A SPARSE EQUIDISTRIBUTION RESULT FOR $(\mathrm{SL}(2,\mathbb{R})/\Gamma_0)^n$

PANKAJ VISHE

Abstract. Let $G = \mathrm{SL}(2,\mathbb{R})^n$, let $\Gamma = \Gamma_0^n$, where Γ_0 is a co-compact lattice in $\mathrm{SL}(2,\mathbb{R})$, let $F(\mathbf{x})$ be a non-singular quadratic form and let $u(x_1,...,x_n) := \left(\begin{smallmatrix} 1 & x_1 \\ 0 & 1 \end{smallmatrix}\right) \times ... \times \left(\begin{smallmatrix} 1 & x_n \\ 0 & 1 \end{smallmatrix}\right)$ denote unipotent elements in G which generate an n dimensional horospherical subgroup. We prove that in the absence of any local obstructions for F, given any $x_0 \in G/\Gamma$, the sparse subset $\{u(\mathbf{x})x_0 : \mathbf{x} \in \mathbb{Z}^n, F(\mathbf{x}) = 0\}$ equidistributes in G/Γ as long as $n \geq 481$, independent of the spectral gap of Γ_0.

1. Introduction

Let G be a Lie group, let Γ be a lattice in G and let $M = G/\Gamma$. Let U be a unipotent subgroup of G. Recently there has been an increased interest in understanding the behaviour of sparse arithmetic subsets of U orbits in M. It is widely considered that under some reasonable assumptions on G and Γ, certain discrete arithmetic subsets of dense unipotent orbits should equidistribute in M, independent of the choice of the starting points of these orbits. These kind of questions falls a category called *discrete analogues* for mixing systems, a term coined after some of the early works of Stein and Waigner [17], [18] among others. Let $\Gamma_0 \subset G_0 = \mathrm{SL}(2,\mathbb{R})$ be a co-compact lattice, let $M_0 = G_0/\Gamma_0$ and given $x \in \mathbb{R}$, let

$$u_0(x) := \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}$$

be the matrices generating a one-parameter horocycle flow on G_0. In this case, a conjecture by Shah [16], later generalised by Margulis [13] predicts that ergodic averages of these unipotent trajectories evaluated at polynomial times equidistribute irrespective of their starting points. Namely, that the set $\{u_0(F(n))x_0 : n \in \mathbb{N}\}$ equidistributes in M_0 for any polynomial F and any $x_0 \in M_0$. This conjecture remains completely open even in the case $F(x) = x^2$. There have been several works which establish results of *metric* nature. Namely, they bound the size of the set of initial points which violate this expectation, as seen by works of Bourgain [2], Ubis and Sarnak [15] and Katz [9] among others. Apart from works of Shah [16], Venkatesh [21], Tanis and the author [19] and Flaminio, Forni and Tanis [4], there haven't been many results available which establish such sparse equidistribution results for *every* such orbit. Some of the newer results in similar directions include those of Katz [10] (on effective sparse equidistribution of horocycle orbits at times lying inside annuli) and that of McAdam [14] (on horocycle flow at almost prime times).

There are several related generalisations that have been studied. Given n commuting ergodic invertible measure preserving one parameter flows $U_1, U_2, \dots, U_n$ on a probability space (X,μ), let $U(\mathbf{t}) := (U_1(t_1), ..., U_n(t_n))$ denote the corresponding flow on the product space $X \times ... \times X$. Jones [8] considered ergodic averages of functions on the expanding spherical sub-orbits $U(B_r)x_0$, where B_r denotes the n dimensional sphere $x_1^2 + ... + x_n^2 = r$. Namely, given $f \in \mathrm{L}^p(X \times ... \times X)$, [8] considers the spherical averages

$$\int_{\mathbf{t}\in B_r} f(U(\mathbf{t})x_0)d\sigma_{B_r}(\mathbf{t}), \tag{1.1}$$

where σ_{B_r} denotes an appropriately normalised Haar probability measure on B_r. Jones proved that the spherical average in (1.1) tends to the spatial average of f as $r \to \infty$ for *almost every* x_0, as long as $n \geq 3$ and $p \geq n/n-1$. Moreover, Jones in [8, Theorem 3.1] also attempts a discrete version of this problem. He was able to prove the equidistribution of discrete averages of integer points lying on the annuli

$$\lim_{k\to\infty} \#\{\mathbf{t} \in \mathbb{Z}^n, r_k \leq |\mathbf{t}| \leq r_k + d_k\}^{-1} \sum_{\substack{\mathbf{t}\in\mathbb{Z}^n \\ r_k - d_k \leq |\mathbf{t}| \leq r_k + d_k}} f(U(\mathbf{t})x_0),$$

for almost every x_0, where $d_k r_k \to \infty$ and the thickness d_k is bounded, as long as $n \geq 5$ and $p \geq n/n-1$. This result thus just falls short of being able to cope with the natural discrete analogue: $\{U(\mathbf{x})x_0 : \mathbf{x} \in \mathbb{Z}^n \cap B_r\}$ as $r \to \infty$. As in the case of $\mathrm{SL}(2,\mathbb{R})/\Gamma_0$, it is believed that under suitable conditions on $X = G/\Gamma$, these sparse arithmetic averages of U orbits should equidistribute for every such orbit. In the special case when $X = \mathbb{R}/\mathbb{Z}$ and $U_i(t)(x) := tx \bmod 1$, Magyar in [12], proves the equidistribution the sparse set

$$\{U(\mathbf{t})x_0 \bmod 1 : x_0 \in (\mathbb{R}/\mathbb{Z})^n, \mathbf{t} \in \mathbb{Z}^n \cap \{F(\mathbf{t}) = \lambda\} \cap (-P,P)^n\},$$

where F is a polynomial with a non-singular leading degree form, as $P \to \infty$ for every *Diophantine* initial point x_0 as long as $n \gg_{F,\lambda} 1$.

We now state the context in this paper. As before, let $\Gamma_0 \subset G_0 = \mathrm{SL}(2,\mathbb{R})$ be a co-compact lattice. Let $G = \mathrm{SL}(2,\mathbb{R})^n$, let $\Gamma = \Gamma_0 \times \cdots \times \Gamma_0$, let $M = G/\Gamma$ and let $d\mu_G = d\mu_{G_0} \times ... \times d\mu_{G_0}$, where $d\mu_{G_0}$ denotes the Haar measure on M_0 normalised such that $\int_{M_0} d\mu_{G_0}(g) = 1$. Let

$$U := \{u(\mathbf{x}) := u_0(x_1) \times ... \times u_0(x_n),\ x_1, ..., x_n \in \mathbb{R}\}$$

denote the expanding horospherical subgroup corresponding to the action of a suitable ray of a standard one parameter geodesic flow. The equidistribution of the whole U orbit for every $x_0 \in M$ follows from Ratner's equidistribution theorems, and in this particular setting by some earlier works of Furstenberg ($n = 1$ case) and more generally Veech [22]. In the vein of the aforementioned conjectures by Shah and Margulis, a question by Lindenstrauss on spherical horospheric averages led Ubis [20] to investigate analogues of (1.1) in this particular setting. Ubis establishes the equidistribution of orbits of the type $U(V)x_0$, where V is any totally curved sub-manifold of $\mathbb{R}^n$ of *low* co-dimension, as long as n is *large enough* depending on the spectral gap of Γ_0 as well as the co-dimension of the manifold. He achieves this by locally approximating pieces of such a manifold by quadratic hypersurfaces and then proving the equidistribution of *every* U-orbit restricted to a quadric hypersurface in $\mathbb{R}^n$.

Here, we work on a natural generalisation of the works of Magyar [12] and Ubis [20]. In particular, we consider the sparse subsets of integer points in U orbits in M, which lie on a quadratic hypersurface in $\mathbb{R}^n$. Let $F(\mathbf{x}) = \mathbf{x}^t L\mathbf{x} \in \mathbb{Z}[\mathbf{x}]$ be a smooth quadratic form in n variables defined by an invertible $n \times n$ matrix L with integer entries. We further assume that F has no local obstructions, i.e. that $F(\mathbf{x}) = 0$ for some $\mathbf{x} \in \mathbb{R}^n \setminus \mathbf{0}$ as well as for some $\mathbf{x} \in \mathbb{Q}_p^n \setminus \mathbf{0}$ for each prime p. Given any parameter $P > 0$, a standard circle method result (see [1, Theorem 1] for example) hands us a constant $0 < \gamma' \ll 1$ such that the following asymptotic formula holds as long as $n \geq 5$:

$$N_F(P) := \#\{\mathbf{x} \in \mathbb{Z}^n : |\mathbf{x}| < P, F(\mathbf{x}) = 0\} = C_F P^{n-2} + O(P^{n-2-\gamma'}). \tag{1.2}$$

The implied constant $C_F > 0$ if and only if F has no local obstructions. Here and throughout, we use the notation $A \ll B$ to denote that $A \leq CB$, for some constant C. Throughout, our implied constants in $\ll$ as well as in $O(\cdot)$ notation are allowed to depend freely on Γ, n and F. Any further dependence will be explicitly denoted via adding a subscript to the corresponding expression.

Our main goal is to prove the following sparse equidistribution/mixing result:

Theorem 1.1. *Let $G_0 = \mathrm{SL}(2,\mathbb{R})$, let Γ_0 be a co-compact lattice in G_0, let $G = \mathrm{SL}(2,\mathbb{R})^n$ and let $\Gamma = \Gamma_0^n$. Then for any non-singular quadratic form $F(\mathbf{x}) \in \mathbb{Z}[x_1, ..., x_n]$ with no local obstructions, any point $x_0 \in G/\Gamma$ and any continuous function $f \in \mathrm{C}(G/\Gamma)$, we have*

$$\lim_{P\to\infty} \frac{1}{N_F(P)} \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n, |\mathbf{x}|<P \\ F(\mathbf{x})=0}} f(u(\mathbf{x})x_0) = \int f(x) d\mu_G(x),$$

as long as $n \geq 481$.

There are two main highlights of this result. Firstly, the equidistribution is established for *every* $x_0 \in G/\Gamma$. Secondly, our bound for n is independent of the spectral gap of G_0/Γ_0. We have also not tried to optimise the lower bound 481 appearing here. An effective version of this result will appear later in Theorem 1.2, however one should note that as expected, the rate of this equidistribution which corresponds to the exponent γ_0 in Theorem 1.2 does depend on the spectral gap of the lattice Γ even though n itself does not. Throughout, we have used results of Flaminio-Forni and Tanis to bound the twisted averages along horocycles. However, one may perfectly use [19, Theorem 1.2] which use much softer techniques rather than [4]. These will provide us with a bound on n independent on the spectral gap, although a very rough calculation shows that this way will lead to a lower bound at worst of size $n \geq 811$ or so.

It is likely that our bounds may be improved to be able to obtain a better result. When the spectral gap of M_0 μ is around the size $1/4$, then using the main results in [4] or [19] is wasteful. One may be able to get a much improved bounds in this case. A natural limit of the process here for in the general case (any $0 < \mu \leq 1/4$) would be $n \geq 457 = 2 \times 228 + 1$, which arises from the second term in our van der Corput bound (3.21). One may be able to get obtain an analogue of Theorem 1.1 for any lattice Γ in G, as long as one is allowed dependence on spectral gaps of irreducible components of G/Γ. However, in the current form it is crucial for us that Γ is completely factorisable of the form $\Gamma_1 \times ... \times \Gamma_n$.

We must highlight that the situation considered here is significantly different than that of $(\mathbb{R}/\mathbb{Z})^n$ or the usual circle method setting which leads to the asymptotic formula in (1.2). Here, we need to consider exponential sums of the type

$$S(\alpha) := \sum_{\mathbf{x}\in\mathbb{Z}^n} w(\mathbf{x}/P) f(u(\mathbf{x})x_0) e(\alpha F(\mathbf{x})), \tag{1.3}$$

where w is a suitable compactly supported function on $\mathbb{R}^n$ and α is a real number. Notice that the extra factor $f(u(\mathbf{x})x_0)$ appearing here is highly oscillatory, and due to this, the usual analytic techniques break down. We therefore need to lower the degree of F using some sort of differencing, and then use bounds for twisted averages of functions along horocycles. One way to do so is to use van der Corput differencing, which hands us exponential integrals of differenced functions (see (3.26)); which we follow up by applying uniform bounds for twisted horocyclic averages in [4, Theorem 1,1], which require taking large, in fact $7+\varepsilon$, number of derivatives of these functions. The final optimisation therefore amounts to a loss by a factor of size $14+\varepsilon$. However, unfortunately, this bound is not enough and we need to invent a new technique, namely, our alternate bound in Lemma 3.1. For further explanation of this technique, we refer the reader to the explanation given following the statement of Theorem 1.2. As far as our knowledge, the differencing technique used to obtain Lemma 3.1 has not yet been used in this setting before. It would be interesting if this bound can be modified to be made to work in the whole *minor arcs* regime. In this case, one may be able to obtain the result using much lower number of variables. However, it should be noted that using analytic methods, one may not expect a result as good as $n \geq 5$. Unless one improves upon the work of Flaminio, Forni and Tanis [4], one would at least require $2\times 6+1 = 13$ variables (or more realistically $2 \times 13 = 26$ variables to allow for a typical loss arising due to differencing).

