



Arithmetic progressions with a pseudorandom step

Elad Aigner-Horev¹

*Mathematics and Computer science Department
Ariel University
Ariel, Israel*

Hiệp Hàn²

*Mathematics Department
Universidad de Chile
Santiago, Chile*

Abstract

Let $\alpha, \sigma > 0$ and let A and S be subsets of a finite abelian group G of densities α and σ , respectively, both independent of $|G|$. Without any additional restrictions, the set A need not contain a 3-term arithmetic progression whose common gap lies in S . What is then the weakest pseudorandomness assumption that if put on S would imply that A contains such a pattern?

More precisely, what is the least integer $k \geq 2$ for which there exists an $\eta = \eta(\alpha, \sigma)$ such that $\|S - \sigma\|_{U^k(G)} \leq \eta$ implies that A contains a non-trivial 3-term arithmetic progression with a common gap in S ? Here, $\|\cdot\|_{U^k(G)}$ denotes the k th Gowers norm.

For $G = \mathbb{Z}_n$ we observe that k must be at least 3. However for $G = \mathbb{F}_p^n$ we show that $k = 2$ is sufficient, where here p is an odd prime and n is sufficiently large.

Keywords: Pseudorandom sets, regularity method, arithmetic progressions.

1 Introduction

Given a set $A \subseteq [n] = \{1, \dots, n\}$ with positive density (dense, hereafter), an additional set $S \subseteq [n]$, and an integer $k \geq 2$ we may enquire whether A contains a k -term arithmetic progression (k AP, hereafter) whose common gap lies in S (k SAP, hereafter). The celebrated Szemerédi's Theorem [13] addresses the case $S = [n]$. Using ergodic methods Bergelson and Leibman [1] proved a far reaching generalisation of Szemerédi's Theorem which has come to be known as the *polynomial Szemerédi's Theorem* where the following set S is considered.

Theorem 1.1 (Polynomial Szemerédi's Theorem [1])

For every $\alpha > 0$ there exists an N_0 such that for all $N \geq N_0$ the following holds. Let $A \subseteq [n]$ have density α and let P_1, \dots, P_k be polynomials with integer coefficients all vanishing at zero. Then there exists a $d \neq 0$ such that A contains the configuration $x + P_1(d), \dots, x + P_k(d)$.

Currently the sole known proof of Theorem 1.1 is the ergodic proof of Bergelson and Leibman. Hence, the qualitative formulation of this result. Without using ergodic methods Green [8] established the following.

Theorem 1.2 (Green [8]) *There exists a constant c such that any subset of $[n]$ of density at least $(\log \log n)^{-c}$ contains the configuration $\{x, x + d_1^2 + d_2^2, x + 2d_1^2 + 2d_2^2\}$ for some integers d_1 and d_2 not both zero.*

To date this result of Green is the sole known non-ergodic proof in the direction of Theorem 1.1.

More recently using ergodic methods, Christ [2] and separately of him Frantzikinakis and Lesigne [3,4], considered the emergence of 3SAPs with S being a random set. Roughly speaking, they establish the following. Let $A \subseteq [n]$ be dense (n sufficiently large) and let $S \subseteq [n]$ be random and of density $\gg n^{-1/2}$. Then, with high probability, A contains a 3SAP.

In this paper, without using ergodic methods, we consider the emergence of 3SAPs in a dense set A in the case that S is dense and pseudorandom to some extent. Requiring that S is dense is clearly insufficient on its own³. Moreover, requiring that S forms a relatively dense subset⁴ of a random set

¹ Email: horev@ariel.ac.il

² Email: han.hiep@googlemail.com

³ Consider: $A = (\frac{2}{3}n, n]$ and $S = (\frac{1}{3}n, \frac{2}{3}n]$.

⁴ We say that a set X is a *relatively dense subset* of a set Y if there exists a $\delta > 0$ independent of $|Y|$ such that $|X| \geq \delta|Y|$.

is insufficient as well⁵. In the infinite setting of this problem (i.e., $A, S \subseteq \mathbb{Z}$), already for the emergence of 2SAPs in A having $S \cap q\mathbb{Z} \neq \emptyset$, for every $q \in \mathbb{Z}$, is a necessary condition.

