Acta Math., 205 (2010), 361–402
DOI: 10.1007/s11511-010-0057-4

Prime and almost prime integral points on principal homogeneous spaces

by

Amos Nevo

Technion – Israel Institute of Technology
Haifa, Israel

Peter Sarnak

Institute for Advanced Study
Princeton, NJ, U.S.A.

1. Introduction

1.1. Prime and almost prime integral matrices of a given determinant

For integers $n\geqslant 2$ and $m\neq 0$ let $V_{m,n}(\mathbb{Z})$ or $\mathcal{O}^{m,n}$ denote the set of $n\times n$ integral matrices of determinant equal to m. We study the points $x\in\mathcal{O}^{m,n}$ for which

$$f(x)=\prod_{1\leqslant i\leqslant j\leqslant n} x_{ij},$$

or more generally any $f\in\mathbb{Q}[x_{ij}]$ which is integral on $\mathcal{O}^{m,n}$, has few (or fewest possible) prime factors. In general, given a set $\mathcal{O}$ of integer points in $\mathrm{Mat}_{n\times n}$ and $d\geqslant 1$, let $\mathcal{O}_d$ denote the reduction of $\mathcal{O}$ in $\mathrm{Mat}_{n\times n}(\mathbb{Z}/d\mathbb{Z})$. We say that f is *weakly primitive* for $\mathcal{O}$ if

$$\gcd f(\mathcal{O}):=\gcd\{f(x):x\in\mathcal{O}\}=1.$$

If f is not weakly primitive then f/N is, where $N=\gcd f(\mathcal{O})$, and we can represent any weakly primitive f as g/N, with $g\in\mathbb{Z}[x_{ij}]$ and $N=\gcd g(\mathcal{O})$.

Define the *saturation number* r_0 of the pair $(\mathcal{O}^{m,n},f)$ to be the least r such that the set of $x\in\mathcal{O}^{m,n}$, for which $f(x)$ has at most r prime factors, is Zariski-dense in the affine variety $V_{m,n}=\{x\in\mathrm{Mat}_{n\times n}:\det x=m\}=\mathrm{Zcl}(\mathcal{O}^{m,n})$. (We denote by Zcl the operation of taking Zariski closure in affine space; see also [S2] for a further discussion and motivation for this set up.) It turns out that r_0 is finite, though this is by no means obvious. The coordinate ring $\mathbb{Q}[x_{ij}]/(\det(x_{ij})-m)$ is a unique factorization domain [Sa], and we factor f into $t=t(f)$ irreducibles $f_1\dots f_t$ in this ring. We assume that the f_j's are distinct

A. N. was supported by the Institute for Advanced Study, Princeton, and ISF grant 975/05. P. S. was supported by an NSF grant and BSF grant 2006254.

and for simplicity that they are irreducible in $\overline{\mathbb{Q}}[x_{ij}]/(\det(x_{ij})-m)=\overline{\mathbb{Q}}[V_{m,n}]$. Clearly $r_0(\mathcal{O}^{m,n},f)\geqslant t$ and, if f and the f_j's have integer coefficients, then $r_0(\mathcal{O}^{m,n},f)=t$ if and only if the set of $x\in\mathcal{O}^{m,n}$ for which $f_j(x)$ are all prime is Zariski-dense in $V_{m,n}$. The general local-to-global conjectures in [BGS], when applied to the pair $(V_{m,n}(\mathbb{Z}),f)$, assert that $r_0(V_{m,n}(\mathbb{Z}),f)=t$. In the case that f and the f_j's are in $\mathbb{Z}[x_{ij}]$, we even expect a "prime number theorem" type of asymptotics as follows.

Let $|\cdot|$ be any norm on the linear space $\mathrm{Mat}_{n\times n}(\mathbb{R})$, and for $T\geqslant 1$ set

$$N_{m,n}(T)=|\{x\in\mathcal{O}^{m,n}:|x|\leqslant T\}|. \tag{1.1}$$

It is known [DRS], [Ma], [GW] that

$$N_{m,n}(T)\sim c(\mathcal{O}^{m,n})T^{n^2-n},\quad \text{as } T\to\infty. \tag{1.2}$$

Here c is a positive constant which is given as a product of local densities associated with $\mathcal{O}^{m,n}$, and in particular c depends also on the norm. Let

$$\pi_{m,n,f}(T)=|\{x\in\mathcal{O}^{m,n}:|x|\leqslant T \text{ and } f_j(x) \text{ is prime for } j=1,...,t\}|. \tag{1.3}$$

Our conjectured asymptotics for $\pi_{m,n,f}$ is then

$$\pi_{m,n,f}(T)\sim\frac{c(\mathcal{O}^{m,n},f)N_{m,n}(T)}{(\log T)^{t(f)}},\quad \text{as } T\to\infty, \tag{1.4}$$

where for a general set of integral points $\mathcal{O}$ we define

$$c(\mathcal{O},f)=c_\infty(\mathcal{O},f)\prod_{p<\infty}\left(1-\frac{|\mathcal{O}_p^f|}{|\mathcal{O}_p|}\right)\left(1+\frac{t(f)}{p}\right), \tag{1.5}$$

and $\mathcal{O}_p^f$ is the subset of $\mathcal{O}_p$ at which $f(x)=0 \pmod p$, while the positive Archimedean factor $c_\infty(\mathcal{O},f)$ is a bit more complicated to describe. We will see that the product of local densities in (1.5) converges absolutely and each factor is non-zero since we are assuming that f is weakly primitive.

The main tool that we develop in this paper is an affine linear sieve for homogeneous spaces and as in the more familiar classical 1-variable sieve [HR], our main results are upper bounds which are sharp up to a multiplicative constant for $\pi_{\mathcal{O},f}(T)$, and lower bounds which are also sharp up to a constant factor, for points $x\in\mathcal{O}$ for which f has at most a fixed number of large prime factors ("almost primes").

In particular, for the set $\mathcal{O}^{m,n}=V_{m,n}(\mathbb{Z})$ of integral $n\times n$ matrices of determinant m the upper bound is given by the following result.

THEOREM 1.1. *Let $V_{m,n}(\mathbb{Z})$ be as above and $f \in \mathbb{Z}[x_{ij}]$ be weakly primitive with $t(f)$ irreducible factors in both $\mathbb{Q}[V_{m,n}]$ and $\overline{\mathbb{Q}}[V_{m,n}]$. Then*

$$\pi_{m,n,f}(T) \ll \frac{N_{m,n}(T)}{(\log T)^{t(f)}},$$

the implied constant depending explicitly on m, n and f.

The lower bound is given by the following result.

THEOREM 1.2. *Let $V_{m,n}(\mathbb{Z})$ be as above and let $f \in \mathbb{Q}[x_{ij}]$ be weakly primitive and taking integer values on $V_{m,n}(\mathbb{Z})$. Assume that f has $t(f)$ distinct irreducible factors in both $\mathbb{Q}[V_{m,n}]$ and $\overline{\mathbb{Q}}[V_{m,n}]$. Let r be the least integer satisfying $r > 18t(f)n^3\eta_n \deg f$. Then*

$$\{x \in V_{m,n}(\mathbb{Z}) : |x| \leqslant T \text{ and } f(x) \text{ has at most } r \text{ prime factors}\} \gg \frac{N_{m,n}(T)}{(\log T)^{t(f)}}. \tag{1.6}$$

Here

$$\eta_n = \begin{cases} 1, & \text{if } n \text{ is odd}, \\ \dfrac{n}{n-1}, & \text{if } n \text{ is even}, \end{cases}$$

and again the implied positive constant depends explicitly on m, n and f.

COROLLARY 1.3. *Under the assumptions and notation in Theorem* 1.2, *the saturation number satisfies the upper bound $r_0(V_{m,n}(\mathbb{Z}), f) \leqslant r$. Namely, the set of $x \in V_{m,n}(\mathbb{Z})$ for which $f(x)$ has at most r prime factors, is Zariski-dense in $V_{m,n}$.*

In the case when $f(x) = \prod_{1 \leqslant i \leqslant j \leqslant n} x_{ij}$, Corollary 1.3 can be sharpened considerably. Exploiting the linearity of the determinant form in the rows and columns of a matrix, we use the method of Vinogradov [Vi] (see [Va]) for handling one linear equation in three or more prime variables to show that we can make all coordinates of the matrix simultaneously prime as long as there is no local obstruction.

THEOREM 1.4. *We have that*

$$f(x) = \prod_{1 \leqslant i \leqslant j \leqslant n} x_{ij}$$

is weakly primitive for $V_{m,n}(\mathbb{Z})$ if and only if $m \equiv 0 \pmod{2^{n-1}}$, and if this is the case and $n \geqslant 3$ then $r_0(V_{m,n}(\mathbb{Z}), f) = n^2$. That is, for $n \geqslant 3$ the set of $x \in V_{m,n}(\mathbb{Z})$ for which each x_{ij} is prime, is Zariski-dense in $V_{m,n}$ if and only if $m \equiv 0 \pmod{2^{n-1}}$.

Remark 1.5. (1) The proof of Theorem 1.4 provides a lower bound for $\pi_{m,n,f}(T)$ which is a power of T but not the one expected in conjecture (1.4).

(2) Theorem 1.4 should hold when $n=2$ but Vinogradov's methods do not apply in this binary case. For m fixed, the infinitude of x_{ij} satisfying $x_{11}x_{22}-x_{12}x_{21}=2m$ and x_{ij} prime, is apparently still open. Recent work of Goldston–Graham–Pintz–Yildrim [GGPY] on small differences of numbers which are products of exactly two primes, shows that the desired set is infinite for at least one number m in $\{1,2,3\}$.

(3) An immediate improvement to the upper estimate above for the value of r_0 arises by choosing the norm to be invariant under the rotation group. This improvement, together with much more significant ones, will be considered systematically in §6.

1.2. Prime and almost prime points on principal homogeneous spaces.

Theorems 1.1 and 1.2 are special cases of more general results which are concerned with finding points on an orbit $\mathcal{O}$ of $v\in\mathbb{Z}^n$ under a subgroup Γ of $\mathrm{GL}_n(\mathbb{Z})$ at which a given polynomial $f\in\mathbb{Q}[x_1,...,x_n]$, which is integral on $\mathcal{O}$, has few prime factors. The approach is based on the "affine linear sieve" introduced recently in [BGS]. Our purpose here is to specialize to Γ being a congruence subgroup of an algebraically simply connected semisimple linear algebraic group $\mathbb{G}\subset\mathrm{GL}_n$ defined over $\mathbb{Q}$ and for which the stabilizer of v in $\mathbb{G}$ is trivial. This allows us to make use of the well-developed analytic methods [DRS], [GN1] for counting points in such orbits in a big Euclidean ball, as well as the strong bounds towards the general Ramanujan conjectures that are known from the theory of automorphic forms (see [Cl1] and [S1]). We assume further that $\mathbb{G}$ is of non-compact type, that is the group of real points of any $\mathbb{Q}$-factor of $\mathbb{G}$ is non-compact. This is needed to ensure that there are enough $\mathbb{G}(\mathbb{Z})$-points for our purposes. We restrict further to principal homogeneous spaces (i.e. the stabilizer of v being trivial) and to $\mathbb{G}$ being algebraically simply connected. Note however that the last two restrictions are not serious ones, as far as the production of a Zariski-dense set of points x at which $f(x)$ has few prime factors is concerned. As explained in [BGS], the dominant $\mathbb{Q}$-morphisms from $\mathbb{G}$ to an orbit $\mathbb{G}/\mathbb{H}$ and from the simply connected cover $\widetilde{\mathbb{G}}$ of $\mathbb{G}$, to $\mathbb{G}$, reduce (by pull-back) the basic saturation problem for orbits of more general congruence groups to the cases that we consider in this paper.

We now describe our results in more detail. Let $\mathbb{G}\subset\mathrm{GL}_n(\mathbb{C})$ be a connected and algebraically simply connected semisimple algebraic matrix group defined over $\mathbb{Q}$. Let $\mathbb{G}(\mathbb{Q})$ be its rational points and $\Gamma=\mathbb{G}(\mathbb{Z})=\mathbb{G}(\mathbb{Q})\cap\mathrm{GL}_n(\mathbb{Z})$ its integral points. Fix $v\in\mathbb{Z}^n$ and let $V=\mathbb{G}v$ be the corresponding orbit which we assume is Zariski-closed in A^n. Since $\mathbb{G}$ is algebraically connected and the stabilizer of v is assumed to be trivial, V is

an absolutely irreducible affine variety defined over $\mathbb{Q}$ and has dimension equal to $\dim \mathbb{G}$. The ring of $\mathbb{G}$-invariants for the action of $\mathbb{G}$ on the n-dimensional space A^n separates the closed $\mathbb{G}$-orbits [BH]. We can choose generators $h_1, ..., h_\nu$ in $\mathbb{Q}[x_1, ..., x_n]$ of this ring so that V is given by

$$V = \{x : h_j(x) = \lambda_j, j = 1, ..., \nu\}, \quad \text{with } \lambda_j \in \mathbb{Q}. \tag{1.7}$$

Let $\mathcal{O}=\Gamma v$ be the Γ-orbit of v in $\mathbb{Z}^n$. According to [B], $\mathcal{O}$ is Zariski-dense in V. The coordinate ring of V, $\mathbb{Q}[x_1, ..., x_n]/(h_1-\lambda_1, ..., h_\nu-\lambda_\nu)$ is a unique factorization domain [Sa]. Hence an $f\in\mathbb{Q}[x_1, ..., x_n]$ factors into irreducibles $f=f_1 ... f_t$ in this ring. We assume that these f_j's are distinct and that they are irreducible in

$$\overline{\mathbb{Q}}[x_1, ..., x_n]/(h_1-\lambda_1, ..., h_\nu-\lambda_\nu),$$

that f takes integer values on $\mathcal{O}$ and that it is $\mathcal{O}$-weakly primitive. The saturation number $r_0(\mathcal{O}, f)$ of the pair $(\mathcal{O}, f)$ is the least r such that the set of $x\in\mathcal{O}$ for which $f(x)$ has at most r prime factors is Zariski-dense in $V(=\mathrm{Zcl}(\mathcal{O}))$.

To order the elements of $\mathcal{O}$ we use the following "height" functions. Let $\|\cdot\|$ be any norm on the linear space $\mathrm{Mat}_{n\times n}(\mathbb{R})$. For $T>0$ and $v\in\mathbb{Q}^n$,

$$\mathcal{O}(T) = \{\gamma v : \|\gamma\| \leqslant T \text{ and } \gamma \in \Gamma\} \tag{1.8}$$

(this depends on v but in an insignificant way).

In many interesting cases the sets $\mathcal{O}(T)$ can be described as $\{x\in\mathcal{O}:|x|\leqslant T\}$, where $|\cdot|$ is a norm on $\mathbb{R}^n$, but this is not true in general. The main term of the asymptotics for $N_\mathcal{O}(T)=|\mathcal{O}(T)|$ is known for any norm ([GW], [Ma]) and takes the form

$$N_\mathcal{O}(T) \sim B_\mathcal{O}(\|\cdot\|)T^a(\log T)^b \tag{1.9}$$

with $B_\mathcal{O}(\|\cdot\|)>0$, $a>0$ and $b\in\mathbb{N}$ being a non-negative integer. The numbers a and b are given explicitly in terms of the data in Theorem 3.1 below (see the discussion following that theorem). They do not depend on the norm chosen, unlike $B_\mathcal{O}(\|\cdot\|)$ which does.

THEOREM 1.6. *Let $\mathcal{O}$ and f be as above and assume that f_j, $j=1, ..., t(f)$, are integral on $\mathcal{O}$. Then, for $T\geqslant 2$,*

$$|\{x \in \mathcal{O}(T) : f_j(x) \text{ is prime for } j = 1, ..., t(f)\}| \ll \frac{N_\mathcal{O}(T)}{(\log T)^{t(f)}},$$

the implied constant depending on f and $\mathcal{O}$.

THEOREM 1.7. *Let $\mathcal{O}$ and f be as above and let r be the least integer satisfying the condition $r>(9t(f)(1+\dim G)^2 2n_e(\Gamma)\deg f)/a$, where $n_e(\Gamma)$ is the integer defined following Theorem* 3.3. *Then, as $T\to\infty$,*

$$|\{x\in\mathcal{O}(T): f(x) \text{ has at most } r \text{ prime factors}\}| \gg \frac{N_{\mathcal{O}}(T)}{(\log T)^{t(f)}},$$

the implied constant depending on f and $\mathcal{O}$.

