INTEGRAL POINTS ON QUADRICS IN THREE VARIABLES WHOSE COORDINATES HAVE FEW PRIME FACTORS

Jianya Liu[1] and Peter Sarnak[2]

Jianya Liu	Peter Sarnak
Department of Mathematics	Department of Mathematics
Shandong University	Princeton University & Institute for Advanced Study
Jinan, Shandong 250100	Princeton, NJ 08544-1000
China	USA
jyliu@sdu.edu.cn	sarnak@math.princeton.edu

Date: September 1, 2007.

[1]Supported by NSFC Grant #10531060, and the Ministry of Education Grant #305009.

[2]Supported by Oscar Veblen Fund (IAS) and an NSF Grant.

1. Introduction

Let $f(x_1, x_2, x_3)$ be an indefinite integral quadratic form with determinant $d(f)$, t a non-zero integer, and assume that $td(f)$ is square-free. We denote by $V = V_{f,t}$ the affine quadric

$$\{\mathbf{x} : f(\mathbf{x}) = t\},$$

where here and throughout bold face letters denote vectors whose dimension is clear from the context. According to Siegel's mass formula for f [26] and the fact that such an f only has one class in its genus ([5], page 203),

$$V(\mathbb{Z}) = \{\mathbf{x} \in \mathbb{Z}^3 : f(\mathbf{x}) = t\} \neq \emptyset$$

if and only if $f(\mathbf{x}) = t (\mathrm{mod} d)$ is solvable for every $d \geq 1$. We assume that this is the case, that is $V(\mathbb{Z}) \neq \emptyset$. Our main result is that the set of $\mathbf{x} \in V(\mathbb{Z})$ for which the product $x_1x_2x_3$ has at most 26 prime factors, is Zariski dense in $V_{f,t}$.

The two cases, f isotropic or not over $\mathbb{Q}$, are best treated separately. The former can be investigated in more elementary terms by exploiting that there are nonconstant polynomial parameterizations of points in $V(\mathbb{Z})$. That is, there are morphisms $P = (P_1, P_2, P_3)$ with P_j an integral polynomial in one variable for which $P(y) \in V(\mathbb{Z})$ if $y \in \mathbb{Z}$. With this one can produce points with few prime factors on such curves in V using standard sieve methods [14]. For example for the cone

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 - x_3^2 = 0, \tag{1.1}$$

Diamond and Halberstam [7] show that there are infinitely many such primitive Pythagorean triples for which $x_1x_2x_3$ has at most 17 prime factors. Using similar arguments, one can do the same for $V_{f,t}$ when f is isotropic.

If f is anisotropic, which is the case that we deal with in this paper, there are no nonconstant integral polynomial morphisms of the line into V, as can be seen by considering the highest degree coefficient of a potential such morphism. Instead, we use the affine linear sieve introduced in [3]. The general theorem in [3] when applied to this setting asserts that the saturation number $r(V(\mathbb{Z}), x_1x_2x_3)$ is finite, where $r(V(\mathbb{Z}), x_1x_2x_3)$ denotes the least number r such that the set of $\mathbf{x} \in V(\mathbb{Z})$, for which $x_1x_2x_3$ is a product of at most r primes, is Zariski dense in V. Our purpose here is to exploit various special features of 3-variable quadrics to give a reasonable upper bound for $r(V(\mathbb{Z}), x_1x_2x_3)$. In contrast to [3], we use an archimedean height to order the points, and we sieve on the orbit on the homogeneous space V rather than on the corresponding spin group. Moreover, the groups that intervene in the analysis are congruence subgroups of the spin group, and we can exploit the sharpest known bound towards Selberg's Eigenvalue Conjecture [19]. For

the abstract 3-dimensional sieve that underlies the analysis, we use the weighted sieve in [8] and [7].

Blomer and Brüdern [1] have investigated this problem in the definite case

$$x_1^2 + x_2^2 + x_3^2 = t.$$

For t large and satisfying suitable congruence conditions, they show that on such a quadric, there is an $\mathbf{x}$ with each x_j having at most 521 prime factors. They employ the vector sieve [4] via theta functions and the theory of holomorphic forms of 3/2-integral weight. In particular, they need nontrivial bounds towards the Ramanujan Conjectures for these forms. Such bounds are known and are critical in many applications [17] [9]. However their quality is poor compared to what is known towards the Selberg Eigenvalue Conjecture. This together with Propositions 3.1 and 3.2, which give sharp and uniform bounds for the remainders in a weighted count of cosets associated with one and two sheeted hyperboloids, lead to the central level of distribution Theorem 2.1. In turn this is responsible for the quality of our results in the indefinite case.

To end the introduction, we clarify a point about counting the number of prime factors r. For $d \geq 1$, let

$$V(\mathbb{Z}/d\mathbb{Z}) = \{\mathbf{x} \in (\mathbb{Z}/d\mathbb{Z})^3 : f(\mathbf{x}) \equiv t(\mathrm{mod} d)\}, \tag{1.2}$$

and

$$V^0(\mathbb{Z}/d\mathbb{Z}) = \{\mathbf{x} \in V(\mathbb{Z}/d\mathbb{Z}) : x_1x_2x_3 \equiv 0(\mathrm{mod} d)\}. \tag{1.3}$$

There may be "bad or ramified primes" p for which

$$V^0(\mathbb{Z}/p\mathbb{Z}) = V(\mathbb{Z}/p\mathbb{Z}). \tag{1.4}$$

We will see that this can only happen if p is in $\{2, 3, 5\}$. If it does, then $p|x_1x_2x_3$ for any $\mathbf{x} \in V^0(\mathbb{Z})$, and hence at least one of the x_j's is divisible by p. In our count of 26 primes factors, we do not include a prime $p \in \{2, 3, 5, 7\}$ for which our chosen orbit $\mathscr{O}$ (see Section 2) in $V(\mathbb{Z})$ is ramified.

The general local to global conjectures [3] when applied to $V(\mathbb{Z})$ after pulling back to the spin group, assert that the saturation number $r(V(\mathbb{Z}), x_1x_2x_3)$ is 3 if and only if there are no bad primes for V. In particular, in this case of no bad primes we conjecture that there are infinitely many $\mathbf{x} \in V(\mathbb{Z})$ whose coordinates are simultaneously prime.

Finally, if we assume the Selberg Eigenvalue Conjecture [24], we can establish our main result with 22 replacing 26.

In the forthcoming paper of Nevo-Sarnak [22], the affine linear sieve is developed further to give effective saturation numbers for orbits of congruence subgroups of semi-simple $\mathbb{Q}$-groups.

2. Statement of theorems

In order to formulate the Theorems, we introduce the central tools. Let SO_f denote the special orthogonal group of 3×3 matrices which preserve f. It is a linear algebraic group defined over $\mathbb{Q}$. Denote by G the spin double cover of SO_f (see [5]). Then G is defined over $\mathbb{Q}$ and, since f is anisotropic, G consists of the elements of norm 1 in a quaternion division algebra D_f over $\mathbb{Q}$. For example if f is diagonal,

$$f(\mathbf{x}) = f_1x_1^2 + f_2x_2^2 + f_3x_3^2 \quad \text{with } f_j \in \mathbb{Z},$$

and the quaternion $\mathbf{u}$ is of the form

$$\mathbf{u} = u_0 + u_1E_1 + u_2E_2 + u_3E_3 \quad \text{with } E_j^2 = -\frac{f_1f_2f_3}{f_j},$$

then the morphism $\rho : G \to SO_f$ is given by

$$\rho(\mathbf{u}) = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix}$$

with

$$\begin{aligned}
g_{11} &= u_0^2 + f_2f_3u_1^2 - f_1f_3u_2^2 - f_1f_2u_3^2, \\
g_{12} &= 2f_1f_3u_1u_2 - 2f_1u_0u_3, \\
g_{13} &= 2f_1u_0u_2 + 2f_1f_2u_1u_3, \\
g_{21} &= 2f_2u_0u_3 + 2f_2f_3u_1u_2, \\
g_{22} &= u_0^2 - f_2f_3u_1^2 + f_1f_3u_2^2 - f_1f_2u_3^2, \\
g_{23} &= 2f_1f_2u_2u_3 - 2f_2u_0u_1, \\
g_{31} &= 2f_2f_3u_1u_3 - 2f_3u_0u_2, \\
g_{32} &= 2f_1f_3u_2u_3 + 2f_3u_0u_1, \\
g_{33} &= u_0^2 - f_2f_3u_1^2 - f_1f_3u_2^2 + f_1f_2u_3^2.
\end{aligned}$$

Here

$$N(\mathbf{u}) = u_0^2 + f_2f_3u_1^2 + f_1f_3u_2^2 + f_1f_3u_3^2 = 1. \tag{2.1}$$

Let Γ be the unit group of integral quaternions, i.e. $u_j \in \mathbb{Z}$ in the above examples. If there is no room for confusion, we drop ρ and write the action of $\gamma \in \Gamma$ on $\mathbf{x} = (x_1, x_2, x_3)$ by $\mathbf{x} \to \mathbf{x}\gamma$.

Now $\rho(\Gamma) \subset SO_f(\mathbb{Z})$ and $V(\mathbb{Z})$ decomposes into finitely many Γ orbits

$$V(\mathbb{Z}) = \bigsqcup_{j=1}^{h} \mathbf{y}^{(j)}\Gamma. \tag{2.2}$$

The quadric $V(\mathbb{Z})$ in 3-variables does not satisfy strong approximation (see Borovoi [2] for a quantification of its failure), so for our analysis, we fixate on each orbit $\mathbf{y}^{(j)}\Gamma$ separately. We fix such an orbit $\mathscr{O} = \mathbf{y}\Gamma$. Let K be a maximal compact subgroup of $SO_f(\mathbb{R})$, and let $|\cdot|$ be a Euclidean norm on $\mathbb{R}^3$ which is K invariant. In Section 3, we construct smooth functions $F_T : \mathbb{R}^3 \to \mathbb{R}$ for $T \geq 10$ which depend only on $|\mathbf{x}|$ and satisfy

$$\begin{cases} \text{(i)} & 0 \leq F_T(\mathbf{x}) \leq 1; \\ \text{(ii)} & F_T(\mathbf{x}) = 1, \qquad \text{if } |\mathbf{x}| \leq T/c_0; \\ \text{(iii)} & F_T(\mathbf{x}) = 0, \qquad \text{if } |\mathbf{x}| \geq c_0 T; \end{cases} \tag{2.3}$$

where c_0 is a positive constant depending only on $\mathscr{O}$.

For $n \geq 0$, set

$$a_n(T) = \sum_{\substack{\mathbf{x} \in \mathscr{O} \\ x_1 x_2 x_3 = \pm n}} F_T(\mathbf{x}). \tag{2.4}$$

Note that $a_n(T) \geq 0$ and $a_n(T) = 0$ if $n \geq c_1 T^3$ for a suitable constant c_1. If

$$X = \sum_{n \geq 1} a_n(T), \tag{2.5}$$

then we show in Section 3 that

$$X \sim c_2(\mathscr{O})T \quad \text{as } T \to \infty, \tag{2.6}$$

where $c_2(\mathscr{O})$ is a positive constant.

The splitting of D_f at infinity yields $D_f \otimes \mathbb{R} \cong M_2(\mathbb{R})$ and this realizes Γ as a co-compact lattice in $SL_2(\mathbb{R})$. Let $0 \leq \theta < 1/4$ be so that the first eigenvalue λ_1 of the Laplacian on the hyperbolic surface $X_{\Gamma(q)} = \Gamma(q)\backslash\mathbb{H}$, where $\Gamma(q)$ is the principal congruence subgroup of Γ of level $q \geq 1$, satisfies

$$\lambda_1(X_{\Gamma(q)}) \geq \frac{1}{4} - \theta^2. \tag{2.7}$$

Applying the Jacquet-Langlands correspondence [18] and the bound of Kim-Sarnak [19], we have that $\theta = 7/64$ is such a number, while Selberg's Eigenvalue Conjecture [24] implies that $\theta = 0$ is admissible.