This together with the observation that the set of allowed gaps S in Theorem 1.2 is dense in $[n]$ (see, e.g., [14, Corollary 4.15] and comments thereafter) suggest that considering the emergence of 3SAPs in a dense set A with S being dense and *adequately pseudorandom* is a natural venue for this problem. Our interest is in quantifying the phrase *adequately pseudorandom*.

Problem 1.3 *Let $\alpha, \sigma > 0$ and let A and S be subsets of a finite abelian group G of densities α and σ , respectively, both independent of $|G|$. What is the least integer $k \geq 2$ for which there exists an $\eta = \eta(\alpha, \sigma)$ such that $\|S - \sigma\|_{U^k} \leq \eta$ implies that A contains a non-trivial 3SAP?*

Here, $\|\cdot\|_{U^k}$ denotes the k th Gowers norm [14]. Using the k th Gowers norm of the *balanced function* of S , i.e., $S - \sigma$ in $L^1(G)$, in order to quantify the pseudorandomness of S follows the traditional definition of pseudorandom sets [14].

With the exception that the characteristic function of a set X is denoted $X(\cdot)$, our notation is that of [14]. Throughout, we write $x = y \pm d$ to denote that $x \in [y - d, y + d]$.

Prior to stating our main result, let us consider Problem 1.3 for sets taken in the group \mathbb{Z}_n . First, let us consider a simpler problem. Given two sets A and S in \mathbb{Z}_n how dense can A be if it contains no 2SAPs? Here a 2SAP consists of two points $x, y \in A$ such that $x - y \in S$. This is a generalisation of the well-known Fürstenberg-Sárközy Theorem [5,12] in which $S = \mathbb{Z}_n$ is taken.

Applying the so called *Hoffman bound* [11] over the size of the largest independent set in a graph to the undirected Cayley graph generated by S leads to the following result; proof of which we omit.

Proposition 1.4 *Let n be a positive integer, let $S \subseteq \mathbb{Z}_n$ be symmetric⁶, and let $A \subseteq \mathbb{Z}_n$. If A contains no 2SAP, then A has density at most $\|S\|_u \|S\|_{L^1(\mathbb{Z}_n)}^{-1}$.*

⁵ Take A to be the even numbers and S to be the intersection of the odd numbers with a dense random set.

⁶ A set $X \subseteq \mathbb{Z}_n$ is called symmetric if $x \in X \iff x^{-1} \in X$.

Here, $\|S\|_u$ denotes the *linear bias*⁷ of S given by

$$\|S\|_u = \sup_{\xi \in \widehat{\mathbb{Z}_n} \setminus \{\widehat{0}\}} |\widehat{S}(\xi)| = \sup_{\xi \in \widehat{\mathbb{Z}_n} \setminus \{\widehat{0}\}} \left| \mathbb{E}_{x \in \mathbb{Z}_n} S(x) \overline{\xi(x)} \right|,$$

where $\widehat{S} : \widehat{\mathbb{Z}_n} \rightarrow \mathbb{C}$ is the Fourier transform of S . For $S = \mathbb{Z}_n$ Proposition 1.4 is meaningless and consequently does not imply the Fürstenberg-Sárközy Theorem [5,12]. However, for dense sets S satisfying $\|S\|_u = o(\|S\|_{L^1(\mathbb{Z}_n)})$ this proposition is meaningful.

Let us now contrast Proposition 1.4 with the emergence of 3SAPs with respect to two sets A and S both taken in \mathbb{Z}_n . Here the restriction over the pseudorandomness of S must be more substantial. In particular one must pose a restriction over the U^3 norm of $S - \|S\|_{L^1(\mathbb{Z}_n)}$, while for 2SAPs a restriction over the U^2 norm of $S - \|S\|_{L^1(\mathbb{Z}_n)}$ was sufficient.

