FINDING CERTAIN ARITHMETIC PROGRESSIONS IN 2-COLOURED CYCLIC GROUPS

MATEI MANDACHE*

Received July 28, 2020
Revised December 1, 2021
Online First October 10, 2022

We say a pair of integers \( (a,b) \) is findable if the following is true. For any \( \delta > 0 \) there exists a \( p_0 \) such that for any prime \( p \geq p_0 \) and any red-blue colouring of \( \mathbb{Z}/p\mathbb{Z} \) in which each colour has density at least \( \delta \), we can find an arithmetic progression of length \( a+b \) inside \( \mathbb{Z}/p\mathbb{Z} \) whose first \( a \) elements are red and whose last \( b \) elements are blue.

Szemerédi’s Theorem on arithmetic progressions implies that \( (0,k) \) and \( (1,k) \) are findable for any \( k \). We prove that \( (2,k) \) is also findable for any \( k \). However, the same is not true of \( (3,k) \). Indeed, we give a construction showing that \( (3,30000) \) is not findable. We also show that \( (14,14) \) is not findable.

1. Introduction

In 1975, Szemerédi [6] proved the following famous theorem about arithmetic progressions:

**Theorem 1.1.** Let \( k \) be an integer and let \( \delta > 0 \). There exists an \( N_0 = N_0(k,\delta) \) such that if \( N \geq N_0 \) and \( A \subseteq \{1,2,\ldots,N\} \) with \( |A| \geq \delta N \), then \( A \) must contain a non-trivial arithmetic progression of length \( k \).

We use the notation \( \{N\} \) to denote the set \( \{1,2,\ldots,N\} \). Several generalisations of Theorem 1.1 have since been proven (see, for example, [2] and [5]). In the current paper, we consider another way of trying to generalise Theorem 1.1. If we consider the set \( \{N\} \) as being coloured red and blue, with

---

* Work completed while the author was at the Mathematical Institute, University of Oxford.
the set $A$ being red and $[N] \setminus A$ being blue, then Theorem 1.1 tells us we can find red arithmetic progressions (under suitable size conditions). But what about arithmetic progressions which mix the two colours? For example, what if we want to find an arithmetic progression of length 4 which is coloured (red, red, blue, blue)?

When considering this question, it is natural to work in cyclic groups modulo a prime, rather than in the set $[N]$, or in arbitrary cyclic groups. This is because in the latter two cases we may have some simple obstructions to finding such patterns. For example, one could colour all the even numbers red and all the odd numbers blue. These obstructions do not happen in $\mathbb{Z}/p\mathbb{Z}$, making the question more interesting.

For any integers $a, b \geq 0$, we define an $(a, b)$-arithmetic progression, or $(a, b)$-AP for short, to be an arithmetic progression of length $a + b$ in which the first $a$ elements are red and the last $b$ elements are blue. Given a pair of non-negative integers $(a, b)$, we say that $(a, b)$ is findable if for all $\delta > 0$ there is a $p_0$ such that for any prime $p \geq p_0$, any red-blue colouring of $\mathbb{Z}/p\mathbb{Z}$ in which each colour class has size at least $\delta p$ contains an $(a, b)$-AP with common difference $d \neq 0$. The question of whether $(k, k)$ is findable was posed by Alon, Fox and Zhao, it is Question 5.5 of [1].

We conclude the introduction by making some basic remarks about which pairs are findable. Findability is symmetric and monotone, in that if $(a, b)$ is findable, then so are $(b, a)$ and $(a', b')$ for any $a' \leq a$ and $b' \leq b$. Theorem 1.1 still holds if we replace $[N]$ with $\mathbb{Z}/p\mathbb{Z}$ (and have a size condition on $p$), so $(k, 0)$ is findable for all $k$. In fact, Theorem 1.1 also implies that $(k, 1)$ is findable for all $k$. Once we have found a red arithmetic progression of length $k$ in $\mathbb{Z}/p\mathbb{Z}$, say $\{a, a + d, \ldots, a + (k - 1)d\}$, let $t$ be the minimal non-negative integer such that $a + td$ is blue. This must exist, because since $d$ and $p$ are coprime, $a + td$ ranges over the whole of $\mathbb{Z}/p\mathbb{Z}$, and $\mathbb{Z}/p\mathbb{Z}$ is not all red. Also, $t \geq k$. So $\{a + (t - k)d, a + (t + 1 - k)d, \ldots, a + (t - 1)d, a + td\}$ is a $(k, 1)$-AP.

This prompts several questions: Is it true that $(a, b)$ is findable for any $a$ and $b$? For which $a$ is $(a, k)$ findable for all $k$? What is the set of findable pairs $(a, b)$? The aim of this paper is to make progress on these questions. Not all pairs are findable: in Section 3, we show that $(14, 14)$ is not findable. Our construction showing $(14, 14)$ is not findable is a fractal-type construction, based on replacement rules which are iterated ad infinitum. We note that this construction has some similarities to constructions used to solve problems about pattern avoiding words, such as those in [3] and [4], which are based upon iterating a replacement rule. However, the construction in our case is somewhat different. It turns out that $(a, k)$ is findable for all $k$ if and only if $a \leq 2$: in Section 2, we show that $(2, k)$ is findable for all $k$, and in
Section 4 we show that $(3, 30000)$ is not findable (and hence $(a, 30000)$ is not findable for any $a \geq 3$ by monotonicity). This leaves open the question of findability for some pairs, but there are only finitely many such pairs. Indeed, combining the results of the following sections with symmetry and monotonicity, we have the following “state of affairs”:

$$(a, b) \begin{cases} 
\text{findable} & \text{if } \min(a, b) \leq 2 \\
\text{not findable} & \text{if } \min(a, b) \geq 14 \text{ or } (\min(a, b) \geq 3 \text{ and } \max(a, b) \geq 30000) \\
\text{unknown} & \text{if } 3 \leq \min(a, b) < 14 \text{ and } \max(a, b) < 30000
\end{cases}$$

The author feels it is possible that further progress could be made in proving pairs are not findable by constructing clever fractals similar to those in Sections 3 and 4, whereas progress in proving pairs are findable likely requires a significant new idea. As such, perhaps the most interesting open question is that of whether $(3, 3)$ is findable. Another possible direction for further research would be to consider colourings with more than two colours, which are not covered in this paper.

2. Finding $(2, k)$

Let $k$ be a positive integer. We aim to show $(2, k)$ is findable. Throughout this section, we assume that $\mathbb{Z}/p\mathbb{Z}$ has been red-blue coloured with at least $\delta p$ elements in each colour class and that this colouring contains no $(2, k)$-AP. We aim to reach a contradiction, provided $p$ is sufficiently large.

Our colouring of $\mathbb{Z}/p\mathbb{Z}$ induces a colouring of $\mathbb{Z}$ via the quotient map $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. The induced colouring is periodic mod $p$, and contains no $(2, k)$-AP. For the rest of this section, unless otherwise stated, we will work with the colouring of $\mathbb{Z}$ thus induced. In this section, we use $[x, y]$ to denote the set $\{n \in \mathbb{Z} | x \leq n \leq y\}$ and $d[x, y]$ to denote the set $\{dn | n \in [x, y]\}$ (where $d$ is an integer; note these sets are arithmetic progressions). We define a gap to be an arithmetic progression whose smallest and largest elements are both red, but all other elements are blue.

First, a few words about the strategy of the proof. We can use Szemerédi’s Theorem to find large gaps (i.e. gaps with many elements). It turns out that, once we have found a gap, we can deduce quite a few things about the rest of the colouring by considering various arithmetic progressions of length $k+2$ and using the fact that they are not $(2, k)$-APs. Roughly speaking, our strategy is to show that if $[0, d]$ is a gap, then in fact all reds must occur at multiples of $d$. By combining this with periodicity mod $p$, we reach a
contradiction. Unfortunately, this only works for some values of \(d\), namely when \(d\) is a sufficiently large prime. However, there are other tricks we can use when \(d\) is not prime, to reduce to the case when \(d\) is prime. But before we can get to that we need to do some preparatory work.

We start off by proving the following lemma.

**Lemma 2.1.** Let \(m \geq 1\) and suppose \([0,d]\) is a gap. Then the number of reds in \([0,md]\) is bounded above by a function depending on \(m\) and \(k\) only.

The idea behind the proof is to work incrementally. Given an interval with a bounded number of reds, the fact there is no \((2,k)\)-AP forces all but a bounded number of points to be blue in an appropriate adjoining interval.

**Proof.** Let \(c\) be the least integer such that \(m \leq \left(\frac{k+1}{k}\right)^c\). We will show that the number of reds in \([0,md]\) is at most \(3(k+1)^c - 1\) by induction on \(c\).

In the base case \(c = 0\), we have \(m = 1\) and the result is trivial. For the induction step, \(c \geq 1\) and we may assume that \(m = \left(\frac{k+1}{k}\right)^c\), since this makes \(m\) bigger without increasing \(c\). We know by the induction hypothesis that there are at most \(3(k+1)^{c-1} - 1\) reds in the interval \([0,\left(\frac{k+1}{k}\right)^{c-1} \cdot d]\). Let \(r\) be the least red with \(r > \left(\frac{k+1}{k}\right)^{c-1} d\). If \(r \geq \left(\frac{k+1}{k}\right)^c d\), we are done.

