

ARTICLE

A generalisation of Varnavides’s theorem

Asaf Shapira

School of Mathematics, Tel Aviv University, Tel Aviv, Israel
Email: asafico@tau.ac.il

(Received 22 January 2024; revised 10 April 2024; accepted 6 May 2024; first published online 29 May 2024)

Abstract

A linear equation E is said to be *sparse* if there is $c > 0$ so that every subset of $[n]$ of size n^{1-c} contains a solution of E in distinct integers. The problem of characterising the sparse equations, first raised by Ruzsa in the 90s, is one of the most important open problems in additive combinatorics. We say that E in k variables is *abundant* if every subset of $[n]$ of size εn contains at least $\text{poly}(\varepsilon) \cdot n^{k-1}$ solutions of E . It is clear that every abundant E is sparse, and Girão, Hurley, Illingworth, and Michel asked if the converse implication also holds. In this note, we show that this is the case for every E in four variables. We further discuss a generalisation of this problem which applies to all linear equations.

Keywords: Varnavides; Ruzsa; Roth; supersaturation

2020 MSC Codes: Primary: 05D99; Secondary: 11B75

1. Introduction

Turán-type questions are some of the most well-studied problems in combinatorics. They typically ask how ‘dense’ should an object be in order to guarantee that it contains a certain small substructure. In the setting of graphs, this question asks how many edges an n -vertex graph should contain in order to force the appearance of some fixed graph H . For example, a central open problem in this area asks, given a bipartite graph H , to determine the smallest $T = T_H(\varepsilon)$ so that for every $n \geq T$ every n -vertex graph with $\varepsilon \binom{n}{2}$ edges contains a copy of H (see [3] for recent progress). A closely related question which also attracted a lot of attention is the *supersaturation* problem, introduced by Erdős and Simonovits [5] in the 80s. In the setting of Turán’s problem for bipartite H , the supersaturation question asks to determine the largest $T_H^*(\varepsilon)$ so that every n -vertex graph with $\varepsilon \binom{n}{2}$ edges contains at least $(T_H^*(\varepsilon) - o_n(1)) \cdot n^h$ labelled copies of H , where $h = |V(H)|$ and $o_n(1)$ denotes a quantity tending to 0 as $n \rightarrow \infty$. One of the central conjectures in this area, due to Sidorenko, suggests that $T_H^*(\varepsilon) = \varepsilon^m$, where $m = |E(H)|$ (see [4] for recent progress).

We now describe two problems in additive number theory, which are analogous to the graph problems described above. We say that a homogenous linear equation $\sum_{i=1}^k a_i x_i = 0$ is *invariant* if $\sum_i a_i = 0$. All equations we consider here will be invariant and homogenous. Given a fixed linear equation E , the Turán problem for E asks to determine the smallest $R = R_E(\varepsilon)$ so that for every $n \geq R$, every $S \subseteq [n] := \{1, \dots, n\}$ of size εn contains a solution to E in distinct integers. For example, when E is the equation $a + b = 2c$ we get the Erdős–Turán–Roth problem on sets avoiding 3-term arithmetic progressions (see [7] for recent progress). Continuing the analogy with the previous paragraph, we can now ask to determine the largest $R_E^*(\varepsilon)$ so that every $S \subseteq [n]$ of size εn contains at least $(R_E^*(\varepsilon) - o_n(1)) \cdot n^{k-1}$ solutions to E , where k is the number of variables in E .

Supported in part by ERC Consolidator Grant 863438 and NSF-BSF Grant 20196.

© The Author(s), 2024. Published by Cambridge University Press.



We now turn to discuss two aspects which make the arithmetic problems more challenging than the graph problems.

Let us say that E is *sparse* if there is $C = C(H)$ so that¹ $R_E(\varepsilon) \leq \varepsilon^{-C}$. The first aspect which makes the arithmetic landscape more varied is that while in the case of graphs it is well known (and easy) that for every bipartite H we have $T_H(\varepsilon) = \text{poly}(1/\varepsilon)$, this is no longer the case in the arithmetic setting. Indeed, while Sidon's equation $a + b = c + d$ is sparse, a well-known construction of Behrend [1] shows that $a + b = 2c$ is not sparse.² The problem of determining which equations E are sparse is a wide open problem due to Ruzsa, see Section 9 in [9].

