), so that we need to count the number of $\gamma_1 \asymp T_1$, $\gamma_2 \asymp T_2$ so that the "c" entry of $\gamma_1\gamma_2$,

$$\langle e_2, \gamma_1\gamma_2 e_1\rangle \equiv r(\bmod q),$$

where e_j are standard basis vectors. This decomposes into two cases according to the size of q.

Case $q < T_2^{\Theta_0/3}$: In this case, we simply apply spectral theory in γ_2 while leaving γ_1 fixed, as follows. Break γ_2 into progressions mod q:

$$\sum_{\gamma_1\asymp T_1}\sum_{\gamma_2\asymp T_2} \mathbf{1}_{\{\langle e_2,\gamma_1\gamma_2 e_1\rangle\equiv r(\bmod q)\}}$$

$$= \sum_{\gamma_1\asymp T_1}\sum_{\gamma_0\in\Gamma(q)\backslash\Gamma}\left[\mathbf{1}_{\{\langle e_2,\gamma_1\gamma_0 e_1\rangle\equiv r(\bmod q)\}} \sum_{\substack{\gamma_2\asymp T_2\\ \gamma_2\equiv\gamma_0(\bmod q)}} 1\right].$$

The bracketed term may be estimated using the uniform spectral gap (see [8]) to give

$$\begin{aligned}&\ll \sum_{\gamma_1 \asymp T_1} \sum_{\gamma_0 \in \Gamma(q)\backslash\Gamma} \mathbf{1}_{\{\langle e_2, \gamma_1\gamma_0 e_1\rangle \equiv r (\operatorname{mod} q)\}} \left[\frac{1}{q^3} T_2^{2\delta} + O(T_2^{2\delta-\Theta_0}) \right] \\ &\ll \frac{1}{q} T_2^{2\delta} T_1^{2\delta} + q^2 T_1^{2\delta} T_2^{2\delta-\Theta_0}.\end{aligned}$$

Here we used that $[\Gamma : \Gamma(q)] \asymp q^3$. This saves $1/q$ (more than claimed) as long as $q < T_2^{\Theta_0/3}$.

Case $q \geq T_2^{\Theta_0/3} = T_1^{10^8}$**:** The overview of the argument is as follows. For any fixed γ_2, we consider the set of $\gamma_1 \asymp T_1$ for which $\langle e_2, \gamma_1\gamma_2 e_1\rangle \equiv r (\operatorname{mod} q)$. Since different integers having the same residue class mod q differ by q, and q is huge compared to T_1, we will show by Nullstellensatz that in fact the modular restriction can be lifted to an absolute restriction $\langle e_2, \gamma_1\gamma_2 e_1\rangle = r_*$, for some integer r_* (depending on γ_2, which is fixed). Then we will relax the absolute restriction back down to a modular one, but with a much smaller modulus, $\langle e_2, \gamma_1\gamma_2 e_1\rangle \equiv r_* (\operatorname{mod} q_*)$, where $q_* \asymp T_1^{\Theta_0/3}$, and apply the previous argument to save a power of q_*, which itself is a tiny power of q.

An issue arises in the use of Nullstellensatz that was overlooked in related arguments in [5, 6]. Write $\gamma_2 e_1 = (u, v)$ and $\gamma_1 e_1 = (c, d)$, so that $\langle e_2, \gamma_1\gamma_2 e_1\rangle = uc + vd$, with $|u|, |v| \leq T_2$ being "large" and fixed, and $|c|, |d| \leq T_1$ being "small" variables. It was claimed that, since $(u, v) = 1$, we may assume that, say, $(u, q) = 1$, and rewrite the modular condition as $c + v\bar{u}d \equiv r\bar{u} (\operatorname{mod} q)$. Unfortunately the obvious linear transformation that allows this rewrite requires changing the coefficients c, d to ones of size bounded by $T_1 T_2 = T$, and this ruins the heights of the polynomials to be used in effective Nullstellensatz. So we need a more delicate argument to control the size of coefficients, as follows.

Suppose $q < N$ has a divisor $\tilde{q} \mid q$ of size $T_1 < \tilde{q} < T_2^{\Theta_0/3}$, say. Then we relax $\langle e_2, \gamma_1\gamma_2 e_1\rangle \equiv r (\operatorname{mod} q)$ to the same congruence mod $\tilde{q}$, and count as in the previous case. This saves $1/\tilde{q} > 1/T_1$, which is a (very small) power of $N > q$, and completes the argument in this case.

Next we suppose that q has no divisor in this range. Let $\tilde{q}$ be the largest divisor of q not exceeding T_1, and write $q_0 := q/\tilde{q}$. We again relax the congruence restriction to $\langle e_2, \gamma_1\gamma_2 e_1\rangle \equiv r (\operatorname{mod} q_0)$; if we can save a small power of q_0, this also saves a small power of q. Then any prime divisor p of q_0 must exceed $T_2^{\Theta_0/3}$, for otherwise either p or $p\tilde{q}$ is a divisor of q which is does not exceed $T_2^{\Theta_0/3}$. Therefore q_0 is "almost-prime", that is, there are primes $p_j \geq T_2^{\Theta_0/3} = N^\eta$, say, so that

$$q_0 = p_1 p_2 \cdots p_\ell,$$

with $\ell < \lceil 1/\eta \rceil$.

Next we consider the values of $u + \alpha v$, for $\alpha = 1, 2, \ldots, \ell + 1$, and claim that at least one such value is coprime to q_0. (The point here is that ℓ depends only on q, and is bounded only in terms of Θ_0, which only depends on Γ.) Consider first the primes p_j which divide either u or v (recall that u and v are coprime); then since

$p_j > T_2^{\Theta_0/3} > \ell + 1$ (for N, and hence T_2, large enough), none of the p_j can divide *any* value of $u + \alpha v$. Now consider the p_j which are coprime to u and v. Then since $u + \alpha v$ is an arithmetic progression of length $\ell + 1 < p_j$, at most one value $\alpha_j \in \{1, 2, \dots, \ell + 1\}$ can satisfy $u + \alpha_j \equiv 0 (\mathrm{mod}\, p_j)$. Since the number of α's exceeds the number of p_j's, there is some α so that $u + \alpha v$ is coprime to all the p_j, and hence coprime to q_0.

Again, this α is bounded absolutely, and depends only on q, u, and v, and not on c and d (which depend on γ_1). Now we proceed with the Nullstellensatz argument. Using the modulus q_0, we fix γ_2, let $(u, v) = \gamma_2 e_1$, and consider the set

$$S = S_{\gamma_2} := \{\gamma_1 \in \Gamma,\ \|\gamma_1\| \le T_1 \text{ with } uc + vd \equiv r (\mathrm{mod}\, q_0)\},$$

where we have set $(c, d) := \gamma_1 e_2$. Using α from the previous argument with $u + \alpha v$ coprime to q_0, we write $uc + vd = (u + \alpha v)c + v(d - \alpha c)$, so that the congruence condition becomes

$$c + v\overline{(u + \alpha v)}(d - \alpha c) \equiv r\overline{(u + \alpha v)}(\mathrm{mod}\, q_0).$$

Now consider the (linear) polynomials $P_{\gamma_1} \in \mathbb{Z}[U, V]$ given by

$$P_{\gamma_1}(U, V) := c + U(d - \alpha c) - V,$$

and consider the affine variety

$$\mathcal{V} := \bigcap_{\gamma_1 \in S} \{P_{\gamma_1} = 0\}.$$

We claim that $\mathcal{V}(\mathbb{C})$ is nonempty. Note that the coefficients of P_{γ_1} are bounded by $(\ell + 2)T_1$. Then if $\mathcal{V}(\mathbb{C})$ is empty, Hilbert's Nullstellensatz, in effective form (see, e.g., [22, Theorem IV]) gives the existence of polynomials $Q_{\gamma_1} \in \mathbb{Z}[U, V]$ and an integer $\mathfrak{d} \ge 1$ so that

$$\sum_{\gamma_1 \in S} P_{\gamma_1}(U, V) Q_{\gamma_1}(U, V) = \mathfrak{d}, \tag{3.9}$$

and with $\mathfrak{d}$ bounded (for N, and hence T_1, large enough) by

$$\mathfrak{d} \le \exp(8^7(\log T_1 + \log(\ell + 2) + 8 \log 8) \le T_1^{10^7}.$$

("Large enough" is in terms of an implied constant depending only on Γ, since ℓ depends only on Θ_0). But if we reduce (3.9) mod q_0 and set $(U, V) \equiv (v\overline{(u + \alpha v)}, r\overline{(u + \alpha v)})$, we get $\mathfrak{d} \equiv 0 (\mathrm{mod}\, q_0)$, which is impossible since $q_0 = q/\tilde{q} > T_1^{10^8 - 1}$.

Therefore $\mathcal{V}(\mathbb{C})$ is nonempty, and hence $\mathcal{V}(\mathbb{Q})$ is nonempty, and so clearing denominators, there exist coprime integers u_*, v_*, r_*, so that for all $\gamma_1 \in S$,

$$u_* c + v_* d = r_*.$$

We have turned our congruence condition into an archimedean condition. Now we take some $q_* \asymp T_1^{\Theta_0/3}$ coprime to u_*, v_*, r_*, relax the archimedean condition back to a modular one, $u_* c + v_* d \equiv r_* (\operatorname{mod} q_*)$, and count the number of $\gamma_1 \asymp T_1$ satisfying this. As before, the spectral argument saves $1/q_*$, which is some small power of q. □

3.3 Singular series preliminaries

Recall that $\gcd(A, B, C, D) = 1$ and $\Delta = AD - BC$. Let c_q denote the Ramanujan sum,

$$c_q(m) := \sideset{}{'}\sum_{a (\operatorname{mod} q)} e_q(am).$$

(There should be no confusion between c_q and bottom left element $c = \gamma_c$ of a typical matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.) We study here sums arising in the singular series analysis, of the form

$$\mathfrak{S}_q(n) := \frac{1}{|\Gamma(q)\backslash\Gamma|} \sum_{\gamma \in \Gamma(q)\backslash\Gamma} c_q(\mathfrak{f}_\gamma(x, y) - n),$$

for fixed $x, y \in \mathbb{Z}$. Note immediately that the sum, being over all $\gamma \in \Gamma(q)\backslash\Gamma$, is independent of x, y, which we may assume are both 0; thus $\mathfrak{f}_\gamma = Aa + Bb + Cc + Dd$. By the structure of $\Gamma(q)\backslash\Gamma$ in (2.2), the sum is multiplicative, so we may assume that $q = p^\ell$ is a prime power. For "good" primes, we have that $\Gamma(q)\backslash\Gamma = \mathrm{SL}_2(q)$. For "bad" primes, we have that $\Gamma(p^k)\backslash\Gamma = \{I\}$ for some "saturation exponent" k, while $\Gamma(p^\ell)\backslash\Gamma$ for $\ell > k$ is the full lift to $\mathrm{SL}_2(p^\ell)$ of the identity in $\mathrm{SL}_2(p^k)$.

3.4 Good primes

Lemma 3.10 *Assume that q is a power of a good prime. Then we have that*

$$\mathfrak{S}_q(n) = \frac{1}{|\Gamma(q)\backslash\Gamma|} \sum_{\gamma \in \Gamma(q)\backslash\Gamma} c_q(a_\gamma + \Delta d_\gamma - n).$$

Proof Recalling that

$$\mathfrak{f}_\gamma(0, 0) = \operatorname{tr}\left[\gamma \left(\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right)\right] = \operatorname{tr}\left[\left(\begin{smallmatrix} \alpha & \beta \\ \kappa & \delta \end{smallmatrix}\right) \gamma \left(\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right) \left(\begin{smallmatrix} \delta & -\beta \\ -\kappa & \alpha \end{smallmatrix}\right)\right],$$

for any $\left(\begin{smallmatrix}\alpha & \beta \\ \kappa & \delta\end{smallmatrix}\right) \in \mathrm{SL}_2$, and the sum being over all $\gamma \in \mathrm{SL}_2(q)$, we may simplify the expression in the following way. Assume WLOG that $(A, q) = 1$; then we can rescale $\left(\begin{smallmatrix}A & C \\ B & D\end{smallmatrix}\right)$ to $\left(\begin{smallmatrix}1 & C\bar{A} \\ BA & DA\end{smallmatrix}\right)$, and continuing by elementary operations, we may replace $\left(\begin{smallmatrix}A & C \\ B & D\end{smallmatrix}\right)$ by $\left(\begin{smallmatrix}1 & 0 \\ 0 & \Delta\end{smallmatrix}\right)$, as claimed. □

3.4.1 Case $\Delta \equiv 0(\mathrm{mod}\, p)$

Lemma 3.11 *Assume $\Delta \equiv 0(p)$. For $q = p$ a good prime, we have that:*

$$\mathfrak{S}_q(n) = \begin{cases} \frac{-1}{p+1} & ifn \equiv 0(p), \\ \frac{1}{p^2-1} & ifn \not\equiv 0(p). \end{cases}$$

Proof Recall that

$$c_p(x) = \begin{cases} p-1 & \text{if } x = 0, \\ -1 & \text{else.} \end{cases}$$

Write $\gamma = \left(\begin{smallmatrix}a & b \\ c & d\end{smallmatrix}\right)$. Assume p is a good prime. From Lemma 3.10, we need to count the number of $\left(\begin{smallmatrix}a & b \\ c & d\end{smallmatrix}\right)$ with $a = n$ or $a \neq n$.

Consider the case $n \equiv 0(\mathrm{mod}\, p)$. Then either $a \equiv n \equiv 0(\mathrm{mod}\, p)$ or not. In the first case, $bc \equiv 1(\mathrm{mod}\, p)$ and d is free ($p(p-1)$ matrices) and $c_p = p-1$ for a total contribution of $p(p-1)^2$. In the second case $a \neq n$, there are $p-1$ choices for a, then p^2 choices for b, c, and $d = (bc+1)\bar{a}$ is determined. This is $p^2(p-1)$ matrices with $c_p = -1$. Combining these contributions gives $(-1)p(p-1)$ when $n \equiv 0$.

Now suppose $n \not\equiv 0$. Then if $a \equiv n$, then b and c are free (with p^2 choices) and d is determined, with $c_p = p-1$, for a net contribution of $p^2(p-1)$. If $a \not\equiv n$, then $c_p = -1$ and we either have $a = 0, bc \equiv 1$ and d free ($p(p-1)$ choices), or $a \neq 0$ (with $p-2$ choices), and b, c free and d determined (p^2 choices). The total contribution is then p when $n \not\equiv 0(p)$.

The size of $\mathrm{SL}_2(p)$ is $p(p-1)(p+1)$, which gives the claim. □

Lemma 3.12 *Assume $\Delta \equiv 0(\mathrm{mod}\, p)$. For $q = p^\ell$ a power of a good prime ($\ell \geq 2$), we have that:*

$$\mathfrak{S}_q(n) = 0.$$

Proof For prime powers, we have that:

$$c_{p^\ell}(x) = \begin{cases} 0 & \text{if } x \not\equiv 0(p^{\ell-1}), \\ -p^{\ell-1} & \text{if } x \not\equiv 0(p^\ell) \text{ but } x \equiv 0(p^{\ell-1}), \\ p^{\ell-1}(p-1) & \text{if } x \equiv 0(p^\ell). \end{cases}$$

So there is no contribution unless $a + \Delta d \equiv n(\mathrm{mod}\, p^{\ell-1})$. Fix $\gamma_0 = \left(\begin{smallmatrix}a_0 & b_0 \\ c_0 & d_0\end{smallmatrix}\right) \in \mathrm{SL}_2(p^{\ell-1})$ which solves $a_0 + \Delta d_0 \equiv n(\mathrm{mod}\, p^{\ell-1})$ and $a_0 d_0 - b_0 c_0 = 1$. Consider any lift $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(p^\ell)$ of γ_0, that is, $a = a_0 + p^{\ell-1}a_1$, etc. The restriction that $ad - bc \equiv 1 (\bmod\, p^\ell)$ becomes:

$$a_1 d_0 + d_1 a_0 - c_1 b_0 - b_1 c_0 \equiv 0 (\bmod\, p). \tag{3.13}$$

(This is just the Jacobian of the determinant.) The above defines a 3-dimensional subspace in a_1, b_1, c_1, d_1. Assume WLOG that $a_0 \not\equiv 0 (\bmod\, p)$. Then (3.13) determines d_1 once a_1, b_1, c_1 are determined. We consider two cases, $a + \Delta d \equiv n (\bmod\, p^\ell)$ or not; since $\Delta \equiv 0 (\bmod\, p)$, this is a restriction on a_1, which leaves b_1, c_1 free (p^2 choices, which is the same count either way). If a_1 is the unique value mod p for which $a + \Delta d \equiv n (\bmod\, p^\ell)$, then $c_{p^\ell} = p^{\ell-1}(p-1)$. But if a_1 is one of the $(p-1)$ values for which $a + \Delta d \not\equiv n (\bmod\, p^\ell)$, then $c_{p^\ell} = -p^{\ell-1}$. The net contribution from these two cases exactly cancels. □

3.4.2 Case $\Delta \not\equiv 0 (\bmod\, p)$

Lemma 3.14 *Assume $\Delta \not\equiv 0 (\bmod\, p)$. For $q = p$ a good prime, we have that:*

$$\mathfrak{S}_q(n) = \frac{1 + p\left(\frac{n^2 - 4\Delta}{p}\right)}{p^2 - 1},$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol.

Proof Again by Lemma 3.10, we need to count the number of $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with $a + \Delta d = n$ or not. We decompose $\mathrm{SL}_2(p)$ according to whether $c = 0$ or not.

If $c = 0$, then $\gamma = \left(\begin{smallmatrix} a & b \\ 0 & \bar{a} \end{smallmatrix}\right)$, and we need to know whether $a + \Delta \bar{a} \equiv n$ or not. This equation is equivalent to $a^2 - na + \Delta \equiv 0$, which has $\left(\frac{n^2 - 4\Delta}{p}\right) + 1$ solutions for a with b free (with p choices), each contributing $c_p = p - 1$ to $\mathfrak{S}_q$. The remaining $\left(p - 1 - \left(\frac{n^2-4\Delta}{p}\right) - 1\right) p$ solutions contribute $c_p = -1$ each.

If $c \not\equiv 0$, then for any choice of d, we either have $a \equiv n - \Delta d$ (with one choice, contributing $c_p = p - 1$) or not ($p - 1$ choices contributing $c_p = -1$). Then c is free ($p - 1$ choices) and $b = (ad - 1)\bar{c}$ is determined. These two contributions exactly cancel.

On using $|\,\mathrm{SL}_2(p)| = p(p-1)(p+1)$, the net contribution is as claimed. □

Lemma 3.15 *Assume $\Delta \not\equiv 0 (\bmod\, p)$. Let $p^L \| (n^2 - 4\Delta)$. For $q = p^\ell$ a power of a good prime ($\ell \geq 2$), we have that:*

$$\mathfrak{S}_q(n) = \begin{cases} 0 & \text{if } L \leq \ell - 2, \text{ or if } \ell \text{ is odd and } L \geq \ell, \\ p^{-(\ell-3)/2}(p^2-1)^{-1}\left(\frac{(n^2-4\Delta)/p^L}{p}\right) & \text{if } \ell \text{ is odd and } L = \ell - 1, \\ p^{-(\ell-2)/2}(p+1)^{-1} & \text{if } \ell \text{ is even and } L \geq \ell, \\ -p^{-(\ell-2)/2}(p^2-1)^{-1} & \text{if } \ell \text{ is even and } L = \ell - 1. \end{cases} \tag{3.16}$$

In any case,

$$\mathfrak{S}_{p^\ell}(n) \ll p^{-\ell/2}. \tag{3.17}$$

Proof We decompose SL_2 according to the value of γ_c:

$$\mathrm{SL}_2(p^\ell) = \bigsqcup_{c \in \mathbb{Z}/p^\ell} \mathscr{C}_c,$$

where

$$\mathscr{C}_c = \{\gamma \in \mathrm{SL}_2(p^\ell) : \gamma_c = c\}.$$

Notice that $\Gamma_\infty = \{n_x \ x \in \mathbb{Z}/p^\ell\}$ acts on the left on $\mathscr{C}_c$, where $n_x = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)$, so we may decompose $\mathscr{C}_c$ into Γ_∞-cosets. The value $\mathfrak{f}_\gamma = a + \Delta d$ changes to $\mathfrak{f}_{n_x\gamma} = a + \Delta d + cx$ when γ is replaced by $n_x\gamma$. If $c \not\equiv 0 (\mathrm{mod}\ p^\ell)$, then this is an arithmetic progression as x varies (otherwise, it is constant). For some Γ_∞-cosets, the values of this progression are never $\equiv n (\mathrm{mod}\ p^{\ell-1})$, in which case there is no contribution to $\mathfrak{S}_q$ since the Ramanujan value c_{p^ℓ} vanishes. If the progression does attain the value $n (\mathrm{mod}\ p^{\ell-1})$, then as x ranges mod p^ℓ, this value in $\mathbb{Z}/p^{\ell-1}$ is attained with equal multiplicities from its p lifts in $\mathbb{Z}/p^\ell$. Exactly one of these lifts is $\equiv n (\mathrm{mod}\ p^\ell)$, which contributes $c_{p^\ell} = p^{\ell-1}(p-1)$, and the other $(p-1)$ lifts contribute $c_{p^\ell} = -p^{\ell-1}$. The two types of contributions exactly cancel.

We are left to study the distribution of the values $a + \Delta\bar{a}$ from $\gamma = \left(\begin{smallmatrix} a & b \\ 0 & \bar{a} \end{smallmatrix}\right)$ ranging in $\mathscr{C}_0$. In particular, we need only consider the values $a + \Delta\bar{a} \equiv n (\mathrm{mod}\ p^{\ell-1})$ and determine which of these are also $\equiv n (\mathrm{mod}\ p^\ell)$. The equation

$$a + \Delta\bar{a} \equiv n (\mathrm{mod}\ p^\ell)$$

is equivalent to

$$a^2 - na + \Delta \equiv 0 (\mathrm{mod}\ p^\ell),$$

which on completing the square gives the equation:

$$(a - \bar{2}n)^2 \equiv \bar{4}(n^2 - 4\Delta)(\mathrm{mod}\ p^\ell). \tag{3.18}$$

We want to consider the number of solutions to (3.18) as compared to the solutions to the same equation but mod $p^{\ell-1}$:

$$(a - \bar{2}n)^2 \equiv \bar{4}(n^2 - 4\Delta)(\mathrm{mod}\ p^{\ell-1}). \tag{3.19}$$

Consider first solutions to (3.19). If $n^2 - 4\Delta$ is not a square mod $p^{\ell-1}$, then (3.19) has no solutions, and $\mathfrak{S}_q = 0$. Assume henceforth that $n^2 - 4\Delta$ is a square mod $p^{\ell-1}$. Let $p^L \| n^2 - 4\Delta$.