Otherwise, suppose \(s\) is red with \(r < s \leq \left(\frac{k+1}{k}\right)^c d\). We aim to show that there are at most \(3k(k+1)^{c-1} - k\) such reds. Consider the arithmetic progression of length \(k+2\) with first term \(s\) and common difference \(r - s\). The first two terms, \(s\) and \(r\), are both red, and so it must contain another red. We pick a \(t \in \{1, 2, \ldots, k\}\) such that \(r + t(r-s)\) is red. Since \(t(s-r) \leq k \left(\left(\frac{k+1}{k}\right)^c - \left(\frac{k+1}{k}\right)^{c-1} d\right) = \left(\frac{k+1}{k}\right)^{c-1} d \leq r\), we have \(0 \leq r + t(r-s) < r\) so \(r + t(r-s)\) must be one of the reds in \([0,\left(\frac{k+1}{k}\right)^{c-1} \cdot d]\), of which there are at most \(3(k+1)^{c-1} - 1\). There are \(k\) possible values for \(t\) and at most \(3(k+1)^{c-1} - 1\) possible values for \(r + t(r-s)\). Since \(t\) and \(r + t(r-s)\) determine \(s\), this means there are at most \(3k(k+1)^{c-1} - k\) possible values for \(s\).

We now do the final sum, by adding the numbers of reds in the interval which are less than \(r\), equal to \(r\) and greater than \(r\). The total is at most \(3(k+1)^{c-1} - 1 + 1 + 3k(k+1)^{c-1} - k = 3(k+1)^c - k \leq 3(k+1)^c - 1\), as required.

Lemma 2.1 has the following corollary.

**Corollary 2.2.** Let \(c\) be a positive real. There exists an \(N = N(c,k) \geq 1\) such that the following holds: for any integers \(m_1,m_2,n_1,n_2\), if \([m_1,m_2]\) and \([n_1,n_2]\) are gaps with \(n_2 - n_1 \geq N(m_2 - m_1)\), then the distance between the gaps is at least \(c(n_2 - n_1)\).
Proof. Without loss of generality, $m_1 \geq n_2$. By translating, we may also assume that $n_1 = 0$. We prove the contrapositive: assume that the gaps are less than $cn_2$ apart, and aim to prove that $(n_2 - n_1)/(m_2 - m_1)$ is bounded by a function of $c$ and $k$. We have $m_1 \leq (c+1)n_2$. If $m_2 > (c+2)n_2$, then $m_2 - m_1 > n_2 - n_1$ and so $(n_2 - n_1)/(m_2 - m_1) < 1$, so we may assume that $m_2 \leq (c+2)n_2$. We apply Lemma 2.1 with $d = n_2$ and $m = c+2$ to deduce that there are at most $b = b(c,k)$ reds in the interval $[0,m_2]$. Let $r_1 \leq r_2 \leq \cdots \leq r_a$ be these reds. So we have $r_1 = 0, r_2 = n_2, r_{a-1} = m_1, r_a = m_2$ and $a \leq b$. For each $2 \leq i \leq a-1$, we must have $r_i - r_{i-1} \leq k(r_{i+1} - r_i)$, since otherwise the arithmetic progression of length $k+2$ with first term $r_{i+1}$ and common difference $r_i - r_{i+1}$ would be a $(2,k)$-AP. By combining these inequalities for all $i$, we get that $r_1 - r_0 \leq k^{a-1}(r_a - r_{a-1})$, which implies that $n_2 - n_1 \leq k^{b-1}(m_2 - m_1)$. So $(n_2 - n_1)/(m_2 - m_1) \leq k^{b-1}$, as required. 

We define a table to be a set of the form $d[-\ell,m]$ where $d$, $\ell$, and $m$ are non-negative integers, which satisfies the following conditions:

- $-\ell d$ and $m d$ are both red,
- if $d[n_1,n_2] \subset d[-\ell,m]$ is a gap, then $n_2 - n_1 \leq \frac{m + \ell}{k+1}$.

Note that $d[n_1,n_2]$ being a gap is not the same as $[dn_1,dn_2]$ being a gap.

The reason behind this definition is the following: once we have a table $d[-\ell,m]$, we can obtain information about the red points in $[-\ell d,md] \setminus d[-\ell,m]$ by repeatedly considering arithmetic progressions of length $k+2$ whose first element is a red in $d[-\ell,m]$ and whose second element is a red in $[-\ell d,md] \setminus d[-\ell,m]$. This will give us information about the reds in $[-\ell d,md] \setminus d[-\ell,m]$, and in some cases (when $d$ is a sufficiently large prime), we will see that in fact no such reds can exist. However, we first need to establish some results concerning the existence of tables.

Lemma 2.3. Suppose $[0,d]$ is a gap, for some integer $d$. There exists an $N_0 = N_0(k)$ such that if $\ell, m \geq N_0$ and $-\ell d, m d$ are both red, then $d[-\ell,m]$ is a table.

Proof. Let $d[n_1,n_2] \subset d[-\ell,m]$ be a gap. By applying Corollary 2.2 inside the set $dZ$ with $n_1,n_2$ as given, $m_1 = 0$, $m_2 = 1$ and $c = k+1$, we get an $N = N(k)$ such that if $n_2 - n_1 \geq N$, then either $n_2 \leq -(k+1)(n_2 - n_1)$ or $n_1 \geq (k+1)(n_2 - n_1) + 1$. In the former case we get that $\ell \geq -n_2 \geq (k+1)(n_2 - n_1)$, and in the latter case we get that $m \geq n_1 \geq (k+1)(n_2 - n_1)$, so either way $\ell + m \geq (k+1)(n_2 - n_1)$, as required. We take $N_0(k) = (k+1)N(k)$. Then, if $n_2 - n_1 \geq N$ we are done by the above argument, while if $n_2 - n_1 < N$ we have $n_2 - n_1 \leq N_0/(k+1) < (\ell + m)/(k+1)$, so the condition holds either way. So $d[-\ell,m]$ is a table.
A table $d[−\ell,m]$ is called extendable if, for any $\ell′ \geq \ell$ and $m′ \geq m$ with $−\ell′d$ and $md$ both red, $d[−\ell′,m′]$ is also a table. Because we want to extend our reasoning to the whole of $\mathbb{Z}$, we are interested in extendable tables, as opposed to mere tables. Luckily Lemma 2.3 has the following corollary.

**Corollary 2.4.** Suppose $[0,d]$ is a gap, for some integer $d$. There exists an $N_0 = N_0(k)$ such that if $\ell,m \geq N_0$ and $−\ell d,md$ both red, then $d[−\ell,m]$ is also an extendable table.

**Proof.** We choose $N_0$ to be the quantity given in Lemma 2.3, so that if we are given $\ell,m \geq N_0$, then we know that $d[−\ell,m]$ is a table. If $\ell′ \geq \ell$ and $m′ \geq m$ have $−\ell′d$ and $m′d$ both red, then $d[−\ell′,m′]$ also satisfies the conditions of Lemma 2.3, and so is also a table. Therefore, $d[−\ell,m]$ is an extendable table.

So, we can find extendable tables, provided we can find suitable reds $−\ell d$ and $md$. The following lemma helps us find suitable reds to use as $−\ell d$ and $md$. It is also useful because it implies that once we have an extendable table, the table can be extended to be arbitrarily large.

**Lemma 2.5.** Suppose $[0,d]$ is a gap, for some integer $d$. Every set of the form $d[n,(k+1)n]$ or $d[−(k+1)n,−n]$ contains a red element.

**Proof.** It is sufficient to find sequences of reds $m_1d < m_2d < \ldots$ and $−\ell_1 d > −\ell_2 d > \ldots$ with the following properties.

- $1 \leq m_1, \ell_1 \leq k + 1$,
- for all $i \geq 1$, we have $m_{i+1} \leq (k+1)m_i$ and $\ell_{i+1} \leq (k+1)\ell_i$.

This is sufficient because it ensures that each $d[n,(k+1)n]$ contains some $m_id$ and each $d[−(k+1)n,−n]$ contains some $−\ell_id$.

We set $m_1 = 1$. To find $\ell_1$, note that the arithmetic progression

$$(d,0,−d,−2d,\ldots,−kd)$$

has the first two terms being red, and therefore must contain another red term somewhere. So there is an $\ell_1 \in \{1,2,\ldots,k\}$ with $−\ell_1 d$ red.

We define the rest inductively. Given $m_i$, consider the arithmetic progression $(0,m_id,2m_id,\ldots,(k+1)m_id)$. This has the first two terms red, so there must be a $2 \leq t \leq k + 1$ with $tm_i d$ red. So set $m_{i+1} = tm_i$. Likewise, given $\ell_i$, consider the arithmetic progression $(0,−\ell_id,−2\ell_id,\ldots,−(k+1)\ell_id)$. This has the first two terms red, so there must be a $2 \leq t \leq k + 1$ with $−t\ell_id$ red. So set $\ell_{i+1} = t\ell_i$. □
It is now time to see tables in action. The idea behind a table is that if \(d[-\ell, m]\) is a table and \(r\) is a red in \([-\ell d, md] \setminus d[-\ell, m]\), then we can find an arithmetic progression of length \(k + 2\) whose first term is a red in \(d[-\ell, m]\) and whose second term is \(r\) such that the arithmetic progression stays in \([-\ell d, md]\). In this way, we can get useful information from the fact there are no \((2, k)\)-APs.

