A PROBABILISTIC THRESHOLD FOR MONOCHROMATIC ARITHMETIC PROGRESSIONS

Aaron Robertson
Department of Mathematics, Colgate University, Hamilton, New York
arobertson@colgate.edu

Abstract
We show that $\sqrt{k^{2k^2}}$ is, roughly, the threshold where, under mild conditions, on one side almost every coloring contains a monochromatic $k$-term arithmetic progression, while on the other side, there are almost no such colorings.

1. Introduction

For $k \in \mathbb{Z}^+$, let $w(k)$ be the minimum integer such that every 2-coloring of $[1, w(k)]$ admits a monochromatic $k$-term arithmetic progression. The existence of such an integer was shown by van der Waerden [4], and these integers are referred to as van der Waerden numbers. Current knowledge places $w(k)$ somewhere between $(k - 1)^{2k^2} - 1$ (for $k - 1$ prime) and $2^{2^{2^{2(k+9)}}}$, with the upper bound being from one of Gowers’ seminal work [3]. A matching of upper and lower bounds appears unlikely in the near (or distant?) future. However, by loosening the restriction that every 2-coloring must have a certain property to almost every (in a probabilistic sense), we are able to home in on the rate of growth of the associated numbers.

In this article, we assume that every 2-coloring of a given interval is equally likely. We refer to a $k$-term arithmetic progression as a $k$-ap and will use the notation $\langle a, d \rangle_k$ to represent $a, a+d, a+2d, \ldots, a+(k-1)d$, where we refer to $d$ as the gap of the $k$-ap. We will use the notation $[1,n] = \{1,2,\ldots,n\}$.

Definition 1. Let $t(k)$ be a function defined on $\mathbb{Z}^+$ with some property $\mathcal{P}$. We say that $t(k)$ is a minimal function (with respect to $\mathcal{P}$) if for every function $s(k)$ defined on $\mathbb{Z}^+$ with property $\mathcal{P}$ we have

$$\lim \inf \frac{t(k)}{s(k)} \leq 1.$$

We say that $t(k)$ is a maximal function (with respect to $\mathcal{P}$) if for every function $s(k)$ defined on $\mathbb{Z}^+$ with property $\mathcal{P}$ we have

$$\lim \sup \frac{t(k)}{s(k)} \geq 1.$$
Definition 2. Let \( N^+(k) \) be a minimal function such that the probability that a randomly chosen 2-coloring of \([1, N^+(k)]\) admits a monochromatic \(k\)-ap tends to 1 as \(k \to \infty\). Let \( N^-(k) \) be a maximal function such that the probability that randomly chosen 2-coloring of \([1, N^-(k)]\) admits a monochromatic \(k\)-ap tends to 0 as \(k \to \infty\).

Brown [1] showed that \( N^+(k) \leq (\log k)2^k g(k) \), while Vijay [5] made a significant improvement by showing that \( N^+(k) \leq k^{3/2}2^{k/2} g(k) \), where \(g(k)\) is any function tending to \(\infty\). Vijay, using the linearity of expectation, also provided a lower bound for \( N^-(k) \) that is not much smaller than his given upper bound:

**Theorem 3.** (Vijay) Let \( f(k) \to 0 \) arbitrarily slowly. Then \( N^-(k) \geq \sqrt{k} 2^{k/2} f(k) \).

We note here that if we consider the set of \(k\)-aps with gaps that are primes larger than \(k\), we can follow Vijay’s argument for an upper bound on \( N^+(k) \) very closely to show that \( N^+(k) \leq k^2 \sqrt{\log k} 2^{k/2} g(k) \) for any function \(g(k) \to \infty\) (we will use the notation \(g(k) \to \infty\) as opposed to \(g(k) \to 0\) as found in [5]). In the next section, we construct a larger family of \(k\)-aps (than Vijay’s and than those \(k\)-aps with prime gap larger than \(k\)) with the aim of lowering this upper bound by a factor of \(\sqrt{k}\).