COROLLARY 1.8. *Let $\mathcal{O}$ and f be as in Theorem* 1.7. *Then $r_0(\mathcal{O},f)\leqslant r$.*

Remark 1.9. (i) The integer $n_e(\Gamma)$ is at least 1 and is determined by the extent to which the representation spaces $L^2_0(G/\Gamma(q))$ weakly contain non-tempered irreducible representations of $G=\mathbb{G}(\mathbb{R})$. Here $L^2_0(G/\Gamma(q))$ is the space of functions with zero integral, and $\Gamma(q)$ is any congruence subgroup of Γ. The non-temperedness is measured by the infimum over all $p>0$ for which the representation space contains a dense subspace of matrix coefficients belonging to $L^p(G)$. Thus $n_e(\Gamma)$ is directly connected to the generalized Ramanujan conjectures for $\mathbb{G}(\mathbb{A})/\mathbb{G}(\mathbb{Q})$ [S1].

(ii) Theorems 1.1 and 1.2 and Corollary 1.3 are connected to the general Theorems 1.6 and 1.7 and Corollary 1.8 as follows: $\Gamma=\mathrm{SL}_n(\mathbb{Z})$ acts on $V_{n,m}(\mathbb{Z})$ by left multiplication. This action has finitely many orbits. Set $\mathbb{G}=\mathrm{SL}_n\subset\mathrm{GL}_M$, $M=n^2$, and (with this action) $\mathcal{O}=\mathbb{G}v$, where $\det v=m$. Theorem 1.1 then follows by applying Theorem 1.6 to each orbit separately. For Theorem 1.2, the difference between the individual orbit $\mathcal{O}$ of $\mathrm{SL}_n(\mathbb{Z})$ and all of $V_{m,n}(\mathbb{Z})$ raises the issue of the weak primitivity of f on $V_{m,n}(\mathbb{Z})$. So one needs to globalize the argument as is explained in §4.

(iii) The assumption of absolute irreducibility of the factors f_j will be used in §4.1, when we estimate the number of their solutions modulo a prime p. However, this assumption is made for convenience and it is possible to discard it by passing to a finite extension field, along the lines of the argument used in §4.3.

The upper estimate of r_0 given in Theorem 1.7 and Corollary 1.8 is by no means optimal. There are various places where the analysis can be modified to give a far better bound. First, by using smooth positive weights instead of the sharp cutoff counting function in (2.9) and Theorem 3.2, we can improve the level of distribution τ in (4.15). We carry this out in §6 for the case when Γ is co-compact in $G=\mathbb{G}(\mathbb{R})$. In Theorem 6.1 we obtain the sharpest possible remainder for such smooth sums in terms of bounds towards the Ramanujan conjectures. We call this smooth weighted sum formula for K-bi-invariant metrics a *Poisson summation formula*. It constitutes the main ingredient of the spectral method for counting integral points in the orbit $\mathcal{O}$. This leads to the improvement in the upper estimate for r_0 that is given in (6.6). A further improvement is gotten by using

a weighted sieve ([HR], [DH]) rather than the simple sieve from §2. This leads to the improved estimate for r_0 given in (6.16). Finally there are cases such as the following for which very strong bounds on the spectrum of $L_0^2(G/\Gamma(q))$ are known and which result in quite a good estimate for r_0. Let D be a division algebra over $\mathbb{Q}$ of degree n (which for technical reasons we assume is prime) and for which $D\otimes\mathbb{R}\cong \mathrm{Mat}_{n\times n}(\mathbb{R})$. Then $D(\mathbb{Q})$ has dimension $M=n^2$ over $\mathbb{Q}$, and choosing a basis gives a $\mathbb{Z}$-structure for D, that is M coordinates x_{ij}. Consider the reduced norm and let $\mathbb{G}$ be the linear algebraic group of elements of reduced norm 1. Let $\Gamma=\mathbb{G}(\mathbb{Z})\subset \mathrm{GL}_M$, where the action is by multiplication on the left. Let $\mathcal{O}=\Gamma v$, where $v\in D(\mathbb{Z})$ has reduced norm $m\neq 0$, and let $f\in\mathbb{Q}[x_{ij}]$ be integral on $\mathcal{O}$ and weakly primitive. Then all the improvements mentioned above apply to the pair $(\mathcal{O}, f)$, and Theorem 1.7 and Corollary 1.8 apply with the following estimate for r_0 (see Theorem 6.4)

$$r_0(\mathcal{O}, f)\leqslant 6\deg(f)+t(f)\log t(f). \tag{1.10}$$

This bound is independent of the dimension and it is of the same quality and shape, in terms of dependence on the degree of f and the number of its irreducible factors, as what is known for r-almost primes for values of $f(x)$ in the classical case of one variable [HR].

Uniform bounds such as those in (1.10) which are independent of the dimension are useful when combined with $\mathbb{Q}$-morphisms. Let $\phi\colon \mathbb{G}\to A^k$ be a $\mathbb{Q}$-morphism of affine varieties for which $\mathcal{O}=\phi(\mathbb{G}(\mathbb{Z}))\subset\mathbb{Z}^k$. Then, if $f\in\mathbb{Q}[x_1, ..., x_k]$ is $\mathcal{O}$-integral and weakly primitive, we have that $\phi^*(f)=f\circ\phi$ is $\mathbb{G}(\mathbb{Z})$-integral and weakly primitive. Moreover, $r_0(\mathcal{O}, f)\leqslant r_0(\mathbb{G}(\mathbb{Z}), \phi^*(f))$. If $\mathcal{O}$ is part of a larger set of integral points that can be swept out by varying ϕ suitably, then the uniformity allows one to give bounds for saturation numbers for the larger set. This of course applies also with $\mathbb{G}=A^1$, in which case one can apply the classical 1-variable sieve. For example, Corollary 1.3 can be approached by this more elementary method. Let $y\in V_{m,n}(\mathbb{Z})$ and let $\phi_y\colon A^1\to V_{m,n}$ be the morphism

$$\phi_y\colon x\longmapsto \begin{bmatrix} 1 & 0 & \dots & 0 & x \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} y.$$

Then $\mathcal{O}=\phi_y(\mathbb{Z})\subset V_{m,n}(\mathbb{Z})$. Apply the classical 1-variable results about almost primes to the pair $(\mathbb{Z}, \phi^*(f))$. For a generic y, the bound for r_0 depends only on t and d, and the set of such y's is Zariski-dense in $V_{m,n}$. In this way one can establish Corollary 1.3 with r_0 comparable to (1.10) above. This approach is possible whenever G has $\mathbb{Q}$-unipotent

elements. However, in the opposite case when $\mathbb{G}(\mathbb{R})/\mathbb{G}(\mathbb{Z})$ is compact, such as in the division algebra examples above, there are no such $\mathbb{Q}$-rational parametric affine curves in $G=\mathbb{G}(\mathbb{R})$, and the general affine linear sieve developed in this paper is the only approach that we know of to obtain Corollary 1.8. In [LS] this technique is developed for anisotropic quadratics in three variables.

2. The combinatorial sieve

To begin with we make use of a simple version of the combinatorial sieve, see [IK, §§6.1–6.4] and [HR, Theorem 7.4]. Later we will use more sophisticated versions. Our formulation is tailored to the applications.

Let $\mathcal{A}=\{a_k\}_{k\geqslant 1}$ be a finite sequence of non-negative numbers. Denote by X the sum

$$X=\sum_k a_k. \tag{2.1}$$

We think of X as large, tending to infinity as the number of elements of $\mathcal{A}$ increases. Fix a finite set B of "ramified" primes which for the most part will be the empty set. For z a large parameter (in applications z will be a small power of X), set

$$P=P_{z,B}=\prod_{\substack{p\leqslant z\\ p\notin B}} p. \tag{2.2}$$

Under suitable assumptions about the sums of the a_k's in progressions with moderately large moduli, the sieve gives upper and lower estimates which are of the same order of magnitude, for the sums of $\mathcal{A}$ over k's which remain after sifting out numbers with prime factors in P.

More precisely, let

$$S(\mathcal{A},P):=\sum_{(k,P)=1} a_k. \tag{2.3}$$

The assumptions for the sums over progressions that we make are as follows:

(A_0) For d square-free and having no prime factors in B ($d<X$), we assume that the sum over multiples of d takes the form

$$\sum_{k\equiv 0 \pmod d} a_k=\frac{\varrho(d)}{d}X+R(\mathcal{A},d), \tag{2.4}$$

where $\varrho(d)$ is multiplicative in d and, for $p\notin B$,

$$0\leqslant\frac{\varrho(p)}{p}\leqslant 1-\frac{1}{c_1}<1. \tag{2.5}$$

The understanding being that $(\varrho(d)/d)X$ is the main term in (2.4) and $R(\mathcal{A},d)$ is smaller, at least on average.

(A_1) $\mathcal{A}$ has level distribution $D=D(X)$, that is for some $\varepsilon>0$ there is $C_\varepsilon<\infty$ such that

$$\sum_{d\leqslant D}|R(\mathcal{A},d)|\leqslant C_\varepsilon X^{1-\varepsilon}. \tag{2.6}$$

If this holds with $D=X^\tau$, we say that the level distribution is τ.

(A_2) $\mathcal{A}$ has sieve dimension t, that is there is C_2 fixed such that

$$-C_2\leqslant\sum_{\substack{(p,B)=1\\ w\leqslant p\leqslant z}}\frac{\varrho(p)\log p}{p}-t\log\frac{z}{w}\leqslant C_2\quad\text{for } 2\leqslant w\leqslant z. \tag{2.7}$$

Assuming (A_0), (A_1) and (A_2), the simple combinatorial sieve that we use asserts that for $s>9t$, $z=D^{1/s}$ and X large enough, one has

$$\frac{X}{(\log X)^t}\ll S(\mathcal{A},P)\ll\frac{X}{(\log X)^t}, \tag{2.8}$$

where the implied constants depend explicitly on t, ε, C_1, C_2 and C_ε (all of which are fixed).

For our application, $V\subset A^n$ is a principal homogeneous space for G and $\mathcal{O}=\mathbb{G}(\mathbb{Z})v$ is an orbit of integral points in V. The polynomial $f\in\mathbb{Q}[x_1,\dots x_n]$ is integral on $\mathcal{O}$ and $\|\cdot\|$ is the norm described in §1. For $k\geqslant 0$, we let $a_k=a_k(T)$ be the number of elements $\gamma\in\mathbb{G}(\mathbb{Z})$ with norm bounded by T, for which $|f(\gamma v)|=k$, namely

$$a_k(T):=\sum_{\substack{\|\gamma\|\leqslant T\\ |f(\gamma v)|=k}}1. \tag{2.9}$$

Under the assumptions (2.4), (2.6) and (2.7) on the level distribution, it follows that

$$a_0(T)\ll\frac{1}{D}X\log X, \tag{2.10}$$

where D is the level and $X=\sum_{k\in\mathbb{N}}a_k(T)$. Hence, for our purposes, we may include $k=0$ into the sieve analysis without affecting (2.8).

A large part of this paper is concerned with verifying (A_0), (A_1) and (A_2) for the sequence $a_k(T)$, and determining an admissible level of distribution.

3. Uniform lattice point count

In this and the next section we will identify the main term in the asymptotics of

$$\sum_{k\equiv 0 \pmod d} a_k(T)$$

(condition (A_0)), as well as estimate the level of distribution D (condition (A_1)). In §4.1 we will establish the multiplicativity of $\varrho(d)/d$, the coefficient of the main term, concluding the proof of (A_0) and (A_1). The most demanding part is establishing an explicit bound on the error terms $\sum_{d\leqslant D}|R(\mathcal{A},d)|$ appearing in condition (A_1). Here the basic ingredient will be an error estimate for the lattice points counting problem which is uniform over all cosets of all congruence groups. This will be established in §3.2 and §3.3.

3.1. Spectral estimates

Let $\mathbb{G}\subset \mathrm{GL}_n(\mathbb{C})$ be a connected semisimple algebraic matrix group defined over $\mathbb{Q}$, with $G=\mathbb{G}(\mathbb{R})$ having no non-trivial compact factors. Fix any norm on the linear space $M_n(\mathbb{R})$. Let $\Gamma(1)=\mathbb{G}(\mathbb{Z})$ be the group of integral points, which is a lattice subgroup of G [BH] and a subgroup of $\mathrm{GL}_n(\mathbb{Z})$. Let $\Gamma(q)$, $q\in\mathbb{N}$, denote the principal congruence subgroup of level q, namely

$$\Gamma(q)=\{\gamma\in\Gamma:\gamma\equiv I \pmod q\}.$$

We begin by stating the following volume asymptotic, established in [GW] and [Ma].

THEOREM 3.1. *Let $\mathbb{G}$ and G be as above and let $\|\cdot\|$ be any norm on $M_n(\mathbb{R})$. Then there exists $a>0$ and a non-negative integer b, both depending only on G, and a constant $B(\|\cdot\|)$, depending also on the norm, such that*

$$\lim_{T\to\infty}\frac{\mathrm{vol}\{g\in G:\|g\|\leqslant T\}}{B(\|\cdot\|)T^a(\log T)^b}=1.$$

The exponent a has the following simple algebraic description [GW], [Ma]. Let A denote a maximally $\mathbb{R}$-split Cartan subgroup of $G=\mathbb{G}(\mathbb{R})$, with Lie algebra $\mathfrak{a}$. Let $\mathcal{C}$ denote the convex hull of the weights of $\mathfrak{a}$ associated with the representation of G in $\mathrm{GL}_n(\mathbb{R})$. Let $2\varrho_G$ denote the sum of all the positive roots (counted with multiplicities) of $\mathfrak{a}$. Then a is the unique positive real number with the property that $2\varrho_G/a\in\partial\mathcal{C}$. The parameter $b+1$ is a positive integer, which is at most the $\mathbb{R}$-rank of G, and is equal to the codimension of a minimal face of the polyhedron $\mathcal{C}$ containing the point $2\varrho_G/a$. In particular, both a and b are independent of the norm.

We now state the following uniform error estimate for the lattice point counting problem, which underlies our estimate of the level of distribution.

THEOREM 3.2. *Let* $\mathbb{G}$, G, $\Gamma(q)$ *and* $\|\cdot\|$ *be as above and normalize Haar measure on* G *so that* $\mathrm{vol}(G/\Gamma(1))=1$. *Then the following uniform error estimate holds* (*for any* $\eta>0$):

$$\frac{|\{w\in\Gamma(q)y:\|w\|\leqslant T\}|}{\mathrm{vol}\{g\in G:\|g\|\leqslant T\}}=\frac{1}{[\Gamma:\Gamma(q)]}+O_\eta(T^{-\theta/(1+\dim G)+\eta}),$$

where

(i) $\theta=a/2n_e(G,\Gamma)>0$ *is explicit and depends only on the bounds towards the generalized Ramanujan conjecture for the homogeneous spaces* $G/\Gamma(q)$ (*see Theorem* 3.3 *below*),

(ii) *the estimate holds uniformly over all cosets of all congruence subgroups* $\Gamma(q)$ *in* $\Gamma(1)$, *namely the implied constant is independent of* q *and* $y\in\Gamma(1)$ (*it depends only on* η *and the chosen norm*).

A general approach to the lattice point counting problem with error estimate for S-algebraic groups is developed in [GN1], based on establishing quantitative mean ergodic theorems for the Haar-uniform averages supported on the norm balls. In §3.2 and §3.3 we follow this approach and give a short proof of Theorem 3.2, thus establishing the uniformity of the error term, which is crucial for our considerations. More general results can be found in [GN2].

Let $B_t\subset G$, $t\in\mathbb{R}_+$, denote the family of subsets

$$B_t=\{g\in G:\|g\|\leqslant e^t\},$$

and let $\pi_{G/\Gamma}(\beta_t)$ denote the following averaging operator acting in $L^2(G/\Gamma)$:

$$\pi_{G/\Gamma}(\beta_t)f(x)=\int_{g\in B_t}f(g^{-1}x)\,dm_G.$$

The subspace $L^2_0(G/\Gamma)$ of functions of zero mean is obviously a G-invariant subspace, and the representation there is denoted by $\pi^0_{G/\Gamma}$.