Our main goal in this paper is to study another aspect which differentiates the arithmetic and graph-theoretic problems. While it is easy to translate a bound for $T_H(\varepsilon)$ into a bound for $T_H^*(\varepsilon)$ (in particular, establishing that $T_H^*(\varepsilon) \geq \text{poly}(\varepsilon)$ for all bipartite H), it is not clear if one can analogously transform a bound for $R_E(\varepsilon)$ into a bound for $R_E^*(\varepsilon)$. The first reason is that while we can average over all subsets of vertices of graphs, we can only average over "structured" subsets of $[n]$. This makes it hard to establish a black-box reduction/transformation between $R_E(\varepsilon)$ and $R_E^*(\varepsilon)$. The second complication is that, as we mentioned above, we do not know which equations are sparse. This makes it hard to directly relate these two quantities. Following [6], we say that E is *abundant* if $R_E^*(\varepsilon) \geq \varepsilon^C$ for some $C = C(E)$. Clearly, if E is abundant then it is also sparse. Girão et al. [6] asked if the converse also holds, that is, if one can transform a polynomial bound for $R_E(\varepsilon)$ into a polynomial bound for $R_E^*(\varepsilon)$. Our aim in this note is to prove the following.

Theorem 1.1. *If an invariant equation E in four variables is sparse, then it is also abundant. More precisely, if $R_E(\varepsilon) \leq \varepsilon^{-C}$ then $R_E^*(\varepsilon) \geq \frac{1}{2}\varepsilon^{8C}$ for all small enough ε .*

Given the above discussion, it is natural to extend the problem raised in [6] to all equations E .

Problem 1.2. *Is it true that for every invariant equation E there is $c = c(E)$, so that for all small enough ε*

$$R_E^*(\varepsilon) \geq 1/R_E(\varepsilon^c).$$

It is interesting to note that Varnavides [11] (implicitly) gave a positive answer to Problem 1.2 when E is the equation $a + b = 2c$. In fact, essentially the same argument gives a positive answer to this problem for all E in three variables. Hence, Problem 1.2 can be considered as a generalisation of Varnavides's theorem. Problem 1.2 was also implicitly studied previously in [2,8]. In particular, Kosciuszko [8], extending earlier work of Schoen and Sisask [10], gave direct lower bounds for R_E^* which, thanks to [7], are quasi-polynomially related to those of R_E .

The proof of Theorem 1.1 is given in the next section. For the sake of completeness, and as a preparation for the proof of Theorem 1.1, we start the next section with a proof that Problem 1.2 holds for equations in three variables. We should point that a somewhat unusual aspect of the proof of Theorem 1.1 is that it uses a Behrend-type [1] geometric argument in order to find solutions, rather than avoid them.

2. Proofs

In the first subsection below, we give a concise proof of Varnavides's theorem, namely, of the fact that Problem 1.2 has a positive answer for equations with three variables. In the second subsection, we prove Theorem 1.1.

¹It is easy to see that this definition is equivalent to the one we used in the abstract.

²More precisely, it shows that in this case $R_E(\varepsilon) \geq (1/\varepsilon)^{c \log 1/\varepsilon}$. Here and throughout this note, all logarithms are base 2.

2.1 Proof of Varnavides’s theorem

Note that for every equation E , there is a constant C such that for every prime $p \geq Cn$ every solution of E with integers $x_i \in [n]$ over \mathbb{F}_p is also a solution over \mathbb{R} . Since we can always find a prime $Cn \leq p \leq 2Cn$, this means that we can assume that n itself is prime³ and count solutions over \mathbb{F}_n . So let S be a subset of \mathbb{F}_n of size εn and let $R = R_E(\varepsilon/2)$. For $b = (b_0, b_1) \in (\mathbb{F}_n)^2$ and $x \in [R]$ let $f_b(x) = b_1x + b_0$ and⁴ $f_b([R]) = \{x \in [R] : f_b(x) \in S\}$. Pick b_0 and b_1 uniformly at random from \mathbb{F}_n and note that for any $x \in [R]$ the integer $f_b(x)$ is uniformly distributed in \mathbb{F}_n . Hence,