2. A Structured Family of Arithmetic Progressions

For \(k, n \in \mathbb{Z}^+\), let \(AP_k(n) = \{(a, d) \in [1, n] : a, d \in \mathbb{Z}^+\} \), i.e., the set of \(k\)-aps in \([1, n]\) and denote by \(A_j(n)\) those elements of \(AP_k(n)\) with \(d = j\), so that \(AP_k(n)\) is the disjoint union of the \(A_j(n)\):

\[
AP_k(n) = \bigsqcup_{d=1}^{n-1} A_d(n).
\]

We will now sieve out elements from each \(A_d(n)\). Since each \(A_d(n)\) is in one-to-one correspondence (via their initial terms) with \([1, n-(k-1)d]\), start with the sequence \(1, 2, \ldots, n-(k-1)d\). Proceeding from \(i = 1\) we sieve by: (1) for integer \(i\) remove \(i + d\) and \(i + 2d\); (2) move to the least integer greater than \(i\); (3) repeat. For example, for \(A_4(n)\), the first few steps are:

\[
\begin{align*}
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, \ldots \\
1, 2, 3, 4, 6, 7, 8, 10, 11, 12, 13, \ldots \\
1, 2, 3, 4, 7, 8, 11, 12, 13, \ldots \\
1, 2, 3, 4, 8, 12, 13, \ldots \\
1, 2, 3, 4, 13, \ldots .
\end{align*}
\]

Denote the set of elements of \(A_d(n)\) that remain after sieving by \(\hat{A}_d(n)\). (Clearly, there are easier ways to describe this set; however, the stated description is given for clarity in proving the main result.) It is easy to check that the following is true.

**Lemma 4.** For any \(d \in \mathbb{Z}^+\), we have \(|\hat{A}_d(n)| \geq \frac{d}{kd} |A_d(n)| = \frac{1}{d} |A_d(n)|\).
The family in which we are interested is

\[ \mathcal{AP}_k(n) = \bigcup_{d=1}^{\frac{n-1}{k-1}} \mathcal{A}_{d}(n). \]  

(1)

We will use the following lemmas.

**Lemma 5.** For \( n > k > 1 \), we have \( |\mathcal{AP}_k(n)| \geq \frac{n^2}{6(k-1)}(1 + o(1)) \).

**Proof.** It is a standard exercise to show that \( |\mathcal{AP}_k(n)| = \frac{n^2}{2(k-1)}(1 + o(1)) \). Coupling this with Lemma 4 and Equation 1 gives the stated bound. \( \square \)

**Lemma 6.** Let \( A = \langle a, b \rangle_k \) and \( C = \langle c, d \rangle_k \) belong to \( \mathcal{AP}_k(n) \). Then \( |A \cap C| \leq k - 3 \). Furthermore,

(i) \( |A \cap C| > \left\lceil \frac{k}{2} \right\rceil \) only if \( b = d \)

(ii) \( \left\lceil \frac{k}{3} \right\rceil \leq |A \cap C| \leq \left\lceil \frac{k}{2} \right\rceil \) only if \( \frac{d}{3} \in \left\{ \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{2}, 1, \frac{3}{2}, 2, 3 \right\} \).

**Proof.** We first argue that in order to have \( |A \cap C| \geq \left\lceil \frac{k}{2} \right\rceil \) we must have \( \frac{b}{3} \in \left\{ \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{2}, 1, \frac{3}{2}, 2, 3 \right\} \). Consider \( b \leq d \). We have \( a + ib = c + j_1d \) and \( a + (i + x)b = c + j_2d \) for some \( i \in [0, k - 2] \) and \( x \in \{1, 2, 3\} \) (else we cannot have enough intersections) and \( j_1 < j_2 \). Thus, \( xb = (j_2 - j_1)d \). Since \( d \geq b \) we must have \( j_2 - j_1 \leq x \). This leaves \( \frac{b}{3} = \frac{j_2 - j_1}{x} \in \left\{ \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1 \right\} \). For \( d > b \), we take reciprocals and achieve the stated goal.

Now, in order for \( A \) and \( C \) to intersect in more than \( \left\lceil \frac{k}{2} \right\rceil \) places, there must be two consecutive elements of, say, \( A \) in the intersection. Let \( a + ib \) and \( a + (i + 1)b \) be two such elements. We must have \( d \leq b \) in order for \( C \) to intersect both of these. So, let \( a + ib = c + j_1d \) and \( a + (i + 1)b = c + j_2d \). These imply that \( b = (\ell - j)d \). If \( \ell - j > 1 \) then \( d \leq \frac{k}{2} \). In this situation, \( C \) intersects \( A \) in at most \( \left\lceil \frac{k}{2} \right\rceil \) places since for every two consecutive terms of \( A \), there exists a term of \( C \) between them. Thus, \( \ell - j = 1 \) and \( b = d \) as stated.