The fundamental spectral estimate in our discussion is given by the following result.

THEOREM 3.3. *Let* $\mathbb{G}$, G, $\Gamma(q)$ *and* B_t *be as above. Then, for* $\theta=a/2n_e(G,\Gamma)>0$, *uniformly for every* $q\in\mathbb{N}$,

$$\left\|\pi_{G/\Gamma(q)}(\beta_t)f-\int_{G/\Gamma(q)}f\,dm_{G/\Gamma(q)}\right\|_{L^2(G/\Gamma(q))}\leqslant C_\eta e^{-(\theta-\eta)t}\|f\|_{L^2(G/\Gamma(q))},$$

where $m_{G/\Gamma(q)}$ *is the* G*-invariant probability measure on* $G/\Gamma(q)$. *Here* $\eta>0$ *is arbitrary*, a *is the exponent in the rate of exponential volume growth of* B_t *and* $n_e(G,\Gamma)$ *is the least even integer* $\geqslant\frac{1}{2}p(G,\Gamma)$, *with* $p(G,\Gamma)$ *being the bound towards the Ramanujan conjecture described in the proof.*

Proof. The proof of Theorem 3.3 consists of two parts. The first part is to note that the bounds towards the generalized Ramanujan conjecture (see [Cl2] and [S1]) imply that there exists an explicit $p=p(G,\Gamma)$ with the property that all the (K-finite) matrix coefficients occurring in $L^2_0(G/\Gamma(q))$ are in $L^{p+\eta}(G)$ for all $\eta>0$, and all $q\in\mathbb{N}$. The second is to use this information to give an explicit estimate for θ.

For the first part, let us note that our lattice $\Gamma=\mathbb{G}(\mathbb{Z})$ is an irreducible lattice, and as a result, in the unitary representation of G in $L^2_0(G/\Gamma)$, the matrix coefficients decay to zero as $g\to\infty$ in G, namely the representation is strongly mixing (recall that we assume that G has no non-trivial compact factors). In particular, if H is an almost-simple component of G, then H has no invariant unit vectors in $L^2_0(G/\Gamma(q))$. We now divide the argument into two cases.

(1) Assume that H has property T. Then there exists $p=p_H$, such that *every* unitary representation of H without H-fixed unit vectors has its (K-finite) matrix coefficients in $L^{p+\eta}$ for all $\eta>0$ [Co]. Thus we can use as a bound for the automorphic spectrum a bound valid for all unitary representations of H. Therefore, in this case the integrability parameter p above depends only on H but not on Γ. In particular, the bound holds in $L^2_0(G/\Gamma(q))$ for all finite-index subgroups $\Gamma(q)$, including the principal congruence subgroups. However, note that for some lattices Γ one can do better; this applies in particular to certain uniform arithmetic lattices, as will be discussed further in §6 below.

For a list of p_H for classical simple groups H with property T, we refer to [Li], and for exceptional groups, to [LiZ] and [LoS] (see also [O]). Thus, for example, $p_{\mathrm{SL}_n(\mathbb{R})}=2(n-1)$, $n\geqslant 3$, and $p_{\mathrm{Sp}(n,\mathbb{R})}=2n$, $n\geqslant 2$.

(2) G is defined over $\mathbb{Q}$, and so are its simple component subgroups. Let now H be a simple algebraic $\mathbb{Q}$-subgroup of real rank 1 which does not satisfy property T, and let $\Gamma=\mathbb{G}(\mathbb{Z})$. Then there still exists $p=p(H,\Gamma)$ such that the set of representations of H obtained as restrictions from the representations of G on $L^2_0(G/\Gamma(q))$ have their (K-finite) matrix coefficients in $L^{p+\eta}(H)$, $\eta>0$, for all $q\in\mathbb{N}$. This fact is established in most cases in [BS], and in the missing cases by [Cl1]. These results yield the following explicit estimate. Let $\varrho_H=\frac{1}{2}(m_1+2m_2)$, where m_1 (resp. m_2) is the multiplicity of the short (resp. long) root in the root system associated with a maximal $\mathbb{R}$-split torus in H. The parameter s of a non-trivial complementary series representation π_s that can occur in the automorphic spectrum of H is constrained to satisfy $s\leqslant\varrho_H-\frac{1}{4}$. Now the volume density on H in radial coordinates is comparable to $e^{2\varrho_H t}$, and the decay of the spherical function φ_s is comparable to $e^{-t/4}$, so that the matrix coefficients are in $L^{p+\eta}(H)$, where $p=8\varrho_H=4(m_1+2m_2)$.

Finally, to conclude the first part of the proof, note that for $p=p(G,\Gamma)$, we may take the maximum of $p(H,\Gamma)$ as H ranges over the almost-simple normal $\mathbb{Q}$-subgroups, since

any (normalized) matrix coefficient has absolute value bounded by 1.

The second part of the proof consists of showing how to derive an explicit estimate for the decay of the operator norms of $\pi^0_{G/\Gamma(q)}(\beta_t)$ from the bound on the automorphic spectrum. Let $p{=}p(G,\Gamma)$ be the minimum value such that every strongly mixing unitary representation weakly contained in $L^2_0(G/\Gamma)$ has its (K-finite) matrix coefficients in $L^{p+\eta}(G)$ for all $\eta{>}0$. Recall that we define $n_e{=}n_e(\Gamma)$ as the least even integer greater than or equal to $\frac{1}{2}p(G,\Gamma)$. By [Ne, Theorem 1],

$$\|\pi^0_{G/\Gamma}(\beta_t)\|_{L^2_0(G/\Gamma)} \leqslant \|\lambda_G(\beta_t)\|^{1/n_e}_{L^2(G)},$$

where λ_G is the regular representation of G on $L^2(G)$. Now, following [Ne, Theorem 4], by the Kunze–Stein phenomenon the norm of the convolution operator $\lambda_G(\beta_t)$ determined by β_t on $L^2(G)$ is bounded by $C'_\eta \operatorname{vol}(B_t)^{-1/2+\eta}$. Taking the volume asymptotics of B_t stated in Theorem 3.1 into account, we conclude that $\theta{=}a/2n_e$ gives the norm bound stated in Theorem 3.3. □

3.2. Averaging operators and counting lattice points

We now turn to the proof of Theorem 3.2, and begin by explicating the connection between the averaging operators associated with β_t and counting lattice points, following the method developed in [GN1].

Consider a bi-K-invariant Riemannian metric on G covering the Riemannian metric associated with the Cartan–Killing form on the symmetric spaces $S{=}G/K$, and let d denote the distance function. Let vol denote the Haar measure defined by the volume form associated with the Riemannian metric. Define

$$O_\varepsilon = \{g \in G : d(g,e) < \varepsilon\}.$$

Recall that $B_t{\subset}G$, $t{\in}\mathbb{R}_+$, is the family of subsets

$$B_t = \{g \in G : \|g\| \leqslant e^t\}.$$

The sets B_t enjoy the following stability and regularity properties.

PROPOSITION 3.4. ([GN1, Theorem 3.15]) *The family B_t is admissible, namely there exist $c{>}0$, $\varepsilon_0{>}0$ and $t_0{>}0$ such that for all $t{\geqslant}t_0$ and $0{<}\varepsilon{<}\varepsilon_0$,*

$$O_\varepsilon B_t O_\varepsilon \subset B_{t+c\varepsilon} \tag{3.1}$$

and

$$\operatorname{vol}(B_{t+\varepsilon}) \leqslant (1{+}c\varepsilon)\operatorname{vol}(B_t). \tag{3.2}$$

Given the lattice $\Gamma=\Gamma(1)=\mathbb{G}(\mathbb{Z})$, fix the unique invariant volume form $\mathrm{vol}=\mathrm{vol}_G$ on G satisfying $\mathrm{vol}(G/\Gamma)=1$. We denote by $\mathrm{vol}_{G/\Gamma(q)}$ the volume form induced on $G/\Gamma(q)$ by the volume form on G. Then $\mathrm{vol}_{G/\Gamma(q)}(G/\Gamma(q))=[\Gamma:\Gamma(q)]$ is the total volume of the locally symmetric space $G/\Gamma(q)$. We also let $m_{G/\Gamma(q)}$ denote the corresponding probability measure on $G/\Gamma(q)$, namely $\mathrm{vol}_{G/\Gamma(q)}/[\Gamma:\Gamma(q)]$.

We note that clearly $\frac{1}{2}\varepsilon^{\dim G}\leqslant\mathrm{vol}(O_\varepsilon)\leqslant 2\varepsilon^{\dim G}$ for $0<\varepsilon\leqslant\varepsilon_0'$. Let

$$\chi_\varepsilon=\frac{\chi_{O_\varepsilon}}{\mathrm{vol}(O_\varepsilon)}.$$

We now fix a congruence subgroup $\Gamma(q)\subset\Gamma=\Gamma(1)$, and define, for every $y\in\Gamma(1)$,

$$\phi_\varepsilon^y(g\Gamma(q))=\sum_{\gamma\in\Gamma(q)}\chi_\varepsilon(g\gamma y).$$

Thus ϕ_ε^y is a measurable bounded function on $G/\Gamma(q)$ with compact support, and

$$\int_G\chi_\varepsilon\,d\mathrm{vol}=1,$$

so that

$$\int_{G/\Gamma(q)}\phi_\varepsilon^y\,d\mathrm{vol}_{G/\Gamma(q)}=1,$$

and

$$\int_{G/\Gamma(q)}\phi_\varepsilon^y\,dm_{G/\Gamma(q)}=\frac{1}{[\Gamma:\Gamma(q)]}.$$

Clearly, for any $\delta>0$, $h\in G$ and $t\in\mathbb{R}_+$, the following are equivalent (for any function on $G/\Gamma(q)$):

$$\left|\pi_{G/\Gamma}(\beta_t)\phi_\varepsilon^y(h\Gamma(q))-\frac{1}{[\Gamma:\Gamma(q)]}\right|\leqslant\delta, \tag{3.3}$$

and

$$\frac{1}{[\Gamma:\Gamma(q)]}-\delta\leqslant\frac{1}{\mathrm{vol}(B_t)}\int_{B_t}\phi_\varepsilon^y(g^{-1}h\Gamma(q))\,d\mathrm{vol}(g)\leqslant\frac{1}{[\Gamma:\Gamma(q)]}+\delta. \tag{3.4}$$

The set where the first inequality holds will be estimated using the quantitative mean ergodic theorem. The integral in the second expression is connected to lattice points as follows.

LEMMA 3.5. *For every $t\geqslant t_0+c\varepsilon_0$, $0<\varepsilon\leqslant\varepsilon_0$, and for every $h\in\mathcal{O}_\varepsilon$,*

$$\int_{B_{t-c\varepsilon}}\phi_\varepsilon^y(g^{-1}h\Gamma(q))\,d\mathrm{vol}(g)\leqslant|B_t\cap\Gamma(q)y|\leqslant\int_{B_{t+c\varepsilon}}\phi_\varepsilon^y(g^{-1}h\Gamma(q))\,d\mathrm{vol}(g).$$

Proof. If $\chi_\varepsilon(g^{-1}h\gamma y)\neq 0$ for some $g\in B_{t-c\varepsilon}$, $h\in O_\varepsilon$ and $\gamma y\in\Gamma(q)y$, then, by (3.1),

$$\gamma y\in h^{-1}B_{t-c\varepsilon}(\operatorname{supp}\chi_\varepsilon)\subset B_t.$$

Hence,

$$\int_{B_{t-c\varepsilon}}\phi_\varepsilon^y(g^{-1}h\Gamma(q))\,d\mathrm{vol}(g)\leqslant\sum_{\gamma y\in B_t\cap\Gamma(q)y}\int_{B_t}\chi_\varepsilon(g^{-1}h\gamma y)\,d\mathrm{vol}(g)\leqslant|B_t\cap\Gamma(q)y|.$$

In the other direction, for $\gamma y\in B_t\cap\Gamma(q)y$ and $h\in O_\varepsilon$,

$$\operatorname{supp}(g\mapsto\chi_\varepsilon(g^{-1}h\gamma y))=h\gamma y(\operatorname{supp}\chi_\varepsilon)^{-1}\subset B_{t+c\varepsilon},$$

and since $\chi_\varepsilon\geqslant 0$, again by (3.1),

$$\int_{B_{t+c\varepsilon}}\phi_\varepsilon^y(g^{-1}h\Gamma(q))\,d\mathrm{vol}(g)\geqslant\sum_{\gamma y\in B_t\cap\Gamma(q)y}\int_{B_{t+c\varepsilon}}\chi_\varepsilon(g^{-1}h\gamma y)\,d\mathrm{vol}(g)\geqslant|B_t\cap\Gamma(q)y|.\quad\square$$

3.3. Uniform error estimates for congruence groups

We now complete the proof of Theorem 3.2, using the method developed in [GN1, §6.6].

For the lattices $\Gamma(q)$, the action of the operators $\pi_{G/\Gamma(q)}(\beta_t)$ on $L^2_0(G/\Gamma(q))$ satisfies the spectral estimate stated in Theorem 3.3, uniformly in q. It follows that for the probability spaces $(G/\Gamma(q),m_{G/\Gamma(q)})$ we have, for all $t>0$ and every $\theta'<\theta$,

$$\left\|\pi_{G/\Gamma(q)}(\beta_t)\phi_\varepsilon^y-\int_{G/\Gamma(q)}\phi_\varepsilon^y\,dm_{G/\Gamma(q)}\right\|_{L^2(m_{G/\Gamma(q)})}\leqslant C_{\theta'}e^{-\theta' t}\|\phi_\varepsilon^y\|_{L^2(m_{G/\Gamma(q)})}.$$

Therefore, for all $\delta>0$, all $t>0$ and $\varepsilon<\varepsilon_0'$,

$$m_{G/\Gamma(q)}\left\{h\Gamma(q):\left|\pi_{G/\Gamma(q)}(\beta_t)\phi_\varepsilon^y(h\Gamma(q))-\frac{1}{[\Gamma:\Gamma(q)]}\right|>\delta\right\}\leqslant C_{\theta'}^2\delta^{-2}e^{-2\theta' t}\|\phi_\varepsilon^y\|^2_{L^2(m_{G/\Gamma(q)})}.$$

Clearly, we can fix ε_0'' such that if $\varepsilon<\varepsilon_0''$ then the translates $O_\varepsilon w$ are disjoint for distinct $w\in\Gamma(1)$. Then the supports of the functions $\chi_\varepsilon(h\gamma y)$ for $\gamma\in\Gamma(q)$ (and a fixed $y\in\Gamma(1)$) do not intersect, and so

$$\begin{aligned}\|\phi_\varepsilon^y\|^2_{L^2(m_{G/\Gamma(q)})}&=\int_{G/\Gamma(q)}\phi_\varepsilon^y(h\Gamma(q))^2\,d\mathrm{vol}_{G/\Gamma(q)}/[\Gamma:\Gamma(q)]\\&=\int_G\chi_\varepsilon^2(g)\,d\mathrm{vol}(g)/[\Gamma:\Gamma(q)]=\frac{1}{\mathrm{vol}(O_\varepsilon)[\Gamma:\Gamma(q)]}\leqslant\frac{2\varepsilon^{-\dim G}}{[\Gamma:\Gamma(q)]}.\end{aligned}$$

We conclude that

$$m_{G/\Gamma(q)}\left\{h\Gamma(q):\left|\pi_{G/\Gamma(q)}(\beta_t)\phi_\varepsilon^y(h\Gamma(q))-\frac{1}{[\Gamma:\Gamma(q)]}\right|>\delta\right\}\leqslant\frac{2C_{\theta'}^2\delta^{-2}\varepsilon^{-\dim G}e^{-2\theta' t}}{[\Gamma:\Gamma(q)]}. \quad (3.5)$$

In particular, the measure of the latter set decays exponentially fast with t. Therefore, it will eventually be strictly smaller than $m_{G/\Gamma(q)}(O_\varepsilon\Gamma(q))$, and for $\varepsilon<\varepsilon_0''$, we clearly have $m_{G/\Gamma(q)}(O_\varepsilon\Gamma(q))=\mathrm{vol}(O_\varepsilon)/[\Gamma:\Gamma(q)]$.