**Proposition 2.6.** Suppose \([0, d]\) is a gap (where \(d\) is an integer), and \(B \geq N_0(k)\) is an integer, where \(N_0(k)\) is as in Corollary 2.4. Then there exists a \(D_0 = D_0(B, k)\) and an extendable table \(d[-\ell, m]\) such that \(\ell, m \geq B\) and all reds in \([-\ell d, md]\) occur at integers of the form \(rd\), where \(r \in \mathbb{Q}\) has denominator at most \(D_0\).

**Proof.** By Lemma 2.5, we can find \(\ell, m \in [B, (k + 1)B]\) such that \(-\ell d\) and \(md\) are both red. The conditions of Corollary 2.4 apply, so \(d[-\ell, m]\) is an extendable table. Since \([0, d]\) is a gap, we can apply Lemma 2.1 to deduce that there is some \(R_0 = R_0(B, k)\) such that there are at most \(R_0/2\) reds in \([0, ((k + 1)B + 1)d]\). Similarly, by reflecting in \(d/2\) and applying the same result, there are at most \(R_0/2\) reds in \([-((k + 1)Bd, d]\). Putting this together, there are at most \(R_0\) reds in \([-((k + 1)Bd, (k + 1)Bd]\) and hence at most \(R_0\) reds in \([-\ell d, md]\).

Let \(S\) be the set of reds in \([-\ell d, md] \setminus d[-\ell, m]\) and let \(T\) be the set of reds in \(d[-\ell, m]\). Note that \(S\) and \(T\) are disjoint and \(S \cup T\) is the set of reds in \([-\ell d, md]\). We therefore have \(|S \cup T| \leq R_0\).

We construct a function \(f : S \to S \cup T\) as follows. Let \(s \in S\). Let \(n_1d\) be the greatest element of \(T\) which is less than \(s\) and \(n_2d\) be the least element of \(T\) which is greater than \(s\) (these must exist because \(-\ell d, md \in T\). \(d[n_1, n_2]\) is a gap, so by the table property of \(d[-\ell, m]\) we have \(n_2 - n_1 \leq m + \ell\).

If \(n_1d + (k + 1)(s - n_1d) \leq md\), then the arithmetic progression \((n_1d, s, n_1d + 2(s - n_1d), \ldots, n_1d + (k + 1)(s - n_1d))\) stays inside \([-\ell d, md]\). This arithmetic progression has its first two terms being red, so some other term must be red. We pick a \(t \in \{2, 3, \ldots, k + 1\}\) such that \(n_1d + t(s - n_1d)\) is red, and define \(f(s) = n_1d + t(s - n_1d)\).

If, however, \(n_1d + (k + 1)(s - n_1d) > md\), we can subtract \(k(n_2 - n_1)d\) from both sides to get \(n_2d + (k + 1)(s - n_2d) > md - k(n_2 - n_1)d > (m - (k + 1)(n_2 - n_1))d \geq -\ell d\), where the last inequality holds because \((k + 1)(n_2 - n_1) \leq \ell + m\). So the arithmetic progression \((n_2d, s, n_2d + 2(s - n_2d), \ldots, n_2d + (k + 1)(s - n_2d))\) will stay inside \([-\ell d, md]\). The first two terms are red, so pick a \(t \in \{2, 3, \ldots, k + 1\}\) such that \(n_2d + t(s - n_2d)\) is red, and define \(f(s) = n_2d + t(s - n_2d)\).

Let \(s \in S\) be arbitrary. We are interested in the sequence \(s, f(s), f(f(s)), \ldots\). Since \(|S| \leq R_0\), this sequence must either land in \(T\) (in
which case it terminates) or repeat itself within the first \( R_0 \) iterations. For each \( s \), \( f(s) \) is given by an equation of the form \( f(s) = hd + ts \), where \( h \) is some integer and \( t \in \{2,3,\ldots,k+1\} \) (both \( h \) and \( t \) depend on \( s \)). Composing a string of such equations, we can see that in general \( f^x(s) \) has equation \( f^x(s) = h'd + t_1t_2\ldots t_xs \), where \( h' \) is some integer and all the \( t_i \) lie in \( \{2,3,\ldots,k+1\} \). One of the following two cases must hold.

1. If \( f^x(s) \in T \) for some \( 0 \leq x \leq R_0 \), then we have that \( h'd + t_1t_2\ldots t_xs \) is a multiple of \( d \), and hence we can write \( s = rd \) where \( r \) is a rational whose denominator divides \( t_1t_2\ldots t_x \). In particular, the denominator of \( r \) is at most \( (k+1)^{R_0} \).

2. If \( f^x(s) = f^y(s) \) with \( 0 \leq x < y \leq R_0 \), then we have \( h'd + t_1t_2\ldots t_xs = h'd + t_1t_2\ldots t_ys \), which can be rearranged as \( t_1t_2\ldots t_x(t_{x+1}t_{x+2}\ldots t_y - 1)s = (h' - h''d). \) Hence, we can write \( s = rd \) where \( r \) is a rational whose denominator divides \( t_1t_2\ldots t_x(t_{x+1}t_{x+2}\ldots t_y - 1) \). Again this implies the denominator is at most \( (k+1)^{R_0} \).

So we have shown that every red in \( [-\ell d, md] \) can be written as \( rd \) where \( r \in \mathbb{Q} \) has denominator at most \( (k+1)^{R_0} \). Therefore, the result holds with \( D_0(B,k) = (k+1)^{R_0(B,k)} \).

We define \( q_0 = q_0(k) \) to be the least prime greater than \( D_0(N_0(k),k) \).

**Claim 2.7.** If \( [0,d] \) is a gap and \( d \geq q_0 \) is prime, then all reds occur at multiples of \( d \).

**Proof.** We apply Proposition 2.6 with \( B = N_0(k) \) to get an extendable table \( d[-\ell, m] \) such that all reds in \( [-\ell d, md] \) occur at integers of the form \( rd \) where \( r \in \mathbb{Q} \) has denominator at most \( D_0(N_0(k),k) < q_0 \). Since \( d \) is a prime greater than the denominator of \( r \), this implies \( r \) must be an integer, and hence all reds in \( [-\ell d, md] \) occur at multiples of \( d \).

It is sufficient to show that all reds in \( [-\ell' d, m'd] \) occur at multiples of \( d \), for any table \( d[-\ell', m'] \) that is an extension of the table \( d[-\ell, m] \) because by Lemma 2.5 we can find arbitrarily large such tables. We will prove this by induction, extending the table one bit at a time. The table \( d[-\ell, m] \) serves as a base case for our induction.

**Induction step:** Suppose \( d[-\ell', m'] \) is a table such that the only reds in \( [-\ell' d, m'd] \) occur at multiples of \( d \). Let \( m''d \) be the smallest multiple of \( d \) greater than \( m'd \) which is red. This exists by Lemma 2.5. If there are no reds apart from \( m'd \) and \( m''d \) in \( [m'd, m''d] \), then we are done. Otherwise, let \( s \) be the least red greater than \( m'd \) (\( s \) will not be a multiple of \( d \)). Since \( [-\ell' d, m''d] \) is a table and \( d[m', m''] \) is a gap, we have
This implies that all the terms in the arithmetic progression \((m''d, s, m''d - 2(m''d - s), \ldots, m''d - (k+1)(m''d - s))\) are in \([-\ell'd, m''d]\). The first two terms \(m''d\) and \(s\) are both red. Therefore, there is a \(t \in \{2, 3, \ldots, (k+1)\}\) such that \(m''d - t(m''d - s)\) is red. \(m''d - t(m''d - s)\) must be a multiple of \(d\), because of the induction hypothesis and the minimality of \(s\). Therefore, \(ts\) is a multiple of \(d\). Since \(d\) is a prime larger than \(t\), we deduce that \(s\) is a multiple of \(d\), a contradiction.

This completes the induction step for \(m'\). Mirroring the same argument tells us we can increment \(\ell'\) as well, completing the proof.

\textbf{Claim 2.8.} Let \(q\) be a prime. There exists a positive integer \(n_q\) such that if \([0, d]\) is a gap with \(q^{n_q} \mid d\), then we can find an arithmetic progression of length at least \(d\) and common difference \(q\) all of whose elements are blue.

\textbf{Proof.} We apply Proposition 2.6 with \(B = \max(q, N_0(k))\) to get a table \(d[-\ell, m]\) with \(\ell, m \geq q\) such that all reds in \([-\ell d, md]\) have the form \(rd\) where \(r \in \mathbb{Q}\) has denominator at most \(D_0(B, k)\). Choose \(n_q\) such that \(q^{n_q} > D_0(B, k)\). We take our arithmetic progression to be the set of elements of \([-\ell d, md]\) that are congruent to 1 mod \(q\). This is an arithmetic progression with common difference \(q\) of length at least \(\frac{d(\ell + m)}{q} \geq d\).

If \(q^{n_q} \mid d\), then all the elements of this progression are blue. Indeed, if one was red, we could write it as \(rd\) with \(r \in \mathbb{Q}\) having denominator less than \(q^{n_q}\). But since \(q^{n_q} \mid d\) and \(q \nmid rd\), \(q^{n_q}\) must divide the denominator of \(r\), a contradiction.