Denote by $\mathscr{O}(\mathbb{Z}/d\mathbb{Z})$ the orbit $\mathbf{y}\Gamma$ in $(\mathbb{Z}/d\mathbb{Z})^3$, and by $\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})$ the subset of $\mathscr{O}(\mathbb{Z}/d\mathbb{Z})$ for which $y_1y_2y_3 \equiv 0(\mathrm{mod} d)$. The following key level distribution theorem is proven in Section 3.

Theorem 2.1. *Let f and $\mathscr{O}$ be as above. Then, for $1 \leq d \leq T$, we have*

$$\sum_{\substack{n\equiv 0(\mathrm{mod}\ d)\\ n\geq 1}} a_n(T) = \frac{|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|}X + O_\varepsilon(d^{1+\varepsilon}T^{1/2+\theta+\varepsilon}). \tag{2.8}$$

As with $V(\mathbb{Z}/d\mathbb{Z})$ and $V^0(\mathbb{Z}/d\mathbb{Z})$, we say that p is "bad" for $\mathscr{O}$ if $\mathscr{O}(\mathbb{Z}/d\mathbb{Z}) = \mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})$. We show in Section 4 that the only possibility for a "bad" p for $\mathscr{O}$ is a p that lies in $\{2, 3, 5, 7\}$. Let $B(\mathscr{O})$ denote the set of bad primes for $\mathscr{O}$. For $r \geq 1$ and B a subset of primes, let $P_r(B)$ denote the set of integers which have at most r prime factors outside B. In Section 6, we use the 3-dimensional weighted sieve of [7] to prove

Theorem 2.2. *Let $B(\mathscr{O})$ be the set of bad primes for $\mathscr{O}$. Then*

$$\sum_{n\in P_{26}(B(\mathscr{O}))} a_n(T) \gg \frac{T}{\log^3 T}.$$

Moreover, if $\theta = 0$ is admissible in (2.7), then the above holds with 26 replaced by 22.

Corollary 2.3. *The set of $\mathbf{x} \in \mathscr{O}$, for which $x_1x_2x_3$ has at most 26 prime factors outside of $B(\mathscr{O})$, is Zariski dense in V. If we assume $\theta = 0$ in (2.7), then 26 may be replaced by 22.*

3. Spectral theory and counting

We need some lemmas which concern counting points in regions for the action of a lattice Γ in $SL_2(\mathbb{R})$. A key point being the uniformity, assuming the spectral gap condition (2.7) as Γ varies over subgroups of a given such group.

The cases where $V_{f,t}(\mathbb{R})$ is a two or one sheeted hyperboloid are handled by an analysis on different spaces, so we deal with them separately. We begin with the two sheeted hyperboloid in which case each sheet is a hyperbolic plane, and we develop some general facts about the counting, with uniform small remainder, of a co-compact lattice Γ in $SL_2(\mathbb{R})$.

Let $\mathbb{H}$ be the upper half non-Euclidian plane. We use the hyperbolic metric for all measurements. Let Γ be a co-compact lattice and assume that there is $C < \infty$ such that

$$|\{\gamma \in \Gamma : d(\gamma z, z) \leq 1\}| \leq C \tag{3.1}$$

for any $z \in \mathbb{H}$, where $d(z, w)$ is the distance from z to w. This condition asserts that $\Gamma\backslash\mathbb{H}$ has no small geodesic loops.

For $w \in \mathbb{H}$, let (ρ, θ), $\rho \geq 0$ and $\theta \in [0, 2\pi)$, denote polar coordinates about w. For $R \geq 10$, let $F_R(\rho)$ be a function satisfying

$$\begin{cases} \text{(i)} & 0 \leq F_R(\rho) \leq 1; \\ \text{(ii)} & F_R(\rho) = 1, \text{ for } \rho \leq R-1; \\ & F_R(\rho) = 0, \text{ for } \rho \geq R+1; \\ \text{(iii)} & \text{The derivatives up to a fixed order } l \text{ of } F_R \text{ with respect to } \rho \\ & \text{are bounded by } c(l). \end{cases} \tag{3.2}$$

Define $K_R(z, w)$ for $z, w \in \mathbb{H}$ by

$$K_R(z, w) = \sum_{\gamma \in \Gamma} F_R(d(\gamma z, w)). \tag{3.3}$$

This will be our main counting function in (2.3); it is a smoothed out version of counting the number of images of z under γ which lie in a ball about w of radius R. Note that $K_R(\gamma z, \delta w) = K_R(z, w)$ for $\gamma, \delta \in \Gamma$ and $z, w \in \mathbb{H}$. Let

$$X = X(R) = \int_{\mathbb{H}} F_R(z, w) dA(z) = 2\pi \int_0^\infty F_R(\rho) \sinh \rho d\rho, \tag{3.4}$$

so that $X \asymp e^R$ as $R \to \infty$.

Proposition 3.1. *If Γ satisfies (2.7) and (3.1), then for any $\varepsilon > 0$,*

$$K_R(z, w) = \frac{X(R)}{\text{Area}(\Gamma \backslash \mathbb{H})} + O_\varepsilon\{e^{(1/2+\theta+\varepsilon)R}\},$$

where the implied constant depends only on ε, on C in (3.1), and on $c(l)$ in (3.2).

Proof. As with the familiar treatment of the hyperbolic lattice point counting problem ([6], [21], [12], [9]), we expand the function K spectrally. Let $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$ be the spectrum of Laplacian on $L^2(\Gamma \backslash \mathbb{H})$ and $\{\phi_0, \phi_1, \phi_2, \dots\}$ a corresponding orthonormal basis of eigenfunctions. Write $\lambda_j = 1/4 + t_j^2$ with $t_j \geq 0$ or $t_j \in i\mathbb{R}^+$. A well-known calculation (see [24]) gives

$$K_R(z, w) = \sum_{j=0}^\infty h_R(t_j) \phi_j(z) \overline{\phi_j(w)}. \tag{3.5}$$

Here

$$\begin{aligned} h_R(t) &= \int_{\mathbb{H}} F_R(z, w) \omega_t(z, w) dA(z) \\ &= 2\pi \int_0^\infty F_R(\rho) \omega_t(\rho) \sinh \rho d\rho, \end{aligned} \tag{3.6}$$

where $\omega_t(z,w)$ is the unique spherical function about w with eigenvalue $\lambda = 1/4+t^2$ normalized with $\omega_t(w,w) = 1$, while $\omega_t(\rho)$ is the same function in polar coordinates (so it does not depend on w).

For $\lambda = 0$ i.e. $t = i/2$, we have $\omega_{i/2}(\rho) \equiv 1$ and

$$h_R\left(\frac{i}{2}\right) = X(R), \tag{3.7}$$

while

$$\phi_0(z) = \frac{1}{\sqrt{\text{Area}(\Gamma\backslash\mathbb{H})}}. \tag{3.8}$$

Thus the contribution from $j = 0$ in (3.5) is the main term in Proposition 3.1. So we must show that

$$\sum_{j\neq 0} h_R(t_j)\phi_j(z)\overline{\phi_j(w)}$$

is bounded as claimed.

It follows from Harish-Chandra's formula for spherical functions [15] and in our setting from the theory of Legendre functions that

$$\omega_t(\rho) \ll (\rho+1)e^{(-1/2+\theta)\rho} \tag{3.9}$$

if λ satisfies (2.7). Hence for $l \geq 0$ an integer

$$\begin{aligned} h_R(t) &= 2\pi\int_0^\infty \omega_t(\rho)F_R(\rho)\sinh\rho d\rho \\ &= 2\pi\lambda^{-l}\int_0^\infty \{\Delta_\rho\omega_t(\rho)\}F_R(\rho)\sinh\rho d\rho \\ &= 2\pi\lambda^{-l}\int_0^\infty \omega_t(\rho)\{\Delta_\rho F_R(\rho)\}\sinh\rho d\rho \\ &\ll_{c(l),\varepsilon} \lambda^{-l}e^{(1/2+\theta+\varepsilon)R} \end{aligned} \tag{3.10}$$

by (3.2) and (3.9). It follows that, for a fixed $l \geq 0$,

$$\sum_{j\neq 0} h_R(t_j)\phi_j(z)\overline{\phi_j(w)} \ll_{\varepsilon,l} e^{(1/2+\theta+\varepsilon)R}\sup_z \sum_{j\neq 0}\lambda_j^{-l}|\phi_j(z)|^2. \tag{3.11}$$

In order to estimate the last sum uniformly in z and Γ, we go back to (3.5), and use positivity and an $F(\rho)$ of small support. For $\varepsilon > 0$, let F_ε be positive definite ε-appropriate identity. That

is

$$\int_{\mathbb{H}} F_\varepsilon(d(z,w))dA(z) = 1, \tag{3.12}$$

and F_ε is supported in $d(z,w) < \varepsilon$ with

$$F_\varepsilon(z,z) \asymp \frac{c_3}{\varepsilon^2}. \tag{3.13}$$

Moreover,

$$h_{F_\varepsilon}(t) \geq 0, \quad \text{for } t \in \mathbb{R} \cup i\mathbb{R}^+, \tag{3.14}$$

and

$$h_{F_\varepsilon}(t) \geq \frac{1}{10}, \quad \text{for } |t| \leq \frac{1}{\varepsilon}. \tag{3.15}$$

Such F_ε's are easily constructed by convolving, in the algebra of point-pair invariants (see [25]), two conjugate functions of smaller support (see for example [23]).

With such an F_ε, (3.5) reads

$$\sum_j h(t_j)|\phi_j(z)|^2 = \sum_{\gamma\in\Gamma} F_\varepsilon(d(\gamma z, z)), \tag{3.16}$$

and hence by (3.1),

$$\sum_{t_j\in i\mathbb{R}^+} |\phi_j(z)|^2 + \sum_{\substack{t_j\in\mathbb{R} \\ |t_j|\leq 1/\varepsilon}} |\phi_j(z)|^2 \ll K_{F_\varepsilon}(z,z) \ll_C \frac{1}{\varepsilon^2}, \tag{3.17}$$

where the implied constant depends on C. Put another way, we have for $\xi \geq 1$,

$$\sum_{\lambda_j\leq\xi} |\phi_j(z)|^2 \ll_C \xi \tag{3.18}$$

with the implied constant depending only on C in (3.1) and not on Γ or z. It follows on summing by parts that, for $l = 2$,

$$\sum_{\lambda_j\geq 1} \lambda_j^{-2}|\phi_j(z)|^2 \ll_C 1. \tag{3.19}$$

Hence from (3.11) we conclude that

$$\sum_{j\neq 0} h_R(t_j)\phi_j(z)\overline{\phi_j(w)} \ll_{\varepsilon,l,C} e^{(1/2+\theta+\varepsilon)R}, \tag{3.20}$$

proving Proposition 3.1. □

For the quadric which is a one-sheeted hyperboloid, the counting problem reduces to one of counting points in the orbit of Γ in a half annulus. After conjugation in $SL_2(\mathbb{R})$, we can assume that our variable Γ contains the fixed hyperbolic cyclic subgroup Γ_ξ where

$$\Gamma_\xi = \left\{ \begin{pmatrix} k^{1/2} & 0 \\ 0 & k^{-1/2} \end{pmatrix}^n : \quad n \in \mathbb{Z} \right\}. \tag{3.21}$$