Case: $L \leq \ell - 2$.

If L is odd, then (3.19) has no solutions. So we assume that $L = 2L_1$ is even. Since $n^2 - 4\Delta$ is a square, we can thus write $n^2 - 4\Delta \equiv s^2 p^{2L_1} (\operatorname{mod} p^{\ell-1})$ for some $s \not\equiv 0 (\operatorname{mod} p)$. Then (3.19) becomes:

$$(a - \bar{2}n - \bar{2}sp^{L_1})(a - \bar{2}n + \bar{2}sp^{L_1}) \equiv 0 (\operatorname{mod} p^{\ell-1}).$$

This equation is equivalent to the existence of $U, V \leq \ell - 1$ with $U + V \geq \ell - 1$ such that

$$a - \bar{2}n - \bar{2}sp^{L_1} \equiv 0 (\operatorname{mod} p^U), \ a - \bar{2}n + \bar{2}sp^{L_1} \equiv 0 (\operatorname{mod} p^V).$$

Assume WLOG that $U \leq V$. Then taking the difference of these equations, we have that $U \geq L_1$. But since s is invertible mod p, we must also have $U \leq L_1$, that is, $U = L_1$, and only the equation mod p^V needs to be solved, which is solved uniquely. Thus are then $2p^{L_1}$ solutions to (3.19), which are all of the form:

$$a = \bar{2}n \pm \bar{2}sp^{L_1} + kp^{\ell-1-L_1},$$

as k ranges in $\mathbb{Z}/p^{L_1}$.

For each such value of a, the question becomes: which of these also solves (3.18)? Letting k range in $\mathbb{Z}/p^{L_1+1}$ and inserting this expression for a into (3.18), we get:

$$(a - \bar{2}n)^2 \equiv \bar{4}s^2 p^{2L_1} \pm skp^{\ell-1} \overset{?}{\equiv} \bar{4}(n^2 - 4\Delta) (\operatorname{mod} p^{\ell}),$$

where we used that $2(\ell - 1 - L_1) \geq \ell$. Since $s \not\equiv 0 (\operatorname{mod} p)$, as k ranges in $\mathbb{Z}/p^{L_1+1}$, the values of $\bar{4}s^2 p^{2L_1} \pm skp^{\ell-1}$ range in an arithmetic progression of step size $p^{\ell-1}$, and so are periodic, taking each value with equal probability. As before, the corresponding Ramanujan values are then such that the contributions to $\mathfrak{S}_q$ exactly cancel.

Case $L \geq \ell - 1$ and ℓ even:

In this case, (3.19) asks for $(a - \bar{2}n)^2 \equiv 0 (\operatorname{mod} p^{\ell-1})$. The solutions to this are

$$a = \bar{2}n + kp^{\ell/2},$$

as k ranges in $\mathbb{Z}/p^{\frac{1}{2}\ell-1}$.

To see which solutions lift to (3.18), we let k range in $\mathbb{Z}/p^{\ell/2}$. Then (3.18) asks whether

$$(a - \bar{2}n)^2 \equiv k^2 p^{\ell} \overset{?}{\equiv} \bar{4}(n^2 - 4\Delta) (\operatorname{mod} p^{\ell}).$$

If $n^2 - 4\Delta \equiv 0 (\operatorname{mod} p^{\ell})$, that is, $L \geq \ell$, then every solution to (3.19) also solves (3.18). So there are $p^{\ell/2}$ values of a in $\left(\begin{smallmatrix} a & b \\ 0 & \bar{a} \end{smallmatrix}\right)$, and another p^{ℓ} values of b which is free. Each such matrix has a Ramanujan value $c_q = p^{\ell-1}(p-1)$, for a net contribution of:

$$\mathfrak{S}_q = \frac{p^{(2-\ell)/2}}{p+1},$$

where we used that $|\operatorname{SL}_2(p^\ell)| = p^{3(\ell-1)}p(p+1)(p-1)$.

If $n^2-4\Delta \not\equiv 0 (\operatorname{mod} p^\ell)$, that is, $L = \ell-1$, then no solution to (3.19) lifts to (3.18). Each matrix as above has a Ramanujan value of $c_q = -p^{\ell-1}$, for a net contribution of:

$$\mathfrak{S}_q = -\frac{p^{(2-\ell)/2}}{p^2-1}.$$

Case $L \geq \ell - 1$ and ℓ odd:

Now the solutions to (3.19) are:

$$a = \bar{2}n + kp^{(\ell-1)/2},$$

as k ranges in $\mathbb{Z}/p^{(\ell-1)/2}$.

Inserting these values into (3.18) and letting k range in $\mathbb{Z}/p^{(\ell+1)/2}$, we are asking whether

$$(a-\bar{2}n)^2 \equiv k^2 p^{(\ell-1)} \overset{?}{\equiv} \bar{4}(n^2-4\Delta)(\operatorname{mod} p^\ell).$$

If $L \geq \ell$, then this equation is satisfied if and only if $k \equiv 0(p)$, so again there is a balance and the contributions to $\mathfrak{S}_q$ cancel.

Lastly, if $L = \ell - 1$, which is even since ℓ is odd, then note that $(n^2-4\Delta)/p^L$ is a non-zero square mod p if and only if $n^2 - 4\Delta$ is a square mod p^ℓ. Whether or not this holds, there are $\left(\frac{(n^2-4\Delta)/p^L}{p}\right) + 1$ solutions for $k(\operatorname{mod} p)$, and every lift of these to $\mathbb{Z}/p^{(\ell+1)/2}$ solves (3.18). The number of these lifts is

$$p^{(\ell-1)/2}\left(\left(\frac{(n^2-4\Delta)/p^L}{p}\right)+1\right),$$

each contributing a Ramanujan value of $c_q = p^{\ell-1}(p-1)$. And of course the complementary number of solutions to (3.19) that do not lift to (3.18) is

$$p^{(\ell-1)/2}\left(p-1-\left(\frac{(n^2-4\Delta)/p^L}{p}\right)\right),$$

each giving a Ramanujan value of $c_q = -p^{\ell-1}$. Recalling that there are p^ℓ values of b in $\left(\begin{smallmatrix} a & b \\ 0 & \bar{a}\end{smallmatrix}\right)$, the total contribution to $\mathfrak{S}_q$ is then:

$$\mathfrak{S}_q = \frac{p^{-(\ell-3)/2}}{p^2-1}\left(\frac{(n^2-4\Delta)/p^L}{p}\right).$$

This completes the proof. □

We summarize this subsection as follows.

Corollary 3.20 *Let p be a good prime for Γ, and let*

$$\mathfrak{S}^{(p)}(n) := 1 + \mathfrak{S}_p(n) + \mathfrak{S}_{p^2}(n) + \cdots$$

be the "local factor" at p. Then for all p,

$$\mathfrak{S}^{(p)}(n) = 1 + \mathfrak{S}_p(n) + \mathfrak{S}_{p^2}(n) + O(p^{-3/2}).$$

Moreover,

- *If $p \mid \Delta$, then*
$$\mathfrak{S}^{(p)}(n) = 1 + O(p^{-1}).$$
- *If $p \nmid \Delta$ and $p \nmid n^2 - 4\Delta$, then*
$$\mathfrak{S}^{(p)}(n) = 1 + \frac{1}{p}\left(\frac{n^2 - 4\Delta}{p}\right) + O(p^{-2}).$$
- *If $p \nmid \Delta$ and $p \| n^2 - 4\Delta$, then*
$$\mathfrak{S}^{(p)}(n) = 1 + O(p^{-2}).$$
- *If $p \nmid \Delta$ and $p^L \| n^2 - 4\Delta$ with $L \geq 2$, then*
$$\mathfrak{S}^{(p)}(n) = 1 + O(p^{-1}).$$

3.5 Bad primes

For bad primes, our strategy is as follows. Rather than evaluating $\mathfrak{S}_q(n)$ explicitly, we show the following "density formula."

Lemma 3.21 *For any $\ell \geq 0$ and any prime p (good or bad), we have that*

$$1 + \mathfrak{S}_p(n) + \cdots + \mathfrak{S}_{p^\ell}(n) = p^\ell \frac{\#\{\gamma \in \Gamma(p^\ell)\backslash\Gamma : \mathfrak{f}_\gamma \equiv n (\mathrm{mod}\, p^\ell)\}}{[\Gamma : \Gamma(p^\ell)]}. \quad (3.22)$$

This will tautologically capture the condition that n is admissible; that is, it clearly vanishes if n is not admissible. And then, for ℓ large enough, we claim that $\mathfrak{S}_{p^\ell}(n) = 0$, so these probabilities stabilize.

Proof of Lemma 3.21 This follows immediately from

$$\mathfrak{S}_{p^m}(n) = \frac{1}{|\Gamma(p^\ell)\backslash\Gamma|} \sum_{\gamma \in \Gamma(p^\ell)\backslash\Gamma} c_{p^m}(\mathfrak{f}_\gamma - n),$$

for any $0 \le m \le \ell$, together with the fact that

$$1 + c_p(x) + \cdots + c_{p^\ell}(x) = \mathbf{1}_{\{x \equiv 0 (\mathrm{mod}\ p^\ell)\}} p^\ell.$$

□

Finally, we show that the densities stabilize.

Lemma 3.23 *Let p be a bad prime, and let $k = k_p$ be the "saturation level" of p, as in (2.4). Let $p^L \| B$ (and recall that $B \neq 0$, and that $k > L$ by (2.6)). If $\ell > 2k$, then $\mathfrak{S}_{p^\ell}(n) = 0$.*

Proof Decompose $\Gamma(p^\ell)\backslash\Gamma$ into disjoint $\Gamma_\infty = \{n_x = \left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right)\}$ cosets; here x ranges in $\mathbb{Z}/p^\ell$ but is restricted (by saturation) to $x \equiv 0(p^k)$. We claim that the Ramanujan values on each coset exactly cancel. Note that $\mathfrak{f}_\gamma = Aa + Bb + Cc + Dd$ changes when $\gamma \mapsto n_x\gamma$ to

$$\mathfrak{f}_{n_x\gamma} = \mathfrak{f}_\gamma + (Ac + Bd)x.$$

Since $c \equiv 0(\mathrm{mod}\ p^k)$ and $d \equiv 1(\mathrm{mod}\ p^k)$, and $p^L \| B$ with $k > L$, we have that

$$p^L \| Ac + Bd.$$

Now as x ranges over p^ℓ subject to $x \equiv 0(\mathrm{mod}\ p^k)$, since $\ell > 2k > k + L$, the values of $\mathfrak{f}_{n_x\gamma}$ range in some non-constant arithmetic progression. The resulting Ramanujan values cancel exactly, as claimed. □

So the high powers of bad primes have vanishing $\mathfrak{S}_q$. For the lower powers, we give the following trivial estimate on $\mathfrak{S}_q$.

Lemma 3.24 *For any prime p (good or bad) and any $\ell \ge 1$, we have:*

$$|\mathfrak{S}_{p^\ell}(n)| \le p^\ell. \tag{3.25}$$

Proof The density formula (3.22) gives upper and lower bounds for its left-hand side of: p^ℓ and 0, respectively. Replace ℓ by $\ell - 1$ and subtract to get the claim. □

3.6 Short sum of $\mathfrak{S}_q$

Define

$$\mathfrak{S}(n) = \sum_{q \ge 1} \mathfrak{S}_q(n).$$

Lemma 3.26 *Assume that $\Delta = 0$. Then the series defining $\mathfrak{S}(n)$ is absolutely convergent,*

$$\sum_{q < Q_0} \mathfrak{S}_q(n) = \mathfrak{S}(n) + O_\varepsilon(Q_0^{-1} n^\varepsilon),$$

as $Q_0 \to \infty$, and satisfies, for n admissible,

$$\frac{1}{\log \log n} \ll \mathfrak{S}(n) \ll 1.$$

Proof By Lemmas 3.11 and 3.12, we have that

$$\sum_{q \geq Q_0} |\mathfrak{S}_q(n)| \ll_\varepsilon n^\varepsilon Q_0^{-1}.$$

For n admissible,

$$\mathfrak{S}(n) \asymp \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

where we also used Lemmas 3.21 and 3.23. The claim follows immediately. □

To prepare for the case $\Delta \neq 0$, we need some preliminaries.

Lemma 3.27 *Let χ be a Dirichlet character of conductor M and fix $\Pi \in \mathbb{Z}$. Then*

$$\left| \sum_{\substack{q \asymp H \\ \text{squarefree} \\ (q,\Pi)=1}} \chi(q) \right| \ll_\varepsilon H^{1/2} M^{3/16} (HM\Pi)^\varepsilon,$$

as $H \to \infty$.

Proof To capture both the squarefree and coprime conditions, we use Möbius inversion. Using $\zeta(s)/\zeta(2s) = \sum_{n \text{ squarefree}} 1/n^s$, we have that $\mu(q)^2 = \sum_{m^2|q} \mu(m)$. Similarly, $\sum_{d|x} \mu(d) = 1$ if $x = 1$ and 0 otherwise. Therefore

$$\begin{aligned}
\sum_{\substack{q \asymp H \\ \text{squarefree} \\ (q,\Pi)=1}} \chi(q) &= \sum_{\substack{q \asymp H \\ (q,\Pi)=1}} \chi(q) \sum_{m^2|q} \mu(m) = \sum_{\substack{m \ll H^{1/2} \\ (m,\Pi)=1}} \mu(m)\chi(m)^2 \sum_{\substack{q \asymp H/m^2 \\ (q,\Pi)=1}} \chi(q) \\
&= \sum_{\substack{m \ll H^{1/2} \\ (m,\Pi)=1}} \mu(m)\chi(m)^2 \sum_{q \asymp H/m^2} \chi(q) \sum_{\substack{d|q \\ d|\Pi}} \mu(d) \\
&= \sum_{\substack{m \ll H^{1/2} \\ (m,\Pi)=1}} \mu(m)\chi(m)^2 \sum_{\substack{d \ll H/m^2 \\ d|\Pi}} \mu(d)\chi(d) \sum_{q \asymp H/(m^2 d)} \chi(q).
\end{aligned}$$

Applying Burgess [9] to the last sum, we have that

$$\left|\sum_{\substack{q\asymp H\\ \text{squarefree}\\ (q,\Pi)=1}} \chi(q)\right| \ll_\varepsilon \sum_{m\ll H^{1/2}} \sum_{\substack{d\ll H/m^2\\ d\mid\Pi}} \frac{H^{1/2}}{md^{1/2}} M^{3/16+\varepsilon} \ll_\varepsilon \Pi^\varepsilon H^{1/2+\varepsilon} M^{3/16+\varepsilon},$$

as claimed. (Slightly better estimates are available today but not needed here.) □

Going forward, we let $\mathfrak{B}_1$ be the (finitely many) primes which are "bad" (for Γ), $\mathfrak{B}_2$ be the primes not in $\mathfrak{B}_1$ which divide Δ, and $\mathfrak{B}_3 = \mathfrak{B}_3(n)$ be the primes not in $\mathfrak{B}_1$ or $\mathfrak{B}_2$ which divide $n^2 - 4\Delta$. Let $\mathfrak{B} = \mathfrak{B}(n) = \sqcup_j \mathfrak{B}_j$, and set

$$\Pi = \Pi(\mathfrak{B}) := \prod_{p\in\mathfrak{B}} p.$$

For all the other primes $p \nmid \Pi$, Lemma 3.14 gives that

$$\mathfrak{S}_p(n) = \frac{1}{p}\left(\frac{n^2-4\Delta}{p}\right) + E_p,$$

where

$$E_p = E_p(n) := \frac{p + \left(\frac{n^2-4\Delta}{p}\right)}{p(p^2-1)} \ll \frac{1}{p^2}.$$

We extend E_p to a multiplicative function E_q supported on square-free q.

Note that we now do not have absolute convergence, and must be much more careful in our analysis.

We break the tail $\sum_{q\ge Q_0}$ of $\mathfrak{S}(n)$ into dyadic regions $\sum_{q\asymp H}$, with $Q_0 \le H \to \infty$.

Lemma 3.28 *Assume that* $\Delta \neq 0$. *Then as* $H \to \infty$,

$$\left|\sum_{q\asymp H} \mathfrak{S}_q(n)\right| \ll_\varepsilon (nH)^\varepsilon n^{3/8} H^{-1/2}, \tag{3.29}$$

for any $\varepsilon > 0$.

Proof From Lemma 3.15, we have that $\mathfrak{S}_q$ vanishes if $(q,\Pi) = 1$ and q is not square-free. For q square-free and coprime to Π, we have then that

$$\mathfrak{S}_q(n) = \prod_{p\mid q}\left(\frac{1}{p}\left(\frac{n^2-4\Delta}{p}\right) + E_p\right) = \sum_{ab=q} \frac{1}{a}\left(\frac{n^2-4\Delta}{a}\right) E_b(n). \tag{3.30}$$

Write any q as

$$q = q_{\mathfrak{B}} \cdot q_1,$$

where

$$q_{\mathfrak{B}} = \prod_{\substack{p^\ell \| q \\ p \mid \Pi}} p^\ell, \text{ and } q_1 = \prod_{\substack{p^\ell \| q \\ (p,\Pi)=1}} p^\ell$$

From multiplicativity, we have that $\mathfrak{S}_q = \mathfrak{S}_{q_{\mathfrak{B}}} \cdot \mathfrak{S}_{q_1}$.

Then we have

$$\begin{aligned} \sum_{q \asymp H} \mathfrak{S}_q(n) &= \sum_{\substack{q_{\mathfrak{B}} \ll H \\ p \mid q_{\mathfrak{B}} \Longrightarrow p \mid \Pi}} \mathfrak{S}_{q_{\mathfrak{B}}}(n) \sum_{\substack{q_1 \asymp H/q_{\mathfrak{B}} \\ \text{square-free} \\ (q_1,\Pi)=1}} \mathfrak{S}_{q_1}(n) \\ &= \sum_{\substack{q_{\mathfrak{B}} \ll H \\ p \mid q_{\mathfrak{B}} \Longrightarrow p \mid \Pi}} \mathfrak{S}_{q_{\mathfrak{B}}}(n) \sum_{\substack{q_1 \asymp H/q_{\mathfrak{B}} \\ \text{square-free} \\ (q_1,\Pi)=1}} \sum_{ab=q_1} \frac{1}{a} \left(\frac{n^2 - 4\Delta}{a} \right) E_b \\ &= \sum_{\substack{q_{\mathfrak{B}} \ll H \\ p \mid q_{\mathfrak{B}} \Longrightarrow p \mid \Pi}} \mathfrak{S}_{q_{\mathfrak{B}}}(n) \sum_{\substack{b \ll H/q_{\mathfrak{B}} \\ \text{square-free, } (b,\Pi)=1}} E_b \sum_{\substack{a \asymp H/(q_{\mathfrak{B}} b) \\ \text{square-free} \\ (a,\Pi)=1}} \frac{1}{a} \left(\frac{n^2 - 4\Delta}{a} \right) \\ &\ll_\varepsilon (nH)^\varepsilon n^{3/8} H^{-1/2} \sum_{\substack{q_{\mathfrak{B}} \ll H \\ p \mid q_{\mathfrak{B}} \Longrightarrow p \mid \Pi}} \mathfrak{S}_{q_{\mathfrak{B}}}(n) q_{\mathfrak{B}}^{1/2}, \end{aligned}$$

where we used Lemma 3.27, partial summation, and $E_b \ll b^\varepsilon / b^2$.

To deal with the remaining $q_{\mathfrak{B}}$ sum, we decompose

$$\mathfrak{S}_{q_{\mathfrak{B}}}(n) q_{\mathfrak{B}}^{1/2} = \mathfrak{S}_{q_{\mathfrak{B}_1}}(n) q_{\mathfrak{B}_1}^{1/2} \cdot \mathfrak{S}_{q_{\mathfrak{B}_2}}(n) q_{\mathfrak{B}_2}^{1/2} \cdot \mathfrak{S}_{q_{\mathfrak{B}_3}}(n) q_{\mathfrak{B}_3}^{1/2},$$

corresponding to $\mathfrak{B} = \mathfrak{B}_1 \sqcup \mathfrak{B}_2 \sqcup \mathfrak{B}_3$.

Since $\mathfrak{B}_1$ is a finite set of primes which are bad for Γ, and only finitely many powers of such primes have non-vanishing $\mathfrak{S}_{q_{\mathfrak{B}_1}}$ by Lemma 3.23, the total contribution from $\mathfrak{B}_1$ is bounded by a constant depending only on Γ and the linear form $\mathscr{L}$, that is, on A, B, C, D.

Recall that $\mathfrak{B}_2$ consists of the good primes dividing Δ; Lemma 3.12 removes any non-square-free $q_{\mathfrak{B}_2}$ contributions, and Lemma 3.11 otherwise gives $\mathfrak{S}_{q_{\mathfrak{B}_2}}(n) \ll 1/q_{\mathfrak{B}_2}$. So again this contribution is bounded.

Finally, for $\mathfrak{B}_3$, we use (3.17) to offset the factor of $q_{\mathfrak{B}_3}^{1/2}$, and (3.16) to kill the contribution from any powers ℓ of p^ℓ in $q_{\mathfrak{B}_3}$ unless $\ell \leq L + 1 \leq 2L$, where $p^L \| n^2 - 4\Delta$. Therefore the only $q_{\mathfrak{B}_3}$ contributing to the sum are divisors of $(n^2 - 4\Delta)^2$, and the number of such is $\ll n^\varepsilon$. This gives the claim. □

Lemma 3.28 is sufficient to show that $\mathfrak{S}(n)$ converges (conditionally, not absolutely), but does not allow us a good enough error estimate for the very short sum $\sum_{q<Q_0}\mathfrak{S}_q(n)$, since H needs to be at least $n^{3/4+}$ for (3.29) to decay. If we replaced our use of Burgess with GRH, we could get good estimates with H as small as Q_0, which is a tiny power of N. Unconditionally, we can only do this on average, as follows.

Theorem 3.31 *As $H\to\infty$, we have*

$$\sum_{n\ll N}\left|\sum_{q\asymp H}\mathfrak{S}_q(n)\right|^2 \ll_\varepsilon H^\varepsilon\left(H+\frac{N}{H^{1/2}}\right),$$

for any $\varepsilon>0$.

