HOW SMALL MUST ILL-DISTRIBUTED SETS BE?

H. A. HELFGOTT AND A. VENKATESH

ABSTRACT. Consider a set $S \subset \mathbb{Z}^n$. Suppose that, for many primes p, the distribution of S in congruence classes $\bmod p$ is far from uniform. How sparse is S forced to be thereby?

A clear dichotomy appears: it seems that S must either be very small or possess much algebraic structure. We show that, if $S \subset \mathbb{Z}^2 \cap [0, N]^2$ occupies few congruence classes $\bmod p$ for many p, then either S has fewer than N^ϵ elements or most of S is contained in an algebraic curve of degree $O_\epsilon(1)$. Similar statements are conjectured for $S \subset \mathbb{Z}^n$, $n \neq 2$.

We follow an approach that combines ideas from the larger sieve of Gallagher [Ga] and from the work of Bombieri and Pila [BP]. All techniques used are elementary.

1. INTRODUCTION

Let S be a subset of $\mathbb{Z}^n$, $n \geq 1$. Assume, now and henceforth, that S is far from being uniformly distributed modulo p, and that this holds for every p in a large set of primes. (For example, let $S_0 \subset \mathbb{Z}$, and assume that, for every prime p greater than a given constant, there are at least 0.01 congruence classes $\bmod p$ on which no element of S_0 lies.) The elements of S are thus, in an average sense, abnormal, and this fact should force S to be small. How small?

Of course, S might be infinite; we may more precisely ask whether S is *sparse*; that is, we desire a bound on how many elements of S there can be in an interval of length N. Large and larger sieves furnish upper bounds on the number of such elements; these bounds are of the form $O(N^\delta)$, for some $\delta > 0$. If, for instance, we consider the set S_0 defined above, Gallagher's larger sieve [Ga, Thm. 1] tells us that the number $|S_0 \cap [0, N]|$ of elements of S_0 between 0 and N is at most

$$|S_0 \cap [0, N]| \ll N^{0.99}, \tag{1.1}$$

where the implied constant is absolute. Large sieves go further than the larger sieve in this instance: [Bo, Thm. 6] gives us the bound

$$|S_0 \cap [0, N]| \ll N^{1/2} (\log N)^c, \tag{1.2}$$

where c and the implied constant are absolute.

Are the upper bounds given by large and larger sieves optimal? In, say, the case of S_0, is (1.2) tight? There are sets for which such bounds are optimal. Take, for example, the set S_0 of all squares. Then $S_0 \bmod p$ avoids, not merely $0.01p$, but $\frac{1}{2}(p-1)$ residue classes $\bmod p$ for every prime p. The number of elements of this set S_0 between 0 and N is, of course, $\lfloor N^{1/2} \rfloor$. Thus, it would seem, (1.2) is optimal.

However, such a set S_0 is clearly not typical: the set S_0 of all squares is algebraic in a very strong sense[1]. Could it be that bounds such as (1.1) and (1.2) are optimal, or can be optimal, only for sets S that are strongly algebraic? That is – given a set S whose distribution $\bmod p$ is far from uniform for every p, is it the case that S must be either strongly algebraic or very sparse? We make some speculations in this direction in Section 4.2.

* * *

For subsets $S \subset \mathbb{Z}^2$ we can prove the following theorem.

Theorem 1.1. *Let $S \subset \mathbb{Z}^2 \cap [0,N]^2$, $N \geq 1$. Suppose that the number of residues $\{(x,y) \bmod p : (x,y) \in S\}$ is at most αp for some fixed $\alpha > 0$ and for every prime p.*

Then, for any $\epsilon > 0$, either

- $|S| \ll_{\alpha,\epsilon} N^\epsilon$ *or*
- *there is an algebraic plane curve C of degree $O_{\alpha,\epsilon}(1)$ such that at least $(1-\epsilon)|S|$ points of S lie on C.*

Let us clarify the quantifiers here: the conclusion means that there are functions $c_1 = c_1(\alpha,\epsilon)$ and $c_2 = c_2(\alpha,\epsilon)$ such that either $|S| \leq c_1(\alpha,\epsilon)N^\epsilon$ holds or $(1-\epsilon)|S|$ points lie on a curve of degree $c_2(\alpha,\epsilon)$. Here and from now on, $|S|$ denotes the number of elements of a set S. The plane curve C may, of course, be reducible.

The assumption that S falls into at most αp congruence classes modulo p is rather strong, even if α is large: a typical subset of $\mathbb{Z}^2 \cap [0,N]^2$ should cover all or almost all of the p^2 congruence classes (x,y), $x, y \in \mathbb{Z}/p\mathbb{Z}$. Still, such a condition is fulfilled for S equal to the intersection of $[0,N]^2$ and the set of integer points on a plane curve C: by Weil's bounds, there are at most $p + O_g(p^{1/2})$ points over $\mathbb{Z}/p\mathbb{Z}$ on C, where g is the genus of C.

