# A course in arithmetic Ramsey theory

Frank de Zeeuw  
May 6, 2017

## 1 Coloring Theorems

1.1 Van der Waerden's Theorem ............................................. 2  
1.2 Bounds for Van der Waerden's Theorem .............................. 3  
1.3 The Hales–Jewett Theorem ............................................. 4  
1.4 Schur's Theorem and Rado's Theorem ................................ 7  

## 2 Roth's Theorem

2.1 Hilbert cubes .............................................................. 9  
2.2 Combinatorial proof of Roth's Theorem ............................... 11  
2.3 Bounds for Roth's Theorem .......................................... 13  
2.4 An alternative proof using the regularity lemma .................. 14  

## 3 The Density Hales–Jewett Theorem

3.1 Introduction .............................................................. 15  
3.2 An outline of the proof ................................................ 16  
3.3 Preparation for the proof ............................................. 19  
3.4 The proof of the density Hales–Jewett theorem .................... 22  
3.5 Consequences ........................................................... 28  

## 4 Arithmetic progressions over finite fields

4.1 Introduction .............................................................. 29  
4.2 The slice rank method ................................................ 30  
4.3 Sunflowers ............................................................... 32  
4.4 Arithmetic progressions in \( \mathbb{F}_3^N \) ............................. 33  

---

This is a partial draft of the lecture notes for the first half of the course *Additive Combinatorics* at EPFL.
Chapter 1

Coloring Theorems

1.1 Van der Waerden’s Theorem

The following theorem of Van der Waerden [37] is a central result in arithmetic Ramsey theory. Historically, it was the third theorem of this kind, after the first steps of Hilbert (Lemma 2.1) and Schur (Theorem 1.8). See Soifer [30] for an exhaustive history of Van der Waerden’s theorem. Our presentation of the proof is based on the exposition in lecture notes of Leader [22].

**Theorem 1.1 (Van der Waerden).** For all \(k, \ell\) there is a number \(W(k, \ell)\) such that any \(k\)-coloring of \([W(k, \ell)]\) has a monochromatic arithmetic progression of length \(\ell\).

**Proof.** We define an \((\ell, m)\)-fan to be a set of \(m\) monochromatic \(\ell\)-term arithmetic progressions \(\{a_i, a_i + d_i, \ldots, a_i + (\ell - 1)d_i\}\) of different colors that all have the same focus \(f\), i.e. \(f = a_i + \ell d_i\) for all \(1 \leq i \leq m\). We prove the following claim for all \(k, \ell, m \leq k\).

There is a number \(V(k, \ell, m)\) such that if \([V(k, \ell, m)]\) is \(k\)-colored, then there is a monochromatic \((\ell + 1)\)-term arithmetic progression, or there is an \((\ell, m)\)-fan.

If this claim holds for \(m = k\), then we directly get an \((\ell + 1)\)-term arithmetic progression, or there is an \((\ell, k)\)-fan. In the second case, one of the progressions in the fan must have the same color as the focus, so again there is a monochromatic \((\ell + 1)\)-term arithmetic progression. In other words, if \(V(k, \ell, k)\) exists, then \(W(k, \ell + 1)\) also exists.

We prove the claim by double induction, with an outside induction on \(\ell\) and an inside induction on \(m\). For \(\ell = 1\), we set \(V(k, 1, m) = k + 1\), so that the pigeonhole principle gives a monochromatic 2-term arithmetic progression. Assuming that \(V(k, \ell - 1, m)\) exists for all \(k, m\), we prove by induction on \(m\) that \(V(k, \ell, m)\) exists for all \(k, m\). For \(m = 1\) the existence of \(V(k, \ell, 1)\) follows from the existence of \(V(k, \ell - 1, k)\), since an \((\ell, 1)\)-fan is just a monochromatic \(\ell\)-term arithmetic progression. Assuming that \(V(k, \ell, m - 1)\) exists, we now prove that \(V(k, \ell, m)\) exists.

We set \(V = V(k, \ell, m - 1)\) and

\[
M = 2 \cdot W(k^V, \ell) \cdot V,
\]

and we let \([M]\) be \(k\)-colored. We split the first half of \([M]\) into \(W(k^V, \ell)\) blocks of length \(V\). Define a coloring of this sequence of blocks with \(k^V\) colors, by letting the color of a block be its coloring.

