SZEMERÉDI'S PROOF OF SZEMERÉDI'S THEOREM

TERENCE TAO

Abstract. In 1975, Szemerédi famously established that any set of integers of positive upper density contained arbitrarily long arithmetic progressions. The proof was extremely intricate but elementary, with the main tools needed being the van der Waerden theorem and a lemma now known as the Szemerédi regularity lemma, together with a delicate analysis (based ultimately on double counting arguments) of limiting densities of sets along multidimensional arithmetic progressions. In this note we present an arrangement of this proof that incorporates a number of notational and technical simplifications. Firstly, we replace the use of the regularity lemma by that of the simpler "weak regularity lemma" of Frieze and Kannan. Secondly, we extract the key inductive steps at the core of Szemerédi's proof (referred to as "Lemma 5", "Lemma 6", and "Fact 12" in that paper) as stand-alone theorems that can be stated with less notational setup than in the original proof, in particular involving only (families of) one-dimensional arithmetic progressions, as opposed to multidimensional arithmetic progressions. Thirdly, we abstract the analysis of limiting densities along the (now one-dimensional) arithmetic progressions by introducing the notion of a family of arithmetic progressions with the "double counting property".

We also present a simplified version of the argument that is capable of establishing Roth's theorem on arithmetic progressions of length three.

1. Introduction

In this paper we adopt the convention that the natural numbers $\mathbb{N} = \{1, 2, \dots\}$ begin at 1, rather than 0. For any natural number N, we let $[N]$ denote the initial segments[1]

$$[N] := \{n \in \mathbb{N} : n \leqslant N\} = \{1, \dots, N\},$$

with the convention that $[0]$ is the empty set. For any natural number K, we define a *length K arithmetic progression*, or *K-AP* for short, to be a K-tuple $\vec{P}$ of the form

$$\vec{P} = a + r \cdot \overrightarrow{[K]} := (a + kr)_{k \in [K]} = (a + r, a + 2r, \dots, a + Kr)$$

for some integer a and natural number r (thus in this paper arithmetic progressions are always strictly increasing). We abbreviate $a + r \cdot \overrightarrow{[K]}$ as $a + \overrightarrow{[K]}$ when $r = 1$, or $\overrightarrow{[K]}$ when $a = 0$ and $r = 1$; we will make a (very) slight distinction between the ordered K-tuple $\overrightarrow{[K]} = (1, \dots, K)$ and the unordered K-element set $[K] = \{1, \dots, K\}$. To each

[1] In general, we will try to use lower case Roman letters to denote elements of initial segments whose length is denoted by the corresponding upper case letter, thus for instance $n \in [N]$, $k \in [K]$, $l \in [L]$, $h \in [H]$, etc.. Sets (particularly sets of integers) will usually be denoted in boldface such as $\mathbf{A}, \mathbf{S}$; tuples (particularly arithmetic progressions) will be denoted by symbols with arrows, such as $\vec{P}$ or $\vec{R}$; and collections of such tuples will be denoted in calligraphic font such as $\mathcal{P}$ or $\mathcal{AP}$.

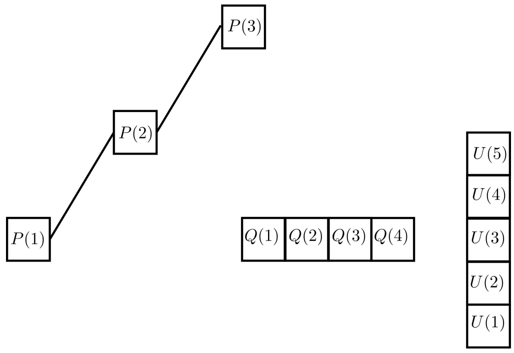

FIGURE 1. Two-dimensional depictions of a 3-AP $\vec{P} = (P(1), P(2), P(3))$, a 4-AP $\vec{Q} = (Q(1), Q(2), Q(3), Q(4))$, and a 5-AP $\vec{U} = (U(1), U(2), U(3), U(4), U(5))$. We will view the upwards and rightwards directions in such depictions as "positive", so that all three arithmetic progressions depicted here are increasing, in agreement with the definition of such progressions in this paper.

K-AP $\vec{P}$, we associate the increasing affine-linear function $P : \mathbb{Z} \to \mathbb{Z}$ defined by

$$P(k) := a + kr,$$

thus $\vec{P} = (P(1), \ldots, P(K))$. We say that a K-AP $\vec{P} = a + r \cdot \overrightarrow{[K]}$ is *contained* in a set of integers $\mathbf{A} \subset \mathbb{Z}$ if $a + kr \in \mathbf{A}$ for all $k \in [K]$, or equivalently if $P([K]) \subset \mathbf{A}$.

Arithmetic progressions lie in the integers, which are a one-dimensional set. However, when depicting such progressions, it will be more convenient to draw them as two-dimensional objects: see Figure 1.

In 1927, van der Waerden [11] proved:

Theorem 1.1 (Van der Waerden's theorem)**.** [11] *Let K, M be natural numbers, and let N be a natural number that is sufficiently large depending on K, M. If $[N]$ is partitioned into at most M colour classes, then one of the colour classes contains a K-AP.*

Equivalently: if the natural numbers are partitioned into finitely many colour classes, one of them will contain arbitrarily long arithmetic progressions.

A significant strengthening of van der Waerden's theorem was established by Szemerédi [9] in 1975. If $\mathbf{A}$ is a set of integers, then its *upper Banach density* $\overline{\mathrm{BD}}(\mathbf{A})$ is defined to be the quantity

$$\overline{\mathrm{BD}}(\mathbf{A}) := \limsup_{N \to \infty} \sup_{h \in \mathbb{Z}} \frac{|\mathbf{A} \cap (h + [N])|}{N}$$

where we use $|\mathbf{E}|$ to denote the cardinality of a finite set $\mathbf{E}$, and $h + \mathbf{E} = \{h + n : n \in \mathbf{E}\}$ to denote a translate of a set of integers $\mathbf{E}$ by a shift $h \in \mathbb{Z}$. In Szemerédi then showed

Theorem 1.2 (Szemerédi's theorem)**.** [9] *Let $\mathbf{A} \subset \mathbb{N}$ have positive upper Banach density. Then $\mathbf{A}$ contains a K-AP for every $K \in \mathbb{N}$.*

The $K = 1, 2$ cases of this theorem are trivial. The $K = 3$ case was established in 1953 by Roth [7] using Fourier-analytic methods. The $K = 4$ case was significantly more difficult, and was first established in 1969 by Szemerédi [8], prior to Szemerédi's 1975 proof in [9] of Theorem 1.2 in full generality. It is not difficult to show that Theorem 1.2 implies Theorem 1.1, but the converse implication is far less clear. One can replace the notion of upper Banach density here by other notions of density and still obtain an equivalent theorem; we leave the description of such variants to the interested reader.

In contrast to Roth's argument, Szemerédi's argument was purely combinatorial, and involved three main ingredients. The first is van der Waerden's theorem (Theorem 1.1). The second was a surprisingly delicate and recursive analysis of densities of sets of integers along certain arithmetic progressions (or higher dimensional analogues of arithmetic progressions). This analysis was quite technically involved, but was ultimately based on the classical combinatorial technique of *double counting*. The final ingredient was a lemma [9, Lemma 1] which, in its modern formulation, is now known as the *Szemerédi regularity lemma*; see e.g. [10].

In the original proof of Szemerédi, these three ingredients were interwoven together in a quite complicated manner (see in particular the diagram[2] on [9, p.202]). Since the original proof of this theorem, several quite different proofs of this theorem have been given; we mention in particular the ergodic theory proof of Furstenberg [2], the higher-order Fourier analytic proof of Gowers [3], and the hypergraph regularity proofs of Gowers [4] and Nagle-Rödl-Schacht-Skokan [5], [6]. Our focus here will however be on the original argument of Szemerédi.

In this paper we present an arrangement of Szemerédi's argument that contains a number of notational and technical simplifications. Firstly, we replace the use of the regularity lemma by a simpler "weak regularity lemma" of Frieze and Kannan [1], which we state here:

Lemma 1.3 (Weak regularity lemma)**.** [1] *Let* $\mathbf{V}, \mathbf{W}$ *be finite sets, let* $\varepsilon > 0$*, and let* $\mathbf{E} \subset \mathbf{V} \times \mathbf{W}$*. Then there exist partitions* $\mathbf{V} = \mathbf{V}_1 \cup \cdots \cup \mathbf{V}_A$ *and* $\mathbf{W} = \mathbf{W}_1 \cup \ldots \mathbf{W}_B$ *with* $A, B = O_\varepsilon(1)$ *and real numbers* $0 \leqslant d_{a,b} \leqslant 1$ *for* $a \in [A], b \in [B]$ *such that*

$$\left| |(\mathbf{F} \times \mathbf{G}) \cap \mathbf{E}| - \sum_{a \in [A]} \sum_{b \in [B]} d_{a,b} |\mathbf{F} \cap \mathbf{V}_a| |\mathbf{G} \cap \mathbf{W}_b| \right| \leqslant \varepsilon |\mathbf{V}||\mathbf{W}| \tag{1.1}$$

for all $\mathbf{F} \subset \mathbf{V}$ *and* $\mathbf{G} \subset \mathbf{W}$*.*

Here we use $O_\varepsilon(1)$ to denote an expression bounded in magnitude by a constant C_ε that depends only on ε. This lemma can be established as a corollary of the Szemerédi regularity lemma; however, by proving Lemma 1.3 directly one can obtain far better quantitative bounds on A, B than those provided by that regularity lemma (of exponential type in $1/\varepsilon$, rather than tower-exponential type). See [1] for further discussion. From a graph theory perspective, one can interpret the triplet $(\mathbf{V}, \mathbf{W}, \mathbf{E})$ as a bipartite

[2] Note that the arrow from Fact 12 to Lemma 6 in that diagram should be reversed.

graph connecting the two vertex sets $\mathbf{V}, \mathbf{W}$, but we will not use graph-theoretic terminology further in this paper. For sake of completeness, we provide a proof of Lemma 1.3 in Appendix A.

Actually, it will be convenient to use the following consequence of the weak regularity lemma, that allows one to approximately compute a large number of statistics $|\mathbf{F} \cap \mathbf{E}_w|$, $w \in \mathbf{W}$ using a much smaller number of statistics $|\mathbf{F} \cap \mathbf{V}_a|$, $a \in [A]$, where A can be significantly smaller than $\mathbf{W}$:

Corollary 1.4 (Consequence of weak regularity lemma). *Let* $\mathbf{V}, \mathbf{W}$ *be finite sets, let* $\varepsilon > 0$*, and for each* $w \in \mathbf{W}$*, let* $\mathbf{E}_w$ *be a subset of* $\mathbf{V}$*. Then there exist a partition* $\mathbf{V} = \mathbf{V}_1 \cup \cdots \cup \mathbf{V}_A$ *with* $A = O_\varepsilon(1)$*, and real numbers* $0 \leqslant c_{a,w} \leqslant 1$ *for* $a \in [A]$ *and* $w \in \mathbf{W}$*, such that for any set* $\mathbf{F} \subset \mathbf{V}$*, one has*

$$\left| |\mathbf{F} \cap \mathbf{E}_w| - \sum_{a \in [A]} c_{a,w} |\mathbf{F} \cap \mathbf{V}_a| \right| \leqslant \varepsilon |\mathbf{V}|$$

for all but $\varepsilon|\mathbf{W}|$ *values of* $w \in |\mathbf{W}|$.

Proof. We apply Lemma 1.3 with $\mathbf{E} := \{(v, w) \in \mathbf{V} \times \mathbf{W} : v \in \mathbf{E}_w\}$, and ε replaced by $\varepsilon^2/2$. This creates partitions $\mathbf{V} = \mathbf{V}_1 \cup \cdots \cup \mathbf{V}_A$ and $\mathbf{W} = \mathbf{W}_1 \cup \cdots \cup \mathbf{W}_B$ with $A, B = O_\varepsilon(1)$ and coefficients $0 \leqslant d_{a,b} \leqslant 1$ for $a \in [A]$, $b \in [B]$ such that

$$\left| |(\mathbf{F} \times \mathbf{G}) \cap \mathbf{E}| - \sum_{a \in [A]} \sum_{b \in [B]} d_{a,b} |\mathbf{F} \cap \mathbf{V}_a| |\mathbf{G} \cap \mathbf{W}_b| \right| \leqslant \frac{\varepsilon^2}{2} |\mathbf{V}||\mathbf{W}|$$

for all $\mathbf{F} \subset \mathbf{V}$ and $\mathbf{G} \subset \mathbf{W}$. If we define $c_{a,w}$ to equal $d_{a,b}$ whenever $a \in [A]$, $b \in [B]$, and $w \in \mathbf{W}_b$, then we can rearrange the left-hand side of the above inequality to obtain

$$\left| \sum_{w \in \mathbf{G}} \left(|\mathbf{F} \cap \mathbf{E}_w| - \sum_{a \in [A]} c_{a,w} |\mathbf{F} \cap \mathbf{V}_a| \right) \right| \leqslant \frac{\varepsilon^2}{2} |\mathbf{V}||\mathbf{W}|.$$

Applying this inequality with $\mathbf{G}$ equal to the set where the summand is positive (resp. negative) and summing, we conclude that

$$\sum_{w \in \mathbf{W}} \left| |\mathbf{F} \cap \mathbf{E}_w| - \sum_{a \in [A]} c_{a,w} |\mathbf{F} \cap \mathbf{V}_a| \right| \leqslant \varepsilon^2 |\mathbf{V}||\mathbf{W}|$$

for all $\mathbf{F} \subset \mathbf{V}$, and the claim now follows from Markov's inequality. $\square$

In this paper, the weak regularity lemma will be combined with van der Waerden's theorem to establish an important "mixing lemma" (see Theorem 4.1 below) which will then become a key ingredient in the proof of Szemerédi's theorem. Neither the regularity lemma nor van der Waerden's theorem will be used outside of the proof of this mixing lemma.

A prominent feature of the original argument of Szemerédi is the frequent use of multidimensional arithmetic progressions

$$(a + n_1 r_1 + \cdots + n_D r_D)_{(n_1,\dots,n_D)\in[N_1]\times\cdots\times[N_D]}$$

for relatively large values of dimension D (in fact to locate K-APs, one needs to use dimensions D of size $D = O(2^K)$). In particular, the three most important propositions in [9], which are labeled Lemma 5, Lemma 6, and Fact 12 in that paper, heavily involve these progressions, making them somewhat difficult to interpret as stand-alone propositions. The second simplification introduced in this paper is to "refactor" the argument so that the only progressions one encounters in the course of the argument are either one-dimensional arithmetic progressions

$$\vec{P} = (a + nr)_{n\in[N]}$$

or two-dimensional "rectangles"

$$\vec{R} = (a + hr + ls)_{(h,l)\in[H]\times[L]}.$$

In particular, the analogues of [9, Lemma 5, Lemma 6, Fact 12] in this paper (namely, Theorem 6.8, Proposition 6.9, and Claim 6.1) only involve (families of) one-dimensional arithmetic progressions, and which are somewhat easier to interpret in a stand-alone fashion. For instance, here are two such propositions which will play a key role in the $K = 3$ case (i.e. in proving Roth's theorem):

Theorem 1.5 ($C(3, \{2\})$)**.** *Let L be a natural number, and let $\mathbf{S}$ be a set of integers of upper Banach density at least $1 - \frac{1}{10L}$. Suppose that $\mathbf{S}$ is partitioned into finitely many colour classes. Then there exists a colour class $\mathbf{A} \subset \mathbf{S}$, together with a family $(\vec{P}_l)_{l\in[L]}$ of 3-APs $\vec{P}_l = (P_l(1), P_l(2), P_l(3))$ indexed by $l \in [L]$, obeying the following properties:*

(i) *For all $l \in [L]$, $\vec{P}_l$ is contained in $\mathbf{S}$.*
(ii) *For all $l \in [L]$, $P_l(1)$ lies in $\mathbf{A}$.*
(iv) *The tuple $(P_l(2))_{l\in[L]}$ is an L-AP.*

(The property (iii) turns out to be redundant in the $k = 3$ case and is thus omitted here; see Theorem 1.7 or Claim 6.1 below.) See Figure 2.

Theorem 1.6 ($C(3, \{3\})$)**.** *Let L be a natural number, and let $\mathbf{S}$ be a set of integers of upper Banach density at least $1 - \frac{1}{10L}$. Suppose that $\mathbf{S}$ is partitioned into finitely many colour classes. Then there exists a colour class $\mathbf{A} \subset \mathbf{S}$, together with a family $(\vec{P}_l)_{l\in[L]}$ of 3-APs $\vec{P}_l = (P_l(1), P_l(2), P_l(3))$ indexed by $l \in [L]$, obeying the following properties:*

(i) *For all $l \in [L]$, $\vec{P}_l$ is contained in $\mathbf{S}$.*
(ii) *For all $l \in [L]$, $P_l(1)$ and $P_l(2)$ lie in $\mathbf{A}$.*
(iv) *The tuple $(P_l(3))_{l\in[L]}$ is an L-AP.*

See Figure 3.

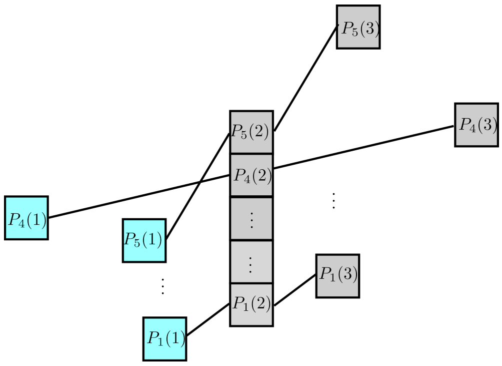

FIGURE 2. A configuration produced by Theorem 1.5, with $L = 5$ and the 3-APs $\vec{P}_2, \vec{P}_3$ omitted. Blue elements denote elements of $\mathbf{A}$; gray elements denote elements of $\mathbf{S}$ (which may potentially also lie in $\mathbf{A}$).