Our procedure allows us to obtain quantitative results on the cardinality of S when S has a large intersection with a curve C of low degree: we recover (in Proposition 3.2) the bounds in [BP], up to an ϵ in the exponent. One may, in fact, see our method as a reinterpretation of [BP] from a local perspective. Seen from another angle, what we have is a generalization of the larger sieve. More precisely, we obtain local data much as in [BP] and combine the data from different primes much as in the larger sieve. We recall the idea of the larger sieve in Section 2 and give the proof of Theorem 1.1 in Section 3.

* * *

In the context of Theorem 1.1, one might expect a bound of the form $|S| \ll_{\alpha,\epsilon} N^\epsilon$ to hold whenever S does not have a large intersection with a rational curve. Such a bound would follow from Theorem 1.1 and the folkloric conjecture that there are at most $O_{d,\epsilon}(N^\epsilon)$

---

[1] Any recursively enumerable set is diophantine [Ma]; that is, just about any reasonable subset of $\mathbb{Z}^n$ is the projection onto $\mathbb{Z}^n$ of the set of integer solutions to some equation of high degree in many variables. It should be intuitively clear that the set of squares is more strongly algebraic than a set that is merely known to be diophantine: the set of all squares is defined by an equation of degree 2 in one variable.

We could take "defined by an equation of low degree in a few variables" to be a working definition of "strongly algebraic", except for the fact that we would want the term "strongly algebraic" to be robust: if S is defined to be the union of (a) a large subset of the set of all squares and (b) any very small set, we still want to be able to say that S is strongly algebraic. We will not attempt to give a formal definition of "strongly algebraic"; it is a term we will use only in informal discussion, and we will do without it in the statements of our results.

integer points in $[0, N]^2$ on any curve of positive genus. At the same time, such a bound would imply the said conjecture; since the latter is reputed to be very hard, we should not hope to prove the former for the while being.

To fall in few congruence classes modulo p for every p is only one way in which a set S may be far from being uniformly distributed modulo p. It is possible to prove results along the lines of Theorem 1.1 for all sets S that are far from being uniformly distributed, even if they occupy all or many congruence classes $\bmod p$. Instead of the cardinality of $S \bmod p$, one could use assumptions on the ℓ_1 and ℓ_2 norms of $P \mapsto |\{Q \in S : Q \equiv P \bmod p\}|$. We shall hew to a treatment in terms of the cardinality of $|S \bmod p|$ for the sake of clarity.

Acknowledgements. The first author was supported by a CRM-ISM postdoctoral fellowship. The second author was supported by a Clay research fellowship, NSF Grants DMS–0245606 and NSF Grants DMS–0111298; he also thanks the Institute for Advanced Study for providing superb working conditions. Many of the ideas of this paper were developed during the CRM summer school "Equidistribution in Number Theory" held in Montréal, 2005. We are grateful to the CRM, as well as to Andrew Granville and Trevor Wooley for their encouragement.

2. Procedural overview

Let us begin by stating what is essentially Gallagher's larger sieve. Suppose a given set $S \subset \mathbb{Z} \cap [0, N]$ intersects $\leq \alpha p$ congruence classes $\bmod p$ for every $p > c$, where $\alpha \in (0, 1)$ and $c > 0$ are given constants. We wish to bound the cardinality of $|S|$.

To this purpose, we consider the product $\Delta = \prod_{x,y \in S, x \neq y} (x - y)$. Take a prime $p > c$. There are at least $\frac{|S|^2}{\alpha p} - |S|$ pairs $(x, y) \in S^2$, $x \neq y$, such that $x \equiv y \bmod p$. Hence $v_p(\Delta) \geq \frac{|S|^2}{\alpha p} - |S|$. By multiplying $p^{v_p(\Delta)}$ over all primes p between c and $|S|$, we obtain that

$$|\Delta| \geq \prod_{c<p\leq|S|} p^{v_p(\Delta)} \geq \prod_{c<p\leq|S|} e^{\left(\frac{|S|^2}{\alpha p} - |S|\right) \log p} \geq e^{\frac{|S|^2}{\alpha} (\log |S| - O(1))}$$

Comparing this lower bound for $|\Delta|$ with the trivial upper bound $|\Delta| \leq \prod_{x,y \in S, x \neq y} N \leq N^{|S|^2}$, we obtain that

$$e^{\alpha^{-1} \log |S| - O(1)} \leq N,$$

and thus

$$|S| = e^{\log |S|} \leq e^{\alpha(\log N + O(1))} \ll_\alpha N^\alpha,$$

as we desired to show. The bound $|S| \ll_\alpha N^\alpha$ is Gallagher's.