Theorem 1.1 can also be seen as a *mixing* type result. If F were a diagonal form instead, then as noted by Ubis in a private communication, a simple Hölder inequality type argument applied to the exponential sums in the spirit of techniques used in ternary Goldbach conjecture (see [7, Lemma 19.4]) can directly establish Theorem 1.1 as soon as $n \geq 5$. A sketch of this argument will be produced in Remark 4.2. The assumption that Γ_0 is co-compact could also be removed with some more technical work, using finer results in [4, Theorem 1.1]. We believe that the method in this paper can be suitably modified to obtain a version of Ubis' result [20] independent of the spectral gap. In this case, possibly a variant of Lemma 3.1 itself could be made to work which may lead to requiring a relatively few number of variables.

The strategy used in this paper is rather *soft* and is capable of establishing a much more general result than the one stated here. For example, let X be a probability space, and let U be an n-dimensional measure preserving flow on X, then techniques here may be used to establish the equidistribution of discrete sparse subsets $\{U(\mathbf{x})x_0 : \mathbf{x} \in \mathbb{Z}^n \cap \{F(\mathbf{x}) = 0\}\}$ (or of its continuous version a.k.a. [20]), for every x_0, as long as one has an effective bound for the twisted averages of the flow U on X and that the dimension n of the flow is *large enough*. One relevant application could be to the case where U denotes a full dimensional horospheric flow on $X = \mathrm{SL}(n,\mathbb{R})/\mathrm{SL}(n,\mathbb{Z})$ as long as n is large enough with respect to the degree d of the polynomial F.

We now move on to the statement of Theorem 1.2, our main tool in proving Theorem 1.1. Let x_0 be an arbitrary fixed point in G/Γ. In order to use Fourier analytic tools effectively, given any parameter $P \geq 1$, given any $f \in \mathrm{C}^\infty(M)$ and any compactly supported function $w \in \mathrm{C}_c^\infty(\mathbb{R}^n)$, we will consider the following smooth average

$$\Sigma(P) = \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n \\ F(\mathbf{x})=0}} w(\mathbf{x}/P) f(u(\mathbf{x})x_0). \tag{1.4}$$

Our main tool for proving Theorem 1.1 will be provided by Theorem 1.2 below. It establishes an effective bound for the smooth sum $\Sigma(P)$ for any smooth factorisable function $f \in \mathrm{C}^\infty(G/\Gamma)$ of zero average and a suitably chosen factorisable function $w \in \mathrm{C}_c^\infty((-1,1)^n)$. More explicitly, in Theorem 1.2, we assume that f is of the form

$$f(g_1,...,g_n) = \prod_{i=1}^{n} f_i(g_i), \text{ where } f_i \in \mathrm{C}^\infty(G_0/\Gamma_0) \text{ and } \int_{M_0} f_1(g_1)d\mu_{G_0}(g_1) = 0. \tag{1.5}$$

Here, $g_i \in G_0/\Gamma_0$.

Similarly, we will work with factorisable functions on $\mathbb{R}^n$. Let $\omega \in \mathrm{C}_c^\infty(\mathbb{R})$ be a smooth compactly supported function on $\mathbb{R}$, whose support is contained in $(-1,1)$ and let

$$w(\mathbf{x}) := \prod_{i=1}^{n} \omega(x_i). \tag{1.6}$$

The implied constants in our final bounds may depend on the measure of the support of ω. The fact that ω is supported in $(-1,1)$ is only assumed to simplify this dependence in Theorem 1.2.

Before we give the statement of Theorem 1.2, we must set some notation for various Sobolev norms appearing there. For any function $w(\mathbf{x})$ in $\mathrm{C}^\infty(\mathbb{R}^n)$, any $k \in \mathbb{Z}_{\geq 0}$ and any real number $p \in [1,+\infty]$, we introduce the standard Sobolev norm $S_{p,k}(w)$, taking values in $\widehat{\mathbb{R}}^2_{\geq 0} \cup \{+\infty\}$, through

$$S_{p,k}(w) = \sum_{j=0}^{k} \sum_{|\beta|=j} \|\partial_{\mathbf{x}}^{\beta} w(\mathbf{x})\|_{\mathrm{L}^p}. \tag{1.7}$$

Here, given $\beta = (k_1,...,k_n) \in \mathbb{Z}^n_{\geq 0}$, let $|\beta| = k_1 + ... + k_n$, and let $\partial_{\mathbf{x}}^{\beta} := \partial_{x_1}^{k_1}...\partial_{x_n}^{k_n}$.

Our norms for functions $f \in \mathrm{C}^\infty(G/\Gamma)$ will be analogous and standard. Let $\{Y, X, Z\}$ be a basis for the Lie algebra $\mathfrak{s}l(2,\mathbb{R})$ given by,

$$Y = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \tag{1.8}$$

Given $p \in [1,\infty]$ and $k \in \mathbb{Z}_{\geq 0}$, by $\|f\|_{\mathrm{L}^p_k}$ we denote the sums of L^p norms of upto "k-derivatives" of f. To formalise this, let $\mathscr{O}_k$ be a collection of vectors $D := (D_1,...,D_n)$, where each co-ordinate D_i is a monomial in $\{Y_i, X_i, Z_i\}$ such that the total order of all these monomials is at most k. Here, X_i (and analogously Y_i and Z_i) denotes the element in the lie algebra of G/Γ which contains X in the i-th co-ordinate and zero everywhere else, i.e., $X_i = (0,..,0,X,0,...)$. Then we define

$$\|f\|_{\mathrm{L}^p_k} := \sum_{D=(D_1,...,D_n)\in\mathscr{O}_k} \|Df\|_{\mathrm{L}^p}, \tag{1.9}$$

where $Df := D_1D_2...D_nf$. Upon interpolation as in [11], the above norms can be extended to hold for all $k \in \mathbb{R}_{\geq 0}$.

We are now set to state Theorem 1.2:

Theorem 1.2. *There exists an absolute constant $\gamma_0 := \gamma_0(\Gamma_0)$ depending on the spectral gap of Γ_0 such that given any non-singular quadratic form $F(\mathbf{x}) \in \mathbb{Z}[x_1,...,x_n]$, any $P \geq 1$, given any*

$w(\mathbf{x}) \in \mathrm{C}_c^\infty((-1,1)^n)$ satisfying (1.6) and any $f \in \mathrm{C}^\infty(G/\Gamma)$ satisfying (1.5), we have

$$|\Sigma(P)| \ll S_{\infty,9n}(w)\|f\|_{\mathrm{L}^\infty_{9n+1}} P^{n-2-\gamma_0},$$

as long as $n \geq 481$.

As before, we have not tried to optimise the Sobolev norms as well as the number 481 appearing in Theorem 1.2. It is possible to make γ_0 explicit using the spectral gap μ_0 of M_0. In fact, after following the steps in the proofs here it may be possible to choose any $0 < \gamma_0 \leq \min\{\gamma/624, 1/3744\}$, where

$$\gamma := \frac{1 - \mathrm{Re}\sqrt{1-4\mu}}{2}, \tag{1.10}$$

where μ denotes the spectral gap of M_0, namely, the first non-zero eigenvalue of the Laplace-Beltrami operator on M.

Let us give an overview of the method that will be used to prove Theorem 1.2. The main tool here will be provided by the Hardy-Littlewood circle method. Given any $m \in \mathbb{Z}$, let

$$\int_0^1 e(mz)dz = \begin{cases} 1, & \text{if } m = 0, \\ 0, & \text{otherwise,} \end{cases}$$

denote the delta function detecting when an integer $m = 0$. Here, $e(\alpha) = \exp(2\pi i\alpha)$, a standard notation. Using this, we start by rewriting $\Sigma(P)$ as

$$\Sigma(P) = \int_0^1 S(\alpha)d\alpha, \tag{1.11}$$

where, $S(\alpha)$ is as defined in (1.3) is an exponential sum. Typically, one needs to estimate $S(\alpha)$ at $\alpha = a/q + z$, where $|z| < q^{-2}$. One of our key ingredients in removing the dependence on the spectral gap is provided by the uniform bounds for twisted averages appearing in [4] and [19]. When q is large or when z is very small ($|z| \leq q^{-2}P^{-2+o(1)}$), we use van der Corput differencing to lower the degree of F along with the bounds in [4], which would hand us Lemma 3.2. This bound itself is unfortunately not enough to remove the dependence on the spectral gap when $q, |z|$ are *mid-range*. Here, we use a novel degree lowering technique. Namely, we split the sum over $\mathbf{x}$ in (1.3) as $\mathbf{x} = \mathbf{x}_1 + N\mathbf{x}_2$ where N is approximately of size $|z|^{-1/2}$. This choice means that the term $zF(\mathbf{x}_1)$ is bounded. For a fixed value of $\mathbf{x}_2$, we then consider the sum over $\mathbf{x}_1$. This trick allows us to lower the degree of F in the exponential integral which typically arises after applying Poisson summation. This is the essence of Lemma 3.1.

Let us briefly compare our work with that of Ubis [20]. The key bound in [20] uses van der Corput differencing to bound the exponential integrals, which is analogous to the bound (3.22) of Lemma 3.2 here. Here, we must point out that the hypersurface $F(\mathbf{x}) = 0$ is of co-dimension one, and therefore, the $n-1$ dimensional volume of the set $\{F(\mathbf{x}) = 0 : |\mathbf{x}| < P\} \sim P^{n-1}$. In this paper however, we are averaging over a sparser subset in this manifold as demonstrated by the counting estimate (1.2). This is one philosophical reason behind why we need to establish the bound in Lemma 3.1 and why this problem is significantly harder to tackle.

1.1. **Acknowledgements.** The author had the inspiration for this project while discussing the aforementioned work of Ubis [20] with Kevin Hughes. We are thankful to him for introducing us to this work. It has also been very helpful to get the input of Kevin Hughes, Asaf Katz and Adrián Ubis on an earlier version of this paper. Their contribution is greatly acknowledged. We thank the anonymous referee for their insightful and comprehensive comments and also for suggesting some further projects. The simpler argument in the diagonal case was pointed out to us by Ubis, and we thank him for this as well.

2. Auxiliary results

In this section, we will gather together various auxiliary lemmas necessary for us.

2.1. Bounds for smooth twisted horocyclic averages on $M_0 = \mathrm{SL}(2,\mathbb{R})/\Gamma_0$.

Here, for the sake of avoiding the complication of introducing a separate notation, throughout this section, given $\omega \in \mathrm{C}_c^\infty(\mathbb{R})$, and $f \in \mathrm{C}^\infty(M_0)$, we will use the same notations $S_{p,k}(\omega)$ and $\|f\|_{\mathrm{L}^p_k}$ to denote the corresponding Sobolev norms. These can be seen to be equal to those in (1.7) and (1.9) in the special case when $n = 1$.

The first result to be obtained below is a smooth variant of a twisted averages result [4, Theorem 1.1].

Lemma 2.1. *Let $\omega \in \mathrm{C}_c^\infty(a,b)$ be a smooth, compactly supported function on $\mathbb{R}$, let $f \in \mathrm{C}^\infty(M_0)$ be a function of zero average and let x_0 be any point in M_0. Then given any $P > 1$, any $\varepsilon > 0$, and any $c \in \mathbb{R}$, we have the following bounds for twisted averages along the horocycle flow:*

$$\begin{aligned} &P^{-1}\left|\int \omega(t/P) f(u_0(t)x_0) e(ct)\,dt\right| \\ &\ll_{N,\varepsilon} (1+|b-a|)\log^{1/2}(P) S_{1,1}(\omega)\min\{\|f\|_{\mathrm{L}^2_{7+\varepsilon}}|P|^{-1/6}(1+|c|^{-1/6}), \|f\|_{\mathrm{L}^2_{3+\varepsilon}}|P|^{-\gamma}\}, \end{aligned} \tag{2.1}$$

where γ is as defined in (1.10).