We are finally in a position to prove that \((2, k)\) is findable. In fact, we prove this slightly stronger result:

\textbf{Theorem 2.9.} Let \(\delta > 0\) and let \(k\) be a positive integer. There is a \(p_0 = p_0(k, \delta)\) such that whenever \(p \geq p_0\) is prime and \(\mathbb{Z}/p\mathbb{Z}\) is coloured red and blue such that at least \(\delta p\) of the elements are blue and at least two of the elements are red, there is a \((2, k)\)-AP.

\textbf{Proof.} By Szemerédi’s theorem, we can pick \(p_0\) such that we are guaranteed to find an arithmetic progression of length at least

\[
\prod_{q \text{ prime}} q^{n_q} \quad \text{for } q < q_0
\]

all of whose elements are blue, where \(q_0\) is as in Claim 2.7 and \(n_q\) is as in Claim 2.8.
Let \((a_1, a_2, \ldots, a_{d-1})\) be an arithmetic progression of maximal length, subject to all the elements being blue (so \(d - 1\) is the maximal length). By applying an automorphism of \(\mathbb{Z}/p\mathbb{Z}\) we can arrange for this progression to be \((1, 2, \ldots, d-1)\). By maximality, 0 and \(d\) must both be red. Since there are at least two red elements, we have \(d < p\).

Let us assume there is no \((2, k)\)-AP. If there is some prime \(q < q_0\) such that \(q^{n_q} \mid d\), we may apply Claim 2.8, working in the colouring induced in \(\mathbb{Z}\), to find a blue arithmetic progression of common difference \(q\) and length at least \(d\). \(q \neq 0\) in \(\mathbb{Z}/p\mathbb{Z}\) since \(p \geq q_0\), and so the arithmetic progression descends to a valid arithmetic progression in \(\mathbb{Z}/p\mathbb{Z}\). This contradicts the maximality of \((a_1, a_2, \ldots, a_{d-1})\).

If, on the other hand, there is no prime \(q < q_0\) such that \(q^{n_q} \mid d\), then consider the prime factorisation of \(d\). Since

\[
d > \prod_{q \text{ prime}} q^{n_q}
\]

there must be some prime \(q \geq q_0\) such that \(q \mid d\). We manipulate the colouring of \(\mathbb{Z}/q\mathbb{Z}\) by applying the automorphism

\[
\phi: \mathbb{Z}/p\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z} \\
x \mapsto xd/q.
\]

Then we will have that 0 and \(q\) are red while \([1, 2, \ldots, q-1]\) are blue. In the colouring induced on \(\mathbb{Z}\), \([0, q]\) is a gap and \(q \geq q_0\) is prime, so we may apply Claim 2.7 to deduce that all reds in the colouring occur at multiples of \(q\). \(p \in \mathbb{Z}\) must be red since it corresponds to \(0 \in \mathbb{Z}/p\mathbb{Z}\), therefore \(q \mid p\). Since \(p\) is prime, we have \(q = p\), but \(q \leq d < p\), a contradiction.

3. Construction without a \((14, 14)\)-AP

We aim to produce colourings for \(\mathbb{Z}/p\mathbb{Z}\) which contain no \((14, 14)\)-AP. To do this, it suffices to give such a colouring of \(\mathbb{R}/\mathbb{Z}\), since this induces colourings of \(\mathbb{Z}/p\mathbb{Z}\) for all \(p\) via the map

\[
\phi: \mathbb{Z}/p\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z} \\
x \mapsto \frac{x}{p}
\]

and the induced colourings contain \((14, 14)\)-APs only if the original colouring contained \((14, 14)\)-APs. The majority of this section is devoted to defining a red-blue colouring of \(\mathbb{R}/\mathbb{Z}\) and proving that this colouring has no \((14, 14)\)-AP.
ARITHMETIC PROGRESSIONS IN 2-COLOURED CYCLIC GROUPS

The configuration $c_0$. The red interval is shown with a thick line and the blue interval is shown with a double line. Uncoloured intervals are shown with an arrow indicating their direction.

This is shown in Figure 1. We now define a replacement rule for directed intervals. Let $(a, b)$ be one of the uncoloured intervals, and $\epsilon$ be its direction. We define the function $f : [0, 1] \rightarrow [a, b]$ as follows:

$$f : x \mapsto \begin{cases} a + (b - a)x & \text{if } \epsilon = +1 \\ b - (b - a)x & \text{if } \epsilon = -1. \end{cases}$$

We replace $(a, b) = f((0, 1))$ with the following:

- $f((0, 1/4))$ uncoloured, direction $\epsilon$
- $f((1/4, 1/2))$ uncoloured, direction $+1$
- $f((1/2, 3/4))$ coloured blue
- $f((3/4, 1))$ uncoloured, direction $-1$
- $f((1/5, 2/5))$ coloured blue
- $f((2/5, 3/5))$ uncoloured, direction $-\epsilon$
- $f((3/5, 4/5))$ coloured red
- $f((4/5, 1))$ uncoloured, direction $\epsilon$.

**Figure 2.** The replacement rule for a directed interval. The red interval is shown with a thick line and the blue interval is shown with a double line. Uncoloured intervals are shown with an arrow indicating their direction. This is shown for the interval $(0, 1)$. The rule for a general interval is obtained by applying the linear function $f$.

### 3.1. Fractal construction

We define a configuration to be a partial colouring of $\mathbb{R}/\mathbb{Z}$ together with a partition of the uncoloured subset into intervals, and an assignment of a direction to each of these intervals. Each interval can either be assigned direction $+1$, i.e. “forwards”, or direction $-1$, i.e. “backwards”. We proceed to define a fractal by specifying an initial configuration $c_0$ and a replacement rule telling us how to go from $c_i$ to $c_{i+1}$. Note that, as a set, $\mathbb{R}/\mathbb{Z}$ may be identified with $[0, 1)$. We will often do this for convenience.

Let $c_0$ be the following configuration:

- $[0, 1/4]$ coloured red
- $(1/4, 1/2)$ uncoloured interval direction $+1$
- $[1/2, 3/4]$ coloured blue
- $(3/4, 1)$ uncoloured interval direction $-1$.

This is shown in Figure 1. We now define a replacement rule for directed intervals. Let $(a, b)$ be one of the uncoloured intervals, and $\epsilon$ be its direction. We define the function $f : [0, 1] \rightarrow [a, b]$ as follows:

$$f : x \mapsto \begin{cases} a + (b - a)x & \text{if } \epsilon = +1 \\ b - (b - a)x & \text{if } \epsilon = -1. \end{cases}$$

We replace $(a, b) = f((0, 1))$ with the following:

- $f((0, 1/4))$ uncoloured, direction $\epsilon$
- $f((1/4, 1/2))$ uncoloured, direction $+1$
- $f((1/2, 3/4))$ coloured blue
- $f((3/4, 1))$ uncoloured, direction $-1$
- $f((1/5, 2/5))$ coloured blue
- $f((2/5, 3/5))$ uncoloured, direction $-\epsilon$
- $f((3/5, 4/5))$ coloured red
- $f((4/5, 1))$ uncoloured, direction $\epsilon$. 
This is shown in Figure 2. We can construct configurations \((c_i)_{i=0}^\infty\) inductively. For each \(i \geq 0\), define \(c_{i+1}\) to be the configuration obtained by applying the replacement rule to all the uncoloured intervals in \(c_i\). Note that in \(c_0\) we split the interval into quarters, whereas at each successive stage we split the uncoloured intervals into fifths. Figure 3 shows \(c_1\) and \(c_2\). These colourings have the property that if a point receives a given colour in \(c_i\), it will also receive that colour in \(c_j\) for all \(j \geq i\). Therefore, we can define a partial colouring \(c_\infty\), by colouring each point red if there is some \(i\) such that \(c_i\) colours that point red, and colouring it blue if there is some \(i\) such that \(c_i\) colours that point blue. This partial colouring \(c_\infty\) will be a fractal. There will be some points (for example, the point \(3/8\)) that will not receive either colour in \(c_\infty\), so this is not yet a full colouring of \(\mathbb{R}/\mathbb{Z}\). We finally construct a colouring \(c'_\infty\) by taking \(c_\infty\) and colouring all the uncoloured points in some arbitrary way (for example, we could colour all the uncoloured points blue). This \(c'_\infty\) is our counterexample colouring.

We now establish a symmetry of the configurations \(c_i\), and use it to deduce a symmetry property of \(c_\infty\). Let \(T\) denote the following operation for a configuration:

- Translate everything by \(1/2\)
- For all coloured points, switch the colour between red and blue
- For all uncoloured intervals, reverse the direction of the interval.

One can see by inspection that \(T(c_0) = c_0\). Also, by inspecting the replacement rule, we also have that, for any \(i\), if \(T(c_i) = c_i\), then \(T(c_{i+1}) = c_{i+1}\). Therefore, by induction, \(T\) is a symmetry of \(c_i\) for all \(i\). By going to the limit, we can deduce that in \(c_\infty\), \(x\) is red if and only if \(x+1/2\) is blue.