For any t such that the measure in (3.5) is sufficiently small, clearly

$$O_\varepsilon\Gamma(q)\cap\left\{h\Gamma(q):\left|\pi_{G/\Gamma(q)}(\beta_t)\phi_\varepsilon^y(h\Gamma(q))-\frac{1}{[\Gamma:\Gamma(q)]}\right|\leqslant\delta\right\}\neq\varnothing, \quad (3.6)$$

and thus, for any h_t such that $h_t\Gamma(q)$ is in the non-empty intersection (3.6), one has

$$\frac{1}{\mathrm{vol}(B_t)}\int_{B_t}\phi_\varepsilon^y(g^{-1}h_t\Gamma(q))\,d\mathrm{vol}(g)\leqslant\frac{1}{[\Gamma:\Gamma(q)]}+\delta.$$

On the other hand, by Lemma 3.5, for any $\varepsilon\leqslant\varepsilon_0$, $t\geqslant t_0+c\varepsilon_0$ and $h\in\mathcal{O}_\varepsilon$,

$$|\Gamma(q)y\cap B_t|\leqslant\int_{B_{t+c\varepsilon}}\phi_\varepsilon^y(g^{-1}h\Gamma(q))\,d\mathrm{vol}(g). \quad (3.7)$$

Combining the foregoing estimates and using (3.2), we conclude that

$$|\Gamma(q)y\cap B_t|\leqslant\left(\frac{1}{[\Gamma:\Gamma(q)]}+\delta\right)\mathrm{vol}(B_{t+c\varepsilon})\leqslant\left(\frac{1}{[\Gamma:\Gamma(q)]}+\delta\right)(1+c\varepsilon)\,\mathrm{vol}(B_t).$$

This estimate holds as soon as (3.6) holds, and so certainly when

$$\frac{2C_{\theta'}^2}{[\Gamma:\Gamma(q)]}\delta^{-2}\varepsilon^{-\dim G}e^{-2\theta' t}\leqslant\frac{1}{2}\frac{\frac{1}{2}\varepsilon^{\dim G}}{[\Gamma:\Gamma(q)]}\leqslant\frac{1}{2}\frac{\mathrm{vol}(\mathcal{O}_\varepsilon)}{[\Gamma:\Gamma(q)]}.$$

Thus we seek to determine the parameters so that $8C_{\theta'}^2\delta^{-2}e^{-2\theta' t}=\varepsilon^{2\dim G}$. In order to balance the two significant parts of the error term, let us take $c\varepsilon=\delta$, and then

$$\delta=C'_{\theta'}e^{-2\theta' t/(2\dim G+2)},$$

and so as soon as $\delta<1$, we have, using also that $[\Gamma:\Gamma(q)]\geqslant 1$,

$$\begin{aligned}\frac{|\Gamma(q)y\cap B_t|}{\mathrm{vol}(B_t)}&\leqslant\left(\frac{1}{[\Gamma:\Gamma(q)]}+\delta\right)(1+c\varepsilon)\leqslant\frac{1}{[\Gamma:\Gamma(q)]}+\delta+c\varepsilon+\delta c\varepsilon\\&\leqslant\frac{1}{[\Gamma:\Gamma(q)]}+3C'_{\theta'}e^{-\theta' t/(\dim G+1)}.\end{aligned}$$

Note that both the estimate (3.4) as well as the comparison argument in Lemma 3.5, give a lower bound in addition to the foregoing upper bound. Thus the same arguments can be repeated to yield also a lower bound for the uniform lattice points count. This concludes the proof of Theorem 3.2.

Remark 3.6. When the admissible family of sets B_t consists of bi-K-invariant sets, namely sets that are invariant under left and right multiplication by a maximal compact subgroup K of G, two improvements are possible in the previous result.

(i) First, the parameter θ which controls the exponential decay of the operator norm $\|\pi^0_{G/\Gamma(q)}(\beta_t)\|$ depends only on the spherical spectrum and can be estimated directly by the spectral theory of spherical functions. The resulting estimate is $\theta = a/p_K$, where $p_K(G,\Gamma)$ is the $L^p(G)$-integrability parameter associated with the spherical functions in $\pi^0_{G/\Gamma}$. This estimate eliminates the lack of resolution that can be caused by the tensor power argument, which gives $\theta = a/2n_e$, where n_e is the least even integer $\geqslant \frac{1}{2}p(G,\Gamma)$.

(ii) Second, when B_t are bi-K-invariant, the arguments in the proof of Theorem 3.2 can be applied in the obvious manner to the symmetric space G/K whose dimension is $\dim G - \dim K$, so that the exponent in the resulting error estimate is

$$\frac{a}{p_K(1+\dim G/K)}.$$

4. Multiplicativity and sieve dimension

As before, let $\mathbb{G} \subset \mathrm{GL}(n,\mathbb{C})$ be an algebraically connected semisimple algebraic matrix group defined over $\mathbb{Q}$ which we now assume is also simply connected and of non-compact type. Denote by $\mathbb{G}(R)$ the points of G with coefficients in a ring R, and set $G = \mathbb{G}(\mathbb{R})$. Fix $v_0 \in \mathbb{Z}^n$ and let $V = \mathbb{G}v_0$ be the corresponding orbit which we assume is Zariski-closed in the affine n-space A^n. We assume further that the stabilizer of v_0 in $\mathbb{G}$ is trivial. Thus V is a principal homogeneous space for $\mathbb{G}$ and it is defined over $\mathbb{Q}$. Since $\mathbb{G}$ is connected, it follows that V is an (absolutely) irreducible affine variety defined over $\mathbb{Q}$ and is of dimension equal to $\dim \mathbb{G}$. The ring of $\mathbb{G}$-invariants for the action of $\mathbb{G}$ on n-dimensional space separates the closed $\mathbb{G}$-orbits (see [BH]), and we may choose generators $h_1, ..., h_\nu$ in $\mathbb{Q}[x_1, ..., x_n]$ of this ring so that V is given by

$$V = \{x : h_j(x) = \lambda_j, j = 1, ..., \nu\}, \quad \text{with } \lambda_j \in \mathbb{Q}. \tag{4.1}$$

Let $V(\mathbb{Z})$ and $V(\mathbb{Q})$ denote the points of V with coordinates in $\mathbb{Z}$ and $\mathbb{Q}$, respectively.

4.1. Congruential analysis

Let $\Gamma = G(\mathbb{Z}) \subset \mathrm{GL}_n(\mathbb{Z})$ and $\mathcal{O} = \Gamma v_0$ be the corresponding orbit in $\mathbb{Z}^n$. Since $\mathrm{Zcl}(\Gamma) = G$ (see [B]), $\mathrm{Zcl}(\mathcal{O}) = V$. For an integer $d \geqslant 1$, let $\mathcal{O}_d$ be the subset of $(\mathbb{Z}/d\mathbb{Z})^n$ which is obtained by reducing $\mathcal{O}$ modulo d. Similarly, let Γ_d be the reduction of Γ in $\mathrm{GL}_n(\mathbb{Z}/d\mathbb{Z})$, and so $\mathcal{O}_d = \Gamma_d v_0 \pmod d$.

For $g \in \mathbb{Z}[x_1, ..., x_n]$, let

$$\mathcal{O}_d^g = \{x \in \mathcal{O}_d : g(x) \equiv 0 \pmod{d}\}. \tag{4.2}$$

According to the strong approximation theorem (see [PR] and [MVW]; recall that we are assuming that $\mathbb{G}(\mathbb{R})$ has no compact factors), the diagonal embedding

$$\Gamma \longrightarrow \prod_p G(\mathbb{Z}_p) \tag{4.3}$$

is dense, where $\mathbb{Z}_p$ are the p-adic integers. Hence, if $(d_1, d_2)=1$, then

$$\Gamma_d = \Gamma_{d_1} \times \Gamma_{d_2} \tag{4.4}$$

as a subgroup of $\mathrm{GL}_n(\mathbb{Z}/d\mathbb{Z}) \cong \mathrm{GL}_n(\mathbb{Z}/d_1\mathbb{Z}) \times \mathrm{GL}_n(\mathbb{Z}/d_2\mathbb{Z})$.

It follows that in $(\mathbb{Z}/d\mathbb{Z})^n \cong (\mathbb{Z}/d_1\mathbb{Z})^n \times (\mathbb{Z}/d_2\mathbb{Z})^n$,

$$\mathcal{O}_d = \mathcal{O}_{d_1} \times \mathcal{O}_{d_2} \tag{4.5}$$

and

$$\mathcal{O}_d^g = \mathcal{O}_{d_1}^g \times \mathcal{O}_{d_2}^g. \tag{4.6}$$

Now let $f \in \mathbb{Q}[x_1, ..., x_n]$, with $f=g/N$, $N \geqslant 1$, where $g \in \mathbb{Z}[x_1, ..., x_n]$ with

$$\gcd g(\mathcal{O}) = N.$$

Note that $f(\mathcal{O}) \subset \mathbb{Z}$. For $d \geqslant 1$, let

$$\varrho_f(d) = \frac{d|\mathcal{O}_{dN}^g|}{|\mathcal{O}_{dN}|}. \tag{4.7}$$

PROPOSITION 4.1. *$\varrho_f(d)$ is multiplicative in d and, for p prime, $0 \leqslant \varrho_f(p) < p$.*

Proof. Let $d=d_1d_2$ with $(d_1, d_2)=1$ and write $N=N_1N_2$, with $(N_1, d_2)=(N_2, d_1)=(N_1, N_2)=1$. Clearly,

$$|\mathcal{O}_{d_1d_2N_1N_2}| = \frac{|\mathcal{O}_{d_1N_1N_2}|\,|\mathcal{O}_{d_2N_2N_1}|}{|\mathcal{O}_N|} = \frac{|\mathcal{O}_{d_1N}|\,|\mathcal{O}_{d_2N}|}{|\mathcal{O}_N|}.$$

Furthermore,

$$|\mathcal{O}_{d_1d_2N_1N_2}^g| = |\mathcal{O}_{d_1N_1}^g|\,|\mathcal{O}_{d_2N_2}^g| = \frac{|\mathcal{O}_{d_1N_1N_2}^g|}{|\mathcal{O}_{N_2}^g|}\frac{|\mathcal{O}_{d_2N_2N_1}^g|}{|\mathcal{O}_{N_1}^g|} = \frac{|\mathcal{O}_{d_1N}^g|\,|\mathcal{O}_{d_2N}^g|}{|\mathcal{O}_N^g|},$$

and hence

$$\varrho_f(d_1d_2)=\frac{|\mathcal{O}^g_{d_1d_2N}|}{|\mathcal{O}_{d_1d_2N|}}=\frac{|\mathcal{O}^g_{d_1N}|\,|\mathcal{O}^g_{d_2N}|}{|\mathcal{O}^g_N|}\frac{|\mathcal{O}_N|}{|\mathcal{O}_{d_1N}|\,|\mathcal{O}_{d_2N}|}=\varrho_f(d_1)\varrho_f(d_2),$$

since $|\mathcal{O}^g_N|=|\mathcal{O}_N|$.

For $d=p$ there is $x\in\mathcal{O}$ such that

$$\frac{g(x)}{N}\not\equiv 0 \pmod p,$$

as $\gcd g(\mathcal{O})=N$.

Hence $x\notin\mathcal{O}^g_{dN}$ and therefore $\varrho_f(p)<p$. □

Factoring $f\in\mathbb{Q}[V]$ into $t=t(f)$ irreducibles, we get $f=f_1\ldots f_t$, where we are assuming further that each f_j is irreducible in $\mathbb{C}[V]$ and that the f_j's are distinct. According to E. Noether's theorem [No], there is a finite set of primes S such that for $p\notin S$, V is absolutely irreducible over $\mathbb{Z}/p\mathbb{Z}=\mathbb{F}_p$. By increasing the set S if necessary, we may assume that for $p\notin S$ the equations defining G also yield an absolutely irreducible variety over $\mathbb{F}_p$. According to Lang's theorem [La], we have that

$$V(\mathbb{Z}/p\mathbb{Z})=V(\mathbb{F}_p)=\mathbb{G}(\mathbb{F}_p)v. \tag{4.8}$$

By strong approximation we have (again increasing S if needed) that for $p\notin S$,

$$\mathbb{G}(\mathbb{Z}/p\mathbb{Z})=\Gamma_p, \tag{4.9}$$

and hence that

$$\mathcal{O}_p=V(\mathbb{F}_p)\quad\text{for } p\notin S. \tag{4.10}$$

Also when p does not divide N, we have that

$$\mathcal{O}^f_p=\{x\in\mathcal{O}_p: f(x)\equiv 0 \pmod p\}$$

is well defined and

$$\frac{|\mathcal{O}^f_p|}{|\mathcal{O}_p|}=\frac{|\mathcal{O}^g_{pN}|}{|\mathcal{O}_{pN}|}. \tag{4.11}$$

Finally, for $1\leqslant j\leqslant t(f)$, let

$$W_j=\{x\in V: f_j(x)=0\}. \tag{4.12}$$

W_j is an (absolutely irreducible) affine variety defined over $\mathbb{Q}$ of dimension $\dim V-1$. Hence again by Noether's theorem, W_j is absolutely irreducible over $\mathbb{F}_p$ for p outside S',

say. For such p we apply a weak form of the Weil conjectures to W_j (see [LW] or [Sch, §V.5] for an elementary treatment) to conclude that

$$|W_j(\mathbb{F}_p)| = p^{\dim V-1} + O(p^{\dim V-3/2}), \tag{4.13}$$

where the implied constant depends on j only.

Furthermore, since the f_j's are distinct irreducibles in $\mathbb{C}[V]$, we have for $i \neq j$,

$$\dim(W_i \cap W_j) \leqslant \dim V - 2. \tag{4.14}$$

Hence, for $p \notin S \cup S'$, we have

$$|\mathcal{O}_p^f| = \sum_{j=1}^{t(f)} |W_j(\mathbb{F}_p)| + O(p^{\dim V-2}) = t(f)p^{\dim V-1} + O(p^{\dim V-3/2}), \tag{4.15}$$

and similarly

$$|\mathcal{O}_p| = |V(\mathbb{F}_p)| = p^{\dim V} + O(p^{\dim V-1/2}). \tag{4.16}$$

Combining the above, we have that for $p \notin S \cup S'$,

$$\frac{|\mathcal{O}_p^f|}{|\mathcal{O}_p|} = \frac{t(f)}{p} + O(p^{-3/2}), \tag{4.17}$$

and hence that

$$|\varrho_f(p) - t(f)| \leqslant Cp^{-1/2}, \tag{4.18}$$

where C depends only on $\mathcal{O}$ and f.

4.2. Applying the sieve

We now turn to consider the sequence $a_k(T)$, $k \geqslant 0$, defined in (2.9) by

$$a_k(T) = \sum_{\substack{\gamma \in \Gamma : \|\gamma\| \leqslant T \\ |f(\gamma v)| = k}} 1. \tag{4.19}$$

The sums on progressions are then, for $d \geqslant 1$ square free

$$\sum_{k \equiv 0 \pmod d} a_k(T) = \sum_{\substack{\gamma \in \Gamma : \|\gamma\| \leqslant T \\ f(\gamma v) \equiv 0 \pmod d}} 1 = \sum_{\substack{\delta \in \Gamma/\Gamma(dN) \\ g(\delta v) \equiv 0 \pmod{dN}}} \sum_{\substack{\gamma \in \Gamma(dN) \\ \|\delta\gamma\| \leqslant T}} 1, \tag{4.20}$$

where $\Gamma(q)$ is the congruence subgroup of Γ of level q and $f = g/N$ as in §4.1.

