ARITHMETIC PROGRESSIONS IN GENERIC SETS

FARUK TEMUR

Abstract. In this short note we aim to study the problem of existence of arithmetic progressions in a generic subset of $\{1, 2, \ldots, N\}$. We observe that by considering almost every set obeying a density lower bound instead of every set we get very different, and much stronger results.

1. Introduction

Study of arithmetic progressions in certain subsets of $[N] := \{1, 2, \ldots, N\}$ spawned the large and fast developing field of additive combinatorics. For a short account of this history see the introduction of [2], for more comprehensive accounts [6, 8]. But all efforts in this field appear to be concentrated on the existence of arithmetic progressions either in every subset in $[N]$ with a lower bound on density, or in specific sets like primes. In this short note we will discuss the existence of arithmetic progressions in generic subsets of $[N]$ with a lower bound on density. We will observe that by considering almost every set obeying a density lower bound instead of every set we get very different, and much stronger results.

We formalize genericity using selector variables, as in [1, 5]. Let $\xi_n, n \in [N]$ be independent, identically distributed random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}[\xi_n = 1] = N^{-\delta}$ and $\mathbb{P}[\xi_n = 0] = 1 - N^{-\delta}$ with $0 < \delta < 1$. In this case $\{n\xi_n\}_{n\in[N]}$ are random subsets of $[N]$. Then, for which δ do we have

$$\lim_{N\to\infty} \mathbb{P}\left[\{n\xi_n\}_{n\in[N]} \text{ has at least one k-term AP}\right] \to 1? \tag{1}$$

It turns out that this question can be answered briefly.

Theorem 1. *The statement* (1) *holds if and only if* $\delta < 2/k$.

This presents a very different picture compared to the deterministic case. By [3], that improves upon the well-known Behrend bound we know that there are sets $A \subseteq [N]$ that contains no three term arithmetic progression with

$$|A| = \Omega\Big(\frac{N \log^{1/4} N}{2^{2\sqrt{2}\sqrt{\log_2 N}}}\Big).$$

Conversely, it is known by the work of Kelley and Meka [4] if A contains no 3-term AP then

$$|A| \le e^{-\Omega(\log^{1/11} N)}.$$

For comparison, in our formulation this means $\delta > 0$ is not possible if we consider every subset of $[N]$.

Date: December 7, 2024.
2020 *Mathematics Subject Classification.* Primary: 11B25, 60C05 .
Key words and phrases. Arithmetic progressions, Generic subsets.

2. Proof of our Theorem

Before going to the proof we present a heuristic argument that may be helpful in guiding future research on similar problems. Suppose we pick an arbitrary ω, then the set $\{n\xi_n\}_{n\in[N]}$ will have essentially $N^{1-\delta}$ elements. So we can pick two elements from this set in essentially $N^{2-2\delta}$ ways. Now once two elements are picked, to complete these to a k-term arithmetic progression, we need $k-2$ essentially fixed elements to be in our set. Each are in the set with probability $N^{-\delta}$, so all of them are in the set with probability $N^{-(k-2)\delta}$. Summing these probabilities over all possible choices of two fixed elements we get $N^{2-k\delta}$. Then we may surmise that an arbitrary set contain an arithmetic progression with high probability if the power $2-k\delta$ is positive.

We now present our proof. It actually yields much more information than is given in the statement of the theorem. The proof proceeds via a multilinear form that counts aritmetic progressions. The product structure of these forms is very suitable for using the independence of selector variables.

Proof. We begin by observing that for any set $\{n\xi_n\}_{n\in[N]}$ the random variable

$$Z_{\delta,k,N} := \sum_{n=1}^{N} \sum_{r=1}^{\lfloor \frac{N-1}{k-1} \rfloor} \prod_{j=0}^{k-1} \xi_{n+jr} \tag{2}$$

counts the number of k-term arithmetic progressions. This formulation is very handy to us as will be seen below. If we take the expectation of this we obtain the average number of k-term AP contained in $\{n\xi_n\}_{n\in[N]}$. By independence of the selector variables we have

$$\mathbb{E}\, Z_{\delta,k,N} = \sum_{n=1}^{N} \sum_{r=1}^{\lfloor \frac{N-n}{k-1} \rfloor} \prod_{j=0}^{k-1} \mathbb{E}\, \xi_{n+jr} = \sum_{n=1}^{N} \sum_{r=1}^{\lfloor \frac{N-n}{k-1} \rfloor} N^{-k\delta} = N^{-k\delta} \sum_{n=1}^{N} \Big\lfloor \frac{N-n}{k-1} \Big\rfloor \tag{3}$$