Proof As before, let $\mathfrak{B}_1$ be the "bad" primes for Γ, and $\mathfrak{B}_2$ be the primes not in $\mathfrak{B}_1$ which divide Δ. Since $\mathfrak{B}_3$ depends on n, we now have to handle it separately. We now write $\mathfrak{B}=\mathfrak{B}_1\sqcup\mathfrak{B}_2$ and $\Pi:=\prod_{p\in\mathfrak{B}}p$ as before, and decompose

$$q=q_{\mathfrak{B}}\cdot q_1,$$

with $(q_1,\Pi)=1$. Furthermore, we will split off the square-full part of q_1, writing $q_1=q_2\cdot q_3$, where

$$q_2:=\prod_{p\|q_1}p,\quad q_3:=q_1/q_2=\prod_{\substack{p^\ell\|q_1\\ \ell\geq 2}}p^\ell.$$

With this decomposition, we open the square and reverse orders:

$$\begin{aligned}
&\sum_{n\ll N}\left|\sum_{q\asymp H}\mathfrak{S}_q(n)\right|^2\\
&=\sum_{n\ll N}\left|\sum_{\substack{q_{\mathfrak{B}}\ll H\\ p|q_{\mathfrak{B}}\Longrightarrow p\in\mathfrak{B}}}\mathfrak{S}_{q_{\mathfrak{B}}}(n)\sum_{\substack{q_3\ll H/q_{\mathfrak{B}}\\ (q_3,\Pi)=1,\ \text{square-full}}}\mathfrak{S}_{q_3}(n)\sum_{\substack{q_2\asymp H/(q_{\mathfrak{B}}q_3)\\ (q_2,\Pi)=1=(q_2,q_3),\ \text{square-free}}}\mathfrak{S}_{q_2}(n)\right|^2\\
&=\sum_{q_{\mathfrak{B}},q_3,q_2,q'_{\mathfrak{B}},q'_3,q'_2}\ \sum_{n\ll N}\mathfrak{S}_{q_{\mathfrak{B}}}(n)\mathfrak{S}_{q'_{\mathfrak{B}}}(n)\mathfrak{S}_{q_3}(n)\mathfrak{S}_{q'_3}(n)\mathfrak{S}_{q_2}(n)\mathfrak{S}_{q'_2}(n)
\end{aligned}$$

Here, instead of using the decomposition (3.30), we return to Lemma 3.14 and write

$$\mathfrak{S}_p(n)=\left(\frac{n^2-4\Delta}{p}\right)\frac{p}{p^2-1}+E_p,$$

where

$$E_p = \frac{1}{p^2 - 1}.$$

The crucial fact for our purposes here is that E_p is now independent of n. Extending E_p to a multiplicative function on square-frees gives

$$E_q \ll \frac{1}{q^2}, \tag{3.32}$$

where we used that $\prod_{p|q}(1 - 1/p^2) \asymp 1$. Then we can write, for any q square-free and coprime to Π, that

$$\mathfrak{S}_q(n) = \prod_{p|q} \left(\left(\frac{n^2 - 4\Delta}{p} \right) \frac{p}{p^2 - 1} + E_p \right) = \sum_{ab=q} \psi(a) \left(\frac{n^2 - 4\Delta}{a} \right) E(b),$$

where ψ is a multiplicative function supported on square-free numbers taking the value $\psi(p) = p/(p^2 - 1)$ on primes. In particular,

$$\psi(q) \ll \frac{1}{q}. \tag{3.33}$$

For q_2 and q_2', we insert this expression to get:

$$\begin{aligned}
\sum_{n \ll N} \left| \sum_{q \asymp H} \mathfrak{S}_q(n) \right|^2 = & \sum_{q_{\mathfrak{B}}, q_3, q_2, q_{\mathfrak{B}}', q_3', q_2'} \sum_{n \ll N} \mathfrak{S}_{q_{\mathfrak{B}}}(n) \mathfrak{S}_{q_{\mathfrak{B}}'}(n) \mathfrak{S}_{q_3}(n) \mathfrak{S}_{q_3'}(n) \\
& \times \sum_{ab=q_2} \psi(a) \left(\frac{n^2 - 4\Delta}{a} \right) E_b \sum_{a'b'=q_2'} \psi(a') \left(\frac{n^2 - 4\Delta}{a'} \right) E_{b'} \\
\leq & \sum_{q_{\mathfrak{B}}, q_3, q_2, q_{\mathfrak{B}}', q_3', q_2'} \sum_{ab=q_2} \sum_{a'b'=q_2'} \psi(a) \psi(a') |E_b E_{b'}| \\
& \times \sum_{n_0 \bmod \tilde{q}} \left| \mathfrak{S}_{q_{\mathfrak{B}}}(n_0) \mathfrak{S}_{q_{\mathfrak{B}}'}(n_0) \mathfrak{S}_{q_3}(n_0) \mathfrak{S}_{q_3'}(n_0) \right| \\
& \times \left| \sum_{\substack{n \ll N \\ n \equiv n_0 (\mathrm{mod}\, \tilde{q})}} \left(\frac{n^2 - 4\Delta}{a} \right) \left(\frac{n^2 - 4\Delta}{a'} \right) \right|,
\end{aligned} \tag{3.34}$$

where we have decomposed n into progressions mod $\tilde{q}$, where

$$\tilde{q} := [q_{\mathfrak{B}}, q_{\mathfrak{B}}', q_3, q_3'].$$

While it is clear at, say, a and q_3 are coprime (by construction, since $a \mid q_2$), we actually also have that there is no contribution unless a and q_3' are coprime. Indeed, if $p \mid a$ and $p \mid q_3'$, then either $n^2 - 4\Delta \equiv 0(p)$, in which case $\left(\frac{n^2-4\Delta}{a}\right)$ vanishes, or else $\mathfrak{S}_{q_3'}(n)$ vanishes from (3.16) and the square-full-ness of q_3'. Therefore, we may restrict the summations to

$$(a, \tilde{q}) = (a', \tilde{q}) = 1.$$

We first analyze the last n sum. Let $\tilde{a} := aa'/(a, a')^2$, so that

$$\left(\frac{n^2 - 4\Delta}{a}\right)\left(\frac{n^2 - 4\Delta}{a'}\right) = \left(\frac{n^2 - 4\Delta}{\tilde{a}}\right),$$

for n such that $n^2 - 4\Delta$ is coprime to (a, a'). (Otherwise the characters vanish.) Breaking the n sum further into residue classes mod $\tilde{a}$ gives:

$$\sum_{\substack{n \ll N \\ n \equiv n_0 (\mathrm{mod}\, \tilde{q}) \\ (n^2 - 4\Delta, a, a') = 1}} \left(\frac{n^2 - 4\Delta}{\tilde{a}}\right) = \sum_{m (\mathrm{mod}\, \tilde{a})} \left(\frac{m^2 - 4\Delta}{\tilde{a}}\right) \left[\sum_{\substack{n \ll N \\ n \equiv n_0 (\mathrm{mod}\, \tilde{q}),\ n \equiv m (\mathrm{mod}\, \tilde{a}) \\ (n^2 - 4\Delta, a, a') = 1}} 1 \right]$$

We wish to get square-root cancellation from Weil in the m summation, but the n sum may be incomplete, which will give too large an error in terms of m. So we separate the roles of n and m by completing the sum.

$$\sum_{\substack{n \ll N \\ n \equiv n_0 (\mathrm{mod}\, \tilde{q}) \\ (n^2 - 4\Delta, a, a') = 1}} \left(\frac{n^2 - 4\Delta}{\tilde{a}}\right) = \frac{1}{\tilde{a}} \sum_{k(\tilde{a})} \left[\sum_{m (\mathrm{mod}\, \tilde{a})} \left(\frac{m^2 - 4\Delta}{\tilde{a}}\right) e_{\tilde{a}}(-km) \right] \times \left[\sum_{\substack{n \ll N \\ n \equiv n_0 (\mathrm{mod}\, \tilde{q}) \\ (n^2 - 4\Delta, a, a') = 1}} e_{\tilde{a}}(kn) \right]. \tag{3.35}$$

Now the m sum is free and is bounded by $\tilde{a}^{1/2+\varepsilon}$ by Weil. We now deal with the last n sum. We remove the gcd condition via Möbius inversion.

$$\begin{aligned}\sum_{\substack{n\ll N\\ n\equiv n_0(\mathrm{mod}\,\tilde{q})\\ (n^2-4\Delta,a,a')=1}} e_{\tilde{a}}(kn) &= \sum_{d|(a,a')}\mu(d)\sum_{\substack{n\ll N\\ n\equiv n_0(\mathrm{mod}\,\tilde{q})\\ n^2-4\Delta\equiv 0(\mathrm{mod}\,d)}} e_{\tilde{a}}(kn)\\ &= \sum_{d|(a,a')}\mu(d)\sum_{\substack{m_0(\mathrm{mod}\,d)\\ m_0^2\equiv 4\Delta(\mathrm{mod}\,d)}}\left[\sum_{\substack{n\ll N\\ n\equiv n_0(\mathrm{mod}\,\tilde{q})\\ n\equiv m_0(\mathrm{mod}\,d)}} e_{\tilde{a}}(kn)\right], \qquad (3.36)\end{aligned}$$

where we decomposed the n sum further into residue classes $m_0 \bmod d$. Note that $d \mid a$ is square-free, and for each $p \mid d$, there are at most two solutions to $m_0^2 \equiv 4\Delta(p)$, so the number of m_0 is at most d^{ε}. The last bracketed sum restricts n to a residue class x, say, mod $\tilde{q}d$ (since $(d,\tilde{q}) = (a,\tilde{q}) = 1$). Changing $n \mapsto x + n\tilde{q}d$, the bracketed term is a geometric series, giving:

$$\begin{aligned}\left|\sum_{\substack{n\ll N\\ n\equiv n_0(\mathrm{mod}\,\tilde{q})\\ n\equiv m_0(\mathrm{mod}\,d)}} e_{\tilde{a}}(kn)\right| &= \left|e_{\tilde{a}}(kx)\sum_{n\ll N/(\tilde{q}d)} e_{\tilde{a}}(k\tilde{q}dn)\right|\\ &\ll \min\left(\frac{N}{\tilde{q}}+1, \frac{1}{\|\frac{k\tilde{q}d}{\tilde{a}}\|}\right), \qquad (3.37)\end{aligned}$$

where $\|\cdot\|$ is the distance to the nearest integer. Inserting (3.37) into (3.36) and into (3.35) gives

$$\left|\sum_{\substack{n\ll N\\ n\equiv n_0(\mathrm{mod}\,\tilde{q})\\ (n^2-4\Delta,a,a')=1}}\left(\frac{n^2-4\Delta}{\tilde{a}}\right)\right| \ll_{\varepsilon} \tilde{a}^{1/2+\varepsilon}\sum_{d|(a,a')} d^{\varepsilon}\frac{1}{\tilde{a}}\sum_{k(\tilde{a})}\min\left(\frac{N}{\tilde{q}}+1, \frac{1}{\|\frac{k\tilde{q}d}{\tilde{a}}\|}\right).$$

Since a and a' are square-free, d is coprime to $\tilde{a}$, and hence $(\tilde{q}d, \tilde{a}) = 1$. So the k sum is invariant under $k \mapsto k\overline{\tilde{q}d}$. This finally gives

$$\left|\sum_{\substack{n\ll N\\ n\equiv n_0(\mathrm{mod}\,\tilde{q})\\ (n^2-4\Delta,a,a')=1}}\left(\frac{n^2-4\Delta}{\tilde{a}}\right)\right| \ll_{\varepsilon} \tilde{a}^{\varepsilon}\left(\frac{N}{\tilde{q}\tilde{a}^{1/2}} + \tilde{a}^{1/2}\right).$$

Returning to (3.34), we get that

$$\begin{aligned}\sum_{n\ll N}\left|\sum_{q\asymp H}\mathfrak{S}_q(n)\right|^2 \ll H^{\varepsilon} & \sum_{q_{\mathfrak{B}},q_3,q_2,q'_{\mathfrak{B}},q'_3,q'_2}\ \sum_{ab=q_2}\ \sum_{a'b'=q'_2}\frac{1}{aa'b^2b'^2}\\ & \times\sum_{n_0 \bmod \tilde{q}}\left|\mathfrak{S}_{q_{\mathfrak{B}}}(n_0)\mathfrak{S}_{q'_{\mathfrak{B}}}(n_0)\mathfrak{S}_{q_3}(n_0)\mathfrak{S}_{q'_3}(n_0)\right|\\ & \times\left(\frac{N(a,a')}{\tilde{q}(aa')^{1/2}}+\frac{(aa')^{1/2}}{(a,a')}\right),\end{aligned}$$

where we used (3.33) and (3.32).

Next we analyze the contributions from $q_{\mathfrak{B}}, q'_{\mathfrak{B}}$. Combining Lemma 3.12 and Lemma 3.11 (for $q_{\mathfrak{B}_2}$) with Lemma 3.23 and (3.25) (for $q_{\mathfrak{B}_1}$), we see that in fact there is no contribution unless $q_{\mathfrak{B}}, q'_{\mathfrak{B}} \ll 1$, and in this case the constribution is $\mathfrak{S}_{q_{\mathfrak{B}}}\mathfrak{S}_{q'_{\mathfrak{B}}} \ll 1$. Therefore $\tilde{q} \asymp [q_3, q'_3]$.

Recall from (3.17) that $\mathfrak{S}_{q_3}(n_0) \ll q_3^{-1/2}$. Finally, we analyze the number of $n_0 (\bmod\, \tilde{q})$ for which $\mathfrak{S}_{q_3}\mathfrak{S}_{q'_3}$ is non-vanishing. Suppose that $p^m \| [q_3, q'_3]$. Then since $\Delta \not\equiv 0(p)$, (3.16) shows that, if $\mathfrak{S}_{p^m}(n_0) \neq 0$, then $n_0^2 - 4\Delta \equiv 0 (\bmod\, p^{m-1})$. The number of such $n_0 (\bmod\, p^m)$ is at most $2p \ll p^{m/2}$, since $m \geq 2$. So the number of $n_0 (\bmod\, \tilde{q})$ which contribute is $\ll_{\varepsilon} \tilde{q}^{1/2+\varepsilon}$.

Putting everything together gives

$$\begin{aligned}\sum_{n\ll N}\left|\sum_{q\asymp H}\mathfrak{S}_q(n)\right|^2 & \ll_{\varepsilon} H^{\varepsilon}\sum_{q_3,q'_3}([q_3,q'_3])^{1/2}q_3^{-1/2}q_3'^{-1/2}\\ & \quad\times\sum_{q_2,q'_2}\sum_{a\mid q_2}\sum_{a'\mid q'_2}\frac{aa'}{q_2^2q_2'^2}\left(\frac{N(a,a')}{[q_3,q'_3](aa')^{1/2}}+\frac{(aa')^{1/2}}{(a,a')}\right).\end{aligned}$$

Let $t := (a, a')$, which is a divisor of (q_2, q'_2), and let $a_1 := a/t$ and $a'_1 := a'/t$. Then

$$\begin{aligned}\sum_{n\ll N}\left|\sum_{q\asymp H}\mathfrak{S}_q(n)\right|^2 & \ll_{\varepsilon} H^{\varepsilon}\sum_{q_3,q'_3}([q_3,q'_3])^{1/2}q_3^{-1/2}q_3'^{-1/2}\\ & \quad\times\sum_{q_2,q'_2}\frac{1}{q_2^2q_2'^2}\sum_{t\mid(q_2,q'_2)}t^2\sum_{a_1\mid\frac{q_2}{t}}\sum_{a'_1\mid\frac{q'_2}{t}}\left(\frac{N(a_1a'_1)^{1/2}}{[q_3,q'_3]}+(a_1a'_1)^{3/2}\right)\\ & \ll H^{\varepsilon}\sum_{\substack{q_3\ll H\\ (q_3,\Pi)=1\\ \text{square-full}}}\ \sum_{\substack{q'_3\ll H\\ (q'_3,\Pi)=1\\ \text{square-full}}}([q_3,q'_3])^{1/2}q_3^{-1/2}q_3'^{-1/2}\end{aligned}$$

$$\times \left(\frac{H}{(q_3 q_3')^{1/2}} + \frac{N}{[q_3, q_3']} \sum_{\substack{q_2 \asymp H/q_3 \\ (q_2,\Pi)=1=(q_2,q_3) \\ \text{square-free}}} \sum_{\substack{q_2' \asymp H/q_3' \\ (q_2',\Pi)=1=(q_2',q_3') \\ \text{square-free}}} \frac{(q_2, q_2')}{(q_2 q_2')^{3/2}} \right).$$

Next we need some cancellation from (q_2, q_2'). Let $d = (q_2, q_2')$ which is a divisor of q_2 such that $q_2' \equiv 0(d)$.

$$\sum_{n \ll N} \left| \sum_{q \asymp H} \mathfrak{S}_q(n) \right|^2 \ll_\varepsilon H^\varepsilon \sum_{\substack{q_3 \ll H \\ (q_3,\Pi)=1 \\ \text{square-full}}} \sum_{\substack{q_3' \ll H \\ (q_3',\Pi)=1 \\ \text{square-full}}} ([q_3, q_3'])^{1/2} q_3^{-1/2} q_3^{\prime -1/2}$$

$$\times \left(\frac{H}{(q_3 q_3')^{1/2}} + \frac{N}{[q_3, q_3']} \sum_{\substack{q_2 \asymp H/q_3 \\ (q_2,\Pi)=1=(q_2,q_3) \\ \text{square-free}}} \sum_{d \mid q_2} \sum_{\substack{q_2' \asymp H/(q_3' d) \\ (q_2',\Pi)=1=(q_2',q_3') \\ \text{square-free}}} \frac{d}{(q_2 q_2' d)^{3/2}} \right)$$

$$\ll_\varepsilon H^\varepsilon \sum_{\substack{q_3 \ll H \\ (q_3,\Pi)=1 \\ \text{square-full}}} \sum_{\substack{q_3' \ll H \\ (q_3',\Pi)=1 \\ \text{square-full}}} \left(H \frac{1}{(q_3 q_3')^{1/2} (q_3, q_3')^{1/2}} + \frac{N}{H} \frac{(q_3, q_3')^{1/2}}{(q_3 q_3')^{1/2}} \right).$$

Finally, we bound $(q_3, q_3') \ll H$ in the numerator and $(q_3, q_3') \geq 1$ in the denominator. It remains to estimate a sum of the form

$$\sum_{\substack{q_3 \ll H \\ \text{square-full}}} \frac{1}{q_3^{1/2}}.$$

Since q_3 is square-full, any such q_3 can be written as $q_3 = k^2 \ell$ where $\ell \mid k$. Then

$$\sum_{\substack{q_3 \ll H \\ \text{square-full}}} \frac{1}{q_3^{1/2}} \ll \sum_{k^2 \ll H} \sum_{\ell \mid k} \frac{1}{k \ell^{1/2}} \ll_\varepsilon \sum_{k^2 \ll H} k^\varepsilon \frac{1}{k} \ll H^\varepsilon.$$

The claim follows immediately. □

Theorem 3.31 allows us to show, for almost all n (with power savings error), that the very short sum $\sum_{q \leq Q_0} \mathfrak{S}_q(n_0)$ (with $Q_0 = N^{\alpha_0}$, $\alpha_0 > 0$ small) is a good approximation (also with power savings error) for $\mathfrak{S}(n)$.

Theorem 3.38 *For any $\eta > 0$ with $\eta < \frac{1}{6}\alpha_0$, there is a set $\mathscr{E}$ of "exceptional" n of cardinality*

$$\mathscr{E} \cap [1, N] \ll N^{1-\eta}$$

such that, for all $n \asymp N$, $n \notin \mathscr{E}$,

$$\sum_{q \le Q_0} \mathfrak{S}_q(n) \;=\; \mathfrak{S}(n) + O(N^{-\eta}).$$

Proof Recall that the series $\mathfrak{S}(n)$ does converge (conditionally) by Lemma 3.28, but the error there is insufficient to approximate it to the required error in all ranges of H.

For $\eta > 0$ fixed, let

$$\mathscr{E}(N) := \left\{ n \in [1, N] : \left| \mathfrak{S}(n) - \sum_{q \le Q_0} \mathfrak{S}_q(n) \right| \ge N^{-\eta} \right\}.$$

We estimate

$$\begin{aligned}
\#\mathscr{E}(N) &= \sum_{\substack{n \ll N \\ \left|\mathfrak{S}(n) - \sum_{q \le Q_0} \mathfrak{S}_q(n)\right| \ge N^{-\eta}}} 1 \;\le\; N^{2\eta} \sum_{n \ll N} \left| \mathfrak{S}(n) - \sum_{q \le Q_0} \mathfrak{S}_q(n) \right|^2 \\
&\ll N^{2\eta} \sum_{n \ll N} \left(\sum_{\substack{Q_0 < H < N^{4/5} \\ \text{dyadic}}} \left| \sum_{q \asymp H} \mathfrak{S}_q(n) \right|^2 + \sum_{\substack{H \ge N^{4/5} \\ \text{dyadic}}} \left| \sum_{q \asymp H} \mathfrak{S}_q(n) \right|^2 \right).
\end{aligned}$$

We apply Theorem 3.31 in the first term, and Lemma 3.28 (individually) in the second term.

$$\begin{aligned}
\#\mathscr{E}(N) &\ll_\varepsilon N^{2\eta+\varepsilon} \left(N^{4/5} + \frac{N}{Q_0^{1/2}} + \sum_{\substack{H \ge N^{4/5} \\ \text{dyadic}}} \sum_{n \ll N} n^{3/4} H^{-1} \right) \\
&\ll N^{2\eta+\varepsilon} \left(N^{4/5} + \frac{N}{Q_0^{1/2}} + N^{7/4} N^{-4/5} \right).
\end{aligned}$$

Since $7/4 - 4/5 = 19/20 < 1$, we have a power savings as long as $2\eta < \frac{1}{2}\alpha_0$, where $Q_0 = N^{\alpha_0}$. As long as $\eta < \frac{1}{6}\alpha_0$, we are guaranteed to have $\mathscr{E}(N) \ll N^{1-\eta}$. This completes the proof. □

Theorem 3.39 *For all admissible $n \asymp N$, and all $\varepsilon > 0$,*

$$\mathfrak{S}(n) \gg_\varepsilon n^{-\varepsilon} L(1, \chi_n),$$

where

$$L(1, \chi_n) := \prod_p \left(1 - \frac{1}{p}\left(\frac{n^2 - 4\Delta}{p}\right)\right)^{-1}. \tag{3.40}$$

The implied constant is effective.