Here $k > 1$ and Γ_ξ is primitive in Γ. We want to count cosets $\Gamma_\xi \backslash \Gamma$ ordered in a natural way or what is the same orbit γz in the fundamental domain for Γ_ξ

$$\mathscr{F}_\xi = \{z \in \mathbb{H} : 1 \leq |z| < k\}. \tag{3.22}$$

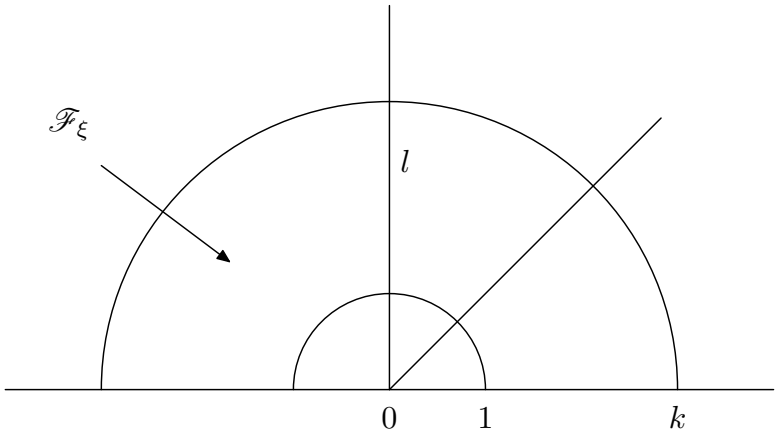

FIGURE 1. $\mathscr{F}_\xi$: the fundamental domain for Γ_ξ

Let (ρ, t), $-\infty < \rho < \infty$ and $0 \leq t \leq \log k = \kappa$, be Fermi coordinates for the "flare" $\mathscr{F}_\xi$ (see Gamburd [11]). This time ρ is the signed distance of $z = (\rho, t) \in \mathscr{F}_\xi$ to the closed geodesic ℓ and $r = e^t = |z|$. The action $z \to \kappa z$ of Γ_ξ becomes

$$(\rho, t) \to (\rho, \kappa + t). \tag{3.23}$$

In these coordinates, the line element takes the form

$$ds^2 = d\rho^2 + (\cosh \rho)^2 dt^2. \tag{3.24}$$

Let $R \geq 10$ and $F_R(z) = F_R(\rho)$ be a smooth even function of ρ satisfying

$$\begin{cases} \text{(i)} & 0 \leq F_R(\rho) \leq 1; \\ \text{(ii)} & F_R(\rho) = 1, \text{ for } \rho \leq R-1; \\ & F_R(\rho) = 0, \text{ for } \rho \geq R+1; \\ \text{(iii)} & \text{The } l\text{-th derivative of } F_R \text{ with respect to } \rho \\ & \text{are bounded by } c(l). \end{cases} \tag{3.25}$$

So,

$$F_R(\gamma z) = F_R(z), \quad \text{for } \gamma \in \Gamma_\xi. \tag{3.26}$$

Define the weighted counting function H_R by

$$H_R(z) := \sum_{\gamma \in \Gamma_\xi \backslash \Gamma} F_R(\gamma z). \tag{3.27}$$

Proposition 3.2. *Fix*

$$\Lambda = \left\{ \begin{pmatrix} k^{1/2} & 0 \\ 0 & k^{-1/2} \end{pmatrix}^n : \quad n \in \mathbb{Z} \right\}.$$

Let Γ be a (variable) subgroup of $SL_2(\mathbb{R})$ satisfying (2.7) and (3.1), and such that $\Gamma \supset \Gamma_\xi = \Lambda$. Then, for any $\varepsilon > 0$,

$$H_R(z) = \frac{X(R)}{\text{Area}(\Gamma \backslash \mathbb{H})} + O_\varepsilon\{e^{(1/2+\theta+\varepsilon)R}\},$$

where the implied constant depends only on ε, C in (3.1), and $c(l)$ in (3.25), but not on z or Γ. Here

$$X(R) = \kappa \int_{-\infty}^{\infty} F_R(\rho) \cosh \rho d\rho. \tag{3.28}$$

Note that $X(R) \asymp e^R$, while if $\theta = 0$ the remainder term is essentially $O(e^{R/2})$.

Proof. The proof is similar to that of Proposition 3.1, the difference being with the special functions and periods that intervene. For simplicity, we assume that $\theta = 0$ in (2.7). From its definition in (3.27) it follows that $H_R(\gamma z) = H_R(z)$ for $\gamma \in \Gamma$, so we expand $H_R(z)$ in the orthonormal basis $\{\phi_1, \phi_2, \ldots\}$,

$$H_R(z) = \sum_{j=0}^{\infty} \langle H_R, \phi_j \rangle \phi_j(z), \tag{3.29}$$

where

$$\begin{aligned}\langle H_R, \phi_j\rangle &= \int_{\mathscr{F}_\Gamma} H_R(z)\phi_j(z)dA(z)\\ &= \int_{\mathscr{F}_\Gamma} \sum_{\gamma\in\Gamma_\xi\backslash\Gamma} F_R(\gamma z)\phi_j(z)dA(z)\\ &= \int_{\mathscr{F}_\xi} F_R(z)\phi_j(z)dA(z)\\ &= \int_{-\infty}^{\infty} F_R(\rho)\bigg(\int_0^\kappa \phi_j(\rho,t)dt\bigg)\cosh\rho d\rho. \end{aligned} \tag{3.30}$$

The inner integral which is a function of ρ satisfies the differential equation

$$\frac{1}{\cosh\rho}\frac{d}{d\rho}\bigg(\cosh\rho\frac{d\psi}{d\rho}\bigg)+\lambda\psi=0. \tag{3.31}$$

Two linearly independent solutions of (3.31) can be chosen to be even and odd with

$$\begin{cases} \psi^e_\lambda(\rho) \text{ even and } \psi^e_\lambda(0)=1;\\ \psi^o_\lambda(\rho) \text{ odd and } \frac{d}{d\rho}\psi^o_\lambda\big|_{\rho=0}=1. \end{cases} \tag{3.32}$$

This determines ψ^e_λ and ψ^o_λ uniquely. In this way, we have

$$\int_0^\kappa \phi_j(\rho,t)dt = A_j\psi^e_{\lambda_j}(\rho)+B_j\psi^o_{\lambda_j}(\rho), \tag{3.33}$$

where A_j is the "period"

$$A_j=\int_0^\kappa \phi_j(0,t)dt=\int_1^k \phi_j(iy)\frac{dy}{y}. \tag{3.34}$$

Returning to (3.29) and recalling that F_R is even, we have

$$H_R(z)=\sum_{j=0}^{\infty}\bigg(\int_1^k\phi_j(iy)\frac{dy}{y}\bigg)\bigg(\int_0^\infty F_R(\rho)\psi^e_{\lambda_j}(\rho)\cosh\rho d\rho\bigg)\phi_j(z). \tag{3.35}$$

The term with $j=0$ gives

$$\frac{\log k}{\text{Area}(\Gamma\backslash\mathbb{H})}\int_{-\infty}^{\infty}F_R(\rho)\cosh\rho d\rho=\frac{X(R)}{\text{Area}(\Gamma\backslash\mathbb{H})}. \tag{3.36}$$

As before, we need to estimate the contribution from $j\neq 0$ in (3.35).

To this end, we identify explicitly the function $\psi_\lambda^e(\rho)$ in terms of the Legendre function P_ν^μ (see Gamburd [11]). If $\lambda = 1/4 + \xi^2$, then

$$\psi_\lambda^e(\rho) = \frac{1}{2P_\nu^\mu(0)} \frac{1}{\sqrt{\cosh \rho}} \bigg(P_\nu^\mu(\tanh \rho) + P_\nu^\mu(-\tanh \rho) \bigg), \tag{3.37}$$

where $\mu = -i\xi, \nu = -1/2$. We use the integral formula ([13], 8.714)

$$P_{-1/2}^{-i\xi}(\cos \phi) = \frac{\Gamma(1+2i\xi)(\sin \phi)^{i\xi}}{2^{i\xi}\Gamma(1+i\xi)\Gamma(1/2+i\xi)\Gamma(-1/2+i\xi)} \int_0^\infty \frac{t^{-1/2+i\xi}dt}{(1+2t\cos\phi+t^2)^{1/2+i\xi}}, \tag{3.38}$$

and ([13], 8.756)

$$P_{-1/2}^{-i\xi}(0) = \frac{2^{-i\xi}\sqrt{\pi}}{\Gamma(3/4+i\xi/2)^2}. \tag{3.39}$$

Hence, by Stirling's series for the Γ-function and an elementary estimation of the integral in (3.38), we have for $\xi \geq 0$,

$$\frac{P_\lambda^e(\cos\phi)}{P_{-1/2}^{-i\xi}(0)} \ll (1+\xi)^3(1+|\rho|). \tag{3.40}$$

Corollary 3.3. *For $\lambda \geq 1/4$ i.e. $\xi \geq 0$,*

$$\psi_{-1/2}^{-i\xi}(\rho) \ll \frac{\lambda^{3/2}(1+|\rho|)}{\sqrt{\cosh\rho}}.$$

For $l \geq 0$ fixed, we have

$$\begin{aligned}
\int_{\mathscr{F}_\xi} F_R(z)\phi_j(z)dA(z) &= \lambda_j^{-l} \int_{\mathscr{F}_\xi} \{\Delta^l \phi_j(z)\} F_R(z) dA(z) \\
&= \lambda_j^{-l} \int_{\mathscr{F}_\xi} \phi_j(z) \{\Delta^l F_R(z)\} dA(z) \\
&= \lambda_j^{-l} \int_{\mathscr{F}_\xi} \phi_j(\rho,t) \bigg(\frac{1}{\cosh\rho} \frac{d}{d\rho} \bigg(\cosh \frac{d}{d\rho} \bigg) \bigg)^\nu F_R(\rho) \cosh\rho d\rho \\
&\ll_{c(l)} \lambda_j^{-l} \bigg| \int_1^k \phi_j(iy) \frac{dy}{y} \bigg| \int_{R-1}^{R+1} |\psi_j^e(\rho)| \cosh\rho d\rho \\
&\ll_{c(l)} \lambda_j^{3/2-l} Re^{R/2} \bigg| \int_1^k \phi_j(iy) \frac{dy}{y} \bigg|.
\end{aligned} \tag{3.41}$$

Hence,

$$\begin{aligned}\sum_{j\neq 0}\langle H_R,\phi_j\rangle\phi_j(z) &\ll Re^{R/2}\sum_{j\neq 0}\lambda_j^{3/2-l}\left|\int_1^k \phi_j(iy)\frac{dy}{y}\right||\phi_j(z)| \\ &\ll Re^{R/2}\int_1^k\sum_{j\neq 0}\lambda_j^{3/2-l}|\phi_j(iy)||\phi_j(z)|\frac{dy}{y} \\ &\ll Re^{R/2}\sup_z\sum_{j\neq 0}\lambda_j^{3/2-l}|\phi_j(z)|^2. \end{aligned} \tag{3.42}$$

Applying (3.18) and summing by parts with $l=3$, we get from (3.42) that

$$\sum_{j\neq 0}\langle H_R,\phi_j\rangle\phi_j(z) \ll_{C,c(l)} Re^{R/2}. \tag{3.43}$$

This completes the proof of Proposition 3.2 when $\theta = 0$. The case of $\theta > 0$ is handled in the same way. □

Next we construct the key weight functions $F_T(\mathbf{x})$ in (2.3) using the corresponding functions $F_R(z)$ in (3.3) and and (3.6). To this end, we make a fixed (f and t are fixed) linear change of variables over $\mathbb{R}$ with $\mathbf{x}$ going to ξ, taking f to g where

$$g(\xi) = g(\xi_1,\xi_2,\xi_3) = \xi_2^2 - \xi_1\xi_3, \tag{3.44}$$

and V_t to

$$\begin{cases} W_1: & g(\xi) = 1/2 \text{ if } V_t \text{ is one-sheeted;} \\ W_{-1}: & g(\xi) = -1/2 \text{ if } V_t \text{ is two-sheeted.} \end{cases} \tag{3.45}$$

In these coordinates, the spin double cover morphism τ from $SL_2(\mathbb{R})$ to $SO_g(\mathbb{R})$ can be given by