By the choice of the number of blocks, there is an \(\ell\)-term monochromatic arithmetic progression of blocks. In other words, the first numbers of these blocks form an arithmetic
progression with common difference \( D \), and each block is colored in the same way. By the choice of the length of the blocks, each of those blocks contains in its first half a copy of the same \((\ell, m - 1)\)-fan, translated by \( jD \) for some \( 0 \leq j \leq \ell - 1 \). This fan consists of \( m - 1 \) monochromatic \( \ell \)-term progressions \( a_i + jd_i \) with common focus \( f = a_i + \ell d_i \), which is contained in the same block as the fan. We can assume that \( f \) has a different color than the progressions of the fan, since otherwise we would have an \((\ell + 1)\)-term arithmetic progression. For each \( i = 1, \ldots, m - 1 \) there is a color such that the numbers

\[
a_i + j_1d_i + j_2D,
\]

for \( j_1 = 0, \ldots, \ell - 1 \) and \( j_2 = 0, \ldots, \ell - 1 \) all have that color. Then for each \( i \) the numbers

\[
a_i + jd_i + jD
\]

with \( j = 0, \ldots, \ell - 1 \) form a monochromatic arithmetic progression with focus \( f + \ell D \). Moreover, the numbers \( f + jD \) form another monochromatic arithmetic progression, of a different color and with the same focus, so we have an \((\ell, m)\)-fan. This proves the claim.

1.2 Bounds for Van der Waerden’s Theorem

The values of \( W(k, \ell) \) that we can obtain from our proof of Theorem 1.1 are ridiculously large. Even for \( k = 3 \), we would get (roughly speaking) an exponential tower of height \( \ell \). Let us discuss some of the more sensible bounds that are known.

We write \( W(k, \ell) \) for the smallest number \( W \) for which the statement of Theorem 1.1 holds, i.e., the smallest \( W \) such that whenever \( \mathbb{Z}/W \) is \( k \)-colored, there is a monochromatic \( \ell \)-term arithmetic progression. The best general bounds in terms of \( k \) and \( \ell \) are

\[
c \ell^{\ell^{-1}k^\ell} \leq W(k, \ell) \leq 2^{2^{2^{L+g}}},
\]

where \( c \) is some constant. The upper bound is due to Gowers \[15\]. He obtained it from a quantitative improvement of Szemerédi’s Theorem (to be mentioned later in this course), which he proved using heavy analytic methods. The lower bound is an application of the probabilistic method and the Lovász Local Lemma; see for instance \[17\] Section 4.3.

There are many improvements in specific cases, like \( k = 2 \) or \( \ell = 3 \). For 3-term arithmetic progressions, the best bounds are

\[
k^{c \log k} \leq W(3, 3) \leq 2^{ck}.\]

The lower bound is due to Gunderson and Rödl \[19\], and is based on Behrend’s construction \[2\], which we will see later in the course. The upper bound follows from a quantitative version of Roth’s Theorem (Theorem 2.4) proved by Sanders \[28\].

To see at least one proof of a more sensible bound, we give an argument of Solymosi for an upper bound on \( W(3, 3) \). It is weaker than the bound mentioned above, but the proof is far simpler. Solymosi wrote down a version of this argument in his lecture notes \[31\] Theorem 7. It is also mentioned in the book of Tao and Vu \[35\] Remark 6.18, and similar reasoning is used to prove a related theorem by Graham and Solymosi \[18\].

\[^1\]To the best of my knowledge...
Theorem 1.2 (Solymosi). For all $k$, any $k$-coloring of $[k^{2^k+2}]$ has a monochromatic 3-term arithmetic progression.

Proof. Assume $[N]$ is $k$-colored without a monochromatic 3-term arithmetic progression. Let $c_1$ be the most popular color, and let $C_1$ be the set of numbers with that color. Then we have $|C_1| \geq N/k$. Let $d_1$ be half the most popular difference in $C_1$, so we can write $2d_1 = b - a$ for at least

$$\frac{1}{3}\left(\frac{N}{2k}\right) \geq \frac{N}{6k^2}$$

pairs $a < b \in C_1$. Here we used the fact that at least one third of the differences determined by any set of integers are even. Set

$$S_1 = \{a : a - d_1, a + d_1 \in C_1\},$$

so we have $|S_1| \geq N/(6k^2)$. Since there is no monochromatic 3-term arithmetic progression, no number in $S_1$ has the color $c_1$.