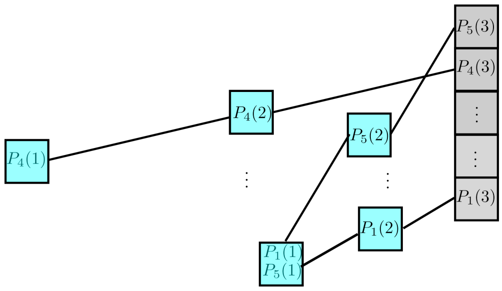

FIGURE 3. A configuration produced by Theorem 1.6, with $L = 5$ and the 3-APs $\vec{P}_2, \vec{P}_3$ omitted, and with the same colouring conventions as Figure 3. In this particular case we have a collision $P_1(1) = P_5(1)$, but such collisions are not prohibited by the theorem.

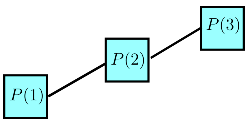

FIGURE 4. A configuration produced by Roth's theorem.

The notations $C(3, \{2\})$ and $C(3, \{3\})$ come from a more general claim $C(K, \Omega)$ that will be defined in Claim 6.1 below. Informally, one can think of $C(3, \{2\})$ as "one third" of Roth's theorem, in that it produces a family of 3-APs with just one element in a good set $\mathbf{A}$; similarly $C(3, \{3\})$ can be thought of as "two thirds" of Roth's theorem, in that it produces a family of 3-APs with two elements in such a good set. (Compare Figure 4 to Figures 2, 3.) A key technical difficulty here is that the set $\mathbf{S}$ has upper Banach density slightly less than one, rather than equal to 1; the claims would follow easily from van der Waerden's theorem in the latter case.

In Section 5 we give a self-contained proof of Roth's theorem, by first giving a short proof of Theorem 1.5, and then using that theorem (together with van der Waerden's theorem and the weak regularity lemma, together with double counting arguments) to prove Theorem 1.6, and finally using double counting arguments to derive Roth's

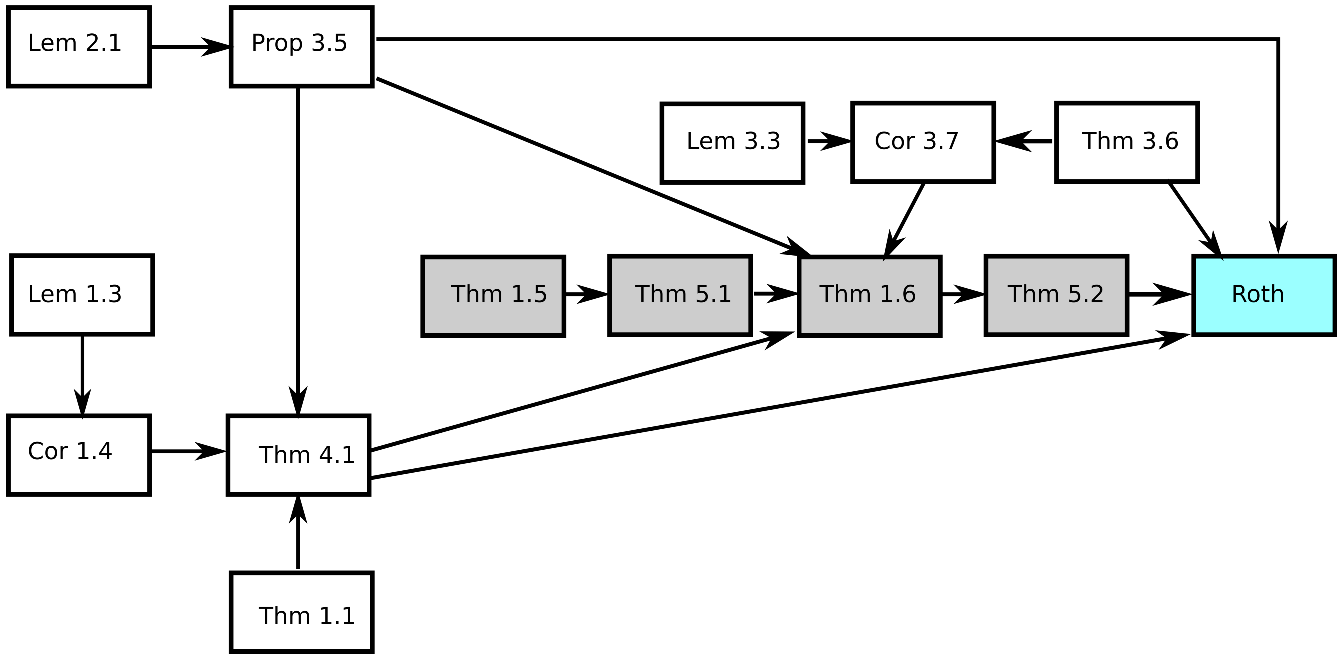

Figure 5. The logical structure of the proof of Roth's theorem (blue) in this paper. Key subtheorems are in gray; general-purpose tools are uncoloured. The structure of dependencies is actually slightly simpler than what is depicted here; for instance, the derivation of Roth's theorem from Proposition 3.5, Theorem 3.6, Theorem 5.2, and Theorem 4.1 does not use the full strength of Theorem 4.1, but rather the simpler component (i) of that theorem, which does not require Theorem 1.1 or Corollary 1.4 to prove.

theorem from Theorem 1.6; see Figure 5. This is by no means the shortest proof of Roth's theorem in the literature, but it serves as a warmup for the proof of the general case of Szemerédi's theorem, which has a slightly different top-level structure but uses essentially the same low-level ingredients; see Figure 6.

The third main simplification in this paper regards the analysis in [9, Section 3] of limiting densities of sets along arithmetic progression. A key technical difficulty arises from a distinction between (various notions of) upper density and density. For instance, suppose $\mathbf{A} \subset \mathbb{N}$ has upper Banach density $\overline{\mathrm{BD}}(\mathbf{A})$ equal to δ. Then there exists a sequence of intervals $h_n + [N_n]$ with $N_n \to \infty$ such that $\frac{|\mathbf{A} \cap (h_n + [N_n])|}{N_n}$ converges to δ. However, for all other intervals $h + [N]$, one only has an upper bound

$$\frac{|\mathbf{A} \cap (h + [N])|}{N} \leqslant \delta + o(1)$$

as $N \to \infty$, but not necessarily the matching lower bound

$$\frac{|\mathbf{A} \cap (h + [N])|}{N} \geqslant \delta - o(1).$$

In particular there can exist arbitrarily large intervals $h + [N]$ in which the density of A is far smaller than δ. In this particular case, one can eliminate this problem by

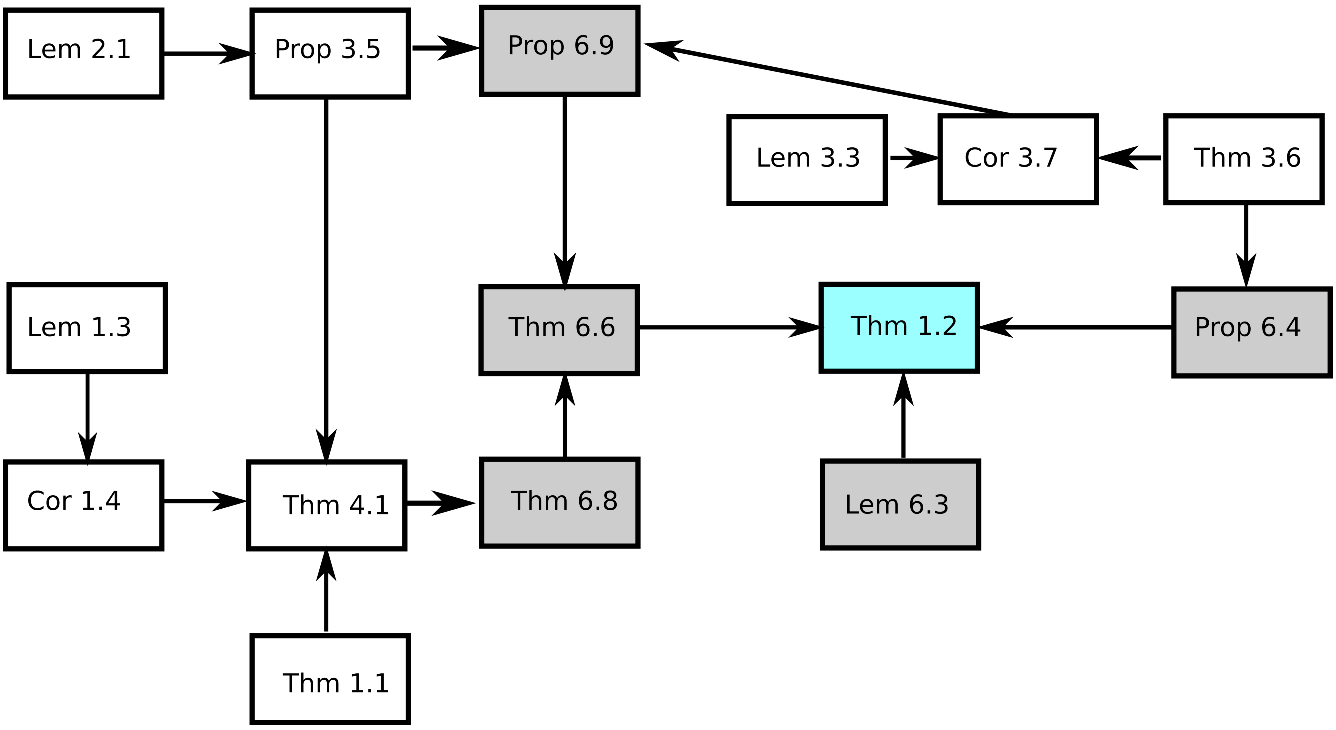

FIGURE 6. The logical structure of the proof of Szemerédi's theorem (blue) in this paper. Key subtheorems are in gray; general-purpose tools are uncoloured.

restricting the class $\{h + [N] : N \in \mathbb{N}\}$ of intervals one is measuring density on to just the subsequence $\{h_n + [N_n] : n \in \mathbb{N}\}$. The set $\mathbf{A}$ will now have a limiting density δ (and not just upper density δ) "along" this subsequence of intervals.

In the argument of Szemerédi [9], after using the previous simplifications to work with one-dimensional progressions, one is measuring upper densities or densities of sets $\mathbf{A}$ of integers along various collections $\mathcal{P}$ of arithmetic progressions $\vec{P}$ (which are not necessarily intervals $\overrightarrow{[N]}$). Analogously to the above discussion, if $\mathbf{A}$ has some upper density δ along such a family $\mathcal{P}$, it is not difficult to pass to a subfamily of progressions $\mathcal{P}'$ along which $\mathbf{A}$ has density δ. However, it is important in [9] that the resulting subfamily $\mathcal{P}'$ of progressions is compatible with a number of "double counting" arguments. In this paper we abstract this compatibility by introducing the concept of the *double counting property* for such a family of progressions $\mathcal{P}'$. We will define this property formally in Definition 3.4 below, but we describe just one typical consequence of this property: if one has a "$H \times N$-rectangle"

$$\vec{R} = (a + hr + ns)_{(h,n) \in [H] \times [N]}$$

for some integers a and natural numbers r, s, H, N, where N is much larger than H (which is in turn also assumed to be large), and all of the H "columns" $(a+hr+ns)_{n \in [N]}$, $h \in [H]$ lie in $\mathcal{P}'$, then almost all of the N "rows" $(a + hr + ns)_{h \in [H]}$, $n \in [N]$ also need to lie in $\mathcal{P}'$; see Proposition 3.5(ii) for a precise statement. The analysis in [9, Section

3] can then be abstracted into a standalone statement (Corollary 3.7 below) that does not require inputs such as van der Waerden's theorem or the regularity lemma to prove.

1.1. **Some ideas of the proof.** We now give an informal discussion of some of the ideas used in the arguments in this paper. We begin with a discussion of a derivation of Roth's theorem from Theorem 1.6, which is a relatively simple implication that nevertheless uses many of the key methods of the paper, omitting the formal details which may be found in Section 5.3. The first step is to observe (by a standard and elementary argument) that Theorem 1.6 implies a "bounded" variant, in which the set $\mathbf{S}$ now has density at least $1 - \frac{1}{10L}$ inside a single, sufficiently large interval $[N]$ (where N is large compared to the number of colours used), as opposed to being an infinite set of upper Banach density at least $1 - \frac{1}{10L}$; see Theorem 5.2 below for a precise statement.

Now let $\mathbf{A}$ be a set of integers with positive upper Banach density. Let δ be the upper density of $\mathbf{A}$ along *all* arithmetic progressions, that is to say

$$\delta := \limsup_{N\to\infty} \sup_{a\in\mathbb{Z}, r\in\mathbb{N}} \frac{1}{N}|\{n \in [N] : a + nr \in \mathbf{A}\}|.$$

Szemerédi's theorem (Theorem 1.2) implies that in fact δ is equal to one, but we cannot invoke that theorem currently as that would be circular; the best we can say for now is that $0 < \delta \leqslant 1$. By construction, $\mathbf{A}$ has density at most $\delta + o(1)$ along any arithmetic progression $\vec{P}$ as the length of that progression goes to infinity, and has density equal to $\delta + o(1)$ for some sequence of arithmetic progressions of length going to infinity.

We will informally[3] refer to a (long) arithmetic progression as "saturated" if $\mathbf{A}$ has density $\delta + o(1)$ along that progression. By a double counting argument, we will be able to find parameters $1 \leqslant L \leqslant H \leqslant N$ (with L large, H much larger than L, and N much larger than H or L), and a N-AP $\vec{U} = (U(1), \ldots, U(N))$, such that if one denotes $\mathbf{S}'$ to be the set of all $n \in [N]$ for which the H-AP $(U(n + h))_{h\in[H]}$ is saturated, then $\mathbf{S}'$ will have density at least $1 - \frac{1}{10L}$. Furthermore, one can colour $\mathbf{S}'$ by 2^H colours by assigning to each n the set $\{h \in [H] : U(n + h) \in \mathbf{A}\}$ as a colour. One can then apply (the bounded version of) Theorem 1.6 to conclude that there is a "perfect" colour class $\mathbf{A}'$ of $\mathbf{S}'$ and a family $(\vec{P}_l)_{l\in[L]}$ of 3-APs $\vec{P}_l$ such that for all $l \in [L]$, $P_l(1)$ and $P_l(2)$ lie in $\mathbf{A}'$ and $P_l(3)$ lie in $\mathbf{S}'$, while the L-tuple $(P_l(3))_{l\in[L]}$ forms an L-AP.

To prove Roth's theorem, we will restrict attention to 3-APs of the form

$$(U(P_l(1) + Q(1)), U(P_l(2) + Q(2)), U(P_l(3) + Q(3)))$$

where $l \in [L]$ and $\vec{Q} = (Q(1), Q(2), Q(3))$ is a 3-AP in $[H]$. It is easy to see that such 3-tuples are automatically 3-APs. Because $P_l(1)$ and $P_l(2)$ lie in the colour class $\mathbf{A}'$, we will have $U(P_l(1) + Q(1)), U(P_l(2) + Q(2)) \in \mathbf{A}$ precisely when $Q(1)$ and $Q(2)$ lie in the "perfect colour" $\mathbf{P} \subset [H]$ associated to the colour class $\mathbf{A}'$. Crucially, this class does not depend on the parameter l. If one then lets $\mathbf{E} \subset [H]$ denote the collection of all numbers

[3]To make this rigorous, one needs to carefully choose an explicit decay rate for $o(1)$; see the proof of Theorem 3.6. The terminology of "saturated" and "perfect" progressions is adapted from the original paper [9] of Szemerédi.

$Q(3)$, where $\vec{Q}$ is a 3-AP in H with $Q(1), Q(2) \in \mathbf{P}$, then it is not difficult to show that $\mathbf{E}$ has positive density in $[H]$ (with the bound depending only on δ). To finish locating a 3-AP in $\mathbf{A}$, it thus suffices to find $l \in [L]$ and $h \in \mathbf{E}$ such that $U(P_l(3) + h) \in \mathbf{A}$.

Now consider the $H \times L$-rectangle

$$\vec{R} := (U(P_l(3) + h))_{(h,l) \in [H] \times [L]}.$$

By construction, all of the L rows $(U(P_l(3) + h))_{h \in [H]}, l \in [L]$ of this rectangle are saturated. A double counting argument then implies that most of the H columns $(U(P_l(3) + h))_{l \in [L]}, h \in [H]$ of this rectangle are also saturated; in particular, we can find $h \in \mathbf{E}$ such that the associated column $(U(P_l(3)+h))_{l \in [L]}$ is saturated. In particular, this column will contain an element of $\mathbf{A}$, and we are done.

A key step in the above argument was the location of a "good" $l \in [L]$ for which the set

$$\{h \in \mathbf{E} : U(P_l(3) + h) \in \mathbf{A}\}$$

was well controlled (and in particular was non-empty). For the more general arguments below (and in particular when trying to derive Theorem 1.6 from Theorem 1.5), it will turn out that we will need an l which is "good" for multiple sets $\mathbf{E}_w \subset [H]$ simultaneously, where w ranges over some large finite index set $\mathbf{W}$. If the index set $\mathbf{W}$ is bounded, then it turns out that one can accomplish this by combining the above double counting arguments with van der Waerden's theorem (Theorem 1.1), roughly speaking because the latter theorem allows one to restrict to a long arithmetic progression of l's in which the behavior of each of the sets $\mathbf{E}_w$ "is constant in l". On its own, this van der Waerden argument is insufficient to treat the case when $\mathbf{W}$ is extremely large; however, if one combines the argument with the weak regularity lemma (or more precisely, Corollary 1.4) then one can treat this case also (at the cost of losing control of a tiny fraction of the $\mathbf{E}_w$), basically by using the regularity lemma to reduce back to the case of boundedly many sets. For the formal details of this argument, see Section 4.