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \xrightarrow{\tau} \begin{pmatrix} \alpha^2 & \alpha\beta & \beta^2 \\ 2\alpha\gamma & \alpha\delta+\beta\gamma & 2\beta\delta \\ \gamma^2 & \gamma\delta & \delta^2 \end{pmatrix} \tag{3.46}$$

with kernel $\pm I$. Then over $\mathbb{R}$ we have that

$$W_1 = \mathbf{v}_1\tau(SL_2(\mathbb{R})) \tag{3.47}$$

where $\mathbf{v}_1 = (0, 1/\sqrt{2}, 0)$, and if W_{-1}^0 is a connected component of W_{-1} then

$$W_{-1}^0 = \mathbf{v}_{-1}\tau(SL_2(\mathbb{R})) \tag{3.48}$$

with $\mathbf{v}_{-1} = (1/\sqrt{2}, 0, 1/\sqrt{2})$. In (3.47), the stabilizer of $\mathbf{v}_1$ is the split torus

$$H = \left\{ \pm \begin{pmatrix} \lambda^{1/2} & 0 \\ 0 & \lambda^{1/2} \end{pmatrix} : \quad \lambda > 0 \right\}, \tag{3.49}$$

while in (3.48) the stabilizer of $\mathbf{v}_{-1}$ is the compact $K = SO_2$, i.e.

$$K = \left\{ \pm \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : \quad 0 \le \theta < 2\pi \right\}. \tag{3.50}$$

Let $|\cdot|$ be the Euclidean norm on ξ,

$$|\xi|^2 = \xi_1^2 + 2\xi_2^2 + \xi_3^2. \tag{3.51}$$

Then

$$|\xi\tau(k)| = |\xi| \quad \text{for } k \in K. \tag{3.52}$$

Hence

$$|\mathbf{v}_1\tau(g)| = |\mathbf{v}_1\tau(hgk)| \tag{3.53}$$

for any $h \in H, k \in K$, and $g \in SL_2(\mathbb{R})$, while

$$|\mathbf{v}_{-1}\tau(g)| = |\mathbf{v}_{-1}\tau(k_1 g k_2)| \tag{3.54}$$

for any $k_1, k_2 \in K$, and $g \in SL_2(\mathbb{R})$.

Now any $g \in SL_2(\mathbb{R})$ can be factored uniquely as

$$g = h \begin{pmatrix} 1 & t/2\sqrt{2} \\ 0 & 1 \end{pmatrix} k \tag{3.55}$$

with $t \in \mathbb{R}, h \in H, k \in K$. It follows from (3.53) that in terms of this factorization

$$|\mathbf{v}_1\tau(g)|^2 = 1 + t^2. \tag{3.56}$$

Similarly, if

$$g = k_1 \begin{pmatrix} \lambda^{1/2} & 0 \\ 0 & \lambda^{-1/2} \end{pmatrix} k_2 \tag{3.57}$$

with $k_1, k_2 \in K$ and $\lambda \ge 1$, then

$$|\mathbf{v}_{-1}\tau(g)|^2 = \frac{1}{2}(\lambda^2 + \lambda^{-2}). \tag{3.58}$$

Using these, we transplant the function F_R in (3.6) and (3.3) to W_1 and W_{-1}^0 respectively. For $\xi \in W_1$, $\xi = \mathbf{v}_1\tau(g)$, set

$$\widetilde{F}_T(\xi) = \widetilde{F}_T(\mathbf{v}_1\tau(g)) := F_R(gi) \tag{3.59}$$

where in the last $z \to gz$ is the usual linear factional action on $\mathbb{H}$, and where

$$T = \sinh R. \tag{3.60}$$

Note that from (3.50), (3.52), (3.53), and (3.54), $\widetilde{F}_T$ is well defined and depends only on $|\xi|$. Also from (3.56) and a simple calculation

$$|\xi| = \cosh \rho \quad \text{ for } \xi \in W_1. \tag{3.61}$$

Thus, $\widetilde{F}_T(\xi)$ with T satisfying (3.60) satisfies the condition (2.4).

For $\xi \in W_{-1}^0$ and $\xi = \mathbf{v}_1 \tau(g)$, set

$$\widetilde{F}_T(\xi) = \widetilde{F}_T(\mathbf{v}_1 \tau(g)) = F_R(gi), \tag{3.62}$$

with

$$T = \sqrt{\cosh 2R}. \tag{3.63}$$

Again, $\widetilde{F}_T(\xi)$ is well defined and depends only on $|\xi|$. From (3.58) and a calculation we see that for $\xi \in W_{-1}^0$,

$$|\xi|^2 = \cosh 2\rho. \tag{3.64}$$

Hence with T satisfying (3.63), we conclude that $\widetilde{F}_R(\xi)$ satisfies (2.4).

Conjugating back to variables $\mathbf{x} \in V_t$, the functions $\widetilde{F}_R(\xi)$ become $F_T(\mathbf{x})$ and these then satisfy (2.4). The group of units $G(\mathbb{Z})$ (as in (2.1)) is realized as a co-compact lattice in $SL_2(\mathbb{R})$, as in the paragraph after (2.6). Let $\mathbf{y} \in V_t(\mathbb{Z})$ be our "origin" which we fix once and for all (as in (2.2)). In the case that $\mathbf{y}$ lies on a one-sheeted hyperboloid, the stabilizer $H_{\mathbf{y}}$ of $\mathbf{y}$ in $SL_2(\mathbb{R})$ under τ is a split real torus. Let $\Lambda_{\mathbf{y}} = G(\mathbb{Z}) \cap H_{\mathbf{y}}$, the corresponding infinite (cyclic) subgroup. In the case of the two-sheeted hyperboloid, set for now $\Lambda_{\mathbf{y}} = \pm\{I\}$. Let Γ be a variable congruence subgroup of $G(\mathbb{Z})$ with $\Gamma \supset \Lambda_{\mathbf{y}}$. Then Γ satisfies (3.1) and (2.8), and hence Proposition 3.1 and (3.3) yield

$$\sum_{\gamma \in \Lambda_{\mathbf{y}} \backslash \Gamma} F_T(\mathbf{y}\tau(\gamma g)) = \frac{X(R)}{\text{Area}(\Gamma \backslash \mathbb{H})} + O_\varepsilon\{e^{(1/2+\theta)R+\varepsilon}\} \tag{3.65}$$

uniformly over such Γ and $g \in SL_2(\mathbb{R})$, where R and T are related by (3.60) and (3.63) respectively, and $X(R)$ is the function in (3.4) or (3.28).

4. Proof of Theorem 2.1

With the weighted counting approximations that we established in Section 3, we are ready to prove the key level distribution Theorem 2.1 and to examine the arithmetical properties of the main term $|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|/|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|$.

Proof of Theorem 2.1. The notation and setup is that of Section 2, and the weight function F_T on $V_t(\mathbb{R})$ is from Section 3. The orbit $\mathscr{O} = \mathbf{y}\tau(\Gamma)$ with $\Gamma = G(\mathbb{Z})$ and $a_n(T)$ is given in (2.4). For $d \geq 1$, we have

$$\begin{aligned} \sum_{\substack{n\geq 1 \\ n\equiv 0(\bmod d)}} a_n(T) &= \sum_{\substack{\mathbf{x}\in\mathscr{O} \\ x_1x_2x_3\equiv 0(\bmod d)}} F_T(\mathbf{x}) \\ &= \sum_{\substack{\gamma\in\Gamma_{\mathbf{y}}\backslash\Gamma \\ f(\mathbf{y}\gamma)\equiv 0(\bmod d)}} F_T(\mathbf{y}\tau(\gamma)), \end{aligned} \tag{4.1}$$

where $f(\mathbf{x}) = x_1x_2x_3$ and $\Gamma_{\mathbf{y}}$ is the stabilizer of $\mathbf{y}$ in Γ through the τ action. If $V_t(\mathbb{R})$ is one-sheeted, then $\Gamma_{\mathbf{y}} = \Lambda_{\mathbf{y}}$ is an infinite cyclic $(\bmod \pm I)$ subgroup of $H_{\mathbf{y}}$, while in the two sheeted case $\Gamma_{\mathbf{y}}$ is a (fixed) finite cyclic subgroup of K. Let $\Gamma_{\mathbf{y}}^{(d)}$ be the subgroup of Γ which stabilizes $\mathbf{y}$ mod d, i.e.

$$\Gamma_{\mathbf{y}}^{(d)} = \{\gamma \in \Gamma : \mathbf{y}\tau(\gamma) \equiv \mathbf{y}(\mathrm{mod}d)\}. \tag{4.2}$$

Then

$$\Lambda_{\mathbf{y}} \subset \Gamma_{\mathbf{y}} \subset \Gamma_{\mathbf{y}}^{(d)} \subset \Gamma. \tag{4.3}$$

Moreover $\Gamma_{\mathbf{y}}^{(d)}$ is clearly a congruence subgroup of Γ, and hence (3.65) applies. We can write (4.1) as

$$\sum_{n\equiv 0(\bmod d)} a_n(T) = \sum_{\substack{\delta\in\Gamma_{\mathbf{y}}^{(d)}\backslash\Gamma \\ f(\mathbf{y}\delta)\equiv 0(\bmod d)}} \sum_{\beta\in\Gamma_{\mathbf{y}}\backslash\Gamma_{\mathbf{y}}^{(d)}} F_T(\mathbf{y}\tau(\beta\delta)). \tag{4.4}$$

As to the inner sum in (4.4), it is equal to the left-hand side of (3.65) with $\Gamma_{\mathbf{y}} = \Lambda_{\mathbf{y}}$ and Γ there being $\Gamma_{\mathbf{y}}^{(d)}$, at least in the case that V_t is one-sheeted. It is equal to $1/|\Gamma_{\mathbf{y}}|$ times the left-hand side of (3.65) in the two sheeted case. By (3.65) the inner sum in (4.4) is equal to

$$\frac{cX(R)}{[\Gamma : \Gamma_{\mathbf{y}}^{(d)}]} + O_\varepsilon\{e^{(1/2+\theta+\varepsilon)R}\}, \tag{4.5}$$

where $c = c(\mathbf{y}, \Gamma)$ is a fixed positive constant independent of d. Hence (4.4) becomes

$$\begin{aligned}\sum_{n \equiv 0(\bmod d)} a_n(T) &= \sum_{\substack{\delta \in \Gamma_{\mathbf{y}}^{(d)} \backslash \Gamma \\ f(\mathbf{y}\delta) \equiv 0(\bmod d)}} \left\{ \frac{cX(R)}{[\Gamma : \Gamma_{\mathbf{y}}^{(d)}]} + O_\varepsilon(e^{(1/2+\theta+\varepsilon)R}) \right\} \\ &= \sum_{\zeta \in \mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})} \left\{ \frac{cX(R)}{[\Gamma : \Gamma_{\mathbf{y}}^{(d)}]} + O_\varepsilon(e^{(1/2+\theta+\varepsilon)R}) \right\} \\ &= \frac{|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|}{[\Gamma : \Gamma_{\mathbf{y}}^{(d)}]} cX(R) + O_\varepsilon\left\{ |\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})| e^{(1/2+\theta+\varepsilon)R} \right\}. \end{aligned} \tag{4.6}$$

Since $|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})| = [\Gamma : \Gamma_{\mathbf{y}}^{(d)}]$, we have

$$\sum_{n \equiv 0(\bmod d)} a_n(T) = \frac{|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|} cX(R) + O_\varepsilon\left\{ |\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})| e^{(1/2+\theta+\varepsilon)R} \right\}. \tag{4.7}$$

