Sets without \(k\)-term progressions can have many shorter progressions

Jacob Fox\(^1\) | Cosmin Pohoata\(^2\)

\(^1\)Department of Mathematics, Stanford University, Stanford, CA, USA
\(^2\)Department of Mathematics, California Institute of Technology, Pasadena, CA, USA

Correspondence
Cosmin Pohoata, Department of Mathematics, California Institute of Technology, Pasadena, CA 91106.
Email: andrei.pohoata@yale.edu

Funding information
This research was supported by the NSF grant, Grant/Award Number: DMS-1855635.

Abstract
Let \(f_{s,k}(n)\) be the maximum possible number of \(s\)-term arithmetic progressions in a set of \(n\) integers which contains no \(k\)-term arithmetic progression. For all fixed integers \(k \geq s \geq 3\), we prove that \(f_{s,k}(n) = n^{2-o(1)}\), which answers an old question of Erdős. In fact, we prove upper and lower bounds for \(f_{s,k}(n)\) which show that its growth is closely related to the bounds in Szemerédi’s theorem.

KEYWORDS
additive combinatorics, arithmetic progressions, probabilistic methods, Szemerédi’s theorem

1 | INTRODUCTION

Let \(k \geq 3\) be an integer. In this paper, a \(k\)-term arithmetic progression of integers will denote as usual a set of the form \(\{x, x+d, \ldots, x+(k-1)d\}\). If \(d \neq 0\), then we say that the progression is nontrivial. If a set \(A\) does not contain any nontrivial \(k\)-term arithmetic progressions, we say that \(A\) is \(k\)-AP free. The study of \(k\)-AP free sets in the integers and other groups has been a central topic in additive combinatorics. Following the standard notation, we will denote by \(r_k(n)\) the size of the largest \(k\)-AP free subset of \(\{1, \ldots, n\}\). The seminal result on this topic is Szemerédi’s Theorem [10], which states that sets of integers with positive density contain arbitrarily long arithmetic progressions, or using the notation above \(r_k(n) = o(n)\).

Since Szemerédi, the problem of finding better quantitative bounds for \(r_k(n)\) has received a lot of attention, with impressive progress that led to many important tools, which in the meantime have become standard. For our application, we will not need the best bounds for each \(k\), so we will limit ourselves to only mentioning Gowers’ theorem [4,5] (see also [6]) that for each \(k \geq 3\) there exists an absolute constant \(c_k > 0\) such that

\[
\left(\frac{n}{(\log \log n)^{c_k}}\right)^{r_k(n)}.
\]
Regarding lower bounds, Rankin [8] showed that there exists a constant $c'_k > 0$ such that

$$r_k(n) \gg \frac{n}{2^{c'_k (\log n)^{(s-2)}}}.$$  \hfill (1.2)

Throughout the paper, all logarithms are base 2 and the signs $\ll$ and $\gg$ are the usual Vinogradov symbols.

Let $\mathbf{A}_k(n)$ be the set of $n$-term nonnegative integer sequences which contain no $k$-term arith-
metric progression as a subsequence. Furthermore, let $f_s(A)$ denote the number of $s$-term arithmetic
progressions in $A$, and finally let $f_{s,k}(n) = \max_{A \in \mathbf{A}_k(n)} f_s(A)$. In [[3], page 119], Erdős observed that

$$\frac{\log f_{3,4}(n)}{\log n} > 1.4649$$

holds for infinitely many $n$ by constructing examples of sequences $A \in \mathbf{A}_4(3^s)$ for which $f_3(A) = 3^s - 1$. Furthermore, he noticed that for each $k > 3$ the limit $\lim_{n \to \infty} \log f_{3,k}(n)/\log n : = f_{3,k}$ exists, and asked whether or not $f_{3,k}$ is always less than 2. In [1], Simmons and Abbott improved on Erdős’ observation by showing that $f_{3,4}(n) \geq n^{1.623}$ holds infinitely often, and also proved that $f_{3,k} \to 2$ as $k$ goes to infinity. Nonetheless, in the regime when $k$ is fixed, there has been no further progress on understanding the limit $f_{3,k}$ as far as we are aware of. In this note, we settle Erdős’ question in the negative by proving the following more general result.

**Theorem 1.1.** For all integers $k > s \geq 3$, we have

$$\lim_{n \to \infty} \frac{\log f_{s,k}(n)}{\log n} = 2.$$

In fact, we prove upper and lower bounds for $f_{s,k}(n)$ which show that its growth is closely related to the bounds in Szemerédi’s theorem.