To see this, fix $\varepsilon < 1/10$, fix an irrational number ϑ , and let n be a sufficiently large integer. Consider the sets $A = \{x \bmod n : \|x^2\vartheta\| < \varepsilon\} \subseteq \mathbb{Z}_n$ and $S = \{d \bmod n : \|2d^2\vartheta\| > 1/2 - \varepsilon\} \subseteq \mathbb{Z}_n$, where for a real number $t \in \mathbb{R}$ we write $\|t\|$ to denote the distance of t to the integers⁸. Both A and S are dense and pseudorandom in the sense that $\|A\|_u = o(|A|)$ and $\|S\|_u = o(|S|)$ see e.g. [7, pp. 9 – 10]. Nevertheless, A contains no 3SAPs. Indeed, let us assume, towards contradiction, that $(x, x+d, x+2d)$ is a 3SAP in A (so that, $d \in S$). Then

$$\| -2(x+d)^2\vartheta \| = \|2(x+d)^2\vartheta\| \leq 2\|(x+d)^2\vartheta\| < 2\varepsilon.$$

Observe that

$$\|2d^2\vartheta\| = \|(x^2 - 2(x+d)^2 + (x+2d)^2)\vartheta\| < 4\varepsilon$$

contradicting the assumption that $d \in S$ and satisfying $\|2d^2\vartheta\| > 1/2 - \varepsilon$.

Our main result is the resolution of Problem 1.3 for sets taken in the group \mathbb{F}_p^n where p is an odd prime and n is sufficiently large. We state this next.

Theorem 1.5 *Let p be an odd prime. For every $\alpha > 0$ and $\sigma > 0$ there exist an $\eta > 0$ a $C > 0$, and an integer $n_0 > 0$ such that for every integer $n \geq n_0$ the following holds.*

Let A and S be subsets of \mathbb{F}_p^n of densities α and σ , respectively, such that $\|S\|_u \leq \eta\sigma$. Then A contains at least $C|S|p^n$ 3SAPs.

⁷ The linear bias of a set S is essentially equivalent to the U^2 norm of the balanced function of S [14].

⁸ That is $\|t\| = \min\{\{t\}, 1 - \{t\}\}$, where $\{t\}$ is the fractional part of t .

Our proof of Theorem 1.5 relies on three results that are used in conjunction. The first is the so called *arithmetic regularity lemma* established by Green and Tao [10]. For our purposes the variant of this lemma found in [9] will be sufficient. This is presented in Section 2. The second result is a *Generalised von Neumann* type lemma that fits for 3SAPs. This is presented in Section 3. The third, is a lemma for counting 3SAPs along certain structured functions and is presented in Section 4. Finally, a sketch of our proof of Theorem 1.5 is presented in Section 5.

2 An arithmetic regularity lemma

The aim of this section is to state Theorem 2.2 [9]. For a σ -algebra \mathcal{B} of \mathbb{F}_p^n and $x \in \mathbb{F}_p^n$, we write $\mathcal{B}(x)$ to denote the atom of \mathcal{B} containing x . Given $f : \mathbb{F}_p^n \rightarrow \mathbb{C}$ we write

$$\mathbb{E}(f|\mathcal{B})(x) = \mathbb{E}_{\mathcal{B}(x)}f = \frac{1}{|\mathcal{B}(x)|} \sum_{y \in \mathcal{B}(x)} f(y)$$

to denote the average of f over the atom of \mathcal{B} containing x .

Let $f_1, \dots, f_k \in \mathbb{C}^{\mathbb{F}_p^n}$ be functions of the form $\mathbb{F}_p^n \rightarrow \mathbb{C}$. A σ -algebra of \mathbb{F}_p^n with each of its atoms of the form $\{x \in \mathbb{F}_p^n : f_1(x) = z_1, \dots, f_k(x) = z_k\}$, where $z_1, \dots, z_k \in \mathbb{C}$, is called a *factor* of \mathbb{F}_p^n . A factor of \mathbb{F}_p^n each of whose atoms has the form $\{x \in \mathbb{F}_p^n : (r_1^T x, \dots, r_k^T x) = a\}$ where $r_1, \dots, r_k \in \mathbb{F}_p^n$ and $a \in \mathbb{F}_p^k$ is called a *linear factor of complexity k* , and we say that this linear factor is *generated* by r_1, \dots, r_k .