According to Theorem 3.2, (4.20) becomes

$$\sum_{k\equiv 0 \pmod d} a_k(T) = \sum_{\substack{\delta\in\Gamma/\Gamma(dN)\\ g(\delta v)\equiv 0 \pmod{dN}}} \left(\frac{\operatorname{vol}\{\|w\|\leqslant T\}}{[\Gamma:\Gamma(dN)]} + O_\varepsilon(T^{a-\theta/(1+\dim G)+\varepsilon})\right)$$
$$= X \sum_{\substack{\delta\in\Gamma/\Gamma(dN)\\ g(\delta v)\equiv 0 \pmod{dN}}} \frac{1}{[\Gamma:\Gamma(dN)]} + O_\varepsilon(|\mathcal{O}_{dN}^g| T^{a-\theta/(1+\dim G)+\varepsilon}),$$

where

$$X = \sum_{k\in\mathbb{N}} a_k(T). \tag{4.21}$$

Now,

$$\mathcal{O}_{dN} = \Gamma_{dN} v \pmod{dN},$$

and hence

$$|\mathcal{O}_{dN}| = \frac{|\Gamma_{dN}|}{|H_{dN}|}, \tag{4.22}$$

where H_{dN} is the stabilizer of v in Γ_{dN}. Also $\Gamma/\Gamma(dN)\cong\Gamma_{dN}$, and so

$$|\{\delta\in\Gamma_{dN} : g(\delta v)\equiv 0 \pmod{dN}\}| = |\mathcal{O}_{dN}^g|\,|H_{dN}|. \tag{4.23}$$

Thus (4.20) becomes

$$\sum_{k\equiv 0 \pmod d} a_k(T) = \frac{X|\mathcal{O}_{dN}^g|\,|H_{dN}|}{|\Gamma_{dN}|} + O_\varepsilon(d^{\dim G} T^{a-\theta/(1+\dim G)+\varepsilon}), \tag{4.24}$$

where we have used $|\mathcal{O}_{dN}^g| \ll d^{\dim G}$, though for later note that, from (4.15) and (4.6),

$$|\mathcal{O}_{dN}^g| \ll d^{\dim G-1}. \tag{4.25}$$

Hence from (4.22) and (4.24) we have

$$\sum_{k\equiv 0 \pmod d} a_k(T) = \frac{\varrho_f(d)}{d} X + R(\mathcal{A}, d), \tag{4.26}$$

with

$$\varrho_f(d) = \frac{d|\mathcal{O}_{dN}^g|}{|\mathcal{O}_{dN}|} \tag{4.27}$$

and

$$\begin{aligned} |R(\mathcal{A},d)| &\underset{\varepsilon}{\ll} d^{\dim G} T^{a(1-\theta/a(1+\dim G))+\varepsilon} \\ &\underset{\varepsilon}{\ll} d^{\dim G} X^{1-\theta/a(1+\dim G)+\varepsilon} \\ &\underset{\varepsilon}{\ll} d^{\dim G} X^{1-1/2n_e(1+\dim G)+\varepsilon}, \end{aligned} \tag{4.28}$$

according to Theorem 3.3, since $\theta = a/2n_e$.

By Proposition 4.1, (4.26) establishes axiom (A_0) in the case when B is the empty set. As for the level distribution (A_1), we have from (4.28) that

$$\sum_{d\leqslant D}|R(\mathcal{A},d)|\underset{\varepsilon}{\ll}D^{1+\dim G}X^{1-1/2n_e(1+\dim G)+\varepsilon}=O(X^{1-\zeta}),\tag{4.29}$$

as long as

$$D\leqslant X^{\tau},\quad \text{with } \tau<\frac{1}{2n_e(1+\dim G)^2}.\tag{4.30}$$

Finally axiom (A_2) follows with a suitable $C_2=C_2(\mathcal{O},f)$ from (4.18). We apply the combinatorial sieve in the form (2.8) to conclude that for

$$z=X^{\alpha},\quad \text{with } \alpha=\frac{1}{9t(f)(1+\dim G)^2 2n_e},\tag{4.31}$$

and for

$$P=\prod_{p\leqslant z}p,$$

with X large enough,

$$\frac{X}{(\log X)^{t(f)}}\ll S(\mathcal{A},P)\ll\frac{X}{(\log X)^{t(f)}},\tag{4.32}$$

where the implied constants depend only on f and the orbit $\mathcal{O}$.

4.3. Completion of proofs of Theorems 1.6 and 1.7 and Corollary 1.8

We begin by establishing the following two lemmas.

LEMMA 4.2. *Assume that $h\in\mathbb{Q}[x_1,\dots,x_n]$ does not vanish identically when restricted to $V=Gv$. Then there is $\delta=\delta(h)>0$ such that*

$$|\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } h(\gamma v)=0\}|\ll T^{a-\delta}.$$

Proof. We may assume that h is not constant on V and hence there is a finite extension E of $\mathbb{Q}$ over which h factors into $h=h_1\dots h_\nu$, where each h_j is absolutely irreducible in $E[V]$. For p large enough and a prime ideal $\mathcal{P}$ in the ring of integers $\mathcal{I}_E$ of E with $\mathcal{P}|(p)$, we have, letting $N(\mathcal{P})$ denote the norm of the ideal $\mathcal{P}$,

$$|\{x\in V(\mathcal{I}_E/\mathcal{P}):h_j(x)=0\}|\ll N(\mathcal{P})^{\dim V-1},\tag{4.33}$$

the implied constant depending on h.

Assume further that p splits completely in E so that $\mathcal{I}_E/\mathcal{P}\cong\mathbb{Z}/p\mathbb{Z}$. Then

$$|\{x\in V(\mathbb{Z}/p\mathbb{Z}): h(x)\equiv 0(p)\}|\ll p^{\dim V-1}. \tag{4.34}$$

Let T be the large parameter in the lemma and choose p as above with $\frac{1}{2}T^{\alpha}\leqslant p\leqslant 2T^{\alpha}$ for $\alpha>0$ small and to be chosen momentarily. Such a p exists by Chebotarev's density theorem [Ch]. With this choice, we have that

$$\begin{aligned}|\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } h(\gamma v)=0\}| &\leqslant |\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } h(\gamma v)\equiv 0 \pmod p\}|\\ &= \frac{|\mathcal{O}_p^h|}{|\mathcal{O}_p|}X+O_{\varepsilon}(T^{a-\theta/(1+\dim G)+\varepsilon}p^{\dim G}),\end{aligned} \tag{4.35}$$

on using (4.24).

Coupled with (4.34), this gives

$$|\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } h(\gamma v)=0\}|\underset{\varepsilon}{\ll}\frac{T^{a+\varepsilon}}{p}+T^{a-\theta/(1+\dim G)+\varepsilon}p^{\dim G}\ll T^{a-\delta}, \tag{4.36}$$

where we choose $\delta=\theta/(1+\dim G)^2$. □

LEMMA 4.3. *Let $f=f_1\dots f_t$, with $f_j\in\mathbb{Z}[x_1,\dots,x_n]$ irreducible as in Theorem* 1.6, *$1\leqslant j\leqslant t(f)$. Then there is $\delta_1>0$ such that for any $m\in\mathbb{Z}$ and any $1\leqslant j\leqslant t(f)$,*

$$|\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } f_j(\gamma v)=m\}|\ll T^{a-\delta_1},$$

the implied constant depending only on f and Γ.

Proof. By assumption, f_j is not constant when restricted to V. Hence, by Lemma 4.2 with $g=f_j-m$, we get Lemma 4.3, but potentially the implied constant depends on m. The only place where this dependence may enter is in (4.33) and for m outside a finite set, f_j-m is irreducible over $\overline{\mathbb{Q}}$. The equations defining $f_j-m=0$ and V will be irreducible over $\overline{\mathbb{F}}_p$ for p outside a fixed finite set and they are of a fixed degree in the variables $(x_1,\dots,x_n)$. Hence, by [LW, Lemma 1], the upper bound in (4.33), with $h=f_j-m$, is uniform in m. □

Proof of Theorem 1.6. We choose $\varepsilon_1>0$ small (but fixed) so that firstly $\varepsilon_1<\delta_1$, where δ_1 is determined by Lemma 4.3. For $1\leqslant j\leqslant t(f)$ and T large, we have from Lemma 4.3 that

$$|\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } |f_j(\gamma v)|\leqslant T^{\varepsilon_1}\}|\ll T^{a-\delta_1+\varepsilon_1}. \tag{4.37}$$

Now,

$$\begin{aligned}&\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } f_j(\gamma v) \text{ is prime for all } j\}\\ &\quad\subset\bigcup_{j=1}^{t(f)}\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } |f_j(\gamma v)|\leqslant T^{\varepsilon_1}\}\\ &\qquad\cup\{\gamma\in\Gamma:\|\gamma\|\leqslant T, |f_j(\gamma v)|\geqslant T^{\varepsilon_1} \text{ and } f_j(\gamma v) \text{ is prime for each } j\}.\end{aligned} \tag{4.38}$$

By (4.37), the cardinality of the union of the first $t(f)$ sets above is at most $O(T^{a-\delta_1+\varepsilon_1})$. The last set on the right-hand side of (4.38) is contained in

$$\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } (f(\gamma v),P_z)=1\},$$

where

$$P_z=\prod_{p\leqslant z} z=T^{\varepsilon_1}. \tag{4.39}$$

The cardinality of the last set is $S(\mathcal{A},P_z)$ and if $\varepsilon_1<\alpha$, where α is the level distribution in (4.31), then we may apply (4.32) to conclude that

$$\{\gamma\in\Gamma:\|\gamma\|\leqslant T \text{ and } f_j(\gamma v) \text{ is prime}\}\ll T^{a-\delta_1+\varepsilon_1}+\frac{X}{(\log X)^{t(f)}}\ll\frac{X}{(\log X)^{t(f)}},$$

since X is asymptotic to $BT^a(\log T)^b$. This completes the proof of Theorem 1.6. □

Proof of Theorem 1.7. Taking α as in (4.31) and $z=X^\alpha$, we have that

$$\sum_{\substack{\|\gamma\|\leqslant T\\(f(\gamma v),P_z)=1}} 1\gg\frac{X}{(\log X)^{t(f)}}, \tag{4.40}$$

where $P_z=\prod_{p\leqslant z}p$.

Now any point $\gamma v\in\mathcal{O}$ which occurs in the sum in (4.40) has $|\gamma v|\ll T$ (where $|\cdot|$ is the usual Euclidean norm on $\mathbb{R}^n$) and hence $|f(\gamma v)|\ll T^{\deg f}$. On the other hand, for such a point $\gamma v\in\mathcal{O}$ in the sum, $f(\gamma v)$ has all its prime factors at least $z\gg X^\alpha\gg T^{a\alpha}$. It follows that for such a point γv, $f(\gamma v)$ has at most

$$r=\frac{\deg f}{a\alpha}=\frac{9t(f)(1+\dim G)^2 2n_e(\Gamma)\deg f}{a}$$

prime factors. □

Proof of Corollary 1.8. Suppose, by way of contradiction, that the points γv that we produced in the previous paragraph are not Zariski-dense in V. Since V is connected, it follows that there is an $h\in\mathbb{Q}[x_1,...,x_n]$ which does not vanish identically on V and such that all our points lie in $V\cap\{x:h(x)=0\}$. But by Lemma 4.2 the total number of points in this intersection with $\|\gamma\|\leqslant T$ is $O(T^{a-\delta})$ with $\delta=\delta(h)>0$. This contradicts the lower bound of $cX/(\log X)^{t(f)}$ (with $c>0$ fixed) for the number of points with at most r prime factors that was produced in Theorem 1.7. □

Finally we apply Theorems 1.6 and 1.7 to the case of $n\times n$ integral matrices of determinant m as in the introduction. Let $G=\mathrm{SL}_n$, $\Gamma=\mathrm{SL}_n(\mathbb{Z})$ and $v\in\mathrm{Mat}_n(\mathbb{Z})$ with $\det v=m\neq 0$. We identify Mat_n with the affine space A^M $(M=n^2)$, with G acting by $x\mapsto gx$. We have that $V_{m,n}=Gv$ and that $V_{m,n}(\mathbb{Z})$ consists of a finite number of Γ-orbits. Note that if $|\cdot|$ is a norm on $\mathrm{Mat}_n(\mathbb{R})$ then for a fixed invertible v as above, $g\mapsto|gv|$ defines a (vector-space) norm on $\mathrm{Mat}_n(\mathbb{R})$ and we may apply Theorems 1.6 and 1.7 to this setting, namely to the individual orbits $\mathcal{O}=\Gamma v$ with v fixed as above. In this case, $\dim G+1=n^2$ and $a=n^2-n$ for any choice of norm as above [DRS], [GW], [Ma], and the norm balls are admissible [GN1]. Furthermore $p(\Gamma)=2(n-1)$ (see [DRS]) and so $n_e=(n-1)$ if n is odd, and $n_e=n$ if $n\geqslant 4$ is even. For $n=2$ we can take $n_e=2$ by [KiS], see (6.18). Thus Theorem 1.7 yields that for n odd r need only satisfy

$$r>\frac{9t(f)n^4 2(n-1)\deg f}{n(n-1)}=18t(f)n^3\deg f, \tag{4.41}$$

and for n even the previous expression is multiplied by $n/(n-1)$.

Thus, for $\mathcal{O}=\Gamma v$ and f primitive on $\mathcal{O}$,

$$|\{x\in\mathcal{O}:|x|\leqslant T \text{ and } f(x) \text{ has a most } r \text{ prime factors}\}|\gg\frac{T^{n^2-n}}{(\log T)^{t(f)}}. \tag{4.42}$$

This proves the Γ-orbit version of Theorem 1.2 and Corollary 1.3.

Now, for $m\neq 0$ fixed, $V_{m,n}(\mathbb{Z})$ consists of a finite number of such Γ-orbits, and Theorem 1.1 follows from the Γ-orbit version. In order to establish Theorem 1.2 and Corollary 1.3 for $V_{m,n}(\mathbb{Z})$, we need to show the following: if f is $V_{m,n}(\mathbb{Z})$-weakly primitive, then for some $v\in V_{m,n}(\mathbb{Z})$, f is $\mathcal{O}$-weakly primitive on the orbit $\mathcal{O}=\Gamma v$.

As in the theory of Hecke correspondences on n-dimensional lattices (see [T]), we decompose $V_{m,n}$ into Γ-orbits

$$V_{m,n}(\mathbb{Z})=\coprod_{j=1}^{k(m)}\mathcal{O}^{(j)}, \tag{4.43}$$

with $\mathcal{O}^{(j)}=\Gamma v_j$ and $v_j\in V_{m,n}(\mathbb{Z})$.

Denote by W the union of the $k(m)$ global Γ-orbits and for $d\geqslant 1$ let $\mathcal{O}_d^{(j)}$ denote the reduction of $\mathcal{O}^{(j)}$ modulo d, which defines a point in the orbit space

$$\mathrm{SL}_n(\mathbb{Z}/d\mathbb{Z})\backslash\mathrm{Mat}_n(\mathbb{Z}/d\mathbb{Z}),$$

where Mat_n is the space of $n\times n$ matrices. Let W_d denote the reduction of W into this space. Note that for $d=p$ with a prime p that does not divide m, W_p consists of a single

point, that is to say the orbits $\mathcal{O}^{(j)}$ all reduce to the same $\mathrm{SL}_n(\mathbb{Z}/d\mathbb{Z})$-orbit modulo p. The key property that we need for these reductions is that if $(d_1,d_2)=1$ then the diagonal embedding

$$W \longrightarrow (\mathrm{SL}_n(\mathbb{Z}/d_1\mathbb{Z})\backslash \mathrm{Mat}_n(\mathbb{Z}/d_1\mathbb{Z}))\times(\mathrm{SL}_n(\mathbb{Z}/d_2\mathbb{Z})\backslash \mathrm{Mat}_n(\mathbb{Z}/d_2\mathbb{Z})) \tag{4.44}$$

is onto $W_{d_1}\times W_{d_2}$.

With (4.44), the weak primitivity property that we need is established as follows. Let f be weakly primitive on $V_{m,n}(\mathbb{Z})$ and for simplicity of notation assume that $N=1$ in §1.1 and that m is square free. So $f\in\mathbb{Z}[x_{ij}]$ and for each prime $p\geqslant 2$ there is an $x\in V_{m,n}(\mathbb{Z})$ such that $f(x)\neq 0 \pmod p$. We claim that there is a $J\in\{1,\dots,k(m)\}$ such that f is weakly primitive for $\mathcal{O}^{(J)}$. That is, for every prime $p\geqslant 2$ there is $x\in\mathcal{O}^{(J)}$ with $f(x)\neq 0 \pmod p$.

Call a prime $p\geqslant 2$ *good* for $\mathcal{O}^{(j)}$ if such an x exists for p. This property is determined locally at p. That is by strong approximation for SL_n, p is good for $\mathcal{O}^{(j)}$ if and only if the local orbit $\mathrm{SL}_n(\mathbb{Z}/p\mathbb{Z})v_j$ in $\mathrm{Mat}_n(\mathbb{Z}/p\mathbb{Z})$ contains an x such that $f(x)\not\equiv 0 \pmod p$. So the condition is one on $\mathcal{O}_p^{(j)}$. Every prime p that does not divide m is good for any $\mathcal{O}^{(j)}$, $j=1,\dots,k(m)$, since $V_{m,n}(\mathbb{Z})$ is good at p and all global orbits reduce to the same local orbit at such a p. Now write $m=p_1\dots p_\ell$, and then

$$W \longrightarrow W_{p_1}\times\dots\times W_{p_\ell} \tag{4.45}$$

is onto.