Proof. We begin by applying integration by parts and using the fact that the boundary terms vanish since $\mathrm{supp}(\omega) \subset (a,b)$ to obtain

$$\int \omega(t/P) f(u_0(t)x_0) e(ct)\,dt = P^{-1}\int_{aP}^{bP} \omega'(t/P)\int_{aP}^{t} f(u_0(z)x_0)e(cz)\,dz\,dt.$$

When $|cP| > e$, an application of [4, Theorem 1.1, Equation (5)] to the inner integral on the right hand side of the above equation implies that this term is

$$\begin{aligned} &\ll P^{-1}\log(P)^{1/2}\int_{aP}^{bP} |\omega'(t/P)|(1+|c|^{-1/6})|t-aP|^{5/6}\,dt \\ &\ll \|f\|_{\mathrm{L}^2_{7+\varepsilon}} P^{5/6}\log(P)^{1/2}\int_a^b |\omega'(t)|(1+|c|^{-1/6})|t-a|^{5/6}\,dt \\ &\ll |b-a|^{5/6} S_{1,1}(\omega)\log(P)^{1/2} P^{5/6}\|f\|_{\mathrm{L}^2_{7+\varepsilon}}(1+|c|^{-1/6}). \end{aligned} \tag{2.2}$$

On the other hand, when $|cP| < e$, an application of a weaker bound obtained at the bottom of [4, Page 1361] hands us a constant γ, depending on the spectral gap of M_0, such that

$$|\int_{aP}^{t} f_0(u_0(z)x_0)e(cz)dz| \ll \|f\|_{\mathrm{L}^2_{3+\varepsilon}} |t-aP|^{1-\gamma} \tag{2.3}$$

giving the remaining bound in (2.1), after following the same steps as in the derivation of (2.2) and further noting that $|b-a|^{1-\gamma} + |b-a|^{5/6} \ll 1 + |b-a|$. □

Note that the explicit dependence on $|b-a|$ in Lemma 2.1 is not necessary for our applications. While applying this result, our function ω will be assumed to be supported in an interval of size $\ll 1$. We now focus our attention to estimating averages of smooth twisted averages. It should be noted this when the function f here is a constant function, the lemma below could be related to standard Weyl differencing type estimates arising in the traditional circle method arguments (see [3, Chapters 11 and 12] for example).

Lemma 2.2. *Given any $f \in \mathrm{C}^\infty(M_0)$, any $x_0 \in M_0$, any $1 \le q \le P$, any function $\omega \in \mathrm{C}_c^\infty(a,b)$, any $c \in \mathbb{R}$, and any $\varepsilon > 0$ we have*

$$\begin{aligned}&\sum_{v\in\mathbb{Z}} |\int \omega(y/P) f(u_0(y)x_0)e((c-v/q)y)dy| \\ &\ll (1+|b-a|)\log^{1/2}(P)\|f\|_{\mathrm{L}^2_{9+\varepsilon}} S_{1,3}(\omega) P((1+\|qc\|P/q)^{-1/6} + qP^{-1/6}),\end{aligned} \tag{2.4}$$

where $\|x\|$ denotes the distance of a real number x to the nearest integer.

Moreover, if $\int_{M_0} f(x)d\mu_{G_0}(x) = 0$, then we may further have

$$\sum_{v\in\mathbb{Z}} |\int \omega(y/P) f(u_0(y)x_0)e((c-v/q)y)dy| \ll (1+|b-a|)\log^{1/2}(P)\|f\|_{\mathrm{L}^2_{9+\varepsilon}} S_{1,3}(\omega) P(P^{-\gamma} + qP^{-1/6}), \tag{2.5}$$

where γ is the constant appearing in the statement of Lemma 2.1.

Proof. Let S denote the sum under investigation, that is, let

$$S := \sum_{v\in\mathbb{Z}} \left| \int \omega(y/P) f(u_0(y)x_0)e((c-v/q)y)dy \right|. \tag{2.6}$$

We first begin by considering the special case when $\int_{M_0} f(x)d\mu_{G_0}(x) = 0$, i.e., when f is a zero average function. When $q \le |qc-v|$, we will apply integration by parts twice, followed by Lemma 2.1, while in the range $1/2 \le |qc-v| < q$, Lemma 2.1 will be directly applied. To this end, given any non-negative integer k and any $c_1 \in \mathbb{R}$, integration by parts k times leads us to

$$\begin{aligned}&\left| \int \omega(y/P) f(u_0(y)x_0)e(c_1 y)dy \right| \\ &\ll_k |c_1|^{-k} \sum_{j=0}^{k} P^{-j} \left| \int \omega^{(j)}(y/P)(X^{k-j}f)(u_0(y)x_0)e(c_1 y)dy \right|.\end{aligned} \tag{2.7}$$

Here, X is as in (1.8), acts on f via the explicit action $Xf(x) := \frac{\partial}{\partial_t}|_{t=0} f(u_0(t)x)$. Lemma 2.1 can now be employed to estimate the inner integrals on the right hand of the above expression to obtain

$$\left|\int \omega(y/P) f(u_0(y)x_0) e(c_1 y) dy\right| \ll_{N,\varepsilon,k} (1+|b-a|) P \log^{1/2}(P) |c_1|^{-k} S_{1,k+1}(\omega) \min\{\|f\|_{\mathrm{L}^2_{7+k+\varepsilon}} P^{-1/6}(1+|c_1|^{-1/6}), \|f\|_{\mathrm{L}^2_{3+k+\varepsilon}} |P|^{-\gamma}\}. \tag{2.8}$$

When $|qc-v| \geq q$, we apply (2.8) with $k=2$ and $c_1 = c - v/q$ and when $1/2 \leq |qc-v| < q$, we again apply (2.8) with $k=0$ and $c_1 = c - v/q$ to obtain

$$\begin{aligned}
&((1+|b-a|)P\log^{1/2}(P))^{-1} \sum_{\substack{v\in\mathbb{Z} \\ |qc-v|\geq 1/2}} \left|\int \omega(y/P) f(u_0(y)x_0) e((c-v/q)y) dy\right| \\
&\ll \|f\|_{\mathrm{L}^2_{9+\varepsilon}} S_{1,3}(\omega) P^{-1/6} \left(\sum_{\substack{v\in\mathbb{Z} \\ 1/2\leq|qc-v|<q}} q^{1/6}|qc-v|^{-1/6} + \sum_{\substack{v\in\mathbb{Z} \\ q\leq|qc-v|}} q^2|qc-v|^{-2} \right) \\
&\ll \|f\|_{\mathrm{L}^2_{9+\varepsilon}} S_{1,3}(\omega) P^{-1/6} q. \qquad (2.9)
\end{aligned}$$

Similarly, when $|qc-v| = \|qc\| < 1/2$, we will apply (2.8) with $k=0$ and $c_1 = \|qc\|/q$ to obtain

$$\begin{aligned}
&((1+|b-a|)P\log^{1/2}(P))^{-1} \left|\int \omega(y/P) f(u_0(y)x_0) e(\|qc\| y/q) dy\right| \\
&\qquad \ll S_{1,1}(\omega) \left(\min\{\|f\|_{\mathrm{L}^2_{9+\varepsilon}} |P/q|^{-1/6} \|qc\|^{-1/6}, \|f\|_{\mathrm{L}^2_{3+\varepsilon}} |P|^{-\gamma}\} \right).
\end{aligned} \tag{2.10}$$

Combing (2.9) and (2.10) together, we establish the Lemma when f is of zero average.

When f is not of zero average, we start by writing $f = f_0 + \int_{M_0} f(x) d\mu_{G_0}(x)$, where f_0 is now a function of zero average. Thus,

$$S \leq S_1 + |\int_{M_0} f(x) d\mu_{G_0}(x)| S_2, \tag{2.11}$$

where

$$S_1 := \sum_{v\in\mathbb{Z}} \left|\int \omega(y/P) f_0(u_0(y)x_0) e((c-v/q)y) dy\right| \text{ and } S_2 := \sum_{v\in\mathbb{Z}} \left|\int \omega(y/P) e((c-v/q)y) dy\right|. \tag{2.12}$$

S_1 can be bound by our analysis above. Note that $f = f_0 + \int_{M_0} f(x) d\mu_{G_0}(x)$ is an orthogonal decomposition of f with respect to the L^2 norm, and therefore, for every $k \geq 0$, we must have $\|f_0\|_{\mathrm{L}^2_k} \ll \|f\|_{\mathrm{L}^2_k}$. As a result, S_1 can be bound by

$$S_1 \ll (1+|b-a|) P \log^{1/2}(P) S_{1,3}(\omega) \|f\|_{\mathrm{L}^2_{9+\varepsilon}} ((1+\|qc\|P/q)^{-1/6} + qP^{-1/6}). \tag{2.13}$$

On the other hand, the sum S_2 is simpler and can be bound via direct integration by parts using

$$S_2 = P\sum_{v\in\mathbb{Z}}\left|\int \omega(y)e(P(qc-v)y/q)dy\right|$$
$$\ll PS_{1,2}(\omega)((1+P\|qc\|/q)^{-1}+(P/q)^{-2}\sum_{|qc-v|\geq 1/2}|qc-v|^{-2}) \ll PS_{1,2}(\omega)(1+P\|qc\|/q)^{-1}.$$

Combining this bound with the one in (2.13), and further noticing that $|\int f(x)d\mu_{G_0}(x)| \leq \|f\|_{\mathrm{L}^2_{9+\varepsilon}}$, we establish (2.4). ☐

It should be noted that since M_0 is assumed to be compact, the bounds here are independent of the choice of x_0.

Let $f \in \mathrm{C}^\infty(M_0)$ be a smooth function. Given any $t\in\mathbb{R}$, we will also need bounds for the Sobolev norms of the function $u_0(t)\cdot f(x) := f(u_0(t)x_0)$. In particular, we would like to make the dependence on t more explicit. We recall the explicit action of the basis (1.8) of the Lie algebra in [19, eq (3.1)],

$$\begin{aligned} X(u_0(t)\cdot f) &= u_0(t)\cdot(Xf)\\ Y(u_0(t)\cdot f) &= u_0(t)\cdot((Y+tX)f)\\ Z(u_0(t)\cdot f) &= u_0(t)\cdot((Z-2tY-t^2X)f). \end{aligned} \tag{2.14}$$

Using this explicit action, followed by induction, we are able to prove that for any monomial $X^{i_1}Y^{i_2}Z^{i_3}$, of order $k=i_1+i_2+i_3$, we must have

$$X^{i_1}Y^{i_2}Z^{i_3}(u_0(t)\cdot f) = \sum_{D\in\mathscr{O}_k} p_D(t)D(u_0(t)\cdot f),$$

where p_D are polynomials of degree at most $2k$, with integer coefficients only depending on i_1, i_2 and i_3. Summing over all such monomials, and using the fact that action of $u_0(t)$ preserves the L^2 norms of functions, for any $s\in\mathbb{Z}_{\geq 0}$ we have

$$\|u_0(t)\cdot f\|_{\mathrm{L}^2_s} \ll |t|^{2s}\|f\|_{\mathrm{L}^2_s}. \tag{2.15}$$

Here, as stated at the beginning of this section, for an integer s, L^2_s denotes the Sobolev norm on $\mathrm{C}^\infty(M_0)$, which can be seen as the analogue of (1.9) in the case $n=1$. Upon interpolation, this bound can be extended to be true for all $s\in\mathbb{R}_{\geq 0}$.