This leads us to study the numbers $|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|$ and $|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|$ as functions of d, for s square-free. Now Γ satisfies strong approximation (see [5], Chapter 10; this is the reason that we define the orbits in terms of the spin group rather than the orthogonal group), and hence the reduction $G(\mathbb{Z}) \to G(\mathbb{Z}/d\mathbb{Z})$ is onto. Moreover, if $d = d_1 d_2$ with $(d_1, d_2) = 1$ then $G(\mathbb{Z}/d\mathbb{Z}) \cong G(\mathbb{Z}/d_1\mathbb{Z}) \times G(\mathbb{Z}/d_2\mathbb{Z})$. Thus, if we denote by $\theta(d)$ the reduction of $\tau(\Gamma)$ (via the morphism in (2.1) and its generalization) in $SL_3(\mathbb{Z}/d\mathbb{Z})$, then $\theta(\mathbb{Z}/d\mathbb{Z}) \cong \theta(\mathbb{Z}/d_1\mathbb{Z}) \times \theta(\mathbb{Z}/d_2\mathbb{Z})$ inside $SL_3(\mathbb{Z}/d\mathbb{Z})$. It follows that the orbit of $\mathbf{y}$ mod d, $\mathscr{O}(\mathbb{Z}/d\mathbb{Z})$ is equal to $\mathscr{O}(\mathbb{Z}/d_1\mathbb{Z}) \times \mathscr{O}(\mathbb{Z}/d_2\mathbb{Z})$ in $(\mathbb{Z}/d_1\mathbb{Z})^3 \times (\mathbb{Z}/d_2\mathbb{Z})^3 = (\mathbb{Z}/d\mathbb{Z})^3$. From this, it follows that $\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z}) = \mathscr{O}^0(\mathbb{Z}/d_1\mathbb{Z}) \times \mathscr{O}^0(\mathbb{Z}/d_2\mathbb{Z})$ as a subset of $(\mathbb{Z}/d\mathbb{Z})^3 = (\mathbb{Z}/d_1\mathbb{Z})^3 \times (\mathbb{Z}/d_2\mathbb{Z})^3$. Hence $|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|$ and $|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|$ are multiplicative in d, as is $\omega(d)$ where

$$\omega(d) := d \frac{|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|}. \tag{4.8}$$

With this, the analysis is reduced to the case when $d = p$, a prime. If $p \nmid d(f)t$, then it follows by Lang's theorem [20], for example, that $\mathscr{O}(\mathbb{Z}/p\mathbb{Z}) = V_t(\mathbb{Z}/p\mathbb{Z})$. That is the orbit of $\mathbf{y}$ under $G(\mathbb{Z}/p\mathbb{Z})$ sweeps out the entire quadric over $\mathbb{Z}/p\mathbb{Z}$. Hence in this case

$$\begin{cases} |\mathscr{O}(\mathbb{Z}/p\mathbb{Z})| = |V_t(\mathbb{Z}/p\mathbb{Z})|; \\ |\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})| = |V_t^0(\mathbb{Z}/p\mathbb{Z})|. \end{cases} \tag{4.9}$$

These cardinalities are easy to compute, and we do so in the Appendix using Gauss sums. One can also proceed directly as in Cassels [5], Exercise 13 on page 31. For $p \nmid d(f)t$,

$$|V_t(\mathbb{Z}/p\mathbb{Z})| = p^2 + \left(\frac{-dt}{p}\right) p, \tag{4.10}$$

while

$$|V_t^0(\mathbb{Z}/p\mathbb{Z})| \leq 3p + 6. \tag{4.11}$$

Hence for d square-free,

$$|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})| \ll_\varepsilon d^{1+\varepsilon}. \tag{4.12}$$

Putting (4.12) in (4.7), we deduce Theorem 2.1. □

We conclude this section with an analysis of the possible p's for which $|\mathscr{O}(\mathbb{Z}/p\mathbb{Z})| = |\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})|$, that is the "bad primes" for the orbit $\mathscr{O}$. Our claim is this can happen only if $p \in \{2, 3, 5, 7\}$. That $2, 3$, and 5 can be bad even for $V_{f,t}(\mathbb{Z})$ and a fortiori for $\mathscr{O}$, is immediate from the following example:

$$f(x_1, x_2, x_3) = 30x_1^2 + 31x_2^2 - x_3^2, \tag{4.13}$$

with $t = 1$. f is indefinite; it is anisotropic over $\mathbb{Q}_{31}$ (the Hilbert norm residue symbol at 31 is $(30, 31)_{31} = (\frac{30}{31}) = -1$), and $d(f) = -2 \cdot 3 \cdot 5 \cdot 31$. One checks directly that $|V_1^0(\mathbb{Z}/p\mathbb{Z})| = |V_1(\mathbb{Z}/p\mathbb{Z})|$ for $p = 2, 3$, and 5.

Returning to our analysis in general of the bad primes for $\mathscr{O}$, we note that if $p \nmid td(f)$ and $p \geq 7$ then from (4.9), (4.10), and (4.11) it follows that $|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})| \neq |\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|$, since

$$|V_t(\mathbb{Z}/p\mathbb{Z})| \geq p^2 - p > 3p + 6 \geq |V_t^0(\mathbb{Z}/p\mathbb{Z})|. \tag{4.14}$$

So what remains to be considered are the cases:

(i) $p|t$ and $p \nmid d(f)$,

(ii) $p|d(f)$ and $p \nmid t$.

In the first case, we show that if $p \geq 7$ then $\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z}) \neq \mathscr{O}(\mathbb{Z}/p\mathbb{Z})$. Since $p \nmid d(f)$, f is isotropic over $\mathbb{F}_p$, and hence by a linear change of variables over $\mathbb{F}_p$ we can bring f in the form

$$f(x_1, x_2, x_3) = x_1x_3 - ax_2^2 \tag{4.15}$$

with $a = 1$ or $a = r$ a quadratic non-residue mod$p$. Say $a = 1$ (the other case is similar), then the spin double cover $G(\mathbb{Z}/p\mathbb{Z}) \cong SL_2(\mathbb{Z}/p\mathbb{Z})$, and its action can be described by (3.46) over $\mathbb{F}_p$. The orbits of $\tau(G(\mathbb{Z}/p\mathbb{Z}))$ on $V_0(\mathbb{F}_p)$ are

$$(0, 0, 0), \quad (1, 0, 0)\tau(G(\mathbb{Z}/p\mathbb{Z})), \quad (r, 0, 0)\tau(G(\mathbb{Z}/p\mathbb{Z})). \tag{4.16}$$

To see this, note that the stabilizer of $(1,0,0)$ and of $(r,0,0)$ in $G(\mathbb{Z}/p\mathbb{Z})$ is

$$\pm\left\{\begin{pmatrix} 1 & 0 \\ \xi & 1 \end{pmatrix} : \xi \in \mathbb{F}_p\right\},$$

and hence has order $2p$. Moreover these points $(1,0,0)$ and $(r,0,0)$ are in distinct orbits. Hence

$$|\mathscr{O}_{(1,0,0)}| = |\mathscr{O}_{(r,0,0)}| = \frac{p(p-1)(p+1)}{2p} = \frac{p^2-1}{2}. \tag{4.17}$$

It is elementary that

$$|V_0(\mathbb{Z}/p\mathbb{Z})| = p^2, \tag{4.18}$$

and hence (4.17) implies (4.16).

Now let L_1, L_2, L_3 be three linearly independent linear forms over $\mathbb{F}_p$, and assume that

$$\left(\bigcup_{j=1}^{3} \ker(L_j)\right) \cap \mathscr{O}(\mathbb{Z}/p\mathbb{Z}) = \mathscr{O}^0(\mathbb{Z}/p\mathbb{Z}). \tag{4.19}$$

For each j, we have that

$$|\ker(L_j) \cap V_0(\mathbb{Z}/p\mathbb{Z})| \le 2p. \tag{4.20}$$

Furthermore, since $r\mathscr{O}_{(1,0,0)}(\mathbb{Z}/p\mathbb{Z}) = \mathscr{O}_{(r,0,0)}(\mathbb{Z}/p\mathbb{Z})$ and $\ker(L_j) \cap V_0(\mathbb{Z}/p\mathbb{Z})$ is invariant under multiplication by any non-zero λ, it follows from (4.20) that

$$|\ker(L_j) \cap \mathscr{O}(\mathbb{Z}/p\mathbb{Z})| \le p. \tag{4.21}$$

Hence

$$|\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})| = \left|\left(\bigcup_{j=1}^{3} \ker(L_j)\right) \cap \mathscr{O}(\mathbb{Z}/p\mathbb{Z})\right| \le 3p. \tag{4.22}$$

According to (4.17),

$$|\mathscr{O}(\mathbb{Z}/p\mathbb{Z})| = \frac{p^2-1}{2} > 3p = |\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})|$$

if $p \ge 7$. This proves our claim in case (i).

In case (ii), we show that $|\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})| \ne |\mathscr{O}(\mathbb{Z}/p\mathbb{Z})|$ for $p \ge 11$, and in fact equality may occur for $p = 7$. In this case $p|d(f)$, and since $d(f)$ is square-free, f may be brought by linear change of variables over $\mathbb{Z}_p$ into the form

$$f(x_1,x_2,x_3) = px_1^2 + bx_2^2 + cx_3^2 \tag{4.23}$$

with $p \nmid bc$ and also $p \nmid t$. Again we can count the number of points on $V_t(\mathbb{Z}/p\mathbb{Z})$:

$$|V_t(\mathbb{Z}/p\mathbb{Z})| = \left(p - \left(\frac{-bc}{p}\right)\right)p. \tag{4.24}$$

According to (2.1) the spin double cover $G(\mathbb{Z}/p\mathbb{Z})$ is given by $\mathbf{u} = (u_0, u_1, u_2, u_3) \in \mathbb{F}_p^4$ with

$$u_0^2 + bcu_1^2 = 1 \tag{4.25}$$

with the action $\tau(\mathbf{u})$ given by

$$\begin{pmatrix} 1 & 0 & 0 \\ 2bu_0u_3 + 2bcu_1u_2 & u_0^2 - bcu_1^2 & -2bu_0u_1 \\ 2bcu_1u_3 - xu_0u_2 & 2cu_0u_1 & u_0^2 - bcu_1^2 \end{pmatrix}. \tag{4.26}$$

Hence

$$|G(\mathbb{Z}/p\mathbb{Z})| = p^2\left(p - \left(\frac{-bc}{p}\right)\right). \tag{4.27}$$

If $(x_0, y_0, z_0) \in V_t(\mathbb{Z}/p\mathbb{Z})$, then clearly $y_0z_0 \neq 0$. Now if $\tau(\mathbf{u})$ stabilizes (x_0, y_0, x_0), then

$$(y_0, z_0)\begin{pmatrix} u_0^2 - bcu_1^2 & -2bu_0u_1 \\ 2cu_0u_1 & u_0^2 - bcu_1^2 \end{pmatrix} = (y_0, z_0). \tag{4.28}$$

So completing $\mathbf{v}_0 = (y_0, z_0)$ to a basis $\{\mathbf{v}_0, \mathbf{v}_1\}$ of $\mathbb{F}_p^2$, the linear transformation above in this new basis takes the form

$$\begin{pmatrix} 1 & 0 \\ \alpha & \beta \end{pmatrix}.$$

Since its determinant is 1, it follows that $\beta = 1$, and hence that its trace is 2. That is

$$u_0^2 - bcu_1^2 = 1. \tag{4.29}$$

We also have (4.25) and thus conclude that $u_0 = \pm 1$. It follows easily from this and (4.26) that the stabilizer in $G(\mathbb{Z}/p\mathbb{Z})$ of (x_0, y_0, z_0) has order $2p$, and hence from (4.27) that

$$|\mathscr{O}_{(x_0,y_0,z_0)}(\mathbb{Z}/p\mathbb{Z})| = \frac{p}{2}\left(p - \left(\frac{-bc}{p}\right)\right) = \frac{V_t(\mathbb{Z}/p\mathbb{Z})}{2}. \tag{4.30}$$