The proof of Szemerédi's theorem (Theorem 1.2) in full generality proceeds along similar lines, except that the statements $C(3, \{2\})$ and $C(3, \{3\})$ that were formalised in Theorem 1.5 and Theorem 1.6 respectively must be replaced by a longer sequence of more complicated statements $C(K, \Omega)$ of this type, in which one works with progressions $\vec{P}_{\vec{l}}$ parameterised by a tuple $\vec{l} \in [L]^\Omega$ rather than a scalar $l \in [L]$. We present a typical such statement here; for the general case, see Claim 6.1 below.

Theorem 1.7 ($C(4, \{2, 4\})$). *Let L be a natural number, and let $\mathbf{S}$ be a set of integers of upper Banach density at least $1 - \varepsilon$, where $\varepsilon > 0$ is sufficiently small depending on L. Suppose that $\mathbf{S}$ is partitioned into finitely many colour classes. Then there exists a colour class $\mathbf{A} \subset \mathbf{S}$, together with a family $(\vec{P}_{l_2,l_4})_{(l_2,l_4) \in [L]^2}$ of 4-APs $\vec{P}_{l_2,l_4} = (P_{l_2,l_4}(1), P_{l_2,l_4}(2), P_{l_2,l_4}(3), P_{l_2,l_4}(4))$ indexed by a pair $(l_2, l_4) \in [L]^2$, obeying the following properties:*

(i) *For all $l_2, l_4 \in [L]$, $\vec{P}_{l_2,l_4}$ is contained in $\mathbf{S}$.*
(ii) *For all $l_2, l_4 \in [L]$, $P_{l_2,l_4}(1)$ lies in $\mathbf{A}$.*

(iii) *The colours of* $P_{l_2,l_4}(2)$ *and* $P_{l_2,l_4}(3)$ *are allowed to depend on* l_2*, but are independent of* l_4*, thus for instance* $P_{l_2,l_4}(3)$ *and* $P_{l_2,l_4'}(3)$ *have the same colour whenever* $l_2, l_4, l_4' \in [L]$.

(iv) *For any* $l_4 \in [L]$*, the tuple* $(P_{l_2,l_4}(2))_{l_2\in[L]}$ *is an* L*-AP. Similarly, for any* $l_2 \in [L]$*, the tuple* $(P_{l_2,l_4}(4))_{l_4\in[L]}$ *is an* L*-AP.*

The bulk of the proof of Theorem 1.2 is then concerned with taking statements $C(K,\Omega)$ (of which Theorem 1.7 above is typical) and using them to establish further statements $C(K,\Omega')$ of the same type. For instance, the statement $C(4,\{2,4\})$ above will be used to prove $C(4,\{1,2,4\})$, which will in turn be used to establish $C(4,\{3,4\})$, and so forth; the procedure here is analogous to that of incrementing a binary decimal (note that $2^{2-1} + 2^{4-1} = 10$ increments to $2^{1-1} + 2^{2-1} + 2^{4-1} = 11$, which in turn increments to $2^{3-1} + 2^{4-1} = 12$). In particular, the proof of Theorem 1.2 for K-APs will require $O(2^K)$ implications of this form. (In the case of Roth's theorem, we will use an *ad hoc* shortcut to get from $C(3,\{2\})$ directly to $C(3,\{3\})$, without the need to pass through an auxiliary statement $C(3,\{1,2\})$.)

1.2. **Acknowledgments.** This work originated from a working group with Shabnam Akhtari, Irfam Alam, Renling Jin, Steven Leth, Karl Mahlburg, Paul Potgieter, and Henry Towsner at the AIM workshop on "Nonstandard methods in combinatorial number theory" during Aug 14–18, 2017; it also draws on some earlier unpublished notes of the author at `www.math.ucla.edu/~tao/preprints/Expository/szemeredi_theorem.dvi` and `terrytao.wordpress.com/2012/03/23`. The author is indebted to the other members of the group for suggestions, encouragement, and much valuable discussion, and to Paul Epstein and the anonymous referee for a careful reading. The author is supported by NSF grant DMS-1266164, the James and Carol Collins Chair, and by a Simons Investigator Award.

2. Notation

As usual, we use the asymptotic notation $X = O(Y)$ to denote the estimate $|X| \leqslant CY$ for an absolute constant C. If instead we need the constant to depend on one or more parameters, we will denote this by subscripts, thus for instance $X = O_k(Y)$ denotes the estimate $|X| \leqslant C_k Y$ for some C_k depending on k.

A prominent feature of Szemerédi's arguments in [9] is the presence of a large number of parameters, most of which are assumed to be either very large or very small with respect to previously introduced parameters. To formalise this, it is convenient to use the following somewhat exotic asymptotic notation: when we write $X \lll Y$, we mean that Y is a real number that is sufficiently large depending on X and all free variables that occur to the left of X, thus we have $Y \geqslant F(X,\dots)$, where F is a suitable fixed function of X and all preceding free variables. We also write $X \lll \frac{1}{Y}$ to denote that Y is a sufficiently small positive real depending on X and all preceding free variables. We illustrate this notation with some examples involving a sequence $f_n : \mathbb{R} \to \mathbb{R}$, $n \in \mathbb{N}$ of functions:

- The functions f_n are individually continuous iff, for every $n \in \mathbb{N}$, every $x \in \mathbb{R}$ and all $0 < \frac{1}{\varepsilon} \ll \frac{1}{\delta}$, one has $|f_n(y) - f_n(x)| \leqslant \varepsilon$ whenever $y \in \mathbb{R}$ with $|y - x| \leqslant \delta$.
- The functions f_n are individually uniformly continuous iff for every $n \in \mathbb{N}$ and for all $0 < \frac{1}{\varepsilon} \ll \frac{1}{\delta}$, one has $|f_n(y) - f_n(x)| \leqslant \varepsilon$ whenever $x, y \in \mathbb{R}$ with $|y - x| \leqslant \delta$.
- The sequence f_n is equicontinuous iff, for every $x \in \mathbb{R}$ and all $0 < \frac{1}{\varepsilon} \ll \frac{1}{\delta}$, one has $|f_n(y) - f_n(x)| \leqslant \varepsilon$ whenever $n \in \mathbb{N}$ and $y \in \mathbb{R}$ with $|y - x| \leqslant \delta$.
- The sequence f_n is uniformly equicontinuous iff, for all $0 < \frac{1}{\varepsilon} \ll \frac{1}{\delta}$, one has $|f_n(y) - f_n(x)| \leqslant \varepsilon$ whenever $n \in \mathbb{N}$ and $x, y \in \mathbb{R}$ with $|y - x| \leqslant \delta$.

We will often concatenate several instances of above asymptotic notation; for instance, if we write $X, Y \ll Z \ll W$, this means that Z is sufficiently large depending on X, Y and preceding free variables, while W is sufficiently large depending on X, Y, Z, and preceding free variables.

If $\vec{P}$ is a K-AP, we write $|\vec{P}| = K$ to denote its length, and use $\mathcal{AP}$ to denote the collection of all arithmetic progressions $\vec{P}$ (of any length). If $\vec{P}$ is a K-AP, we define the *uniform probability measure* $\mu_{\vec{P}}$ on $\vec{P}$ to be the measure on $\mathbb{Z}$ defined by the formula

$$\mu_{\vec{P}}(\mathbf{A}) := \frac{1}{K}|\{i \in [K] : P(i) \in \mathbf{A}\}|$$

for all $\mathbf{A} \subset \mathbb{Z}$. As the name suggests, $\mu_{\vec{P}}$ is nothing more than the uniform probability measure on $\{P(1), \ldots, P(K)\}$.

Let L, H be natural numbers. We define a *rectangle of length L and height H*, or a *$L \times H$-rectangle* for short, to be a tuple of the form

$$\vec{R} = (a + lr + hs)_{(l,h)\in[L]\times[H]}$$

with $a \in \mathbb{Z}$ and $r, s \in \mathbb{N}$. Note that we do not insist that the LH integers $a + lr + hs$ for $(l, h) \in [L] \times [H]$ are all distinct, although this will automatically be the case if we hold l fixed, or if we hold h fixed. As with progressions, the rectangle $\vec{R}$ induces an affine-linear increasing function $R : \mathbb{Z}^2 \to \mathbb{Z}$ from $\mathbb{Z}^2$ (with the product ordering) to $\mathbb{Z}$, defined by

$$R(l, h) := a + lr + hs$$

for $(l, h) \in \mathbb{Z}$. We define the *rows* $\vec{R}_h$, $h \in [H]$ of the rectangle $\vec{R}$ to be the L-APs

$$\vec{R}_h := (R(l, h))_{l\in[L]}$$

and similarly define the *columns* $\vec{R}^l$, $l \in [L]$ of the rectangle to be the H-APs

$$\vec{R}^l := (R(l, h))_{h\in[H]};$$

see Figure 7. We define the *uniform probability measure* $\mu_{\vec{R}}$ on the rectangle $\vec{R}$ by the formula

$$\mu_{\vec{R}}(\mathbf{A}) := \frac{1}{LH}|\{(l, h) \in [L] \times [H] : R(l, h) \in \mathbf{A}\}|.$$

We caution that this is the uniform measure on the *multi-set* $\{R(l, h) : (l, h) \in [L]\times[H]\}$, in which one takes into account the possible multiplicity of representation of a given

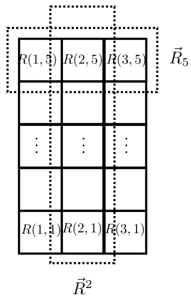

FIGURE 7. A rectangle $\vec{R}$ of length $L = 3$ and height $H = 5$ and a selection of its entries, together with the row $\vec{R}_5$ and column $\vec{R}^2$.

integer in the form $R(l, h)$ for $(l, h) \in [L] \times [H]$. We observe the basic *double counting identity*

$$\mu_{\vec{R}} = \frac{1}{H} \sum_{h \in [H]} \mu_{\vec{R}_h} = \frac{1}{L} \sum_{l \in [L]} \mu_{\vec{R}^l}. \tag{2.1}$$

Given two probability measures μ, ν on the integers $\mathbb{Z}$, we define the *total variation distance* $d_{TV}(\mu, \nu)$ between them by the formula

$$d_{TV}(\mu, \nu) := \sup_{\mathbf{A} \subset \mathbb{Z}} |\mu(\mathbf{A}) - \nu(\mathbf{A})|. \tag{2.2}$$

We have the following simple but important computation, that allows one to approximate the uniform distribution on a long arithmetic progression by the uniform distribution on a rectangle with one long side and one much shorter side:

Lemma 2.1. *Let $1 \leqslant H \leqslant N$ be natural numbers, and let $\vec{P}$ be an N-AP. Let $\vec{R}$ be the $H \times N$-rectangle*

$$\vec{R} := (P(i + j))_{(j,i) \in [H] \times [N]}.$$

Then $d_{TV}(\mu_P, \mu_R) \leqslant 2\frac{H}{N}$.

Proof. Let $\mathbf{A} \subset \mathbb{Z}$. Then

$$\mu_{\vec{R}}(\mathbf{A}) = \frac{1}{NH} \sum_{h \in [H]} |\{n \in [N] + h : P(n) \in \mathbf{A}\}|$$

and hence

$$\mu_{\vec{R}}(\mathbf{A}) - \mu_{\vec{P}}(\mathbf{A}) = \frac{1}{NH} \sum_{h \in [H]} (|\{n \in [N] + h : P(n) \in \mathbf{A}\}| - |\{n \in [N] : P(n) \in \mathbf{A}\}|).$$

But $[N]$ and $[N] + h$ differ by at most $2H$ elements, giving the claim. $\square$

3. Density along progressions

Let $\mathcal{P} \subset \mathcal{AP}$ be a collection of arithmetic progressions. We say that $\mathcal{P}$ is *unbounded* if the set $\{|\vec{P}| : \vec{P} \in \mathcal{P}\}$ is unbounded, that is to say $\mathcal{P}$ contains arithmetic progressions of arbitrarily large length. If $\mathcal{P}$ is unbounded and $\mathbf{A} \subset \mathbb{Z}$, we define the *upper density* $\overline{d}_{\mathcal{P}}(\mathbf{A})$ of $\mathbf{A}$ along $\mathcal{P}$ by the formula

$$\overline{d}_{\mathcal{P}}(\mathbf{A}) := \limsup_{|\vec{P}| \to \infty : \vec{P} \in \mathcal{P}} \mu_{\vec{P}}(\mathbf{A}) = \inf_{N \in \mathbb{Z}} \sup_{\vec{P} \in \mathcal{P} : |\vec{P}| \geqslant N} \mu_{\vec{P}}(\mathbf{A})$$

and similarly define the *lower density*

$$\underline{d}_{\mathcal{P}}(\mathbf{A}) := \liminf_{|\vec{P}| \to \infty : \vec{P} \in \mathcal{P}} \mu_{\vec{P}}(\mathbf{A}) = \sup_{N \in \mathbb{Z}} \inf_{\vec{P} \in \mathcal{P} : |\vec{P}| \geqslant N} \mu_{\vec{P}}(\mathbf{A}).$$

Clearly one has

$$0 \leqslant \underline{d}_{\mathcal{P}}(\mathbf{A}) \leqslant \overline{d}_{\mathcal{P}}(\mathbf{A}) \leqslant 1.$$

If we have $\underline{d}_{\mathcal{P}}(\mathbf{A}) = \overline{d}_{\mathcal{P}}(\mathbf{A})$, then we denote this quantity by $d_{\mathcal{P}}(\mathbf{A})$, and say that A has *density* $d_{\mathcal{P}}(\mathbf{A})$ along $\mathcal{P}$. Otherwise we say that $\mathbf{A}$ does not have a density along $\mathcal{P}$.

Examples 3.1. If $\mathcal{P}$ is the collection of initial segments $\overrightarrow{[N]}$, $N \in \mathbb{N}$, and $\mathbf{A}$ is a subset of $\mathbb{N}$, then $\overline{d}_{\mathcal{P}}(\mathbf{A})$, $\underline{d}_{\mathcal{P}}(\mathbf{A})$, and (if it exists) $d_{\mathcal{P}}(\mathbf{A})$ are the upper natural density, lower natural density, and natural density of $\mathbf{A}$ respectively. Similarly, if $\mathcal{P}$ is the collection of shifted intervals $a + \overrightarrow{[N]}$ with $a \in \mathbb{Z}$ and $N \in \mathbb{N}$, and $\mathbf{A} \subset \mathbb{Z}$, then $\overline{d}_{\mathcal{P}}(\mathbf{A})$ is the upper Banach density $\overline{\mathrm{BD}}(\mathbf{A})$ of $\mathbf{A}$. One could make a similar remark for lower Banach density or Banach density, but we will not use these concepts in this paper.

Remark 3.2. Szemerédi's theorem (Theorem 1.2) implies (and is in fact equivalent to) the assertion that any set of positive upper Banach density will have density 1 along $\mathcal{AP}$. Unfortunately, we are not allowed to use this fact in the proof of Theorem 1.2, as it will be circular! Nevertheless, this observation does imply that many of the results in this paper become rather trivial to prove once one is permitted to invoke Szemerédi's theorem.

One can interpret upper density or density using the asymptotic notation from the previous section. Indeed, if $\mathcal{P}$ is an unbounded family of progressions and $\mathbf{A} \subset \mathbb{Z}$, then whenever one has parameters

$$1 \leqslant \frac{1}{\varepsilon} \ll N$$

then one has

$$\mu_{\vec{P}}(\mathbf{A}) \leqslant \overline{d}_{\mathcal{P}}(\mathbf{A}) + \varepsilon \tag{3.1}$$

whenever $\vec{P}$ is an arithmetic progression in $\mathcal{P}$ of length at least N; furthermore, we have the stronger claim

$$|\mu_{\vec{P}}(\mathbf{A}) - \overline{d}_{\mathcal{P}}(\mathbf{A})| \leqslant \varepsilon \tag{3.2}$$

for at least one such progression $\vec{P}$. If $\mathbf{A}$ has a density along $\mathcal{P}$, then in fact we have

$$|\mu_{\vec{P}}(\mathbf{A}) - d_{\mathcal{P}}(\mathbf{A})| \leqslant \varepsilon \tag{3.3}$$

for *all* progressions $\vec{P}$ in $\mathcal{P}$ of length at least N.

It is easy to verify the subadditivity property

$$\overline{d}_{\mathcal{P}}(\mathbf{A} \cup \mathbf{B}) \leqslant \overline{d}_{\mathcal{P}}(\mathbf{A}) + \overline{d}_{\mathcal{P}}(\mathbf{B}) \tag{3.4}$$

of upper density whenever $\mathbf{A}, \mathbf{B} \subset \mathbb{Z}$ and $\mathcal{P}$ is an unbounded family of progressions. Iterating this, we immediately conclude

Lemma 3.3 (Pigeonhole principle for upper density)**.** *Let $\mathcal{P} \subset \mathcal{AP}$ be an unbounded family of arithmetic progressions. If $\mathbf{S} \subset \mathbb{Z}$ has positive upper density along $\mathcal{P}$, and $\mathbf{S}$ is partitioned into finitely many colour classes, then at least one of the colour classes will also have positive upper density along $\mathcal{P}$.*

We now come to a key property of certain families of arithmetic progressions.