2.2. **A lattice sum bound.** In the proof of Lemma 3.1, we will need a bound for the following lattice sum, which we derive next:

Lemma 2.3. *Let L be a fixed invertible $n\times n$ matrix with $\mathbb{Z}$ entries and let $1\leq P, H$ be real numbers satisfying $0\leq H\leq P$. Then, given any $0<|z|<1$, any $0<C$ and any $0<\delta<1$,*

$$\sum_{0\leq y_i\leq P}\prod_{i=1}^n((1+H\|z(L\mathbf{y})_i\|)^{-\delta}+C) \ll_L P^n\prod_{i=1}^n(1/P+|z|+H^{-\delta}+(H|z|P)^{-\delta}+C).$$

Proof. The bound is obvious if $1 \ll |z| \ll 1$. So it is enough to assume that $0 < |z| < 1/2$, say. By changing the variables to $\mathbf{z} = L\mathbf{y}$, it is enough to bound

$$\prod_{i=1}^{n} \sum_{|z_i| \leq |L| P} ((1 + H\|zz_i\|)^{-\delta} + C).$$

To bound the above expression, without loss of generality, we may assume that z is positive. Let N denote the nearest integer to $1/z$, which means $|N - 1/z| \leq 1/2$. Moreover, since $0 < z < 1/2$, $N \geq 2$ and therefore, $|z - 1/N| \leq z/(2N) < 1/N^2$. We now write $z = 1/N + z'$, where $|z'| < 1/N^2$. We begin by noting that for any real number r, and for all but at most one integer x satisfying $|x| < N/2$, we must have

$$\|r + zx\| = \|r + x/N + xz'\| \gg \|r + x/N\|, \tag{2.16}$$

since $|xz'| < 1/(2N)$. Since, L is assumed to be fixed throughout, our constants are free to depend on it, and therefore it is enough to look at

$$\prod_{i=1}^{n} \sum_{-P \leq z_i \leq P} ((1 + H\|zz_i\|)^{-\delta} + C).$$

If $P \geq N/2$, we begin by writing $z_i = z_{i,1} + \lceil N/2 \rceil z_{i,2}$, where $|z_{i,1}| < N/2$. In the light of our observation (2.16), for a fixed i,

$$\begin{aligned}
&\sum_{|z_i| \leq P} ((1 + (H\|zz_i\|))^{-\delta} + C) \ll PC + \sum_{0 \leq |z_{i,2}| \leq P/N} \sum_{0 \leq |z_{i,1}| < N/2} (1 + H\|zz_{i,1} + z\lceil N/2 \rceil z_{i,2}\|)^{-\delta} \\
&\ll PC + \sum_{0 \leq |z_{i,2}| \leq P/N} (1 + \sum_{0 \leq |z_{i,1}| < N/2} (1 + H\|z_{i,1}/N + z\lceil N/2 \rceil z_{i,2}\|)^{-\delta}) \\
&\ll PC + \sum_{0 \leq |z_{i,2}| \leq P/N} (1 + \sum_{0 \leq |z_{i,1}| < N/2} (1 + |Hz_{i,1}/N|)^{-\delta}) \\
&\ll P/N(1 + NH^{-\delta}) + PC \ll P(1/N + H^{-\delta} + C).
\end{aligned}$$

On the other hand if $P < N/2$, then

$$\begin{aligned}
&\sum_{0 \leq z_i \leq P} ((1 + (H\|zz_i\|))^{-\delta} + C) \ll \sum_{0 \leq z_i \leq P} ((1 + (H|z_i/N|))^{-\delta} + C) \\
&\ll PC + 1 + \sum_{0 < |z_i| \leq P} (H|z_i/N|)^{-\delta} \ll 1 + PC + H^{-\delta} N^{\delta} P^{1-\delta} \\
&\ll P(1/P + (HP/N)^{-\delta} + C).
\end{aligned}$$

Therefore,

$$\prod_{i=1}^{n} \left(\sum_{|z_i| \leq P} ((1 + H\|zz_i\|)^{-\delta} + C) \right) \ll P^n \prod_{i=1}^{n} (1/P + |z| + (HP|z|)^{-\delta} + H^{-\delta} + C),$$

which implies the lemma. □

3. Exponential sum estimates

In this section, we will assume that f and w satisfy (1.5) and (1.6) respectively. Throughout, let $x_0 \in M$ be an arbitrary point and let

$$x_0 = (x_{0,1}, ..., x_{0,n}), \quad \text{where } x_{0,i} \in M_0 \text{ for } i = 1, ..., n.$$

Given any $P > 1$ and and any $\alpha \in \mathbb{R}$, our prime focus in this section will be to establish bounds for the exponential sum $S(\alpha)$ defined in (1.3):

$$S(\alpha) := \sum_{\mathbf{x} \in \mathbb{Z}^n} w(\mathbf{x}/P) f(u(\mathbf{x})x_0) e(\alpha F(\mathbf{x})).$$

Recall that (1.4) allows us to consider the integral $\Sigma(P) = \int_0^1 S(\alpha) d\alpha$. Let $Q = P^{\Delta}$ where $0 < \Delta < 1$ be a parameter to be chosen later in the proof Lemma 4.1. An application of Dirichlet approximation hands us:

$$(0,1) \subseteq \bigcup_{q=1}^{Q} \bigcup_{\substack{0 \le a < q \\ \gcd(a,q)=1}} \{|a/q - \alpha| < (qQ)^{-1}\}. \tag{3.1}$$

Therefore, for a fixed value of $1 \le q \le Q$ and for a fixed $1 \le a < q$ co-prime to q, we need to estimate integrals of the form

$$\int_{|z|<1/(qQ)} |S(a/q + z)| dz.$$

Therefore throughout, let $\alpha = a/q + z$, where $|z| < (qQ)^{-1}$. We would need to bound $S(a/q + z)$ in two different ways, which will be our focus in this section. For our first bound, i.e. Lemma 3.1, we will begin by splitting the sum over $\mathbf{x}$ as $\mathbf{x}_1 + N\mathbf{x}_2$, for a suitable choice of N, depending on z. For a fixed choice of $\mathbf{x}_2$, we will estimate the corresponding exponential sum separately, and gain from the fact that for most of the values of $\mathbf{x}_2$, we would be able to bound the exponential sum satisfactorily. The second bound (Lemma 3.2) will be provided by van der Corput differencing. The first bound will be useful to deal with mid-ranges of z and the latter will be used to deal when z is small or relatively large.

Lemma 3.1. *Let $P \in \mathbb{Z}_{>0}$, let $f \in \mathrm{C}^{\infty}(M)$ and $w \in \mathrm{C}_c^{\infty}((-1,1)^n)$ satisfying* (1.5) *and* (1.6) *respectively, and let $\alpha \in \mathbb{R}$ satisfying $\alpha = a/q + z$, where $|z| \le q^{-1}Q^{-1}$ where $1 \le q \le Q = P^{\Delta}$, say. Then, given any $0 < \varepsilon \ll_{\Delta} 1$ we have*

$$|S(\alpha)| \ll_{\varepsilon,\Delta} S_{\infty,3n}(w) \|f\|_{\mathrm{L}^{\infty}_{9n+\varepsilon}} P^{n+\varepsilon}(q^{n/2}(|z| + 1/P)^{n/6} + q^{-n/2}(1 + |Pz^{1/2}|)^{-n/6}). \tag{3.2}$$

Proof. Before we begin the proof of the Lemma, let us give an intuitive idea of the proof. We will begin by splitting the sum $|\mathbf{x}| \ll P$ appearing in the definition of $S(\alpha)$ in (1.3) as $\mathbf{x} = \mathbf{z} + N\mathbf{y}$, where N will be chosen roughly of size $|z|^{-1/2}$. Thus, for any fixed value of $\mathbf{y} = \mathbf{y}_0$, we will treat $F(\mathbf{z}+N\mathbf{y}_0)$ as a quadratic polynomial with leading quadratic part $F(\mathbf{z})$. This decomposition is done in such a way that the quadratic term $zF(\mathbf{z})$ can essentially be disregarded as it stays bounded. On the other hand, the factor $e(aF(\mathbf{z})/q)$ is periodic modulo q. Therefore, we further split $\mathbf{z} = \mathbf{z}_0 + q\mathbf{z}_1$, where the sum over $\mathbf{z}_1$ is roughly of size $O(N/q)$. We then treat $\mathbf{z}_0$ and $\mathbf{y}_0$ as fixed and use Lemma

2.3 to estimate twisted averages of f along the arithmetic progressions with gaps of size q with total length $O(N/q)$ along the horocycle orbit:

$$\{u((\mathbf{z}_0 + N\mathbf{y}_0 + q\mathbf{z}_1)x_0) : |\mathbf{z}_1| \ll (N/q)\}.$$

Note that in the extreme case when q is roughly of size Q and when z is roughly of size $1/Q^2$, N is roughly of size Q which means the sum over $\mathbf{z}_1$ is only of size $O(1)$. Thus, we are unable to extract a useful bound using this method, which is reflected in the first bound on the right hand side of (3.2). However, this bound is powerful as long as we are away from this extreme case.

We now begin the precise proof. Let $0 < \varepsilon < \Delta/2$ be a small, positive number. Let $N = \min\{\lfloor P^{-\varepsilon}|z|^{-1/2}\rfloor, P\}$. Throughout, we treat Δ as fixed and our constants in $O(\cdot)$ and $\ll$ are allowed to depend on it. The condition $|z| \le q^{-1}P^{-\Delta}$ and that $\varepsilon < \Delta/2$ implies that $1 \le N$. We begin by splitting the sum over $\mathbf{x}$ in (1.3) into $O((P/N)^n)$ sums of length N each. [5, Lemma 2] hands us an elegant and smooth way of doing so. [5, Lemma 2] gives us that for any $0 < \delta \le 1$, there is a smooth function ω_δ satisfying

$$\omega(x) = \delta^{-1}\int \omega_\delta(\frac{x-y}{\delta}, y)dy. \tag{3.3}$$

The function ω_δ further satisfies

$$|\partial^\beta_{x,y}\omega_\delta(x,y)| \ll_\beta S_{\infty,|\beta|}(\omega). \tag{3.4}$$

Moreover, for a fixed y, the x-support of $w_\delta(x,y)$, is contained in the set $\{|x| \le 1\}$, and the support of $\omega_\delta(\frac{x-y}{\delta}, y)$ is contained in the support of ω for every y. Since ω is supported in $(-1,1)$, this implies that ω_δ is supported in the set $[-1,1] \times [-1-\delta, 1+\delta]$.

Using our definition of the function w in (1.6), we may then analogously obtain

$$w(\mathbf{x}) = \delta^{-n}\int w_\delta(\frac{\mathbf{x}-\mathbf{y}}{\delta}, \mathbf{y})d\mathbf{y}, \tag{3.5}$$

where

$$w_\delta(\mathbf{x},\mathbf{y}) = \prod_{i=1}^n \omega_\delta(x_i, y_i). \tag{3.6}$$

Thus, for any $0 < \delta \le 1$, and any $\mathbf{x} \in \mathbb{R}^n$, we have

$$\begin{aligned} w(\mathbf{x}/P) &= \delta^{-n}\int w_\delta(\frac{\mathbf{x}}{P\delta} - \frac{\mathbf{y}}{\delta}, \mathbf{y})d\mathbf{y} = \int w_\delta(\frac{\mathbf{x}}{P\delta} - \mathbf{y}, \delta\mathbf{y})d\mathbf{y} \\ &= \sum_{\mathbf{y}_0 \in \mathbb{Z}^n}\int_{|\mathbf{y}_1|<1/2} w_\delta(\frac{\mathbf{x}}{P\delta} - \mathbf{y}_0 - \mathbf{y}_1, \delta(\mathbf{y}_0 + \mathbf{y}_1))d\mathbf{y}_1 \\ &= \sum_{\mathbf{y}_0 \in \mathbb{Z}^n} W_{\delta,\mathbf{y}_0}(\frac{\mathbf{x} - P\delta\mathbf{y}_0}{P\delta}), \end{aligned}$$

where

$$W_{\delta,\mathbf{y}}(\mathbf{x}) = \int_{|\mathbf{y}_1|<1/2} w_\delta(\mathbf{x} - \mathbf{y}_1, \delta(\mathbf{y} + \mathbf{y}_1))d\mathbf{y}_1. \tag{3.7}$$

Since the support of w_δ is contained in the hypercube $[-1,1]^n \times [-1-\delta, 1+\delta]^n$, the sum over $\mathbf{y}_0$ is contained in the set $|\mathbf{y}_0| \ll \delta^{-1}$ and for such $\mathbf{y}_0$'s the function $W_{\delta,\mathbf{y}_0}(\mathbf{x})$ is supported in the set $\{|\mathbf{x}| < 3/2\}$.

We now choose $\delta = N/P$, where N as chosen at the beginning of the proof. Using this choice of δ, we thus arrive at

$$S(\alpha) = \sum_{\mathbf{y}_0 \in \mathbb{Z}^n} \sum_{\mathbf{x} \in \mathbb{Z}^n} W_{\delta,\mathbf{y}_0}(\frac{\mathbf{x} - N\mathbf{y}_0}{N}) f(u(\mathbf{x})x_0) e(\alpha F(\mathbf{x})).$$

At this point we introduce $\mathbf{z} = \mathbf{x} - N\mathbf{y}_0$. The above expression can be rewritten as

$$S(\alpha) = \sum_{\mathbf{y}_0 \in \mathbb{Z}^n} \sum_{\mathbf{z} \in \mathbb{Z}^n} W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}}{N}) f(u(\mathbf{z} + N\mathbf{y}_0)x_0) e(\alpha F(\mathbf{z} + N\mathbf{y}_0)). \tag{3.8}$$

Note that for a fixed value of $\mathbf{y}_0 \in \mathbb{Z}^n$, the function $W_{\delta,\mathbf{y}_0}(\mathbf{z})$ is a smooth function supported in the set $\{|\mathbf{z}| < 2\}$. Moreover, using the bounds on the derivatives of ω_δ in (3.4), we further have

$$|\partial_{\mathbf{z}}^{\beta} W_{\delta,\mathbf{y}_0}(\mathbf{z})| \ll_\beta S_{\infty,|\beta|}(w). \tag{3.9}$$

The sum over $\mathbf{y}_0$ is supported in the set $\{|\mathbf{y}_0| \ll P/N\}$.