Definition 2.1 Let $r_1, \dots, r_{d_1} \in \mathbb{F}_p^n$ and let M_1, \dots, M_{d_2} be symmetric $n \times n$ matrices over \mathbb{F}_p . Let \mathcal{B}_1 be the linear factor generated by r_1, \dots, r_{d_1} , and let \mathcal{B}_2 be the factor generated⁹ by the quadratic forms $x^T M_1 x, \dots, x^T M_{d_2} x$ and the linear forms $r_1^T x, \dots, r_{d_1}^T x$. The pair $(\mathcal{B}_1, \mathcal{B}_2)$ is called a *quadratic factor of complexity (d_1, d_2)* .

Let $(\mathcal{B}_1, \mathcal{B}_2)$ be a quadratic factor of complexity (d_1, d_2) . The atoms of \mathcal{B}_1 are indexed using the elements of $\mathbb{F}_p^{d_1}$. The atoms of \mathcal{B}_2 are indexed using the

⁹ The atoms of \mathcal{B}_2 have the form

$$\{x \in \mathbb{F}_p^n : r_1^T x = c_1, \dots, r_{d_1}^T x = c_{d_1} \text{ and } x^T M_1 x = z_1, \dots, x^T M_{d_2} x = z_{d_2}\},$$

where $(r_1, \dots, r_{d_1}) \in \mathbb{F}_p^{d_1}$ and $(z_1, \dots, z_{d_2}) \in \mathbb{F}_p^{d_2}$.

elements of $\mathbb{F}_p^{d_1} \times \mathbb{F}_p^{d_2}$. We map an $x \in \mathbb{F}_p^n$ to the pair

$$(\Gamma(x), \Phi(x)) = ((r_1^T x, \dots, r_{d_1}^T x), (x^T M_1 x, \dots, x^T M_{d_2} x)) \in \mathbb{F}_p^{d_1} \times \mathbb{F}_p^{d_2},$$

so that $(\Gamma(x), \Phi(x))$ is the atom of the quadratic factor containing x .

We write $\text{rk}M$ to denote the rank of a matrix M . A quadratic factor of complexity (d_1, d_2) satisfying $\text{rk}(\lambda_1 M_1 + \dots + \lambda_{d_2} M_{d_2}) \geq r$ for any $\lambda_1, \dots, \lambda_{d_2} \in \mathbb{F}_p$ not all zero, where M_1, \dots, M_{d_2} are the symmetric matrices involved in its generation, is said to have *rank* at least r .

Theorem 2.2 ([9, Proposition 3.12], [6, Theorem 3.5])

For every real $\delta > 0$ and every two growth functions $\omega_{\text{rk}}, \omega_{\text{uni}} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ (which may be independent of δ) there exists an n_0 such that for every integer $n \geq n_0$ the following holds.

For every function $f : \mathbb{F}_p^n \rightarrow [-1, 1]$ there exists a constant d_0 , a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$, and a decomposition $f = f_{\text{str}} + f_{\text{uni}} + f_{\text{neg}}$ satisfying the following terms.

- (i) *The complexity of $(\mathcal{B}_1, \mathcal{B}_2)$ is at most (d_1, d_2) where $d_1, d_2 \leq d_0$;*
- (ii) *the rank of $(\mathcal{B}_1, \mathcal{B}_2)$ is at least $\omega_{\text{rk}}(d_1 + d_2)$;*
- (iii) *and*

$$f_{\text{str}} = \mathbb{E}(f|\mathcal{B}_2), \quad \|f_{\text{neg}}\|_{L^2(\mathbb{F}_p^n)} \leq \delta, \quad \text{and} \quad \|f_{\text{uni}}\|_{U^3(\mathbb{F}_p^n)} \leq 1/\omega_{\text{uni}}(d_1 + d_2).$$

3 A generalised von Neumann type lemma

The aim of this section is to state a generalised von Neumann type lemma for k SAPs with respect to two sets A and S taken in an arbitrary finite abelian group G . Proof of this is omitted.