**Theorem 1.2.** There exist absolute positive constants $c$ and $C$ such that, for integers $k > s \geq 3$ and every sufficiently large integer $n$, we have

$$\left( \frac{c \cdot r_k(n)}{n} \right) 2^{(s-2)} \cdot n^2 \leq f_{s,k}(n) \leq \left( \frac{r_k(n)}{n} \right)^c \cdot n^2.$$

In light of the bounds on $r_k(n)/n$ provided by (1.1) and (1.2), it is easy to check that Theorem 1.1 follows from Theorem 1.2; therefore, it suffices to prove the latter. The proof of Theorem 1.2 will require a few ingredients from additive combinatorics, but we will state them in full as we will get to apply them, as they do not require much preparation.

**2 | PROOF OF THEOREM 1.2**

We first prove the desired upper bound on $f_{s,k}(n)$. For $s \geq 3$, we have $f_{s,k}(n) \leq f_{3,k}(n)$, so in order to prove the upper bound it suffices to show that

$$f_{3,k}(n) \leq \left( \frac{r_k(n)}{n} \right)^c n^2.$$
holds for some absolute constant \( C > 0 \) and sufficiently large \( n \). We will in fact show this claim for \( C = 1/25 \). Let \( A \in \mathcal{A}_k(n) \) and let \( pn^2 \) denote the number of three-term arithmetic progressions in \( A \), where \( p \) is some positive real number (which is strictly less than 1); i.e. \( f_3(A) = pn^2 \).

To upper bound \( p \), we will require the following variant of the Balog-Szemerédi-Gowers theorem (see [4], Proposition 7.3, p. 503) or [2, Section 5.1]).

**Theorem 2.1.** If \( A \) and \( B \) are sets of \( n \) integers and \( G \) is a bipartite graph between \( A \) and \( B \) with \( pn^2 \) edges such that the partial sumset \( A + G B \) has size at most \( K|A| \), then there is a subset \( A' \) of \( A \) with \( |A'| \geq pn/4 \) and

\[
|A' - A'| \ll K^4 p^{-5} n.
\]

Here \( A+G B \) denotes as usual the sumset restricted to the edges coming from \( G \), namely

\[
A+G B = \{ a + b : a \in A, b \in B, (a, b) \in E(G) \}.
\]

It is perhaps important to mention that Theorem 2.1 is a somewhat nonstandard version of the Balog-Szemerédi-Gowers theorem, which outputs directly a large set \( A' \subset A \) with small difference set, without applying any Ruzsa-type inequality. One can derive this version from the following lemma.

**Lemma 2.2.** If a bipartite graph \( G = (A, B, E) \) with \( |A| = |B| = n \) has \( pn^2 \) edges with \( p \geq 30n^{-1/2} \), then there is a subset \( A' \) of \( A \) of size at least \( pn/4 \) such that every pair of vertices in \( A' \) have at least \( \Omega(p^5 n^3) \) paths of length four connecting them.

For the sake of completeness, we include a quick proof Lemma 2.2. We apply Lemma 5.1 from [2] with \( e = 1/10 \) and \( c = p \) to obtain a subset \( U \) of \( A \) with \( |U| \geq pn/2 \) such that at least a .9 fraction of pairs of vertices in \( U \) have at least \( p^2 n/20 \) common neighbors. Consider the auxiliary graph \( F \) on \( U \) where two vertices are adjacent if in the original graph they have less than \( p^2 n/20 \) common neighbors. By construction, the average degree in \( F \) is at most \( .1(|U|−1) \), so there are at most \( |U|/2 \) vertices of degree at most twice the average degree, which is at most \( .2(|U|−1) \). Let \( A' \) be the \( |U|/2 \) vertices of minimum degree in \( F \). Then for every two vertices \( a, a' \) in \( A' \), their number of common neighbors \( d'' \) in the complement of \( F \) is at least \( |U|/2 − 4(|U|−1) \geq |U|/2 \). Then any choice of this common neighbor \( d'' \) in \( F \) can be used as the middle vertex of at least \( (p^2 n/20)(p^2 n/20−1) \) paths of length four between \( a \) and \( a' \) in the original graph, giving a total of at least \( \frac{|U|}{2} (p^2 n/20)(p^2 n/20−1) \geq p^5 n^3/2000 \) paths of length three between \( a \) and \( a' \).