3.2. Proof there is no \((14,14)\)-AP

When we have a partial colouring, we define an \((a,b)\)-AP to be an arithmetic progression of length \(a+b\) such that none of the first \(a\) elements are blue and none of the last \(b\) elements are red. When the partial colouring is in fact a full colouring, this notion coincides with the usual notion of an \((a,b)\)-AP.
We start off by noting that no \((14, 14)\)-AP in \(c'_{\infty}\) can have common difference 0 or 1/2, because in the former case all points have the same colour, and in the latter case the colouring has period 2. So it is sufficient to focus on the case where \(d \in (-1/2, 0) \cup (0, 1/2)\). We will prove the following:

**Theorem 3.1.** \(c_{\infty}\) does not contain any \((14, 14)\)-AP with common difference \(d \in (-1/2, 0) \cup (0, 1/2)\).

This implies that \(c'_{\infty}\) does not contain any \((14, 14)\)-AP with \(d \in (-1/2, 0) \cup (0, 1/2)\).

**Proof.** Firstly, we observe that we may restrict attention to \(d \in (0, 1/2)\). Indeed, if \(d \in (-1/2, 0)\), then any \((14, 14)\)-AP \(\{a, a + d, \ldots, a + 27d\}\) with common difference \(d\) will give rise to another \((14, 14)\)-AP, 

\[
\{a + 27d + 1/2, a + 26d + 1/2, \ldots, a + d + 1/2, a + 1/2\},
\]

with common difference \(-d\). If the former is a \((14, 14)\)-AP, then so is the latter because of the symmetry property of \(c_{\infty}\).

We split into the following cases, according to where in the interval \((0, 1/2)\) the common difference \(d\) lies:

1. \(d \in \left(\frac{1}{4^{k+1} \cdot 5^k}, \frac{1}{4^k}\right]\) for some integer \(k \geq 0\)
2. \(d \in \left[\frac{1}{3}, \frac{5}{16}\right]\)
3. \(d \in \left]\frac{5}{16}, \frac{7}{20}\right]\)
4. \(d \in \left]\frac{7}{20}, \frac{2}{5}\right]\)
5. \(d \in \left[\frac{1}{2} - \frac{1}{8^{k+1}}, \frac{1}{2} - \frac{1}{8^k}\right]\) for some integer \(k \geq 0\).

Note that at least one of these cases applies. In each case we will show that, for some given \(i\) and \(a, b \leq 14\), \(c_i\) does not contain an \((a, b)\)-AP with common difference \(d\). This implies \(c_{\infty}\) does not contain a \((14, 14)\)-AP with common difference \(d\).

**Case 1.** Let \(k \geq 0\) be an integer. We claim \(c_{k+1}\) does not contain a \((14, 14)\)-AP with \(d \in \left(\frac{1}{4^{k+1} \cdot 5^k}, \frac{1}{4^k}\right]\).

We start off by considering \(c_k\). \(c_k\) consists of the following repeating pattern:

- a blue interval of length at least \(\frac{1}{4^k}\)
- an uncoloured interval of length \(\frac{1}{4^k}\)
- a red interval of length at least \(\frac{1}{4^k}\)
- an uncoloured interval of length \(\frac{1}{4^k}\).
This repeats itself over and over as we go around $\mathbb{R}/\mathbb{Z}$. The lengths of the red and blue intervals will vary as the pattern repeats itself. The pattern has the additional property that in any pair of adjacent coloured intervals, at least one of the coloured intervals has length exactly $\frac{1}{4.5^k}$. All these properties can easily be verified by induction on $k$.

We say an arithmetic progression jumps over an interval if it contains points before and after the interval, but does not contain any points in the interval. Obviously, this can only happen if the length of the interval is smaller than the common difference. In this case, we assume $d \leq \frac{1}{4.5^k}$, so the arithmetic progression cannot jump over any of the intervals in $c_k$.

Let us assume for contradiction that \( \{a, a+d, a+2d, \ldots, a+26d, a+27d\} \) is a \((14, 14)\)-AP. Since $a+13d$ is not blue and $a+14d$ is not red, the interval $[a+13d, a+14d]$ cannot be contained in a coloured interval. Therefore, we can pick some $t \in [13, 14]$ such that $a+td$ is in an uncoloured interval of $c_k$. Let this uncoloured interval be $(b, b + \frac{1}{4.5^k})$. Since $d > \frac{1}{4.5^k+1}$, our arithmetic progression must contain points both before and after this uncoloured interval. Therefore, the uncoloured interval $(b, b + \frac{1}{4.5^k})$ is preceded by a red interval and succeeded by a blue interval (it cannot be the other way around because the first 14 points are not allowed to be blue and intervals in $c_k$ are too long to be jumped over). At least one of these two coloured intervals must have length exactly $\frac{1}{4.5^k}$. Because the situation so far is symmetric, we may assume without loss of generality that the red interval has length $\frac{1}{4.5^k}$. So we have the following situation, which is shown in Figure 4:

- $[b - \frac{3}{4.5^k}, b - \frac{2}{4.5^k}]$ is blue
- $(b - \frac{1}{4.5^k}, b)$ is uncoloured
- $[b, b + \frac{1}{4.5^k}]$ is red
- $(b, b + \frac{1}{4.5^k})$ is uncoloured
- $[b + \frac{1}{4.5^k}, b + \frac{2}{4.5^k}]$ is blue.

The first 14 terms of the arithmetic progression must be contained in the interval $(b - \frac{2}{4.5^k}, b + \frac{1}{4.5^k})$, since they are not allowed to be blue and the blue intervals are too big to be jumped over. Now we consider the colouring given by $c_{k+1}$. This is shown in Figure 5. The interval $(b - \frac{2}{4.5^k}, b + \frac{1}{4.5^k})$ is partitioned into fifteen intervals of width $\frac{1}{4.5^k+1}$, two of which are coloured blue. Since $d > \frac{1}{4.5^k+1}$, each of these intervals contains at most one of

---

**Figure 4.** The configuration $c_k$ around the area of interest. The red interval is shown with a thick line and blue intervals are shown with double lines
Figure 5. The configuration \( c_{k+1} \) around the area of interest, and the partition of \((b - \frac{2}{5^k}, b + \frac{1}{5^k})\) into fifteen intervals. Red intervals are shown with thick lines and blue intervals are shown with double lines.

\( \{a, a+d, \ldots, a+12d, a+13d\} \). This implies at least one of the two blue intervals contains a point of \( \{a, a+d, \ldots, a+12d, a+13d\} \), because you cannot put 14 points in 13 intervals without putting two in the same interval. But none of these points are allowed to be blue, a contradiction. This concludes case (1).

For cases (2)-(4), we need the notion of a ladder. We define a ladder to be a finite non-decreasing sequence of real numbers, such that the difference between consecutive terms is at most \( \frac{1}{4} \), and the difference between the first and last terms is 1. These are useful because, since \([1/2, 3/4]\) is blue in \( c_0 \), any ladder must contain a blue point of \( c_0 \).

**Case 2.** If \( d \in \left[ \frac{1}{4}, \frac{5}{16} \right] \), then \( 4d-1, 1-3d \in [0, 1/4] \), so

\[
\begin{align*}
a, a+4d-1, a+d, a+5d-1, a+2d, a+6d-1, a+3d, a+1
\end{align*}
\]

is a ladder for any \( a \). \( c_0 \) cannot contain a \((7, 0)\)-AP with first term \( a \) and \( d \in \left[ \frac{1}{4}, \frac{5}{16} \right] \), because if it did, the above ladder would contain no blue points.

**Case 3.** If \( d \in \left[ \frac{5}{16}, \frac{7}{20} \right] \), then \( 5d-3/2, 3/2-4d, 1/2-d, 2d-1/2 \in [0, 1/4] \), so

\[
\begin{align*}
a, a+5d-\frac{3}{2}, a+d, a+3d-\frac{1}{2}, a+2d, a+4d-\frac{1}{2}, a+1
\end{align*}
\]

is a ladder for any \( a \). \( c_0 \) cannot contain a \((3, 3)\)-AP with first term \( a \) and \( d \in \left[ \frac{5}{16}, \frac{7}{20} \right] \), because if it did, the above ladder would contain no blue points. (Note that \( a+3d, a+4d, a+5d \) not being red is the same as \( a+3d-\frac{1}{2}, a+4d-\frac{1}{2} \) and \( a+5d-\frac{1}{2} \) not being blue, because of the symmetry).

**Case 4.** If \( d \in \left[ \frac{7}{20}, \frac{2}{5} \right] \), then \( 3d-1, 2-5d \in [0, 1/4] \), so

\[
\begin{align*}
a, a+3d-1, a+6d-2, a+d, a+4d-1, a+7d-2, a+2d, a+5d-1, a+1
\end{align*}
\]

is a ladder for any \( a \). \( c_0 \) cannot contain an \((8, 0)\)-AP with first term \( a \) and \( d \in \left[ \frac{7}{20}, \frac{2}{5} \right] \), because if it did, the above ladder would contain no blue points.