Moreover, by our assumption on f, for each p_i, $i=1,\dots,\ell$, there is j_{p_i} such that $\mathcal{O}_{p_i}^{(j_{p_i})}$ is good at p_i. Hence, by (4.45), there is a $J\in\{1,\dots,k(m)\}$ such that $\mathcal{O}_{p_i}^{(J)}$ is good for each $i=1,\dots,\ell$. Thus, f is weakly primitive for $\mathcal{O}^{(J)}=\Gamma v_J$.

5. Zariski density of prime matrices

Fix $n\geqslant 3$. We say that an $n\times n$ integral matrix is *prime* if all of its coordinates are prime numbers. For an integer m, $V_{m,n}$ denotes as usual the affine variety given by $\{x\in\mathrm{Mat}_n(\mathbb{R}):\det x=m\}$. We are interested in the set of prime matrices being Zariski-dense in $V_{m,n}$. For this to happen we must clearly allow x to have all its coordinates x_{ij} to be odd numbers. Such a matrix x satisfies $\det x\equiv 0 \pmod{2^{n-1}}$. It turns out that this is the only obstruction to producing many primes in $V_{m,n}(\mathbb{Z})$. As an application of Vinogradov's methods for analyzing linear equations in primes with three or more variables, we show the following result.

THEOREM 5.1. *Fix $n\geqslant 3$. Then the set of prime matrices x in $V_{m,n}(\mathbb{Z})$ is Zariski-dense in $V_{m,n}$ if and only if $m\equiv 0 \pmod{2^{n-1}}$.*

The proof of Theorem 5.1 can be extended to prove a special case of the general local to global conjectures for primes in orbits of actions of certain groups [BGS].

To state the result, let Λ be a finite index subgroup of $\mathrm{SL}_n(\mathbb{Z})$, $n\geqslant 3$. For an $n\times n$ integral matrix A with $\det A=m\neq 0$, let $\mathcal{O}_A$ denote the Λ-orbit ΛA. Thus $\mathcal{O}_A$ is contained in $V_{m,n}(\mathbb{Z})$ and is Zariski-dense in $V_{m,n}$.

THEOREM 5.2. *The set of prime matrices x in $\mathcal{O}_A$ is Zariski-dense in $V_{m,n}$ if and only if there are no local congruence obstructions* (*see the remark below*).

Remark 5.3. (A) We are using here that for $n\geqslant 3$ every finite-index subgroup of $\mathrm{SL}_n(\mathbb{Z})$ is a congruence subgroup ([Me], [BMS]).

(B) The general orbit conjecture for this action asserts that Theorem 5.2 holds for a subgroup Λ of $\mathrm{SL}_n(\mathbb{Z})$ which is Zariski-dense in SL_n and with the coordinate functions x_{ij}, $i,j=1,...,n$, replaced by any set $f_1,...,f_t$ of primes in the coordinate ring $\mathbb{Q}[x_{ij}]/(\det(x_{ij})-m)$. In this setting, the local congruence obstructions that need to be passed are that for any $q\geqslant 2$ there is an x in $\mathcal{O}_A$ (mod q), the reduction of $\mathcal{O}_A$ modulo q, such that $f_1(x)\,...\,f_t(x)\in(\mathbb{Z}/q\mathbb{Z})^*$.

An example of an orbit in Theorem 5.2 for which there are no local obstructions for any Γ is

$$\mathcal{O}=\Gamma\begin{bmatrix} 1 & 1 & 1 & ... & 1 & 1\\ -1 & 1 & 1 & ... & 1 & 1\\ -1 & -1 & 1 & ... & 1 & 1\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ -1 & -1 & -1 & ... & 1 & 1\\ -1 & -1 & -1 & ... & -1 & 1 \end{bmatrix}\subset V_{2^{n-1},n}.$$

This is similar to $a=1$ (or $a=-1$) in Dirichlet's theorem, i.e., there are infinitely many $p\equiv 1$ (mod q) for any q.

Proof of necessity of the congruence condition in Theorem 5.1. If the set of matrices with prime entries is Zariski-dense, then of course the set of matrices with odd entries is Zariski-dense. But then if x is $n\times n$ integral and has odd entries, then writing the columns of x as $a_1,...,a_n$, we have

$$\det x=\det[a_1,...,a_n]=\det[a_1,a_2-a_1,a_3-a_1,...,a_n-a_1]=\det[a_1,2b_2,...,2b_n],$$

with b_j integral. Hence $\det x=2^{n-1}\det[a_1,b_2,...,b_n]\equiv 0$ (mod 2^{n-1}). □

To demonstrate the sufficiency of the congruence condition in Theorem 5.1, we will consider the simplest case when $m=2^{n-1}$. In general one needs to impose further congruence conditions in the construction below.

LEMMA 5.4. *For $n\geqslant 2$ let*

$$\mathcal{Y}=\left\{\begin{bmatrix} x_{21} & x_{22} & x_{23} & \dots & x_{2n}\\ x_{31} & x_{32} & x_{33} & \dots & x_{3n}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nn}\end{bmatrix}\right\},$$

which we identify with the affine $A^{(n-1)n}$ space. We denote by $A_j(y)$ the $(n-1)\times(n-1)$ minor of y gotten by deleting the j-th column. Let $\mathcal{G}$ be the set of $y\in\mathcal{Y}$ for which

(i) $(A_1(y),\dots,A_n(y))=2^{n-2}$;

(ii) $A_1(y)\dots A_n(y)\neq 0$;

(iii) $A_1(y)+\dots+A_n(y)\equiv 0 \pmod{2^{n-1}}$;

and the x_{ij} (where $(x_{ij})=y$) are all prime. Then $\mathcal{G}$ is Zariski-dense in $\mathcal{Y}$.

Proof. We use Dirichlet's theorem repeatedly and proceed by induction on n.

For $n=2$, $y=[x_{21},x_{22}]$ and we seek $(x_{21},x_{22})=1$, $x_{21}+x_{22}\equiv 0 \pmod 2$, and x_{21} and x_{22} both prime. Clearly the set $\mathcal{G}$ of such vectors is Zariski-dense in A^2.

For $n\geqslant 3$ we assume the lemma for $n-1$ and construct the $y\in\mathcal{G}$ as follows: for

$$z=\begin{bmatrix} z_2\\ \vdots\\ z_n\end{bmatrix},\quad \xi=\begin{bmatrix} \xi_2\\ \vdots\\ \xi_n\end{bmatrix}$$

and an $(n-1)\times(n-2)$ matrix w, we write

$$y=[z\xi w].$$

By induction, the set of w's in the space $\mathcal{W}$ of such $(n-1)\times(n-2)$ matrices, for which w_{ij} are all prime and such that

$$C_2(w)+\dots+C_n(w)\equiv 0 \pmod{2^{n-2}} \tag{5.1}$$

and

$$(C_2(w),\dots,C_n(w))=2^{n-3}, \tag{5.2}$$

is Zariski-dense in $\mathcal{W}$. Here $C_i(w)$ is the $(n-2)\times(n-2)$ minor of w obtained by deleting the ith row of w. For such a w, we seek ξ satisfying

$$A_1=\xi_2C_2-\xi_3C_3+\dots+(-1)^n\xi_nC_n\equiv 2^{n-2} \pmod{2^{n-1}} \tag{5.3}$$

and

$$(\xi_j,2)=1. \tag{5.4}$$

In view of (5.1) and (5.2), this amounts to

$$\xi_2 C_2' - \xi_3 C_3' + \ldots + (-1)^n \xi_n C_n' \equiv 2 \pmod 4, \tag{5.5}$$

where

$$(C_2', \ldots, C_n') = 1 \quad \text{and} \quad C_2' + \ldots + C_n' \equiv 0 \pmod 2. \tag{5.6}$$

According to (5.6), the number ℓ of C_j''s which are $\equiv \pm 1 \pmod 4$ is even and positive. Collecting these C_j''s on the left, renumbering the indices and replacing ξ_j by $-\xi_j$ suitably, leads to solving

$$\xi_2 + \ldots + \xi_{\ell+1} \equiv b \pmod 4, \tag{5.7}$$

where b is either 0 or 2 modulo 4. If $b \equiv 0 \pmod 4$, choose $\xi_j = (-1)^j$, $2 \leqslant j \leqslant \ell+1$, while if $b \equiv 2 \pmod 4$, choose $\xi_2 = \xi_3 = 1$ and $\xi_j = (-1)^j$ for $4 \leqslant j \leqslant \ell+1$. Since ℓ is even, these choices solve (5.5).

Having found such a $\xi \pmod{2^{n-1}}$ satisfying (5.3) and (5.4), we choose ξ integral satisfying this congruence and for which ξ_j are all prime. This is possible by Dirichlet's theorem and their choice is Zariski-dense in the ξ-space.

For each choice of w and ξ above, we choose z as follows. First, we have

$$A_1 \equiv 2^{n-2} \pmod{2^{n-1}}, \tag{5.8}$$

and hence $A_1 \neq 0$. For each odd prime p dividing A_1, let

$$t^{(p)} = \begin{bmatrix} t_2^{(p)} \\ \vdots \\ t_n^{(p)} \end{bmatrix}$$

be chosen with $t_j^{(p)} \in (\mathbb{Z}/p\mathbb{Z})^*$ and satisfying

$$A_2 := t_2^{(p)} C_2 - t_3^{(p)} C_3 + \ldots + (-1)^n t_n^{(p)} C_n \not\equiv 0 \pmod p. \tag{5.9}$$

It is clear that such a $t^{(p)}$ can be found, since $(C_2, \ldots, C_n) = 2^{n-3}$ and $p \geqslant 3$.

Next, let $q_3, ..., q_n$ be distinct primes different from 2, from any prime divisor of A_1 and from any entry of w. We choose z to satisfy the following congruences:

$$\begin{bmatrix} z_2 \\ \vdots \\ z_n \end{bmatrix} \equiv \begin{bmatrix} w_{2j} \\ \vdots \\ w_{nj} \end{bmatrix} \pmod{q_j}, \qquad \text{for } 3 \leqslant j \leqslant n. \tag{5.10}$$

$$\begin{bmatrix} z_2 \\ \vdots \\ z_n \end{bmatrix} \equiv \begin{bmatrix} t_2^{(p)} \\ \vdots \\ t_n^{(p)} \end{bmatrix} \pmod{p}, \qquad \text{for } p|A_1,\ p \text{ odd}, \tag{5.11}$$

$$\begin{bmatrix} z_2 \\ \vdots \\ z_n \end{bmatrix} \equiv \begin{bmatrix} \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \pmod{2^{n-1}}. \tag{5.12}$$

The conditions (5.10), (5.11) and (5.12) involve distinct prime moduli and the numbers on the right are all prime to their moduli, hence by Dirichlet's theorem we can choose z_j to be prime and to satisfy the congruences (5.10), (5.11) and (5.12). Moreover, the set of choices for these z's is Zariski-dense in the space of z's. This produces matrices $y=[z\xi w]$ which we check satisfy the requirement of the lemma. First, by (5.8),

$$A_1(y) \equiv 2^{n-2} \pmod{2^{n-1}}. \tag{5.13}$$

Second, by (5.9), $A_1(y)$ and $A_2(y)$ have no odd prime common factor. Finally, by (5.12),

$$A_2(y) \equiv A_1(y) \pmod{2^{n-2}}, \tag{5.14}$$

so we conclude that

$$(A_1(y), A_2(y)) = 2^{n-2}. \tag{5.15}$$

Note that, for $3 \leqslant j \leqslant n$, $A_j(y) \equiv 0 \pmod{2^{n-2}}$, since A_j is the determinant of an $(n-1)\times(n-1)$ matrix with odd entries. Hence

$$A_1(y) + ... + A_n(y) \equiv 0 \pmod{2^{n-1}}. \tag{5.16}$$

Thus, together with (5.13), we deduce that

$$(A_1(y), ..., A_n(y)) = 2^{n-2}. \tag{5.17}$$

From (5.13) and (5.14) we conclude that $A_1(y)A_2(y) \neq 0$, while from (5.10) we have that $A_j(y) \equiv \pm A_1(y) \pmod{q_j}$ for $3 \leqslant j \leqslant n$, and hence $A_3(y) \dots A_n(y) \neq 0$. All this, coupled with the fact that the entries of y are prime and that the y's can be chosen to be Zariski-dense in $\mathcal{Y}$, completes the proof of Lemma 5.4. □

We now appeal to Vinogradov's methods [Vi] for studying the solvability of

$$H_{A_0,\dots,A_n}=\{(s_1,\dots,s_n): A_1s_1-A_2s_2+\dots+(-1)^{n+1}A_ns_n-A_0=0\}, \tag{5.18}$$

with s_j prime (i.e., (s_j) a prime ideal in $\mathbb{Z}$). The treatment in Vaughan [Va, p. 37], shows that if $n\geqslant 3$ and $A_1\dots A_n\neq 0$, then the number of solutions to (5.18), with $|s_j|\leqslant T$ and s_j prime, satisfies

$$R(T)\gg C\frac{T^{n-1}}{(\log T)^n}+O_\nu\left(\frac{T^{n-1}}{(\log T)^\nu}\right) \tag{5.19}$$

for any fixed (large) ν. Moreover, the critical number C given by the singular series is non-zero if and only if the following local conditions are satisfied:

$$(A_0,A_1,\dots,A_{n-1})=(A_0,A_1,\dots,A_{n-2},A_n)=(A_1,A_2,\dots,A_n)=(A_0,A_1,\dots,A_n) \tag{5.20}$$

and

$$A_0+\dots+A_n\equiv 0 \pmod{2(A_0,\dots,A_n)}. \tag{5.21}$$

If any of these conditions fail, for example if there is a prime p and $e\geqslant 1$ with $p^e|A_j$ for $j=0,\dots,n-1$ but $p^e\nmid A_n$, then any solution to (5.18) must have $p|s_n$. Hence the set of solutions to (5.18) with s_n prime is not Zariski-dense in $H_{A_0,\dots,A_n}$. Thus the conditions (5.20) and (5.21) are necessary for the Zariski density of $(s_1,\dots,s_n)$, s_j prime, in $H_{A_0,\dots,A_n}$. These conditions are also sufficient. Indeed $H_{A_0,\dots,A_n}$ is connected and hence if these points are not Zariski-dense, then there is a polynomial $f(s_1,\dots,s_n)$ which is non-constant on $H_{A_0,\dots,A_n}$ such that all the s's lie in

$$H_{A_0,\dots,A_n}\cap\{s: f(s)=0\}.$$

It is elementary that the number of integer points in this intersection and for which $|s_j|\leqslant T$ is $O(T^{n-2})$. Hence, if $C\neq 0$, then (5.19) gives a contradiction to the points all lying in $\{s: f(s)=0\}\cap H_A$. We conclude that

$$\{(s_1,\dots,s_n): s_j \text{ is prime and } s\in H_{A_0,\dots,A_n}\} \tag{5.22}$$

is Zariski-dense in $H_{A_0,\dots,A_n}$ if and only if (5.20) and (5.21) hold.

Let $\mathcal{Y}$ be the space in Lemma 5.4 and $\mathcal{G}$ the set of y's constructed in that lemma. Set $A_0=2^{n-1}$. Then, for $y\in\mathcal{G}$,

$$\begin{aligned}&(A_1(y),\dots,A_n(y))=2^{n-2}\\ &A_1(y)+\dots+A_n(y)\equiv 0 \qquad (\mathrm{mod}\ 2^{n-1}),\end{aligned}$$

and $A_1(y)\dots A_n(y)\neq 0$. Hence $(A_0, A_1(y), ..., A_n(y))=2^{n-2}$ and the number of $1\leqslant j\leqslant n$ for which $2^{n-2}|A_j(y)$ is even and positive. It follows that

$$(A_0, A_1(y), ..., A_{n-1}(y)) = (A_0, A_1(y), ..., A_{n-2}(y), A_n(y)) = (A_1(y), ..., A_n(y)) = 2^{n-2}$$

and

$$A_0+A_1(y)+...+A_n(y)\equiv 0 \pmod{2(A_0, A_1(y), ..., A_n(y))}.$$

Thus (5.20) and (5.21) are satisfied and so, by (5.22), it follows that for any $y\in\mathcal{G}$ the set of $s\in H_{A_0,A_1(y),...,A_n(y)}$, for which all s_j are prime, is Zariski-dense in $H_{A_0,A_1(y),...,A_n(y)}$. For each such y and s, the matrix x given by

$$\begin{bmatrix} s_1 & \dots & s_n \\ & y & \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ z_2 & \xi_2 & & \\ \vdots & \vdots & & w \\ z_n & \xi_n & & \end{bmatrix}$$

is a prime matrix in $V_{2^{n-1},n}$.