Let $\alpha = a/q + z$, as given. We now make a further change of variables $\mathbf{z} = \mathbf{z}_0 + q\mathbf{z}_1$ to write $S(\alpha)$ as:

$$S(\alpha) = \sum_{\mathbf{y}_0 \in \mathbb{Z}^n} \sum_{\mathbf{z}_0 \in [0,q-1]^n} \sum_{\mathbf{z}_1 \in \mathbb{Z}^n} W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}_0 + q\mathbf{z}_1}{N}) f(u(\mathbf{z})x_1) e(\alpha F(\mathbf{z}_0 + q\mathbf{z}_1 + N\mathbf{y}_0))). \tag{3.10}$$

Here, the notation $\mathbf{z}_0 \in [0, q-1]^n$ mean that each co-ordinate of $\mathbf{z}_0$ is an integer between (and including) 0 and $q-1$. Here,

$$x_1 := u(N\mathbf{y}_0)x_0. \tag{3.11}$$

We begin by noting that

$$\begin{aligned}
&e((a/q + z)F(\mathbf{z}_0 + q\mathbf{z}_1 + N\mathbf{y}_0)) \\
&= e((a/q + z)(F(\mathbf{z}_0 + q\mathbf{z}_1) + 2N(L\mathbf{y}_0) \cdot (\mathbf{z}_0 + q\mathbf{z}_1) + N^2 F(\mathbf{y}_0)) \\
&= e((a/q + z)N^2 F(\mathbf{y}_0)) e_q(a(F(\mathbf{z}_0) + 2N(L\mathbf{y}_0) \cdot \mathbf{z}_0)) e(z(F(\mathbf{z}_0 + q\mathbf{z}_1) + 2N(L\mathbf{y}_0 \cdot (\mathbf{z}_0 + q\mathbf{z}_1))),
\end{aligned}$$

where $e_q(x) := \exp(2\pi i x/q)$ as is a standard notation. Recall here that L is the $n \times n$ integer matrix defining F. For now, we will treat $\mathbf{y}_0$ as fixed and concentrate on the exponential sum

$$\begin{aligned}
S_1 := S_1(z, \mathbf{y}_0) := &\sum_{\mathbf{z}_0 \in [0,q-1]^n} e_q(a(F(\mathbf{z}_0) + 2N(L\mathbf{y}_0) \cdot \mathbf{z}_0)) \times \\
&\sum_{\mathbf{z}_1 \in \mathbb{Z}^n} W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}_0 + q\mathbf{z}_1}{N}) f(u(\mathbf{z}_0 + q\mathbf{z}_1)x_1) e(z(F(\mathbf{z}_0 + q\mathbf{z}_1) + 2N(L\mathbf{y}_0 \cdot (\mathbf{z}_0 + q\mathbf{z}_1)).
\end{aligned}$$

We may now apply Poisson summation formula to the sum over $\mathbf{z}_1$ to obtain

$$\begin{aligned}
&\sum_{\mathbf{z}_1\in\mathbb{Z}^n} W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}_0+q\mathbf{z}_1}{N})f(u(\mathbf{z}_0+q\mathbf{z}_1)x_1)e(z(F(\mathbf{z}_0+q\mathbf{z}_1)+2N(L\mathbf{y}_0\cdot(\mathbf{z}_0+q\mathbf{z}_1))\\
&=\sum_{\mathbf{v}\in\mathbb{Z}^n}\int W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}_0+q\mathbf{z}_1}{N})f(u(\mathbf{z}_0+q\mathbf{z}_1)x_1)e(z(F(\mathbf{z}_0+q\mathbf{z}_1)+2N(L\mathbf{y}_0\cdot(\mathbf{z}_0+q\mathbf{z}_1))e(-\mathbf{v}\cdot\mathbf{z}_1)d\mathbf{z}_1\\
&=\sum_{\mathbf{v}\in\mathbb{Z}^n} q^{-n}\int W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}}{N})f(u(\mathbf{z})x_1)e(z(F(\mathbf{z})+2N(L\mathbf{y}_0\cdot\mathbf{z})e(-\mathbf{v}\cdot(\mathbf{z}-\mathbf{z}_0)/q)d\mathbf{z}\\
&=\sum_{\mathbf{v}\in\mathbb{Z}^n} q^{-n}e(\mathbf{v}\cdot\mathbf{z}_0/q)\int W_{\delta,\mathbf{y}_0}(\frac{\mathbf{z}}{N})f(u(\mathbf{z})x_1)e(z(F(\mathbf{z})+2N(L\mathbf{y}_0\cdot\mathbf{z})e(-\mathbf{v}\cdot\mathbf{z}/q)d\mathbf{z},
\end{aligned}$$

where in the second last step we have substituted $\mathbf{z}=\mathbf{z}_0+q\mathbf{z}_1$. This leads us to

$$S_1 = q^{-n}\sum_{\mathbf{v}\in\mathbb{Z}^n} S_q(a,\mathbf{v})I(z,-2NzL\mathbf{y}_0+\mathbf{v}/q),$$

where

$$S_q(a,\mathbf{v}) := \sum_{\mathbf{x} \bmod q} e_q(a(F(\mathbf{x})+2NL\mathbf{y}_0\cdot\mathbf{x})+\mathbf{x}\cdot\mathbf{v}), \tag{3.12}$$

is a standard quadratic exponential sum and

$$I(z,\mathbf{v}) := \int W_{\delta,\mathbf{y}_0}(\mathbf{z}/N)f(u(\mathbf{z})x_1)e(zF(\mathbf{z})-\mathbf{v}\cdot\mathbf{z})d\mathbf{z}, \tag{3.13}$$

is the corresponding exponential integral.

The exponential sum we encounter in (3.12) is a standard quadratic exponential sum. A standard bound that leads to [5, Lemma 25] hands us square root cancellations in the exponential sums for all $\mathbf{v}$'s. This follows essentially from squaring and further changing the variable to $\mathbf{x}_3=\mathbf{x}_2-\mathbf{x}_1$:

$$\begin{aligned}
|S_q(a,\mathbf{v})|^2 &= |\sum_{\mathbf{x} \bmod q} e_q(a(F(\mathbf{x})+2NL\mathbf{y}_0\cdot\mathbf{x})+\mathbf{x}\cdot\mathbf{v})|^2\\
&= |\sum_{\mathbf{x}_1,\mathbf{x}_2 \bmod q} e_q(a(F(\mathbf{x}_2)-F(\mathbf{x}_1))+(2aNL\mathbf{y}_0+\mathbf{v})\cdot(\mathbf{x}_2-\mathbf{x}_1))|\\
&\ll \sum_{\mathbf{x}_1 \bmod q} |\sum_{\mathbf{x}_3 \bmod q} e_q(a(F(\mathbf{x}_1+\mathbf{x}_3)-F(\mathbf{x}_1))+(2aNL\mathbf{y}_0+\mathbf{v})\cdot\mathbf{x}_3)|\\
&\ll \sum_{\mathbf{x}_1 \bmod q} |\sum_{\mathbf{x}_3 \bmod q} e_q((aM\mathbf{x}_1+2aNL\mathbf{y}_0+\mathbf{v})\cdot\mathbf{x}_3)|\\
&\ll q^n\{\mathbf{x}_1 \bmod q : q \mid (2aM\mathbf{x}_1+2aNL\mathbf{y}_0+\mathbf{v})\} \ll q^n\{\mathbf{x} \bmod q : q \mid 2M\mathbf{x}\}.
\end{aligned}$$

Here, to get the last inequality, we have used that if $\mathbf{y}=\mathbf{y}_1$ and $\mathbf{y}_2$ are two solutions of $q \mid 2aM\mathbf{y}+2aNL\mathbf{y}_0+\mathbf{v}$, then their difference must satisfy $q \mid 2M(\mathbf{y}_1-\mathbf{y}_2)$. Using the Smith normal form for $M=SDT$, where S,D,T are matrices with integer entries, where S,T have determinant ±1 and D is a diagonal matrix, we are able to obtain:

$$\{\mathbf{x}_1 \bmod q : q \mid 2M\mathbf{x}_1\} \ll_F 1.$$

To sum up, for any integer q, any a satisfying $\gcd(a,q)=1$, and any $\mathbf{v}\in\mathbb{Z}^n$ we have

$$|S_q(a,\mathbf{v})| \ll_F q^{n/2}, \tag{3.14}$$

where the implied constant only depends on the discriminant of the form F. The reader may also refer to [23, Lemma 2.5] where (3.14) is proved in the function field setting. A minor modification of this bound will work here. We now turn to bounding the exponential integral. Note that the exponential integral we encounter here will turn out to be simpler than the typical quadratic exponential integral which shows up in the circle method considerations. This is due to the fact that we have truncated the the sum over $\mathbf{x}$ to ensure that the integral over $\mathbf{z}$ is over a box of smaller size. As a result, $|zF(\mathbf{z})| \ll P^{-\varepsilon}$, for all $|\mathbf{z}|\le 2N$. We may now use a Taylor series expansion along with the fact that $Q=P^{\Delta}$, where $0<\Delta<1$, to write

$$e(zF(\mathbf{z})) = e(F(z^{1/2}\mathbf{z})) = \sum_{|\beta|\le k/\varepsilon} c_\beta (z^{1/2}\mathbf{z})^\beta + O_{k,\varepsilon}(P^{-k}), \tag{3.15}$$

where, the constants c_β are absolutely bounded

$$|c_\beta| \ll_\beta 1,$$

and given $\beta=(\beta_1,\dots,\beta_n)\in\mathbb{N}^n$, and any vector $\mathbf{z}\in\mathbb{R}^n$, $\mathbf{z}^\beta$ denote the monomial

$$\mathbf{z}^\beta := \prod_{i=1}^n z_i^{\beta_i}.$$

In light of (3.15), assuming that $\log q \ll \log P$, we have

$$\begin{aligned} |S_1| &= q^{-n}\sum_{|\beta|\le k/\varepsilon}|c_\beta|\sum_{\mathbf{v}\in\mathbb{Z}^n} S_q(a,\mathbf{v}) I_\beta(z,-2NzL\mathbf{y}_0+\mathbf{v}/q)| + O_{k,\varepsilon}(S_{\infty,0}(w)\|f\|_{\mathrm{L}^\infty}N^nP^{-k}) \\ &\ll_{k,\varepsilon} q^{-n/2}\sum_{|\beta|\le k/\varepsilon}\sum_{\mathbf{v}\in\mathbb{Z}^n}|I_\beta(z,-2NzL\mathbf{y}_0+\mathbf{v}/q)| + O_{k,\varepsilon}(S_{\infty,0}(w)\|f\|_{\mathrm{L}^\infty}N^nP^{-k}), \end{aligned} \tag{3.16}$$

where I_β is the exponential integral:

$$I_\beta(z,\mathbf{v}) := \int (z^{1/2}\mathbf{z})^\beta W_{\delta,\mathbf{y}_0}(\mathbf{z}/N) f(u(\mathbf{z})x_1) e(-\mathbf{v}\cdot\mathbf{z})d\mathbf{z}. \tag{3.17}$$

Note here that $|z|^{1/2}<1/N$, and therefore $|z^{1/2}N|<1$. We next write

$$\begin{aligned} I_\beta(z,\mathbf{v}) &= (z^{1/2}N)^{|\beta|}\int (\mathbf{z}/N)^\beta W_{\delta,\mathbf{y}_0}(\mathbf{z}/N) f(u(\mathbf{z})x_1)e(-\mathbf{v}\cdot\mathbf{z})d\mathbf{z} \\ &= (z^{1/2}N)^{|\beta|}\int W_{\delta,\mathbf{y}_0,\beta}(\mathbf{z}/N) f(u(\mathbf{z})x_1)e(-\mathbf{v}\cdot\mathbf{z})d\mathbf{z}, \end{aligned} \tag{3.18}$$

where $W_{\delta,\mathbf{y}_0,\beta}(\mathbf{z}) := \mathbf{z}^\beta W_{\delta,\mathbf{y}_0}(\mathbf{z})$ is a smooth function whose derivatives are $\ll$ those of $W_{\delta,\mathbf{y}_0}$ and further applying (3.9) we have

$$|\partial_{\mathbf{z}}^{\beta'} W_{\delta,\mathbf{y}_0,\beta}(\mathbf{z})| \ll_{\beta',\beta} S_{\infty,|\beta'|}(w). \tag{3.19}$$

The main advantage of the Taylor expansion in (3.15) is that the integral in (3.18) now splits as a product of n separate one dimensional integrals. We may now invoke Lemma 2.2 to bound each of these one dimensional integrals. We thus end up with

$$\begin{aligned}&\sum_{\mathbf{v}\in\mathbb{Z}^n}|I_\beta(z,-2NzL\mathbf{y}_0+\mathbf{v}/q)|\\ &\ll_\varepsilon N^{n+\varepsilon}S_{\infty,3n}(w)\prod_{i=1}^n\|f_i\|_{\mathrm{L}^\infty_{9+\varepsilon}}\left((1+\|-2qzN(L\mathbf{y}_0)_i\|N/q)^{-1/6}+qN^{-1/6}\right).\end{aligned}\tag{3.20}$$