We repeat the procedure starting with $S_1$. Let $C_2 \subset S_1$ be the largest color class, with color $c_2$, so we have $|C_2| \geq |S_1|/k \geq N/(4k^3)$. Let $d_2$ be half the most popular difference in $C_2$, so $2d_2 = b - a$ for at least

$$\frac{1}{3}\left(\frac{|C_2|}{2}\right) \geq \frac{2}{3}\left(\frac{N/4k^3}{2}\right) \geq \frac{N}{4^4 k^6}$$

pairs $a < b \in C_2$. Set

$$S_2 = \{s : s - d_2, s + d_2 \in C_2\},$$

so $|S_2| \geq N/(4^4 k^6)$. As above, no number in $S_2$ has the color $c_2$. But the color $c_1$ also does not occur in $S_2$, because $s \in S_2$ forms an arithmetic progression with $s - d_2 - d_1, s + d_2 + d_1$, both of which have color $c_1$ (since $s - d_2, s + d_2 \in C_2 \subset S_1$).

We keep repeating this until we have a set $S_k$, which has none of the colors, so must be empty. But we have (with some calculation and rough overestimation)

$$|S_k| \geq \frac{N}{4k^{2^k+2}} > \frac{N}{k^{2^k+2}},$$

which gives a contradiction if $N \geq k^{2^k+2}$. \hfill $\square$

1.3 The Hales–Jewett Theorem

We now prove a generalization of Van der Waerden’s Theorem, due to Hales and Jewett [20]. The statement may appear a little odd at first, but it turns out to be just the right formulation to obtain various other generalizations of Van der Waerden, which would be hard to obtain otherwise (see for instance Corollary 1.5). We will see that the proof is almost the same as the one we gave for Theorem 1.1.

We say that $L \subset [d]$ is a Hales–Jewett line if there is a nonempty set $I \subset [d]$ and numbers $a_i \in [d]$ for all $i \notin I$, such that

$$L = \{(x_i) \in [d]^d : x_i = x_j \text{ for all } i, j \in I, x_i = a_i \text{ for } i \notin I\}.$$  

For instance, $\{(2, 1, 1, 3, 1), (2, 2, 2, 3, 1), (2, 3, 3, 3, 1)\}$ is a line in $[3]^5$. A line can be represented as a word $w \in ([d] \cup \{\ast\})^d$ that contains at least one star. A line is then obtained
from a word by replacing all the stars in the word by the same number \( j \), for each \( j \in [\ell] \).

The example above can be written as \( 2**31 \).

We call the "last" point of the line, i.e. the one where \( * \) is replaced by \( \ell \), the focus of the line. Given two lines \( L_1 \subset [\ell]^d, L_2 \subset [\ell]^d \), we write \( L \otimes L' \) for the line in \([\ell]^{d_1 + d_2}\) that is obtained by concatenating the words of \( L \) and \( L' \). We use the same notation when one of the lines is a point (a word without stars).

One can think of \([\ell]^d\) as a grid in \( \mathbb{R}^d \), and then a Hales–Jewett line is the intersection of that grid with a line; but it is not true that every geometric line gives a Hales–Jewett line. For example, \{\(1, 3, 2\), \(2, 2, 2\), \(3, 1, 2\)\} lies on a geometric line but is not a Hales–Hewett line. Nevertheless, when it is clear from the context, we use the word "line" for a Hales–Jewett line.

**Theorem 1.3** (Hales–Jewett). For all \( k, \ell \) there is a number \( \text{HJ}(k, \ell) \) such that if \([\ell]^H(k, \ell)\) is \( k \)-colored, then there is a monochromatic Hales–Jewett line.

**Proof.** We define an \( m \)-fan in \([\ell]^d\) to be a set of \( m \) lines in \([\ell]^d\) which are monochromatic in \([\ell − 1]^d\) with different colors, and which all have the same focus \( f \in [\ell]^d \). We will prove the following claim for all \( k, \ell, m \).