Proof By the multiplicativity of $\mathfrak{S}_q$, we have that

$$\mathfrak{S}(n) = \prod_p \left(1 + \mathfrak{S}_p(n) + \mathfrak{S}_{p^2}(n) + \cdots\right).$$

(Since the series only converges conditionally, we argue by considering the functions $s \mapsto \sum_{q \in \mathbb{N}} \mathfrak{S}_q(n) q^{-s}$ and $s \mapsto \prod_p (1 + \mathfrak{S}_p(n) p^{-s} + \mathfrak{S}_{p^2}(n) p^{-2s} + \cdots)$; for $\mathfrak{Re}(s) > 0$, both converge absolutely and coincide, and hence their limiting values as $s \to 0^+$ do too.)

For $p \in \mathfrak{B}_1$ a "bad" prime for Γ, this is a finite sum (Lemma 3.23) which is non-vanishing only if n is admissible by Lemma 3.21. For the other primes p, the Euler factor is $(1 + \mathfrak{S}_p(n) + \mathfrak{S}_{p^2}(n))(1 + O(p^{-3/2}))$. Recall that $\mathfrak{B}_2$ contains the (finite list of) primes $p \mid \Delta$. By Lemma 3.11 and Lemma 3.12, we have

$$\prod_{p \in \mathfrak{B}_2} (1 + \mathfrak{S}_p(n)) \gg \frac{1}{\log \log n}.$$

For all other primes we apply Corollary 3.20. If $p \nmid \Delta$ and $p \nmid n^2 - 4\Delta$, we have

$$1 + \mathfrak{S}_p(n) + \mathfrak{S}_{p^2}(n) = 1 + \frac{1}{p}\left(\frac{n^2 - 4\Delta}{p}\right) + O(p^{-2}),$$

while if $p \nmid \Delta$ but $p \mid n^2 - 4\Delta$,

$$1 + \mathfrak{S}_p(n) + \mathfrak{S}_{p^2}(n) = 1 + O(p^{-1}).$$

The product of the latter (finite set of primes) is $\gg_\varepsilon n^{-\varepsilon}$. □

By Siegel's theorem, $L(1, \chi_n) \gg_\varepsilon n^{-\varepsilon}$ with an *ineffective* implied constant. But since we anyway only prove our result on average over n, we want to make this constant effective.

Theorem 3.41 *There is an exceptional set $\mathscr{E}$ with the following property. For all admissible $n \asymp N$ outside of $\mathscr{E}$, and any $\varepsilon > 0$, we have*

$$\mathfrak{S}(n) \gg_\varepsilon n^{-\varepsilon}.$$

Moreover

$$\#\mathscr{E} \ll_{\varepsilon} N^{\varepsilon}. \tag{3.42}$$

The implied constants are all effective. (But the exact determination of the exceptional set $\mathscr{E}$ is ineffective!)

Proof Consider the characters $\chi_n = \left(\frac{n^2-4\Delta}{\cdot}\right)$ appearing in (3.40). These need not be primitive, and are induced from characters $\left(\frac{q_n}{\cdot}\right)$, where $q_n := \mathrm{sqf}(n^2 - 4\Delta) \ll N^2$ is the square-free part of $n^2 - 4\Delta$; that is,

$$n^2 - 4\Delta = q_n m^2, \tag{3.43}$$

for some integer m. Group admissible $n \asymp N$ according to the values of q_n; that is, for a given square-free $q \ll N^2$, let

$$\mathcal{N}_q := \{n \asymp N : \mathrm{sqf}(n^2 - 4\Delta) = q\}.$$

If (n, m) is a solution to $n^2 - qm^2 = 4\Delta$, then the ideal $(n + \sqrt{q}m)$ in $\mathbb{Z}[\sqrt{q}]$ has norm $|4\Delta|$. The prime ideals $\mathfrak{p}$ dividing $(n + \sqrt{q}m)$ and their multiplicities are bounded in terms of those of the rational primes dividing 4Δ (which is fixed). Therefore there are $\ll 1$ inequivalent solutions to (3.43), and equivalent solutions grow exponentially in terms of the units in $\mathbb{Z}[\sqrt{q}]$. Therefore

$$\max_{q \ll N^2} \#\mathcal{N}_q \ll_{\varepsilon} N^{\varepsilon}, \tag{3.44}$$

for any $\varepsilon > 0$ with absolute implied constants.

By Landau's theorem (see, e.g., [14, Theorem 5.28]), there is an absolute constant $A > 0$, such that for all distinct primitive real characters χ, χ' of conductors q, q' (resp.), with L-functions $L(s, \chi)$, $L(s, \chi')$ having largest real zeros β, β' (resp.), we have:

$$\max(\beta, \beta') \leq 1 - \frac{A}{\log(qq')}.$$

Therefore, there is at most a single exceptional $\mathfrak{q} \ll N^2$ such that, for all other square-free $q \ll N^2$ and their corresponding largest real zeros β (if any such exist), we have

$$\beta \leq 1 - \frac{A'}{\log N},$$

where $A' > 0$ is another absolute constant.

We then define the exceptional set $\mathscr{E} := \mathcal{N}_{\mathfrak{q}}$, so that the bound (3.42) is confirmed by (3.44), again with absolute constants. (Though we cannot effectively determine the elements of $\mathscr{E}$, we can effectively control their cardinality.)

Then we use standard arguments (see, e.g., [13]), and take into account the imprimitive factors, to show that $L(1,\chi_n) \gg_\varepsilon N^{-\varepsilon}$ with absolute implied constants, for all $n \notin \mathscr{E}$. This gives the claim. □

4 Minor arc technical estimates

We collect here various lemmata needed in the analysis of the minor arcs. We begin by defining the exponential sum

$$\mathcal{S}_q(r,k,\ell;\gamma) \;=\; \frac{1}{q^2}\sum_{x(q)}\sum_{y(q)} e_q(r\mathfrak{f}_\gamma(x,y)+kx+\ell y). \tag{4.1}$$

Lemma 4.2 *Assume that* $(r,q)=1$. *Write* $q_1 := (BcP^2,q)$, $q=q_1q_2$, *and* $BcP^2 = q_1E$, *with* $E\bar{E}\equiv 1(\mathrm{mod}\, q_2)$. *Then*

$$\mathcal{S}_q(r,k,\ell;\gamma) = \frac{(BcP^2,q)}{q} e_q(r(Aa+Bb+Cc+Dd))\mathbf{1}_{\left\{ {-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}\, q_1) \atop -k\equiv Pr(Ac+Bd)(\mathrm{mod}\, q_1)} \right\}}$$
$$\times e_{qq_1}\left(-\bar{r}\bar{E}(Pr(Ba+Dc)+\ell)(Pr(Ac+Bd)+k)\right)$$

Note that the last exponential term is well-defined by the congruence conditions on ℓ *and* k, *and independent of the lifts of* $\bar{r}$, $\bar{E}$ *to* $\mathbb{Z}/(qq_1)$.

Proof Write $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$ and insert (2.8):

$$\mathcal{S}_q(r,k,\ell;\gamma) = \frac{1}{q^2} e_q(r(Aa+Bb+Cc+Dd))\sum_{x(q)} e_q(r(Ac+Bd)Px+kx)$$
$$\times \sum_{y(q)} e_q(y[r((Ba+Dc)P+BcP^2x)+\ell]).$$

The y sum vanishes unless

$$BcP^2x \equiv -\ell\bar{r} - (Ba+Dc)P \pmod{q}, \tag{4.3}$$

in which case the sum contributes q.

Let $q_1 = \gcd(BcP^2,q)$ and write $q=q_1q_2$ and $BcP^2 = q_1E$. Then (4.3) has a solution only if the right hand side is congruent to zero mod q_1. If this is the case, then x is determined mod q_2,

$$x \equiv x_0 := -\overline{E}\frac{\ell\bar{r}+(Ba+Dc)P}{q_1} \pmod{q_2}.$$

So $x \equiv x_0 + q_2 x'$, where $x' \in \mathbb{Z}/q_1$. Thus we have:

$$\begin{aligned}\mathcal{S}_q(r,k,\ell;\gamma) = {} & e_q(r(Aa+Bb+Cc+Dd))\mathbf{1}_{\{-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}(BcP^2,q))\}}\\ & \times\frac{1}{q}e_q(x_0(r(Ac+Bd)P+k))\sum_{x'(q_1)}e_{q_1}(x'(r(Ac+Bd)P+k)).\end{aligned}$$

The x' sum vanishes unless

$$-k \equiv Pr(Ac+Bd)(\mathrm{mod}\, q_1),$$

in which case it contributes q_1. □

Next we need cancellation over the r sum on the product of two such. A preliminary calculation is the following.

Lemma 4.4 *Assume that* $(r,q)=1$ *as before, and also use* $\left(\begin{smallmatrix}a' & b'\\ c' & d'\end{smallmatrix}\right)=\gamma'$. *Write* $q_1 := (BcP^2, q)$, $q = q_1 q_2$, *and* $BcP^2 = q_1 E$, *with* $E\bar{E}\equiv 1(\mathrm{mod}\, q_2)$. *Similarly, set* $q_1' := (Bc'P^2, q)$, $q = q_1' q_2'$, *and* $Bc'P^2 = q_1' E'$, *with* $E'\bar{E}'\equiv 1(\mathrm{mod}\, q_2')$. *Then*

$$\begin{aligned}&\sideset{}{'}\sum_{r(q)}\mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\\ &\quad=\frac{(BcP^2,q)}{q}\frac{(Bc'P^2,q)}{q}\frac{\phi(q)}{\phi(qq_1q_1')}\\ &\qquad\sideset{}{'}\sum_{r(qq_1q_1')}\mathbf{1}_{\substack{\{-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}\, q_1)\\ -k\equiv Pr(Ac+Bd)(\mathrm{mod}\, q_1)\}}}\mathbf{1}_{\substack{\{-\ell'\equiv Pr(Ba'+Dc')(\mathrm{mod}\, q_1')\\ -k'\equiv Pr(Ac'+Bd')(\mathrm{mod}\, q_1')\}}}\\ &\qquad\times e_{qq_1q_1'}(rJ+\bar{r}K+L),\end{aligned}\tag{4.5}$$

where

$$\begin{aligned}J :=&q_1q_1'(Aa+Bb+Cc+Dd)-q_1q_1'(Aa'+Bb'+Cc'+Dd')\\ &-q_1'\bar{E}(q_1EaA+aB^2dP^2+Ac^2DP^2+q_1EdD)\\ &+q_1\bar{E}'(q_1'E'a'A+a'B^2d'P^2+Ac'^2DP^2+q_1'E'd'D),\end{aligned}$$

$$K := -q_1'\bar{E}\ell k + q_1\bar{E}'\ell'k',$$

and

$$\begin{aligned}L := -&q_1'\bar{E}P(Ac\ell+Bak+Bd\ell+Dck)\\ &+q_1\bar{E}'P(Ac'\ell'+Ba'k'+Bd'\ell'+Dc'k').\end{aligned}$$

Moreover,

$$J \equiv \bar{E}\bar{E}'BP^4(c-c')(B^2-\Delta cc') \qquad (\mathrm{mod}(q_2,q_2')),\tag{4.6}$$

where $\Delta = AD - BC$.

Note that for every value of $r(\mathrm{mod}\, qq_1q_1')$ *occurring in (4.5), we have that*

$$rJ + \bar{r}K + L \equiv 0(\mathrm{mod}\, q_1q_1'). \tag{4.7}$$

Proof Inserting Lemma 4.2, and extending the r sum to modulus qq_1q_1' (which overcounts by a factor of $\phi(qq_1q_1')/\phi(q)$), we have that

$$\begin{aligned}
&\sum_{r(q)}{}' \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\\
&\quad= \frac{(BcP^2,q)}{q}\frac{(Bc'P^2,q)}{q}\frac{\phi(q)}{\phi(qq_1q_1')}\\
&\qquad\sum_{r(qq_1q_1')}{}' \mathbf{1}_{\substack{\{-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}\, q_1)\\ -k\equiv Pr(Ac+Bd)(\mathrm{mod}\, q_1)\}}}\mathbf{1}_{\substack{\{-\ell'\equiv Pr(Ba'+Dc')(\mathrm{mod}\, q_1')\\ -k'\equiv Pr(Ac'+Bd')(\mathrm{mod}\, q_1')\}}}\\
&\qquad e_{qq_1q_1'}(rq_1q_1'(Aa+Bb+Cc+Dd))\\
&\qquad e_{qq_1q_1'}(-rq_1q_1'(Aa'+Bb'+Cc'+Dd'))\\
&\qquad e_{qq_1q_1'}\left(-\bar{r}q_1'\bar{E}(Pr(Ba+Dc)+\ell)(Pr(Ac+Bd)+k)\right)\\
&\qquad e_{qq_1q_1'}\left(\bar{r}q_1\bar{E}'(Pr(Ba'+Dc')+\ell')(Pr(Ac'+Bd')+k')\right).
\end{aligned}$$

By the congruence restrictions on r, the values of $\bar{E}$ and $\bar{E}'$ are independent of their lifts to $\mathbb{Z}/(qq_1q_1')$. Collecting terms gives (4.5).

In the modulus qq_1q_1', we do not know that, for example, $E\bar{E} \equiv 1$, since we took arbitrary lifts. But this does hold when reduced mod q_2, or any divisor thereof. Therefore to prove (4.6), we compute J mod (q_2, q_2'), as follows.

$$\begin{aligned}
J &\equiv -q_1'\bar{E}P^2(B^2+\Delta c^2) + q_1\bar{E}'P^2(B^2+\Delta c'^2)\\
&\equiv \bar{E}\bar{E}'BP^4(c-c')(B^2-\Delta cc') \qquad (\mathrm{mod}(q_2,q_2')),
\end{aligned}$$

where we used $ad - bc = a'd' - b'c' = 1$.

To see (4.7), observe that in the above analysis, the exponential sum modulus has actually been q all along, with some artificial replacements of terms like $e_q(X)$ by $e_{qq_1}(q_1X)$ etc. □

As a corollary, we record a simplified version of this lemma.

Corollary 4.8 *With the same notation as Lemma 4.4, we have:*

$$\left|\sum_{r(q)}{}' \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\right| \ll \frac{(q_1,q_1')}{q}.$$

Proof Returning to Lemma 4.4, we estimate

$$\left|\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\right| \le \frac{1}{q^2}\sideset{}{'}\sum_{r(qq_1q_1')} \mathbf{1}_{\substack{\{-\ell\equiv Pr(Ba+Dc)(\bmod q_1)\ ,\\ -\ell'\equiv Pr(Ba'+Dc')(\bmod q_1')\}}}$$

where we used that $\phi(q)/\phi(qq_1q_1') = 1/(q_1q_1')$, since every prime dividing q_1q_1' also divides q.

Since B, P are fixed throughout, consider the condition

$$-\ell \equiv Pr(Ba+Dc)(\bmod q_1), \implies -\ell \equiv rPBa(\bmod(q_1,c)).$$

Since $\det\gamma = 1$, we have that a is invertible mod c, so r is restricted to a bounded (in terms of B, P) number of residues mod q_1. Similarly, r is also restricted to a bounded number of residue classes mod q_1'. Therefore the total number of r mod qq_1q_1' satisfying the congruence is at most $qq_1q_1'/[q_1,q_1']$. This gives the claim. □

Next we record a Kloosterman-type estimate necessary in what follows.

Lemma 4.9 *Fix any $J, K, L \in \mathbb{Z}$ and let $q_0 \mid q$. Then*

$$\left|\sideset{}{'}\sum_{\substack{r(q)\\ r\equiv r_0(\bmod q_0)}} e_q\left(Jr+K\bar{r}+L\right)\right| \ll_\varepsilon \min\left(\frac{q}{q_0}, q^{3/4+\varepsilon}\frac{1}{q_0^{1/4}}\gcd(q/q_0,J,K)^{1/4}\right).$$

Proof If $q_0 = 1$, this is just Kloosterman's estimate, so we assume $q_0 > 1$. The first bound in the minimum is just the trivial bound, and is sometimes better than the second bound.

Following Kloosterman's method, we take the fourth moment, and consider

$$\mathcal{U} = \sum_{J',K' \bmod q}\left|\sideset{}{'}\sum_{\substack{r(q)\\ r\equiv r_0(\bmod q_0)}} e_q\left(J'r+K'\bar{r}+L\right)\right|^4. \tag{4.10}$$

We open the power and evaluate.

$$\mathcal{U} = \sum_{J',K' \bmod q}\ \sideset{}{'}\sum_{\substack{r_1,r_2,r_3,r_4(q)\\ r_j\equiv r_0(\bmod q_0)}} e_q\left(J'(r_1+r_2-r_3-r_4)+K'(\bar{r}_1+\bar{r}_2-\bar{r}_3-\bar{r}_4)\right) \tag{4.11}$$

The J', K' sum is a complete sum over all of $\mathbb{Z}/q$, which vanishes unless

$$r_1+r_2-r_3-r_4 \equiv 0(\bmod q), \quad \bar{r}_1+\bar{r}_2-\bar{r}_3-\bar{r}_4 \equiv 0(\bmod q),$$

in which case they contribute q each. So we have that

$$\mathcal{U} = q^2 \sum^{\prime}_{\substack{r_1,r_2,r_3,r_4(q)\\ r_j\equiv r_0(\mathrm{mod}\, q_0)}} \mathbf{1}_{\substack{\{r_1+r_2-r_3-r_4\equiv 0(\mathrm{mod}\, q)\\ \bar{r}_1+\bar{r}_2-\bar{r}_3-\bar{r}_4\equiv 0(\mathrm{mod}\, q)\}}}$$

We need to count the number of r_j contributing to the remaining sum. The count is multiplicative, so we may assume that q is a prime power. Set $\widetilde{q} := q/q_0$, and let $R \in \mathbb{Z}/\widetilde{q}$ be defined by: $r_1 - r_3 = Rq_0$; the first condition above is that we also have $r_4 - r_2 \equiv Rq_0$. For the second condition on the r_j, we multiply through by $r_1r_2r_3r_4$, getting the condition:

$$(r_2 - r_3)\, R\, (q_0R + r_2 + r_3) \equiv 0(\mathrm{mod}\, \widetilde{q}).$$

Recall that $q_0 > 1$, and notice that $q_0R + r_2 + r_3 \equiv 2r_0(\mathrm{mod}\, q_0)$ is then invertible $\mathrm{mod}\, q$ (except perhaps when $2 \mid q$, in which case an extra constant factor contributes to the estimate below). We now evaluate the count as follows. First sum over divisors $\mathfrak{q} \mid \widetilde{q}$, then over those R with $(R, \widetilde{q}) = \mathfrak{q}$. The above condition becomes

$$r_3 \equiv r_2(\mathrm{mod}\, \widetilde{q}/\mathfrak{q}).$$

Then r_3 has $\ll q\mathfrak{q}/\widetilde{q} = q_0\mathfrak{q}$ possible values. In total, we have:

$$\begin{aligned}
&\sum^{\prime}_{\substack{r_1,r_2,r_3,r_4(q)\\ r_j\equiv r_0(\mathrm{mod}\, q_0)}} \mathbf{1}_{\substack{\{r_1+r_2-r_3-r_4\equiv 0(\mathrm{mod}\, q)\\ \bar{r}_1+\bar{r}_2-\bar{r}_3-\bar{r}_4\equiv 0(\mathrm{mod}\, q)\}}}\\
&\ll \sum_{\mathfrak{q}\mid\widetilde{q}} \sum_{\substack{R(\mathrm{mod}\, \widetilde{q})\\ (R,\widetilde{q})=\mathfrak{q}}} \sum_{\substack{r_2(\mathrm{mod}\, q)\\ r_2\equiv r_0(\mathrm{mod}\, q_0)}} \sum_{\substack{r_3(\mathrm{mod}\, q)\\ r_3\equiv r_0(\mathrm{mod}\, q_0)}} \mathbf{1}_{\{r_3\equiv r_2(\mathrm{mod}\, \widetilde{q}/\mathfrak{q})\}}\\
&\ll \sum_{\mathfrak{q}\mid\widetilde{q}} \frac{\widetilde{q}}{\mathfrak{q}}\widetilde{q}q_0\mathfrak{q} \ll_\varepsilon \frac{q^{2+\varepsilon}}{q_0}.
\end{aligned}$$

In summary, we obtain the following estimate:

$$\mathcal{U} \ll_\varepsilon q^2\frac{q^{2+\varepsilon}}{q_0}.$$

Next we determine the multiplicity of the size of the original sum (that is, when $(J', K') = (J, K)$) contributing to $\mathcal{U}$. Any change of variables $r \mapsto rs$ with $s \in (\mathbb{Z}/q)^\times$ and $s \equiv 1(\mathrm{mod}\, q_0)$ corresponds to a change in the coefficients $\iota_s : (J, K) \mapsto (Js, K\bar{s})$. Another invariance comes from the map $\sigma_{u,v} : (J, K) \mapsto (J+u\widetilde{q}, K+v\widetilde{q})$, because

$$\sideset{}{'}\sum_{\substack{r(q)\\ r\equiv r_0(\mathrm{mod}\, q_0)}} e_q\left((J+u\widetilde{q})r+(K+v\widetilde{q})\bar{r}+L\right)$$
$$=e_{q_0}\left(ur_0+v\bar{r}_0\right)\sideset{}{'}\sum_{\substack{r(q)\\ r\equiv r_0(\mathrm{mod}\, q_0)}} e_q\left(Jr+K\bar{r}+L\right),$$

with both sides having the same magnitude.

Next we must determine the number of distinct (J', K') obtained by the above transformations which contribute the same magnitude to $\mathcal{U}$ as (J, K). Assume that $\gcd(J, q) \le \gcd(K, q)$. We use ι_s to produce as many values of J' as possible, and for each such, we use $\sigma_{0,v}$ to construct distinct K's.

The $s \in (\mathbb{Z}/q)^\times$ with $s \equiv 1(\mathrm{mod}\, q_0)$ which give distinct values of $Js(\mathrm{mod}\, q)$ are determined by solving

$$J \equiv Js(q).$$

The number of distinct values of $Js(\mathrm{mod}\, q)$ is then $q/\gcd(q, q_0J)$. For each such value of sJ, applying $\sigma_{0,v}$ produces a distinct pair (J', K') where v ranges in $\mathbb{Z}/q_0$.

In total, we have that:

$$\frac{q}{\gcd(q, q_0J, q_0K)}q_0\left|\sideset{}{'}\sum_{\substack{r(q)\\ r\equiv r_0(\mathrm{mod}\, q_0)}} e_q\left(Jr+K\bar{r}+L\right)\right|^4 \le \mathcal{U} \ll_\varepsilon q^2\frac{q^{2+\varepsilon}}{q_0},$$

from which the claim follows. □

Lemma 4.12 *With the same notation as Lemma 4.4, we have that*

$$\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}$$
$$\ll_\varepsilon q^{-5/4+\varepsilon}(q_1q_1')^{1/2}(q_1,q_1')^{1/4}\gcd\left(q(q_1,q_1'),J,K\right)^{1/4},$$

for any $\varepsilon > 0$.