Here, since M_0 is compact, we have used the L^∞ bound to replace the L^2 norm, and similarly used (3.9) to bound the norm of $W_{\delta,\mathbf{y}_0}$ appearing there. Moreover, using (1.5), we may replace $\prod_{i=1}^n\|f_i\|_{\mathrm{L}^\infty_{9+\varepsilon}}$ simply by $\|f\|_{\mathrm{L}^\infty_{(9+\varepsilon)n}}$. Substituting the bound in (3.20) to (3.16) and further summing over $\mathbf{y}_0$ in (3.10), we obtain

$$\begin{aligned}&|S(\alpha)|\ll_{k,\varepsilon}S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{(9+\varepsilon)n}}\times\\ &\left(\sum_{\substack{\mathbf{y}_0\in\mathbb{Z}^n\\|\mathbf{y}_0|\ll P/N}}N^{n+\varepsilon}q^{-n/2}\prod_{i=1}^n\left((1+\|-2qzN(L\mathbf{y}_0)_i\|N/q)^{-1/6}+qN^{-1/6}\right)+P^{n-k}\right).\end{aligned}$$

Since we are free to choose k, we may henceforth choose $k=n$. Therefore, note that the first term in the above equation is always dominant in this case and hence the term $P^{n-n}=1$ can be disregarded. When $N=P$, we simplify the above bound to get

$$\begin{aligned}|S(\alpha)|&\ll_\varepsilon S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{(9+\varepsilon)n}}\sum_{\substack{\mathbf{y}_0\in\mathbb{Z}^n\\|\mathbf{y}_0|\ll P/N}}N^{n+\varepsilon}q^{-n/2}\left(1+qN^{-1/6}\right)^n\\ &\ll_\varepsilon S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{(9+\varepsilon)n}}P^{n+\varepsilon}(q^{-n/2}+q^{n/2}P^{-n/6}).\end{aligned}$$

On the other hand, when $N=O(|z|^{-1/2}P^{-\varepsilon})<P$, we may employ Lemma 2.3 to obtain

$$\begin{aligned}|S(\alpha)|&\ll_\varepsilon S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{(9+\varepsilon)n}}P^{n+\varepsilon}q^{-n/2}(N/P+|qNz|+|N/q|^{-1/6}+|zNP|^{-1/6}+qN^{-1/6})^n\\ &\ll_\varepsilon P^{n+(n+1)\varepsilon}S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{(9+\varepsilon)n}}(q^{-n/2}|z^{1/2}P|^{-n/6}+q^{n/2}|z|^{n/6}).\end{aligned}$$

Combining these two bounds, and choosing an $\varepsilon\ll_{\Delta,n}1$, we get (3.2). ☐

The above bound would need to be supplemented by a van der Corput bound, which we will obtain in Lemma 3.2. Before we present this result, let us give a short explanation of the van der Corput differencing process to be used here. This method starts with a clever use of Cauchy-Schwarz inequality to bound the exponential sum $S(\alpha)$. This amounts to considering a convolution of the original sum (see (3.24)). The convolution amounts to multiplication on the spectral (Fourier transform) side. Upon choosing the H parameter correctly, we both "smooth" the exponential integral as well as choose a certain set of appearing frequencies which are controlled by (3.33). To deal with the frequencies which are relatively large (see (3.34)), one utilizes (2.4) which is a consequence of the general oscillation bound in (2.1). This helps us save significantly in the *generic* frequency setting. On the other hand, while dealing with low frequencies (see (3.35)), one may bootstrap the cancellation of the integral using effective equidistribution of the horocycle flow, namely, the γ term in (2.1). This bound depends on the spectral gap parameter γ. However we win from the fact that the proportion of such small frequencies is relatively low.

Lemma 3.2. *Let* $\alpha = a/q + z$*, where* $1 \le q \le P$*,* $\gcd(a,q) = 1$ *and* $|z| \le 1/q^2$*. Then for all* $0 < \varepsilon \ll 1$*, we have*

$$|S(\alpha)| \ll_\varepsilon S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{9n+\varepsilon}} P^{n+\varepsilon}(q^{-1/2} + (P/q)^{-1/228})^n. \tag{3.21}$$

Moreover, there exists $\gamma_1 := \gamma_1(\Gamma_0)$ *such that for any* α *as before, we have*

$$|S(\alpha)| \ll S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{9n+1}} P^{n-\gamma_1}. \tag{3.22}$$

Proof. We start by noticing that for any $H \in \mathbb{Z}_{>0}$,

$$H^n S(\alpha) := \sum_{\mathbf{x}} \sum_{0 \le \mathbf{h} < H} G(\mathbf{x} + \mathbf{h}),$$

say, where

$$G(\mathbf{x}) = w(\mathbf{x}/P) f(u(\mathbf{x})x_0) e(\alpha F(\mathbf{x})). \tag{3.23}$$

Here $0 \le \mathbf{h} < H$ is a shorthand notation to denote that $h_i \in \mathbb{Z}$ satisfying $0 \le h_i < H$ for all $1 \le i \le n$. Recall that w is assumed to be supported in $(-1,1)^n$. Throughout, we will assume that $H \le P/2$. Thus, the sum over $\mathbf{x}$ is supported in the set $-P \ll \mathbf{x} \ll P$. We may now use this fact and use Cauchy-Schwarz inequality for the sum over $\mathbf{x}$ to get

$$\begin{aligned} H^{2n}|S(\alpha)|^2 &\ll P^n \sum_{\mathbf{h}_1,\mathbf{h}_2} \sum_{\substack{\mathbf{x} \in \mathbb{Z}^n \\ 0 \le \mathbf{x}+\mathbf{h}_1, \mathbf{x}+\mathbf{h}_2 < P}} G(\mathbf{x}+\mathbf{h}_1)\overline{G(\mathbf{x}+\mathbf{h}_2)} \\ &\ll P^n \sum_{|\mathbf{h}|<H} N(\mathbf{h}) \sum_{\mathbf{x} \in \mathbb{Z}^n} G(\mathbf{x}+\mathbf{h})\overline{G(\mathbf{x})}, \end{aligned} \tag{3.24}$$

where

$$N(\mathbf{h}) := \#\{0 \le \mathbf{h}_1, \mathbf{h}_2 < H : \mathbf{h} = \mathbf{h}_1 - \mathbf{h}_2\} \le H^n.$$

Thus,

$$\begin{aligned} |S(\alpha)|^2 &\ll P^n H^{-2n} \sum_{|\mathbf{h}|<H} N(\mathbf{h}) \sum_{\mathbf{x} \in \mathbb{Z}^n} w_{\mathbf{h}}(\mathbf{x}/P) f_{\mathbf{h}}(u(\mathbf{x})x_0) e(\alpha(F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}))) \\ &\ll P^n H^{-n} \sum_{|\mathbf{h}|<H} \left| \sum_{\mathbf{x} \in \mathbb{Z}^n} w_{\mathbf{h}}(\mathbf{x}/P) f_{\mathbf{h}}(u(\mathbf{x})x_0) e(2(\alpha L\mathbf{h}) \cdot \mathbf{x}) \right|, \\ &\ll P^n H^{-n} \sum_{|\mathbf{h}|<H} \prod_{i=1}^n \left| \sum_{x_i \in \mathbb{Z}^n} \omega_{h_i}(\mathbf{x}/P) f_{i,h_i}(u_0(x_i)x_{0,i}) e((2\alpha L\mathbf{h})_i x_i) \right|, \end{aligned} \tag{3.25}$$

where since both f and w are assumed to be factorisable (see (1.5) and (1.6)), for any $y \in M$,

$$f_{\mathbf{h}}(y) := f(u(\mathbf{h})y)\overline{f(y)} := \prod_{i=1}^n f_{i,h_i}(y_i) := \prod_{i=1}^n f_i(u_0(h_i)y_i)\overline{f_i(y_i)}, \tag{3.26}$$

and

$$w_{\mathbf{h}}(\mathbf{x}) := w(\mathbf{x} + \mathbf{h}/P)\overline{w(\mathbf{x})} := \prod_{i=1}^n \omega_{h_i}(x_i) := \prod_{i=1}^n \omega(x_i + h_i/P)\overline{\omega(x_i)}.$$

Our main bound here will come from applying Poisson summation to the inner sums in (3.25), i.e., we obtain:

$$\begin{aligned}&\sum_{x_i\in\mathbb{Z}^n}\omega_{h_i}(x_i/P)f_{i,h_i}(x_i)e((2\alpha L\mathbf{h})_i x_i)\\ &\ll\sum_{v_i\in\mathbb{Z}}\left|\int\omega_{h_i}(x_i/P)f_{i,h_i}(u_0(x_i)x_{0,i})e(((2\alpha L\mathbf{h})_i-v_i))x_i)dx_i\right|.\end{aligned}\tag{3.27}$$

We now estimate the sum on the right hand side of (3.27) via Lemma 2.2. Therefore, for any $1\le i\le n$, and any $\varepsilon>0$, we have

$$\begin{aligned}&\sum_{v_i\in\mathbb{Z}}\left|\int\omega_{h_i}(x_i/P)f_{i,h_i}(u_0(x_i)x_{0,i})e(((2\alpha L\mathbf{h})_i-v_i)x_i)dx_i\right|\\ &\ll S_{1,3}(\omega_{h_i})\|f_{i,h_i}\|_{\mathrm{L}^2_{9+\varepsilon}}P\log^{1/2}(P)((1+\|(2\alpha L\mathbf{h})_i\|P)^{-1/6}+P^{-1/6}).\end{aligned}\tag{3.28}$$

We begin by bounding the derivatives of $f_{i,\mathbf{h}}$. Using the relation (2.15), for any $k\in\mathbb{Z}_{\ge0}$, and an element in the Lie algebra D of order k,

$$\|Df_{i,h_i}(x)\|_{\mathrm{L}^2}\ll(1+|h_i|)^{2k}\sum_{D_1,D_2:\mathrm{ord}(D_1)+\mathrm{ord}(D_2)=k}\|D_1(f_i)(u_0(h_i)x_i)D_2(f_i)(x_i)\|_{\mathrm{L}^2}.$$

As a result, we obtain

$$\|f_{i,h_i}\|_{\mathrm{L}^2_k}\ll(1+|h_i|)^{2k}\|f_i\|^2_{\mathrm{L}^\infty_k}.\tag{3.29}$$

Upon interpolation, this bound can be assumed to be true for all $k\in\mathbb{R}_{\ge0}$. Similarly,

$$S_{\infty,3}(\omega_{h_i})\ll S_{\infty,3}(\omega)^2.\tag{3.30}$$

Substituting (3.29) back in (3.28), we get

$$\begin{aligned}&\sum_{v_i\in\mathbb{Z}}\left|\int\omega_{h_i}(x_i/P)f_{i,h_i}(u_0(x_i)x_{0,i})e(((2\alpha L\mathbf{h})_i-v_i)x_i)dx_i\right|\\ &\ll H^{18}S_{\infty,3}(\omega)^2\|f_i\|^2_{\mathrm{L}^\infty_{9+\varepsilon}}P^{1+\varepsilon}((1+\|(2\alpha L\mathbf{h})_i\|P)^{-1/6}+P^{-1/6}).\end{aligned}\tag{3.31}$$

The above expression holds for ε small enough. Note that since $H\ll P$, the extra powers of H^ε have been absorbed into the term P^ε. When $(L\mathbf{h})_i=0$, we bypass Poisson summation and directly use the following bound:

$$|\sum_{x_i\in\mathbb{Z}}\omega_{h_i}(x_i/P)f_{i,h_i}(u_0(x_i)x_{0,i})|\ll\|f_{i,h_i}\|_{\mathrm{L}^\infty}\sum_{x_i\in\mathbb{Z}}|\omega_{h_i}(x_i/P)|\ll P\|f_i\|^2_{\mathrm{L}^\infty}S_{\infty,0}(\omega)^2.\tag{3.32}$$

Therefore, for $\varepsilon > 0$ small enough, we have

$$\begin{aligned}
&\sum_{|\mathbf{h}|<H}\left|\sum_{\mathbf{x}\in\mathbb{Z}^n} w_{\mathbf{h}}(\mathbf{x}/P)f_{\mathbf{h}}(u(\mathbf{x})x_0)e(2(\alpha L\mathbf{h})\cdot\mathbf{x})\right| \\
&\ll P^{n+\varepsilon}\sum_{|\mathbf{h}|<H}\prod_{i=1}^{n} S_{\infty,3}(\omega)^2\|f_i\|^2_{\mathrm{L}^{\infty}_{9+\varepsilon}}(\delta_{(L\mathbf{h})_i\neq 0}H^{18}((1+\|(2\alpha L\mathbf{h})_i\|P)^{-1/6}+P^{-1/6})+\delta_{(L\mathbf{h})_i=0}) \\
&\ll S_{\infty,3n}(w)^2\|f\|^2_{\mathrm{L}^{\infty}_{(9+\varepsilon)n}}P^{n+\varepsilon}\sum_{|\mathbf{h}|<|L|H}\prod_{i=1}^{n}(\delta_{h_i\neq 0}H^{18}((1+\|2\alpha h_i\|P)^{-1/6}+P^{-1/6})+\delta_{h_i=0}).
\end{aligned}$$