That is, there are two $\tau(G(\mathbb{Z}/p\mathbb{Z}))$ orbits. Now let L_1, L_2, L_3 be our three (linearly independent) linear forms in x_1, x_2, x_3, for which we want to find an $(x_1, x_2, x_3) \in \mathscr{O}_{(x_0,y_0,z_0)}$ such that $L_j(\mathbf{x}) \neq 0$ for $j = 1, 2, 3$. The problematic case, which we examine, is when two of three forms, say L_2 and L_3, involve only x_2 and x_3. In this case, consider only the x_2 and x_3 coordinates and the action (4.28) on these. We have seen that using u_0 and u_1 we can produce $(p - (\frac{-bc}{p}))/2$ distinct points (x_2, x_3) (ignoring the first coordinate x_1) in the orbit of (4.28). So if $p \geq 11$, there are

at least 5 such points. On the other hand, the number of of points (x_2, x_3) on $f(x_2, x_3) = t$ and for which $L_j(x_2, x_3) = 0$ with $j = 2$ (respectively 3) is at most 2. Hence there is a point (x_1, x_2, x_3) in $\mathscr{O}_{(x_0,y_0,z_0)}(\mathbb{Z}/p\mathbb{Z})$ such that $L_2(\mathbf{x}) \neq 0$ and $L_3(\mathbf{x}) \neq 0$. Now using the full group (4.26) to vary x_1 (keeping x_2 and x_3 fixed), it is easy to see that we can arrange for a point $(x_1, x_2, x_3) \in \mathscr{O}_{(x_0,y_0,z_0)}(\mathbb{Z}/p\mathbb{Z})$ such that $L_1(\mathbf{x})L_2(\mathbf{x})L_3(\mathbf{x}) \neq 0$. This completes the analysis that for $p|d(f)$ and $p \geq 11$, we have

$$\begin{aligned} \mathscr{O}^0(\mathbb{Z}/p\mathbb{Z}) &\leq 4p < \frac{p^2-p}{2} \leq \frac{1}{2}\left\{p^2 - \left(\frac{-bc}{p}\right)p\right\} \\ &= \mathscr{O}(\mathbb{Z}/p\mathbb{Z}). \end{aligned} \tag{4.31}$$

For $p = 7$, the above argument fails, and that is because it may happen that $\mathscr{O}^0(\mathbb{Z}/7\mathbb{Z}) = \mathscr{O}(\mathbb{Z}/7\mathbb{Z})$. Let

$$f(x_1, x_2, x_3) = 7x_1^2 + x_2^2 + x_3^2, \quad t = 1.$$

The subgroup $\tau(G(\mathbb{Z}/p\mathbb{Z}))$ in (4.26) has the form

$$\begin{pmatrix} 1 & 0 & 0 \\ * & & \\ * & & U \end{pmatrix}$$

where

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Hence the orbit $\mathscr{O}_{(0,0,1)}(\mathbb{Z}/7\mathbb{Z})$ consists of points of the form $(*, 0, \pm 1)$ or $(*, \pm 1, 0)$. In particular, if $L_2(\mathbf{x}) = x_2$ and $L_3(\mathbf{x}) = x_3$ then $L_2(\mathbf{x})L_3(\mathbf{x}) = 0$ on $\mathscr{O}_{(0,0,1)}(\mathbb{Z}/7\mathbb{Z})$.

5. Sieves of dimension $\kappa > 1$

Let $\mathscr{A}$ be a finite sequence of real numbers $a_n \geq 0$, and B a fixed finite set of primes. We are interested in a reasonable lower-bound estimate for

$$\sum_{n \in P_r(B)} a_n, \tag{5.1}$$

where $P_r(B)$ is the set of positive integers with at most r prime divisors outside B. To estimate (5.1), we need to know how $\mathscr{A}$ is distributed to each of the arithmetic progression $0(\mathrm{mod}d)$, where d is square-free and $(d, B) = 1$. To this end, let d be a square-free number, and write

$$\mathscr{A}_d = \{a_n \in \mathscr{A} : n \equiv 0(\mathrm{mod}d)\}. \tag{5.2}$$

We note that $\mathscr{A}_1 = \mathscr{A}$. Suppose there exists an approximation X to $|\mathscr{A}| := \sum a_n$ and a non-negative multiplicative function $\omega(d)$ satisfying

$$\begin{cases} \omega(1) = 1; \\ 0 \le \omega(p) < p, & \text{if } p \notin B; \\ \omega(p) = 0, & \text{if } p \in B; \end{cases} \tag{5.3}$$

and for some fixed (independent of z, z_1) constants $\kappa > 1$ and $A \ge 2$,

$$\prod_{z_1 \le p < z} \left(1 - \frac{\omega(p)}{p}\right)^{-1} \le \left(\frac{\log z}{\log z_1}\right)^{\kappa} \left(1 + \frac{A}{\log z_1}\right), \quad \text{for } 2 \le z_1 < z. \tag{5.4}$$

Write

$$R_d = |\mathscr{A}_d| - \frac{\omega(d)}{d} X.$$

The quantity $\frac{\omega(d)}{d} X$ is considered as an approximation to $|\mathscr{A}_d|$, and therefore we suppose that the errors R_d are small on average, in the sense that for some constants τ with $0 < \tau < 1$, $A_1 \ge 1$, and $A_2 \ge 2$,

$$\sum_{\substack{d < X^{\tau} \log^{-A_1} X \\ (d,B)=1}} \mu^2(d) 4^{\nu(d)} |R_d| \le A_2 \frac{X}{\log^{\kappa+1} X}, \tag{5.5}$$

where $\nu(d)$ denotes the number of prime factors of d. Finally we introduce a constant μ such that

$$\max_{a_n \in \mathscr{A}} n \le X^{\tau\mu} \tag{5.6}$$

The following two lemmas are essentially Theorems 0 and 1 in Diamond-Halberstam [7].

Lemma 5.1. *Let $\kappa > 1$ be given, and let $\sigma_\kappa(u)$ be the continuous solution of the differential-difference problem*

$$\begin{cases} u^{-\kappa}\sigma(u) = A_\kappa^{-1}, & \text{for } 0 < u \le 2,\ A_\kappa = (2e^{\gamma})^{\kappa}\Gamma(\kappa+1), \\ (u^{-\kappa}\sigma(u))' = -\kappa u^{-\kappa-1}\sigma(u-2), & \text{for } 2 < u; \end{cases} \tag{5.7}$$

here γ denotes the Euler constant. Then there exist two numbers α_κ and β_κ satisfying

$$\alpha_\kappa \ge \beta_\kappa \ge 2 \tag{5.8}$$

such that the simultaneous differential-difference system

$$\begin{cases} F(u) = 1/\sigma_\kappa(u), & \text{for } 0 < u \le \alpha_\kappa, \\ f(u) = 0, & \text{for } 0 < u \le \beta_\kappa, \\ (u^\kappa F(u))' = \kappa u^{\kappa-1} f(u-1), & \text{for } u > \alpha_\kappa, \\ (u^\kappa f(u))' = \kappa u^{\kappa-1} F(u-1), & \text{for } u > \beta_\kappa, \end{cases} \tag{5.9}$$

has continuous solutions $F_\kappa(u)$ and $f_\kappa(u)$ with the properties that

$$F_\kappa(u) = 1 + O(e^{-u}), \quad f_\kappa(u) = 1 + O(e^{-u}), \tag{5.10}$$

and that $F_\kappa(u)$ and $f_\kappa(u)$, respectively, decreases and increases monotonically towards 1 as $u \to \infty$.

Lemma 5.2. *Let $\mathscr{A}$ and B be described as above. Then, for any two real numbers u and v satisfying*

$$\frac{1}{\tau} < u \le v, \quad \beta_\kappa < \tau v, \tag{5.11}$$

we have

$$\sum_{n \in P_r(B)} a_n \gg X \prod_{p < X^{1/v}} \left(1 - \frac{\omega(p)}{p}\right) \tag{5.12}$$

provided only that

$$r > \tau\mu u - 1 + \frac{\kappa}{f_\kappa(\tau v)} \int_1^{v/u} F_\kappa(\tau v - s) \left(1 - \frac{u}{v} s\right) \frac{ds}{s}. \tag{5.13}$$

6. Proof of Theorem 2.2 via a weighted three dimensional sieve

We are ready to establish Theorem 2.2.

Proof of Theorem 2.2. Under the assumption of Theorem 2.2, we may specify in Section 5

$$\mathscr{A} = \{a_n(T) : a_n(T) \text{ are defined as in (2.4)}\}. \tag{6.1}$$

With X defined as in (2.5), we have

$$|\mathscr{A}| = \sum_{n \ge 1} a_n(T) = X.$$

Let $B(\mathscr{O})$ be the set of bad primes for $\mathscr{O}$. Then the discussion in Section 5 states that

$$B(\mathscr{O}) \subset \{2, 3, 5, 7\}.$$

For these $\mathscr{A}$ and $B(\mathscr{O})$, (5.2) takes the form

$$\mathscr{A}_d = \{a_n(T) \in \mathscr{A} : n \equiv 0 (\mathrm{mod} d)\},$$

where d is square-free and $(d, B(\mathscr{O})) = 1$. Following (4.8), we define, for square-free d,

$$\omega(d) = \begin{cases} d\frac{|\mathscr{O}^0(\mathbb{Z}/d\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/d\mathbb{Z})|}, & \text{if } (d, B(\mathscr{O})) = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

It follows from the argument before (4.8) that $\omega(d)$ is multiplicative, and (5.3) holds. To verify (5.4), we recall that $td(f)$ is square-free, and therefore for $p \notin B(\mathscr{O})$ there are three cases:

(i) $p \nmid d(f)$ and $p|t$;

(ii) $p|d(f)$ and $p \nmid t$;

(iii) $p \nmid d(f)$ and $p \nmid t$.

Consequently,

$$\prod_{z_1 \le p < z} \left(1 - \frac{\omega(p)}{p}\right)^{-1} = \varpi_1 \varpi_2 \varpi_3,$$

where $\varpi_1, \varpi_2, \varpi_3$ denote, respectively, products over primes satisfying the above three conditions (i), (ii), and (iii). If p satisfies the first condition, then by (4.17) and (4.22),

$$\frac{\omega(p)}{p} = \frac{|\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/p\mathbb{Z})|} \le \frac{6p}{p^2 - 1},$$

and consequently

$$\left(1 - \frac{\omega(p)}{p}\right)^{-1} \le 1 + \frac{100}{p}. \tag{6.3}$$

It follows that

$$\begin{aligned} \varpi_1 &= \prod_{\substack{z_1 \le p < z \\ p \nmid d(f), p|t}} \left(1 - \frac{\omega(p)}{p}\right)^{-1} \le \prod_{\substack{p \ge z_1 \\ p|t}} \left(1 + \frac{100}{p}\right) \\ &\le 1 + 100^{\nu(t)} \sum_{\substack{d \ge z_1 \\ d|t}} \frac{1}{d} \le 1 + \frac{200^{\nu(t)}}{z_1}. \end{aligned}$$

In the second case, we apply (4.31), to get

$$\frac{\omega(p)}{p} = \frac{|\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/p\mathbb{Z})|} \le \frac{8}{p-1},$$

and therefore (6.3) still holds in case (ii). Hence, similarly

$$\varpi_2 \le 1 + \frac{200^{\nu(t)}}{z_1}.$$

In the third case, (4.9), (7.7) and (7.10) are applicable, giving

$$\frac{\omega(p)}{p} = \frac{|\mathscr{O}^0(\mathbb{Z}/p\mathbb{Z})|}{|\mathscr{O}(\mathbb{Z}/p\mathbb{Z})|} = \frac{|V^0(\mathbb{Z}/p\mathbb{Z})|}{|V(\mathbb{Z}/p\mathbb{Z})|} = \frac{3}{p} + O\left(\frac{1}{p^2}\right).$$

Consequently, by the Mertens theorem, there is an absolute constant $c_4 \ge 1$ such that

$$\begin{aligned} \varpi_3 &\le \prod_{\substack{z_1 \le p < z \\ p \nmid B(\mathscr{O})}} \left(1 - \frac{3}{p} + O\left(\frac{1}{p^2}\right)\right)^{-1} \\ &\le \left(\frac{\log z}{\log z_1}\right)^3 \left(1 + \frac{c_4}{\log z_1}\right), \quad \text{for } 2 \le z_1 < z. \end{aligned}$$