Using Lemma 2.2, one can then deduce Theorem 2.1 in the usual way. First observe that we may assume \( p \geq 30n^{-1/2} \) since otherwise Theorem 2.1 is trivial taking \( A' = A \). Applied to the graph from the setup of Theorem 2.1, Lemma 2.2 produces \( A' \subset A \) of size at least \( pn/4 \) such that every pair of vertices in \( A \) have at least \( \Omega(p^5 n^3) \) paths of length four connecting them. This set happens to also satisfy \( |A' - A'| \ll K^4 p^{-5} n \). Indeed, for each \( a, a' \in A' \), consider a path of length four in \( G \) between them, say \( (a, b, a'', b', a') \). For \( y := a - a' \in A' - A' \), we can then write

\[
a - a' = (a + b) - (a'' + b) + (a'' + b') - (a' + b') = x_1 - x_2 + x_3 - x_4,
\]

where \( x_1 = a + b, x_2 = a'' + b, x_3 = a'' + b', \) and \( x_4 = a' + b' \) are all elements of \( A + G B \). Since for every \( a, a' \in A' \) there are at least \( \Omega(p^5 n^3) \) paths of length four between \( a \) and \( a' \), this means every \( y \in A' - A' \) can be written as \( x_1 - x_2 + x_3 - x_4 \) for at least \( \Omega(p^5 n^3) \) quadruples \((x_1, x_2, x_3, x_4) \in (A + G B)^4 \). However, \( |A + G B| \leq Kn \) holds by assumption, so there are at most \( K^4 n^4 \) such quadruples. By the pigeonhole
principle, it then follows that the number of distinct elements \( y \in A' - A' \) is at most \( O(K^4 p^{-5} n) \), as claimed.

Returning to the task of deriving the upper bound in Theorem 1.2, we apply Theorem 2.1 to the graph \( G \) where \( A \) and \( B \) are chosen to be two copies of our \( k \)-AP free \( A \) and with an edge between \( (a, b) \in A \times A \) if \( a + b = 2c \) for some \( c \in A \). This graph has precisely \( pn^2 \) edges and we can apply Theorem 2.1 to it with \( K = 1 \) since

\[
|A + _G A| = |\{2a : a \in A\}| = |A|.
\]

This yields a subset \( A' \subseteq A \) with \( |A'| \geq p|A|/4 \) and \( |A' - A'| \ll p^{-5} n \ll p^{-6} |A'| \). At this point, we recall a version of the so-called Freiman-Ruzsa modeling lemma (see for instance [9], Theorem 2.3.5, p. 127).

**Lemma 2.3.** Let \( S \) be a finite set of integers and let \( r \geq 2 \) be an arbitrary integer. Then, there is a set \( S'' \subseteq S \) with \( |S''| \geq |S|/r^2 \) which is Freiman \( r \)-isomorphic to a set of integers \( T \) such that

\[
T \subseteq \left\{1, 2, \ldots, \left[ \frac{1}{r} \cdot |rS - rS| \right] \right\}.
\]

Here \( rS - rS \) denotes the sumset \( S + \cdots + S - S - \cdots - S \), where \( S \) appears \( 2r \) times. For the reader’s convenience, we also recall that for any two commutative groups \( G_1 \), \( G_2 \) two sets \( S \subseteq G_1 \) and \( T \subseteq G_2 \) are said to be Freiman \( r \)-isomorphic if there exists a one to one map \( \phi : S \to T \) such that for every \( x_1, \ldots, x_r, y_1, \ldots, y_r \) in \( S \) (not necessarily distinct) the equation

\[
x_1 + \cdots + x_r = y_1 + \cdots + y_r
\]

holds if and only if

\[
\phi(x_1) + \cdots + \phi(x_r) = \phi(y_1) + \cdots + \phi(y_r).
\]

We combine Lemma 2.3 with (a consequence of) the classical Plünnecke-Ruzsa inequality, for which a simple proof can be found in [7].

**Lemma 2.4.** Let \( S \) and \( T \) be finite sets of reals such that \( |S + T| \leq \alpha |S| \), and let \( r, r' \) be positive integers. Then

\[
|rT - r'T| \leq \alpha^{r + r'} |S|.
\]

Indeed, if we apply this with \( S = A' \), \( T = -A' \), \( r = r' = 2 \), and \( \alpha = p^{-6} \), we have

\[
|2A' - 2A'| \leq p^{-24} |A'| \leq p^{-24} n.
\]