There is a number \( M = M(k, \ell, m) \) such that if \([\ell]^M\) is \( k \)-colored, then there is a monochromatic Hales–Jewett line in \([\ell]^M\) or there is an \( m \)-fan in \([\ell]^M\).

If this claim holds for \( m = k \), then we directly get a line in \([\ell]^M\), or there is a \( k \)-fan. In the second case, one of the lines in the fan must have the same color in \([\ell − 1]^M\) as the focus of the fan, so again there is a monochromatic line in \([\ell]^M\). In other words, if \( \text{HJ}(k, \ell, k) \) exists, then \( \text{HJ}(k, \ell) \) also exists.

We prove the claim by double induction, with an outside induction on \( \ell \) and an inside induction on \( m \). Assuming that \( \text{HJ}(k, \ell − 1, m) \) exists for all \( k, m \), we prove by induction on \( m \) that \( \text{HJ}(k, \ell, m) \) exists for all \( k, m \). For \( m = 1 \) the existence of \( \text{HJ}(k, \ell, 1) \) follows from the existence of \( \text{HJ}(k, \ell − 1) \), since a \( 1 \)-fan in \([\ell]^M\) is basically the same as a monochromatic line in \([\ell − 1]^M\). Assuming that \( \text{HJ}(k, \ell, m − 1) \) exists, we now prove that \( \text{HJ}(k, \ell, m) \) exists.

We set

\[
M = \alpha + \beta \quad \text{with} \quad \alpha = \text{HJ}(k^{\ell_1(k, \ell, m − 1)}, \ell − 1), \quad \beta = \text{HJ}(k, \ell, m − 1).
\]

We factor

\[
[\ell]^M = A \times B \quad \text{with} \quad A = [\ell]^\alpha, \quad B = [\ell]^\beta.
\]

Let \([\ell]^M\) be \( k \)-colored. We color \( A \) with \( k^\alpha \) colors by letting the color of \( a \in A \) be the \( k \)-coloring of \( B \) induced by the \( k \)-coloring of \( a \times B \). Since the dimension of \( A \) is \( \text{HJ}(k^\alpha, \ell − 1) \), there is a line \( L \) in \( \ell \) whose restriction to \([\ell − 1]^\alpha\) is monochromatic for this coloring. In other words, for each \( a \in L \), except for the focus of \( L \), we get the same \( k \)-coloring on \( B \), which we denote by \( c_L \).

Because the dimension of \( B \) is \( \text{HJ}(k, \ell, m − 1) \), there is a monochromatic line \( L' \) in \( B \), or there is an \( (m − 1) \)-fan \( F \) in \( B \) for the coloring \( c_L \). In the first case, the line \( f_1 \otimes L' \) is a monochromatic line in \([\ell]^M\).

We obtain an \( m \)-fan as follows. Let \( f_1 \in A \) be the focus of \( L \), and let \( f_2 \in B \) be the focus of \( F \). If \( c_L \) gives \( f_2 \) the same color as one of the lines in \( F \), then we have a monochromatic line \( L' \) in \( B \) for \( c_L \), and \( L \otimes L' \) is a monochromatic line in \( \alpha \times \beta = [\ell]^M \). So we can assume that the \( c_L \)-color of \( f_2 \) is different from the lines in \( F \). For each line \( L_i \) of the \( m − 1 \) lines in \( F \), \( L \otimes L_i \) is a line in \([\ell]^M\) with focus \( f_1 \otimes f_2 \), and these lines are monochromatic in \([\ell − 1]^M\) with different colors. Moreover, \( L \otimes f_2 \) is another monochromatic line with focus \( f_1 \otimes f_2 \), and its color is different from the other \( m − 1 \) lines. 

\( \square \)
Corollary 1.4 (Van der Waerden). For all \( k, \ell \) there is a number \( W(k, \ell) \) such that any \( k \)-coloring of \( [W(k, \ell)] \) has a monochromatic \( \ell \)-term arithmetic progression.