Proof Applying (4.5), we decompose the sum on r mod qq_1q_1' into residue classes mod $q_0 := [q_1, q_1']$ to catch the indicator functions.

$$\begin{aligned}
&\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\\
&\quad= \frac{(BcP^2,q)}{q}\frac{(Bc'P^2,q)}{q}\frac{\phi(q)}{\phi(qq_1q_1')}\\
&\qquad \sideset{}{'}\sum_{r_0(q_0)} \mathbf{1}_{\substack{\{-\ell\equiv Pr_0(Ba+Dc)(\bmod q_1)\\ -k\equiv Pr_0(Ac+Bd)(\bmod q_1)\}}}\mathbf{1}_{\substack{\{-\ell'\equiv Pr_0(Ba'+Dc')(\bmod q_1')\\ -k'\equiv Pr_0(Ac'+Bd')(\bmod q_1')\}}}\\
&\qquad \sideset{}{'}\sum_{\substack{r(qq_1q_1')\\ r\equiv r_0(q_0)}} e_{qq_1q_1'}(rJ+\bar{r}K+L).
\end{aligned}\tag{4.13}$$

On the last summation, we apply Lemma 4.9.

$$\begin{aligned}
&\ll_\varepsilon \frac{(BcP^2,q)}{q}\frac{(Bc'P^2,q)}{q}\frac{\phi(q)}{\phi(qq_1q_1')}\\
&\qquad \sideset{}{'}\sum_{r_0(q_0)} \mathbf{1}_{\substack{\{-\ell\equiv Pr_0(Ba+Dc)(\bmod q_1)\\ -k\equiv Pr_0(Ac+Bd)(\bmod q_1)\}}}\mathbf{1}_{\substack{\{-\ell'\equiv Pr_0(Ba'+Dc')(\bmod q_1')\\ -k'\equiv Pr_0(Ac'+Bd')(\bmod q_1')\}}}\\
&\qquad (qq_1q_1')^{3/4+\varepsilon}\frac{1}{q_0^{1/4}}\gcd(qq_1q_1'/q_0,J,K)^{1/4}.
\end{aligned}$$

Finally, we estimate the number of r_0 contributing. Recall that $q_1=(BcP^2,q)$. Since B, P are fixed throughout, consider the condition

$$-\ell\equiv Pr_0(Ba+Dc)(\bmod q_1),\implies -\ell\equiv r_0PBa(\bmod (q_1,c)).$$

Since $\det\gamma=1$, we have that a is invertible mod c, so r_0 is restricted to a bounded (in terms of B, P) number of residues mod q_1. Similarly, r_0 is bounded mod q_1', and hence the sum on r_0 has a bounded number of contributions. The claim follows immediately. □

If $c\neq c'$, this analysis will suffice. But we need more work if $c=c'$, since then J in (4.6) will be 0 mod q_2. (Note here that in this case, $q_1=q_1',\ q_2=q_2'$, and $E=E'$.)

Lemma 4.14 *With notation as in Lemma 4.4 and assuming $c=c'$, we have that:*

$$\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \ll_\varepsilon q^{-5/4+\varepsilon}q_1^2\gcd\left(q,\ell'k'-\ell k\right)^{1/4},$$

Proof Applying Lemma 4.12 gives

$$\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \ll_\varepsilon q^{-5/4+\varepsilon}q_1^{3/2}\gcd\left(q,K\right)^{1/4},$$

where

$$K = q_1 \bar{E}(\ell' k' - \ell k),$$

which gives the claim on bounding $\gcd(q, \bar{E})$ by q_1, since E is invertible mod q_2. □

This suffices as long as $k\ell \neq k'\ell'$. In the final case that both $c = c'$ and $k\ell = k'\ell'$, we have

Lemma 4.15 *Assume that $c = c'$ and $k\ell = k'\ell'$. Then*

$$\sum_{r(q)}{}' \mathcal{S}_q(r, k, \ell; \gamma)\overline{\mathcal{S}_q(r, k', \ell'; \gamma')} \ll \left(\frac{(BcP^2, q)}{q}\right)^2 \sum_{r(q)}{}' \mathbf{1}_{\substack{\{-\ell \equiv Pr(Ba+Dc) (\mathrm{mod}\, q_1)\cdot \\ -k \equiv Pr(Ac+Bd) (\mathrm{mod}\, q_1)\}}}$$

Proof Inserting Lemma 4.2 and estimating the r sum trivially gives:

$$\begin{aligned}&\sum_{r(q)}{}' \left|\mathcal{S}_q(r, k, \ell; \gamma)\overline{\mathcal{S}_q(r, k', \ell'; \gamma')}\right| \\ &\quad \le \left(\frac{(BcP^2, q)}{q}\right)^2 \sum_{r(q)}{}' \mathbf{1}_{\substack{\{-\ell \equiv Pr(Ba+Dc) (\mathrm{mod}\, q_1) \\ -k \equiv Pr(Ac+Bd) (\mathrm{mod}\, q_1)\}}} \mathbf{1}_{\substack{\{-\ell' \equiv Pr(Ba'+Dc) (\mathrm{mod}\, q_1)\cdot \\ -k' \equiv Pr(Ac+Bd') (\mathrm{mod}\, q_1)\}}}\end{aligned}$$

Dropping the conditions on k', ℓ' gives the claim. □

Now we need an estimate where we average over q itself. To this end, we will first need the following result.

Lemma 4.16 *Given positive integers $R \ge S \ge W,\ U, V, X$, we have that*

$$\begin{aligned}\sum_{\substack{(q,U)=1, q \equiv V(W) \\ R \le q \le R+S}} \frac{\phi(q)}{q^2} e_q(\bar{U}X) \ll_\varepsilon\ & \frac{SU^\varepsilon (U, (U, X)W)}{RUW} + \frac{S^2}{R^2} S^\varepsilon \\ & + \frac{[U, W]\log R}{R} + \frac{XS}{R^2 UW},\end{aligned} \tag{4.17}$$

for every $\varepsilon > 0$.

Proof If $\gcd(V, W)$ is not coprime to U, then the sum is empty, whence (4.17) holds trivially. So we assume that $(V, W, U) = 1$. Observe that the trivial bound is $S/(RW)$, so this is what must be improved upon.

Since $(q, U) = 1$, there exist x, y with

$$qx + Uy = 1, \text{ or } \frac{x}{U} + \frac{y}{q} = \frac{1}{qU},$$

so that $\bar{U} \equiv y (\operatorname{mod} q)$. Then

$$e_q(\bar{U}X) = e\left(\frac{y}{q}X\right) = e\left(-\frac{x}{U}X\right) + O(X/(qU)) = e_U(-xX) + O(X/(qU)).$$

We have that the left hand side of (4.17) is:

$$LHS = \sum_{\substack{(q,U)=1, q\equiv V(W) \\ R\leq q\leq R+S}} \frac{\phi(q)}{q^2} e_U(-\bar{q}X) + O\left(\frac{XS}{R^2UW}\right). \tag{4.18}$$

Leaving the last error term aside, we break q into residue classes mod $U_1 := [U, W]$.

$$LHS_1 = \sum_{\substack{q_0 (\operatorname{mod} U_1) \\ (q_0,U)=1, q_0\equiv V(W)}} e_U(-\bar{q}_0 X) \left[\sum_{\substack{q\equiv q_0(U_1) \\ R\leq q\leq R+S}} \frac{\phi(q)}{q^2}\right]. \tag{4.19}$$

Our point will be that the bracketed sum is independent of q_0, to first order, and therefore we can get cancellation from the first q_0 sum. To this end, next use Möbius inversion in the form $\phi(n) = n \sum_{d|n} \mu(d)/d$,

$$\left[\,\cdot\,\right] = \sum_{\substack{q\equiv q_0(U_1) \\ R\leq q\leq R+S}} \frac{1}{q^2} q \sum_{d|q} \frac{\mu(d)}{d} = \sum_{d\leq R+S} \frac{\mu(d)}{d} \sum_{\substack{q\equiv q_0(U_1), q\equiv 0(d) \\ R\leq q\leq R+S}} \frac{1}{q}. \tag{4.20}$$

Introduce a parameter $0 < D < R + S$ and break the sum on d according to $d \leq D$ or not. We deal with the large d first.

$$\begin{aligned}\sum_{D<d\leq R+S} \frac{\mu(d)}{d} \sum_{\substack{qd\equiv q_0(U_1) \\ R\leq dq\leq R+S}} \frac{1}{dq} &\ll \frac{1}{R} \sum_{D<d\leq R+S} \frac{1}{d} \sum_{R\leq dq\leq R+S} 1 \\ &= \frac{1}{R} \sum_{D<d\leq R+S} \frac{1}{d}\left(\frac{S}{d}+1\right) \\ &\ll \frac{S}{RD} + \frac{\log R}{R},\end{aligned}$$

which saves either D or S over the trivial bound.

Next we handle small d's. Observe that since $(q_0, U) = 1$, we have that $(q_0, U_1) \mid W$. But we must also have $q_0 \equiv V(W)$, and thus $(q_0, U_1) = (V, W) =: V_1$, say. Since $q \equiv 0(d)$, we let $t := q/d$. Then

$$\sum_{d\leq D} \frac{\mu(d)}{d} \sum_{\substack{td\equiv q_0(U_1) \\ R\leq dt\leq R+S}} \frac{1}{dt} = \sum_{d\leq D} \frac{\mu(d)}{d^2} \sum_{\substack{td\equiv q_0(U_1) \\ \frac{R}{d}\leq t\leq \frac{R}{d}+\frac{S}{d}}} \frac{1}{t}.$$

The condition $dt \equiv q_0(U_1)$ admits a solution in t iff $d_1 := (d, U_1)$ divides $(q_0, U_1) = V_1$. Write $d = d_1 d_2$ with $(d_2, U_1/d_1) = 1$. For d satisfying this condition, the restriction on t becomes $t \equiv \bar{d}_2(q_0/d_1)(\bmod U_1/d_1)$, which of course is now uniquely determined modulo U_1/d_1. Thus

$$\begin{aligned}
\sum_{d\le D}\frac{\mu(d)}{d}\sum_{\substack{td\equiv q_0(U_1)\\ R\le dt\le R+S}}\frac{1}{dt} &= \sum_{\substack{d\le D, d=d_1d_2\\ d_1=(d,U_1), d_1|V_1}}\frac{\mu(d)}{d^2}\sum_{\substack{t\equiv\bar{d}_2(q_0/d_1)(\bmod U_1/d_1)\\ \frac{R}{d}\le t\le\frac{R}{d}+\frac{S}{d}}}\frac{1}{t}\\
&= \sum_{\substack{d\le D, d=d_1d_2\\ d_1=(d,U_1), d_1|V_1}}\frac{\mu(d)}{d^2}\sum_{\substack{t\equiv\bar{d}_2(q_0/d_1)(\bmod U_1/d_1)\\ \frac{R}{d}\le t\le\frac{R}{d}+\frac{S}{d}}}\frac{1}{R/d}\left(1+O\left(\frac{S}{R}\right)\right)\\
&= \frac{1}{R}\sum_{\substack{d\le D, d=d_1d_2\\ d_1=(d,U_1), d_1|V_1\\ (d_2,U_1/d_1)=1}}\frac{\mu(d)}{d}\left(\frac{Sd_1}{dU_1}+O(1)\right)\left(1+O\left(\frac{S}{R}\right)\right).
\end{aligned} \tag{4.21}$$

The conditions $(d_2, U_1/d_1) = 1$ and $(d_2, d_1) = 1$ (from Möbius) are together equivalent to $(d_2, U_1) = 1$. This allows to separate the d_1 and d_2 sums, and extend the d_2 sum to infinity. The "main" contribution becomes:

$$\begin{aligned}
&\frac{1}{R}\sum_{\substack{d\le D, d=d_1d_2\\ d_1=(d,U_1), d_1|V_1 (d_2,U_1/d_1)=1}}\frac{\mu(d)}{d}\frac{Sd_1}{dU_1}\\
&= \frac{S}{RU_1}\sum_{d_1|V_1}\frac{\mu(d_1)}{d_1}\sum_{\substack{d_2\le D/d_1\\ (d_2,U_1)=1}}\frac{\mu(d_2)}{d_2^2}\\
&= \frac{S}{RU_1}\sum_{d_1|V_1}\frac{\mu(d_1)}{d_1}\left[\sum_{\substack{d_2\le\infty\\ (d_2,U_1)=1}}\frac{\mu(d_2)}{d_2^2}+O(d_1/D)\right]\\
&= \frac{S}{RU_1}\sum_{d_1|V_1}\frac{\mu(d_1)}{d_1}\left[M_{U_1}+O(d_1/D)\right],
\end{aligned}$$

where

$$M_{U_1} := \prod_{p, p\nmid U_1}\left(1-\frac{1}{p^2}\right) \asymp 1, \tag{4.22}$$

is $1/\zeta(2)$ with the primes of U_1 removed.

Continuing the analysis gives

$$\begin{aligned}&= \frac{S}{RU_1}\sum_{d_1|V_1}\frac{\mu(d_1)}{d_1}M_{U_1} + O_\varepsilon\left(\frac{SW^\varepsilon}{DRU_1}\right)\\ &= \frac{S}{RU_1}M_{U_1}V_2 + O_\varepsilon\left(\frac{SW^\varepsilon}{DRU_1}\right),\end{aligned}$$

where

$$V_2 := \prod_{p|V_1}\left(1-\frac{1}{p}\right)$$

satisfies

$$V_1^{-\varepsilon} \ll_\varepsilon V_2 \le 1. \tag{4.23}$$

We return to handle the error terms of (4.21). The first is:

$$\frac{1}{R}\sum_{\substack{d\le D, d=d_1d_2\\ d_1=(d,U_1), d_1|V_1\\ (d_2,U_1/d_1)=1}}\frac{1}{d} \ll \frac{\log D}{R},$$

saving about S/W over the trivial bound. The second is:

$$\frac{1}{R}\sum_{\substack{d\le D, d=d_1d_2\\ d_1=(d,U_1), d_1|V_1\\ (d_2,U_1/d_1)=1}}\frac{1}{d}\frac{Sd_1}{dU_1}\frac{S}{R} \ll_\varepsilon \frac{S^2}{R^2U_1}D^\varepsilon,$$

which saves about R/S.

Putting everything together into (4.20) gives:

$$\begin{aligned}(4.20) &= \frac{S}{RU_1}M_{U_1}V_2 + O_\varepsilon\left(\frac{SW^\varepsilon}{DRU_1}+\frac{\log D}{R}+\frac{S^2}{R^2U_1}D^\varepsilon\right) + O\left(\frac{S}{RD}+\frac{\log R}{R}\right)\\ &= \frac{S}{RU_1}M_{U_1}V_2 + O_\varepsilon\left(\frac{S^2}{R^2U_1}D^\varepsilon + \frac{S}{RD} + \frac{\log R}{R}\right),\end{aligned}$$

which is, at least in the main term, independent of q_0, as desired. Inserting this into (4.19) now gives

$$
\begin{aligned}
LHS_1 &= \sum_{\substack{q_0 (\bmod U_1)\\ (q_0,U)=1, q_0 \equiv V(W)}} e_U(-\bar{q}_0 X)\left[\frac{S}{RU_1} M_{U_1} V_2 + O_\varepsilon\left(\frac{S^2}{R^2 U_1} D^\varepsilon + \frac{S}{RD} + \frac{\log R}{R}\right)\right]\\
&= \frac{S}{RU_1} M_{U_1} V_2 \left[\sum_{\substack{q_0 (\bmod U_1)\\ (q_0,U)=1, q_0 \equiv V(W)}} e_U(-\bar{q}_0 X)\right] + O_\varepsilon\left(\frac{S^2}{R^2} D^\varepsilon + \frac{SU_1}{RD} + \frac{U_1 \log R}{R}\right).
\end{aligned}
\tag{4.24}
$$

We analyze the bracketed summation by first decomposing it into residue classes mod U:

$$
\left[\,\cdot\,\right] = \sideset{}{'}\sum_{\substack{q \bmod U\\ q \equiv V(\bmod (W,U))}} e_U(-\bar{q} X) \sum_{\substack{q_0 (\bmod U_1)\\ q_0 \equiv q(U), q_0 \equiv V(W)}} 1.
$$

Using $U_1 = [U, W]$ (recall that this is the lcm) and the compatibility condition $q \equiv V(\bmod (W, U))$, the Chinese Remainder Theorem gives that the last summation has exactly one q_0 contributing. Therefore only the first summation remains. Let $X_1 := (U, X)$ and write $X = X_1 X_2$ and $U = X_1 U_2$ with $(X_2, U_2) = 1$. Then we break into residues mod U_2

$$
\left[\,\cdot\,\right] = \sideset{}{'}\sum_{\substack{q_2 (U_2)\\ q_2 \equiv V(\bmod (W,U_2))}} e_{U_2}(-\bar{q}_2 X_2) \sideset{}{'}\sum_{\substack{q \bmod U\\ q \equiv q_2(U_2), q \equiv V(\bmod (W,U))}} 1.
$$

In the last summation, we again get a unique contribution from the compatibility condition on q_2 which together with the Chinese Remainder Theorem determines q uniquely mod $[U_2, (W, U)]$. Therefore the second summation evaluates to: $\phi(U)/\phi([U_2, (W, U)])$. In summary, we have that

$$
\left[\,\cdot\,\right] = \frac{\phi(U)}{\phi([U_2, (W, U)])} \sideset{}{'}\sum_{\substack{q_2 (U_2)\\ q_2 \equiv V(\bmod (W,U_2))}} e_{U_2}(-\bar{q}_2 X_2).
$$

Looking locally, it is easy to see that the remaining summation either vanishes or is 1 in absolute value.

Returning to (4.24) and inserting the above argument gives that:

$$
LHS_1 \ll_\varepsilon \frac{S}{R[U, W]}\left[\frac{\phi(U)}{\phi([U/(U, X), (W, U)])}\right] + \frac{S^2}{R^2} D^\varepsilon + \frac{SU_1}{RD} + \frac{U_1 \log R}{R},
$$

where we used (4.22) and (4.23). We choose $D = S$ for simplicity. Returning all the way to the left hand side of (4.17), we have from (4.18) that:

$$LHS \ll_\varepsilon \frac{S}{R[U,W]}\left[\frac{\phi(U)}{\phi([U/(U,X),(W,U)])}\right] + \frac{S^2}{R^2}S^\varepsilon + \frac{[U,W]\log R}{R} + \frac{XS}{R^2UW}.$$

Finally recall the well-known fact that $m/\phi(m) \ll_\varepsilon m^\varepsilon$, so:

$$\phi([U/(U,X),(W,U)]) \gg_\varepsilon U^{-\varepsilon}([U/(U,X),(W,U)]).$$

The claim then follows. □

Now we can give the final estimate, as follows.

Lemma 4.25 *Assume that $c = c'$ and $k\ell = k'\ell'$. Then for parameters $Q \geq V \to \infty$, and any $\varepsilon > 0$, we have that:*

$$\left|\sum_{Q\leq q\leq Q+V}\sideset{}{'}\sum_{r(q)}\mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\right| \ll Q^\varepsilon \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\mathrm{mod}(q_1,c))\}}\mathcal{N}_{q_1}\sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^\infty,q_1)|Q_1\\ (E,q_1/Q_1)=1}} \times\left[\frac{V(EQ_1,Z)}{QEQ_1}+\frac{V^2c}{Q^2}+\frac{c^3}{Q}+\frac{V|Z|}{Q^2}\right],$$

where

$$\begin{aligned} Z &= Z(k,\ell,k',\ell',\gamma,\gamma') \\ &:= (Ac\ell+Bak+Bd\ell+Dck)-(Ac\ell'+Ba'k'+Bd'\ell'+Dck'), \end{aligned} \tag{4.26}$$

and

$$\begin{aligned} \mathcal{N}_m &= \mathcal{N}_m(k,\ell,k',\ell',\gamma,\gamma') \\ &:= \#\{r\in(\mathbb{Z}/m)^\times : {\scriptstyle -\ell\equiv Pr(Ba+Dc), -k\equiv Pr(Ac+Bd), \atop -\ell'\equiv Pr(Ba'+Dc), -k'\equiv Pr(Ac+Bd')}\}. \end{aligned} \tag{4.27}$$

Proof We apply Lemma 4.4, but with the special condition $q_1 = q_1'$ and $K = 0$:

$$\sum_{Q\leq q\leq Q+V}\sideset{}{'}\sum_{r(q)}\mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} = \sum_{Q\leq q\leq Q+V}\frac{(BcP^2,q)^2}{q^2}\frac{\phi(q)}{\phi(qq_1)} \sideset{}{'}\sum_{r(qq_1)}\mathbf{1}_{\{-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}\,q_1) \atop -k\equiv Pr(Ac+Bd)(\mathrm{mod}\,q_1)\}}\mathbf{1}_{\{-\ell'\equiv Pr(Ba'+Dc)(\mathrm{mod}\,q_1) \atop -k'\equiv Pr(Ac+Bd')(\mathrm{mod}\,q_1)\}}e_{qq_1}(rJ+L), \tag{4.28}$$

where under these conditions, we obtain:

$$\begin{aligned} J = &\, q_1(Aa + Bb + Dd) - q_1(Aa' + Bb' + Dd') \\ &-\bar{E}(q_1 EaA + aB^2dP^2 + q_1 EdD) \\ &+\bar{E}(q_1 Ea'A + a'B^2d'P^2 + q_1 Ed'D), \end{aligned}$$

and

$$L = -\bar{E}P(Ac\ell + Bak + Bd\ell + Dck) + \bar{E}P(Ac\ell' + Ba'k' + Bd'\ell' + Dck'). \tag{4.29}$$