**Case 5.** Let \( k \geq 0 \) be an integer. We claim \( c_k \) does not contain a \((13, 0)\)-AP with \( d \in \left[ \frac{1}{2} - \frac{1}{8^k}, \frac{1}{2} - \frac{1}{8^k + n} \right) \). Suppose for contradiction \( \{a, a+d, \ldots, a+11d, a+12d\} \) is a \((13, 0)\)-AP.
Let \( d' = 1/2 - d \) and let \( a' = a + 12d \). For any integer \( t \) with \( 0 \leq t \leq 12 \), \( a' + td' = a + (12 - t)d + t/2 \). Since \( a + (12 - t)d \) is not blue, \( a' + td' \) is not blue if \( t \) is even and not red if \( t \) is odd. We also have that \( d' \in \left( \frac{1}{8.5^k + 1}, \frac{1}{8.5^k} \right) \). Recall \( c_k \) consists of the following repeating pattern:

- a blue interval of length at least \( \frac{1}{4.5^k} \)
- an uncoloured interval of length \( \frac{1}{4.5^k} \)
- a red interval of length at least \( \frac{1}{4.5^k} \)
- an uncoloured interval of length \( \frac{1}{4.5^k} \).

Two consecutive points in the arithmetic progression

\[ \{a', a' + d', \ldots, a' + 11d', a' + 12d'\} \]

cannot lie in the same coloured interval of \( c_k \), since one of these two points is not allowed to be red, and the other is not allowed to be blue. Suppose \( a' + kd' \) lies in some coloured interval. If \( k \neq 0, 12 \), then either \( a' + (k - 1)d' \) or \( a' + (k + 1)d' \) must lie in the same coloured interval, giving a contradiction, because the coloured interval has length at least \( \frac{1}{4.5^k} \) and \( d' \leq \frac{1}{8.5^k} \).

Therefore, all of \( a' + d', a' + 2d', \ldots, a' + 11d' \) must be in uncoloured intervals. Like in case (1), the common difference is too small to jump over the coloured intervals, so these 11 points must all lie in the same uncoloured interval. But \( a' + 11d' - (a' + d') = 10d' > \frac{1}{4.5^k} \) so \( a' + 11d' \) and \( a' + d' \) are too far from each other to lie in the same uncoloured interval. We have a contradiction.

This concludes the proof of Theorem 3.1.

So we now know that our colouring \( c_{\infty} \) of \( \mathbb{R}/\mathbb{Z} \) does not contain a \((14,14)\)-AP, and therefore the same must also be true of the induced colourings of \( \mathbb{Z}/p\mathbb{Z} \).

Let \( r_p \) and \( b_p \) be the proportions of red and blue points respectively, in the colouring on \( \mathbb{Z}/p\mathbb{Z} \). The red and blue sets in \( c_{\infty} \) are both countable unions of intervals of total measure 1/2. Since the proportion of points in \( \mathbb{Z}/p\mathbb{Z} \) above some interval of length \( \ell \) in \( \mathbb{R}/\mathbb{Z} \) is \( \ell + O(p^{-1}) \), this implies that as \( p \to \infty \), \( \liminf r_p \geq 1/2 \) and \( \liminf b_p \geq 1/2 \). As \( r_p + b_p = 1 \), This implies \( r_p \) and \( b_p \) must both tend to 1/2. So \((14,14)\) is not findable, even for \( \delta \) arbitrarily close to 1/2.

4. Construction without a \((3,30000)\)-AP

4.1. Fractal construction

We will use the same kind of construction as in Section 3.1, except that the fractal will have a more complicated structure. We aim to construct a
colouring $c'_\infty$ of $\mathbb{R}/\mathbb{Z}$ which has no (3, 30000)-AP, as this will induce suitable colourings of $\mathbb{Z}/p\mathbb{Z}$. We start by defining an infinite series of configurations $c_0, c_1, c_2, \ldots$, each one obtained from the previous one by certain replacement rules. Since the fractal involved is more complicated, we need to use two kinds of uncoloured intervals, each with their own replacement rule.

For the purposes of this section, we define a configuration as follows. A configuration is a partial colouring of $R/\mathbb{Z}$ together with a partition of the uncoloured subset into intervals, and an assignment of a type and a direction to each of these intervals. Each interval can either be assigned Type I or Type II, independently of its direction, and the direction can be either $+1$, i.e. “forwards”, or $-1$, i.e. “backwards”. Let $c_0$ be the following configuration:

- $(0, 1/6)$ Type I uncoloured interval directed backwards
- $[1/6, 1/3]$ coloured blue
- $(1/3, 1/2)$ Type I uncoloured interval directed forwards
- $[1/2, 1]$ coloured blue.

This is shown in Figure 6. The replacement rules are as follows. Let $(a, b)$ be one of the uncoloured intervals, and $\epsilon$ be its direction. Let $f : [0, 1] \rightarrow [a, b]$ be given by

$$f : x \mapsto \begin{cases} a + (b - a)x & \text{if } \epsilon = +1 \\ b - (b - a)x & \text{if } \epsilon = -1. \end{cases}$$

If $(a, b)$ is of Type I, we replace $(a, b) = f((0, 1))$ with the following:

- $f \left((0, \frac{1}{288})\right)$ Type II uncoloured, direction $-\epsilon$
- $f \left(\left[\frac{1}{288}, \frac{7}{288}\right]\right)$ coloured red
- $f \left(\left[\frac{7}{288}, \frac{1}{36}\right]\right)$ Type II uncoloured, direction $\epsilon$
- $f \left(\left[\frac{1}{36}, \frac{1}{3}\right]\right)$ coloured blue
- $f \left(\left(\frac{1}{3}, 1\right)\right)$ Type I uncoloured, direction $\epsilon$.

Note that the values $\frac{1}{288}$ and $\frac{7}{288}$ are chosen so that the middle $\frac{3}{4}$ of the first $\frac{1}{36}$ of the interval is red.

If $(a, b)$ is of Type II, we replace $(a, b) = f((0, 1))$ with the following:

- $f \left((0, \frac{1}{101})\right)$ Type II uncoloured, direction $\epsilon$
- $f \left(\left[\frac{1}{101}, \frac{51}{101}\right]\right)$ coloured blue
- $f \left(\left[\frac{51}{101}, 1\right]\right)$ Type I uncoloured, direction $\epsilon$.

This is shown in Figure 7. In the limit, we will get a partial colouring $c_\infty$ of $\mathbb{R}/\mathbb{Z}$. We then form our colouring $c'_\infty$ by arbitrarily colouring the uncoloured elements from $c_\infty$. 

Figure 6. The configuration $c_0$. Blue intervals are shown with double lines. Type I uncoloured intervals are shown in black with an arrow indicating their direction.

Figure 7. The replacement rules. Red intervals are shown with thick lines and blue intervals are shown with double lines. Type I uncoloured intervals are shown in black with an arrow indicating their direction. Type II uncoloured intervals are shown in green with a double arrow indicating their direction. The drawings are not to scale. For the first replacement rule, we zoom in on the first $\frac{1}{36}$ of the interval to show detail. These are shown for the interval $(0,1)$. The rules for a general interval are obtained by applying the linear function $f$.

4.2. Proof there is no $(3, 30000)$-AP

As in the previous section, we will use the convention that, when dealing with a partial red-blue colouring, we define an $(a, b)$-AP to be an arithmetic progression of length $a + b$ such that none of the first $a$ elements are blue and none of the last $b$ elements are red. It suffices to prove that $c_\infty$ contains no non-trivial $(3, 30000)$-AP under this definition.

First, we establish the following properties of the configurations $c_k$:

1. (a) Every Type I uncoloured interval of length $\ell$ in $c_k$ is preceded by a blue interval of length at least $\frac{11\ell}{24}$
   (b) Every Type I uncoloured interval of length $\ell$ in $c_k$ is succeeded by a blue interval of length at least $3\ell$

2. (a) Every Type II uncoloured interval of length $\ell$ in $c_k$ is preceded by a red interval of length at least $6\ell$
   (b) Every Type II uncoloured interval of length $\ell$ in $c_k$ is succeeded by a blue interval of length at least $6\ell$

3. Behind a Type I uncoloured interval of length $\ell$ in $c_k$, within a distance of $\frac{25\ell}{24}$, we can find a red interval of $c_{k+1}$ of length at least $\frac{\ell}{48}$

4. If a Type II uncoloured interval of length $\ell$ in $c_k$ came from a Type I uncoloured interval of $c_{k-1}$, then in front of that Type II uncoloured interval, within a distance of $300\ell$, we can find a red interval of $c_{k+1}$ of length at least $4\ell$
Note that the meanings of the terms “preceded”, “succeeded”, “in front of” and “behind” are determined by the direction of the uncoloured interval. We will prove all of these properties together, by induction on \( k \).

**Proof.** We start off with the base case \( k = 0 \). Since there are no Type II uncoloured intervals in \( c_0 \), we only need to check properties 1 and 3. \( c_0 \) has two uncoloured intervals of Type I, and each has length \( \frac{1}{6} \). They are \((0, \frac{1}{6})\) directed backwards and \((\frac{1}{3}, \frac{1}{2})\) directed forwards. They are each preceded by a blue interval of length \( \frac{1}{6} \) and succeeded by a blue interval of length \( \frac{1}{2} \), so Property 1 holds.

Since \( c_0 \) is symmetric about \( \frac{1}{4} \), it suffices to check Property 3 holds for \((\frac{1}{3}, \frac{1}{2})\). When we apply the replacement rule to \((0, \frac{1}{6})\), we get a red interval \((\frac{281}{1728}, \frac{287}{1728})\) of length \( \frac{1}{288} = \frac{1}{48} \times \frac{1}{5} \). This interval is in \( c_1 \) and is within \( \frac{25}{24} \) of \( \frac{1}{3} \) (as \( \frac{1}{3} \) - \( \frac{25}{24} \times \frac{1}{6} = \frac{23}{144} = \frac{276}{1728} \)).