In conclusion,

$$\begin{aligned} \prod_{z_1 \le p < z} \left(1 - \frac{\omega(p)}{p}\right)^{-1} &= \varpi_1 \varpi_2 \varpi_3 \\ &\le \left(\frac{\log z}{\log z_1}\right)^3 \left(1 + \frac{A}{\log z_1}\right), \quad \text{for } 2 \le z_1 < z, \end{aligned} \tag{6.4}$$

where $A = A(t) \ge 2$ is a constant depending on t only. This establishes (5.4) with

$$\kappa = 3. \tag{6.5}$$

A formula like (5.5) is provided by Theorem 2.1; with the definition of $\omega(d)$ in (6.2) and X as in (2.6), Theorem 2.1 states that, for $(d, B(\mathscr{O})) = 1$,

$$|\mathscr{A}_d| = \sum_{n \equiv 0 (\bmod d)} a_n(T) = \frac{\omega(d)}{d} X + R_d(T) \tag{6.6}$$

with

$$R_d(T) \ll_\varepsilon d^{1+\varepsilon} T^{1/2+\theta}. \tag{6.7}$$

If we write

$$D = T^{1/4 - \theta/2 - \varepsilon}, \tag{6.8}$$

then (6.7) yields

$$\sum_{\substack{d\le D\\(d,B(\mathscr{O}))=1}} \mu^2(d)4^{\nu(d)}|R_d(T)| \ll T^{1-\varepsilon}. \tag{6.9}$$

Recall $X \asymp T$; then (6.8) and (6.9) establish (5.5) with

$$\tau = \frac{1}{4} - \theta. \tag{6.10}$$

Also, since $1 \le n \ll T^3$ for $a_n \in \mathscr{A}$, (5.6) holds with

$$\tau\mu = 3. \tag{6.11}$$

This finishes checking the requirements of Lemmas 5.1 and 5.2, and therefore we have

$$\sum_{n\in P_r(B(\mathscr{O}))} a_n(T) \gg X \prod_{p<X^{1/v}} \left(1-\frac{\omega(p)}{p}\right) \tag{6.12}$$

provided that

$$r > \tau\mu u - 1 + \frac{3}{f_3(\tau v)}\int_1^{v/u} F_3(\tau v - s)\left(1-\frac{u}{v}s\right)\frac{ds}{s}. \tag{6.13}$$

Our aim is to find the smallest r satisfying (6.13).

Although it is difficult to compute $F_\kappa(u)$ or $f_\kappa(u)$ by hand, the following estimate is quite effective in practice: For general $0 < \tau \le 1$ and $\kappa > 1$, and any $0 < \zeta < \beta_\kappa$, put

$$\tau u = 1 + \zeta - \frac{\zeta}{\beta_\kappa}, \quad \tau v = \frac{\beta_\kappa}{\zeta} + \beta_\kappa - 1. \tag{6.14}$$

Then

$$\frac{\kappa}{f_\kappa(\tau v)}\int_1^{v/u} F_\kappa(\tau v - s)\left(1-\frac{u}{v}s\right)\frac{ds}{s} \le (\kappa+\zeta)\log\frac{\beta_\kappa}{\zeta} - \kappa + \zeta\frac{\kappa}{\beta_\kappa}. \tag{6.15}$$

This follows from (10.1.10), (10.2.4), and (10.2.7) in Halberstam-Richert [14].

Now we insert (6.14) and (6.15) into (6.13), and then specify $\kappa = 3$, to get

$$r > (1+\zeta)\mu - 1 + (3+\zeta)\log\frac{\beta_3}{\zeta} - 3 - \zeta\frac{\mu-3}{\beta_3} =: m(\zeta). \tag{6.16}$$

Note that

$$\beta_3 = 6.6408 \tag{6.17}$$

by Appendix III on p.345 in [7]. The minimum of the function $m(\zeta)$ is easily determined by hand or by Mathematica. It turns out that, for $\theta = 7/64$ (i.e. $\tau = 1/4 - 7/128$),

$$\min_{0<\zeta<\beta_3} m(\zeta) = m(0.1866...) = 25.2615...; \tag{6.18}$$

and for $\theta = 0$ (i.e. $\tau = 1/4$),

$$\min_{0<\zeta<\beta_3} m(\zeta) = m(0.2306...) = 21.3105.... \tag{6.19}$$

Therefore, $r = 26$ is acceptable unconditionally, and $r = 22$ under Selberg's Eigenvalue Conjecture.

When $\zeta = 0.1866...$ or $0.2306...$, we can decide the corresponding values of v by the second equation in (6.14), and consequently (6.12) and (6.4) give

$$\sum_{n\in P_r(B(\mathscr{O}))} a_n(T) \gg X \prod_{\substack{p<X^{1/v}\\ p\notin B(\mathscr{O})}} \left(1 - \frac{3}{p} + O\left(\frac{1}{p^2}\right)\right) \gg \frac{X}{\log^3 X}.$$

This completes our proof of Theorem 2.2. □

We end this section with a proof of Corollary 2.3.

Proof of Corollary 2.3. The analogous passage from a lower bound in Theorem 2.2 to the Zariski dense statement, when ordering points on an orbit combinatorially by word length, is described in detail in [3], so we provide only an outline here. $V_{f,t}$ is irreducible as a variety in $\mathbb{A}^3$ and hence if the points produced in Theorem 2.2 are not Zariski dense in $V_{f,t}$, then there is a polynomial $g \in \mathbb{Q}[x_1, x_2, x_3]$ which is non-zero when restricted to V and such that all the points in question lie in $\mathscr{O} \cap \{\mathbf{x} : g(\mathbf{x}) = 0\}$. Let T be our large parameter and choose p a large prime of size T^δ for δ a small constant to be chosen shortly. As in our analysis of $V^0(\mathbb{Z}/p\mathbb{Z})$, we have

$$|V^g(\mathbb{Z}/p\mathbb{Z})| := |\{\mathbf{x} \in V(\mathbb{Z}/p\mathbb{Z}) : g(\mathbf{x}) \equiv 0 (\mathrm{mod} p)\}| \ll p. \tag{6.20}$$

Proceeding as in the derivation of Theorem 2.1 but with $x_1x_2x_3$ replaced by $g(\mathbf{x})$, we get using (6.20) that

$$\sum_{\substack{\mathbf{x}\in\mathscr{O}\\ |\mathbf{x}|\le T\\ g(\mathbf{x})\equiv 0(\bmod p)}} F_T(\mathbf{x}) \ll_\varepsilon \frac{T}{p} + p^{1+\varepsilon}T^{1/2+\theta+\varepsilon}. \tag{6.21}$$

Hence under our assumption that the points produced are in $\{\mathbf{x} : g(\mathbf{x}) = 0\}$, we have

$$\sum_{\substack{\mathbf{x}\in\mathscr{C}\\ |\mathbf{x}|\le T\\ x_1x_2x_3\in P_{26}(B(\mathscr{C}))}} F_T(\mathbf{x}) \le \sum_{\substack{\mathbf{x}\in\mathscr{C}\\ |\mathbf{x}|\le T\\ g(\mathbf{x})=0}} F_T(\mathbf{x}) \le \sum_{\substack{\mathbf{x}\in\mathscr{C}\\ |\mathbf{x}|\le T\\ g(\mathbf{x})\equiv 0(\bmod p)}} F_T(\mathbf{x})$$

$$\ll \frac{T}{p} + p^{1+\varepsilon}T^{1/2+\theta+\varepsilon}. \tag{6.22}$$

Hence choosing δ small enough and positive, the upper bound on the right of (6.22) contradicts the lower bound for the left-hand side of (6.22) that is provided by Theorem 2.2. □

7. Appendix

In this appendix, we use properties of Gauss sums to compute $|V(\mathbb{Z}/p\mathbb{Z})|$ and $|V^0(\mathbb{Z}/p\mathbb{Z})|$, and then decide those primes p such that $|V(\mathbb{Z}/p\mathbb{Z})| = |V^0(\mathbb{Z}/p\mathbb{Z})|$. This information is needed by the sieve of Sections 5 and 6. Here we work in a slightly more general context, that is we do not require $f(x_1, x_2, x_3)$ to be anisotropic, or $t \neq 0$.

Lemma 7.1. *Let $f(x_1, x_2, x_3)$ be an integral quadratic form with square-free determinant $d(f)$, and t an integer. Define $|V(\mathbb{Z}/p\mathbb{Z})|$ and $|V^0(\mathbb{Z}/p\mathbb{Z})|$ as in (1.2) and (1.3). If $|V(\mathbb{Z}/p\mathbb{Z})| = |V^0(\mathbb{Z}/p\mathbb{Z})|$, then $p|(d(f), t)$ or $p = 2, 3, 5$.*

Proof. Without loss of generality, we may suppose

$$f(x_1, x_2, x_3) = ax_1^2 + bx_2^2 + cx_3^2,$$

where a, b, c are non-zero integers, and the determinant $d(f) = abc$ is square-free. Thus, $|V(\mathbb{Z}/p\mathbb{Z})|$ denotes the number of solutions of

$$ax_1^2 + bx_2^2 + cx_3^2 \equiv t(\bmod p), \tag{7.1}$$

and $|V^0(\mathbb{Z}/p\mathbb{Z})|$ the number of solution of (7.1) with $x_1x_2x_3 \equiv 0(\bmod p)$. To determine the bad primes, we compute $|V(\mathbb{Z}/p\mathbb{Z})|$ and $|V^0(\mathbb{Z}/p\mathbb{Z})|$ using Gauss sums.

The Gauss sum we are going to use is defined as

$$S(m, p) = \sum_{x=0}^{p-1} e\left(\frac{mx^2}{p}\right), \tag{7.2}$$

and the sum $S(1, p)$ will play a special role in carrying out the computation below. A classical result states that (see for example Hua [16], Theorems 7.5.4 and 7.5.5)

$$S(1, p) = \sum_{x=1}^{p-1} \left(\frac{x}{p}\right) e\left(\frac{x}{p}\right) = \begin{cases} \sqrt{p}, & \text{if } q \equiv 1(\bmod 4);\\ i\sqrt{p}, & \text{if } q \equiv 3(\bmod 4).\end{cases}$$

It follows from this and $(\frac{-1}{p}) = (-1)^{(p-1)/2}$ that

$$S(1,p)^2 = \left(\frac{-1}{p}\right)p. \tag{7.3}$$

Also if p is odd, and $p \nmid m$, then $S(m,p)$ can be expressed in terms of $S(1,p)$ in the following way:

$$S(m,p) = \left(\frac{m}{p}\right) S(1,p). \tag{7.4}$$

To compute $|V(\mathbb{Z}/p\mathbb{Z})|$ and $|V^0(\mathbb{Z}/p\mathbb{Z})|$, we distinguish between two cases.

Case 1. p is odd, and $p \nmid d(f)$. In view that $d(f) = abc$, we have $p \nmid a, p \nmid b$, and $p \nmid c$.