Definition 3.4 (Double counting property)**.** A collection $\mathcal{P} \subset \mathcal{AP}$ of arithmetic progressions is said to have the *double counting property* if, whenever one has

$$1 \leqslant \frac{1}{\varepsilon} \lll L_1 \lll L_2, \tag{3.5}$$

then whenever $\vec{R}_1$ is an $L_1 \times H_1$-rectangle for some H_1, and $\vec{R}_2$ is an $L_2 \times H_2$-rectangle for some H_2, with the property that

$$d_{TV}(\mu_{\vec{R}_1}, \mu_{\vec{R}_2}) \leqslant \frac{1}{L_1}, \tag{3.6}$$

and such that all the H_2 rows $(\vec{R}_2)_i, i \in [H_2]$ of $\vec{R}_2$ lie in $\mathcal{P}$, then all but at most εH_1 of the H_1 rows $(\vec{R}_1)_i, i \in [H_1]$ of $\vec{R}_1$ lie in $\mathcal{P}$.

Informally, the double counting property asserts that if an arithmetic progression of L_2-APs $(\vec{R}_2)_i$ are in $\mathcal{P}$, and one "rearranges" these progressions (up to a small error) into a progression of much shorter L_1-APs $(\vec{R}_1)_i$, then most of these shorter progressions will also lie in $\mathcal{P}$.

Our main application of the double counting property will proceed via the following proposition:

Proposition 3.5. *Let $\mathcal{P} \subset \mathcal{AP}$ be a family of arithmetic progressions with the double counting property. Select parameters*

$$1 \leqslant \frac{1}{\varepsilon} \lll H \lll N.$$

(i) *If $\vec{P}$ is an N-AP in $\mathcal{P}$, then for all but at most εN of the elements n of $[N]$, the H-AP*
$$(P(n+h))_{h \in [H]}$$
lies in $\mathcal{P}$.

(ii) *If $\vec{R}$ is a $H \times N$-rectangle such that all H columns $\vec{R}^h$, $h \in [H]$ of $\vec{R}$ lie in $\mathcal{P}$, then all but at most εN of the rows $\vec{R}_n, n \in [N]$ lie in $\mathcal{P}$.*

Proof. To prove (i), we apply Definition 3.4 with R_1 being the $H \times N$-rectangle

$$R_1 := (P(n+h))_{(h,n) \in [H] \times [N]}$$

and R_2 being the $N \times 1$-rectangle

$$R_2 := (P(n))_{(n,1) \in [N] \times [1]}.$$

The claim then follows from Lemma 2.1.

To prove (ii), we apply Definition 3.4 with $\vec{R}_1$ being the $H \times N$-rectangle $\vec{R}$, and $\vec{R}_2$ being the $N \times H$-rectangle

$$\vec{R}_2 := (R(h,n))_{(n,h)\in[N]\times[H]}$$

which can be thought of as a "transpose" of $\vec{R}$. The claim then follows from (2.1). □

Clearly $\mathcal{AP}$ has the double counting property. The main result of this section is that one can effectively upgrade upper density to density while retaining the double counting property:

Theorem 3.6 (Upgrading upper density to density)**.** *Let $\mathcal{P} \subset \mathcal{AP}$ be an unbounded family of progressions with the double counting property. Let $\mathbf{A} \subset \mathbb{Z}$ have upper density δ along $\mathcal{P}$ for some $0 \leqslant \delta \leqslant 1$. Then there exists a unbounded subfamily $\mathcal{P}' \subset \mathcal{P}$ of $\mathcal{P}$ with the double counting property, such that $\mathbf{A}$ has density δ along $\mathcal{P}'$.*

Proof. We need a decreasing sequence $c(N) > 0$ of positive numbers that goes to zero as $N \to \infty$ in a sufficiently slow fashion; this will be specified more later. We define $\mathcal{P}'$ to be the set of all arithmetic progressions $\vec{P} \in \mathcal{P}$ such that

$$|\mu_{\vec{P}}(\mathbf{A}) - \delta| \leqslant c(|\vec{P}|). \tag{3.7}$$

(In the language of Szemerédi's paper [9], $\mathcal{P}'$ consists of those arithmetic progressions in $\mathcal{P}$ that are "saturated" with respect to $\mathbf{A}$.)

By (3.2), we see that for any $\varepsilon > 0$, there exist arithmetic progressions $\vec{P} \in \mathcal{P}$ of arbitrarily large length with $|\mu_{\vec{P}}(\mathbf{A}) - \delta| \leqslant \varepsilon$. Thus, if the sequence c is sufficiently slowly decaying, $\mathcal{P}'$ will also contain arithmetic progressions of arbitrarily large length, and will thus be unbounded. As c goes to zero, we see from (3.7) that $\mathbf{A}$ has density δ along $\mathcal{P}'$.

It remains to show that $\mathcal{P}'$ has the double counting property. Let $\varepsilon, L_1, L_2, H_1, H_2, \vec{R}_1, \vec{R}_2$ be as in Definition 3.4. For the purposes of (3.5), we consider the family $\mathcal{P}$, the function c, and the set $\mathbf{A}$ to be free variables; thus for instance, L_2 is assumed sufficiently large depending on $L_1, \varepsilon, c, \mathcal{P}, \mathbf{A}$.

From (3.7) and (3.5) we have

$$\mu_{(\vec{R}_2)_h}(\mathbf{A}) \geqslant \delta - \frac{1}{L_1}$$

for all $h \in [H_2]$. Averaging in h using (2.1), we conclude that

$$\mu_{\vec{R}_2}(\mathbf{A}) \geqslant \delta - \frac{1}{L_1},$$

and thus by (3.6), (2.2) and the triangle inequality, we have

$$\mu_{\vec{R}_1}(\mathbf{A}) \geqslant \delta - \frac{2}{L_1}.$$

Applying (2.1) again, we obtain

$$\sum_{h\in[H_1]} \mu_{(\vec{R}_1)_h}(A) \geqslant \left(\delta - \frac{2}{L_1}\right) H_1.$$

As $\mathcal{P}$ had the double counting property, we see that all but $\eta(L_1)H_1$ of the rows $(\vec{R}_1)_h$, $h \in [H_1]$ lie in $\mathcal{P}$, where $\eta(L_1)$ is a quantity that goes to zero as $L_1 \to \infty$, but does not depend on c. Thus we have

$$\sum_{h\in[H_1]:(\vec{R}_1)_h\in\mathcal{P}} \mu_{(\vec{R}_1)_h}(\mathbf{A}) \geqslant \left(\delta - \frac{2}{L_1} - \eta(L_1)\right) H_1. \tag{3.8}$$

Next, from (3.1) one has

$$\mu_{(\vec{R}_1)_h}(\mathbf{A}) \leqslant \delta + \eta'(L_1)$$

for all h in the above sum, where $\eta'(L_1)$ is a quantity that goes to zero as $L_1 \to \infty$, but does not depend on c. We can then rewrite (3.8) as

$$\sum_{h\in[H_1]:(\vec{R}_1)_h\in\mathcal{P}} (\delta + \eta'(L_1) - \mu_{(R_1)_h}(\mathbf{A})) \leqslant \left(\frac{2}{L_1} + \eta(L_1) + \eta'(L_1)\right) H_1.$$

and hence by Markov's inequality, one will have

$$0 \leqslant \delta + \eta'(L_1) - \mu_{(\vec{R}_1)_h}(\mathbf{A}) \leqslant \frac{2}{\varepsilon}\left(\frac{2}{L_1} + \eta(L_1) + \eta'(L_1)\right) \tag{3.9}$$

for all but at most $\varepsilon H_1/2$ of the $h \in [H_1]$ with $(\vec{R}_1)_h \in \mathcal{P}$. For c sufficiently slowly growing, we then have from (3.5) that

$$|\mu_{(R_1)_h}(\mathbf{A}) - \delta| \leqslant c(L_1)$$

for all but at most εH_1 of the $h \in [H_1]$, and the claim follows. □

We will rely primarily on the following corollary of Theorem 3.6, which roughly corresponds to the content of Section 3 of Szemerédi's original paper [9].

Corollary 3.7 (Saturated progressions and the perfect colour)**.** *Let $\mathbf{S} \subset \mathbb{Z}$ be a set of integers with a positive upper density $\sigma > 0$ along $\mathcal{AP}$. Suppose that there is a colouring $c : \mathbf{S} \to \mathbf{C}$ of $\mathbf{S}$ by a finite collection $\mathbf{C}$ of colours. Then there exists a "perfect" colour $p \in \mathbf{C}$, with associated colour class $\mathbf{A} := \{s \in \mathbf{S} : c(s) = p\}$, as well as an unbounded family $\mathcal{P} \subset \mathcal{AP}$ of "saturated" arithmetic progressions with the double counting property, such that $\mathbf{A}$ has a positive density $\delta > 0$ along $\mathcal{P}$, and additionally $\mathbf{S}$ has density σ along $\mathcal{P}$.*

Proof. Applying Theorem 3.6 with $\mathcal{P}$ replaced by $\mathcal{AP}$ and $\mathbf{A}$ replaced by $\mathbf{S}$, we can find an unbounded family $\mathcal{P}' \subset \mathcal{AP}$ of arithmetic progressions with the double counting property, such that $\mathbf{S}$ has density σ along $\mathcal{P}'$. By Lemma 3.3, there exists a colour $p \in \mathbf{C}$ whose associated colour class $\mathbf{A}$ has a positive upper density $\delta > 0$ along $\mathcal{P}'$. Applying Theorem 3.6 a second time (with $\mathcal{P}$ replaced by $\mathcal{P}'$), we can find an unbounded subfamily $\mathcal{P} \subset \mathcal{P}'$ of $\mathcal{P}'$ with the double counting property, such that $\mathbf{A}$ has density δ along $\mathcal{P}$. The density σ of $\mathbf{S}$ along $\mathcal{P}'$ is of course inherited by the unbounded subfamily $\mathcal{P}$, and the claim follows. □

4. Using Van der Waerden's theorem and the regularity lemma

To motivate the main results of this section we begin with an informal discussion. Suppose one has an $L \times H$-rectangle $\vec{R}$ (with $1 \ll L \ll H$), and suppose that $\mathbf{A}$ is a set of integers that has density close to δ on all the columns $\vec{R}^l, l \in [L]$ of $\vec{R}$, thus for each $l \in [L]$ one has

$$|\{h \in [H] : R(l,h) \in \mathbf{A}\}| \approx \delta H.$$

If $\mathbf{E}$ is a given subset of $[H]$, then if $\mathbf{A}$ was suitably "mixing" in nature, one may then expect to have

$$|\{h \in \mathbf{E} : R(l,h) \in \mathbf{A}\}| \approx \delta |\mathbf{E}|$$

for a "typical" choice of $l \in [L]$. In general, one would not expect this sort of claim to be true for any given set $\mathbf{E}$, let alone for a large family $\mathbf{E}_w$ of such sets. However, it turns out that thanks to van der Waerden's theorem (Theorem 1.1) with the weak regularity lemma (in the form of Corollary 1.4), one can obtain a useful result of this type for at least one choice of $l \in [L]$, if the rows $\vec{R}^i$ of the rectangle belong to an unbounded family of arithmetic progressions with the double counting property, and along which $\mathbf{A}$ has density δ. More precisely, in this section we prove

Theorem 4.1 (Mixing lemma). *Let $\mathcal{P}$ be an unbounded family of arithmetic progressions with the double counting property, and let $\mathbf{A} \subset \mathbb{Z}$ have a density δ along $\mathcal{P}$. Let*

$$1 \leqslant \frac{1}{\varepsilon} \ll L \ll H,$$

and let $\vec{R}$ be an $L \times H$-rectangle such that all the columns $\vec{R}^l, l \in [L]$ lie in $\mathcal{P}$.

(i) *(Single mixing) If $\mathbf{E} \subset [H]$, then there exists $l \in [L]$ such that*

$$|\{h \in \mathbf{E} : R(l,h) \in \mathbf{A}\}| \geqslant \delta |\mathbf{E}| - \varepsilon H.$$

(ii) *(Multiple mixing) If A is a natural number with $A \ll L$, and $\mathbf{E}_1, \ldots, \mathbf{E}_A \subset [H]$, then there exists $l \in [L]$ such that*

$$||\{h \in \mathbf{E}_a : R(l,h) \in \mathbf{A}\}| - \delta |\mathbf{E}_a|| \leqslant \varepsilon H$$

for all $a \in [A]$.

(iii) *(Highly multiple mixing) If $(\mathbf{E}_w)_{w \in \mathbf{W}}$ is a family of subsets $\mathbf{E}_w \subset [H]$ of $[H]$ indexed by some finite set $\mathbf{W}$, then there exists $l \in [L]$ such that*

$$||\{h \in \mathbf{E}_w : R(l,h) \in \mathbf{A}\}| - \delta |\mathbf{E}_w|| \leqslant \varepsilon H$$

for all but at most $\varepsilon |\mathbf{W}|$ of the $w \in \mathbf{W}$.

A key point in (iii) (as compared to (ii)) is that there is no upper bound on the cardinality of $\mathbf{W}$. The ability to make assertions that are uniform in such "vertex sets" $\mathbf{W}$, and thus pass from (ii) to (iii), is entirely thanks to the (weak) regularity lemma. Meanwhile, the ability to pass from assertions (i) about a single set, to assertions (ii) about multiple sets, is entirely thanks to the van der Waerden theorem. This mixing lemma corresponds (very roughly) to the proof of [9, Lemma 4] and portions of [9, Lemma 5].

Proof. We begin with (i), which is proven by a standard double-counting argument. By Proposition 3.5(ii), all but $\frac{\varepsilon}{2}H$ of the rows $\vec{R}_h$, $h \in [H]$ lie in $\mathcal{P}$. By (3.3), one has $\mu_{\vec{R}_h}(\mathbf{A}) \geqslant \delta - \frac{\varepsilon}{2}$ for all such rows. In particular, for all but at most $\frac{\varepsilon}{2}H$ of the $h \in \mathbf{E}$, we have

$$|\{l \in [L] : R(l,h) \in \mathbf{A}\}| \geqslant \left(\delta - \frac{\varepsilon}{2}\right) L$$

and hence on summing over these h and then double counting,

$$\sum_{l \in [L]} |\{h \in \mathbf{E} : R(l,h) \in \mathbf{A}\}| \geqslant \delta L |\mathbf{E}| - \varepsilon L H.$$

By the pigeonhole principle, we can thus find $l \in [L]$ such that

$$|\{h \in \mathbf{E} : R(l,h) \in \mathbf{A}\}| \geqslant \delta |\mathbf{E}| - \varepsilon H,$$

giving (i).

Now we prove (ii). We may assume that there is a natural number L' with

$$\frac{1}{\varepsilon}, A \ll L' \ll L.$$

We can assign to each $l \in [L]$ a colour $c(l) = (c_a(l))_{a \in [A]} \in \{-1, 0, +1\}^A$, where for any $a \in [A]$, $c_a(l)$ is defined to equal -1 if

$$|\{h \in \mathbf{E}_a : R(l,h) \in \mathbf{A}\}| < \delta |\mathbf{E}_a| - \varepsilon H,$$

to equal $+1$ if

$$|\{h \in \mathbf{E}_a : R(l,h) \in \mathbf{A}\}| > \delta |\mathbf{E}_a| + \varepsilon H,$$

and to equal 0 if

$$||\{h \in \mathbf{E}_a : R(l,h) \in \mathbf{A}\}| - \delta |\mathbf{E}_a|| \leqslant \varepsilon H.$$

The number of colour classes here is 3^A, which is small compared with L' or L. By Theorem 1.1, there exists an L'-AP $\vec{Q}$ in $[L]$ and a colour $c = (c_a)_{a \in [A]} \in \{-1, 0, +1\}^A$ such that $c(Q(l')) = c$ for all $l' \in [L']$.

Suppose that $c_a = -1$ for some $a \in [A]$. Then we have

$$|\{h \in \mathbf{E}_a : R(Q(l'),h) \in \mathbf{A}\}| < \delta |\mathbf{E}_a| - \varepsilon H$$

for all $l' \in [L']$. But then the $L' \times H$-rectangle

$$(R(Q(l'),h))_{(l',h) \in [L'] \times [H]}$$

contradicts part (i) (with $\mathbf{E}$ replaced by $\mathbf{E}_a$). If instead $c_a = +1$ for some $a \in [A]$, then we have for all $l' \in [L']$ that

$$|\{h \in \mathbf{E}_a : R(Q(l'),h) \in \mathbf{A}\}| > \delta |\mathbf{E}_a| + \varepsilon H$$

while from (3.3) we have

$$|\{h \in [H] : R(Q(l'),h) \in \mathbf{A}\}| \leqslant \delta H + \frac{1}{2}\varepsilon H$$

and hence

$$|\{h \in [H] \backslash \mathbf{E}_a : R(Q(l'),h) \in \mathbf{A}\}| < \delta |[H] \backslash \mathbf{E}_a| - \frac{1}{2}\varepsilon H.$$

This contradicts part (i) (with $\mathbf{E}$ now replaced instead by $[H]\backslash\mathbf{E}_a$ and ε replaced by $\varepsilon/2$). Thus we must have $c_a = 0$ for all $a \in [A]$. Setting l to be one of the elements of Q (e.g $l = Q(1)$), we obtain the claim.

Finally, we prove (iii). Using Corollary 1.4, we can find a partition $[H] = \mathbf{V}_1 \cup \cdots \cup \mathbf{V}_A$ with $A = O_\varepsilon(1)$, and real numbers $0 \leqslant c_{a,w} \leqslant 1$ for $a \in [A]$ and $w \in \mathbf{W}$, such that for any set $\mathbf{C} \subset [H]$, one has

$$\left| |\mathbf{C} \cap \mathbf{E}_w| - \sum_{a \in [A]} c_{a,w} |\mathbf{C} \cap \mathbf{V}_a| \right| \leqslant \frac{\varepsilon}{3} H \tag{4.1}$$

for all but $\varepsilon|\mathbf{W}|/2$ values of $w \in \mathbf{W}$.