Our task is to develop a two-dimensional analogue of the above method, viz., an upper bound for $|S|$ when $S \subset \mathbb{Z}^2 \cap [0, N]^2$ is a set such that $S \bmod p$ is $O(p)$ for every p. In the proof just given, the function $w(x, y) = x - y$ detects whether two numbers x, y are distinct. In the two dimensional case, we replace it by a function of several variables $W(P_1, P_2, \ldots, P_n)$, where each variable P_j is in $\mathbb{R}^2$. The function W will detect whether $P_1, P_2, \ldots, P_n$ fail to be in general position, i.e., whether they all lie on a curve of low degree.

Using W much like $w(x, y) = x - y$ was used in the larger sieve, we will show that we are in an either-or situation: either $|S|$ is small or a large proportion of all tuples

$(P_1, P_2, \ldots, P_n) \subset S^n$ fail to be in general position. In the latter case, a little work suffices to show that a large proportion of the points on S lie on a single curve of low degree.

3. A two-dimensional larger sieve

Let $\mathscr{W}$ be a set consisting of finitely many linearly independent polynomials $f : \mathbb{Z}^2 \to \mathbb{Z}$, each with integral coefficients. Assume that $\mathscr{W}$ contains the map $(x, y) \mapsto 1$, and that the elements of $\mathscr{W}$ separate points; that is, we assume that, for all $P_1, P_2 \in \mathbb{Z}^2$, there is an $f \in \mathscr{W}$ such that $f(P_1) \neq f(P_2)$.

All polynomials f descend to congruence classes; that is, for any $P_1, P_2 \in \mathbb{Z}^2$ and any p such that $P_1 \equiv P_2 \bmod p$, we have $f(P_1) \equiv f(P_2) \bmod p$. While in principle we might state our results using more general f having this same property, practice compels us to use sets $\mathscr{W}$ composed exclusively of polynomial maps: while there are non-polynomial functions $f : \mathbb{Z}^2 \to \mathbb{Z}$ that descend to congruence classes ([Ha, Thm. 1]), they all grow extremely rapidly ([Ha, Thm. 3]), and thus they will not do for our purposes.

Write $d_{\mathscr{W}}$ for the total degree of all elements of $\mathscr{W}$, and, as usual, $|\mathscr{W}|$ for the cardinality of $\mathscr{W}$. We define a *$\mathscr{W}$-curve* to be an affine algebraic curve described by a single equation $g(x, y) = 0$, where g belongs to the linear span of $\mathscr{W}$.

We may consider the following two examples.

(a) Let $\mathscr{W}$ be the set of monomials $x^i y^j$ with $i + j \leq d$, where $d \geq 0$ is given. Then $|\mathscr{W}| = \frac{(d+1)(d+2)}{2}$ and $d_{\mathscr{W}} = \frac{d(d+1)(d+2)}{2}$. The $\mathscr{W}$-curves are the plane curves of degree $\leq d$.

(b) Let $\mathscr{W}$ be the set of monomials $x^i y^j$ with $i \leq d$ and $j \leq M$, where d and M are given. Then $|\mathscr{W}| = (d + 1)(M + 1)$ and $d_{\mathscr{W}} = (d + 1)(M + 1)\frac{d+M}{2}$. The $\mathscr{W}$-curves are the plane curves having degrees $\leq d$ and $\leq M$ in x and y, respectively.

The choice of $\mathscr{W}$ in (a) may seem natural, and it will in fact be used (with d approaching ∞) to derive Theorem 1.1. However, for the purpose of proving bounds on the number of integral points on an algebraic curve, it will be best to apply (b) with M approaching infinity. (The choice (b) is taken directly from the work of Bombieri and Pila ([BP], [Pi]).)

Let us state our main intermediate result for general $\mathscr{W}$.

Proposition 3.1. *Let $S \subset \mathbb{Z}^2 \cap [0, N]^2$, $N \geq 1$. Suppose that the number of residues $\{(x, y) \bmod p : (x, y) \in S\}$ is at most αp for some fixed $\alpha > 0$ and for every prime p larger than a constant c.*

Let $\mathscr{W}$ be a set consisting of finitely many linearly independent polynomials $f : \mathbb{Z}^2 \to \mathbb{Z}$, each with integral coefficients, and including the map $(x, y) \mapsto 1$. Assume that the elements of $\mathscr{W}$ separate points.