We can estimate this last sum from above by

$$\sum_{n=1}^{N} \frac{N-n}{k-1} N^{-k\delta} = \frac{N^{-k\delta}}{k-1} \sum_{n=1}^{N} N-n = \frac{N^{-k\delta}}{k-1} \Big[N^2 - \frac{N(N+1)}{2} \Big] = \frac{N^{2-k\delta}}{2(k-1)} - \frac{N^{1-k\delta}}{2(k-1)},$$

and estimate it from below by

$$\sum_{n=1}^{N} \Big[\frac{N-n}{k-1} - 1 \Big] N^{-k\delta} = \frac{N^{2-k\delta}}{2(k-1)} - \frac{N^{1-k\delta}}{2(k-1)} - N^{1-k\delta}.$$

By the Markov inequality

$$\mathbb{P}[Z_{\delta,k,N} \geq 1] \leq \mathbb{E}\, Z_{\delta,k,N} \leq \frac{N^{2-k\delta} - N^{1-k\delta}}{2(k-1)}.$$

As $N \to \infty$ the right hand side goes to zero if $\delta > 2/k$, and to $1/(2k-2)$ if $\delta = 2/k$. This makes it plain that for $\delta \geq 2/k$ the statement (1) is wrong.

It remains to prove that the statement holds for $\delta < 2/k$. This can be done using concentration inequalities, which investigate how much a random variable concentrates around its mean. Since we already know the mean of our random variable, if there is enough concentration this will suffice. Fortunately the simplest of concentration inequalities, the Chebyshev inequality suffices. It asserts that if a random variable X has finite expectation and variance, then

$$\mathbb{P}\left[|X - \mathbb{E}\, X| \geq \lambda \sqrt{\mathbb{V}\, X} \right] \leq \lambda^{-2}.$$

From this it immediately follows that

$$\mathbb{P}\left[X \in \left(\mathbb{E}\,X - \lambda\sqrt{\mathbb{V}\,X}, \mathbb{E}\,X + \lambda\sqrt{\mathbb{V}\,X}\right)\right] \geq 1 - \lambda^{-2}.$$

For this to be useful for us we need $1 \leq \mathbb{E}\,X - \lambda\sqrt{\mathbb{V}\,X}$, which will be the case for us. Since $\mathbb{V}\,X = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2$ we calculate $\mathbb{E}\,X^2$.

$$\mathbb{E}\,Z_{\delta,k,N}^2 = \sum_{n,m=1}^{N} \sum_{r,s=1}^{\lfloor \frac{N-1}{k-1} \rfloor} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right]. \tag{4}$$

To evaluate the expectation we will decompose the sum. We let

$$\begin{aligned}
A &:= \left\{(n,m,r,s) :\ 1 \leq n,m \leq N,\quad 1 \leq r \leq \left\lfloor \frac{N-n}{k-1} \right\rfloor,\quad 1 \leq s \leq \left\lfloor \frac{N-m}{k-1} \right\rfloor\right\} \\
B &:= \left\{(n,m,r,s) \in A : n + jr = m + is \text{ for some } 0 \leq j,i \leq k-1\right\} \\
D &:= \left\{(n,m,r,s) \in A : \text{for pairs } \ (i_1,j_1) \neq (i_2,j_2),\quad 0 \leq j_l, i_l \leq k-1,\quad n + j_l r = m + i_l s\right\}.
\end{aligned}$$

So the set A is all possible quadruples in our sum, it is the cartesian product of the set

$$\left\{(n,r) : 1 \leq n \leq N,\quad 1 \leq r \leq \left\lfloor \frac{N-n}{k-1} \right\rfloor\right\}$$

with itself, and we have calculated its cardinality as a sum in (3). The set B are those quadruples that lie on hyperplanes given by equations $n + jr = m + is$. For fixed i,j the number of quadruples in A that satisfy this inequality cannot be larger than N^3, since fixing three of the variables fix the last one. So $|B| \leq k^2N^3$. Finally the elements of D lie on two separate hyperplanes, and for two fixed equations $n + j_l r = m + i_l s$ with $(i_1,j_1) \neq (i_2,j_2)$ there are at most N^2 quadruples in A satisfying both of these. Therefore we have $|D| \leq k^4N^2$. Then we can partition our last sum in (4)

$$= \sum_{A\setminus B} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right] + \sum_{B\setminus D} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right] + \sum_{D} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right].$$