By part (ii) (with the $\mathbf{E}_a$ replaced by $\mathbf{V}_a$), we can find $l \in [L]$ such that

$$||\{h \in \mathbf{V}_a : R(l,h) \in \mathbf{A}\}| - \delta|\mathbf{V}_a|| \leqslant \frac{\varepsilon}{3A} H$$

for all $a \in [A]$. In particular, from (4.1) (with $\mathbf{C} := \{h \in [H] : R(l,h) \in \mathbf{A}\}$) and the triangle inequality one has

$$\left| |\{h \in \mathbf{E}_w : R(l,h) \in \mathbf{A}\}| - \delta \sum_{a \in [A]} c_{a,w} |\mathbf{V}_a| \right| \leqslant \frac{2\varepsilon}{3} H$$

for all but $\varepsilon|\mathbf{W}|/2$ values of $w \in \mathbf{W}$; from a further application of (4.1) (with $\mathbf{C} := [H]$) we also have

$$\left| |\mathbf{E}_w| - \sum_{a \in [A]} c_{a,w} |\mathbf{V}_a| \right| \leqslant \frac{\varepsilon}{3} H$$

for all but $\varepsilon|\mathbf{W}|/2$ values of $w \in \mathbf{W}$. The claim (iii) now follows from the triangle inequality. $\square$

5. Roth's theorem

To demonstrate the above tools in action, we now prove Theorem 1.5, Theorem 1.6, and Roth's theorem (the $K = 3$ case of Theorem 1.2). Our proofs will be slightly *ad hoc* in nature in order to achieve certain shortcuts in the proof, and will thus differ slightly from the lengthier but more systematic arguments in the next section used to prove the general case of Theorem 1.2.

5.1. Proof of Theorem 1.5.

Let L be a natural number, let $\mathbf{S} \subset \mathbb{Z}$ have upper Banach density at least $1 - \frac{1}{10L}$, and suppose $\mathbf{S}$ is partitioned into finitely many colour classes. Then there exists an interval $h + [N]$ with $L \leqslant N$, such that

$$|\mathbf{S} \cap (h + [N])| \geqslant \left(1 - \frac{1}{9L}\right) N \tag{5.1}$$

(say). In particular, we can find $n_0 \in [\lfloor \frac{N}{8L} \rfloor]$ such that $h + n_0 \in \mathbf{S}$. We let $\mathbf{A}$ denote the colour class that $h + n_0$ belongs to.

Call a natural number $r \in [\lfloor \frac{N}{3} \rfloor]$ *good* if $h + n_0 + r, h + n_0 + 2r$ both lie in $\mathbf{S}$, and *bad* otherwise. Since $h + n_0 + r$ and $h + n_0 + 2r$ both lie in $h + [N]$, we see from (5.1) that there are at most $\frac{2}{9L}N$ bad elements of $[\lfloor \frac{N}{3} \rfloor]$. By the pigeonhole principle, we conclude that $[\lfloor \frac{N}{3} \rfloor]$ contains an interval $n_1 + [L]$ that consists entirely of good elements. If one now defines

$$\vec{P}_l := (h + n_0, h + n_0 + (n_1 + l), h + n_0 + 2(n_1 + l))$$

for all $l \in [L]$, we see that all the required conclusions of Theorem 1.5 are satisfied. (In fact we have constructed a configuration in which one has the collision $\vec{P}_1(1) = \cdots = \vec{P}_L(1)$, but this is not prohibited in the statement of Theorem 1.5.)

5.2. **Proof of Theorem 1.6.** Before we begin the proof of Theorem 1.6, we observe that Theorem 1.5 implies the following more "bounded" version.

Theorem 5.1 ($C'(3, \{2\})$). *Let*

$$1 \ll L \ll M \ll N,$$

and let $\mathbf{S}$ be a subset of $[N]$ of cardinality at least $(1 - \frac{1}{10L})N$. Suppose that $\mathbf{S}$ is partitioned into M colour classes. Then there exists a colour class $\mathbf{A} \subset \mathbf{S}$, together with a family $(\vec{P}_l)_{l \in [L]}$ of 3-APs $\vec{P}_l = (P_l(1), P_l(2), P_l(3))$ indexed by $l \in [L]$, obeying the properties (i), (ii), (iv) of Theorem 1.5.

Proof. Suppose for contradiction that Theorem 5.1 failed. Then there exist parameters

$$1 \ll L \ll M,$$

a sequence N_i, $i \in \mathbb{N}$ going to infinity, and sets $\mathbf{S}_i \subset [N_i]$ of cardinality $|\mathbf{S}_i| \geq (1 - \frac{1}{10L})N_i$ with M-colourings $c_i : \mathbf{S}_i \to [M]$, such that the conclusions of the theorem do not hold with $N, \mathbf{S}$ replaced by $N_i, \mathbf{S}_i$ for any i. We may recursively find translates $h_i + [N_i]$ of $[N_i]$ such that the intervals $h_i + [N_i]$ are disjoint, and such that for any $i \in \mathbb{N}$, any 3-AP or L-AP that contains at least two elements in $\bigcup_{i' < i} h_{i'} + [N_{i'}]$ will be disjoint from $h_i + [N_i]$; in particular, as $L \geq 3$, any 3-AP or L-AP that lies in $\bigcup_{i \in \mathbb{N}} h_i + [N_i]$ must in fact lie in a single one of the $h_i + N_i$.

Now set $\mathbf{S} := \bigcup_{i \in \mathbb{N}} h_i + \mathbf{S}_i$, then $\mathbf{S}$ has upper Banach density at least $1 - \frac{1}{10L}$, and any K-AP or L-AP that lies in $\mathbf{S}$ in fact must lie in a single one of the $h_i + \mathbf{S}_i$. We can M-colour $\mathbf{S}$ by assigning to each $h_i + n_i \in h_i + \mathbf{S}_i$ the colour $c_i(n_i) \in [M]$.

Applying Theorem 1.5, we can find a family $(\vec{P}_l)_{l \in [L]}$ of K-APs $\vec{P}_l$ that obey the conclusions (i), (ii), (iv) of that claim for $\mathbf{S}$. By construction of $\mathbf{S}$ and conclusion (i), each of the $\vec{P}_l$ must lie in exactly one of the $h_i + \mathbf{S}_i$; by conclusion (iii), this index i is independent of l. If one now translates all of the $\vec{P}_l$ by $-h_i$, one contradicts the claim that $\mathbf{S}_i$ does not obey the conclusions of the theorem, and the claim follows. $\square$

Actually, as the proof of Theorem 1.5 was so short, it is not difficult to modify it to provide a direct proof of Theorem 5.1; we leave the details to the interested reader. However, we give the argument above instead, as we will use it later in this paper.

Now we prove Theorem 1.6. Let L be a natural number, let $\mathbf{S} \subset \mathbb{Z}$ have upper Banach density at least $1 - \frac{1}{10L}$, and suppose $\mathbf{S}$ is partitioned into finitely many colour classes. Write σ for the upper density of $\mathbf{S}$ along $\mathcal{AP}$, thus[4] $\sigma \geqslant 1 - \frac{1}{10L}$. Applying Corollary 3.7, one can find a colour class $\mathbf{A}$ of $\mathbf{S}$ and an unbounded family $\mathcal{P} \subset \mathcal{AP}$ of arithmetic progressions with the double counting property, such that $\mathbf{A}$ has a positive density $\delta > 0$ along $\mathcal{P}$, and $\mathbf{S}$ has density σ along $\mathcal{P}$.

Now choose parameters

$$L, \frac{1}{\delta} \ll L' \ll H \ll N'.$$

As $\mathcal{P}$ is unbounded, it contains an N-AP U for some $N \geqslant N'$. Let $\mathbf{S}' \subset [N]$ denote the set of all $n \in [N]$ such that the H-AP $(U(n+h))_{h \in [H]}$ lies in $\mathcal{P}$; also for technical reasons we require also that

$$(U(n+h))_{0.9H \leqslant h \leqslant H : h = 0 \bmod 2}$$

and

$$(U(n+h))_{0.9H \leqslant h \leqslant H : h = 1 \bmod 2}$$

also lie in $\mathcal{P}$. From Proposition 3.5(i), we have

$$|\mathbf{S}'| \geqslant \left(1 - \frac{1}{10L'}\right) N.$$

We can also colour $\mathcal{S}'$ in 2^H colours, by assigning to each $n \in \mathcal{S}'$ the colour

$$\{h \in [H] : U(n+h) \in \mathbf{A}\} \subset [H].$$

Applying Theorem 5.1, one can find a colour class $\mathbf{A}'$ of $\mathbf{S}'$, together with a family $(\vec{P}_{l'})_{l' \in [L']}$ of 3-APs, obeying the following properties:

(i) For all $l' \in [L']$, $\vec{P}_{l'}$ is contained in $\mathbf{S}'$.
(ii) For all $l' \in [L']$, $P_{l'}(1)$ lies in $\mathbf{A}'$.
(iv) The tuple $(P_{l'}(2))_{l' \in [L']}$ is an L'-AP.

See Figure 8.

Let $\mathbf{P} \subset [H]$ denote the colour associated to $\mathbf{A}'$, thus by property (ii) we have

$$\{h \in [H] : U(P_{l'}(1) + h) \in \mathbf{A}\} = \mathbf{P}$$

for all $l' \in [L']$. By property (i), the H-AP $(U(P_{l'}(1) + h))_{h \in [H]}$ lies in $\mathcal{P}$; as $\mathbf{A}$ has density δ along $\mathcal{P}$, we conclude from (3.3) that

$$|\mathbf{P}| \geqslant \frac{\delta}{2} H$$

(say).

[4] In [9, Fact 6], a separate application of van der Waerden's theorem is used to exclude the case $\sigma = 1$, but in the arrangement of the argument given here, there is no need to do so, leading to a slight additional simplification.

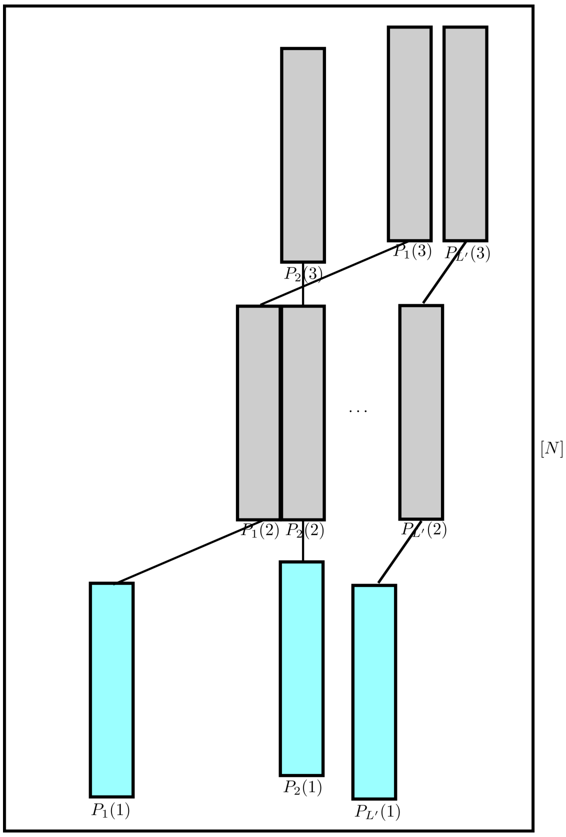

FIGURE 8. Each element $P_{l'}(k)$ for $l' \in [L']$ and $k \in [3]$ gives rise to an H-AP $(P_{l'}(k) + h)_{h \in [H]}$, which will almost certainly lie in $[N]$; these are depicted here as tall rectangles. When $k = 2, 3$, these rectangles are associated to $\mathbf{S}'$ (and are depicted here in grey), implying in particular that the H-AP $(U(P_{l'}(k) + h))_{h \in [H]}$ is "saturated" in the sense that it is almost entirely occupied by S and also occupied with positive density by A. When $k = 1$, these rectangles additionally lie in the colour class $\mathbf{A}'$ (and are depicted here in blue), implying in particular that the set $\mathbf{P} = \{h \in [H] : U(P_{l'}(1) + h) \in \mathbf{A}\}$ is "perfect" and does not vary with l'. The H-APs $(U(P_{l'}(2) + h))_{h \in [H]}$ for $l' \in [L']$ combine to form an $L' \times H$-rectangle, to which Theorem 4.1 may be profitably applied.

For technical reasons we need to remove the topmost portion of $\mathbf{P}$ from consideration. From Proposition 3.5(i), we see that for $l' \in [L']$, the L'-APs $(U(P_{l'}(1) + h + l))_{l \in [L']}$ lie in $\mathcal{P}$ for all but $\frac{\delta}{100}H$ of the $h \in [H]$, which from (3.3) implies that

$$|\mathbf{P} \cap (h + [L'])| \leqslant 2\delta L$$

for all but $\frac{\delta}{100}H$ of the $h \in [H]$. By a covering argument, this implies that

$$\left|\left\{h \in \mathbf{P} : h \geqslant \frac{9}{10}H\right\}\right| \leqslant \frac{\delta}{4}H,$$

and hence

$$\left|\left\{h \in \mathbf{P} : h < \frac{9}{10}H\right\}\right| \geqslant \frac{\delta}{4}H.$$

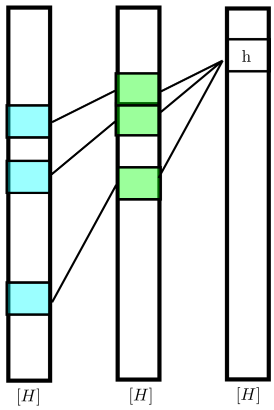

FIGURE 9. Three copies of the interval $[H]$, displayed separately for visual clarity. The left-most copy of $[H]$ is for parameterizing the "perfect" H-APs $(U(P_{l'}(1) + h_1))_{h_1 \in [H]}$ appearing in blue in Figure 8; the blue portion of this copy depicts the set appearing in (5.2); we have $U(P_{l'}(1) + h_1) \in \mathbf{A}$ for all h_1 in that set. The right-most copy of $[H]$ is for parameterizing the H-APs $(U(P_{l'}(3) + h))_{h \in [H]}$; we will only utilise the upper tenth of these H-APs, in which $h \geqslant \frac{9}{10}H$. The middle copy of $[H]$ parameterises the vertical components $(U(P_{l'}(2) + h_2))_{h_2 \in [H]}$ of the rectangle $\vec{R}$; The green portion of this copy depicts the set $\mathbf{E}_h$ associated to some $h \in \mathbf{W}$.

In particular, by the pigeonhole principle there exists a parity $i \in \mathbb{Z}/2\mathbb{Z}$ such that

$$\left|\left\{h \in \mathbf{P} : h < \frac{9}{10}H; h = i \bmod 2\right\}\right| \geqslant \frac{\delta}{8}H. \tag{5.2}$$

Let $\mathbf{W}$ consist of all natural numbers h with $\frac{9}{10}H \leqslant h \leqslant H$ and $h = i \bmod 2$, thus

$$|\mathbf{W}| \geqslant \frac{1}{30}H \tag{5.3}$$

(say). From (5.2) we see that for all $h \in \mathbf{W}$, there are at least $\frac{\delta}{8}H$ 3-APs $\vec{Q}$ in $[H]$ such that $Q(1) \in \mathbf{P}$ and $Q(3) = h$. Let $\mathbf{E}_h \subset [H]$ denote the set of all $Q(2)$, where $\vec{Q}$ ranges over the previously mentioned 3-APs, thus we have

$$|\mathbf{E}_h| \geqslant \frac{\delta}{8}H \tag{5.4}$$

for all $h \in \mathbf{W}$. See Figure 9.

By property (iv), the tuple

$$\vec{R} := (U(P_{l'}(2) + h))_{(l',h) \in [L'] \times [H]}$$

is an $L' \times H$-rectangle, and by property (i), all the columns $\vec{R}^{l'}$, $l' \in [L']$ of this rectangle lie in $\mathcal{P}$. Applying Theorem 4.1(iii) and (5.4), we may thus find $l' \in L'$ such that

$$|\{h' \in \mathbf{E}_h : U(P_{l'}(2) + h') \in \mathbf{A}\}| > 0 \tag{5.5}$$

for all but $\frac{1}{100L}H$ of the $h \in \mathbf{W}$. Also, from (i) we see that the H-AP $(U(P_{l'}(3)+h))_{h\in W}$ lies in $\mathcal{P}$, which implies in particular (since $\mathbf{S}$ has density at least $1-\frac{1}{10L}$ along $\mathcal{P}$) that

$$U(P_{l'}(3)+h) \in \mathbf{S} \tag{5.6}$$

for all but at most $\frac{1}{9L}H$ of the $h \in W$.

Comparing these facts with (5.3), we see that $\mathbf{W}$ must contain an L-AP $h_0 + 2 \cdot \overrightarrow{[L]}$ such that (5.5) and (5.6) hold for all $h \in h_0 + 2 \cdot \overrightarrow{[L]}$. Thus, for each $h \in h_0 + 2 \cdot \overrightarrow{[L]}$, one can find a 3-AP $\vec{Q}_h$ in $[H]$ such that $Q_h(1) \in \mathbf{P}$, $U(P_{l'}(2)+Q_h(2)) \in \mathbf{A}$, and $Q_h(3) = h$. One can then verify that the family

$$((U(P_{l'}(1)+Q_{h_0+2l}(1)), U(P_{l'}(2)+Q_{h_0+2l}(2)), U(P_{l'}(3)+Q_{h_0+2l}(3))))_{l\in[L]}$$

of 3-APs satisfy the required properties for Theorem 1.6.