Then, for every $\delta \in (0, 1)$, either

(a) *there is a $\mathscr{W}$-curve containing at least $\delta|S|$ points of S, or*

(b) $|S| \ll_{c,\delta,\mathscr{W}} N^{\frac{2\alpha d_{\mathscr{W}}}{w(w-1)} + O_{\alpha,\mathscr{W}}(\delta)}$, *where $w = |\mathscr{W}|$.*

Proof. Fixing an arbitrary ordering $f_1, f_2, \ldots, f_w$ for the elements of $\mathscr{W}$, we define a function $W : (\mathbb{R}^2)^w \to \mathbb{R}$ by

$$W(P_1, \ldots, P_w) = \det\left(f_i(P_j)\right)_{1 \leq i,j \leq w}.$$

We note for future reference the following property of $W(P_1, \ldots, P_w)$: if the number of distinct points among the set $(P_1, \ldots, P_w) \bmod p$ is no greater than k, then $W(P_1, \ldots, P_w)$ is divisible by p^{w-k}.

We shall use the notation $\mathbf{P}$ as shorthand for a w-tuple $(P_1, \ldots, P_w)$ of points in S. Consider the product

$$\Delta = \prod_{\mathbf{P}}^{*} |W(\mathbf{P})|, \tag{3.1}$$

where $\prod^*$ denotes a product taken over all tuples $\mathbf{P}$ with $W(\mathbf{P}) \neq 0$. We shall henceforth refer to such $\mathbf{P}$ as *admissible*.

For all $\mathbf{P} \in S^w$, one has $|W(\mathbf{P})| \ll_{\mathscr{W}} N^{d_{\mathscr{W}}}$, so that

$$\frac{\log \Delta}{|S|^w} \leq d_{\mathscr{W}} \log(N) + O_{\mathscr{W}}(1). \tag{3.2}$$

We will now bound Δ from below by a product of local terms. It will then be easy to show that the only way for the upper and lower bounds for Δ to be compatible is either for $|S|$ to be small, or for there to be "relatively few" admissible tuples. The latter possibility will force a large fraction of S to lie on a $\mathscr{W}$-curve.

We assume in what follows that the first possibility of the Proposition does not occur, i.e., any $\mathscr{W}$-curve contains at most $\delta.|S|$ points of S.

Fix any prime $p \leq Q$, where Q is a quantity to be set later. For each $x \in (\mathbb{Z}/p\mathbb{Z})^2$, let ρ_x be the fraction of points in S that reduce to x mod p. For each $\mathbf{P}$, let $\kappa(\mathbf{P}) \in \{0, 1, \ldots, w-1\}$ be such that $w - \kappa(\mathbf{P})$ is the number of distinct points among the P_i mod p. We can bound the p-valuation of Δ – henceforth denoted $\operatorname{ord}_p \Delta$ – from below:

$$\operatorname{ord}_p \Delta \geq \sum^{*} \kappa(\mathbf{P}), \tag{3.3}$$

where $\star$ denotes that we sum only over admissible $\mathbf{P}$.

Let us first analyse the sum $\sum \kappa(\mathbf{P})$ taken over all $\mathbf{P} \in S^w$, admissible or not; we shall substract the non-admissible terms later. Here one obtains that

$$\frac{\sum \kappa(\mathbf{P})}{|S|^w} = w - \sum_{x \in (\mathbb{Z}/p\mathbb{Z})^2} (1 - (1 - \rho_x)^w) = \sum_{x \in (\mathbb{Z}/p\mathbb{Z})^2} ((1 - \rho_x)^w + w\rho_x - 1) \tag{3.4}$$

To see this, consider $\mathbf{P}$ as a random variable; let it have the uniform distribution on its $|S|^w$ possible values. Then $|S|^{-w} \sum (w - \kappa(\mathbf{P}))$ is the expected value of the number of distinct points among the P_i mod p. This number equals the sum of the variables Y_x, where $Y_x = 1$ if at least one of the P_i is congruent to x mod p, and $Y_x = 0$ if none is. The expected value $\mathbb{E}(\sum_x Y_x)$ of $\sum_x Y_x$ equals $\sum_x \mathbb{E}(Y_x)$, and $\mathbb{E}(Y_x)$ is simply the probability that at least one of the P_i be congruent to x mod p. Since the probability that none of the P_i be congruent to x mod p equals $\prod_i \operatorname{Prob}(P_i \not\equiv x \mod p) = \prod_i (1 - \operatorname{Prob}(P_i \equiv x \mod p)) = \prod_i (1 - \rho_x) = (1 - \rho_x)^w$, we are done proving (3.4).

We must now estimate the sum of $\kappa(\mathbf{P})$ over all non-admissible $\mathbf{P}$. Consider the set of all non-admissible $\mathbf{P}$ with $\kappa(\mathbf{P}) > 0$. For such a $\mathbf{P}$, at least one of the following must occur:

(a) There is (i, j) such that $P_i = P_j$;

(b) There is (i,j) such that $P_i \equiv P_j \bmod p$, but $P_i \neq P_j$.