To show that \( |A \cap C| \leq k - 3 \), note that we have proved parts (i) and (ii) so we need only consider \( k \)-aps with the same gap, i.e., those in the same \( \mathcal{A}_g(n) \) for some gap \( g \). In order for two such \( k \)-aps to intersect in more than \( k - 3 \) places, their starting elements must be within \( 2g \) of each other. But by construction of \( \mathcal{A}_g(n) \), this is not possible. \( \square \)

**Lemma 7.** For a given \( A = \langle a, b \rangle_k \in \mathcal{AP}_k(n) \), the number of \( \langle c, d \rangle_k \in \mathcal{AP}_k(n) \) with \( c \geq a \) that intersect \( A \) in \( p \) places is

(i) 0 for \( p > k - 3 \);

(ii) 1 for each \( p \in \left[ \left\lceil \frac{k}{2} \right\rceil + 1, k - 3 \right] \);

(iii) at most 7 for each \( p \in \left[ \left\lceil \frac{k}{3} \right\rceil, \left\lceil \frac{k}{2} \right\rceil \right] \).
Proof. Part (i) is just a restatement of part of Lemma 4. For part (ii), by Lemma 4(i), we must have \( b = d \). For a given \( p \), we have \( c = a + (k - p)d \) and the result follows. For part (iii), since \( b \) is fixed, \( d \) must be one of 7 gaps that adhere to Lemma 4(ii). In order to intersect in exactly \( p \) places, \( c \) is determined. \( \square \)

With these lemmas under our belt, we are now ready to move onto the main result.

3. The Result

We incorporate Theorem 3 into the main result, which we now state.

**Theorem 8.** Let \( f(k) \to 0 \) and \( g(k) \to \infty \) arbitrarily slowly. Then,

\[
\sqrt{k^{2/2} f(k)} \leq N^-(k) < N^+(k) \leq \sqrt{k^{2/2} g(k)}.
\]

**Proof.** We only need to prove the upper bound on \( N^+(k) \) and do so by using the family defined in the previous section, along with techniques from [1, 5]. To this end, let \( n = \sqrt{k^{2/2} g(k)} \) and partition \([1, n]\) into intervals of length \( s = \left\lceil \frac{n}{g(k)^{2/3}} \right\rceil \), where the last interval may be shorter. For any of these subintervals, enumerate the \( k \)-aps from \( \hat{AP}_k(s) \) in the subinterval and let \( X_i \) be the event that the \( i \)th \( k \)-ap, \( 1 \leq i \leq \frac{s^2}{6(k-1)} \) (we suppress lower order terms), is monochromatic under a given 2-coloring. We let \( p \) be the probability that a random 2-coloring of a given interval of length \( s \) admits a monochromatic \( k \)-ap, where each integer is equally likely to be either of 2 colors. Via one of the Bonferroni inequality, we have

\[
p = \left| P \left( \bigcup_{i=1}^{s^{2/6(k-1)}} X_i \right) \right| \geq \sum_{i=1}^{s^{2/6(k-1)}} P(X_i) - \sum_{1 \leq i < j \leq s^{2/6(k-1)}} P(X_i \cap X_j).
\]

Hence,

\[
p \geq \frac{s^2}{6(k-1)} \cdot \frac{1}{2^{k-1}} - \sum_{1 \leq i < j \leq s^{2/6(k-1)}} P(X_i \cap X_j).
\]

We now focus on the double summation. With a slight abuse of notation, we rewrite this as

\[
\sum_{b \in \hat{AP}_k(s)} \sum_{a \in \hat{AP}_k(s)} P(X_a \cap X_b),
\]

where the initial term of \( b \) is at most as large as the initial term of \( a \). For a given \( b \in \hat{AP}_k(s) \) with gap \( g \), we define \( S_b \) to be those \( k \)-aps in \( \hat{AP}_k(s) \) with gap \( g \) that intersect \( b \); \( T_b \) to be those that intersect \( b \) in at least \( \left\lceil \frac{h}{2} \right\rceil \) places but at most \( \left\lceil \frac{h}{3} \right\rceil \) places and have a gap \( h \) such that \( \frac{b}{h} \in \{ \frac{1}{3}, \frac{1}{4}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 2, 3 \} \) (note that we don’t consider \( g = h \) here as they are accounted for in \( S_b \)); \( Q_b \) to be those that intersect \( b \)
but are not in $S_b \sqcup T_b$; and $R_b$ to be those that do not intersect $b$. With these definitions, the double summation becomes

$$\sum_{b \in \hat{AP}_k(s)} \left( \sum_{a \in S_b} P(X_a \cap X_b) + \sum_{a \in T_b} P(X_a \cap X_b) + \sum_{a \in Q_b} P(X_a \cap X_b) + \sum_{a \in R_b} P(X_a)P(X_b) \right).$$