Given $S \subseteq G$ with $\sigma = \|S\|_{L^1(G)}$ set

$$\mu_S(x) = S(x) (\mathbb{E}_{x \in G} S(x))^{-1} = S(x)/\sigma,$$

Lemma 3.1 *Let $k \geq 3$ be an integer, let $\mathcal{F} = \{f_1, \dots, f_k\}$ be a collection of complex valued functions over G , let $S \subseteq G$, and fix $g \in \mathcal{F}$. If $\|f\|_\infty \leq 1$ for each $f \in \mathcal{F} \setminus \{g\}$, then*

$$\begin{aligned} |\mathbb{E}_{x \in G, d \in G} f_1(x) \cdots f_k(x + (k-1)d) \mu_S(d)| \leq \\ \left(\|S\|_{L^1(G)}^2 + \|S\|_u \|S\|_{L^1(G)} \right)^{1/2^{k-1}} \|g\|_{U^k(G)} \|S\|_{L^1(G)}^{-1} \end{aligned} \quad (1)$$

4 Counting 3SAPs over atoms of quadratic factors

The aim of this section is to state Lemma 4.2 proof of which is omitted. Roughly speaking, this lemma estimates the number of 3SAPs with respect to a function of the form of f_{str} (see Section 2).

Definition 4.1 Let $S \subseteq \mathbb{F}_p^n$, and let $(\mathcal{B}_1, \mathcal{B}_2)$ be a quadratic factor of \mathbb{F}_p^n of complexity at most (d_1, d_2) . A quadruple of atoms $((a^{(0)}, b^{(0)}), (a^{(1)}, b^{(1)}), (a^{(2)}, b^{(2)}), (a^{(3)}, b^{(3)}))$ satisfying

$$(a^{(0)}, a^{(1)}, a^{(2)}, a^{(3)}) \text{ is a } 4\Gamma(S)\text{AP in } \mathbb{F}_p^{d_1},$$

and

$$b^{(0)} - 3b^{(1)} + 3b^{(2)} - b^{(3)} = 0$$

is called *viable*, where Γ is as in Section 2 and $\Gamma(S)$ is the image of S under Γ .

Given a viable quadruple, the following lemma estimates the number of 3SAPs contained within the first three atoms of the quadruple. The need for viable quadruples is too technical to be motivated in this short note.

Lemma 4.2 Let $S \subseteq \mathbb{F}_p^n$, let $(\mathcal{B}_1, \mathcal{B}_2)$ be a quadratic factor of \mathbb{F}_p^n of rank at least r and complexity at most (d_1, d_2) , let $((a^{(0)}, b^{(0)}), (a^{(1)}, b^{(1)}), (a^{(2)}, b^{(2)}), (a^{(3)}, b^{(3)}))$ be viable, and let

$$X = \{(x, d) \in \mathbb{F}_p^n \times S : x + jd \in (a^{(j)}, b^{(j)}), 0 \leq j \leq 2\}$$

denote the set of 3SAPs found within the first three atoms of the quadruple. Then,

$$|X| = \left[p^{-2d_1 - 3d_2} \pm \left(\|S\|_u \|S\|_{L^1(\mathbb{F}_p^n)}^{-1} + 4p^{-r/2} \right) \right] p^n |S|.$$

5 Sketch of our proof of Theorem 1.5

Given A, S, α , and σ as in Theorem 1.5 we show that

$$\mathbb{E}_{x, d \in \mathbb{F}_p^n} A(x)A(x+d)A(x+2d)\mu_S(d) \geq \alpha^4/2^6,$$

where $\mu_S(x) = S(x)\|S\|_{L^1(\mathbb{F}_p^n)}^{-1}$.