To complete the proof of Theorem 5.1 with $m=2^{n-1}$, we must show that the set of x's constructed above is dense in $V_{2^{n-1},n}$. Let

$$\mathcal{Y}^0 = \{y\in\mathcal{Y} : A_j(y)\neq 0 \text{ for some } 1\leqslant j\leqslant n\}.$$

$\mathcal{Y}^0$ is an open irreducible subset of $A^{(n-1)\times n}$ and is quasi-affine. Let $\Upsilon: V_{m,n}\to\mathcal{Y}^0$ be the surjective morphism

$$x\longmapsto \begin{pmatrix} x_{21} & \dots & x_{2n} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}. \tag{5.23}$$

If U is a non-empty open subset of $V_{m,n}$ then, since $V_{m,n}$ is connected, U is dense and hence $\Upsilon(U)$ is dense in $\Upsilon(V_{m,n})=\mathcal{Y}^0$. Also $\Upsilon(U)$ is constructible and contains an open dense subset of $\mathcal{Y}^0$. According to Lemma 5.4, there is a $y\in\mathcal{G}\cap\Upsilon(U)$. Now $U\cap\Upsilon^{-1}(y)$ is a non-empty open subset of $H_{2^{n-1},A_1(y),...,A_n(y)}$. According to the analysis above, it contains a point $(p_1, ..., p_n)$ all of whose coordinates are prime. Hence

$$x = \begin{bmatrix} p \\ y \end{bmatrix}$$

is a prime matrix and it is in U. This proves Theorem 5.1.

6. Spectral estimates for uniform lattices

We now turn to explain an alternative approach to estimating the level of distribution. Indeed, rather than giving an error term for the number of lattice points in a ball, it suffices to estimate the deviation of a positive weighted sum over the lattice points. This allows one to take a smooth weight and to estimate its Fourier transform directly via a convergent eigenfunction expansion of the corresponding automorphic kernel. This method gives a sharper result for the level of distribution and the improvement is most significant when the lattice is co-compact. Since the latter assumption also allows us to avoid the analysis of Eisenstein series necessary for the estimates of eigenfunction expansions, we will present the method only for co-compact lattices. In this case it can be naturally viewed as providing the error estimate in the non-Euclidean version of a Poisson summation formula for compact locally symmetric spaces. To simplify matters further, we will assume that the weight functions are radial, allowing us to reduce matters to spectral estimates associated with spherical functions. We note that the estimate of the deviation we give below is in fact sharp: it gives the best possible result for a smooth weighted sum.

We retain the notation of §§3.1-3.3, and again let G be a connected semisimple Lie group with finite center and no compact factors. $\mathcal{S}$ denotes the symmetric space $\mathcal{S}=G/K$, where K is a maximal compact subgroup. We take the Riemannian structure on $\mathcal{S}$ induced by the Cartan–Killing form on G and let d denote the associated G-invariant distance. We let $B_t(z)$ denote the ball of radius t and center z. Consider the family of kernels on $\mathcal{S}\times\mathcal{S}$ given by

$$L_t(z,w)=\chi_{[0,t]}(d(z,w)),$$

where $\chi_{[0,t]}$ is the characteristic function of an interval. Fix a smooth function $b(w)$ on $\mathcal{S}$ which is non-negative, positive definite, supported in a ball of radius t_0 with center w_0, invariant under K_{w_0} and with integral 1. Define the following smooth function, which is supported on the set of points whose distance from w_0 is at most $t+t_0$:

$$W_t(z)=\int_{\mathcal{S}} L_t(z,w)b(w)\,d\mathrm{vol}(w). \tag{6.1}$$

Let Γ be any uniform lattice in G which satisfies, for any z,

$$|\{\gamma\in\Gamma : d(\gamma z,z)\leqslant 1\}|\leqslant C_\Gamma,$$

with C_Γ fixed. Our version of the error estimate in the Poisson summation formula is as follows.

THEOREM 6.1. (Poisson summation formula) *Let Γ be a uniform irreducible lattice in G as above. Then for $\eta>0$ fixed,*

$$\sum_{\gamma\in\Gamma} W_t(\gamma z) = \frac{\mathrm{vol}(B_t)}{\mathrm{vol}(\Gamma\backslash G)} + O_\eta(\mathrm{vol}(B_t)^{1-1/p+\eta}),$$

where

(1) *the result holds uniformly for arbitrary z and w_0;*

(2) *$p=p(G,\Gamma)$ is the integrability parameter of the representation in $L^2_0(\Gamma\backslash G)$ as in Theorem* 3.3; *in particular*,

$$1-\frac{1}{p}=\frac{1}{2}$$

if and only if all representations weakly contained in $L^2_0(\Gamma\backslash G)$ are tempered.

Let us note the following. First, it is indeed the case that $p(G,\Gamma)<\infty$ for every irreducible lattice [Cl1] (we refer to [KeS] for discussion and references). Second, Theorem 6.1 holds for general point-pair-invariant kernels.

The proof is based on establishing uniform control on the pointwise spectral expansion of the smooth function W_t. We will use the spectral expansion associated with the commutative algebra $\mathcal{D}$ of G-invariant differential operators on the symmetric space (recall that the Laplacian generates this algebra if and only if the real rank of G is 1). Being G-invariant, these differential operators descend to operators on $M=\Gamma\backslash\mathcal{S}$, and admit a joint spectral resolution. The eigenvalues are given by (infinitesimal) characters $\omega_\lambda\colon\mathcal{D}\to\mathbb{C}$, parametrized by

$$\lambda\in\Sigma\subset\mathrm{Hom}(\mathfrak{a},\mathbb{C})/\mathcal{W}.$$

Here $G=NAK$ is an Iwasawa decomposition, $\mathfrak{a}$ is the Lie algebra of A, $\mathcal{W}$ is the Weyl group of $(\mathfrak{g},\mathfrak{a})$ and Σ parametrizes the positive-definite spherical functions $\Psi_\lambda\colon G\to\mathbb{C}$. We let $\|\cdot\|$ denote the Euclidean norm associated with the inner product on $\mathfrak{a}$ given by the restriction of the Killing form. Finally recall that every (normalized) joint eigenfunction ϕ on $\Gamma\backslash\mathcal{S}$ of the algebra $\mathcal{D}$ is also a joint eigenfunction of the commutative convolution algebra $L^1(K\backslash G/K)$ of bi-K-invariant kernels on G. The eigenvalue constitutes a complex homomorphism $\widetilde{\omega}_\lambda$ of this algebra, which corresponds uniquely to a spherical function Ψ_λ. Thus, if F is bi-K-invariant,

$$\pi_{\Gamma\backslash G}(F)(\phi)=\widetilde{\omega}_\lambda(F)\phi,\quad \text{where } \widetilde{\omega}_\lambda(F)=\int_G F(g)\Psi_\lambda(g)\,dm_G(g).$$

We begin with the following result.

PROPOSITION 6.2. (The average size of eigenfunctions) *Let Γ be any uniform lattice in G, satisfying the condition preceding Theorem* 6.1, *and let ϕ_j, $j\in\mathbb{N}$, be an orthonormal basis for $L^2(M)$ consisting of joint eigenfunctions of the ring $\mathcal{D}$ of G-invariant differential operators on M. Denote by Ψ_{λ_j} the spherical function associated with ϕ_j. Then there exists a constant $C(\mathcal{S})$, depending only on $\mathcal{S}$ and independent of Γ, such that*

$$\max_{z\in M}\sum_{\|\lambda_j\|\leqslant\|\lambda\|}|\phi_j(z)|^2\leqslant C(\mathcal{S})(1+\|\lambda\|)^{\dim(\mathcal{S})}.$$

Proof. Let $\ell_\varepsilon(z,w)$ be a smooth non-negative positive-definite kernel on $\mathcal{S}\times\mathcal{S}$, depending only on $d(z,w)$ and supported in $d(z,w)\leqslant\varepsilon$ with unit integral. The automorphic kernel

$$A_\varepsilon(\Gamma z,\Gamma w)=\sum_{\gamma\in\Gamma}\ell_\varepsilon(\gamma z,w)$$

on $M\times M$ can be expanded in terms of the joint eigenfunctions ϕ_j of $\mathcal{D}$ on $L^1(K\backslash G/K)$ as

$$A_\varepsilon(\Gamma z,\Gamma w)=\sum_{j=1}^{\infty}h_{A_\varepsilon}(\lambda_j)\phi_j(\Gamma z)\overline{\phi_j(\Gamma w)},$$

where $h_{A_\varepsilon}(\lambda_j)$ are the eigenvalues of the operator defined on $L^2(M)$ by the automorphic kernel A_ε. As noted above, these eigenvalues are given by the Selberg and Harish-Chandra spherical transform (normalized at $w=w_0$)

$$h_{A_\varepsilon}(\lambda_j)=\int_{\mathcal{S}}\ell_\varepsilon(z,w_0)\Psi_{\lambda_j}(z)\,d\mathrm{vol}(z),$$

where Ψ_{λ_j} is the spherical function associated with ϕ_{λ_j}, normalized by $\Psi_\lambda(e)=1$, and viewed as a function on $\mathcal{S}$.

We claim that

$$|h_{A_\varepsilon}(\lambda)-1|\leqslant\left|\int_{z\in B_\varepsilon(w_0)}\ell_\varepsilon(z,w_0)|\Psi_\lambda(z)-1|\,d\mathrm{vol}(z)\right|\leqslant C_1(\mathcal{S})(1+\|\lambda\|)\varepsilon.$$

Clearly, this estimate follows from the fact that for all H in the unit sphere in $\mathfrak{a}$ and $|t|\leqslant1$ (say), the first derivative of the normalized positive definite spherical functions Ψ_λ satisfy

$$\left|\frac{d}{dt}\Psi_\lambda(\exp(tH))\right|\leqslant C_1(\mathcal{S})(1+\|\lambda\|).$$

This estimate is a consequence of the Harish-Chandra power series expansion for the spherical functions, together with the fact that normalized positive definite spherical functions are all bounded by 1. The estimate follows from e.g. [GV, Proposition 4.6.2].

We conclude that if $(1+\|\lambda\|)\varepsilon<\frac{1}{2}C_1(\mathcal{S})$ then $h_{A_\varepsilon}(\lambda)\geqslant\frac{1}{2}$, and therefore we obtain the following upper bound on the average size of the eigenfunctions

$$\sum_{\gamma\in\Gamma}\ell_\varepsilon(\gamma z,z)=\sum_{j=1}^{\infty}h_{A_\varepsilon}(\lambda_j)|\phi_j(z)|^2\geqslant\frac{1}{2}\sum_{\|\lambda_j\|<C_2(\mathcal{S})/\varepsilon}|\phi_j(z)|^2.$$

On the other hand, we can clearly obtain a pointwise upper bound of the form

$$\sum_{\gamma\in\Gamma}\ell_\varepsilon(\gamma z,z)\leqslant C_3(\mathcal{S})\varepsilon^{-\dim\mathcal{S}}.$$

Indeed, this follows when the kernel is defined by a bump function which satisfies the obvious upper bound of being $\ll\varepsilon^{-\dim\mathcal{S}}$, and taking also into account the fact that there are at most c lattice points in a ball of radius $\varepsilon\leqslant1$, this coming from our assumption on Γ. Combining the two estimates, we can conclude that

$$\sum_{\lambda_j<\lambda=C_2(\mathcal{S})/\varepsilon}|\phi_j(z)|^2\leqslant C_3(\mathcal{S})\varepsilon^{-\dim\mathcal{S}}\leqslant C(\mathcal{S})(1+\|\lambda\|)^{\dim\mathcal{S}}.$$

The proof of Proposition 6.2 is now complete. □

Proof of Theorem 6.1. Consider the identity

$$\sum_{\gamma\in\Gamma}\int_{\mathcal{S}}L_t(\gamma z,w)b(w)\,dw=\sum_{\gamma\in\Gamma}W_t(\gamma z)=\sum_{j=0}^{\infty}h_{L_t}(\lambda_j)\phi_j(z)\int_{\mathcal{S}}\overline{\Psi_{\lambda_j}(w)}b(w)\,dw.\tag{6.2}$$

The eigenvalue $\lambda_0=0$ associated with the constant function $\phi_0=1/\mathrm{vol}(M)$ (the unique Δ-eigenfunction with this eigenvalue) gives the main contribution to the infinite sum, which is $\mathrm{vol}(B_t)/\mathrm{vol}(M)=h_{L_t}(0)$. We must therefore estimate the contribution of all other terms.

Now note that since the bump function $b(w)$ is a fixed smooth function, and ϕ_j is an eigenfunction of the Laplacian Δ with eigenvalue $\omega_{\lambda_j}(\Delta)$, m integrations by parts give, for any fixed m and all $j\in\mathbb{N}$,

$$\int_{\mathcal{S}}b(w)\overline{\Psi_{\lambda_j}(w)}\,dw\leqslant C_m(1+\|\lambda_j\|)^{-m}.$$

Let us set

$$\hat{b}(\lambda_j)=\int_{\mathcal{S}}b(w)\overline{\Psi_{\lambda_j}(w)}\,dw.$$

Recall that λ_j, $j\neq0$, is a discrete set, and thus have a fixed positive distance from 0, due to our spectral gap assumption. As a consequence, the spherical functions Ψ_{λ_j} all have a fixed rate of decay, which can be expressed as a negative power of the volume of B_t.

Now,

$$h_{L_t}(\lambda) = \int_{\mathcal{S}} L_t(z, w_0)\overline{\Psi_\lambda(z)}\, dz$$

is the average of the spherical function on a ball of radius t and center w_0, which by Hölder's inequality is estimated by $\mathrm{vol}(B_t^\delta)$, where $\delta = 1 - 1/p + \eta$.

Therefore, using (6.2), we can write

$$\left| \sum_{\gamma\in\Gamma} W_t(\gamma z) - \frac{\mathrm{vol}(B_t)}{\mathrm{vol}(M)} \right| \leqslant \mathrm{vol}(B_t)^\delta \sum_{j\neq 0} |\phi_j(z)|\, |\hat{b}(\lambda_j)|.$$

Now, $|\hat{b}(\lambda_j)| \leqslant C_m(1+\|\lambda_j\|)^{-m}$, and by Proposition 6.2,

$$\sum_{\lambda_j \leqslant \lambda} |\phi_j(z)|^2 \leqslant C(\mathcal{S})(1+\|\lambda\|)^{\dim \mathcal{S}},$$

so that, upon choosing k large enough,

$$\sum_{j\neq 0} |\phi_j(z)|\, |\hat{b}(\lambda_j)| \leqslant \left(\sum_{j\neq 0} (1+\|\lambda_j\|)^{-2k} |\phi_j(z)|^2 \right)^{1/2} \left(\sum_{j\neq 0} (1+\|\lambda_j\|)^{2k} |\hat{b}(\lambda_j)|^2 \right)^{1/2} < \infty.$$

This concludes the proof of Theorem 6.1. □

Remark 6.3. The error estimate in the Poisson summation formula can be similarly established when Γ is non-uniform, using the foregoing arguments and the theory of Eisenstein series.

Next, we apply Theorem 6.1 in the context of sieving as in §§2–4. For the purpose of the lower bound sieve, we can use the non-negative weight function W_t in (6.1). Using our previous setup and notation, let us work with the distance parameter $T = e^t$, where t denotes the distance in the symmetric space $\mathcal{S}$. But notice that since we are now working with symmetric space distance and not with a norm, in general the exponent of volume growth is now $2\|\varrho_G\|$, namely the rate of volume growth for Riemannian balls in $\mathcal{S}$. Recall also that $t(f)$ denotes the number of irreducible factors of the polynomial f. Now consider

$$S_{W_T}(\mathcal{A}, P) := \sum_{\substack{\gamma\in\Gamma \\ (f(\gamma v), P)=1}} W_T(\gamma), \tag{6.3}$$

where $W_T(\gamma) := W_T(\gamma z_0)$ for a fixed $z_0 \in \Gamma\backslash G/K$.