We have

$$|V(\mathbb{Z}/p\mathbb{Z})| = \frac{1}{p}\sum_{m=0}^{p-1} S(am,p)S(bm,p)S(cm,p)e\left(-\frac{mt}{p}\right). \tag{7.5}$$

In the above sum, we may single out the term with $m = 0$, and then apply (7.4) to the terms with $m \neq 0$, to get

$$|V(\mathbb{Z}/p\mathbb{Z})| = p^2 + \left(\frac{abc}{p}\right)\frac{S(1,p)^3}{p}\sum_{m=1}^{p-1}\left(\frac{m}{p}\right)e\left(-\frac{mt}{p}\right). \tag{7.6}$$

The last sum is 0 if $p|t$; it equals $(\frac{-t}{p})S(1,p)$ if $p \nmid t$. Therefore, by (7.3),

$$|V(\mathbb{Z}/p\mathbb{Z})| = \begin{cases} p^2, & \text{if } p|t; \\ p^2 + \left(\frac{-abct}{p}\right)p, & \text{if } p \nmid t. \end{cases} \tag{7.7}$$

Writing

$$\sigma(m,p) = S(am,p)S(bm,p)S(cm,p) - \{S(am,p)-1\}\{S(bm,p)-1\}\{S(cm,p)-1\},$$

we see that

$$|V^0(\mathbb{Z}/p\mathbb{Z})| = \frac{1}{p}\sum_{m=0}^{p-1}\sigma(m,p)e\left(-\frac{mt}{p}\right). \tag{7.8}$$

Expanding $\sigma(m,p)$ in (7.8),

$$\begin{aligned} |V^0(\mathbb{Z}/p\mathbb{Z})| &= Y(p;a,b) + Y(p;b,c) + Y(p;c,a) \\ &\quad -Z(p;a) - Z(p;b) - Z(p;c) + \frac{1}{p}\sum_{m=0}^{p-1} e\left(-\frac{mt}{p}\right), \end{aligned} \tag{7.9}$$

where

$$Y(p;a,b)=\frac{1}{p}\sum_{m=0}^{p-1}S(am,p)S(bm,p)e\left(-\frac{mt}{p}\right),\quad Z(p;a)=\frac{1}{p}\sum_{m=0}^{p-1}S(am,p)e\left(-\frac{mt}{p}\right).$$

Now (7.4) and (7.3) give

$$\begin{aligned}
Y(p;a,b) &= p+\frac{1}{p}\left(\frac{ab}{p}\right)S(1,p)^2\sum_{m=1}^{p-1}e\left(-\frac{mt}{p}\right)\\
&= \begin{cases} p+\frac{1}{p}\left(\frac{ab}{p}\right)S(1,p)^2(p-1), & \text{if } p|t\\ p-\frac{1}{p}\left(\frac{ab}{p}\right)S(1,p)^2, & \text{if } p\nmid t\end{cases}\\
&= \begin{cases} p+\left(\frac{-ab}{p}\right)(p-1), & \text{if } p|t;\\ p-\left(\frac{-ab}{p}\right), & \text{if } p\nmid t.\end{cases}
\end{aligned}$$

Similarly,

$$\begin{aligned}
Z(p;a) &= 1+\frac{1}{p}\left(\frac{a}{p}\right)S(1,p)\sum_{m=1}^{p-1}\left(\frac{m}{p}\right)e\left(-\frac{mt}{p}\right)\\
&= \begin{cases} 1, & \text{if } p|t;\\ 1+\frac{1}{p}\left(\frac{-at}{p}\right)S(1,p)^2, & \text{if } p\nmid t.\end{cases}\\
&= \begin{cases} 1, & \text{if } p|t;\\ 1+\left(\frac{at}{p}\right), & \text{if } p\nmid t.\end{cases}
\end{aligned}$$

The last term in (7.9) is $\delta_{p|t}$, the characteristic function of $p|t$. Inserting these into (7.9),

$$\begin{aligned}
&|V^0(\mathbb{Z}/p\mathbb{Z})|\\
&=\begin{cases} 3p-2+\{(\frac{-ab}{p})+(\frac{-bc}{p})+(\frac{-ca}{p})\}(p-1), & \text{if } p|t;\\ 3p-3-\{(\frac{-ab}{p})+(\frac{-bc}{p})+(\frac{-ca}{p})+(\frac{at}{p})+(\frac{bt}{p})+(\frac{ct}{p})\}, & \text{if } p\nmid t.\end{cases}
\end{aligned}\tag{7.10}$$

To see that 3 and 5 are the only possible bad odd primes, we need to show that $|V(\mathbb{Z}/p\mathbb{Z})|-|V^0(\mathbb{Z}/p\mathbb{Z})|>0$ for all $p\geq 7$. To this end, we treat $p|t$ and $p\nmid t$ separately. If $p|t$, then it follows from (7.7) and (7.10) that $|V^0(\mathbb{Z}/p\mathbb{Z})|\leq 6p-5$, and consequently

$$|V(\mathbb{Z}/p\mathbb{Z})|-|V^0(\mathbb{Z}/p\mathbb{Z})|\geq p^2-(6p-5)>0,\qquad \text{for } p\geq 7.$$

Now suppose $p \nmid t$. Then (7.7) and (7.10) give $|V(\mathbb{Z}/p\mathbb{Z})| \geq p^2 - p$ and $|V^0(\mathbb{Z}/p\mathbb{Z})| \leq 3p + 3$, and therefore,

$$|V(\mathbb{Z}/p\mathbb{Z})| - |V^0(\mathbb{Z}/p\mathbb{Z})| \geq p^2 - 4p - 3 > 0, \quad \text{for } p \geq 7.$$

This proves the Lemma in the case of $p \nmid abc$.

Case 2. p is odd, and $p|d(f)$. In view that $d(f) = abc$ is square-free, we may suppose that $p|a, p \nmid b$, and $p \nmid c$.

Since $p|a$, we have $S(am, p) = p$ for all m, and hence by (7.5),

$$\begin{aligned} |V(\mathbb{Z}/p\mathbb{Z})| &= \frac{1}{p}\sum_{m=0}^{p-1} pS(bm,p)S(cm,p)e\left(-\frac{mt}{p}\right) \\ &= p^2 + \left(\frac{bc}{p}\right)S(1,p)^2\sum_{m=1}^{p-1} e\left(-\frac{mt}{p}\right) \\ &= \begin{cases} p^2 + (\frac{bc}{p})S(1,p)^2(p-1), & \text{if } p|t \\ p^2 - (\frac{bc}{p})S(1,p)^2, & \text{if } p \nmid t \end{cases} \\ &= \begin{cases} p^2 + (\frac{-bc}{p})p(p-1), & \text{if } p|t; \\ p^2 - (\frac{-bc}{p})p, & \text{if } p \nmid t. \end{cases} \end{aligned} \tag{7.11}$$

To compute $|V^0(\mathbb{Z}/p\mathbb{Z})|$, we write

$$\sigma(m,p) = pS(bm,p)S(cm,p) - (p-1)\{S(bm,p)-1\}\{S(cm,p)-1\},$$

and hence

$$|V^0(\mathbb{Z}/p\mathbb{Z})| = \frac{1}{p}\sum_{m=0}^{p-1}\sigma(m,p)e\left(-\frac{mt}{p}\right). \tag{7.12}$$

Expanding $\sigma(m,p)$ in (7.12),

$$\begin{aligned} |V^0(\mathbb{Z}/p\mathbb{Z})| &= (p-1)Z(p;b) + (p-1)Z(p;c) + Y(p;b,c) \\ &\quad -\frac{1}{p}\sum_{m=0}^{p-1}(p-1)e\left(-\frac{mt}{p}\right). \end{aligned} \tag{7.13}$$

The last term in (7.13) is $-(p-1)\delta_{p|t}$. Using the results for Y and Z before,

$$\begin{aligned} &|V^0(\mathbb{Z}/p\mathbb{Z})| \\ &= \begin{cases} 2p - 1 + (\frac{-bc}{p})(p-1), & \text{if } p|t; \\ 3p - 2 + \{(\frac{bt}{p}) + (\frac{ct}{p})\}(p-1) - (\frac{-bc}{p}), & \text{if } p \nmid t. \end{cases} \end{aligned} \tag{7.14}$$

To decide the possible bad primes, we still need to consider $p|t$ and $p \nmid t$ separately. If $p|t$, then by (7.11) and (7.14),

$$|V(\mathbb{Z}/p\mathbb{Z})| - |V^0(\mathbb{Z}/p\mathbb{Z})| \geq p^2 - 2p - 1 - (p-1)^2 = -2.$$

That is, each odd prime divisor p of (abc, t) is possibly a bad odd prime. If $p \nmid t$, then

$$\begin{aligned}
&|V(\mathbb{Z}/p\mathbb{Z})| - |V^0(\mathbb{Z}/p\mathbb{Z})| \\
&= p^2 - 3p + 2 - \left\{ \left(\frac{-bc}{p}\right) + \left(\frac{bt}{p}\right) + \left(\frac{ct}{p}\right) \right\} (p-1) \\
&\geq p^2 - 6p + 5 > 0, \qquad \text{for } p \geq 7.
\end{aligned}$$

This establishes Lemma 7.1 in the second case. □

Note that in Lemma 7.1, t may be any integer, and therefore $t = 0$ is allowed. An application of Lemma 7.1 with $t = 0$ gives that the only bad primes for the Pythagorean equation (1.1) are $2, 3, 5$.

References

[1] V. Blomer and J. Brüdern, A three squares theorem with almost primes, Bull. London Math. Soc. 37(2005), 507-513.

[2] M. Borovoi, On representations of integers by indefinite ternary quadratic forms, J. Number Theory 90(2003), 281-293.

[3] J. Bourgain, A. Gamburd, and P. Sarnak, Sieving and expanders, C. R. Math. Acad. Sci. Paris 343(2006), 155-159; An affine linear sieve, expanders and sum product, in preparation.

[4] J. Brüdern and E. Fouvry, Largrange's four squares theorem with almost prime variables, J. Riene Angew. Math. 454(1994), 59-96.

[5] J. W. S. Cassels, *Rational quadratic forms,* London Mathematical Society Monographs 13, Academic Press, London-New York, 1978.

[6] J. Delsarte, Sur le gitter fuchsian, C. R. Math. Acad. Sci. Paris 214(1942), 147-179.

[7] H. Diamond and H. Halberstam, Some applications of sieves of dimension exceeding 1, in *Sieve methods, exponential sums, and their applications in number theory (Cardiff, 1995)*, 101-107, London Math. Soc. Lecture Notes Ser., 237, Cambridge Univ. Press, Cambridge, 1997.

[8] H. Diamond, H. Halberstam, and H.-E. Richert, Combinatorial sieves of dimension exceeding 1, J. Number Theory 28(1988), 306-346.

[9] W. Duke, Hyperbolic distribution problems and half-integral weight Maass forms, Invent. Math. 92(1988), 73-90.

[10] W. Duke, Z. Rudnick, and P. Sarnak, Density of integer points on affine homogeneous varieties, Duke Math. J. 71(1993), 143-179.

[11] A. Gamburd, On spectral gap for infinite index "congruence" subgroups of $SL_2(\mathbb{Z})$, Israel J. Math. 127(2002), 157-200.

[12] A. Good, Local analysis of Selberg's trace formula, Lecture Notes Math. 1040, Springer-Verlag 1983.
[13] I. S. Gradsbteyn and L. M. Ryzbik, Table of integrals, series, and products, 6th edition, Academic Press 2000.
[14] H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press 1974.
[15] Harish-Chandra, Spherical functions on a semisimple Lie group. I, Amer. J. Math. 80(1958), 241-310; II. Amer. J. Math. 80(1958), 553-613.
[16] L. K. Hua, *Introduction to number theory*, Springer 1990.
[17] H. Iwaniec, Fourier coefficients of modular forms of half-integral weight, Invent. Math. 87(1987), 385-401.
[18] H. Jacquet and R. P. Langlands, Automorphic forms on GL(2), Lecture Notes in Mathematics, Vol. 114, Springer-Verlag, Berlin-New York, 1970.
[19] H. Kim and P. Sarnak, Refined estimates towards the Ramanujan and Selberg conjectures, J. Amer. Math. Soc. 16(2003), 175-181.
[20] S. Lang, Algebraic groups over finite fields, Amer. J. Math. 78(1956), 555-563.
[21] P. D. Lax and R. S. Phillips, The asymptotic distribution of lattice points in Euclidean and non-Euclidean space, J. Funct. Anal. 46(1982) 280-350.
[22] A. Nevo and P. Sarnak, Prime and almost prime matrices, in preparation.
[23] P. Sarnak and X. Xue, Bounds for multiplicities of automorphic representations, Duke Math. J. 64 (1991), 207-227.
[24] A. Selberg, On the estimation of Fourier coefficients of modular forms, Proc. Sympos. Pure Math., Vol. VIII, pp. 1-15, Amer. Math. Soc., Providence, R. I. 1965.
[25] A. Selberg, Harmonic analysis on weakly symmetric Riemannian spaces with applications to Dirichlet series, J. Indian Math. Soc. 20(1956), 47-87.
[26] C. L. Siegel, *Lectures on the analytic theory of quadratic forms*, third edition, Robert Peppermuller, Göttingen 1963.