$$\mathbb{E}|f_b([R])| = \varepsilon R.$$

It is also easy to see that for every $x \neq y$ the random variables $f_b(x)$ and $f_b(y)$ are pairwise independent. Hence,

$$\text{Var}|f_b([R])| \leq \varepsilon R.$$

Therefore, by Chebyshev’s Inequality we have

$$\mathbb{P}\left[|f_b([R])| \leq \frac{\varepsilon}{2}R\right] \leq \frac{\varepsilon R}{\varepsilon^2 R^2/4} \leq 1/2.$$

In other words, at least $n^2/2$ choices of b are such that $|f_b([R])| \geq \frac{\varepsilon}{2}R$. By our choice of R , this means that $f_b([R])$ contains three distinct integers x_1, x_2, x_3 which satisfy E and such that $f_b(x_i) \in S$. Note that if x_1, x_2, x_3 satisfy E , then so do $f_b(x_1), f_b(x_2), f_b(x_3)$. Let us denote the triple $(f_b(x_1), f_b(x_2), f_b(x_3))$ by s_b . We have thus obtained $n^2/2$ solutions s_b of E in S . To conclude the proof, we just need to estimate the number of times we have double counted each solution s_b . Observe that for every choice of $s_b = \{s_1, s_2, s_3\}$ and *distinct* $x_1, x_2, x_3 \in [R]$, there is exactly one choice of $b = (b_0, b_1) \in (\mathbb{F}_n)^2$ for which $b_1x_i + b_0 = s_i$ for every $1 \leq i \leq 3$. Since $[R]$ contains at most R^2 solutions of E this means that for every solution $s_1, s_2, s_3 \in S$, there are at most R^2 choices of b for which $s_b = \{s_1, s_2, s_3\}$. We conclude that S contains at least $n^2/2R^2$ distinct solutions, thus completing the proof.

2.2 Proof of Theorem 1.1

As in the proof above, we assume that n is a prime and count the number of solutions of the equation $E: \sum_{i=1}^d a_i x_i = 0$ over \mathbb{F}_n . Let S be a subset of \mathbb{F}_n of size εn , and let d and t be integers to be chosen later and let X be some subset of $[t]^d$ to be chosen later as well. For every $b = (b_0, \dots, b_d) \in (\mathbb{F}_n)^{d+1}$ and $x = (x_1, \dots, x_d) \in X$, we use $f_b(x)$ to denote $b_0 + \sum_{i=1}^d b_i x_i$ and $f_b(X) = \{x \in X : f_b(x) \in S\}$. We call b *good* if $|f_b(X)| \geq \varepsilon|X|/2$. We claim that at least half of all possible choices of b are good. To see this, pick $b = (b_0, \dots, b_d)$ uniformly at random from $(\mathbb{F}_n)^{d+1}$, and note that for any $x \in X$ the integer $f_b(x)$ is uniformly distributed in \mathbb{F}_n . Hence,

$$\mathbb{E}|f_b(X)| = \varepsilon|X|.$$

It is also easy to see that for every $x \neq y \in X$ the random variables $f_b(x)$ and $f_b(y)$ are pairwise independent. Hence,

$$\text{Var}|f_b(X)| \leq \varepsilon|X|.$$

Therefore, by Chebyshev’s Inequality we have⁵

$$\mathbb{P}\left[|f_b(X)| \leq \frac{\varepsilon}{2}|X|\right] \leq 4/\varepsilon|X| \leq 1/2, \tag{1}$$

³The factor $2C$ loss in the density of S can be absorbed by the factor c in Problem 1.2.

⁴Since $f([R])$ is a subset of $[R]$ (rather than S), it might have been more accurate to denote $f([R])$ by $f^{-1}([R])$ but we drop the -1 to make the notation simpler.