The number of $\mathbf{P}$ that satisfy the first condition is at most $O_w(|S|^{w-1})$. To bound the number of inadmissible $\mathbf{P}$ satisfying the second condition, we permute the entries of $\mathbf{P}$ so that $i=1$ and $j=2$, and permute the w_i so that $w_1 = 1, w_2(P_i) \neq w_2(P_j)$. (The former operation will force us to multiply our bound by $w(w-1)/2$, which will be absorbed by the implied constant. The latter operation is possible because the maps $w \in \mathscr{W}$ are assumed to separate points.) The determinant $\det(w_i(P_j))_{1\leq i,j\leq \ell}$ is nonvanishing for $\ell = 2$; choose the maximal ℓ for which it is nonvanishing. Then $P_{\ell+1}$ lies on a $\mathcal{W}$-curve determined by $P_1, P_2, \ldots, P_\ell$. Therefore there are at most $\delta|S|$ possible values for $P_{\ell+1}$. We obtain that the number of inadmissible $\mathbf{P}$s satisfying the second condition is $O_w(\delta|S|^{w-2}\Delta)$, where Δ is the number of pairs $(P,Q) \in S^2$ which reduce to the same point, mod p. We can express Δ in terms of the proportions ρ_x: clearly, $\Delta = |S|^2 \sum_x \rho_x^2$. We conclude that there are, in total, at most $|S|^w \cdot O_w(|S|^{-1} + \delta \sum_x \rho_x^2)$ inadmissible $\mathbf{P}$ with $\kappa(\mathbf{P}) > 0$.

By (3.3) and (3.4), we finally obtain

$$\frac{\operatorname{ord}_p \Delta}{|S|^w} \geq \left(\sum_x ((1-\rho_x)^w + w\rho_x - 1) - O_w(\delta \sum_x \rho_x^2 + |S|^{-1})\right) \tag{3.5}$$

We will now give a lower bound for the right side of (3.5), using the fact that at most αp congruence classes are occupied by S modulo p (for $p > c$).

There are two cases to be considered.

Case 1: For all $x \in (\mathbb{Z}/p\mathbb{Z})^2$, $\rho_x < \frac{\delta}{w}$. Then, for every x, we know that $(1-\rho_x)^w + w\rho_x - 1 \geq (\binom{w}{2} - O_w(\delta))\rho_x^2$. By Cauchy's inequality, $\sum_x \rho_x^2 \geq \frac{1}{\alpha p}(\sum \rho_x)^2 = \frac{1}{\alpha p}$. Thus

$$\frac{\operatorname{ord}_p \Delta}{|S|^w} \geq \left(\binom{w}{2} - O_w(\delta)\right)\frac{1}{\alpha p} + O_w(|S|^{-1}). \tag{3.6}$$

Case 2: There is an $x \in (\mathbb{Z}/p\mathbb{Z})^2$ such that $\rho_x \geq \frac{\delta}{w}$. Since $\frac{\partial}{\partial z}((1-z)^w + wz - 1) = w(1-(1-z)^{w-1}) \geq w(1-(1-z)) = wz$, we know that $(1-\rho_x)^w + w\rho_x - 1 \geq \frac{1}{2}w\rho_x^2 \geq \frac{1}{2}w \cdot \left(\frac{\delta}{w}\right)^2$. Since $(1-\rho'_x)^w + w\rho'_x - 1 \geq 0$ for $x' \neq x$ and $\sum_x \rho_x^2 \leq 1$, we conclude that

$$\frac{\operatorname{ord}_p \Delta}{|S|^w} \geq \frac{\delta^2}{2w} - O_w(\delta + |S|^{-1}). \tag{3.7}$$

For p greater than a constant $c_{w,\delta}$ depending on w and δ, the bound (3.7) implies the bound (3.6), which we shall henceforth use. (The constants implied by O_w in (3.7) and (3.6) need not be the same.)

We now know (3.6) holds in either case. We multiply both sides of (3.6) by $\log p$ and sum over all p with $\max(c, c_{w,\delta}) < p \leq Q$. We obtain

$$\left(\frac{w(w-1)}{2\alpha} + O_{\alpha,w}(\delta)\right)(\log Q - \log c_{c,w,\delta}) + O(Q|S|^{-1}) \leq \frac{\log \Delta}{|S|^w},$$

where $c_{c,w,\delta}$ depends only on c, w and δ. By (3.2), $\frac{\log \Delta}{|S|^w} \leq d_{\mathscr{W}} \log N + O_{\mathscr{W}}(1)$. Set $Q = |S|$. We obtain