5.3. **Proof of Roth's theorem.** To prove Roth's theorem, we first observe that Theorem 1.6 implies a bounded version:

Theorem 5.2 $(C'(3,\{3\}))$. *Let*

$$1 \ll L \ll M \ll N,$$

and let $\mathbf{S}$ *be a subset of* $[N]$ *of cardinality at least* $(1-\frac{1}{10L})N$. *Suppose that* $\mathbf{S}$ *is partitioned into* M *colour classes. Then there exists a colour class* $\mathbf{A} \subset \mathbf{S}$, *together with a family* $(\vec{P}_l)_{l\in[L]}$ *of 3-APs* $\vec{P}_l = (P_l(1), P_l(2), P_l(3))$ *indexed by* $l \in [L]$, *obeying the properties (i), (ii), (iv) of Theorem 1.6.*

The derivation of Theorem 5.2 from Theorem 1.6 is identical to the derivation of Theorem 5.1 from Theorem 1.5 and is omitted.

Now let A be a set of positive Banach density. Let δ be the upper density of A along $\mathcal{AP}$, thus $0 < \delta \leqslant 1$. By Theorem 3.6, we may find an unbounded family $\mathcal{P} \subset \mathcal{AP}$ of arithmetic progressions with the double counting property such that A has density δ along $\mathcal{P}$.

Now select parameters

$$\frac{1}{\delta} \ll L \ll H \ll N'.$$

As $\mathcal{P}$ is unbounded, we can find an N-AP $\vec{U} = (U(n))_{n\in[N]}$ in $\mathcal{P}$ for some $N \geqslant N'$. Let $\mathbf{S}' \subset [N]$ denote the set of all $n \in [N]$ such that the H-AP $(U(n+h))_{h\in[H]}$ lies in $\mathcal{P}$; also for technical reasons we require that $(U(n+h))_{h\in[H/2]}$ also lies in $\mathcal{P}$. From Proposition 3.5(i) as before, we have

$$|\mathbf{S}'| \geqslant \left(1 - \frac{1}{10L}\right) N.$$

Once again, we colour $\mathbf{S}'$ in 2^H colours, by assigning to each $n \in \mathbf{S}'$ the colour

$$\{h \in [H] : U(n+h) \in \mathbf{A}\} \subset [H].$$

Applying Theorem 5.2, one can find a colour class $\mathbf{A}'$ of $\mathbf{S}'$, together with a family $(\vec{P}_l)_{l\in[L]}$ of 3-APs, obeying the following properties:

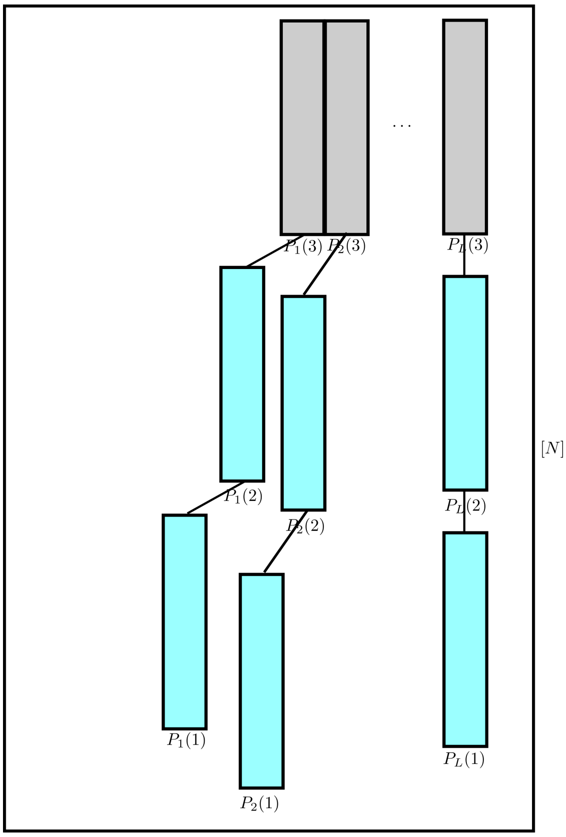

FIGURE 10. The analogue of Figure 10 after applying Theorem 5.2. Now it is the $L \times H$ rectangle $(U(P_l(3)+h))_{(h,l)\in[H]\times L}$, to which Theorem 4.1 may be profitably applied.

(i) For all $l \in [L]$, $\vec{P_l}$ is contained in $\mathbf{S}'$.
(ii) For all $l \in [L]$, $P_l(1)$ and $P_l(2)$ lie in $\mathbf{A}'$.
(iv) The tuple $(P_l(3))_{l\in[L]}$ is an L-AP.

See Figure 10.

Let $\mathbf{P} \subset [H]$ denote the colour associated to $\mathbf{A}'$, thus by property (ii) we have

$$\{h \in [H] : U(P_l(1)+h) \in \mathbf{A}\} = \{h \in [H] : U(P_l(2)+h) \in \mathbf{A}\} = \mathbf{P}$$

for all $l \in [L]$. By property (i) and (3.3) as before, we have

$$|\mathbf{P}| \geqslant \frac{\delta}{2} H \tag{5.7}$$

and also (since δ is the upper density of $\mathbf{A}$ along $\mathcal{AP}$)

$$|\mathbf{P} \cap [H/2]| \geqslant \frac{\delta}{4} H.$$

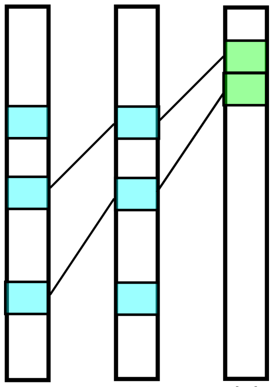

FIGURE 11. The analogue of Figure 9. The green portion of the third copy of $[H]$ depicts the set $\mathbf{E}$.

Let $\mathbf{E} \subset [H]$ denote the set of all numbers of the form $Q(3)$, where $\vec{Q} = (Q(1), Q(2), Q(3))$ is a 3-AP in $[H]$ with $Q(1), Q(2) \in \mathbf{P}$; see Figure 11. By choosing $Q(1), Q(2)$ to be elements of $\mathbf{P} \cap [H/2]$ in increasing order, we see that

$$|\mathbf{E}| \gg \delta^2 H. \tag{5.8}$$

By property (iv), the tuple

$$\vec{R} := (U(P_l(3) + h))_{(l,h)\in[L]\times[H]}$$

is an $L \times H$-rectangle, and by property (i), all the columns $\vec{R}^l$, $l \in [L]$ of this rectangle lie in $\mathcal{P}$. Applying Theorem 4.1(i), we can find $l \in L$ such that

$$|\{h \in \mathbf{E} : U(P_l(3) + h) \in \mathbf{A}\}| > 0,$$

thus we can find a 3-AP $\vec{Q}$ in $[H]$ with $Q(1), Q(2) \in \mathbf{P}$ and $U(P_l(3) + Q(3)) \in \mathbf{A}$. The 3-AP

$$(U(P_l(1) + Q(1)), U(P_l(2) + Q(2)), U(P_l(3) + Q(3)))$$

then lies in $\mathbf{A}$, proving Roth's theorem.

6. Szemerédi's theorem

We now present the proof of Theorem 1.2 in full generality, using a variant of the arguments of the preceding section. The following key claim $C(K, \Omega)$ (essentially "Fact 12" from [9]), defined for any $K \geqslant 3$ and $\Omega \subset [K]$, will play a crucial role in the argument.

Claim 6.1 ($C(K, \Omega)$). *Let*

$$1 \ll L \ll \frac{1}{\varepsilon}.$$

Let $\mathbf{S} \subset \mathbb{Z}$ have upper Banach density at least $1 - \varepsilon$, and let $c : \mathbf{S} \to \mathbf{C}$ be a colouring of $\mathbf{S}$ by some finite set $\mathbf{C}$ of colours. Then there exists a "perfect" colour $p \in \mathbf{C}$, and a family $(\vec{P}_{\vec{l}})_{\vec{l}\in[L]^\Omega}$ of K-APs $\vec{P}_{\vec{l}}$ parameterised by tuples $\vec{l} = (l_k)_{k\in\Omega}$ with $l_k \in [L]$ for all $k \in \Omega$, obeying the following axioms:

(i) *For all* $\vec{l} \in [L]^\Omega$*, the progression* $\vec{P}_{\vec{l}}$ *is contained in* S*.*
(ii) *If* Ω *is non-empty with minimal element* k_0*, then for all* $\vec{l} \in [L]^\Omega$ *and* $1 \leqslant k < k_0$*,* $P_{\vec{l}}(k)$ *has the perfect colour* p *(that is,* $c(P_{\vec{l}}(k)) = p$*). (If* Ω *is empty, we omit this conclusion.)*
(iii) *If* $k \in [K]$*, the colour of* $P_{\vec{l}}(k)$ *only depends on those components* $l_{k'}$ *of* $\vec{l}$ *with* $k' \leqslant k$*. That is, if* $\vec{l}, \vec{l'} \in [L]^\Omega$ *with* $\vec{l}_{k'} = \vec{l}'_{k'}$ *for all* $k' \in \Omega \cap [k]$*, then* $c(P_{\vec{l}}(k)) = c(P_{\vec{l'}}(k))$*.*
(iv) *If* $k \in \Omega$*, and one fixes all components of* $\vec{l} \in [L]^\Omega$ *except for* l_k*, then* $P_{\vec{l}}(k)$ *traces out an* L*-AP. That is to say, if* $\vec{l'} \in [L]^{\Omega\backslash\{k\}}$*, then there is an* L*-AP* $\vec{Q}_{k,\vec{l'}}$ *such that* $Q_{k,\vec{l'}}(l_k) = P_{\vec{l}}(k)$ *whenever* $\vec{l} = (l_{k'})_{k'\in\Omega} \in [L]^\Omega$ *agrees with* $\vec{l'}$ *on* $\Omega\backslash\{k\}$ *(thus* $l_{k'} = l'_{k'}$ *for all* $k' \in \Omega\backslash\{k\}$*.*

The claim $C(K, \varnothing)$ is trivially true: indeed, if S has upper density at least $1 - \varepsilon$, then from a simple counting argument it will contain an interval $P = a + \overrightarrow{[K]}$, which already gives (i), and (ii), (iii), (iv) are vacuously true in the $\Omega = \varnothing$ case. Note that the claims $C(3, \{2\})$ and $C(3, \{3\})$ are essentially Theorem 1.5 and Theorem 1.6 respectively (using the explicit choice $\varepsilon = \frac{1}{10L}$ of ε), while $C(4, \{2, 4\})$ is Theorem 1.7. As one may infer from the statements of Theorems 1.5, 1.6, one should be able to make the dependence of ε on L quite explicit (in particular, this dependence will not involve quantitative bounds for van der Waerden's theorem or the regularity lemma), but we will not do so here in general. The claims $C(K, \Omega)$ will be our substitute for the crucial "Fact 12" in [9]. Generally speaking, $C(K, \Omega)$ becomes harder to prove when the set Ω is larger and/or contains larger numbers; see Theorem 6.6 below (together with the iteration scheme immediately following that theorem) for a more precise statement.

It will be convenient to also use the following "bounded" variant of $C'(K, \Omega)$:

Claim 6.2 ($C'(K, \Omega)$)**.** *Let*

$$1 \ll L \ll \frac{1}{\varepsilon} \ll M \ll N.$$

Let $\mathbf{S} \subset [N]$ *have cardinality* $|\mathbf{S}| \geqslant (1 - \varepsilon)N$*, and let* $c : \mathbf{S} \to [M]$ *be a colouring of* $\mathbf{S}$ *by* M *colours. Then there exists a "perfect" colour* $p \in [M]$*, and a family* $(\vec{P}_{\vec{l}})_{\vec{l}\in[L]^\Omega}$ *of* K*-APs* $\vec{P}_{\vec{l}}$ *parameterised by tuples* $\vec{l} = (l_k)_{k\in\Omega}$ *with* $l_k \in [L]$ *for all* $k \in \Omega$*, obeying the conclusions (i)-(iv) of claim* $C(K, \Omega)$*.*

We now adapt the proof of Theorem 5.1 (or Theorem 5.2) to obtain

Lemma 6.3. *For any* $K \geqslant 3$ *and* $\Omega \subset [K]$*,* $C(K, \Omega)$ *implies* $C'(K, \Omega)$*.*

It is also easy to establish the converse implication of $C(K, \Omega)$ from $C'(K, \Omega)$, but we will not need this direction here.

Proof. Suppose for contradiction that $C(K, \Omega)$ was true but $C'(K, \Omega)$ failed. Then there exists

$$1 \ll L \ll \frac{1}{\varepsilon} \ll M,$$

a sequence N_i, $i \in \mathbb{N}$ going to infinity, and sets $\mathbf{S}_i \subset [N_i]$ of cardinality $|\mathbf{S}_i| \geqslant (1-\varepsilon)N_i$ with M-colourings $c_i : \mathbf{S}_i \to [M]$, such that the conclusions of $C'(K,\Omega)$ do not hold with $N, \mathbf{S}$ replaced by $N_i, \mathbf{S}_i$ for any i. We may recursively find translates $h_i + [N_i]$ of $[N_i]$ such that the intervals $h_i + [N_i]$ are disjoint, and such that for any $i \in \mathbb{N}$, any K-AP or L-AP that contains at least two elements in $\bigcup_{i'<i} h_{i'} + [N_{i'}]$ will be disjoint from $h_i + [N_i]$; in particular, as $K, L \geqslant 3$, any K-AP or L-AP that lies in $\bigcup_{i \in \mathbb{N}} h_i + [N_i]$ must in fact lie in a single one of the $h_i + N_i$.

Now set $\mathbf{S} := \bigcup_{i\in\mathbb{N}} h_i + \mathbf{S}_i$, then $\mathbf{S}$ has upper Banach density at least $1-\varepsilon$, and any K-AP or L-AP that lies in $\mathbf{S}$ in fact must lie in a single one of the $h_i + \mathbf{S}_i$. We can M-colour $\mathbf{S}$ by assigning to each $h_i + n_i \in h_i + \mathbf{S}_i$ the colour $c_i(n_i) \in [M]$.

Applying the claim $C(K,\Omega)$, we can find a family $(\vec{P}_{\vec{l}})_{\vec{l}\in[L]^\Omega}$ of K-APs $\vec{P}_{\vec{l}}$ that obey the conclusions (i)-(iv) of that claim for $\mathbf{S}$. By construction of $\mathbf{S}$ and conclusion (i), each of the $\vec{P}_{\vec{l}}$ must lie in exactly one of the $h_i + \mathbf{S}_i$. In principle, the index i could depend on $\vec{l}$; but by conclusion (iv), we see that i does not change if one varies just one of the components l_k of $\vec{l}$ holding all other components fixed, and so i is in fact independent of $\vec{l}$. If one now translates all of the $\vec{P}_{\vec{l}}$ by $-h_i$, one contradicts the claim that $\mathbf{S}_i$ does not obey the conclusions of $C'(K,\Omega)$, and the claim follows. □

Next, we observe that the claim $C'(K,\{K\})$ can be used to prove Szemerédi's theorem:

Proposition 6.4. *Suppose that $K \geqslant 3$ is such that $C'(K,\{K\})$ holds. Then any set $\mathbf{A}$ of integers of positive upper Banach density contains a $(K-1)$-AP.*

Proof. Write δ for the upper density of $\mathbf{A}$ along $\mathcal{AP}$, thus $0 < \delta \leqslant 1$. By Theorem 3.6, there exists an unbounded family $\mathcal{P}$ of arithmetic progressions with the double counting property such that $\mathbf{A}$ has density δ along $\mathcal{P}$.

Let
$$1 \ll L \ll \frac{1}{\varepsilon} \ll H \ll N'.$$
As $\mathcal{P}$ is unbounded, it contains an N-AP $\vec{P}$ for some $N \geqslant N'$. By Proposition 3.5(i), all but εN of the H-APs $(P(n+h))_{h\in[H]}$, $n \in [N]$ lie in $\mathcal{P}$. If one then defines the set
$$\mathbf{S} := \{n \in [N] : (P(n+h))_{h\in[H]} \in \mathcal{P}\},$$
then $|\mathbf{S}| \geqslant (1-\varepsilon)N$. On the other hand, we can colour $\mathbf{S}$ by at most 2^H colours by assigning to each element n of $\mathbf{S}$ the colour
$$\{h \in [H] : P(n+h) \in \mathbf{A}\} \subset [H].$$
Applying the claim $C'(K,\{K\})$ (and isolating just one $\vec{Q}$ of the K-APs produced by this claim), we conclude from conclusions (i) and (ii) that there exists a K-AP $\vec{Q}$ contained in $\mathbf{S}$ such that $Q(1), \dots, Q(K-1)$ all have the same colour p, thus
$$\{h \in [H] : P(Q(1)+h) \in \mathbf{A}\} = \cdots = \{h \in [H] : P(Q(K-1)+h) \in \mathbf{A}\} = p. \qquad (6.1)$$
Since all the H-APs $(P(Q(k)+h))_{h\in[H]}$, $k \in [K]$ are in $\mathcal{P}$, and $\mathbf{A}$ has density $\delta > 0$ along $\mathcal{P}$, we conclude from (3.3) that the sets in (6.1) are non-empty. Thus there exists

$h \in [H]$ such that $P(Q(k)+h) \in \mathbf{A}$ for all $k \in [K-1]$. Thus A contains the $(K-1)$-AP $(P(Q(k)+h))_{k\in[K-1]}$, and the claim follows. □

Remark 6.5. By a slightly longer argument using the techniques in Section 5.3, one can upgrade the conclusion of Proposition 6.4 by showing that $\mathbf{A}$ contains a K-AP rather than a $(K-1)$-AP. This leads to a slight reduction in the total length of the proof (one needs Theorem 6.6 below about 2^{K-1} times, rather than 2^K), and was utilised already in the proof of Roth's theorem. However, we do not use this shortcut here as it does not significantly simplify the proof if Szemerédi's theorem in full generality.