Therefore, by Lemma 2.3, there is a subset \( A^+ \subseteq A' \) with \( |A^+| \gg pn \) which is Freiman 2-isomorphic to a set of integers \( \phi(A^+) \) contained in the interval \( \{1, \ldots, \lceil p^{-24} n \rceil \} \). In particular, since \( \phi \) preserves \( k \)-term arithmetic progressions,

\[
|\phi(A^+)| \leq |A^+| = r_k(\lceil p^{-24} n \rceil).
\]

Lastly, recall that \( r_k(n) \) is subadditive as a function of \( n \), namely the inequality \( r_k(n + n') \leq r_k(n) + r_k(n') \) holds for all positive integers \( n, n' \). In particular, \( r_k(\lceil p^{-24} n \rceil) \ll p^{-24} r_k(n) \), hence \( pn \ll p^{-24} r_k(n) \), or
equivalently $p^{25} \ll r_k(n)/n$. This means that $A$ contains at most $O((r_k(n)/n)^{1/25} n^2)$ three-term arithmetic progressions. This completes the proof of the upper bound.

We next prove the desired lower bound on $f_{s,k}(n)$ in Theorem 1.2. We begin by revisiting some further simple properties of $r_k(n)$ as a function of $n$. In addition to being subadditive, we also recall that $r_k(n)$ is an increasing function, so $r_k(m) \leq r_k(n)$ if $m \leq n$. Together these imply that if $n \geq m$, we have $r_k(n) \leq \lceil \frac{n}{m} \rceil r_k(m)$, so

$$\frac{r_k(n)}{2n} \leq \frac{r_k(m)}{m}. \quad (2.1)$$

For all positive integers $m$ and $n$, we have

$$r_k(2mn) \geq r_k(m)r_k(n). \quad (2.2)$$

Indeed, if $U$ is a subset of $\{1, \ldots, m\}$ without a $k$-term arithmetic progression and $V$ is a subset of $\{1, \ldots, n\}$ without a $k$-term arithmetic progression, then the set

$$W = \{2u(n-1) + v : u \in U, v \in V\}$$

is a $k$-AP free subset of $\{1, \ldots, 2mn\}$ of size $|U||V|$, so (2.2) follows.

In particular, if $n \geq N^{1/2}$, letting $m = \lceil \frac{N}{2n} \rceil$, we have

$$r_k(N) \geq r_k(2mn) \geq r_k(n)r_k(m) \geq r_k(n)\frac{m}{2n}r_k(n) \geq \frac{N}{8} \left( \frac{r_k(n)}{n} \right)^2,$$

where the first inequality follows from $r_k(n)$ being an increasing function, the second inequality is by (2.2), the third inequality is by (2.1) using $n \geq m$, and finally the fourth inequality is by substituting in $n \leq 4mN$. It thus follows that

$$\frac{r_k(N)}{N} \geq \frac{1}{8} \left( \frac{r_k(n)}{n} \right)^2. \quad (2.3)$$

Let $N = N_{n,k,s}$ be the least positive integer such that $r_k(N) = \lceil n/s \rceil$. Such an $N$ exists since, for every $m$, $r_k(m+1) = r_k(m)$ or $r_k(m+1) = r_k(m) + 1$ and $\lim_{m \to \infty} r_k(m) = \infty$. We will show that for $k > s \geq 3$ and $n$ sufficiently large in terms of $k$, we have

$$f_{s,k}(n) \geq \left( \frac{n}{300sN} \right)^{s-2} n^2. \quad (2.4)$$

For $n$ sufficiently large in terms of $k$, we have $n \geq N^{1/2}$ holds (for instance by (1.2)), so (2.3) implies that $n/N \geq s \cdot r_k(N)/N \geq s \cdot (1/8) \cdot (r_k(n)/n)^2$, and hence the lower bound from Theorem 1.2 follows from (2.4). We next prove (2.4) using a probabilistic construction of a $k$-AP free set $A$ of $n$ integers with many $s$-term arithmetic progressions.

For each $1 \leq i \leq s$, let $d_i$ be an integer chosen uniformly and independently at random from the set $\{1, \ldots, 2N\}$. Let $S \subset \{1, \ldots, N\}$ be a $k$-AP free set of cardinality $r_k(N) = \lceil n/s \rceil$, and $S_i$ denote the translate $\{x + 6(i-1)N - 1 + d_i : x \in S\}$, i.e. $S_i := S + (6(i-1)N - 1 + d_i)$.