$$\left(\frac{w(w-1)}{2\alpha} + O_{\alpha,w}(\delta)\right)(\log|S| - \log c_{c,w,\delta}) \leq d_{\mathscr{W}} \log N + O_{\mathscr{W}}(1)$$

and thus

$$|S| \ll_{c,\delta,w,\mathscr{W}} N^{\frac{2d_{\mathscr{W}}\alpha}{w(w-1)}+O_{\alpha,\mathscr{W}}(\delta)}, \tag{3.8}$$

as desired. (In so far as we use that $(w(w-1)/(2\alpha)+O_{\alpha,w}(\delta))^{-1} - (2\alpha)/(w(w-1))$ is $\ll_{\alpha,w} \delta$, we are assuming that δ is smaller than a constant depending on w and α; we may assume as much by adjusting the constant implied by $O_{\alpha,\mathscr{W}}$ in (3.8) – the bound $|S| \leq N^2$ is trivial.) □

We can now prove our main result.

Proof of Theorem 1.1. We shall apply Proposition 3.1 with $\mathscr{W}$ equal to the set of monomials $x^i y^j$ with $i+j \leq d$, where d is chosen so large that the exponent $\frac{2d_{\mathscr{W}}\alpha}{w(w-1)} < \epsilon/2$. For this value of d and sufficiently small δ, the quantity $O_{\alpha,\mathscr{W}}(\delta)$ that appears in the exponent of N in Proposition 3.1 is also $\leq \epsilon/2$.

Applying Proposition 3.1 with these parameters, we see that either $|S| \ll_{\alpha,\epsilon} N^{\epsilon}$ (and we are done) or there is a $\mathscr{W}$-curve C containing at least $\delta|S|$ points of S. Assume the latter holds. Let S' be the set of points of S not on C. If $|S'| \leq \epsilon|S|$, then $|S \setminus S'| \geq (1-\epsilon)|S|$, and, since $S \setminus S'$ lies on C, we are done. Suppose, then, that $|S'| > \epsilon|S|$. Apply Proposition 3.1 to S' with the same d and $\mathscr{W}$ as before. Either $|S'| \ll_{\alpha,\epsilon} N^{\epsilon}$ (and, by virtue of $|S| < \frac{1}{\epsilon}|S'| \ll_{\alpha,\epsilon} N^{\epsilon}$, we are done) or there is a $\mathscr{W}$-curve C containing at least $\delta|S'|$ points of S'. Recur as before until a set $S^{(j)}$ with either $|S^{(j)}| \ll_{\alpha,\epsilon} N^{\epsilon}$ or $|S^{(j)}| \leq \epsilon|S|$ is attained. Since $|S'| \leq (1-\delta)|S|$, $|S''| \leq (1-\delta)|S'|$, etc., we see that $j \leq \frac{\log \epsilon}{\log(1-\delta)}$.

If $|S^{(j)}| \leq \epsilon|S|$, write

$$S \setminus S^{(j)} = (S \setminus S') \cup (S' \setminus S^{(2)}) \cup \cdots \cup (S^{(j-1)} \setminus S^{j}),$$

and recall that each of the sets $(S^{(k)} \setminus S^{(k+1)})$ on the right lies on a curve of degree $O_{\alpha,\epsilon}(1)$. Let C be the union of all such curves. Then C is itself a curve of degree $O_{\alpha,\epsilon}(1)$, and so we are done.

If $|S^{(j)}|$ is not less than $\epsilon|S|$, we still have $|S^{(j)}| \ll_{\alpha,\epsilon} N^{\epsilon}$, from which we obtain $|S| \ll_{\alpha,\epsilon} N^{\epsilon}$, and are done. □

As was said before, we can reproduce the Bombieri-Pila bounds. Our method of proof is, of course, very closely linked to the original proof of [BP].

Proposition 3.2. *Let C be an irreducible curve of degree d over $\mathbb{Q}$. Let S be the set of points in $\mathbb{Z}^2 \cap [0,N]^2$ on C. Then*

$$|S| \ll_{d,\epsilon} N^{1/d+\epsilon}.$$

Proof. We may assume $|S| > (d+1)^2$, for otherwise we are done. The curve C is defined by the zero-locus $f(x,y) = 0$, where $f \in \mathbb{Z}[x,y]$ is irreducible of degree d. By interpolation, we may assume that the coefficients of f are bounded above by $N^{O_d(1)}$. (This type of argument is used by Heath-Brown in a similar context ([HBR]).) If the degree of f on x is less than d, begin by applying a linear transformation on x and y so as to make the degree of f on x equal to d; otherwise, simply proceed. This linear transformation may be chosen of the form $(x,y) \mapsto (a_1x+a_2y, a_3x+a_4y)$ where $\max(|a_1|,|a_2|,|a_3|,|a_4|) = O_d(1)$; in particular, it suffices to prove the claimed bound for the transformed curve.