To finish the proof of Szemerédi's theorem, it thus suffices to prove the following inductive step.

Theorem 6.6 (Inductive step)**.** *Let $K \geqslant 3$, and let $\Omega \subset [K]$ contain $[k_0]$ but not k_0+1 for some $0 \leqslant k_0 < K$. Then $C'(K,\Omega)$ implies $C(K,\Omega\backslash[k_0] \cup \{k_0\})$.*

Indeed, suppose Theorem 6.6 was true. We assign to each $\Omega \subset [K]$ the "weight" $\sum_{k\in\Omega} 2^{k-1}$; this gives a one-to-one correspondence between the subsets Ω of $[K]$ and the integers between 0 and $2^K - 1$ inclusive, as can be seen by binary expansion. If Ω has weight less than $2^K - 1$, then it contains $[k_0]$ but not k_0+1 for some $0 \leqslant k_0 < K$. From Lemma 6.3 and Theorem 6.6, we conclude that $C(K,\Omega)$ implies $C(K,\Omega\backslash[k_0] \cup \{k_0\})$. Now note that the weight of $\Omega\backslash[k_0] \cup \{k_0\}$ is one more than the weight of Ω. Iterating this observation starting with $\Omega = \varnothing$ (which has weight 0), we conclude that $C(K,\Omega)$ is true for all $\Omega \subset [K]$. Applying Proposition 6.4, we obtain Szemerédi's theorem.

Example 6.7. If $K = 3$, then from Theorem 6.6 and Lemma 6.3 we obtain the chain of implications

$$C(3,\varnothing) \implies C(3,\{1\}) \implies C(3,\{2\}) \implies C(3,\{1,2\}) \implies C(3,\{3\}).$$

As it turns out, the proof of Theorem 6.6 can be modified (and made slightly more complicated) to obtain the variant

$$C'(K,\Omega\backslash\{1\}) \implies C(K,\Omega\backslash[k_0] \cup \{1,k_0\})$$

whenever $K \geqslant 3$ and $\Omega \subset [K]$ contains $[k_0]$ but not k_0+1 for some $1 \leqslant k_0 < K$, so one also has the shorter chain of implications

$$C(3,\varnothing) \implies C(3,\{2\}) \implies C(3,\{3\}).$$

Proposition 6.4 then shows that $C(3,\{3\})$ implies the $K = 2$ case of Szemerédi's theorem; by Remark 6.5, one can in fact obtain Roth's theorem. This essentially recovers the argument structure of Section 5.

It remains[5] to prove Theorem 6.6. A key step in the proof will be to establish the following "counting lemma", which will be a consequence of multiple applications of Theorem 4.1, and which roughly corresponds to [9, Lemma 5]:

[5]The reader may wish, as a warmup, to try to adapt the arguments in Section 5.2 to prove the $(K,\Omega) = (4,\{1,2,4\})$ case of this theorem, that is to say to use $C'(4,\{1,2,4\})$ to prove $C(4,\{3,4\})$, as this case is already quite typical.

Theorem 6.8 (Counting lemma)**.** *Let $K \geqslant 3$, and let $\Omega \subset [K]$ contain $[k_0]$ but not k_0+1 for some $0 \leqslant k_0 < K$. Let $0 < \varepsilon \leqslant 1$, let $\mathcal{P}$ be an unbounded family of arithmetic progressions with the double counting property, let $\mathbf{S} \subset \mathbb{Z}$ be a set with a density σ along $\mathcal{P}$ for some $\sigma \geqslant 1-\varepsilon$, and let $\mathbf{A} \subset \mathbf{S}$ be a set with a density δ along $\mathcal{P}$ for some $0 < \delta \leqslant 1$. Let*

$$\frac{1}{\varepsilon}, \frac{1}{\delta} \ll \frac{1}{\kappa} \ll L' \ll H,$$

let r be a natural number, and suppose that there is a family $(\vec{P}_{\vec{l}'})_{\vec{l}' \in [L']^\Omega}$ of K-APs $\vec{P}_{\vec{l}'}$ indexed by tuples $\vec{l}' = (l'_k)_{k \in \Omega}$ with $l'_k \in [L']$ for all $k \in \Omega$, obeying the following axioms:

(i) *For all $\vec{l}' \in [L']^\Omega$ and $k \in [K]$, the H-AP $\vec{P}_{\vec{l}'}(k) + r \cdot [H]$ lies in $\mathcal{P}$.*
(iii) *If $k \in [K]$, the set $\{h \in [H] : \vec{P}_{\vec{l}'}(k) + rh \in \mathbf{A}\} \subset [H]$ only depends on those components $l'_{k'}$ of $\vec{l}'$ with $k' \leqslant k$.*
(iv) *If $k \in \Omega$, and one fixes all components of $\vec{l}'$ except for l'_k, then $\vec{P}_{\vec{l}'}(k)$ traces out an L'-AP.*

Let $\mathcal{Q}$ denote the collection of K-APs $\vec{Q}$ contained in $[H]$, and for every $k \in [K]$ and $h \in [H]$, let $\mathcal{Q}_k(h) \subset \mathcal{Q}$ denote the subcollection

$$\mathcal{Q}_k(h) := \{\vec{Q} \in \mathcal{Q} : Q(k) = h\}.$$

Let $0 \leqslant k_1 \leqslant k_0$ be an integer. Then there exist $\vec{l}^* \in [L']^\Omega$, such that for every $k' \in [K] \backslash [k_1]$, one has the homogeneity property

$$|\{Q \in \mathcal{Q}_{k'}(h) : \vec{P}_{\vec{l}^*}(k) + rQ(k) \in \mathbf{A} \text{ for all } k \in [k_1]\}| = \delta^{k_1} |\mathcal{Q}_{k'}(h)| + O_K(\kappa H) \quad (6.2)$$

for all but $O_K(\kappa H)$ elements h of $[H]$.

The intuition here is that as $\mathbf{A}$ has density δ along $\mathcal{P}$, the "probability" that the event $\vec{P}_{\vec{l}^*}(k) + rQ(k) \in \mathbf{A}$ occurs for a given $k \in [k_1]$ should be approximately equal to δ for "typical" choices of $\vec{l}^*$ and Q. The idea is to use Theorem 4.1 repeatedly to then locate a good choice of $\vec{l}^*$ in which this intuition is correct for $k = 1, \ldots, k_1$ in turn.

Proof. We induct on k_1; we will allow the implied constants in the $O_K()$ notation to vary with this induction, but this is harmless since k_1 will only increase at most K times. The case $k_1 = 0$ is trivial (just choose $\vec{l}^* \in [L']^\Omega$ arbitrarily), so suppose that $1 \leqslant k_1 \leqslant k_0$ and that the claim has already been proven for $k_1 - 1$. Thus, there exists $\vec{l}^{**} \in [L']^\Omega$ such that for every $k' \in [K] \backslash [k_1 - 1]$, one has

$$|\{Q \in \mathcal{Q}_{k'}(h) : P_{\vec{l}^{**}}(k) + rQ(k) \in \mathbf{A} \text{ for all } k \in [k_1 - 1]\}| = \delta^{k_1 - 1} |\mathcal{Q}_{k'}(h)| + O_K(\kappa H)$$

for all but $O_K(\kappa H)$ elements h of $[H]$.

For any $h \in [H]$ and $k' \in [K] \backslash [k_1]$, let $\mathbf{E}_{k',h} \subset [H]$ denote the set

$$\mathbf{E}_{k',h} := \{Q(k_1) : Q \in \mathcal{Q}_{k'}(h); P_{\vec{l}^{**}}(k) + rQ(k) \in \mathbf{A} \text{ for all } k \in [k_1 - 1]\}, \quad (6.3)$$

thus we have

$$|\mathbf{E}_{k',h}| = \delta^{k_1 - 1} |\mathcal{Q}_{k'}(h)| + O_K(\kappa H) \quad (6.4)$$

for all but $O_K(\kappa H)$ elements h of $[H]$.

For $l' \in [L']$, let $\vec{l}^{*,l'}$ be the element of $[L']^\Omega$ formed from $\vec{l}^{**}$ by replacing the k_1 coordinate with l'. By the hypothesis (iv) of Theorem 6.8, the tuple

$$(P_{\vec{l}^{*,l'}}(k_1))_{l' \in [L']}$$

is an L'-AP, and hence the tuple

$$(P_{\vec{l}^{*,l'}}(k_1) + rh)_{(l',h) \in [L'] \times [H]}$$

is an $L' \times H$-rectangle. By the hypothesis (i) of Theorem 6.8, all the L' columns of this rectangle lie in $\mathcal{P}$. By Theorem 4.1(iii), we may thus find $l' \in [L']$ such that

$$|\mathbf{E}_{k',h} \cap \{h : P_{\vec{l}^{*,l'}}(k_1) + rh \in \mathbf{A}\}| = \delta |\mathbf{E}_{k',h}| + O(\kappa H)$$

for all but at most κH of the tuples $(k', h) \in ([K] \backslash [k_1]) \times [H]$. By (6.4), we thus have that for any $k' \in [K] \backslash [k_1]$, one has

$$|\mathbf{E}_{k',h} \cap \{h : P_{\vec{l}^{*,l'}}(k_1) + rh \in \mathbf{A}\}| = \delta^{i_1} |\mathcal{Q}_{k'}(h)| + O_K(\kappa H)$$

for all but $O_K(\kappa H)$ elements h of $[H]$. By (6.3), the left-hand side may be written as

$$|\{Q \in \mathcal{Q}_{k'}(h) : P_{\vec{l}^{*,l'}}(k) + rQ(k) \in \mathbf{A} \text{ for all } k \in [k_0]\}|.$$

Setting $\vec{l}^* := \vec{l}^{*,l'}$, we obtain the claim. □

To conclude the proof of Theorem 6.6 (and hence Szemerédi's theorem), we now show the following implication, which is a variant of [9, Lemma 6].

Proposition 6.9. *Theorem 6.8 implies Theorem 6.6.*

Proof. This will be a double counting argument (in the spirit of the proof of Theorem 1.5), relying primarily on the fact that the density of $\mathbf{S}$ is at least $1-\varepsilon$ for a fairly small value of ε to eliminate the contribution of those k-APs which are not fully contained in $\mathbf{S}$. A key technical difficulty will be the appearance (via (6.2)) of the quantity δ, which will probably be much smaller than ε; however, all the factors of δ will safely cancel each other out in the final analysis.

We turn to the details. Let $K \geqslant 3$, and let $\Omega \subset [K]$ contain $[k_0]$ but not $k_0 + 1$ for some $0 \leqslant k_0 < K$, and suppose that $C'(K, \Omega)$ holds. Let

$$1 \ll L \ll \frac{1}{\varepsilon},$$

let $\mathbf{S} \subset \mathbb{Z}$ be a set of upper Banach density at least $1 - \varepsilon$, and suppose that one has a finite colouring $c : \mathbf{S} \to \mathbf{C}$ of S. Let σ denote the upper density of $\mathbf{S}$ along $\mathcal{AP}$, then $\sigma \geqslant 1 - \varepsilon$. By Corollary 3.7, there exists an unbounded family $\mathcal{P} \subset \mathcal{AP}$ of arithmetic progressions with the double counting property and a "perfect" colour $p \in \mathbf{C}$, such that $\mathbf{S}$ has density σ along $\mathcal{P}$, and the colour class $\mathbf{A} := \{s \in \mathbf{S} : c(s) = p\}$ has a positive density $\delta > 0$ along $\mathcal{P}$.

Now select parameters

$$\frac{1}{\varepsilon}, \frac{1}{\delta}, |\mathbf{C}| \ll \frac{1}{\kappa} \ll L' \ll \frac{1}{\varepsilon'} \ll H \ll N'.$$

As $\mathcal{P}$ is unbounded, it contains an N-AP $U = a + r \cdot \overrightarrow{[N]}$ for some $N \geqslant N'$, integer a, and natural number r. By Proposition 3.5(i), all but $\varepsilon' N$ of the H-APs

$$(U(n+h))_{h \in [H]} = U(n) + r \cdot \overrightarrow{[H]}$$

for $n \in [N]$ lie in $\mathcal{P}$. If one then defines the set

$$\mathbf{S}' := \{n \in [N] : U(n) + r \cdot \overrightarrow{[H]} \in \mathcal{P}\},$$

then $\mathbf{S}'$ is a subset of $[N]$ with $|\mathbf{S}'| \geqslant (1-\varepsilon')N$. On the other hand, we can colour $\mathbf{S}'$ by at most $(|\mathbf{C}|+1)^H$ colours by assigning to each $n \in \mathbf{S}'$ the colour $c'(n) \in (\mathbf{C} \cup \{\perp\})^H$, defined as the tuple

$$c'(n) := (c(U(n) + rh))_{h \in [H]},$$

where we extend c outside of S by setting $c(n) = \perp$ for all $n \notin S$ and some symbol $\perp$ outside of $\mathbf{C}$. Applying the hypothesis $C'(K,\Omega)$ (with $L, \varepsilon, \mathbf{S}, c$ replaced by $L', \varepsilon', \mathbf{S}', c'$ respectively), and discarding the conclusion (ii) of that claim (which is the only conclusion involving the perfect colour), we obtain a family $(\vec{P}'_{\vec{l}'})_{\vec{l}' \in [L']^\Omega}$ of K-APs $\vec{P}'_{\vec{l}'}$ obeying the following axioms:

(i) For all $\vec{l}' \in [L']^\Omega$, the progression $\vec{P}'_{\vec{l}'}$ is contained in $\mathbf{S}'$.

(iii) If $k \in [K]$, the tuple $\left(c(U(\vec{P}'_{\vec{l}'}(k)) + rh)\right)_{h \in [H]}$ only depends on those components $l'_{k'}$ of $\vec{l}' = (l'_{k'})_{k' \in \Omega}$ with $k' \leqslant k$.

(iv) If $k \in \Omega$, and one fixes all components of $\vec{l}' = (l'_{k'})_{k' \in \Omega}$ except for $l'_{k'}$, then $\vec{P}'_{\vec{l}'}(k)$ traces out an L'-AP.

We then define the K-AP $\vec{P}_{\vec{l}'}$ for each $\vec{l}' \in [L']^\Omega$ by composing the two increasing affine-linear maps $P'_{\vec{l}'} : \mathbb{Z} \to \mathbb{Z}$ and $U : \mathbb{Z} \to \mathbb{Z}$:

$$\vec{P}_{\vec{l}'} := \left(U(P'_{\vec{l}'}(k))\right)_{k \in [K]}.$$

We then see that the $(\vec{P}_{\vec{l}'})_{\vec{l}' \in [L']^\Omega}$ obey the axioms (i), (iii), (iv) of Theorem 6.8. In fact we have a stronger claim than (iii):

(iii') If $k \in [K]$, the tuple $(c(P_{\vec{l}'}(k) + rh))_{h \in [H]}$ only depends on those components $l'_{k'}$ of $\vec{l}' = (l'_{k'})_{k' \in \Omega}$ with $k' \leqslant k$.

Note that (iii') implies (iii), since $P_{\vec{l}'}(k) + rh \in \mathbf{A}$ if and only if $c(P_{\vec{l}'}(k) + rh) = p$.

Applying Theorem 6.8 with $k_1 = k_0$, there exists $\vec{l}^* = (l^*_k)_{k \in \Omega} \in [L']^\Omega$, such that for every $k' \in [K] \backslash [k_0]$, one has the homogeneity property

$$\left|\left\{Q \in \mathcal{Q}_{k'}(h) : P_{\vec{l}^*}(k) + rQ(k) \in \mathbf{A} \text{ for all } k \in [k_0]\right\}\right| = \delta^{k_0} |\mathcal{Q}_{k'}(h)| + O_K(\kappa H) \quad (6.5)$$

for all but $O_K(\kappa H)$ elements h of $[H]$.

Let $\mathbf{L}$ denote the set of all $\vec{l'} = (l'_k)_{k\in\Omega} \in [L']^\Omega$ such that $l'_k = l^*_k$ for all $k \in [k_0]$, and $l'_k \in [L]$ for $k \in \Omega\backslash[k_0]$ (note carefully here that we are restricting l'_k here to the short interval $[L]$ rather than the long interval $[L']$). Clearly

$$|\mathbf{L}| \leqslant L^K. \tag{6.6}$$

Let $\mathcal{Q}'$ denote the collection of k-APs $\vec{Q} \in \mathcal{Q}$ such that $P_{\vec{l}^*}(k) + rQ(k) \in \mathbf{A}$ for all $k \in [k_0]$. Call a K-AP $\vec{Q} \in \mathcal{Q}'$ *good* if one has

$$P_{\vec{l'}}(k') + rQ(k') \in \mathbf{S} \text{ for all } k' \in [K]\backslash[k_0] \text{ and } \vec{l'} \in \mathbf{L},$$

and *bad* otherwise. Clearly, the number of bad $\vec{Q}$ is at most

$$\sum_{\vec{l'}\in\mathbf{L}} \sum_{k'\in[K]\backslash[k_0]} \sum_{h\in[H]:P_{\vec{l'}}(k')+rh\notin\mathbf{S}} |\mathcal{Q}' \cap \mathcal{Q}_{k'}(h)|.$$

From (3.3) applied to the H-AP $P_{\vec{l'}}(k') + r \cdot [H]$, which is in $\mathcal{P}$, we have

$$|\{h \in [H] : P_{\vec{l'}}(k') + rh \in \mathbf{S}\}| \geqslant (1 - 2\varepsilon)H,$$

thus the inner sum in the above expression is over $O(\varepsilon H)$ elements h. From (6.5) and the trivial bound $|\mathcal{Q}_{i'}(h)| \leqslant H$, we have $|\mathcal{Q}' \cap \mathcal{Q}_{k'}(h)| = O(\delta^{k_0} H)$ for all but $O_K(\kappa H)$ elements h of $[H]$. This implies that

$$\sum_{h\in[H]:P_{\vec{l'}}(k')+rh\notin\mathbf{S}} |\mathcal{Q}' \cap \mathcal{Q}_{k'}(h)| = O(\varepsilon\delta^{k_0} H^2)$$

and hence the number of bad $\vec{Q}$ is at most $O_{K,L}(\varepsilon\delta^{k_0} H^2)$.