We have

$$0 \leqslant W_T(\gamma) \leqslant 1,$$

and for T large

$$W_T(\gamma)=\begin{cases}1, & \text{if } \|\gamma\|\leqslant \frac{1}{2}T,\\ 0, & \text{if } \|\gamma\|\geqslant 2T,\end{cases} \tag{6.4}$$

where $\|\cdot\|$ is a bi-K-invariant norm on G.

Theorem 6.1 (assuming, as we do from now on, that Γ is co-compact) gives the conclusion that uniformly for $y\in\Gamma$,

$$\frac{1}{\mathrm{vol}(B_T)}\sum_{\gamma\in\Gamma(q)}W_T(\gamma y)=\frac{1}{[\Gamma:\Gamma(q)]}+O_\varepsilon(T^{-a/p+\varepsilon}). \tag{6.5}$$

This corresponds to Theorem 3.2 with $\theta/(1+\dim G)=a/2n_e(1+\dim G)$ replaced by a/p, where $p=p(G,\Gamma)$. Running the rest of the sieving analysis with this positive smooth weight W_T to the end of §4 yields an improvement in Theorem 1.7 and Corollary 1.8 with the condition on r replaced by

$$r>\frac{9t(f)p(G,\Gamma)(\dim G)\deg f}{a}. \tag{6.6}$$

Note that we have incorporated the small improvement (4.25) of (4.24) as well.

To further improve this value of r, we use the weighted sieve ([HR, Chapter 10]) in place of the elementary sieve in §2. The form which is convenient for us is as follows:

Let $a_k\geqslant 0$ be a finite sequence and assume that for $d\geqslant 1$,

$$\sum_{k\equiv 0 \pmod d}a_k=\frac{\varrho(d)}{d}X+R(\mathcal{A},d), \tag{6.7}$$

with $R(\mathcal{A},1)=0$, $\varrho(1)=1$ and ϱ multiplicative, and that $\varrho(p)$ satisfies (2.5) for all $p\geqslant 2$. Concerning the sieve dimension t, assume that for $2\leqslant z_1\leqslant z$ we have

$$\prod_{z_1\leqslant p<z}\left(1-\frac{\varrho(p)}{p}\right)^{-1}\leqslant\left(\frac{\log z}{\log z_1}\right)^t\left(1+\frac{A}{\log z_1}\right) \tag{6.8}$$

for some fixed constant A.

Assume that we have a level distribution τ, that is for $\varepsilon>0$,

$$\sum_{d\leqslant X^\tau}|R(\mathcal{A},d)|\underset{\varepsilon}{\ll}X^{1-\varepsilon}. \tag{6.9}$$

Define μ by

$$\max_{a_n\in\mathcal{A}}n\leqslant X^{\tau\mu}. \tag{6.10}$$

Let P_r denote the set of positive integers with at most r-prime factors. Then for any $0<\zeta<\nu_t$ (where ν_t is the sieve limit in dimension t) and for any r satisfying

$$r>\left(1+\zeta-\frac{\zeta}{\nu_t}\right)\mu-1+(t+\zeta)\log\frac{\nu_t}{\zeta}-t+\frac{\zeta t}{\nu_t}, \tag{6.11}$$

there is $\delta=\delta(t,\mu,r,\zeta)>0$ such that

$$\sum_{k\in P_r} a_k \geqslant \delta\frac{X}{(\log X)^t}. \tag{6.12}$$

For more on the sieve limit ν_t in dimension t see [HR], where a table is given and the fact that $\nu_t\leqslant 4t$ is established. The latter fact is what we will use below.

We apply this to our sequence

$$a_k(T)=\sum_{|f(\gamma v)|=k} W_T(\gamma). \tag{6.13}$$

By (6.5) and the analysis in §4, we have

$$\tau=\frac{1}{p\dim G}, \tag{6.14}$$

$$\mu=\frac{p(\dim G)\deg f}{a}. \tag{6.15}$$

Taking $\zeta=1$ in (6.11) for simplicity leads to (6.12) holding for

$$r>\frac{2p(\dim G)\deg f}{a}-1+(t(f)+1)\log 4t(f)-t(f)+\frac{1}{4}. \tag{6.16}$$

In particular, Theorems 1.7 and Corollary 1.8 are valid for such r.

As noted in Remark 6.3, these considerations also apply to the case G/Γ non-compact. In particular to $\Gamma=\mathrm{SL}_n(\mathbb{Z})$ and to $V_{n,m}(\mathbb{Z})$. In this case $p=2(n-1)$ for $n\geqslant 3$ and it is estimated in (6.18) for $n=2$. Finally $a=n(n-1)$, so that Theorem 1.2 and Corollary 1.3 are valid with

$$r>4n\deg f-1+(t(f)+1)\log 4t(f)-t(f)+\frac{1}{4} \tag{6.17}$$

for $n>2$.

Our final improvement comes in the cases where much stronger bounds towards the Ramanujan conjectures are valid, especially with n large. Let $D/\mathbb{Q}$ be a division algebra of degree n which, for the reasons below, we assume is itself prime. Assume that $D\otimes\mathbb{R}\cong\mathrm{Mat}_{n\times n}(\mathbb{R})$ and let N_r denote the reduced norm on D. Let

$$V_{m,D}=\{x\in D(\mathbb{Z}):N_r(x)=m\},$$

with $m\neq 0$. Here the $\mathbb{Z}$-structure is given by the defining equations of $D/\mathbb{Q}$ in A^N, $N=n^2$, $G=\{x:N_r(x)=1\}$ and Γ is the set of the integral elements of reduced norm equal to 1. These act on $V_{D,m}$ making it into a principal homogeneous space. Let $f\in\mathbb{Q}[x_{ij}]$ which is weakly primitive on $V_{D,m}(\mathbb{Z})$. The discussion of this section applies to the question of the saturation number $r_0(V_{D,m}(\mathbb{Z}),f)$. What is pleasant about such compact quotients $G(\mathbb{R})/\Gamma(q)$ coming from these division algebras is that we have very good upper bounds for their corresponding p's. Specifically any representation π occuring in $L^2_0(G(\mathbb{R})/\Gamma(q))$ corresponds to an automorphic representation occuring in $L^2(D(\mathbb{A})/D(\mathbb{Q}))$ which in turn, via the Jacquet–Langlands correspondence [JL] if $n=2$ and Arthur–Clozel [AC] if $n>2$, lifts to an automorphic cuspidal representation π of $\mathrm{GL}_n(\mathbb{A})/\mathrm{GL}_n(\mathbb{Q})$ (it is here that we assume that n is prime so that π is not a residual Eisenstein series [MW]). Applying the best known bounds towards the Ramanujan conjectures "at infinity" (see [S1] for a survey) for such π, we conclude that

$$p_n := p(K\backslash G(\mathbb{R})/\Gamma(q)) \leqslant \begin{cases} \dfrac{64}{25}, & \text{if } n=2,\\ \dfrac{28}{9}, & \text{if } n=3,\\ \dfrac{2n}{n-2}, & \text{if } n\geqslant 4 \text{ is even},\\ \dfrac{2(n+1)}{n-1}, & \text{if } n\geqslant 5 \text{ is odd}. \end{cases} \tag{6.18}$$

These follow from (24), (22) and (13) in [S1] by computing p using Theorem 8.48 in [K]. For $V_{D,m}(\mathbb{Z})$ we have $a=n^2-n$ and $\dim G=n^2-1$. Hence, (6.16) leads to the following result.

THEOREM 6.4. *Let $V_{D,m}(\mathbb{Z})\subset A^N$ be the set of integral points of norm m in D, which we assume is non-empty. Let $f\in\mathbb{Q}[x_{ij}]$ be of degree d and assume that f factors into t irreducible factors in the coordinate ring $\overline{\mathbb{Q}}[V_{D,m}]$ and that f is $V_{D,m}(\mathbb{Z})$-weakly primitive. Then*,

$$r_0(V_{D,m}(\mathbb{Z}),f)\leqslant 2p_n\frac{n+1}{n}d+(t+1)\log 4t-t.$$

References

[AC] ARTHUR, J. & CLOZEL, L., *Simple Algebras, Base Change, and the Advanced Theory of the Trace Formula*. Annals of Mathematics Studies, 120. Princeton University Press, Princeton, NJ, 1989.

[BMS] BASS, H., MILNOR, J. & SERRE, J. P., Solution of the congruence subgroup problem for SL_n ($n\geqslant 3$) and Sp_{2n} ($n\geqslant 2$). *Inst. Hautes Études Sci. Publ. Math.*, 33 (1967), 59–137.

[B] Borel, A., Density properties for certain subgroups of semi-simple groups without compact components. *Ann. of Math.*, 72 (1960), 179–188.

[BH] Borel, A. & Harish-Chandra, Arithmetic subgroups of algebraic groups. *Ann. of Math.*, 75 (1962), 485–535.

[BGS] Bourgain, J., Gamburd, A. & Sarnak, P., Affine linear sieve, expanders, and sum-product. *Invent. Math.*, 179 (2010), 559–644.

[BS] Burger, M. & Sarnak, P., Ramanujan duals. II. *Invent. Math.*, 106 (1991), 1–11.

[Ch] Chebotarev, N. G., Determination of the density of the set of prime numbers, belonging to a given substitution class. *Izv. Ross Akad. Nauk*, 17 (1923), 205–250 (Russian).

[Cl1] Clozel, L., Démonstration de la conjecture τ. *Invent. Math.*, 151 (2003), 297–328.

[Cl2] — Spectral theory of automorphic forms, in *Automorphic Forms and Applications*, IAS/Park City Math. Ser., 12, pp. 43–93. Amer. Math. Soc., Providence, RI, 2007.

[Co] Colin de Verdière, Y., Ergodicité et fonctions propres du laplacien. *Comm. Math. Phys.*, 102 (1985), 497–502.

[DH] Diamond, H. G. & Halberstam, H., *A Higher-Dimensional Sieve Method*. Cambridge Tracts in Mathematics, 177. Cambridge University Press, Cambridge, 2008.

[DRS] Duke, W., Rudnick, Z. & Sarnak, P., Density of integer points on affine homogeneous varieties. *Duke Math. J.*, 71 (1993), 143–179.

[GV] Gangolli, R. & Varadarajan, V. S., *Harmonic Analysis of Spherical Functions on Real Reductive Groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete, 101. Springer, Berlin–Heidelberg, 1988.

[GGPY] Goldston, D. A., Graham, S. W., Pintz, J. & Yildirim, C. Y., Small gaps between products of two primes. *Proc. Lond. Math. Soc.*, 98 (2009), 741–774.

[GN1] Gorodnik, A. & Nevo, A., *The Ergodic Theory of Lattice Subgroups*. Annals of Mathematics Studies, 172. Princeton University Press, Princeton, NJ, 2010.

[GN2] — Counting lattice points. To appear in *J. Reine Angew. Math.*

[GW] Gorodnik, A. & Weiss, B., Distribution of lattice orbits on homogeneous varieties. *Geom. Funct. Anal.*, 17 (2007), 58–115.

[HR] Halberstam, H. & Richert, H. E., *Sieve Methods*. London Mathematical Society Monographs, 4. Academic Press, London–New York, 1974.

[IK] Iwaniec, H. & Kowalski, E., *Analytic Number Theory*. American Mathematical Society Colloquium Publications, 53. Amer. Math. Soc., Providence, RI, 2004.

[JL] Jacquet, H. & Langlands, R. P., *Automorphic Forms on* GL(2). Lecture Notes in Mathematics, 114. Springer, Berlin–Heidelberg, 1970.

[KeS] Kelmer, D. & Sarnak, P., Strong spectral gaps for compact quotients of products of $\mathrm{PSL}(2,\mathbb{R})$. *J. Eur. Math. Soc.* (*JEMS*), 11 (2009), 283–313.

[KiS] Kim, H. H. & Sarnak, P., Appendix 2 in Kim, H. H., Functoriality for the exterior square of GL_4 and the symmetric fourth of GL_2. *J. Amer. Math. Soc.*, 16 (2003), 139–183.

[K] Knapp, A. W., *Representation Theory of Semisimple Groups*. Princeton Mathematical Series, 36. Princeton University Press, Princeton, NJ, 1986.

[La] Lang, S., Algebraic groups over finite fields. *Amer. J. Math.*, 78 (1956), 555–563.

[LW] Lang, S. & Weil, A., Number of points of varieties in finite fields. *Amer. J. Math.*, 76 (1954), 819–827.

[Li] Li, J.-S., The minimal decay of matrix coefficients for classical groups, in *Harmonic Analysis in China*, Math. Appl., 327, pp. 146–169. Kluwer, Dordrecht, 1995.

[LiZ] Li, J.-S. & Zhu, C.-B., On the decay of matrix coefficients for exceptional groups. *Math. Ann.*, 305 (1996), 249–270.

[LS] Liu, J. & Sarnak, P., Integral points on quadrics in three variables whose coordinates have few prime factors. *Israel J. Math.*, 178 (2010), 393–426.

[LoS] Loke, H. Y. & Savin, G., Rank and matrix coefficients for simply laced groups. *J. Reine Angew. Math.*, 599 (2006), 201–216.

[MVW] Matthews, C. R., Vaserstein, L. N. & Weisfeiler, B., Congruence properties of Zariski-dense subgroups. I. *Proc. London Math. Soc.*, 48 (1984), 514–532.

[Ma] Maucourant, F., Homogeneous asymptotic limits of Haar measures of semisimple linear groups and their lattices. *Duke Math. J.*, 136 (2007), 357–399.

[Me] Mennicke, J. L., Finite factor groups of the unimodular group. *Ann. of Math.*, 81 (1965), 31–37.

[MW] Mœglin, C. & Waldspurger, J. L., Le spectre résiduel de GL(n). *Ann. Sci. École Norm. Sup.*, 22 (1989), 605–674.

[Ne] Nevo, A., Spectral transfer and pointwise ergodic theorems for semi-simple Kazhdan groups. *Math. Res. Lett.*, 5 (1998), 305–325.

[No] Noether, E., Ein algebraisches Kriterium für absolute Irreduzibilität. *Math. Ann.*, 85 (1922), 26–40.

[O] Oh, H., Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants. *Duke Math. J.*, 113 (2002), 133–192.

[PR] Platonov, V. P. & Rapinchuk, A. S., *Algebraic Groups and Number Theory*. Nauka, Moscow, 1991 (Russian).

[Sa] Sansuc, J. J., Groupe de Brauer et arithmétique des groupes algébriques linéaires sur un corps de nombres. *J. Reine Angew. Math.*, 327 (1981), 12–80.

[S1] Sarnak, P., Notes on the generalized Ramanujan conjectures, in *Harmonic Analysis, the Trace Formula, and Shimura Varieties*, Clay Math. Proc., 4, pp. 659–685. Amer. Math. Soc., Providence, RI, 2005.

[S2] — Equidistribution and primes. *Astérisque*, 322 (2008), 225–240.

[Sch] Schmidt, W. M., *Equations over Finite Fields. An Elementary Approach*. Lecture Notes in Mathematics, 536. Springer, Berlin–Heidelberg, 1976.

[T] Terras, A., *Harmonic Analysis and Symmetric Spaces and Applications*. II. Springer, Berlin–Heidelberg, 1988.

[Va] Vaughan, R. C., *The Hardy–Littlewood Method*. Cambridge Tracts in Mathematics, 80. Cambridge University Press, Cambridge, 1981.

[Vi] Vinogradov, I. M., Representations of an odd number as a sum of three primes. *Dokl. Akad. Nauk SSSR*, 15 (1937), 291–294 (Russian).

Amos Nevo
Technion – Israel Institute of Technology
Technion City
Haifa 32000
Israel
anevo@tx.technion.ac.il

Peter Sarnak
School of Mathematics
Institute for Advanced Study
Einstein Drive
Princeton, NJ 08540
U.S.A.
sarnak@math.princeton.edu

Received November 6, 2008
Received in revised form October 5, 2009