If p is any prime, and $\bar{f} \in \mathbb{F}_p[x,y]$ the reduced polynomial, factor f into $\mathbb{F}_p$-irreducible factors $\bar{f} = \bar{f}_1 \dots \bar{f}_{e_p}$. The Weil bounds show that the number of points on each irreducible components is at most $p + O_d(\sqrt{p})$, i.e.:

$$|\{(x,y) \in \mathbb{F}_p^2 : \bar{f}_i(x,y) = 0\}| \leq p + O_d(\sqrt{p}).$$

The set of primes $\mathcal{P}$ for which $\bar{f}$ is reducible satisfies $\prod_{p\in\mathcal{P}} p \leq N^{O_d(1)}$. Indeed, these are precisely the primes that divide a suitable "discriminant", which is a polynomial in the coefficients of f. Partition the points in S according to which irreducible component of $\bar{f} = 0$ they reduce to modulo each $p \in \mathcal{P}$. The number of irreducible factors of $\bar{f}$ for each prime p is at most d. Therefore, S is covered by sets $S_1, \dots, S_k$, where $k \leq d^{|\mathcal{P}|} \ll_{d,\epsilon} N^{\epsilon}$, and such that each set S_k intersects at most $p + O_d(\sqrt{p})$ residue classes modulo every prime p. In particular, for any $\epsilon > 0$ and $p \geq O_\epsilon(1)$, the set S_k intersects at most $(1+\epsilon/2)p$ residue classes mod p.

It will suffice to prove the conclusion of the Proposition with S replaced by any S_j (for $1 \leq j \leq k$). We make this replacement and proceed.

Apply Proposition 3.1 to S with $\alpha = (1+\epsilon/2)$, with $d-1$ instead of d, and with $\mathscr{W}$ chosen as in the example (b) listed before the statement of the Proposition. Since C is irreducible and of degree d on x, and all $\mathscr{W}$-curves have degree $d-1$ or less on x, the intersection of S with any $\mathscr{W}$-curve has no more than $d(d-1)$ points. Thus, option (a) in the conclusion of Proposition 3.1 would imply that $|S| < \delta^{-1} d(d-1)$. Assume that option (b) holds. Then

$$|S| \ll_{\epsilon,d,\delta,M} N^{\frac{(1+\epsilon/2)(d-1+M)}{d(M+1)-1} + O_{\epsilon,d,M}(\delta)}.$$

We set M to a sufficiently large value, and obtain

$$|S| \ll_{\epsilon,d,\delta} N^{\frac{1}{d} + 3\epsilon/4 + O_{d,\epsilon}(\delta)}.$$

We let δ be small enough for $O_{d,\epsilon}(\delta)$ to be less than $\epsilon/4$, and are done. (The bound $|S| < \delta^{-1} d(d-1)$ in option (a) becomes $|S| \ll_{d,\epsilon} 1$.) □

4. Final remarks

We return to the setting discussed in the introduction: $S \subset \mathbb{Z}^n$ is ill-distributed modulo p, for many p.

4.1. Subsets of $\mathbb{Z}^n$ for $n \geq 3$. The case $S \subset \mathbb{Z}^n \cap [0,N]^n$, $n > 2$, may seem no harder than the case of $n = 2$, yet it does not seem simple to produce a result of strength comparable to that of Theorem 1.1. One can, in fact, derive similar conclusions from the same assumption $|S \bmod p| \leq \alpha p$ as before; however, one would expect these same conclusions to follow from $|S \bmod p| \leq \alpha p^{k-1}$, and, while we believe this to be the case, it is hard to see how one might be able to prove as much by our methods.

4.2. Subsets of $\mathbb{Z}$: a guess. Consider now the case $S \subset \mathbb{Z} \cap [0,N]$. We now return to the speculation made in the Introduction: need such a set be either "strongly algebraic" or "very sparse"? As we are not ready to make a conjecture, let us say we are simply guessing.

Guess. *Let $S \subset \mathbb{Z} \cap [0, N]$, $N \geq 1$, $\epsilon > 0$, $\alpha \in (0,1)$. Suppose that the number of residues $\{x \bmod p : x \in S\}$ is at most αp for every prime p.*

Then, for any $\epsilon > 0$, either

(a) *$|S| \ll_{\alpha,\epsilon} N^\epsilon$ or*

(b) *there is a plane curve $f(x,y) = 0$ such that at least $(1-\epsilon)|S|$ of the points of S lie on the projection of the solution set $\{(x,y) \in \mathbb{Z}^2 : f(x,y) = 0\}$ onto the x axis.*

Moreover, f may be chosen so that the curve defined by $f(x,y) = 0$ does not contain any lines, so that the degree of f is $O_{\alpha,\epsilon}(1)$, and so that all the coefficients of f are integers bounded by $N^{O_{\alpha,\epsilon}(1)}$ in absolute value.