The condition (4.6) here becomes:

$$J \equiv 0 (\mathrm{mod}\, q_2).$$

Returning to (4.29), we have that

$$L \equiv -\bar{E}PB(ak - a'k' + d\ell - d'\ell')(\mathrm{mod}(q_1, c)).$$

But the restrictions on r in (4.28) require that

$$PrB \equiv -d\ell \equiv -ak \equiv -d'\ell' \equiv -a'k'(\mathrm{mod}(q_1, c)),$$

where we used that $ad \equiv a'd' \equiv 1 (\mathrm{mod}\, c)$. Therefore

$$L \equiv 0 (\mathrm{mod}(q_1, c)). \tag{4.30}$$

The sum on r is multiplicative with respect to the modulus, but q_1 and q_2 are not necessarily coprime. To fix this, introduce a new parameter

$$Q_2 := \prod_{\substack{p^u \| q \\ p | q_2}} p^u,$$

so that $Q_2 \mid q$, and $q_2 \mid Q_2$. Then let Q_1 be defined by

$$q = Q_1 Q_2,$$

and it is easy to see that $(Q_1, Q_2) = 1$ and $Q_1 \mid q_1$. Moreover, if $p \mid Q_1$, then $(p^\infty, q_1) \mid Q_1$; that is, the largest prime power of any prime dividing Q_1 occurs in Q_1. It will be convenient to define

$$m = (q_1, Q_2).$$

Decompose q_1 further as

$$q_1 = (Q_1, q_1)(Q_2, q_1) = Q_1 m.$$

Note that $(E, q_2) = 1$, and $m \mid q_2$, so $(E, q_1/Q_1) = 1$. Then

$$q q_1 = Q_1^2 Q_2 m.$$

Let

$$Q_2' := Q_2 m \qquad \text{and} \qquad Q_1' = Q_1^2,$$

so that

$$q q_1 = Q_1' Q_2',$$

with $(Q_1', Q_2') = 1$. The lift $\bar{E}$ can be chosen so that $E\bar{E} \equiv 1 (\operatorname{mod} Q_2')$. Observe for later use that

$$(q_1, Q_1') = Q_1.$$

Then the same calculation leading to (4.6) gives

$$J \equiv 0 (\operatorname{mod} Q_2').$$

Now we split the r sum according to these moduli. Let s_1, s_2 be determined by:

$$\frac{s_1}{Q_1'} + \frac{s_2}{Q_2'} = \frac{1}{Q_1' Q_2'} = \frac{1}{q q_1},$$

that is, $s_1 Q_2' \equiv 1 (\operatorname{mod} Q_1')$, and $s_2 \equiv \overline{Q_1'} \operatorname{mod} Q_2'$, which implies that

$$s_2 \equiv \overline{Q_1}^2 \operatorname{mod} Q_2. \tag{4.31}$$

We will also need the basic fact that, if $a \mid b$, then $\phi(ab) = \phi(b) \cdot a$. Then we can write

$$\begin{aligned}
&\sum_{Q \le q \le Q+V} \sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r, k, \ell; \gamma) \overline{\mathcal{S}_q(r, k', \ell'; \gamma')} \\
&= \sum_{\substack{E | BcP^2 \\ q_1 = BcP^2/E}} \sum_{\substack{Q \le q \le Q+V \\ (q, BcP^2) = q_1,\ q q_1 = Q_1' Q_2'}} \mathbf{1}_{\{d\ell \equiv ak \equiv d'\ell' \equiv a'k' (\operatorname{mod}(q_1, c))\}} \frac{q_1^2}{q^2} \frac{\phi(q)}{\phi(q) \cdot q_1} \\
&\quad \times \left(\sideset{}{'}\sum_{r(Q_1')} \mathbf{1}_{\substack{\{-\ell \equiv Pr(Ba+Dc)(\operatorname{mod}(q_1, Q_1')) \\ -k \equiv Pr(Ac+Bd)(\operatorname{mod}(q_1, Q_1'))\}}} \mathbf{1}_{\substack{\{-\ell' \equiv Pr(Ba'+Dc)(\operatorname{mod}(q_1, Q_1')) \\ -k' \equiv Pr(Ac+Bd')(\operatorname{mod}(q_1, Q_1'))\}}} e_{Q_1'}(s_1(rJ + L)) \right)
\end{aligned}$$

$$\times \left(e_{Q_2'}(s_2L) \sum_{r(Q_2')}^{\prime} \mathbf{1}_{\substack{\{-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}(q_1,Q_2')) \\ -k\equiv Pr(Ac+Bd)(\mathrm{mod}(q_1,Q_2'))\}}} \mathbf{1}_{\substack{\{-\ell'\equiv Pr(Ba'+Dc)(\mathrm{mod}(q_1,Q_2')) \\ -k'\equiv Pr(Ac+Bd')(\mathrm{mod}(q_1,Q_2'))\}}} \right)$$

$$= \sum_{\substack{E\mid BcP^2 \\ q_1=BcP^2/E}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\mathrm{mod}(q_1,c))\}} \sum_{\substack{Q_1\mid q_1 \\ p\mid Q_1\Longrightarrow(p^\infty,q_1)\mid Q_1 \\ Q_1'=Q_1^2 \\ (E,q_1/Q_1)=1}} \frac{q_1}{Q_1^2}$$

$$\sum_{\substack{(Q_2,E)=(Q_2,Q_1)=1,Q_1Q_2\equiv 0(q_1) \\ Q\le Q_1Q_2\le Q+V \\ q=Q_1Q_2 \\ q_2=Q_1Q_2/q_1 \\ Q_2'=qq_1/Q_1'}} \frac{1}{Q_2^2}$$

$$\times \left(\sum_{r_1(Q_1')}^{\prime} \mathbf{1}_{\substack{\{-\ell\equiv Pr_1(Ba+Dc)(\mathrm{mod}\ Q_1) \\ -k\equiv Pr_1(Ac+Bd)(\mathrm{mod}\ Q_1)\}}} \mathbf{1}_{\substack{\{-\ell'\equiv Pr_1(Ba'+Dc)(\mathrm{mod}\ Q_1) \\ -k'\equiv Pr_1(Ac+Bd')(\mathrm{mod}\ Q_1)\}}} e_{Q_1'}(\overline{Q_2'}(r_1J+L)) \right)$$

$$\times \left(e_{Q_2'}(s_2L) \sum_{r(Q_2')}^{\prime} \mathbf{1}_{\substack{\{-\ell\equiv Pr(Ba+Dc)(\mathrm{mod}(q_1,Q_2')) \\ -k\equiv Pr(Ac+Bd)(\mathrm{mod}(q_1,Q_2'))\}}} \mathbf{1}_{\substack{\{-\ell'\equiv Pr(Ba'+Dc)(\mathrm{mod}(q_1,Q_2')) \\ -k'\equiv Pr(Ac+Bd')(\mathrm{mod}(q_1,Q_2'))\}}} \right). \tag{4.32}$$

Next we make the following two claims: (i) that the first sum on r_1 only depends only on the value of Q_2 modulo q_1; and (ii), that we can count the number of solutions in the r sum mod Q_2'. We first work on (ii). Observe that

$$m = (q_1, Q_2) = (q_1, Q_2') = \frac{q_1}{Q_1},$$

which is a divisor of Q_2'. That is, $(m, Q_1) = 1$. Recalling the definition (4.27) of $\mathcal{N}_m$, the sum on $r(Q_2')$ clearly contributes

$$\sum_{r(Q_2')}^{\prime} = \mathcal{N}_m \frac{\phi(Q_2')}{\phi(m)} = \phi(Q_2)\mathcal{N}_m \frac{(q_1, Q_2)}{\phi((q_1, Q_2))}.$$

Next we argue (i). Recall from the analogue of (4.7) in this setting that any r_1 occurring in the first summation satisfies:

$$r_1J + L \equiv 0(Q_1).$$

Therefore the r_1 summation in (4.32) is:

$$\left(Q_1 \sum_{r_1(Q_1)}^{\prime} \mathbf{1}_{\substack{\{-\ell\equiv Pr_1(Ba+Dc)(\mathrm{mod}\ Q_1) \\ -k\equiv Pr_1(Ac+Bd)(\mathrm{mod}\ Q_1)\}}} \mathbf{1}_{\substack{\{-\ell'\equiv Pr_1(Ba'+Dc)(\mathrm{mod}\ Q_1) \\ -k'\equiv Pr_1(Ac+Bd')(\mathrm{mod}\ Q_1)\}}} e_{Q_1}(\overline{Q_2(Q_2,q_1)}\frac{r_1J+L}{Q_1}) \right).$$

The term $\overline{Q}_2$ only depends on the residue class, Q_2^0, say, mod Q_1, so we break the sum according to these residue classes. Returning to the original expression, we have:

$$
\begin{aligned}
&\sum_{Q\leq q\leq Q+V}\ \sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \\
&\sum_{\substack{E|BcP^2\\ q_1=BcP^2/E}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\text{mod}(q_1,c))\}} \sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^\infty,q_1)|Q_1\\ Q_1'=Q_1^2\\ m=q_1/Q_1,\ (m,Q_1)=1\\ (E,q_1/Q_1)=1}} \frac{q_1}{Q_1^2}\mathcal{N}_m\frac{m}{\phi(m)} \\
&\times \sum_{Q_2^0(\text{mod } Q_1)'} \left(Q_1 \sideset{}{'}\sum_{r_1(Q_1)} \mathbf{1}_{\substack{\{-\ell\equiv Pr_1(Ba+Dc)(\text{mod } Q_1)\\ -k\equiv Pr_1(Ac+Bd)(\text{mod } Q_1)\}}} \mathbf{1}_{\substack{\{-\ell'\equiv Pr_1(Ba'+Dc)(\text{mod } Q_1)\\ -k'\equiv Pr_1(Ac+Bd')(\text{mod } Q_1)\}}} e_{Q_1}(\overline{Q_2^0}\overline{m}\frac{r_1J+L}{Q_1}) \right) \\
&\times \left[\sum_{\substack{(Q_2,EQ_1)=1,\, Q_2\equiv 0(m),\, Q_2\equiv Q_2^0(\text{mod } Q_1)\\ Q\leq Q_1Q_2\leq Q+V\\ q=Q_1Q_2\\ q_2=Q_1Q_2/q_1\\ Q_2'=qq_1/Q_1'}} \frac{\phi(Q_2)}{Q_2^2} e_{Q_2'}(s_2L) \right].
\end{aligned}
$$

Let

$$
Q_1'' := (Q_1, c),
$$

with $Q_1/Q_1'' \asymp 1$. (Recall here that P is a constant depending only on the group Γ, and implied constants may depend on Γ and the fixed parameters A, B, C, D.) From (4.30), we have that $L \equiv 0(\text{mod } Q_1'')$, since $Q_1'' \mid (q_1, c)$. Let

$$
m_1 := (m, c),
$$

so that $m/m_1 \asymp 1$. By the same argument, we also have that $L \equiv 0(\text{mod } m_1)$, and since $(m_1, Q_1'') = 1$, we have that $L \equiv 0(\text{mod } Q_1''m_1)$.

So using (4.31) the last sum can be written as

$$
\begin{aligned}
\left[\,\cdot\,\right] &= \sum_{\substack{(Q_2,EQ_1)=1,\, Q_2\equiv 0(m),\, Q_2\equiv Q_2^0(\text{mod } Q_1)\\ Q\leq Q_1Q_2\leq Q+V}} \frac{\phi(Q_2)}{Q_2^2} e_{Q_2}(s_2\frac{L}{m}) \\
&= \sum_{\substack{(Q_2,U)=1,\, Q_2\equiv 0(m),\, Q_2\equiv Q_2^0(\text{mod } Q_1)\\ Q\leq Q_1Q_2\leq Q+V}} \frac{\phi(Q_2)}{Q_2^2} e_{Q_2}(\bar{U}X),
\end{aligned}
$$

where $U = EQ_1(Q_1/Q_1'')(m/m_1) \asymp EQ_1$, and

$$
X = \frac{-P(Ac\ell + Bak + Bd\ell + Dck) + P(Ac\ell' + Ba'k' + Bd'\ell' + Dck')}{Q_1''m_1},
$$

by (4.29). Note that

$$X = \frac{PZ}{Q_1''m_1},$$

where Z is defined as in (4.26). It is at this point that we apply Lemma 4.16, with $R = Q/Q_1$, $S = V/Q_1$, and $W = [m, Q_1] = q_1$. Note that

$$(U, X) = K(EQ_1, X) = K\frac{1}{mQ_1''}(EQ_1mQ_1'', PZ) = K\frac{1}{q_1}(Eq_1Q_1, Z).$$

Here K is an absolute constant, not the same in each occurrence. Then

$$(U, (U, X)W) = K(EQ_1, EQ_1q_1, Z) \ll (EQ_1, Z).$$

Now applying Lemma 4.16 gives:

$$\left[\ \cdot\ \right] \ll_\varepsilon Q^\varepsilon \left(\frac{V(EQ_1, Z)}{QEQ_1q_1} + \frac{V^2}{Q^2} + \frac{Q_1c}{Q} + \frac{V|X|}{Q^2c}\right).$$

Returning to the original summation, we have:

$$\begin{aligned}
&\left|\sum_{Q\le q\le Q+V}\ \sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r, k, \ell; \gamma)\overline{\mathcal{S}_q(r, k', \ell'; \gamma')}\right| \\
&\ll_\varepsilon\ Q^\varepsilon \sum_{\substack{E|BcP^2 \\ q_1=BcP^2/E}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\mathrm{mod}(q_1,c))\}} \\
&\qquad \sum_{\substack{Q_1|q_1 \\ p|Q_1\Longrightarrow(p^\infty,q_1)|Q_1 \\ Q_1'=Q_1^2 \\ m=q_1/Q_1,\ (m,Q_1)=1,(E,q_1/Q_1)=1}} q_1\mathcal{N}_m\frac{m}{\phi(m)} \\
&\quad \times \left(\sideset{}{'}\sum_{r_1(Q_1)} \mathbf{1}_{\substack{\{-\ell\equiv Pr_1(Ba+Dc)(\mathrm{mod}\ Q_1) \\ -k\equiv Pr_1(Ac+Bd)(\mathrm{mod}\ Q_1)\}}}\mathbf{1}_{\substack{\{-\ell'\equiv Pr_1(Ba'+Dc)(\mathrm{mod}\ Q_1) \\ -k'\equiv Pr_1(Ac+Bd')(\mathrm{mod}\ Q_1)\}}}\right)\left[\frac{V(EQ_1, Z)}{QEQ_1q_1}\right. \\
&\quad \left. +\frac{V^2}{Q^2} + \frac{Q_1c}{Q} + \frac{V|X|}{Q^2c}\right].
\end{aligned}$$

Note that the $r_1\,(\mathrm{mod}\ Q_1)$ summation is exactly $\mathcal{N}_{Q_1}$, and since $(m,\ Q_1) = 1$, we have that $\mathcal{N}_m \cdot \mathcal{N}_{Q_1} = \mathcal{N}_{q_1}$. Now we have, crudely, that

$$\begin{aligned}&\left|\sum_{Q\le q\le Q+V}\ \sideset{}{'}\sum_{r(q)}\mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\right|\\ &\ll_\varepsilon Q^\varepsilon \sum_{\substack{E\mid BcP^2\\ q_1=BcP^2/E}}\mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\mathrm{mod}(q_1,c))\}}\sum_{\substack{Q_1\mid q_1\\ p\mid Q_1\Longrightarrow(p^\infty,q_1)\mid Q_1\\ (E,q_1/Q_1)=1}}\mathcal{N}_{q_1}\\ &\quad\times\left[\frac{V(EQ_1,Z)}{QEQ_1}+\frac{V^2c}{Q^2}+\frac{c^3}{Q}+\frac{V|Z|}{Q^2}\right],\end{aligned}$$

from which is the claim. □

Next we analyze the size of $\mathcal{N}_m$ in (4.27).

Lemma 4.33 *Let $m \mid BcP^2$. Then*

$$\mathcal{N}_m \ll 1. \tag{4.34}$$

Proof Our goal is to count the number of r in $(\mathbb{Z}/m)^\times$ satisfying $-\ell\equiv Pr(Ba+Dc)$, $-k\equiv Pr(Ac+Bd)$, and also $-\ell'\equiv Pr(Ba'+Dc)$, $-k'\equiv Pr(Ac+Bd')$. Let

$$m_1 := (m, c),$$

so that $m_1 \mid c$, and $m \ll m_1$ (since B and P are fixed). Reducing the moduli mod m_1 gives the equations:

$$-\bar{a}\ell\equiv rPB,\ -\bar{d}k\equiv rPB,\ -\bar{a}'\ell'\equiv rPB,\ -\bar{d}'k'\equiv rPB(\mathrm{mod}\, m_1).$$

Here we used that $(a,c)=1$ since $\gamma\in\mathrm{SL}_2(\mathbb{Z})$, etc. There are clearly a bounded number of solutions in r to the above, which gives the claim. □

And lastly, we analyze the number of elements in $\mathrm{SL}_2(\mathbb{Z})$ with a given value of c and satisfying a congruence in Z in (4.26).

Lemma 4.35 *Let $\gamma\in\mathrm{SL}_2(\mathbb{Z})\cap B_T$ be given with $\gamma_c=c$, and fix a divisor $Z_1\mid c$ with $c\ll T$. Also fix k,ℓ,k',ℓ' with $k\ell=k'\ell'$. Then the number of $\gamma'\in\mathrm{SL}_2(\mathbb{Z})$ with $\gamma'_c=c$ and $|a'|,|b'|,|d'|\asymp T$ and satisfying $Z\equiv 0(\mathrm{mod}\, Z_1)$ is bounded by*

$$\ll\frac{T}{Z_1}(Z_1,\ell d-ka), \tag{4.36}$$

as $T\to\infty$.

Proof In the variables a',d', we have the pair of equations: $a'd'\equiv 1(\mathrm{mod}\, Z_1)$ and

$$Z=(Ac\ell+Bak+Bd\ell+Dck)-(Ac\ell'+Ba'k'+Bd'\ell'+Dck')\equiv 0(Z_1),$$

or

$$Bk'a'^2 - (Bak + Bd\ell)a' + B\ell' \equiv 0(Z_1).$$

Let $B_1 := (B, Z_1)$ and set $Z_2 := Z_1/B_1$ and $B_2 := B/B_1$, with $(B_2, Z_2) = 1$. Dividing through by B_1, the equation reduces to

$$k'a'^2 - (ak + d\ell)a' + \ell' \equiv 0(Z_2).$$

Let

$$\tilde{Z} := (k', \ell', Z_2).$$

Since $k\ell = k'\ell'$, we have that $\tilde{Z}^2 \mid k\ell$. Working locally, suppose that $\tilde{Z} \mid \ell$. Then reducing mod $\tilde{Z}$ gives the equation

$$-(ak)a' \equiv 0(\tilde{Z}).$$

But $(a, \tilde{Z}) = (a, c) = 1$ and a' is also coprime to $\tilde{Z}$, which implies that $k \equiv 0(\tilde{Z})$. Therefore there are no solutions unless both $\tilde{Z} \mid k$ and $\tilde{Z} \mid \ell$. In this case, we can divide the whole equation by $\tilde{Z}$. Set $k_1 := k/\tilde{Z}, \ \ldots, \ \ell_1' := \ell'/\tilde{Z}$ and $Z_3 := Z_2/\tilde{Z}$. Then the equation becomes

$$k_1'a'^2 - (ak_1 + d\ell_1)a' + \ell_1' \equiv 0(Z_3),$$

with $(k_1', \ell_1', Z_3) = 1$. Working locally, we may assume that $(k_1', Z_3) = 1$. Now we can simply solve the equation. Assuming for simplicity that Z_3 is odd (with minor modifications otherwise), we have that

$$(a' - \bar{2}\bar{k}_1'(ak_1 + d\ell_1))^2 \equiv -\bar{k}_1'\ell_1' + \bar{4}\bar{k}_1'^2(ak_1 + d\ell_1)^2 \equiv (\bar{2}\bar{k}_1'(ak_1 - d\ell_1))^2 (\operatorname{mod} Z_3),$$

where we again used that $k\ell = k'\ell'$ and $ad \equiv 1(\operatorname{mod} Z_3)$. The equation is now a difference of squares, so

$$(a' - \bar{k}_1'\ell_1 d)(a' - \bar{k}_1'k_1 a) \equiv 0(\operatorname{mod} Z_3).$$

Again working locally, suppose that $Z_3 = p^U$. Then for $V + W = U$, we get a solution

$$a' \equiv \bar{k}_1'\ell_1 d(\operatorname{mod} p^V) \text{ and } a' \equiv \bar{k}_1'k_1 a(\operatorname{mod} p^W).$$

Assume WLOG that $V \leq W$. Then for there to be any solutions, it must be the case that $\ell_1 d - k_1 a \equiv 0(\operatorname{mod} p^V)$, and if this is the case, then there are $p^{U-W} = p^V$ solutions for a'. Let

$$Z_4 := (Z_3, \ell_1 d - k_1 a).$$

By the above discussion, the number of solutions for $a'(\text{mod } Z_3)$ is at most

$$\min(Z_4, \sqrt{Z_3}).$$

So once the value of a' mod Z_3 is fixed, the total number of such $a' \asymp T$ is

$$\ll \frac{T}{Z_3} \ll \frac{T\tilde{Z}}{Z_2} \ll \frac{T(k', \ell', Z_1)}{Z_1}.$$

Thus we can bound the total number of values of a' by

$$Z_4 T(k', \ell', Z_1)/Z_1 \ll T(Z_3, \ell_1 d - k_1 a)(k', \ell', Z_1)/Z_1 \ll T(Z_1, \ell d - ka)/Z_1.$$

With a' and c fixed, the number of d', b' is $\ll 1$, since they are all of order T and $a'd' - b'c = 1$. This completes the proof. □

5 Major arc analysis

Theorem 5.1 *There is an $\eta > 0$ and a set $\mathscr{E} \subset \mathbb{Z}$ of "exceptional" n, of zero density,*

$$\frac{1}{N}\#(\mathscr{E} \cap [-N, N]) = O(N^{-\eta}),$$

such that, for $n \notin \mathscr{E}$, $n \asymp N$, we have that

$$\mathcal{M}_N(n) \gg \mathfrak{S}(n)\frac{\widehat{\mathcal{R}_N(0)}}{N} + O\left(\frac{\widehat{\mathcal{R}_N(0)}}{N} N^{-\eta}\right),$$

as $N \to \infty$, where, for admissible $n \notin \mathscr{E}$ and for any $\varepsilon > 0$, the "singular series" $\mathfrak{S}(n)$ satisfies

$$\mathfrak{S}(n) \gg_{\varepsilon} |n|^{-\varepsilon}.$$

The implied constants are absolute.