Proof. Set \( W = \ell \cdot HJ(k, \ell) \). We define a map \( \varphi : [\ell]^{HJ(k, \ell)} \to [W] \) by

\[
\varphi(x_1, \ldots, x_{HJ(k, \ell)}) = \sum_{i \in [HJ(k, \ell)]} x_i.
\]

Given a \( k \)-coloring \( c \) of \([W]\), we get a \( k \)-coloring \( c^* \) of \([\ell]^{HJ(k, \ell)}\) by setting \( c^*(x) = c(\varphi(x)) \). Then there is a monochromatic Hales–Jewett line for \( c^* \), which corresponds to a monochromatic \( \ell \)-term arithmetic progression for \( c \) in \([W]\). Indeed, if the word for this line has \( d \) stars, and the first point is \((a_1, \ldots, a_{HJ(k, \ell)})\), then the progression is \((\sum a_i) + jd\) for \( j \in [\ell] \).

Corollary 1.5 (Gallai). For all \( d, k \) and every finite set \( S \subset \mathbb{N}^d \), there is a number \( G = G(d, k, S) \) such that if \(|G|^d\) is \( k \)-colored, then there is a monochromatic set of the form \( v + \lambda S \) for some \( v \in \mathbb{N}^d \) and \( \lambda \in \mathbb{N} \).

Proof. Set \( G = |S| \cdot HJ(k, |S|) \). Write \( S = \{s_1, \ldots, s_{|S|}\} \) (the order does not matter). We define a map \( \varphi : [|S|]^{HJ(k, |S|)} \to [G] \) by

\[
\varphi(x_1, \ldots, x_{HJ(|S|)}) = \sum_{i \in [HJ(|S|)]} s_i.
\]

Given a \( k \)-coloring \( c \) of \(|G|^d\), \( c^*(x) = c(\varphi(x)) \) gives a \( k \)-coloring \( c^* \) of \(|[S]|^{HJ(k, |S|)}\). Then there is a monochromatic line for \( c^* \), which gives a monochromatic homothetic copy of \( S \).

Corollary 1.6 (Squares). For all \( k \) there is a number \( Q = Q(k) \) such that if \([Q]^2\) is \( k \)-colored, then there is a monochromatic square.

Proof. Take \( S = \{(0, 0), (0, 1), (1, 0), (1, 1)\} \) in Corollary 1.5.

We define a Hales–Jewett space in \([\ell]^d\) of dimension \( \delta \) to be a set represented by a word \( w \in (\ell \cup \{*, \ldots, \#\})^d \), with each star \(*_i\) occurring at least once. For instance, in \([3]^2\) the word \(*_12*_22\) represents the two-dimensional Hales–Jewett space

\[
\{1211, 1222, 1233, 2211, 2222, 2233, 3211, 3222, 3233\}.
\]

Corollary 1.7 (Multidimensional Hales–Jewett). For all \( k, \ell, \delta \) there is a number \( M = MHJ(k, \ell, \delta) \) such that if \([\ell]^M\) is \( k \)-colored, then there is a monochromatic \( \delta \)-dimensional Hales–Jewett space.

Proof. The key is that a \( \delta \)-dimensional Hales–Jewett space in \([\ell]^{\delta d}\) can be represented as a Hales–Jewett line in \([\ell^\delta]^d\). Set \( H = HJ(k, \ell^\delta) \) and \( M = \delta \cdot H \). We take any bijection \( \psi : [\ell]^{\delta} \to [\ell^\delta] \), and then define a map \( \varphi : [\ell]^{\delta d} \to [\ell^\delta]^H \) by

\[
\varphi(x_1, \ldots, x_{\delta H}) = (\psi(x_1, \ldots, x_\delta), \ldots, \psi(x_{\delta H - 1}, \ldots, x_{\delta H})).
\]

Given a coloring \( c \) of \([\ell]^M\), \([\ell]^{\delta H}\), we get a coloring \( c^* \) of \([\ell^\delta]^H\) by setting \( c^*(x) = c(\varphi^{-1}(x)) \). By Theorem 1.3, there is a monochromatic line \( L \) for \( c^* \) in \([\ell^\delta]^H\). Then \( \varphi^{-1}(L) \) is a \( \delta \)-dimensional Hales–Jewett space in \([\ell]^{\delta H}\). □
1.4 Schur’s Theorem and Rado’s Theorem

The following theorem was proved by Schur [29] in 1916, before Van der Waerden’s Theorem. A 3-term arithmetic progression \((x, y, z)\) is a solution of the equation \(x + y = 2z\), and Schur’s Theorem is a Ramsey result for a similar linear equation. But note that in an arithmetic progression \(x, y, z\) must be distinct, which we do not require in Schur’s Theorem.