⁵We will make sure $|X| \geq 8/\varepsilon$.

implying that at least half of the b 's are good. To finish the proof, we need to make sure that every such choice of a good b will “define” a solution s_b in S in a way that s_b will not be identical to too many other $s_{b'}$. This will be achieved by a careful choice of d, t , and X .

We first choose X to be the largest subset of $[t]^d$ containing no three points on one line. We claim that

$$|X| \geq t^{d-2}/d. \tag{2}$$

Indeed, for an integer r let B_r be the points $x \in [t]^d$ satisfying $\sum_{i=1}^d x_i^2 = r$. Then every point of $[t]^d$ lies on one such B_r , where $1 \leq r \leq dt^2$. Hence, at least one such B_r contains at least t^{d-2}/d of the points of $[t]^d$. Furthermore, since each set B_r is a subset of a sphere, it does not contain three points on one line.

We now turn to choose t and d . Let C be such that $R_E(\varepsilon) \leq (1/\varepsilon)^C$. Set $a = \sum_{i=1}^4 |a_i|$ and pick t and d satisfying

$$(1/\varepsilon)^{2C} \geq t^d \geq \left(\frac{2dt^2 a^d}{\varepsilon}\right)^C. \tag{3}$$

Taking $t = 2\sqrt{\log 1/\varepsilon}$ and $d = 2C\sqrt{\log 1/\varepsilon}$ satisfies⁶ the above for all small enough ε . Note that by (3) and our choice of C , we have $R_E\left(\frac{\varepsilon}{2dt^2 a^d}\right) \leq t^d$.

Let us call a collection of four vectors $x^1, x^2, x^3, x^4 \in X$ *helpful* if they are distinct, and they satisfy E in each coordinate, that is, for every $1 \leq i \leq d$ we have $\sum_{j=1}^4 a_j x_i^j = 0$. We claim that for every good r , there are useful $x^1, x^2, x^3, x^4 \in f_r(X)$. To see this let M denote the integers $1, \dots, (at)^d$ and note that (2) along with the fact that r is good implies that

$$|f_r(X)| \geq \varepsilon|X|/2 \geq \frac{\varepsilon t^d}{2t^2 d} = \frac{\varepsilon}{2dt^2 a^d} \cdot |M| \tag{4}$$

Now think of every d -tuple $x \in X$ as representing an integer $p(x) \in [M]$ written in base at . So we can also think of $f_r(X)$ as a subset of $[M]$ of density at least $\varepsilon/2dt^2 a^d$. By (3), we have

$$M = (at)^d \geq t^d \geq R_E\left(\frac{\varepsilon}{2dt^2 a^d}\right),$$

implying that there are *distinct* $x^1, x^2, x^3, x^4 \in f_r(X)$ for which $\sum_{j=1}^4 a_j \cdot p(x^j) = 0$. But note that since the entries of x^1, x^2, x^3, x^4 are from $[t]$, there is no carry when evaluating $\sum_{j=1}^4 a_j \cdot p(x^j)$ in base at , implying that x^1, x^2, x^3, x^4 satisfy E in each coordinate. Finally, the fact that $\sum_j a_j = 0$ and that $x_i^1, x_i^2, x_i^3, x_i^4$ satisfy E for each $1 \leq i \leq d$ allows us to deduce that

$$\sum_{j=1}^4 a_j \cdot f_b(x^j) = \sum_{j=1}^4 a_j \cdot \left(b_0 + \sum_{i=1}^d b_i x_i^j\right) = \sum_{i=1}^d b_i \cdot \left(\sum_{j=1}^4 a_j x_i^j\right) = 0,$$

which means that $f_b(x^1), f_b(x^2), f_b(x^3), f_b(x^4)$ forms a solution of E . So for every good b , let s_b be (some choice of) $f_b(x^1), f_b(x^2), f_b(x^3), f_b(x^4) \in S$ as defined above. We know from (1) that at least $n^{d+1}/2$ of all choices of b are good, so we have thus obtained $n^{d+1}/2$ solutions s_b of E in S . To finish the proof, we need to bound the number of times we have counted the same solution in S , that is, the number of b for which s_b can equal a certain 4-tuple in S satisfying E .