Let $[H]^\circ := \{h \in [H] : H/3 \leqslant h \leqslant 2H/3\}$ denote the middle third of $[H]$. For $h \in [H]^\circ$, the cardinality $|\mathcal{Q}_{k_0+1}(h)|$ is at least $\frac{H}{3K}$. From (6.5), we conclude that

$$|\mathcal{Q}' \cap \mathcal{Q}_{k_0+1}(h)| \geqslant \delta^{k_0} \frac{H}{4K}$$

(say) for all but $O_K(\kappa H)$ elements h of $[H]^\circ$. Comparing this against the number of bad $\vec{Q}$, we conclude that the number of $h \in [H]^\circ$ such that all elements of $\mathcal{Q}' \cap \mathcal{Q}_{k_0+1}(h)$ are bad cannot exceed $O_{k,L}(\varepsilon H)$; crucially, all the factors of δ have been cancelled out. This is significantly less than $\frac{|[H]^\circ|}{L}$, so we conclude that $[H]^\circ$ contains an interval $h_0 + [L]$ with the property that for each $h_0 + l \in h_0 + [L]$, there is a good K-AP $\vec{Q}_l$ in $\mathcal{Q}' \cap \mathcal{Q}_{k_0+1}(h_0 + l)$.

Now we can prove $C(K, \Omega\backslash[k_0] \cup \{k_0 + 1\})$. For $\vec{l} = (l_k)_{k\in\Omega\backslash[k_0]\cup\{k_0+1\}} \in [L]^{\Omega\backslash[k_0]\cup\{k_0+1\}}$, we define $\underline{\vec{P}}_{\vec{l}}$ to be the K-AP

$$\underline{\vec{P}}_{\vec{l}} := (P_{\underline{\vec{l}}}(k) + rQ_{h_0+l_{k_0+1}}(k))_{k\in[K]}, \tag{6.7}$$

where $\underline{\vec{l}} \in [L']^\Omega$ has the same coordinates as $\vec{l}$ on $\Omega\backslash[k_0]$, and the same coordinates as $\vec{l}^*$ on $[k_0]$.

It is clear that the $\underline{\vec{P}}_{\vec{l}}$ are K-APs. To finish the proof of $C(K, \Omega\backslash[k_0] \cup \{k_0 + 1\})$, we need to verify the axioms (i)-(iv) (with $\vec{P}_{\vec{l}}$ replaced by $\underline{\vec{P}}_{\vec{l}}$, and Ω replaced by $\Omega\backslash[k_0] \cup \{k_0 + 1\}$).

We first verify (i). Since $Q_{h_0+l_{k_0+1}}$ lies in $\mathcal{Q}'$, we have

$$P_{\vec{l}^*}(k) + rQ_{h_0+l_{k_0+1}}(k) \in \mathbf{A} \subset \mathbf{S}$$

for all $k \in [k_0]$; as $\underline{\vec{l}}$ and $\vec{l}^*$ agree on the first $[k_0]$ coordinates, we conclude from the hypothesis (iii) of Theorem 6.8 that

$$\underline{\vec{P}}_{\vec{l}}(k) = P_{\underline{\vec{l}}}(k) + rQ_{h_0+l_{k_0+1}}(k) \in \mathbf{A} \subset \mathbf{S} \tag{6.8}$$

for all $k \in [k_0]$. For $k' \in [K]\backslash[k_0]$, we use the fact that $Q_{h_0+l_{k_0+1}}$ is good and that $\underline{\vec{l}} \in \mathbf{L}$ to conclude that

$$\underline{\vec{P}}_{\vec{l}}(k') = P_{\underline{\vec{l}}}(k') + rQ_{h_0+l_{k_0+1}}(k') \in \mathbf{S}.$$

This proves (i).

The claim (ii) follows from (6.8), so we turn to (iii). Let $k \in [K]$, and suppose that $\vec{l} = (l_{k'})_{k' \in \Omega\backslash[k_0]\cup\{k_0+1\}}$ and $\vec{l}' = (l'_{k'})_{k' \in \Omega\backslash[k_0]\cup\{k_0+1\}}$ are elements of $[L]^{\Omega\backslash[k_0]\cup\{k_0+1\}}$ such that $l_{k'} = l'_{k'}$ for all $k' \in \Omega \cap [k]$. We wish to show that

$$c(\underline{\vec{P}}_{\vec{l}}(k)) = c(\underline{\vec{P}}_{\vec{l}'}(k)). \tag{6.9}$$

If $k \in [k_0]$ then this follows from claim (ii) (or (6.8)), so we may assume that $k \in [k]\backslash[k_0]$. In particular we have $l_{k_0+1} = l'_{k_0+1}$, and hence

$$Q_{h_0+l_{k_0+1}}(k) = Q_{h_0+l'_{k_0+1}}(k). \tag{6.10}$$

On the other hand, if we define $\underline{\vec{l}'} \in \mathbf{L}$ similarly to $\underline{\vec{l}}$ (but with $\vec{l}$ replaced by $\vec{l}'$), we see from construction that $\underline{\vec{l}'}$ and $\underline{\vec{l}}$ agree in all coordinates less than or equal to k, and hence by the property (iii') we have

$$(c(P_{\underline{\vec{l}}}(k) + rh))_{h \in [H]} = (c(P_{\underline{\vec{l}'}}(k) + rh))_{h \in [H]}.$$

From this, (6.10), and (6.7) we obtain (6.9) as required.

Finally, we verify (iv). We need to show that for all $k \in \Omega\backslash[k_0] \cup \{k_0+1\}$, the quantity $\underline{\vec{P}}_{\vec{l}}(k)$ depends in an increasing affine-linear fashion on the coordinate l_k of $\vec{l}$, if all other components of $\vec{l}$ are held fixed. For $k = k_0 + 1$ this is immediate from the definition (6.7), and for $k > k_0 + 1$, the claim follows from the property (iv) of Theorem 6.8. This concludes the proof. □

Appendix A. Proof of weak regularity lemma

In this appendix we prove Lemma 1.3. Our main tool will be the singular value decomposition. We follow the arguments laid out in `terrytao.wordpress.com/2012/12/03`.

Let $\mathbf{V}, \mathbf{W}, \varepsilon, \mathbf{E}$ be as in that lemma. Without loss of generality we may assume that $\mathbf{V} = [V]$ and $\mathbf{W} = [W]$ for some natural numbers V, W (the case when $\mathbf{V}$ or $\mathbf{W}$ empty being trivial). Applying the singular value decomposition to the adjacency matrix $(1_{\mathbf{E}}(v,w))_{v\in[V],w\in[W]}$, one can obtain a decomposition

$$1_{\mathbf{E}}(v,w) = \sum_{i=1}^{n} \lambda_i f_i(v) g_i(w) \tag{A.1}$$

for all $v \in [V], w \in [W]$ and some finite sequence $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ of singular values, some singular functions $f_1, \ldots, f_n : [V] \to \mathbb{R}$ that are orthonormal in the sense that $\frac{1}{|V|}\sum_{v\in[V]} f_i(v) f_j(v) = 1_{i=j}$ for all $i, j \in [n]$, and some singular functions $g_1, \ldots, g_n : [W] \to \mathbb{R}$ that are orthonormal in the sense that $\frac{1}{|W|}\sum_{w\in[W]} g_i(w) g_j(w) = 1_{i=j}$ for all $i, j \in [n]$.

Squaring both sides of (A.1) and averaging in v, w, we obtain the Frobenius norm identity

$$\frac{1}{VW} \sum_{v\in[V]} \sum_{w\in[W]} 1_{\mathbf{E}}(v,w)^2 = \sum_{i=1}^{n} \lambda_i^2.$$

The left-hand side is clearly bounded by 1, hence

$$\sum_{i=1}^{n} \lambda_i^2 \leqslant 1. \tag{A.2}$$

In particular, since the λ_i are decreasing, we have

$$\lambda_i \leqslant \frac{\varepsilon}{4}$$

whenever $i \geqslant 16/\varepsilon^2$.

Now suppose that i is such that $\lambda_i > \varepsilon/4$. Multiplying (A.1) by $g_i(w)$ and averaging in w, we conclude that

$$\frac{1}{W} \sum_{w\in[W]} 1_{\mathbf{E}}(v,w) g_i(w) = \lambda_i f_i(v)$$

for all $v \in V$. Since $\frac{1}{W}\sum_{w\in[W]} g_i(w)^2 = 1$, we see from the Cauchy-Schwarz inequality that the left-hand side is at most 1 in magnitude. As we are assuming $\lambda_i > \varepsilon/4$, we conclude the pointwise bound

$$|f_i(v)| < \frac{4}{\varepsilon}$$

for all $v \in [V]$. Similarly we have

$$|g_i(w)| < \frac{4}{\varepsilon}$$

for all $w \in [W]$.

By subdividing the interval $(-\frac{4}{\varepsilon}, \frac{4}{\varepsilon})$ into at most $\frac{128}{\varepsilon^3}$ intervals of length at most $\frac{\varepsilon^2}{16}$, we see that for each i with $\lambda_i > \varepsilon/4$, we can partition $[V]$ into at most $\frac{128}{\varepsilon^3}$ sets, such that the function $f_i(v)$ only fluctuates by at most $\varepsilon^2/16$ on each set. As there are at most $16/\varepsilon^2$ such i, by combining all these partitions together we see that we can find a partition $[V] = \mathbf{V}_1 \cup \cdots \cup \mathbf{V}_A$ into non-empty sets for some $A \leqslant (128/\varepsilon^3)^{16/\varepsilon^2}$ such that for each $a \in [A]$ and each i with $\lambda_i > \varepsilon/4$, f_i fluctuates by at most $\varepsilon^2/16$ on $\mathbf{V}_a$. Similarly, we can partition $[W] = \mathbf{W}_1 \cup \cdots \cup \mathbf{W}_B$ into non-empty sets for some $B \leqslant (128/\varepsilon^3)^{16/\varepsilon^2}$

such that for each $b \in [B]$ and each i with $\lambda_i > \varepsilon/4$, g_i fluctuates by at most $\varepsilon^2/16$ on $\mathbf{W}_b$.

For each $a \in [A]$ and $b \in [B]$, define the density

$$d_{a,b} := \frac{1}{|\mathbf{V}_a|}\frac{1}{|\mathbf{W}_b|}\sum_{v\in\mathbf{V}_a}\sum_{w\in\mathbf{W}_b} 1_{\mathbf{E}}(a,b),$$

thus the function $(v,w) \mapsto 1_{\mathbf{E}}(v,w) - d_{a,b}$ has mean zero on $\mathbf{V}_a \times \mathbf{W}_b$. Clearly we have $0 \leqslant d_{a,b} \leqslant 1$ for all $a \in [A]$ and $b \in [B]$.

Now let $\mathbf{F} \subset [V]$ and $\mathbf{W} \subset [W]$. The expression

$$|(\mathbf{F}\times\mathbf{G}) \cap \mathbf{E}| - \sum_{a\in[A]}\sum_{b\in[B]} d_{a,b}|\mathbf{F}\cap\mathbf{V}_a||\mathbf{G}\cap\mathbf{W}_b|$$

appearing in (1.1) may be written as

$$\sum_{a\in[A]}\sum_{b\in[B]}\sum_{v\in\mathbf{V}_a}\sum_{w\in\mathbf{W}_b} 1_{\mathbf{F}}(v)1_{\mathbf{G}}(w)(1_{\mathbf{E}}(v,w) - d_{a,b}). \tag{A.3}$$

Writing

$$\alpha_a := \frac{1}{|\mathbf{V}_a|}\sum_{v\in\mathbf{V}_a} 1_{\mathbf{F}}(v)$$

and

$$\beta_b := \frac{1}{|\mathbf{W}_b|}\sum_{w\in\mathbf{W}_b} 1_{\mathbf{G}}(v)$$

for the means of $1_{\mathbf{F}}$ and $1_{\mathbf{G}}$ on $\mathbf{V}_a$ and $\mathbf{W}_b$ respectively, we may decompose

$$1_{\mathbf{F}}(v)1_{\mathbf{G}}(w) = (1_{\mathbf{F}}(v) - \alpha_a)1_{\mathbf{G}}(w) + \alpha_a(1_{\mathbf{G}}(w) - \beta_b) + \alpha_a\beta_b.$$

The third term makes no contribution to (A.3) since $(v,w) \mapsto 1_{\mathbf{E}}(v,w) - d_{a,b}$ has mean zero. Thus, by the triangle inequality, the left hand side of (1.1) is bounded by the sum of

$$\left|\sum_{a\in[A]}\sum_{b\in[B]}\sum_{v\in\mathbf{V}_a}\sum_{w\in\mathbf{W}_b} (1_{\mathbf{F}}(v) - \alpha_a)1_{\mathbf{G}}(w)(1_{\mathbf{E}}(v,w) - d_{a,b})\right| \tag{A.4}$$

and

$$\left|\sum_{a\in[A]}\sum_{b\in[B]}\sum_{v\in\mathbf{V}_a}\sum_{w\in\mathbf{W}_b} \alpha_a(1_{\mathbf{G}}(w) - \beta_b)(1_{\mathbf{E}}(v,w) - d_{a,b})\right|. \tag{A.5}$$

We first estimate (A.4). Since $1_{\mathbf{F}}(v) - \alpha_a$ has mean zero, we may remove the $d_{a,b}$ factor and write this as

$$\left|\sum_{a\in[A]}\sum_{b\in[B]}\sum_{v\in\mathbf{V}_a}\sum_{w\in\mathbf{W}_b} (1_{\mathbf{F}}(v) - \alpha_a)1_{\mathbf{G}}(w)1_{\mathbf{E}}(v,w)\right|.$$

This can in turn be rewritten as

$$|\sum_{v\in[V]}\sum_{w\in[W]} F(v)G(w)1_{\mathbf{E}}(v,w)|$$

where $F(v) := \sum_{a\in[A]}(1_{\mathbf{F}}(v) - \alpha_a)1_{\mathbf{V}_a}(v)$ and $G(w) := 1_{\mathbf{G}}(w)$. Using the singular value decomposition (A.1), this can be written in turn as

$$|\sum_{i=1}^{n}\lambda_i(\sum_{v\in[V]}F(v)f_i(v))(\sum_{w\in[W]}G(w)g_i(w))|.$$

From Bessel's inequality (and the observation that F is bounded in magnitude by 1) one has

$$\sum_{i=1}^{n}\left|\frac{1}{V}\sum_{v\in[V]}F(v)f_i(v)\right|^2 \leqslant 1$$

and similarly

$$\sum_{i=1}^{n}\left|\frac{1}{W}\sum_{g\in[W]}G(w)g_i(w)\right|^2 \leqslant 1.$$

From the Cauchy-Schwarz inequality, we conclude that

$$\left|\sum_{i:\lambda_i\leqslant\varepsilon/4}\lambda_i(\sum_{v\in[V]}F(v)f_i(v))(\sum_{w\in[W]}G(w)g_i(w))\right| \leqslant \frac{\varepsilon}{4}|V||W|.$$

On the other hand, if $\lambda_i > \varepsilon/4$, then by construction $f_i(v)$ fluctuates by at most $\frac{\varepsilon^2}{16}$ on each $\mathbf{V}_a$, while $F(v)$ has mean zero and is bounded in magnitude by 1 on each $\mathbf{V}_a$. This implies that

$$|\sum_{v\in[V]}F(v)f_i(v)| \leqslant \frac{\varepsilon^2}{16}|V|$$

for each such i; meanwhile, from Cauchy-Schwarz one has

$$|\sum_{w\in[W]}G(w)g_i(w)| \leqslant |W|.$$

Finally, from (A.2) one has

$$\sum_{i:\lambda_i>\varepsilon/4}\lambda_i < \frac{4}{\varepsilon}$$

and thus

$$|\sum_{i:\lambda_i>\varepsilon/4}\lambda_i(\sum_{v\in[V]}F(v)f_i(v))(\sum_{w\in[W]}G(w)g_i(w))| \leqslant \frac{\varepsilon}{4}|V||W|.$$

Thus the expression (A.4) does not exceed $\frac{\varepsilon}{2}|V||W|$. A similar argument applies for (A.5), and the claim follows.

References

[1] A. Frieze, K. Kannan, *The regularity lemma and approximation schemes for dense problems*, 37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996), 12–20, IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[2] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[3] T. Gowers, *A new proof of Szemerédi's theorem*, GAFA **11** (2001), 465–588.

[4] T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946.

[5] B. Nagle, V. Rödl and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures and Algorithms **28** (2005), 113–179.

[6] V. Rödl, J. Skokan, *Regularity lemma for uniform hypergraphs*, Random Structures and Algorithms **25** (2004), 1–42.

[7] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245–252.

[8] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.

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

[10] E. Szemerédi, *Regular partitions of graphs,* Problémes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), pp. 399–401, Colloq. Internat. CNRS, **260**, CNRS, Paris, 1978.

[11] B. L. van der Waerden, B. L, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk, **15** (1927), 212–216.

UCLA Department of Mathematics, Los Angeles, CA 90095-1555.

Email address: `tao@math.ucla.edu`