**Theorem 1.8** (Schur). For all \(k\) there is a number \(S(k)\) such that any \(k\)-coloring of \([S(k)]\) has a monochromatic solution of \(x + y = z\).

**Proof.** Given a \(k\)-coloring of \([N]\), we construct a complete graph with vertex set \([N + 1]\), and we color the edge between \(i\) and \(j\) with the color of \(|i - j|\). Since

\[
(i - j) + (j - k) = i - k,
\]

a monochromatic triangle in this graph corresponds to a monochromatic solution of \(x + y = z\) with \(x, y, z \in [N]\). So it suffices to prove the claim that there is a number \(R(k)\) such that if a complete graph on \(R(k)\) vertices is \(k\)-colored, then it has a monochromatic triangle.

We use induction on \(k\). For \(k = 1\) the claim is trivial. Given that \(R(k - 1)\) exists, we show that the claim holds for \(R(k) = k \cdot R(k - 1) + 1\). Pick any vertex \(v\). It has \(R(k) = k \cdot R(k - 1)\) incident edges, so at least \(R(k - 1)\) of these have the same color; let \(E\) be the set of those edges. If two edges of \(E\) have endpoints that are connected by another edge of the same color, then we have a monochromatic triangle. Otherwise, the \(R(k - 1)\) endpoints of \(E\) form a complete subgraph that is colored with \(k - 1\) colors, so we are done by induction. \(\square\)

Rado’s Theorem [26] is a generalization of Schur’s Theorem to other linear equations. We will prove Rado’s Theorem only for a single homogeneous equation, but Rado [26] proved a similar statement for systems of nonhomogeneous equations. Van der Waerden’s Theorem can be seen as a special case of that general version, since arithmetic progressions of any length can be considered as a solution of a system of linear equations (again with the requirement that the components of the solution are distinct).

To prove Rado’s Theorem, it is convenient to first deduce the following generalization of Schur’s Theorem 1.8 from Van der Waerden’s Theorem 1.1. This exposition is based on that of Leader [22].

**Lemma 1.9.** Let \(r \in \mathbb{Q}\). For all \(k\) there is a number \(N\) such that any \(k\)-coloring of \([N]\) has a monochromatic solution of \(x + ry = z\).

**Proof.** We can assume that \(r > 0\), by rearranging the equation if necessary. Write \(r = s/t\) with \(s\), \(t\) positive integers. We use induction on \(k\); for \(k = 1\) the statement is easy.

We show that if \(N\) is big enough for \(k - 1\) colors, then \(W(Ns + 1, k)\) suffices for \(k\) colors. Given a \(k\)-coloring of \([W(k, Ns + 1)]\), there is a monochromatic arithmetic progression \(a, a + d, \ldots, a + Nsd\) of color \(c\). If an element \(itd\) of the progression \(td, 2td, \ldots, Ntd\) has color \(c\), then we are done, because \((a, itd, a + isd)\) would be a monochromatic solution of \(x + (s/t)y = z\). Otherwise, the progression \(td, 2td, \ldots, Ntd\) is colored with \(k - 1\) colors, so by induction it contains a monochromatic solution. \(\square\)

**Theorem 1.10** (Rado). An equation

\[
\sum_{i \in [m]} a_i x_i = 0
\]

with all \(a_i \in \mathbb{Z}\) is regular if and only if there is a non-empty \(l \subset [m]\) such that \(\sum_{i \in l} a_i = 0\).
Proof. First we prove that if $\sum_{i \in I} a_i = 0$ for some $I \subset [m]$, then the equation $\sum a_i x_i = 0$ is regular. Pick any $i_0 \in I$. Consider the equation

$$a_{i_0} x + \left( \sum_{i \not\in I} a_i \right) y + \left( \sum_{i \in I \setminus \{i_0\}} a_i \right) z = 0.$$  \hfill (1.1)

Since $\sum_{i \in I} a_i = 0$, this equation is the same as

$$a_{i_0} x + \left( \sum_{i \not\in I} a_i \right) y - a_{i_0} z = 0,$$

which we can rearrange to

$$x + \frac{a_{i_0}}{a_{i_0}} y = z.$$

By Lemma 1.9 for any $k$ there is an $N$ such that this equation has a monochromatic solution $(x, y, z)$, which is then also a monochromatic solution to (1.1). Then we can construct a monochromatic solution $(x_i)_{i \in [m]}$ to $\sum a_i x_i = 0$ by setting $x_{i_0} = x$, $x_i = y$ for $i \in [m] \setminus I$, and $x_i = z$ for $i \in I \setminus \{i_0\}$.