Finally, let us consider the set $A \subset \{1, \ldots, 6sN\}$ defined by

$$A := \bigcup_{i=1}^{s} S_i.$$
We first check that such a (random) set must be \( k \)-AP free. Indeed, the sets \( S_1, \ldots, S_s \) are pairwise disjoint since, for each \( 1 \leq i \leq s \), we have

\[
S_i \subset \{6(i-1)N+1, \ldots, 6(i-1)N+3N-1\}.
\]

Furthermore, these sets are spaced out so that if an arithmetic progression contains an element from \( S_i \) and an element of \( S_j \) with \( i \neq j \), then its common difference is at least \( 3N+2 \), in which case the arithmetic progression cannot contain two elements in the same \( S_i \). In particular, every arithmetic progressions in \( A \) of length longer than \( s \) must be a subset of one of the \( S_i \), and hence \( A \) is \( k \)-AP free.

Finally, \( |A| = s|S| = s\binom{N}{2} \leq n \), so \( A \) is indeed in \( A_k(n) \), or it can be artificially augmented to a set in \( A_k(n) \) by adding some elements that do not create \( k \)-term arithmetic progressions.

We next lower bound the expected number of \( s \)-term arithmetic progressions in \( A \). The number of \( s \)-term arithmetic progressions \( a, a+D, \ldots, a+(s-1)D \) with \( a+(i-1)D \in \{6(i-1)N+N+1, \ldots, 6(i-1)N+2N\} \) for \( 1 \leq i \leq s \) is the same as the number of \( s \)-term arithmetic progressions in \( \{1, \ldots, N\} \) with any integer common difference, which is

\[
N + 2 \sum_{a=1}^{N-1} \left\lfloor \frac{N-a}{s} \right\rfloor \geq \frac{1}{s} \binom{N}{2}.
\]

For each such \( s \)-term arithmetic progression \( a, a+D, \ldots, a+(s-1)D \) and for each sequence \( (a_1, \ldots, a_s) \) of \( s \) elements from \( S \), there is a choice of \( d_1, \ldots, d_s \in \{1, \ldots, 2N\} \) such that \( a + 6(i-1)N-1 + d_i = a + (i-1)D \) for \( 1 \leq i \leq s \). Hence, the expected number of \( s \)-term arithmetic progressions in \( A \) is at least

\[
\frac{1}{s} \binom{N}{2} |S|^{s}(2N)^{-s} \geq \frac{1}{4s} N^2 \left( \frac{|S|}{2N} \right)^s \geq \left( \frac{n}{300sN} \right)^{s-2} n^2.
\]

Thus, there must exist a choice of such an \( A \) for which the number of \( s \)-term arithmetic progressions is at least this lower bound on the expected number, which completes the proof of (2.4) and hence Theorem 1.2.

**ACKNOWLEDGMENTS**

The first author was supported by a Packard Fellowship and by NSF grant DMS-1855635. The second author would like to thank Oliver Roche-Newton for helpful conversations.

**REFERENCES**

1. G. J. Simmons and H. L. Abbott, *How many 3-term arithmetic progressions can there be if there are no long ones?* Amer. Math. Monthly 84 (1977), 633–635.
2. J. Fox and B. Sudakov, *Dependent random choice*, Random Structures Algorithms 38 (2011), 68–99.
3. P. Erdős, *Problems and results on combinatorial number theory*, in *A survey of Combinatorial Theory* (Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., 1971), North-Holland, Amsterdam, 1973, 117–138.
4. W. T. Gowers, *A new proof of Szemerédi’s theorem for progressions of length four*, Geom. Func. Anal. 8 (1998), 529–551.
5. W. T. Gowers, *A new proof of Szemerédi’s theorem*, Geom. Func. Anal. 11 (2001), 465–588.
6. B. Green and T. Tao, *New bounds for Szemerédi’s theorem, III: A polylogarithmic bound for \( r_4(N) \)*, Mathematika 63 (2017), 944–1040.
7. G. Petridis, *New proofs of Plünnecke-type estimates for product sets in groups*, Combinatorica 32 (2012), 721–733.
8. R. A. Rankin, *Sets of integers containing not more than a given number of terms in arithmetical progression*, Proc. Roy. Soc. Edinburgh Sect. A **65** (1961), 332–344.

9. I. Z. Ruzsa, *Sumsets and structure*, in *Combinatorial number theory and additive group theory*, Adv. Courses Math. CRM Barcelona, Birkhäuser Verlag, Basel, 2009, 87–210.

10. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

**How to cite this article:** Fox J, Pohoata C. Sets without $k$-term progressions can have many shorter progressions. *Random Struct Alg.* 2021;58:383–389. https://doi.org/10.1002/rsa.20984