We observe that

$$N^{-2k\delta} \leq \mathbb{E} \prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is} \leq N^{-k\delta}.$$

for there are between k and $2k$ different selector variables in this product, and a selector variable's natural number powers are the same as itself. Therefore

$$\sum_{D} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right] \leq k^4N^2N^{-k\delta} = k^4N^{2-k\delta}.$$

As in the product over $A \setminus B$ no two selector variables are the same, all are independent and we have

$$\sum_{A\setminus B} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right] \leq |A|N^{-2k\delta} = \left[\sum_{n=1}^{N} \left\lfloor \frac{N-n}{k-1} \right\rfloor\right]^2 N^{-2k\delta}.$$

Finally, on $B \setminus D$ only two selector variables can be the same, so

$$\sum_{B\setminus D} \mathbb{E}\left[\prod_{j=0}^{k-1} \xi_{n+jr} \prod_{i=0}^{k-1} \xi_{m+is}\right] \leq k^2N^3N^{-(2k-1)\delta} = k^2N^{3-(2k-1)\delta}.$$

Let $\alpha = \min\{2-k\delta, 1-\delta\}$. Since $\delta < 2/k$ this number is positive. Combining our findings

$$\mathbb{V}\, Z_{\delta,k,N} \le k^4 N^{2-k\delta} + k^2 N^{3-(2k-1)\delta} \le C_k N^{-\alpha}\big(\mathbb{E}\, Z_{\delta,k,N}\big)^2,$$

Hence picking $\lambda = N^{\alpha/8}$

$$\sqrt{\mathbb{V}\, Z_{\delta,k,N}} \le N^{-\alpha/4}\, \mathbb{E}\, Z_{\delta,k,N} \implies \mathbb{E}\, Z_{\delta,k,N} - N^{\alpha/8}\sqrt{\mathbb{V}\, Z_{\delta,k,N}} \ge (1-N^{-\alpha/8})\, \mathbb{E}\, Z_{\delta,k,N},$$

from which we conclude

$$\mathbb{P}[Z_{\delta,k,N} \ge 1] \ge \mathbb{P}\left[|Z_{\delta,k,N} - \mathbb{E}\, Z_{\delta,k,N}| \le N^{\alpha/8}\sqrt{\mathbb{V}\, Z_{\delta,k,N}}\right] \ge 1 - N^{-\alpha/4}.$$

□

3. Future directions

In additive combinatorics many different questions are pursued regarding arithmetic progressions such as finding generalized arithmetic progressions, polynomial patterns, affine copies etc. in sets with a lower bound on density, or in fixed special sets. All of these questions may be pursued for generic sets as in this article.

Also we would like to highlight another approach, recently introduced in [7], that randomizes sets via stochastic processes rather than selector variables. Study of arithmetic progressions in generic sets can be carried out with this method as well. Especially if we want to randomize a fixed set, such as squares, cubes etc., this method is perfectly suited. Consult [2] for studies on arithmetic progressions in such sets.

4. acknowledgements

The author thanks his colleague Haydar Göral for very fruitful seminars and workshops he organized on additive combinatorics. It was during one of these seminars that this question occurred to the author.

References

[1] J. Bourgain, *Bounded orthogonal systems and the $\Lambda(p)$-set problem*, Acta Math. **162**, no. 3-4, (1989), 227–245.
[2] Ş. Çam Çelik, S. Eyidoğan, H. Göral, D.C. Sertbaş, *On classification of sequences containing arbitrarily long arithmetic progressions*, Int. J. of Number Theory, 2023
[3] M. Elkin, *An improved construction of progression-free sets*, Israel J. Math. 184 (2011), 93–128
[4] Z. Kelley, R. Meka,*Strong Bounds for 3-Progressions*
[5] G. Mockenhaupt and W. Schlag, *On the Hardy-Littlewood majorant problem for random sets*, Journal of functional analysis **256**, (2009), 1189-1237.
[6] T. Tao, V. Vu, *Additive combinatorics*, Cambridge University press
[7] F. Temur, C. Sahillioğulları, Random exponential sums and lattice points in regions, arXiv preprint arXiv:2411.06517
[8] Y. Zhao,*Graph Theory and Additive Combinatorics*, Cambridge University press, 2023

Department of Mathematics, Izmir Institute of Technology, Urla, Izmir, 35430, Turkey
Email address: `faruktemur@iyte.edu.tr`