Here, to obtain the last equation, we have made a change of variable to replace $L\mathbf{h}$ by $\mathbf{h}$. Eventually, we will choose $H \le q/(4|L|)$, which means that since $|z| < q^{-2}$,

$$|2zh_i| < 1/(2q), \forall |h_i| < |L|H. \tag{3.33}$$

Thus, if $q \nmid h_i$, then

$$\|2\alpha h_i\| \gg 1/q. \tag{3.34}$$

However, if $|h_i| \le |L|H \le q|L|/(4|L|) = q/2$, then $q \mid h_i$ if and only if $h_i = 0$. Therefore, when $h_i \neq 0$, where $|h_i| \le |L|H$, we may use (3.34). Therefore,

$$\begin{aligned}
&|S(\alpha)|^2 \\
&\ll P^{2n+\varepsilon}H^{-n}S_{\infty,3n}(w)^2\|f\|^2_{\mathrm{L}^{\infty}_{(9+\varepsilon)n}}\sum_{|\mathbf{h}|<|L|H}\prod_{i=1}^{n}(\delta_{h_i\neq 0}H^{18}((1+\|2\alpha h_i\|P)^{-1/6}+P^{-1/6})+\delta_{h_i=0}) \\
&\ll P^{2n+\varepsilon}S_{\infty,3n}(w)^2\|f\|^2_{\mathrm{L}^{\infty}_{(9+\varepsilon)n}}(H^{18}(|P/q|^{-1/6})+H^{-1})^n \\
&\ll P^{2n+\varepsilon}S_{\infty,3n}(w)^2\|f\|^2_{\mathrm{L}^{\infty}_{(9+\varepsilon)n}}\prod_{i=1}^{n}(H^{18}q^{1/6}P^{-1/6}+H^{-1}).
\end{aligned} \tag{3.35}$$

We now choose

$$H = \min\{q/(4|L|), (P/q)^{1/114})\}, \tag{3.36}$$

to get

$$|S(\alpha)| \ll P^{n+\varepsilon}(q^{-1/2}+(P/q)^{-1/228})^n S_{\infty,3n}(w)\|f\|_{\mathrm{L}^{\infty}_{9n+\varepsilon}}.$$

When $|z|$ and q are small, we only hope to exploit from the sum over $i=1$ and apply the second bound in (2.5). More explicitly, we begin with the following variant of (3.25)

(3.37)

$$\begin{aligned}
&|S(\alpha)|^2\\
&\ll P^{2n-1}H^{-n}\left(\prod_{i=2}^{n} S_{\infty,0}(\omega)^2\|f_i\|^2_{\mathrm{L}^\infty_0}\right)\sum_{|\mathbf{h}|<H}\left|\sum_{x_1\in\mathbb{Z}}\omega_{h_1}(x_1/P)f_{h_1,1}(u_0(x_1)x_{0,1})e(2(\alpha L\mathbf{h})_1x_1)\right|\\
&\ll P^{2n-1}H^{-n}\left(\prod_{i=2}^{n} S_{\infty,0}(\omega)^2\|f_i\|^2_{\mathrm{L}^\infty_0}\right)\sum_{|\mathbf{h}|<H}\sum_{v_1\in\mathbb{Z}}\left|\int \omega_{h_1}(x_1/P)f_{h_1,1}(u_0(x_1)x_{0,1})e(((2\alpha L\mathbf{h})_1-v_1)x_1)dx_1\right|\\
&\ll P^{2n+\varepsilon}S_{\infty,3n}(w)^2\|f\|^2_{\mathrm{L}^\infty_{9n+\varepsilon}}H^{18}(P^{-\gamma}+P^{-1/6}).
\end{aligned}$$

Here, we have applied (2.5) to bound the sum over x_1. Note that the worse Sobolev norms appearing here are only chosen to match with our bounds in (3.35). The second part of the lemma now follows from choosing $H=P^{\min\{\gamma,1/6\}/36}$, setting $\gamma_1=\min\{\gamma,1/6\}/78$ and by choosing $\varepsilon\ll_{\gamma,n}1$. □

4. Proof of Theorem 1.2

Recall again that (1.4) writes $\Sigma(P)$ as

$$\Sigma(P)=\int_0^1 S(\alpha)d\alpha.$$

Let $0<\Delta<1$ be a parameter to be chosen later in due course. The Dirichlet approximation theorem (3.1) hands us

$$\Sigma(P)\le\sum_{q=1}^{Q}\sum_{\substack{0\le a<q\\ \gcd(a,q)=1}}\int_{|a/q-\alpha|\le 1/(qQ)}|S(\alpha)|d\alpha. \tag{4.1}$$

We now split the right hand side into integral over two regions which typically correspond to the major and minor arc regimes in the circle method setting. Let ε_0 be a small parameter to be chosen in due course. We define

$$\begin{aligned}
\mathfrak{m}_1 &:= \bigcup_{q=1}^{Q}\bigcup_{\substack{0\le a<q\\ \gcd(a,q)=1}}\{|a/q-\alpha|<q^{-2}P^{-2+\varepsilon_0}\}\text{ and}\\
\mathfrak{m}_2 &:= \bigcup_{q=1}^{Q}\bigcup_{\substack{0\le a<q\\ \gcd(a,q)=1}}\{q^{-2}P^{-2+\varepsilon_0}\le|a/q-\alpha|<(qQ)^{-1}\}.
\end{aligned} \tag{4.2}$$

When $\alpha\in\mathfrak{m}_1$ the bound from (3.22) will suffice. On the other hand, when $\alpha\in\mathfrak{m}_2$, we will use a combination of the bounds in (3.2) and (3.21).

Lemma 4.1. *For any $n \geq 481$ and any $0 < \varepsilon_0 \leq 1/240$, we have*

$$\int_{\mathfrak{m}_2} |S(\alpha)| d\alpha \ll P^{n-2-\varepsilon_0/4} S_{\infty,3n}(w) \|f\|_{\mathrm{L}^{\infty}_{9n+1}}.$$

Proof. Let $Q = P^{\Delta}$ and let $0 < \varepsilon \ll_{\Delta} 1$ be an arbitrarily small number to be chosen later. We begin by combining bounds in (3.2) and (3.21) for any $\alpha = a/q + z$, where $|z| < (qQ)^{-1}$:

$$\begin{aligned} &|S(a/q+z)| \ll_{\varepsilon} P^{n+\varepsilon} S_{\infty,3n}(w) \|f\|_{\mathrm{L}^{\infty}_{9n+\varepsilon}} \times \\ &\left(\min\{q^{1/2}|z|^{1/6}, q^{-1/2}\} + q^{1/2}P^{-1/6} + q^{-1/2}(1+|Pz^{1/2}|)^{-1/6} + (P/q)^{-1/228}\right)^n \quad (4.3)\\ &\ll_{\varepsilon} P^{n+\varepsilon} S_{\infty,3n}(w) \|f\|_{\mathrm{L}^{\infty}_{9n+\varepsilon}} \left(|z|^{1/12} + q^{1/2}P^{-1/6} + q^{-1/2}(1+|Pz^{1/2}|)^{-1/6} + (P/q)^{-1/228}\right)^n, \end{aligned}$$

where we have used a geometric mean to bound the first term inside the brackets on the right side.

We start first by examining the second last term:

$$\begin{aligned} &\sum_{q=1}^{Q} \sum_{\substack{0 \leq a < q \\ \gcd(a,q)=1}} \int_{q^{-2}P^{-2+\varepsilon_0} \leq |z| < (qQ)^{-1}} q^{-n/2}(1+|Pz^{1/2}|)^{-n/6} dz \\ &\ll \sum_{q=1}^{Q} \sum_{\substack{0 \leq a < q \\ \gcd(a,q)=1}} q^{-n/3} \int_{q^{-2}P^{-2+\varepsilon_0} \leq |z| < (qQ)^{-1}} |qPz^{1/2}|^{-n/6} dz \quad (4.4)\\ &\ll \sum_{q=1}^{Q} q^{1-n/3} q^{-2} P^{-2} \int_{P^{\varepsilon_0} \leq |z| < \infty} |z|^{-n/12} dz \ll P^{-2-\varepsilon_0}, \end{aligned}$$

as long as $n \geq 13$. On the other hand,

$$(P/q)^{-n/228} + q^{n/2}P^{-n/6} \ll (P^{-1/228}Q^{1/228})^n + P^{-n/6}Q^{n/2}.$$

Since $|z| \leq (qQ)^{-1}$, the term $|z|^{1/12}$ term may simply be bound by

$$|z|^{1/12} \ll Q^{-1/12}. \quad (4.5)$$

At this point, we choose Q such that $Q^{1/12} = P^{1/228}/Q^{1/228}$, i.e., when $Q = P^{1/20}$ which means $\Delta = 1/20$. For this choice of Q,

$$(P/q)^{-n/228} + q^{n/2}P^{-n/6} + |z|^{n/12} \ll P^{-n/240}. \quad (4.6)$$

Thus, as long as $n = 481 \geq 2 \times 240 + 1$,

$$\left(\min\{q^{1/2}|z|^{1/6}, q^{-1/2}\} + (P/q)^{-1/228} + P^{-1/6}q^{1/2}\right)^n \ll P^{-2-1/240}.$$

Since the measure of $\mathfrak{m}_2$ is at most 1, this leads to

$$\int_{\mathfrak{m}_2} \left(\min\{q^{1/2}|z|^{1/6}, q^{-1/2}\} + (P/q)^{-1/228} + P^{-1/6}q^{1/2}\right)^n d\alpha \ll P^{-2-1/240}, \quad (4.7)$$

as long as $n \geq 481$. Lemma 4.1 now follows from combining bounds in (4.4) and (4.7) and further suitably choosing $\varepsilon \leq \varepsilon_0/4$. □

Proof. (Proof of Theorem 1.2) In order to prove Theorem 1.2, it is enough to bound the contribution from $\alpha \in \mathfrak{m}_1$. Note that the total measure of $\mathfrak{m}_1$ is estimated as

$$\mathrm{meas}(\mathfrak{m}_1) = \sum_{q=1}^{Q} 2q^{-2}P^{-2+\varepsilon_0} \ll P^{-2+\varepsilon_0}.$$

Therefore, a pointwise application of bound (3.22) would suffice us here. Therefore, using (3.22), for any ε_0 we have

$$\int_{\mathfrak{m}_0} |S(\alpha)|d\alpha \ll P^{n-\gamma_1}\mathrm{meas}(\mathfrak{m}_1)S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{9n+1}} \ll P^{n-2-\gamma_1+\varepsilon_0}S_{\infty,3n}(w)\|f\|_{\mathrm{L}^\infty_{9n+1}}. \tag{4.8}$$

Combining the results in Lemma 4.1 and (4.8), and choosing $\varepsilon_0 = \min\{1/240, \gamma_1/2\}$ and setting $\gamma_0 = \varepsilon_0/4$, we establish Theorem 1.2. ☐

Remark 4.2. As mentioned in the introduction, the situation of diagonal forms is significantly easier. We will give a quick sketch of this argument here. In fact, it would be enough to have $F(\mathbf{x}) = F_1(\mathbf{x}_1) + F_2(\mathbf{x}_2) + F_3(\mathbf{x}_3)$, where $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ where $\mathbf{x}_1, \mathbf{x}_2$ are at least two dimensional, and $\mathbf{x}_3$ being at least one dimensional. In this case, the exponential sum $S(\alpha)$ naturally splits as

$$S(\alpha) = S_1(\alpha)S_2(\alpha)S_3(\alpha),$$

where S_i's denote the corresponding exponential sums for the forms F_i for $i = 1, 2, 3$. We now apply the Hölder's inequality:

$$|\Sigma(P)| \le \int_0^1 |S_1(z)S_2(z)S_3(z)|dz \le \|S_3\|_{\mathrm{L}^\infty}\|S_1\|_{\mathrm{L}^2}\|S_2\|_{\mathrm{L}^2}.$$

Using (3.22), we have $\|S_3\|_{\mathrm{L}^\infty} \ll_f P^{n_3-\gamma_1}$, for some $\gamma_1 > 0$. Furthermore, for $i = 1, 2$, given any $\varepsilon > 0$, one may easily obtain

$$\begin{aligned}\int_0^1 |S_i(z)|^2 dz &= \int_0^1 \sum_{\mathbf{x}_i,\mathbf{x}_i' \in \mathbb{Z}^{n_i}} W_i(\mathbf{x}_i/P)\overline{W_i(\mathbf{x}_i'/P)}f(u(\mathbf{x}_i)x_0)\overline{f(u(\mathbf{x}_i')x_0)}e(z(F(\mathbf{x}_i) - F(\mathbf{x}_i')))dz \\ &= \sum_{\mathbf{x}_i,\mathbf{x}_i' \in \mathbb{Z}^{n_i}} W_i(\mathbf{x}_i/P)\overline{W_i(\mathbf{x}_i'/P)}f(u(\mathbf{x}_i)x_0)\overline{f(u(\mathbf{x}_i')x_0)} \int_0^1 e(z(F(\mathbf{x}_i) - F(\mathbf{x}_i')))dz \\ &\ll_f \#\{(\mathbf{x}_i, \mathbf{x}_i') \in \mathbb{Z}^{n_i} \times \mathbb{Z}^{n_i} : F_i(\mathbf{x}_i) - F_i(\mathbf{x}_i') = 0, |\mathbf{x}_i|, |\mathbf{x}_i'| \ll P\} \ll_{f,\varepsilon} P^{2n_i-2+\varepsilon},\end{aligned}$$

where in the final bound we have used [6, Theorem 2], which applies as long as $n_1, n_2 \ge 2$. Combining these bounds, we end up with

$$|\Sigma(P)| \ll_{\varepsilon,f} P^{n-2-\gamma_1+\varepsilon},$$

as long as $n_1, n_2 \ge 2$ and $n_3 \ge 1$.