Appealing to Lemmas 4 and 5, we find the following (for $k$ sufficiently large):

$$\sum_{a \in S_b} P(X_a \cap X_b) \leq \sum_{i=2}^{k-1} \frac{1}{2^{k+1}} = \frac{1}{2^{k+1}} - \frac{1}{2^{2k-1}} \leq \frac{1}{2^{k+1}}; \quad (2)$$

$$\sum_{a \in T_b} P(X_a \cap X_b) \leq \frac{7 \left( \left\lceil \frac{k}{2} \right\rceil - \left\lceil \frac{k}{3} \right\rceil + 1 \right)}{2^{k+1/2}} \leq \frac{1}{3 \cdot 2^{k+1}}; \quad (3)$$

$$\sum_{a \in Q_b} P(X_a \cap X_b) \leq \frac{3sk}{2^{k+2k/3}} \leq \frac{1}{3 \cdot 2^{k+1}}; \quad (4)$$

$$\sum_{a \in R_b} P(X_a)P(X_b) \leq \frac{s^2}{6(k-1)2^{k-1}} \leq \frac{1}{3 \cdot 2^{k+1}}; \quad (5)$$

where: (2) holds since there is exactly 1 such $k$-ap that intersects $b$ in $k - 1 - i$ places and Lemma 5(i) gives $i \geq 2$; (3) follows from Lemma 4(ii) and Lemma 5(iii); (4) holds from the lower bound in Lemma 4(ii) and since $3sk$ or fewer $k$-aps intersect a given $k$-ap (this is a standard bound typically used with the Lovasz Local Lemma; see, e.g., [2]); and (5) holds by independence since the two $k$-aps do not share any element.

Using these bounds, we have

$$\sum_{a \in S_b} P(X_a \cap X_b) + \sum_{a \in T_b} P(X_a \cap X_b) + \sum_{a \in Q_b} P(X_a \cap X_b) + \sum_{a \in R_b} P(X_a)P(X_b) \leq \frac{2}{2^{k+1}} = \frac{1}{2^k}.$$

Hence,

$$\sum_{b \in \hat{AP}_k(s)} \sum_{a \in \hat{AP}_k(s)} P(X_a \cap X_b) \leq \frac{s^2}{6(k-1)} \cdot \frac{1}{2^k}$$

so that

$$p \geq \frac{s^2}{6(k-1)} \cdot \frac{1}{2^{k-1}} - \frac{s^2}{6(k-1)} \cdot \frac{1}{2^k} = \frac{s^2}{6(k-1)} \cdot \frac{1}{2^k}.$$

This gives us that the probability that a given interval of length $s$ has no monochromatic $k$-ap from $\hat{AP}_k(s)$ is at most

$$1 - \frac{s^2}{6(k-1)2^k}.$$
Thus, the probability that $[1,n]$ has no monochromatic $k$-ap from $\hat{AP}_k(n)$ is, for $k$ sufficiently large, at most

$$(1 - \frac{s^2}{6(k-1)^2})^{g^{4/3}(k)} \approx \exp \left( -\frac{g^{4/3}(k)s^2}{6(k-1)^2} \right) = \exp \left( -\frac{n^2}{g^{4/3}(k)6(k-1)^2} \right)$$

$$= \exp \left( -\frac{k}{6(k-1)}g^{2/3}(k) \right) \to 0 \quad \text{as} \quad k \to \infty.$$ 

Since the probability that $[1,n]$ contains a monochromatic $k$-ap from $\hat{AP}_k(n)$ tends to 1, the same holds for the full family of $k$-aps in $[1,n]$. \hfill \Box

References

[1] Tom Brown, A pseudo upper bound for the van der Waerden function, *J. Combin. Theory Ser. A* 87 (1999), 233-238.

[2] Tom Brown, Bruce Landman, and Aaron Robertson, Bounds on some van der Waerden numbers *J. Combin. Theory Ser. A* 115, 1304-1309.

[3] W. T. Gowers, A new proof of Szemerédi’s theorem, *Geom. Funct. Anal.* 11 (2001), no. 3, 465-588.

[4] B. L. Van der Waerden, Beweis einer baudetschen Vermutung, *Nieuw Archief voor Wiskunde* 15 (1927), 212-216.

[5] Sujith Vijay, Monochromatic progressions in random colorings, *J. Combin. Theory Ser. A* 119 (2012), 1078-1080.