Proof We begin with (2.17):

$$\begin{aligned}
\mathcal{M}_N(n) &= \int_0^1 \mathfrak{M}(\theta)\widehat{\mathcal{R}_N}(\theta)e(-n\theta)d\theta \\
&= \int_0^1 \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \sum_{m\in\mathbb{Z}} \psi((\beta+m)\tfrac{N}{K_0})\widehat{\mathcal{R}_N}(\tfrac{r}{q}+\beta)e(-n(\tfrac{r}{q}+\beta))d\beta \\
&= \int_{\mathbb{R}} \sum_{q<Q_0} \sideset{}{'}\sum_{r(q)} \widehat{\mathcal{R}_N}(\tfrac{r}{q}+\beta)e(-n(\tfrac{r}{q}+\beta))\psi(\beta\tfrac{N}{K_0})d\beta.
\end{aligned}$$

We have that

$$\mathcal{R}_N(\tfrac{r}{q}+\beta)$$
$$=\sum_{x,y\in\mathbb{Z}}\Upsilon\left(\frac{x}{X}\right)\Upsilon\left(\frac{y}{X}\right)\sum_{\gamma_0\in\Gamma(q)\backslash\Gamma}e_q(r\mathfrak{f}_{\gamma_0}(x,y))\left[\sum_{\substack{\gamma\in\mathscr{F}_T\\ \gamma\equiv\gamma_0(\mathrm{mod}\,q)}}e(\beta\mathfrak{f}_\gamma(x,y))\right].$$

For the bracketed term, we apply Lemma 3.4, together with $|\beta|<K_0/N<1/X^2$.

$$\mathcal{R}_N(\tfrac{r}{q}+\beta)=\sum_{x,y\in\mathbb{Z}}\Upsilon\left(\frac{x}{X}\right)\Upsilon\left(\frac{y}{X}\right)\sum_{\gamma_0\in\Gamma(q)\backslash\Gamma}e_q(r\mathfrak{f}_{\gamma_0}(x,y))$$
$$\left[\frac{1}{[\Gamma:\Gamma(q)]}\sum_{\gamma\in\mathscr{F}_T}e(\beta\mathfrak{f}_\gamma(x,y))+O(|\mathscr{F}_T|N^{-\Theta})\right].$$

Inserting this into $\mathcal{M}_N$ gives:

$$\mathcal{M}_N(n)=\sum_{x,y\in\mathbb{Z}}\Upsilon\left(\frac{x}{X}\right)\Upsilon\left(\frac{y}{X}\right)\left[\sum_{q<Q_0}\frac{1}{[\Gamma:\Gamma(q)]}\sum_{\gamma_0\in\Gamma(q)\backslash\Gamma}\sideset{}{'}\sum_{r(q)}e_q(r(\mathfrak{f}_{\gamma_0}(x,y)-n))\right]$$
$$\times\left[\sum_{\gamma\in\mathscr{F}_T}\int_{\mathbb{R}}\psi(\beta\tfrac{N}{K_0})e(\beta(\mathfrak{f}_\gamma(x,y)-n))d\beta\right]$$
$$+O\left(\frac{\widehat{\mathcal{R}_N}(0)}{N}N^{-\Theta}Q_0^5K_0\right),$$

where we have split into modular and archimedean components. The proof then follows on applying Theorem 3.38 and Theorem 3.41 to the modular component, and Lemma 3.5 to the archimedean part, together with the choice of parameters in (2.15). □

6 Minor arc analysis

Our goal is to estimate $\mathcal{E}_N$ in ℓ^2, or what is the same (by Parseval), to bound

$$\|\mathcal{E}_N\|^2 = \|\widehat{\mathcal{E}_N}\|^2 = \int_0^1|1-\mathfrak{M}(\theta)|^2|\widehat{\mathcal{R}_N}(\theta)|^2d\theta.$$

The main result of this section is the following.

Theorem 6.1 *There exists some $\eta>0$ so that, as $N\to\infty$,*

$$\|\mathcal{E}_N\|^2\ll\frac{\widehat{|\mathcal{R}_N(0)|^2}}{N}N^{-\eta}.$$

A standard argument concludes the main Theorem 1.17 (and hence Theorem 1.11) from Theorems 5.1 and 6.1. We begin the analysis as follows. In Dirichlet's approximation theorem, we choose the level

$$M = TX,$$

so that for every $\theta \in [0, 1]$, there is a $q < M$ and $(r, q) = 1$ so that $\theta = \frac{r}{q} + \beta$ with

$$\left|\theta - \frac{r}{q}\right| = |\beta| < \frac{1}{qM}.$$

Now we decompose the circle into dyadic regions of the form

$$W_Q = \left\{\theta = \frac{r}{q} + \beta \ \ q \asymp Q, (r, q) = 1, |\beta| \ll \frac{1}{QM}\right\},$$

so that

$$\|\mathcal{E}_N\|^2 \ll \sum_{\substack{Q<M \\ dyadic}} \int_{W_Q} |1 - \mathfrak{M}(\theta)|^2 |\widehat{\mathcal{R}_N}(\theta)|^2 d\theta.$$

This decomposes further into three ranges, according to whether Q satisfies: $Q < Q_0$, or $Q_0 < Q < X/Y$, or $X/Y < Q < XT = M$. Here we have set

$$Y := T^{(2\delta-1)/10} = N^y, \qquad y > 0, \tag{6.2}$$

to be a small power of N.

On the latter two ranges, the weight $|1 - \mathfrak{M}(\theta)|^2$ is exactly 1. To keep track, we define the integrals:

$$\begin{aligned} \mathcal{I}_{Q_0,K_0} &:= \int_{\substack{\theta=\frac{r}{q}+\beta \\ q<Q_0, |\beta|<K_0/N}} \left|\beta \frac{N}{K_0}\right|^2 |\widehat{\mathcal{R}_N}(\theta)|^2 d\theta, \\ \mathcal{I}_{Q_0} &:= \int_{\substack{\theta=\frac{r}{q}+\beta \\ q<Q_0, K_0/N\leq|\beta|<1/(qM)}} |\widehat{\mathcal{R}_N}(\theta)|^2 d\theta, \\ \mathcal{I}_{Q} &:= \int_{\substack{\theta=r/q+\beta \\ Q\leq q<2Q, (r,q)=1, |\beta|<1/(QM)}} \left|\widehat{\mathcal{R}_N}(\theta)\right|^2 d\theta. \end{aligned} \tag{6.3}$$

6.1 Preliminaries

We first estimate $\widehat{\mathcal{R}_N}(\theta)$ for $\theta = \frac{r}{q} + \beta$ as follows. We begin by decomposing x and y according to their residue classes mod q and applying Poisson summation in x and

y gives:

$$\begin{aligned}\mathcal{R}_N(\theta) &= \sum_{\gamma\in\mathscr{F}_T}\sum_{x,y\in\mathbb{Z}} \Upsilon\left(\frac{x}{X}\right)\Upsilon\left(\frac{y}{X}\right) e\left(\left(\frac{r}{q}+\beta\right)\mathfrak{f}_\gamma(x,y)\right)\\ &= X^2\sum_{\gamma\in\mathscr{F}_T}\sum_{k,\ell\in\mathbb{Z}} \mathcal{S}_q(r,k,\ell;\gamma)\mathcal{J}_X(\beta,k,\ell,q;\gamma), \end{aligned}\tag{6.4}$$

where $\mathcal{S}_q$ is given in (4.1) and

$$\mathcal{J}_X(\beta,k,\ell,q;\gamma) \;=\; \iint_{x,y\in\mathbb{R}} \Upsilon(x)\Upsilon(y)\, e(\beta\mathfrak{f}_\gamma(xX,yX) - \tfrac{X}{q}(kx+\ell y))dxdy.$$

Lemma 6.5 *Suppose that $q < X/Y$. Then for any $L<\infty$, we have that*

$$|\mathcal{J}_X(\beta,k,\ell,q;\gamma)| \;\ll_L\; X^{-L},\tag{6.6}$$

unless $|k|, |\ell| \ll 1$, in which case, we have:

$$|\mathcal{J}_X(\beta,k,\ell,q;\gamma)| \ll \min\left(1, \frac{1}{N|\beta|}\right).\tag{6.7}$$

Alternatively, if $X/Y \le q$, then we have the same arbitrary cancellation (6.6), unless $|k|, |\ell| \ll Y\frac{q}{X}$ (in which case, we only need the trivial bound).

Proof The phase of $\mathcal{J}_X$ can be written as $e(g)$ where

$$g(x,y) = \beta\mathfrak{f}_\gamma(xX,yX) - \frac{X}{q}(kx+\ell y).$$

Inputting (2.8), the partial derivatives of g are

$$\partial_x g(x,y) = \frac{X}{q}\left(\beta q B c_\gamma P^2 X y - k\right) + O(|\beta| T X)\tag{6.8}$$

and similarly

$$\partial_y g(x,y) = \frac{X}{q}\left(\beta q B c_\gamma P^2 X x - \ell\right) + O(|\beta| T X).\tag{6.9}$$

First consider the case that $q < X/Y$. Recalling that $|\beta| \ll 1/(QTX)$, $B, P, x, y \asymp 1$, and $c_\gamma \asymp T$, we have that $\nabla g \neq 0$ unless

$$|k|, |\ell| \ll |\beta| q T X \ll 1.$$

Outside of this range, not only does ∇g not vanish, but it is of order at least Y, since $q < X/Y$. Thus, we may apply non-stationary phase, giving (6.6).

Now suppose that the pair (k, ℓ) is such that ∇g does vanish at some point p in the support of $\Upsilon \times \Upsilon$. In this case, we apply stationary phase to show that

$$|\mathcal{J}_X| \ll \min(1, \Delta_p^{-1/2}),$$

where Δ_p is absolute determinant of the Hessian of g at p. Since

$$\Delta_p = |\det(\partial_{i,j}(g))| = (|\beta| B c_\gamma P^2 X^2)^2 \asymp (|\beta| T X^2)^2,$$

we have that

$$\Delta_p^{-1/2} \ll \frac{1}{TX^2|\beta|} = \frac{1}{N|\beta|},$$

which gives (6.7). (In fact, since the form is bilinear in the variables, it is possible to evaluate the integrals explicitly, though this is not needed here.)

In the case that $X/Y \leq q$, we can only apply non-stationary phase if the phase is actually growing, which is the case if $\max(|k|, |\ell|) > Y\frac{q}{X}$. This completes the proof. □

6.2 Minor arcs I: case $q < Q_0$

Proposition 6.10 *Assume that $q < X$. Then*

$$\left|\widehat{\mathcal{R}_N}\left(\frac{r}{q} + \beta\right)\right| \ll \frac{X^2|\mathscr{F}_T|}{N|\beta|}. \tag{6.11}$$

Proof Inserting Lemma 6.5 gives

$$\widehat{\mathcal{R}_N}\left(\frac{r}{q} + \beta\right) \ll X^2 \sum_{\gamma \in \mathscr{F}_T} \sum_{k,\ell \ll 1} |\mathcal{S}_q(r, k, \ell; \gamma)| \frac{1}{N|\beta|},$$

which gives the result on trivially estimating $|\mathcal{S}_q| \leq 1$. □

Corollary 6.12 *We have:*

$$\mathcal{I}_{Q_0,K_0} \ll \frac{|\widehat{\mathcal{R}_N}(0)|^2}{N} \frac{Q_0^2}{K_0},$$

and

$$\mathcal{I}_{Q_0} \ll \frac{|\widehat{\mathcal{R}_N}(0)|^2}{N} \frac{Q_0^2}{K_0}.$$

Proof Inserting the L^∞ bound (6.11) gives:

$$\mathcal{I}_{Q_0,K_0} \ll \int_{\substack{\theta=\frac{r}{q}+\beta \\ q<Q_0,|\beta|<K_0/N}} \left|\beta\frac{N}{K_0}\right|^2 \frac{X^4|\mathscr{F}_T|^2}{N^2|\beta|^2}d\theta \ll \frac{|\widehat{\mathcal{R}_N}(0)|^2}{K_0^2}Q_0^2\frac{K_0}{N},$$

giving the claim. Similarly,

$$\mathcal{I}_{Q_0} \ll \int_{\substack{\theta=\frac{r}{q}+\beta \\ q<Q_0,K_0/N<|\beta|}} \left|\frac{X^2|\mathscr{F}_T|}{N|\beta|}\right|^2 d\theta \ll \frac{|\widehat{\mathcal{R}_N}(0)|^2}{N^2}\frac{Q_0^2N}{K_0},$$

as claimed. □

The choice of parameters (2.14) ensures that these are power savings, as required in Theorem 6.1.

6.3 Minor arcs II: case $Q_0 \leq Q < X/Y$

Next take the intermediate range, where $Q_0 \leq Q < X/Y$. We need to estimate (6.3). Inserting (6.4) and opening the square gives

$$\begin{aligned}\mathcal{I}_Q &= \int_{\substack{\theta=r/q+\beta \\ Q\leq q<2Q,(r,q)=1,|\beta|<1/(QM)}} \left|X^2\sum_{\gamma\in\mathscr{F}_T}\sum_{k,\ell\in\mathbb{Z}}\mathcal{S}_q(r,k,\ell;\gamma)\mathcal{J}_X(\beta,k,\ell,q;\gamma)\right|^2 d\theta \\ &= X^4\sum_{\gamma,\gamma'\in\mathscr{F}_T}\sum_{k,\ell,k',\ell'\in\mathbb{Z}}\sum_{q\asymp Q}\left[\sum_{r(q)}{}^{'}\mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\right] \\ &\qquad \int_{|\beta|<1/(QM)}\mathcal{J}_X(\beta,k,\ell,q;\gamma)\overline{\mathcal{J}_X(\beta,k',\ell',q;\gamma')}d\beta.\end{aligned}$$

We apply Lemma 6.5 to get

$$\mathcal{I}_Q \ll \frac{X^4}{N}\sum_{\gamma,\gamma'\in\mathscr{F}_T}\sum_{k,\ell,k',\ell'\ll 1}\sum_{q\asymp Q}\left|\sum_{r(q)}{}^{'}\mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')}\right|. \tag{6.13}$$

Now we introduce a parameter

$$H := Q_0^{\eta_0/4}, \tag{6.14}$$

where η_0 is the constant in (3.8), and decompose

$$\mathcal{I}_Q \leq \mathcal{I}_Q^{(<)} + \mathcal{I}_Q^{(\geq)},$$

according to whether $\gcd(c,c') < H$ or $\gcd(c,c') \geq H$. We first deal with the large gcd.

Proposition 6.15 *There exists some* $\eta > 0$ *so that:*

$$\mathcal{I}_Q^{(\geq)} \ll \frac{|\widehat{\mathcal{R}_N}(0)|^2}{N} N^{-\eta}.$$

Proof Let $h \geq H$ be the gcd of c and c'. Then applying Corollary 4.8 (and notation therein) to (6.13), we have

$$\mathcal{I}_Q^{(\geq)} \ll \frac{X^4}{N} \sum_{\gamma \in \mathscr{F}_T} \sum_{\substack{h | c_\gamma \\ h \geq H}} \sum_{\substack{q_1 | BcP^2 \\ q_1 \ll Q}} \sum_{\substack{\gamma' \in \mathscr{F}_T \\ c' \equiv 0 (h)}} \sum_{\substack{q_1' | Bc'P^2 \\ q_1' \ll Q}} \sum_{\substack{q \asymp Q \\ q \equiv 0 ((q_1, q_1'))}} \frac{(q_1, q_1')}{q},$$

where in the last sum, we weakened the condition that q is divisible by both q_1 and q_1' to just being divisible by their gcd. Now estimating divisor sums and applying Nullstellensatz (3.8) in the γ' summation gives

$$\mathcal{I}_Q^{(\geq)} \ll_\varepsilon N^\varepsilon \frac{X^4}{N} H^{-\eta_0} |\mathscr{F}_T|^2,$$

from which the claim follows. □

Next we handle the small gcd.

Proposition 6.16 *There exists some* $\eta > 0$ *so that:*

$$\mathcal{I}_Q^{(<)} \ll \frac{|\widehat{\mathcal{R}_N}(0)|^2}{N} N^{-\eta}.$$

Proof We begin with the observation that $(c, c') < H$ implies that $c \neq c'$, since $c \asymp T$ and $H = o(T)$. We apply Lemma 4.12, estimate $\gcd\left(q(q_1, q_1'), J\right)$ by $\gcd\left((q_2, q_2'), J\right) \frac{q(q_1, q_1')}{(q_2, q_2')}$, and apply (4.6), giving

$$\mathcal{I}_Q^{(<)} \ll_\varepsilon \frac{X^4}{N} \sum_{\substack{\gamma, \gamma' \in \mathscr{F}_T \\ (c, c') < H}} \sum_{q \asymp Q} q^{-1+\varepsilon} \frac{1}{(q_2, q_2')^{1/4}} (q_1 q_1')^{1/2} (q_1, q_1')^{1/2}$$
$$\left(\gcd\left((q_2, q_2'), (c - c')(B^2 - \Delta c c')\right)\right)^{1/4}.$$

Since $q = q_1 q_2 = q_1' q_2'$, we have that $(q_2, q_2') = q / [q_1, q_1']$. We crudely estimate

$$\mathcal{I}_Q^{(<)} \ll_\varepsilon N^\varepsilon \frac{X^4}{N} \sum_{q \asymp Q} q^{-5/4} \sum_{\substack{\gamma, \gamma' \in \mathscr{F}_T \\ c \neq c' \\ q_1 = (BcP^2, q) \\ q_1' = (Bc'P^2, q)}} (q_1 q_1') \left(\gcd\left(q, BP^2(c - c')(B^2 - \Delta c c')\right)\right)^{1/4}. \tag{6.17}$$

Let $h = (q_1, q_1')$, then $h \mid BP^2(c, c')$, and $h \ll \min(Q, H)$. Then we can write $q_1 = hg, q_1' = hg'$, with $(g, g') = 1$. Also note that $h \mid (q, BP^2(c-c')(B^2 - \Delta cc'))$, so we can write

$$(q, BP^2(c - c')(B^2 - \Delta cc')) = h\tilde{g}.$$

Then it follows that

$$(\tilde{g}, g) \ll (BP^2(c - c')/h, g) \cdot (BP^2(B^2 - \Delta cc'), g).$$

The first gcd on the right hand side above is 1, since a factor of both should have been included in h. Since $g \mid BP^2\Delta cc'$, the second gcd is equal to $(BP^2B^2, g) \ll 1$. A similar argument shows that

$$(\tilde{g}, g') \ll 1,$$

and hence we have the bound

$$[hg, hg', h\tilde{g}] \gg gg'\tilde{g}$$

on their least common multiple. (Since h may be small, it will not help us in this estimate.)

Similarly, we note that

$$(\tilde{g}, c) \ll (BP^2(c - c'), c) \cdot (B^2 - \Delta cc', c).$$

The second gcd is again bounded, while the first is bounded by H, by the definition of $\mathcal{I}_Q^{(<)}$.

Thus we can estimate:

$$\begin{aligned}
\mathcal{I}_Q^{(<)} \ll_\varepsilon & N^\varepsilon \frac{X^4}{N} Q^{-5/4} \sum_{\gamma \in \mathscr{F}_T} \sum_{\substack{\gamma' \in \mathscr{F}_T \\ c \neq c', (c,c') < H}} \sum_{\substack{h \mid BP^2(c,c') \\ h \ll Q}} \sum_{\substack{g \mid BP^2 c \\ g \ll Q}} \sum_{\substack{g' \mid BP^2 c' \\ g' \ll Q}} \\
& \times \sum_{\substack{\tilde{g} \mid BP^2(c-c')(B^2-\Delta cc') \\ gg'\tilde{g} \ll Q, \ (\tilde{g},c) \ll H}} (hghg')\,(h\tilde{g})^{1/4} \sum_{\substack{q \asymp Q \\ q \equiv 0([hg,hg',h\tilde{g}])}} 1 \\
\ll_\varepsilon & N^\varepsilon \frac{X^4}{N} Q^{-5/4} H^{9/4} \sum_{\gamma \in \mathscr{F}_T} \sum_{\substack{\tilde{g} \ll Q \\ (\tilde{g},c) \ll H}} \frac{Q}{\tilde{g}^{3/4}} \sum_{\substack{\gamma' \in \mathscr{F}_T \\ BP^2(c-c')(B^2-\Delta cc') \equiv 0(\tilde{g})}} 1.
\end{aligned}$$

For fixed c, let $\mathcal{R} = \mathcal{R}(\tilde{g}) \subset \mathbb{Z}/\tilde{g}\mathbb{Z}$ denote the set of roots mod $\tilde{g}$ of the polynomial $BP^2(c - x)(B^2 - \Delta cx)$. If $\Delta = 0$, then this is a linear polynomial with leading coefficient BP^2, so $\#\mathcal{R} \ll 1$. If $\Delta \neq 0$, then this is a reducible quadratic polynomial with leading coefficient $BP^2\Delta c$; since $(\tilde{g}, c) \ll H$, it follows that $\#\mathcal{R} \ll_\varepsilon H^{1+\varepsilon}$. Finally, we have that

$$\mathcal{I}_Q^{(<)} \ll_\varepsilon N^\varepsilon \frac{X^4}{N} Q^{-5/4} H^{9/4} \sum_{\gamma \in \mathscr{F}_T} \sum_{\substack{\tilde{g} \ll Q \\ (\tilde{g},c) \ll H}} \frac{Q}{\tilde{g}^{3/4}} \sum_{\alpha \in \mathcal{R}(\tilde{g})} \sum_{\substack{\gamma' \in \mathscr{F}_T \\ c' \equiv \alpha(\tilde{g})}} 1.$$

We apply Nullstellensatz (3.8) in the last summation to obtain:

$$\begin{aligned} \mathcal{I}_Q^{(<)} &\ll_\varepsilon N^\varepsilon \frac{X^4}{N} Q^{-5/4} H^{9/4} \sum_{\gamma \in \mathscr{F}_T} \sum_{\tilde{g} \ll Q} \frac{Q}{\tilde{g}^{3/4}} H \frac{1}{\tilde{g}^{\eta_0}} |\mathscr{F}_T| \\ &\ll_\varepsilon N^\varepsilon \frac{X^4}{N} Q^{-1/4} H^{13/4} Q^{1/4-\eta_0} |\mathscr{F}_T|^2. \end{aligned}$$

The choice of the parameter H in (6.14) ensures that we have saved a power of $Q \geq Q_0$. The claim follows since Q_0 is a power of N (by (2.13)). □

These two propositions establish Theorem 6.1 in the intermediate range of Q.

6.4 Minor arcs III: case $X/Y \leq Q < M$

In this largest range, we return to the exact evaluation:

$$\mathcal{I}_Q = X^4 \sum_{\gamma,\gamma' \in \mathscr{F}_T} \sum_{k,\ell,k',\ell' \in \mathbb{Z}} \sum_{q \asymp Q} \left[\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma) \overline{\mathcal{S}_q(r,k',\ell';\gamma')} \right] \int_{|\beta|<1/(QM)} \mathcal{J}_X(\beta,k,\ell,q;\gamma) \overline{\mathcal{J}_X(\beta,k',\ell',q;\gamma')} d\beta.$$

Now we break

$$\mathcal{I}_Q \leq \mathcal{I}_Q^{=} + \mathcal{I}_Q^{\neq}$$

depending on whether $c = c'$ or not. We first handle the latter case.