5. Proof of Theorem 1.1

We are now set to prove Theorem 1.1, which will follow from Theorem 1.2. Throughout, we will assume that $n \geq 481$. We start by writing

$$\sum_{\substack{\mathbf{x}\in\mathbb{Z}^n,|\mathbf{x}|<P \\ F(\mathbf{x})=0}} f(u(\mathbf{x})x_0) = \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n \\ F(\mathbf{x})=0}} W(\mathbf{x}/P)f(u(\mathbf{x})x_0), \tag{5.1}$$

where W denotes the characteristic function of the hypercube $(-1,1)^n$. Since F is supposed to have no local obstructions, the asymptotic formula (1.2) implies that Theorem 1.1 is equivalent to proving that

$$\lim_{P\to\infty} \frac{1}{P^{n-2}} \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n \\ F(\mathbf{x})=0}} W(\mathbf{x}/P)f(u(\mathbf{x})x_0) = 0, \tag{5.2}$$

for any continuous function f of zero average.

In order to invoke Theorem 1.2, we will approximate W by a smooth function, and further approximate f by a sum of factorisable functions of zero average. We start with the latter. Since f is continuous and M compact, using the Stone–Weierstrass theorem for compact manifolds, given any $\varepsilon > 0$, we may write

$$f(g) = \sum_{i=1}^{m} h_i(g) + O(\varepsilon),$$

where m may depend on ε, and each h_i is a smooth, factorisable function, that is, it is of the form

$$h_i(g_1, ..., g_n) = h_{i,1}(g_1)...h_{i,n}(g_n). \tag{5.3}$$

Since f is of zero average and M is compact, we must further have

$$|\sum_{i=1}^{m} \int_M h_i(g)d\mu_G(g)| \ll \varepsilon.$$

Using this, we further reach:

$$f(g) = \sum_{i=1}^{m} h'_i(g) + O(\varepsilon),$$

where

$$h'_i(g) = h_i(g) - \int h_i(x)d\mu_G(x), \tag{5.4}$$

is a function of zero average. Note that since h_i is factorisable, $\int_M h_i(x)d\mu_G(x) = \prod_{j=1}^n \int_{M_0} h_{i,j}(x_j)d\mu_{G_0}(x_j)$. Now, we may next write

(5.5)
$$\begin{aligned} h_i'(g) &= \prod_{j=1}^n h_{i,j}(g_j) - \prod_{j=1}^n \int h_{i,j}(x_j)d\mu_{G_0}(x_j) \\ &= \prod_{j=1}^n \left(\left((h_{i,j}(g_j) - \int h_{i,j}(x_j)d\mu_{G_0}(x_j) \right) + \int h_{i,j}(x_j)d\mu_{G_0}(x_j) \right) - \prod_{j=1}^n \int h_{i,j}(x_j)d\mu_{G_0}(x_j). \end{aligned}$$

Note that each function $h_{i,j}(g_j) - \int h_{i,j}(x_j)d\mu_{G_0}(x_j)$ is smooth and of zero average. After expanding out the product over j in (5.5) and noticing that the constant term $\prod_{j=1}^n \int h_{i,j}(x_j)dx_j$ cancels out, we then write h_i' as a sum of factorizable functions of zero average. Therefore, we may now assume that

$$f(g) = \sum_{i=1}^{m_1} \phi_i(g) + O(\varepsilon), \tag{5.6}$$

where ϕ_i's are factorisable functions of zero average. Note that the derivatives of ϕ_i also satisfy

$$\|\phi_i\|_{\mathrm{L}_k^\infty} \ll_{\varepsilon,k,f} 1. \tag{5.7}$$

Therefore, we end up with

$$\begin{aligned} \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n \\ F(\mathbf{x})=0}} W(\mathbf{x}/P)f(u(\mathbf{x})x_0) &= \sum_{i=1}^{m_1} \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n \\ F(\mathbf{x})=0}} W(\mathbf{x}/P)\phi_i(u(\mathbf{x})x_0) + O_f(\varepsilon N_F(P)) \\ &= \sum_{i=1}^{m_1} \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n \\ F(\mathbf{x})=0}} W(\mathbf{x}/P)\phi_i(u(\mathbf{x})x_0) + O_f(\varepsilon P^{n-2}), \end{aligned} \tag{5.8}$$

using the asymptotic formula (1.2). Now let us focus on the sums corresponding to each ϕ_i. In order to invoke Theorem 1.2, W needs to be approximated by a smooth function. In order to do so, let $0 < \delta < 1$ be a parameter to be chosen in due course. Let w be a smooth factorisable function of the type (1.6) supported in $(-1,1)^n$. We may further assume that w is a non-negative function taking values in the closed interval $[0,1]$, it takes value 1 on the hypercube $(-1+\delta, 1-\delta)^n$, and that the derivatives of w satisfy

$$S_{\infty,k}(w) \ll \delta^{-k}. \tag{5.9}$$

The asymptotic formula (1.2) holds for any P, and therefore it hands us a constant $\gamma' > 0$ depending only on n and F such that

$$\begin{aligned}
&\sum_{\substack{\mathbf{x}\in\mathbb{Z}^n\\ F(\mathbf{x})=0}} W(\mathbf{x}/P)\phi_i(u(\mathbf{x})x_0)\\
&= \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n\\ F(\mathbf{x})=0}} w(\mathbf{x}/P)\phi_i(u(\mathbf{x})x_0) + O(\#\{\mathbf{x}\in\mathbb{Z}^n : (1-\delta)P \le |\mathbf{x}| < P, F(\mathbf{x})=0\})\\
&= \sum_{\substack{\mathbf{x}\in\mathbb{Z}^n\\ F(\mathbf{x})=0}} w(\mathbf{x}/P)\phi_i(u(\mathbf{x})x_0) + O(\|f\|_{\mathrm{L}^\infty}\delta P^{n-2}) + O(\|f\|_{\mathrm{L}^\infty}P^{n-2-\gamma'}).
\end{aligned}$$

Since ϕ_i is a factorisable function of zero average, without loss of generality we can assume that it is of type (1.5). We are now able to apply Theorem 1.2 to obtain

$$\left|\sum_{\substack{\mathbf{x}\in\mathbb{Z}^n\\ F(\mathbf{x})=0}} W(\mathbf{x}/P)\phi_i(u(\mathbf{x})x_0)\right| \ll_{f,\varepsilon} \delta^{-9n}P^{n-2-\gamma_0} + \delta P^{n-2} + P^{n-2-\gamma'}.$$

At this point, we choose $\delta = P^{-\gamma_2}$, where $\gamma_2 = \min\{\gamma', \gamma_0/(9n+1)\}$, and combine this bound with that in (5.8) to obtain

$$\left|\sum_{\substack{\mathbf{x}\in\mathbb{Z}^n\\ F(\mathbf{x})=0}} W(\mathbf{x}/P)f(u(\mathbf{x})x_0)\right| \ll_f \varepsilon P^{n-2} + C_\varepsilon P^{n-2-\gamma_2},$$

where C_ε denotes a constant which depends only on ε, F, n and Γ. Since γ_2 is independent of ε, for large enough P, we must have

$$P^{-(n-2)}\left|\sum_{\substack{\mathbf{x}\in\mathbb{Z}^n\\ F(\mathbf{x})=0}} W(\mathbf{x}/P)f(u(\mathbf{x})x_0)\right| \ll_f \varepsilon.$$

Since ε was chosen to be arbitrary, this establishes Theorem 1.1.

References

[1] B. J. Birch. Forms in many variables. *Proc. Roy. Soc. Ser. A*, 265:245–263, 1961/1962.

[2] Jean Bourgain. Pointwise ergodic theorems for arithmetic sets. *Inst. Hautes Études Sci. Publ. Math.*, (69):5–45, 1989. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[3] H. Davenport. Analytic methods for Diophantine equations and Diophantine inequalities. *Cambridge Mathematical Library.* second edition, With a foreword by R. C. Vaughan, D. R. Heath-Brown and D. E. Freeman, Edited and prepared for publication by T. D. Browning, Cambridge University Page, Cambridge, 2005.

[4] Livio Flaminio, Giovanni Forni, and James Tanis. Effective equidistribution of twisted horocycle flows and horocycle maps. *Geom. Funct. Anal.*, 26(5):1359–1448, 2016.

[5] D. R. Heath-Brown. A new form of the circle method, and its application to quadratic forms. *J. Reine Angew. Math.*, 481:149–206, 1996.

[6] D. R. Heath-Brown, The density of rational points on curves and surfaces (with an appendix by J.-L. Colliot-Thélène), Annals of Mathematics, Pages 553-598, Volume 155 (2002).
[7] H. Iwaniec, E. Kowalski, Analytic Number Theory (Colloquium Publications), Publisher : American Mathematical Society (15 July 2004), ISBN-10 : 0821836331.
[8] Roger L. Jones, Ergodic averages on spheres, *Journal d'Analyse Mathématique* volume 61, pages 29–45(1993)
[9] A. Katz. On mixing and sparse ergodic theorems. To appear in *Journal of Modern Dynamics.* arXiv:1612.01146.
[10] A. Katz. Applications of joinings to sparse equidistribution problems. Preprint, arXiv:1910.04674.
[11] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I.* Springer-Verlag, New York-Heidelberg, 1972. Translated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181.
[12] A. Magyar, On the distribution of solutions to Diophantine equations. *A panorama of discrepancy theory*, 487–538, Lecture Notes in Math., 2107, Springer, Cham, 2014.
[13] Gregory Margulis. Problems and conjectures in rigidity theory. In *Mathematics: frontiers and perspectives*, pages 161–174. Amer. Math. Soc., Providence, RI, 2000.
[14] T. McAdam. Almost-primes in horospherical flows on the space of lattices. . Preprint. arXiv:1802.08764.
[15] Peter Sarnak and Adrián Ubis. The horocycle flow at prime times. *J. Math. Pures Appl. (9)*, 103(2):575–618, 2015.
[16] Nimish A. Shah. Limit distributions of polynomial trajectories on homogeneous spaces. *Duke Math. J.*, 75(3):711–732, 1994.
[17] E. M. Stein; S. Wainger. Discrete analogues of singular Radon transforms. *Bull. Amer. Math. Soc.*, 23 (1990), no. 2, 537–544.
[18] E. M. Stein; S. Wainger. Discrete analogues in harmonic analysis. I. l2 estimates for singular Radon transforms. *Amer. J. Math.* 121 (1999), no. 6, 1291–1336.
[19] James Tanis and Pankaj Vishe. Uniform bounds for period integrals and sparse equidistribution. *Int. Math. Res. Not. IMRN*, (24):13728–13756, 2015.
[20] Adrián Ubis. Effective equidistribution of translates of large submanifolds in semisimple homogeneous spaces. *Int. Math. Res. Not. IMRN*, (18):5629–5666, 2017.
[21] Akshay Venkatesh. Sparse equidistribution problems, period bounds and subconvexity. *Ann. of Math. (2)*, 172(2):989–1094, 2010.
[22] William Veech. Unique ergodicity of horospherical flows, *Amer. J. Math.*, 99 (4), 827–859, 1977.
[23] P. Vishe. Rational points on complete intersections on $\mathbb{F}_q(t)$. arXiv:1907.07097.

Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, United Kingdom

Email address: pankaj.vishe@durham.ac.uk