⁶Recalling (2), we see that since $C \geq 1$ (indeed, a standard probabilistic deletion method argument shows that if an equation has k variables, then $C(E) \geq 1 + \frac{1}{k-2}$), we indeed have $|X| \geq 8/\varepsilon$ as we promised earlier.

Fix $s = \{s_1, s_2, s_3, s_4\}$ and recall that $s_b = s$ only if there is a helpful 4-tuple x^1, x^2, x^3, x^4 (as defined just before equation (4)) such that $f_b(x^i) = s_i$. We claim that for every helpful 4-tuple x^1, x^2, x^3, x^4 , there are at most n^{d-2} choices of $b = (b_0, \dots, b_d)$ for which $s_b = s$. Indeed recall that by our choice of X the vectors x^1, x^2, x^3 are distinct and do not lie on one line. Hence, they are affine independent⁷ over \mathbb{R} . But since the entries of x^i belong to $[t]$ and $t \leq 1/\varepsilon$, we see that for large enough n the vectors x^1, x^2, x^3 are also affine independent over \mathbb{F}_n . This means that the system of three linear equations:

$$\begin{aligned} b_0 + b_1 x_1^1 + \dots + b_d x_d^1 &= s_1 \\ b_0 + b_1 x_1^2 + \dots + b_d x_d^2 &= s_2 \\ b_0 + b_1 x_1^3 + \dots + b_d x_d^3 &= s_3 \end{aligned}$$

(in $d + 1$ unknowns b_0, \dots, b_d over \mathbb{F}_n) has only n^{d-2} solutions, implying the desired bound on the number of choices of b . Since $|X| \leq t^d \leq (1/\varepsilon)^{2C}$ by (3), we see that X contains at most $(1/\varepsilon)^{8C}$ helpful 4-tuples. Altogether this means that for every $s_1, s_2, s_3, s_4 \in S$ satisfying E , there are at most $(1/\varepsilon)^{8C} n^{d-2}$ choices of b for which $s_b = s$. Since we have previously deduced that S contains at least $\frac{1}{2} n^{d+1}$ solutions s_b , we get that S contains at least $\frac{1}{2} \varepsilon^{8C} n^3$ distinct solutions, as needed.

Acknowledgement

I would like to thank Yuval Wigderson and an anonymous referee for helpful comments and suggestions.

References

- [1] Behrend, F. A. (1946) On sets of integers which contain no three terms in arithmetical progression. *Proc Nat. Acad. Sci. USA* **32**(12) 331–332.
- [2] Bloom, T. F. (2012) Translation invariant equations and the method of Sanders. *Bull. Lond. Math. Soc.* **44**(5) 1050–1067.
- [3] Bukh, B. (to appear) Extremal graphs without exponentially-small bicliques. *Duke Math. J.*
- [4] Conlon, D., Kim, J. H., Lee, C. and Lee, J. (2018) Some advances on Sidorenko’s conjecture. *J. London Math. Soc.* **98**(3) 593–608.
- [5] Erdős, P. and Simonovits, M. (1983) Supersaturated graphs and hypergraphs. *Combinatorica* **3**(2) 181–192.
- [6] Girão, A., Hurley, E., Illingworth, F. and Michel, L. Abundance: asymmetric graph removal lemmas and integer solutions to linear equations. arXiv: 2310.18202.
- [7] Kelley, Z. and Meka, R. (2023) Strong bounds for 3-progressions. arXiv: 2302.05537.
- [8] Kosciuszko, T. Counting solutions to invariant equations in dense sets. arXiv: 2306.08567.
- [9] Ruzsa, I. Z. (1993) Solving a linear equation in a set of integers I. *Acta Arith.* **65**(3) 259–282.
- [10] Schoen, T. and Sisask, O. (2016) Roth’s theorem for four variables and additive structures in sums of sparse sets. *Forum Math. Sigma* **4** e5,28.
- [11] Varnavides, P. (1959) On certain sets of positive density. *J. London Math. Soc.* **34**(3) 358–360.

⁷That is, if we turn these three d -dimensional vectors into $(d + 1)$ -dimensional vectors, by adding a new coordinate whose value is 1, we get three linearly independent vectors.