

A Note on Elkin's Improvement of Behrend's Construction

Ben Green and Julia Wolf

To Mel Nathanson

Summary We provide a short proof of a recent result of Elkin in which large subsets of $\{1, \dots, N\}$ free of three-term progressions are constructed.

Keywords Arithmetic progressions · Roth's theorem

Mathematics Subject Classifications (2010). Primary 11B25, Secondary 11B75

1 Introduction

Write $r_3(N)$ for the cardinality of the largest subset of $\{1, \dots, N\}$ not containing three distinct elements in arithmetic progression. A famous construction of Behrend [1] shows, when analysed carefully, that

$$r_3(N) \gg \frac{1}{\log^{1/4} N} \cdot \frac{N}{2^{2\sqrt{2}\sqrt{\log_2 N}}}.$$

In a recent preprint [2], Elkin was able to improve this 62-year old bound to

$$r_3(N) \gg \log^{1/4} N \cdot \frac{N}{2^{2\sqrt{2}\sqrt{\log_2 N}}}.$$

B. Green

Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, England
e-mail: b.j.green@dpmms.cam.ac.uk

J. Wolf

Mathematical Sciences Research Institute, 17 Gauss Way, Berkeley, CA 94720, USA
e-mail: julia.wolf@cantab.net

The first author holds a Leverhulme Prize and is grateful to the Leverhulme Trust for their support. This paper was written while the authors were attending the special semester in ergodic theory and additive combinatorics at MSRI.

Our aim in this note is to provide a short proof of Elkin's result. It should be noted that the only advantage of our approach is brevity: it is based on ideas morally close to those of Elkin, and moreover, his argument is more constructive than ours.

Throughout the paper, $0 < c < 1 < C$ denote absolute constants which may vary from line to line. We write $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ for the d -dimensional torus.

2 The Proof

Let d be an integer to be determined later, and let $\delta \in (0, 1/10)$ be a small parameter (we will have $d \sim C \sqrt{\log N}$ and $\delta \sim \exp(-C \sqrt{\log N})$). Given $\theta, \alpha \in \mathbb{T}^d$, write $\Psi_{\theta, \alpha} : \{1, \dots, N\} \rightarrow \mathbb{T}^d$ for the map $n \mapsto \theta n + \alpha \pmod{1}$.

Lemma 2.1. *Suppose that n is an integer. Then $\Psi_{\theta, \alpha}(n)$ is uniformly distributed on \mathbb{T}^d as θ, α vary uniformly and independently over \mathbb{T}^d . Moreover, if n and n' are distinct positive integers, then the pair $(\Psi_{\theta, \alpha}(n), \Psi_{\theta, \alpha}(n'))$ is uniformly distributed on $\mathbb{T}^d \times \mathbb{T}^d$ as θ, α vary uniformly and independently over \mathbb{T}^d .*

Proof. Only the second statement requires an argument to be given. Perhaps the easiest proof is via Fourier analysis, noting that

$$\int e^{2\pi i(k \cdot (\theta n + \alpha) + k' \cdot (\theta n' + \alpha))} d\theta d\alpha = 0$$

unless $k + k' = kn + k'n' = 0$. Provided that k and k' are not both zero, this cannot happen for distinct positive integers n, n' . Since the exponentials $e^{2\pi i(kx + k'x')}$ are dense in $L^2(\mathbb{T}^d \times \mathbb{T}^d)$, the result follows. \square

Let us identify \mathbb{T}^d with $[0, 1)^d$ in the obvious way. For each $r \leq \frac{1}{2}\sqrt{d}$, write $S(r)$ for the region

$$\{x \in [0, 1/2]^d : r - \delta \leq \|x\|_2 \leq r\}.$$

Lemma 2.2. *There is some choice of r for which $\text{vol}(S(r)) \geq c\delta 2^{-d}$.*

Proof. First note that if (x_1, \dots, x_d) is chosen at random from $[0, 1/2]^d$ then, with probability at least c , we have $\|x\|_2 - \sqrt{d/12} \leq C$. This is a consequence of standard tail estimates for sums of independent identically distributed random variables, of which $\|x\|_2^2 = \sum_{i=1}^d x_i^2$ is an example. The statement of the lemma then immediately follows from the pigeonhole principle. \square