Proposition 6.18 *There is an $\eta > 0$ so that*

$$\mathcal{I}_Q^{\neq} \ll \frac{|\widehat{\mathcal{R}}_N(0)|^2}{N} N^{-\eta},$$

as $N \to \infty$.

Proof To begin, we can use the last part of Lemma 6.5 to show that:

$$\mathcal{I}_Q^{\neq} \ll \frac{X^4}{QM} \sum_{\gamma,\gamma'\in\mathscr{F}_T} \sum_{|k|,|\ell|,|k'|,|\ell'|\ll\frac{Yq}{X}} \sum_{q\asymp Q} \left| \sum_{r(q)}{}' \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \right| + O(N^{-100}). \tag{6.19}$$

We will omit this last term henceforth. In the case $c \neq c'$, we estimate using Lemma 4.12, crudely (e.g., $q_1 \ll T$, etc) giving:

$$\begin{aligned}\mathcal{I}_Q^{\neq} &\ll \frac{X^4}{QM} \sum_{q\asymp Q} \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T \\ q_1=(BcP^2,q),q_1'=(Bc'P^2,q)}} \\ &\quad\times \sum_{|k|,|\ell|,|k'|,|\ell'|\ll\frac{Yq}{X}} q^{-5/4+\varepsilon}(q_1q_1')^{1/2}(q_1,q_1')^{1/4}\gcd\left(q(q_1,q_1'),J,K\right)^{1/4} \\ &\ll \frac{N^\varepsilon X^4 T^{3/2}}{Q^{9/4}M} \sum_{q\asymp Q} \sum_{\gamma,\gamma'\in\mathscr{F}_T} \sum_{|k|,|\ell|,|k'|,|\ell'|\ll\frac{Yq}{X}} \gcd(q,J,K)^{1/4}.\end{aligned}$$

Recall from Lemma 4.4 that J does not depend on k, k', ℓ, ℓ' but K does. Then

$$\mathcal{I}_Q^{\neq} \ll \frac{Y^4X^4T^{3/2}}{Q^{9/4}M}\left[\frac{Q^4}{X^4}+1\right] \sum_{q\asymp Q} \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T \\ c\neq c'}} \gcd(q,J)^{1/4}.$$

By (4.6), we have that:

$$\gcd(q,J) \ll T^2\gcd\left((q_2,q_2'),J\right) \ll T^2\gcd\Big((q_2,q_2'),(c-c')(B^2-\Delta cc')\Big) \ll T^5,$$

since $B^2 - \Delta cc'$ is never zero. Then

$$\begin{aligned}\mathcal{I}_Q^{\neq} &\ll \frac{N^\varepsilon X^4T^{3/2}}{Q^{9/4}M}\left[\frac{Q^4}{X^4}+1\right] Q|\mathscr{F}_T|^2T^{5/4} \\ &\ll Y^4\left[T^{3/2}M^{7/4}+\frac{X^4T^{3/2}}{X^{5/4}M}\right]|\mathscr{F}_T|^2T^{5/4} \\ &\ll Y^4|\mathscr{F}_T|^2X^2\frac{T^5}{X^{1/4}} \ll Y^4\frac{|\widehat{\mathcal{R}_N(0)}|^2}{N}\frac{T^6}{X^{1/4}}.\end{aligned}$$

The claim again follows due to the large power of X savings (relative to the small loss of powers of T and Y); see (2.9) and (6.2). □

Next we analyze the case that $c = c'$. At this stage, we decompose $\mathcal{I}_Q^{=}$ further according to whether $k\ell = k'\ell'$ or not,

$$\mathcal{I}_Q^{=} \ll \mathcal{I}_Q^{=,=} + \mathcal{I}_Q^{=,\neq}.$$

We first analyze the case that $k\ell \neq k'\ell'$.

Proposition 6.20 *There is an* $\eta > 0$ *so that*

$$\mathcal{I}_Q^{=,\neq} \ll \frac{|\widehat{\mathcal{R}}_N(0)|^2}{N} N^{-\eta},$$

as $N \to \infty$.

Proof We again apply Lemma 6.5 as in (6.19). In this case, we then apply Lemma 4.14, which gives:

$$\mathcal{I}_Q^{=,\neq} \ll \frac{X^4}{QM} \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'| \ll \frac{Yq}{X} \\ k\ell \neq k'\ell'}} \sum_{q \asymp Q} \left[q^{-5/4+\varepsilon} q_1^2 \gcd\left(q, \ell' k' - \ell k\right)^{1/4} \right].$$

The gcd is bounded by $|k\ell| \ll (Yq/X)^2$, giving:

$$\begin{aligned}
\mathcal{I}_Q^{=,\neq} &\ll \frac{Y^{1/2} X^4}{QM} \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'| \ll \frac{Yq}{X} \\ k\ell \neq k'\ell'}} \sum_{q \asymp Q} \left[q^{-5/4+\varepsilon} q_1^2 (q/X)^{1/2} \right] \\
&\ll Y^5 \frac{X^4}{QM} |\mathscr{F}_T|^2 \left[\frac{Q^4}{X^4} + 1 \right] Q Q^{-5/4} T^2 Q^{1/2} X^{-1/2} \\
&\ll Y^5 \frac{|\widehat{\mathcal{R}_N}(0)|^2}{N} \frac{T^6}{X^{1/4}}.
\end{aligned}$$

The claim again follows due to the large power of X savings. □

The last case is when $c = c'$ and $k\ell = k'\ell'$; here we will fight not for powers of Q but powers of the much smaller parameter T. We can save a factor of $T^{2\delta-1}$ from the fact that $c = c'$ (and hence there are only T values for γ', not $T^{2\delta} \asymp |\mathscr{F}_T|$). But this is insufficient for a power gain in the end. So new ideas are needed.

Proposition 6.21 *There is an* $\eta > 0$ *so that*

$$\mathcal{I}_Q^{=,=} \ll \frac{|\widehat{\mathcal{R}}_N(0)|^2}{N} N^{-\eta},$$

as $N \to \infty$.

Before beginning the proof, we return to the original formulation:

$$\mathcal{I}_Q^{=,=} = X^4 \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T\\ c=c'}} \sum_{\substack{k,\ell,k',\ell'\in\mathbb{Z}\\ k\ell=k'\ell'}} \sum_{q\asymp Q} \left[\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \right]$$
$$\int_{|\beta|<1/(QM)} \mathcal{J}_X(\beta,k,\ell,q;\gamma)\overline{\mathcal{J}_X(\beta,k',\ell',q;\gamma')}d\beta.$$

We apply the last part of Lemma 6.5 to truncate the k, ℓ, k', ℓ' range:

$$\mathcal{I}_Q^{=,=} = X^4 \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T\\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} \sum_{q\asymp Q} \left[\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \right]$$
$$\int_{|\beta|<1/(QM)} \mathcal{J}_X(\beta,k,\ell,q;\gamma)\overline{\mathcal{J}_X(\beta,k',\ell',q;\gamma')}d\beta.$$

Over this range of k, ℓ, k', ℓ', we need to level out the q dependence from the archimedean component $\mathcal{J}$. But the range of $q \asymp Q$ is too long for this purpose, so we decompose the sum into U intervals, where U is a parameter chosen to be

$$U = Q^{1/2}.$$

Each interval is of length $Q/U = Q^{1/2}$, which is much larger than T. Let Q_1 range over the starting points of these intervals. Then we have:

$$\mathcal{I}_Q^{=,=} = X^4 \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T\\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} \sum_{\substack{j\in\{1,\dots,U\}\\ Q_1=Q+j\frac{Q}{U}}}$$
$$\sum_{Q_1\le q\le Q_1+\frac{Q}{U}} \left[\sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r,k,\ell;\gamma)\overline{\mathcal{S}_q(r,k',\ell';\gamma')} \right]$$
$$\int_{|\beta|<1/(QM)} \mathcal{J}_X(\beta,k,\ell,q;\gamma)\overline{\mathcal{J}_X(\beta,k',\ell',q;\gamma')}d\beta.$$

On each of these sub-intervals, we will replace q in $\mathcal{J}$ by Q_1, thereby freeing the q variable for a purely modular analysis, as follows.

Lemma 6.22 *For any $\varepsilon > 0$, we have that:*

$$\mathcal{I}_Q^{=,=} \ll_\varepsilon \frac{1}{QM} X^4 \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'| \ll \frac{YQ}{X} \\ k\ell = k'\ell'}} \sum_{\substack{j \in \{1,\ldots,U\} \\ Q_1 = Q + j\frac{Q}{U}}} \left| \sum_{Q_1 \leq q \leq Q_1 + \frac{Q}{U}} \right.$$

$$\left. \sideset{}{'}\sum_{r(q)} \mathcal{S}_q(r, k, \ell; \gamma) \overline{\mathcal{S}_q(r, k', \ell'; \gamma')} \right|$$

$$+ Y^4 \frac{|\widehat{\mathcal{R}}_N(0)|^2}{N} \frac{T}{U}.$$

Proof Returning to the definition of $\mathcal{J}$, we see that

$$|\mathcal{J}_X(\beta, k, \ell, q; \gamma) - \mathcal{J}_X(\beta, k, \ell, Q_1; \gamma)| \ll \frac{Y}{U},$$

since $k, \ell \ll YQ/X$. We replace the two appearances of q in $\mathcal{J}$ one at a time.

Each time, we apply Lemma 4.15 to the resulting difference, which is bounded by

$$\ll Y^2 \frac{1}{U} \frac{X^4}{QM} \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'| \ll \frac{Yq}{X} \\ k\ell = k'\ell'}} \sum_{\substack{q \asymp Q \\ q_1 = (c,q)}} \left[\frac{(c,q)^2}{q^2} \sideset{}{'}\sum_{r(q)} \mathbf{1}_{\substack{\{-\ell \equiv Pr(Ba+Dc) (\mathrm{mod}\, q_1) \\ -k \equiv Pr(Ac+Bd) (\mathrm{mod}\, q_1)\}}} \right]$$

$$\ll Y^2 \frac{1}{U} \frac{X^4}{Q^3 M} \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{q_1 | c \\ q_1 \ll Q}} q_1^2 \sum_{\substack{q \asymp Q \\ q \equiv 0 (\mathrm{mod}\, q_1)}} \sum_{|k|,|\ell| \ll \frac{Yq}{X}} \sideset{}{'}\sum_{r(q)} \mathbf{1}_{\substack{\{-\ell \equiv PrBa (\mathrm{mod}\, q_1), \\ -k \equiv PrBd (\mathrm{mod}\, q_1)\}}}$$

where we used that $c \equiv 0(q_1)$. Since $(a, c) = (c, d) = 1$, for given ℓ, the value of r is determined up to constants mod q_1, so there are $\ll q/q_1$ values of r contributing. For each value of (ℓ, r), we have that k is determined mod q_1, but we won't use this. In total, we bound the difference by

$$\ll Y^2 \frac{1}{U} \frac{X^4}{Q^3 M} \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{q_1 | c \\ q_1 \ll Q}} q_1^2 \sum_{\substack{q \asymp Q \\ q \equiv 0 (\mathrm{mod}\, q_1)}} \frac{Q}{X} \frac{Q}{q_1} \frac{Q}{X}$$

$$\ll Y^2 \frac{1}{U} \frac{X^4}{Q^3 M} \sum_{\substack{\gamma,\gamma' \in \mathscr{F}_T \\ c=c'}} \sum_{\substack{q_1 | c \\ q_1 \ll Q}} q_1^2 \frac{Q}{q_1} \frac{Q}{X} \frac{Q}{q_1} \frac{Q}{X}$$

$$\ll Y^2 \frac{|\widehat{\mathcal{R}}_N(0)|^2}{N} \frac{T}{U}.$$

Finally, we estimate trivially that

$$\int_{|\beta|<1/(QM)} \mathcal{J}_X(\beta, k, \ell, Q_1; \gamma)\overline{\mathcal{J}_X(\beta, k', \ell', Q_1; \gamma')}d\beta \ll \frac{1}{QM},$$

whence the claim follows, since $Q \geq X/Y$. □

Now we have leveled out the sum, and are in position to apply the crucial Lemma 4.25.

Proof of Proposition 6.21 Inserting Lemma 4.25 into Lemma 6.22 gives:

$$\begin{aligned}\mathcal{I}_Q^{=,=} &\ll_\varepsilon \frac{1}{QM} X^4 \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T\\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} U\\ &\quad\times Q^\varepsilon \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k' (\text{mod}(q_1,c))\}}\mathcal{N}_{q_1} \sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^\infty,q_1)|Q_1\\ (E,q_1/Q_1)=1}}\\ &\quad\times\left[\frac{(EQ_1,Z)}{UEQ_1}+\frac{c}{U^2}+\frac{c^3}{Q}+\frac{|Z|}{UQ}\right]\\ &\quad+N^\varepsilon\frac{|\widehat{\mathcal{R}}_N(0)|^2}{N}\frac{T}{U},\end{aligned}\tag{6.23}$$

where Z and $\mathcal{N}_{q_1}$ are as defined in the statement of Lemma 4.25.

We first handle the contribution from the latter three terms in (6.23)

$$\begin{aligned}&\frac{Q^\varepsilon}{QM}X^4 \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T\\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} \left[\frac{c}{U}+\frac{c^3U}{Q}+\frac{|Z|}{Q}\right]\\ &\ll \frac{N^\varepsilon Y^2}{QM}X^4|\mathscr{F}_T|^2\left(\frac{Q}{X}\right)^2\left[\frac{T}{U}+\frac{T^3U}{Q}+\frac{T^2}{Q}\right]\\ &\ll N^\varepsilon Y^2\frac{|\mathcal{R}_N(0)|^2}{N}\left[\frac{T^2}{U}+\frac{T^4U}{X}+\frac{T^3}{X}\right],\end{aligned}$$

where we bounded $Z \ll TYQ/X \ll YT^2$ from (4.26) and used (4.34). With $U = Q^{1/2}$, this is sufficient savings if T is small enough relative to $Q \geq X/Y$; these are all power savings. Only the first term of (6.23) remains to be handled.

$$\frac{Q^{\varepsilon}}{QM}X^4 \sum_{\substack{\gamma,\gamma'\in\mathscr{F}_T\\ c=c'}} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\operatorname{mod}(q_1,c))\}}$$
$$\sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^{\infty},q_1)|Q_1\\ (E,q_1/Q_1)=1}} \frac{(EQ_1,Z)}{EQ_1}$$
$$\ll \frac{Q^{\varepsilon}}{QM}X^4 \sum_{\gamma\in\mathscr{F}_T} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\operatorname{mod}(q_1,c))\}}$$
$$\sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^{\infty},q_1)|Q_1\\ (E,q_1/Q_1)=1}} \frac{1}{EQ_1} \sum_{Z_1|EQ_1} Z_1 \sum_{\substack{\gamma'\in\mathscr{F}_T\\ c=c'}} \mathbf{1}_{\{Z(\gamma')\equiv 0(\operatorname{mod} Z_1)\}}.$$

Now we apply (4.36) to the last summation, expanding $\gamma'\in\mathscr{F}_T$ to all of $\mathrm{SL}_2(\mathbb{Z})$ (recalling that in $\mathscr{F}_T$, all entries are $\asymp T$). This gives

$$\frac{Q^{\varepsilon}}{QM}X^4 \sum_{\gamma\in\mathscr{F}_T} \sum_{\substack{|k|,|\ell|,|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \mathbf{1}_{\{d\ell\equiv ak\equiv d'\ell'\equiv a'k'(\operatorname{mod}(q_1,c))\}}$$
$$\sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^{\infty},q_1)|Q_1\\ (E,q_1/Q_1)=1}} \frac{1}{EQ_1} \sum_{Z_1|EQ_1} T(Z_1,\ell d-ka)$$
$$\ll \frac{Q^{\varepsilon}}{QM}TX^4 \sum_{\gamma\in\mathscr{F}_T} \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^{\infty},q_1)|Q_1\\ (E,q_1/Q_1)=1}} \frac{1}{EQ_1} \sum_{Z_1|EQ_1}\sum_{Z_2|Z_1} Z_2$$
$$\sum_{\substack{|k|,|\ell|\ll\frac{YQ}{X}\\ d\ell\equiv ak(\operatorname{mod}(q_1,c))\\ d\ell\equiv ak(\operatorname{mod} Z_2)}} \sum_{\substack{|k'|,|\ell'|\ll\frac{YQ}{X}\\ k\ell=k'\ell'}} 1.$$

The k', ℓ' sum is a divisor sum. Note that $(q_1,c)\asymp q_1$. Replace the condition $d\ell\equiv ak \operatorname{mod}(q_1,c)$ by $d\ell\equiv ak \operatorname{mod}(q_1/Q_1,c)$. Note that q_1/Q_1 is coprime to EQ_1. So with k fixed, ℓ is restricted to a residue class mod $(q_1/Q_1,c)Z_2$. This gives

$$\frac{Y^4}{QM}TX^4 \sum_{\gamma\in\mathscr{F}_T} \sum_{\substack{q_1|BcP^2\\ E=BcP^2/q_1}} \sum_{\substack{Q_1|q_1\\ p|Q_1\Longrightarrow(p^{\infty},q_1)|Q_1\\ (E,q_1/Q_1)=1}} \frac{1}{EQ_1} \sum_{Z_2|EQ_1} Z_2$$
$$\frac{Q}{X}\left(\frac{Q}{X(q_1/Q_1,c)Z_2}+1\right) \ll Y^2X^2|\mathscr{F}_T| \ll Y^2\frac{|\widehat{\mathcal{R}_N}(0)|^2}{N}\frac{1}{T^{2\delta-1}}.$$

This gives the claim, by the choice of Y in (6.2). □

Theorem 6.1 has now been established in all ranges of Q, thus completing the proof Theorem 1.11.

Acknowledgements We thank Jens Marklof, Curt McMullen, Lillian Pierce, Alan Reid, Zeev Rudnick, and Peter Sarnak for many discussions and insightful comments which greatly improved this work. We are grateful to the referees for comments and corrections.

Data Availability There is no data associated with this article.

References

1. Beardon, A.F.: The Hausdorff dimension of singular sets of properly discontinuous groups in N-dimensional space. Bull. Am. Math. Soc. **71**, 610–615 (1965)
2. Bourgain, J., Gamburd, A., Sarnak, P.: Generalization of Selberg's 3/16th theorem and affine sieve. Acta Math. **207**, 255–290 (2011)
3. Bourgain, J., Kontorovich, A.: On representations of integers in thin subgroups of SL(2, **Z**). GAFA **20**(5), 1144–1174 (2010)
4. Bourgain, J., Kontorovich, A.: On Zaremba's conjecture. Ann. Math. **180**(1), 137–196 (2014)
5. Bourgain, J., Kontorovich, A.: On the local-global conjecture for integral Apollonian gaskets. Invent. Math. **196**(3), 589–650 (2014)
6. Bourgain, J., Kontorovich, A.: The Affine Sieve Beyond Expansion I: Thin Hypotenuses. Int. Math. Res. Not. IMRN **19**, 9175–9205 (2015)
7. Bourgain, J., Kontorovich, A.: Beyond expansion IV: Traces of thin semigroups. Discrete Anal. Paper No. 6, 27 (2018)
8. Bourgain, J., Kontorovich, A., Sarnak, P.: Sector estimates for hyperbolic isometries. GAFA **20**(5), 1175–1200 (2010)
9. Burgess, D.A.: On character sums and L-series. Proc. Lond. Math. Soc. **3**(12), 193–206 (1962)
10. Bourgain, J., Varjú, P.P.: Expansion in $SL_d(Z/qZ)$, q arbitrary. Invent. Math. **188**(1), 151–173 (2012)
11. Chowla, S., Cowles, J., Cowles, M.: On the number of conjugacy classes in SL(2, Z). J. Number Theory **12**(3), 372–377 (1980)
12. Fuchs, E., Stange, K.E., Zhang, X.: Local-global principles in circle packings. Compos. Math. **155**(6), 1118–1170 (2019)
13. Goldfeld, D.M.: A simple proof of Siegel's theorem. Proc. Nat. Acad. Sci. U.S.A. **71**, 1055 (1974)
14. Iwaniec, H., Kowalski, E.: Analytic Number Theory. American Mathematical Society Colloquium Publications, vol. 53. American Mathematical Society, Providence, RI (2004)
15. Kontorovich, A.: From Apollonius to Zaremba: local-global phenomena in thin orbits. Bull. Am. Math. Soc. (N.S.) **50**(2), 187–228 (2013)
16. Kontorovich, A.: Levels of distribution and the affine sieve. Ann. Fac. Sci. Toulouse Math. (6) **23**(5), 933–966 (2014)
17. Kontorovich, A.: Applications of Thin Orbits. In: Dynamics and analytic number theory, volume 437 of *London Math. Soc. Lecture Note Ser.*, pp. 289–317. Cambridge Univ. Press, Cambridge (2016)
18. Lax, P.D., Phillips, R.S.: The asymptotic distribution of lattice points in Euclidean and non-Euclidean space. J. Funct. Anal. **46**, 280–350 (1982)
19. Marklof, J.: On multiplicities in length spectra of arithmetic hyperbolic three-orbifolds. Nonlinearity **9**(2), 517–536 (1996)
20. McMullen, C.: Dynamics of units and packing constants of ideals (2012). Online lecture notes, http://www.math.harvard.edu/~ctm/expositions/home/text/papers/cf/slides/slides.pdf
21. McMullen, C.T.: Billiards, heights and the arithmetic of non-arithmetic groups (2020). Preprint
22. Masser, D.W., Wüstholz, G.: Fields of large transcendence degree generated by values of elliptic functions. Invent. Math. **72**(3), 407–464 (1983)
23. Ogrodnik, B.: On the local-global conjecture for commutator traces (2021). (Rutgers University PhD Thesis)
24. Brooke Logan Ogrodnik: On the local-global conjecture for commutator traces. J. Number Theory **239**, 365–401 (2022)

25. Patterson, S.J.: The limit set of a Fuchsian group. Acta Math. **136**, 241–273 (1976)
26. Sullivan, D.: Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. Acta Math. **153**(3–4), 259–277 (1984)
27. Weisfeiler, B.: Strong approximation for Zariski-dense subgroups of semisimple algebraic groups. Ann. Math. (2) **120**(2), 271–315 (1984)
28. Zhang, X.: On the local-global principle for integral Apollonian 3-circle packings. J. Reine Angew. Math. (2015). https://doi.org/10.1515/crelle-2015-0042
29. Zhang, X.: On representation of integers from thin subgroups of $SL(2, \mathbb{Z})$ with parabolics. Int. Math. Res. Not. IMRN **18**, 5611–5629 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.