One could posit, more ambitiously, that, unless case (a) holds, S is largely contained in the set of values $f(n)$, $n \in \mathbb{Z}$, of a polynomial map $f : \mathbb{Z} \to \mathbb{Z}$ of degree d. We have the same situation as before: the weaker statement (namely, the guess stated above) together with the standard conjecture on the small number ($\ll_\epsilon N^\epsilon$) of points on an irrational curve would imply the stronger statement (namely, what we have just posited); at the same time, the stronger statement implies the standard conjecture, which is quite hard.

One might contrarily think that our guess is too ambitious, in that S should be allowed to resemble the projection of a surface in n-dimensional space (where $n \ll_{k,\alpha} 1$) to one of its coordinates. It seems to us that the statement might then be too weak to be interesting: sets that are not algebraic in any intuitive sense can be construed as the projections of surfaces in n-dimensional space, where n is large but fixed (see [Ma]). A similar rationale lies behind our specification that all coefficients be of size at most $N^{O_{\alpha,\epsilon}(1)}$.

It is tempting to venture that the bound of $\ll_{\alpha,\epsilon} N^\epsilon$ in our guess (or in Thm. 1.1) is not best, but it is difficult to see what would be best. One can construct a set S obeying the assumptions of Conj. 1, yet lacking any visible algebraic structure; the number of elements $|S|$ of this S is $\sim c \log N$ (see Sec. 4.3 below). Is there a similar example with $|S| \gg (\log N)^2$, say? We do not know.

4.3. **Ill-distributed sets of size** $\log N$**.** It is easy to construct a set $S \subset [1, N]$, of logarithmic size, without any visible algebraic structure, such that S is contained in few residue classes modulo each prime p. Indeed, choose a prime $Q \sim \log(N)$ such that the product of all primes $< Q$ is at most N. Let $R = \prod_{p<Q} p$. By the Chinese remainder theorem, the set of integers in $[1, R]$ that reduce to ± 1 modulo each $p < Q$ has size $2^{\pi(Q)}$. Take S to be any subset of this set of cardinality less than $Q/2$. Then S intersects at most $p/2$ residue classes for each prime p: for $p < Q$, we have a much stronger bound from the construction, and for $p \geq Q$, we have $|S| < Q/2 \leq p/2$. It would be interesting to know whether one can construct a set S like this one but of size $(\log N)^2$ or greater.

4.4. **Pseudo-polynomials.** The results of this paper are also related to *pseudopolynomials.* (See [Ha].) A pseudo-polynomial is a function $f : \mathbb{N} \to \mathbb{Z}$ with the property that $f(x) \equiv f(y) \bmod k$ whenever $x \equiv y \bmod k$, for any integer $k \geq 1$. For such a function, the graph of f, i.e. $\{(x, f(x)) : x \in \mathbb{N}\}$ intersects at most p residue classes modulo each prime p, and therefore Theorem 1.1 shows that this graph must have strong algebraic structure, away from a small set. In this case, however, even stronger results have been known for a long time – see [Ha, Thm. 3] for a proof that a pseudopolynomial that is not a polynomial must grow very rapidly.

References

[Bo] Bombieri, E., Le grand crible dans la théorie analytique des nombres, 2nd. ed., *Astérisque* **18** (1987).

[BP] Bombieri, E., and J. Pila, The number of integral points on arcs and ovals, *Duke Math. J.* **59** (1989), 337–357.

[Ga] Gallagher, P. X., A larger sieve, *Acta Arith.* **18** (1971), 77–81.

[Ha] Hall, R. R., On pseudopolynomials, *Mathematika* **18** (1971), 71–77.

[HBR] Heath-Brown, D. R., The density of rational points on curves and surfaces, *Ann. of Math.* **155** (2002), 553–595.

[Ma] Matijasevič, Ju. V., The Diophantineness of enumerable sets, *Dokl. Akad. Nauk. SSSR* **191** (1970), 279–282; English translation in *Soviet Math. Dokl.* **11** (1970), 354–358.

[Pi] Pila, J., Density of integer points on plane algebraic curves, *IMRN* **18** (1996).

[Sa] Salberger, P., Counting rational points on hypersurfaces of low dimension, *Ann. Scient. Éc. Norm. Sup.* **38** (2005), 93–115.

H. A. Helfgott, Département de Mathématiques et Statistique, Université de Montréal, Montréal, QC H3C 3J7, Canada

E-mail address: `helfgott@dms.umontreal.ca`

A. Venkatesh, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA.

E-mail address: `venkatesh@cims.nyu.edu`