Write $S := S(r)$ for the choice of r whose existence is guaranteed by the preceding lemma; thus $\text{vol}(S) \geq c\delta 2^{-d}$. Write \tilde{S} for the same set S but considered now as a subset of $[0, 1/2]^d \subseteq \mathbb{R}^d$. Since there is no “wraparound”, the three-term progressions in S and \tilde{S} coincide and henceforth we abuse notation, regarding S as a subset of \mathbb{R}^d and dropping the tildes. (To use the additive combinatorics jargon, S and \tilde{S} are *Freiman isomorphic*.) Suppose that (x, y) is a pair for which $x - y, x$ and $x + y$ lie in S . By the parallelogram law

$$2\|x\|_2^2 + 2\|y\|_2^2 = \|x + y\|_2^2 + \|x - y\|_2^2$$

and straightforward algebra we have

$$\|y\|_2 \leq \sqrt{r^2 - (r - \delta)^2} \leq \sqrt{2\delta r}.$$

It follows from the formula for the volume of a sphere in \mathbb{R}^d that the volume of the set $B \subseteq \mathbb{T}^d \times \mathbb{T}^d$ in which each such pair (x, y) must lie is at most $\text{vol}(S)C^d(\delta/\sqrt{d})^{d/2}$.

The next lemma is an easy observation based on Lemma 2.1.

Lemma 2.3. *Suppose that N is even. Define $A_{\theta,\alpha} := \{n \in [N] : \Psi_{\theta,\alpha}(n) \in S\}$. Then*

$$\mathbb{E}_{\theta,\alpha}|A_{\theta,\alpha}| = N \text{vol}(S)$$

whilst the expected number of nontrivial three-term arithmetic progressions in $A_{\theta,\alpha}$ is

$$\mathbb{E}_{\theta,\alpha}T(A_{\theta,\alpha}) = \frac{1}{4}N(N - 5)\text{vol}(B).$$

Proof. The first statement is an immediate consequence of the first part of Lemma 2.1. Now each nontrivial three-term progression is of the form $(n - d, n, n + d)$ with $d \neq 0$. Since N is even there are $N(N - 5)/4$ choices for n and d , and each of the consequent progressions lies inside $A_{\theta,\alpha}$ with probability $\text{vol}(B)$ by the second part of Lemma 2.1. □

To finish the argument, we just have to choose parameters so that

$$\frac{1}{3}\text{vol}(S) \geq \frac{1}{4}(N - 5)\text{vol}(B). \tag{2.1}$$

Then, we shall have

$$\mathbb{E} \left(\frac{2}{3}|A_{\theta,\alpha}| - T(A_{\theta,\alpha}) \right) \geq \frac{1}{3}N \text{vol}(S).$$

In particular, there is a specific choice of $A := A_{\theta,\alpha}$ for which both $T(A) \leq 2|A|/3$ and $|A| \geq \frac{1}{2}N \text{vol}(S)$. Deleting up to two thirds of the elements of A , we are left with a set of size at least $\frac{1}{6}N \text{vol}(S)$ that is free of three-term arithmetic progressions.

To do this, it suffices to have $C^d (\delta/\sqrt{d})^{d/2} \leq c/N$, which can certainly be achieved by taking $\delta := c\sqrt{d}N^{-2/d}$. For this choice of parameters we have, by the earlier lower bound on $\text{vol}(S)$, that

$$|A| \geq \frac{1}{6}N \text{vol}(S) \geq c\sqrt{d}2^{-d}N^{1-2/d}.$$

Choosing $d := \lceil \sqrt{2 \log_2 N} \rceil$ we recover Elkin's bound. \square

3 A Question of Graham

The authors did not set out to try and find a simpler proof of Elkin's result. Rather, our concern was with a question of Ron Graham (personal communication to the first-named author, see also [3, 4]). Defining $W(2; 3, k)$ to be the smallest N such that any red-V-blue colouring of $[N]$ contains either a three-term red progression or a k -term blue progression, Graham asked whether $W(2; 3, k) < k^A$ for some absolute constant A or, even more ambitiously, whether $W(2; 3, k) \leq Ck^2$. Our initial feeling was that the answer was surely no, and that a counterexample might be found by modifying the Behrend example in such a way that its complement does not contain long progressions. Reinterpreting the Behrend construction in the way that we have done here, it seems reasonably clear that it is not possible to provide a negative answer to Graham's question in this way.

Acknowledgement The authors are grateful to Tom Sanders for helpful conversations.

References

- [1] F. Behrend. *On sets of integers which contain no three terms in arithmetic progression*, Proc. Nat. Acad. Sci., **32**:331–332, 1946
- [2] M. Elkin, *An improved construction of progression-free sets*, available at <http://arxiv.org/abs/0801.4310>
- [3] R. Graham, *On the growth of a van der Waerden-like function*, Integers, **6**:#A29, 2006
- [4] B. Landman, A. Robertson and C. Culver, *Some new exact van der Waerden numbers*, Integers, **5**(2):#A10, 2005