Now for the induction step. For \( k > 0 \), every uncoloured interval of \( c_k \) must arise from applying a replacement rule to an uncoloured interval of \( c_{k-1} \), so we work case by case by case.

**Property 1.** First case: an uncoloured interval of Type I that came from an uncoloured interval of Type I and length \( \ell_0 \) in \( c_{k-1} \). The interval has length \( \ell = \frac{2}{3} \ell_0 \), and is preceded by a blue interval of length \( \frac{11}{36} \ell_0 = \frac{11}{24} \ell \). It is succeeded by a blue interval of length at least \( 3 \ell_0 > 3 \ell \) (using Property 1(b) for \( c_{k-1} \)), so Property 1 holds.

Second case: an uncoloured interval of Type I that came from an uncoloured interval of Type II and length \( \ell_0 \) in \( c_{k-1} \). The interval has length \( \ell = \frac{50}{101} \ell_0 \), and is preceded by a blue interval of length \( \frac{50}{101} \ell_0 = \ell > \frac{11}{24} \ell \). It is succeeded by a blue interval of length at least \( 6 \ell_0 > 3 \ell \) (using Property 2(b) for \( c_{k-1} \)), so Property 1 holds.

**Property 2.** First case: an uncoloured interval of Type II that came from an uncoloured interval of Type I and length \( \ell_0 \) in \( c_{k-1} \). The interval has length \( \ell = \frac{1}{288} \ell_0 \), and is preceded by a red interval of length \( \frac{1}{48} \ell_0 = 6 \ell \).

We have two subcases, either it is “pointing inwards”, in which case it is succeeded by a blue interval of length \( \frac{11}{36} \ell_0 > 6 \ell \), or it is “pointing outwards” in which case it is succeeded by a blue interval of length at least \( \frac{11}{24} \ell_0 > 6 \ell \) (using Property 1(a) for \( c_{k-1} \)). So Property 2 holds.

Second case: an uncoloured interval of Type II that came from an uncoloured interval of Type II and length \( \ell_0 \) in \( c_{k-1} \). The interval has length \( \ell = \frac{1}{101} \ell_0 \), and is preceded by a red interval of length at least \( 6 \ell_0 > 6 \ell \) (using Property 2(a) for \( c_{k-1} \)). It is succeeded by a blue interval of length \( \frac{50}{101} \ell_0 > 6 \ell \), so Property 2 holds.
Property 3. First case: an uncoloured interval of Type I that came from an uncoloured interval of Type I and length $\ell_0$ in $c_{k-1}$. The interval has length $\ell = \frac{2}{3}\ell_0$, and we have a red interval of length $\frac{1}{15}\ell_0 = \frac{1}{32}\ell$ also arising from the same Type I uncoloured interval of $c_{k-1}$. It is within $\frac{1}{3}\ell_0 = \frac{1}{2}\ell$ of the rear end of our interval, so Property 3 holds.

Second case: an uncoloured interval of Type I that came from an uncoloured interval of Type II and length $\ell_0$ in $c_{k-1}$. Let $(m, m + \ell_0)$ be this Type II interval of $c_{k-1}$, which we may assume without loss of generality is directed forwards. Our Type I interval is $(m + \frac{51}{101}\ell_0, m + \ell_0)$. It has length $\ell = \frac{50}{101}\ell_0$. By applying property 2(a) in $c_{k-1}$, we can see that $[m - \frac{1}{96}\ell_0, m]$ is red. This has length $\frac{1}{96}\ell_0 = \frac{101}{4800}\ell > \frac{1}{48}\ell$, and is within $(\frac{51}{101} + \frac{1}{96})\ell_0 = (\frac{51}{101} + \frac{101}{4800})\ell < \frac{25}{24}\ell$, so Property 3 holds.

Property 4. Consider an uncoloured interval of Type II that came from an uncoloured interval of Type I and length $\ell_0$ in $c_{k-1}$. The interval has length $\ell = \frac{1}{288}\ell_0$. Here we have two subcases, the “pointing outwards” and “pointing inwards” cases. In the “pointing outwards” case, we use property 3 in $c_{k-1}$ to deduce that there is a red interval of length at least $\frac{1}{48}\ell_0 = 6\ell$ within $\frac{25}{24}\ell_0 = 300\ell$. In the “pointing inwards” case, we note that there is a Type I interval of length $\frac{2}{3}\ell_0$ that came from the same Type I interval of $c_{k-1}$. In $c_{k+1}$, this will give us a red interval of length $\frac{2}{3} \times \frac{1}{48}\ell_0 = \frac{1}{72}\ell_0 = 4\ell$. It is within a distance of $\ell_0 = 288\ell$ in front of our Type II interval, so Property 4 holds.

This completes the induction step and concludes our proof. 

Now we can proceed with the proof that $c_\infty$ contains no non-trivial (3,30000)-AP. Suppose, for a contradiction, that it does contain one, and it has first term $a$ and common difference $d \neq 0$. Because the initial configuration $c_0$ is symmetric by a reflection in $1/4$, it follows that $c_i$ is also symmetric for all $i \leq \infty$. Therefore, we may assume without loss of generality that $d \in (0,1/2)$. Also, the (3,30000)-AP is a (3,30000)-AP in $c_i$ for all $i$, because the partial colouring in each $c_i$ is a subcolouring of $c_\infty$.

We start off by considering just the first three points of the arithmetic progression, and where they might lie. None of these three points are allowed to be blue.

Firstly, all three of them must lie in the same Type I uncoloured interval of $c_0$. This is because the not-blue set in $c_0$ is $(0,1/6) \cup (1/3,1/2)$, and the only way you can have an arithmetic progression of length three in $\mathbb{R}/\mathbb{Z}$ with all three points lying in $(0,1/6) \cup (1/3,1/2)$ is to either have all three of them lying in $(0,1/6)$ or all three lying in $(1/2,1/3)$. Let $k$ be maximal such that $c_k$ contains a Type I uncoloured interval which contains all three
must lie in all shown in Figure 8. The three terms of the arithmetic progression cannot happen.

All blue points of the first 36 intervals. We argue that if the first three points are all in $I_1$ but not all in $I_2$, then they are all in $I_1$.

points. Call this interval $I_1$, and say it starts at $s$ and ends at $s + t$ (t will be negative if $I_1$ is directed backwards). Let the first $\frac{1}{36}$ of $I_1$ be $J_1$, and the last $\frac{2}{3}$ be $I_2$. In $c_{k+1}$, $J_2$ is a Type I uncoloured interval, directed in the same direction as $I_1$, and $J_1 \cup J_2$ is the subset of $I_1$ which is not blue. We similarly define $J_2$ to be the first $\frac{1}{36}$ of $I_2$, $I_3$ to be the last $\frac{2}{3}$ of $I_2$, $J_3$ to be the first $\frac{1}{36}$ of $I_3$ and $I_4$ to be the last $\frac{2}{3}$ of $I_3$. $J_1 \cup J_2 \cup J_3$ contains all the not blue points of $c_{k+2}$ from $I_1$, while $J_1 \cup J_2 \cup J_3 \cup I_4$ contains all the not blue points of $c_{k+3}$ from $I_1$. So all three points must lie in $J_1 \cup J_2 \cup J_3 \cup I_4$. This is all shown in Figure 8. The three terms of the arithmetic progression cannot all lie in $I_2$ by the maximality of $k$, so at least one must lie in $J_1$. We claim that in fact all three of them lie in $J_1$.

Say the three points are $p_1 = s + u_1 t$, $p_2 = s + u_2 t$ and $p_3 = s + u_3 t$, with $u_1, u_2, u_3 \in (0, 1)$ in increasing order. We have $2u_2 = u_1 + u_3$. Note that $p_1$ must lie in $J_1$, so $u_1 \in \left(0, \frac{1}{36}\right)$. If $p_2 \in J_1$ as well, then we must have $p_3 \in J_1$ since the gap between $J_1$ and $I_2$ is wider than $J_1$ is. Then all three points lie in $J_1$, as required.

There are two other options: $p_2$ can lie in $J_2$ or $I_3$. $I_3$ goes from $s + \frac{5}{9} t$ to $s + t$, so in the latter case $u_2 > \frac{5}{9}$ and $u_1 < \frac{1}{36}$ so $u_3 > \frac{10}{9} - \frac{1}{36} > 1$ which cannot happen. $J_2$ goes from $s + \frac{1}{3} t$ to $s + \frac{19}{54} t$, so if $p_2$ lies in $J_2$, then $u_3 \in \left(\frac{2}{3} - \frac{1}{36}, \frac{19}{27}\right) = \left(\frac{23}{36}, \frac{19}{27}\right)$. However, $J_3$ ends at $s + \frac{46}{81} t$ and we have $\frac{46}{81} < \frac{23}{36}$ while $I_4$ starts at $s + \frac{19}{27} t$, which shows that in this case the point $p_3$ will lie in between $J_3$ and $I_4$, which is not allowed.