Roughly speaking, we apply Theorem 2.2 and obtain a decomposition $A =$

$f_{\text{str}} + f_{\text{uni}} + f_{\text{neg}}$ and consequently obtain

$$\begin{aligned}
& |\mathbb{E}_{x,d \in \mathbb{F}_p^n} A(x)A(x+d)A(x+2d)\mu_S(d)| = \\
& \quad |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{str}}(x)f_{\text{str}}(x+d)f_{\text{str}}(x+2d)\mu_S(d)| \\
& \quad \pm |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{neg}}(x)A(x+d)A(x+2d)\mu_S(d)| \\
& \quad \pm |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{uni}}(x)A(x+d)A(x+2d)\mu_S(d)| \\
& \quad \pm |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{str}}(x)f_{\text{neg}}(x+d)A(x+2d)\mu_S(d)| \\
& \quad \pm |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{str}}(x)f_{\text{uni}}(x+d)A(x+2d)\mu_S(d)| \\
& \quad \pm |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{str}}(x)f_{\text{str}}(x+d)f_{\text{neg}}(x+2d)\mu_S(d)| \\
& \quad \pm |\mathbb{E}_{x,d \in \mathbb{F}_p^n} f_{\text{str}}(x)f_{\text{str}}(x+d)f_{\text{uni}}(x+2d)\mu_S(d)|.
\end{aligned}$$

The terms involving f_{neg} are bounded using the assumption that $\|f_{\text{neg}}\|_{L^2(\mathbb{F}_p^n)}$ is small (see Theorem 2.2). The terms involving f_{uni} are bounded using (1) and the assumption that f_{uni} is pseudorandom (see Theorem 2.2). A lower bound for the main term involving only occurrences of f_{str} is obtained through Lemma 4.2 and is consequently shown to dominate all other terms.

References

- [1] Bergelson, V. and A. Leibman, *Polynomial extensions of van der Waerden’s and Szemerédi’s theorems*, J. Amer. Math. Soc. **9** (1996), pp. 725–753.
- [2] Christ, M., *On random multilinear operator inequalities* (2011).
- [3] Frantzikinakis, N., E. Lesigne and M. Wierdl, *Random sequences and pointwise convergence of multiple ergodic averages*, Indiana Univ. Math. J. **61** (2012), pp. 585–617.
- [4] Frantzikinakis, N., E. Lesigne and M. Wierdl, *Random differences in Szemerédi’s theorem and related results* (2013).
- [5] Fürstenberg, H., “Recurrence in ergodic theory and combinatorial number theory,” Princeton University Press, Princeton, N.J., 1981, xi+203 pp., m. B. Porter Lectures.
- [6] Gowers, W. T. and J. Wolf, *The true complexity of a system of linear equations*, Proc. Lond. Math. Soc. (3) **100** (2010), pp. 155–176.
- [7] Granville, A. and Z. Rudnick, editors, “Equidistribution in number theory, an introduction,” NATO Science Series II: Mathematics, Physics and Chemistry **237**, Springer, Dordrecht, 2007.

- [8] Green, B., *On arithmetic structures in dense sets of integers*, Duke Math. J. **114** (2002), pp. 215–238.
- [9] Green, B., *Montréal notes on quadratic Fourier analysis*, in: *Additive combinatorics*, CRM Proc. Lecture Notes **43**, Amer. Math. Soc., Providence, RI, 2007 pp. 69–102.
- [10] Green, B. and T. Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, in: *An irregular mind*, Bolyai Soc. Math. Stud. **21**, János Bolyai Math. Soc., Budapest, 2010 pp. 261–334.
- [11] Hoffman, A. J., *On eigenvalues and colorings of graphs*, in: *Graph Theory and its Applications (Proc. Advanced Sem., Math. Research Center, Univ. of Wisconsin, Madison, Wis., 1969)*, Academic Press, New York, 1970 pp. 79–91.
- [12] Sárközy, A., *On difference sets of sequences of integers. I*, Acta Math. Acad. Sci. Hungar. **31** (1978), pp. 125–149.
- [13] Szemerédi, E., *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), pp. 199–245, collection of articles in memory of Juriĭ Vladimirovič Linnik.
- [14] Tao, T. and V. H. Vu, “Additive combinatorics,” Cambridge Studies in Advanced Mathematics **105**, Cambridge University Press, Cambridge, 2010, xviii+512 pp., paperback edition [of MR2289012].