So all three points lie in $J_1$. In $c_{k+1}$, the middle $\frac{3}{4}$ of $J_1$ is red, while the first $\frac{1}{8}$ and last $\frac{1}{8}$ are Type II uncoloured intervals. Let $J'_1$ be the middle $\frac{3}{4}$ of $J_1$. We claim that at least one of the three points lies in $J'_1$. 

Figure 8. Configurations $c_k$ through $c_{k+3}$, between $s$ and $s+t$. Blue intervals are shown with double lines. The intervals $I_i$ are Type I uncoloured in each case, while the $J_i$ contain a variety of intervals. We argue that if the first three points are all in $I_1$ but not all in $I_2$, then they are all in $I_1$. 


Suppose this is not the case. One of the Type II uncoloured intervals must contain all three points, because the arithmetic progression cannot jump over $J'_1$. Let $m$ be maximal such that all three points lie in the same Type II uncoloured interval of $c_m$. Let $K_1$ be this Type II interval. We have $m \geq k+1$. In $c_{m+1}$, we cannot have all three points lying in the same Type II interval, or in the same Type I interval, by the maximality of $k$ and $m$. In $c_{m+1}$, $K_1$ is replaced by a Type II uncoloured interval, followed by a blue interval, followed by a Type I uncoloured interval, as shown in Figure 7. The three points are not allowed to be blue, but they also cannot jump over the blue interval of $c_{m+1}$, so they must either all lie in the Type I uncoloured interval of $c_{m+1}$ or the Type II uncoloured interval of $c_{m+1}$. But this contradicts the maximality of $k$ and $m$. Hence, at least one of the three points must lie in $J'_1$.

Let $r$ be the length of the interval $J_1$. Since the first three points of our $(3, 30000)$-AP are contained in $J_1$, we have that $d \leq \frac{r}{2}$. To conclude our proof, we will apply the following lemma a couple of times.

**Lemma 4.1.** Suppose we know that $d \leq d_{\text{max}}$ for some $d_{\text{max}}$, and that in front of the interval $J'_1$ in the positive direction there is at some distance a blue interval $L_1$ of length at least $d_{\text{max}}$, and at some further distance there is a red interval $L_2$ of length at least $d_{\text{max}}$. Let $\ell$ be the distance from the end of $J'_1$ to the start of $L_2$. Then we have

$$d \leq \frac{\ell}{29999}.$$  

**Proof.** If our $(3, 30000)$-AP contains any points after the start of $L_2$, then it must contain at least one point of the interval $L_2$, and also at least one point of the interval $L_1$, because both of these intervals are too long to jump over. So the $(3, 30000)$-AP contains a red point in $J'_1$, a blue point in $L_1$ and a red point in $L_2$, and they occur in that order. But this cannot happen. Therefore, all of the points must occur before the start of $L_2$. Also, the last 30000 points are not allowed to be red, so they must occur after the start of $J'_1$. Therefore, we have 30000 points of the arithmetic progression occurring within an interval of length $\ell$, and the conclusion must hold.

Since $d \leq \frac{r}{2}$, we start off by applying Lemma 4.1 with $d_{\text{max}} = \frac{r}{2}$. The last $\frac{1}{8}$ of $J_1$ is a Type II interval of $c_{k+1}$ that came from a Type I interval of $c_k$. It has length $\frac{r}{8}$. By applying Property 2(b), we know that it is followed by a blue interval of length at least $\frac{3r}{4} > d_{\text{max}}$. By applying Property 4, we also know that within a distance of $\frac{300r}{8}$, there is a red interval of length at least $\frac{r}{2} = d_{\text{max}}$. We apply Lemma 4.1 using these intervals, and deduce that $d \leq \frac{300r}{8 \cdot 29999} \leq \frac{r}{799}$. 

We now claim that, for any integer \( n \geq 0 \), if \( d \leq \frac{r}{799(101)^{n+1}} \), then \( d \leq \frac{r}{799(101)^{n+1}} \). We do this by applying Lemma 4.1 again, using \( d_{\text{max}} = \frac{r}{799(101)^{n}} \).

Let \( b \) be the endpoint of \( J'_{ij} \). Start off by considering the Type II interval of \( c_{k+1} \) which starts at \( b \). When we pass to \( c_{k+n+1} \), the first \( \frac{1}{101n} \) of this original interval will be a Type II interval oriented forwards. This interval is \( (b, b + \frac{r}{8(101)^{n+1}}) \). We then consider what happens in this interval in the configuration \( c_{k+n+3} \) (i.e., after applying the replacement rules two more times). We will have, among other things:

- a blue interval \( (b + \frac{r}{8(101)^{n+1}}, b + \frac{51r}{8(101)^{n+1}}) \), whose length \( \frac{50r}{8(101)^{n+1}} \) is greater than \( d_{\text{max}} \)
- a red interval \( (b + \frac{51r}{8(101)^{n+1}} + \frac{50r}{8(101)^{n+1}}, b + \frac{51r}{8(101)^{n+1}} + \frac{350r}{8(101)^{n+1}}) \), whose length \( \frac{50r}{8(101)^{n+1}} \) is also greater than \( d_{\text{max}} \).

Applying Lemma 4.1 with these intervals, we get that

\[
\begin{align*}
    d & \leq \frac{1}{29999} \left( \frac{51r}{8(101)^{n+1}} + \frac{50r}{8 \cdot 288(101)^{n+1}} \right) \leq \frac{r}{799(101)^{n+1}},
\end{align*}
\]

as required.

Since \( d \leq \frac{r}{799(101)^{n}} \) implies \( d \leq \frac{r}{799(101)^{n+1}} \) for all integers \( n \geq 0 \), and we already know \( d \leq \frac{r}{799} \), we must have that \( d = 0 \) contradicting the assumption that the \((3,30000)\)-AP was non-trivial. This concludes our proof that there is no \((3,30000)\)-AP in \( c'_{\infty} \).

Although it is clear that the red and blue densities in \( c_{\infty} \) are positive, and hence they will both have a lower bound when we go to \( \mathbb{Z}/p\mathbb{Z} \), let us calculate the densities to see what quantitative result our theorem produces.

Let \( r \) and \( b \) be the red and blue densities in \( c_{\infty} \). Let \( r_1 \) and \( b_1 \) be the red and blue densities in \( c_{\infty} \) in an interval that is a Type I uncoloured interval in some \( c_i \) (this clearly doesn’t depend on which interval we pick), and define \( r_2 \) and \( b_2 \) similarly. The initial configuration and replacement rules give us the following system of linear equations:

\[
\begin{align*}
    r &= \frac{1}{3} r_1 && \quad b = \frac{2}{3} + \frac{1}{3} b_1 \\
    r_1 &= \frac{1}{48} + \frac{2}{3} r_1 + \frac{1}{144} r_2 && \quad b_1 = \frac{11}{36} + \frac{2}{3} b_1 + \frac{1}{144} b_2 \\
    r_2 &= \frac{50}{101} r_1 + \frac{1}{101} r_2 && \quad b_2 = \frac{50}{101} + \frac{50}{101} b_1 + \frac{1}{101} b_2.
\end{align*}
\]
Solving gives

\begin{align*}
r &= \frac{2}{95} \\
1 &= \frac{6}{95} \\
r_2 &= \frac{3}{95}
\end{align*}

\begin{align*}
b &= \frac{93}{95} \\
b_1 &= \frac{92}{95} \\
b_2 &= \frac{92}{95}.
\end{align*}

Let \( r_p \) and \( b_p \) be the densities of red and blue points in the colourings of \( \mathbb{Z}/p\mathbb{Z} \) induced from \( c_\infty' \). By the same reasoning as in the previous section, \( r_p \) and \( b_p \) tend to \( r = \frac{2}{95} \) and \( b = \frac{93}{95} \), respectively. Therefore, we contradict the findability of \((3,30000)\) for all \( \delta < \frac{2}{95} \).

We end by remarking that, unlike our construction for \((14,14)\)-APs, this construction does not have density close to \( \frac{1}{2} \), so it is an open question whether there exists a \( k \) such that \((3,k)\) is not findable in colourings of density close to \( \frac{1}{2} \).

Acknowledgements. I would like to thank Ben Green for some helpful comments.

References

[1] N. Alon, J. Fox and Y. Zhao: Efficient arithmetic regularity and removal lemmas for induced bipartite patterns, *Discrete Analysis* 3, 2019.

[2] V. Bergelson and A. Leibman: Polynomial extensions of van der Waerden’s and Szemerédi’s theorems, *J. Amer. Math. Soc.* 9 (1996), 725–753.

[3] J. Cassaigne, J. D. Currie, L. Schaeffer and J. Shallit: Avoiding three consecutive blocks of the same size and same sum, *J. ACM* 61 (2014), Art. 10, 17.

[4] F. M. Dekking: Strongly nonrepetitive sequences and progression-free sets, *J. Combin. Theory Ser. A* 27 (1979), 181–185.

[5] H. Furstenberg and Y. Katznelson: An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.* 34 (1979), 275–291.

[6] E. Szemerédi: On sets of integers containing no \( k \) elements in arithmetic progression, *Acta Arith.* 27 (1975), 199–245. Collection of articles in memory of Juriǐ Vladimirovič Linnik.

Matei Mandache

*Department of Pure Mathematics and Mathematical Statistics*,
*University of Cambridge*

matei.mandache@gmail.com