Next we prove that if $\sum a_i x_i = 0$ is regular, then there is some subsum $\sum_{i \in I} a_i = 0$. To do this we define a special coloring using a prime number $p > \sum |a_i|$. For $x \in \mathbb{N}$, let $c(x)$ be the last non-zero digit in the base $p$ expansion of $x$. More precisely, we write $x = c_i p^i + \cdots + c_1 p + c_0$ with $0 \leq c_j < p$, we set $L(x) = \min\{j : c_j \neq 0\}$, and we set $c(x) = c_{L(x)}$. Then $c$ defines a $(p - 1)$-coloring of $\mathbb{N}$. Since $\sum a_i x_i = 0$ is regular, there is a monochromatic solution $(x_1, \ldots, x_m)$, with $c(x_i) = c_{j_0}$ for all $i$. Set $L = \min\{L(x_i) : i \in [m]\}$ and $I = \{i : L(x_i) = L\}$. Considering $\sum_{i \in [m]} a_i x_i$ in base $p$, we see that $\sum_{i \in I} a_i c_{j_0} \equiv 0 \mod p$, which implies $\sum_{i \in I} a_i \equiv 0 \mod p$, since $0 < c_{j_0} < p$. Because by assumption we have $\sum |a_i| < p$, this implies that $\sum_{i \in I} a_i = 0$, as desired. \qed
Chapter 2
Roth’s Theorem

After having seen several coloring theorems, we now move on to density theorems. Our goal in this chapter is to prove Roth’s Theorem, which states that any sufficiently dense set of integers contains a 3-term arithmetic progression. In the proof of Roth’s Theorem, we will use an easier density theorem for different objects called Hilbert cubes, which we prove first, together with its coloring version.

2.1 Hilbert cubes

A Hilbert cube of dimension \( d \) in \( \mathbb{N} \) is a set of the form

\[
\{x_0 + \sum_{i \in I} x_i : I \subset [d]\}
\]

with generators \( x_0, \ldots, x_d \in \mathbb{N} \) (not necessarily distinct). Note that an arithmetic progression \( a, a+b, \ldots, a+db \) is an example of a Hilbert cube (set \( x_0 = a, x_1 = \cdots = x_d = b \)), so it follows from Van der Waerden’s Theorem that we can find a monochromatic Hilbert cube of any dimension for any coloring of a sufficiently large \( \mathbb{N} \). Nevertheless, we give a direct proof, which shows that cubes are an ideal object for inductive proofs.

Moreover, this lemma has historical value, since it is the very first Ramsey-type theorem. Hilbert [21] proved it in 1892, as a lemma in his proof of what is now known as Hilbert’s Irreducibility Theorem (it states that for every irreducible integer polynomial there are infinitely many ways to substitute integers in a subset of the variables and get another irreducible polynomial; see [36] for an exposition).

**Lemma 2.1 (Hilbert).** For all \( k, d \) there is a number \( H(k, d) \) such that if \([H(k, d)]\) is \( k \)-colored, then there is a monochromatic Hilbert cube of dimension \( d \).

**Proof.** We use induction on \( d \). Assume that \( H = H(k, d - 1) \) exists, and set \( K = k^H + H \). Given a coloring of \([K]\), consider each of the intervals \([i, i + H - 1] \) for \( i \in [k^H + 1] \). There are \( k^H \) ways to color such an interval, so by the pigeonhole principle there are two intervals that are colored in the same way, say the ones starting with \( i_1, i_2 \). By the choice of \( H \), there is a monochromatic \((d - 1)\)-dimensional cube in \([i_1, i_1 + H - 1] \), say with generators \( x_0, \ldots, x_{d-1} \), and its translation by \( i_2 - i_1 \) must then also be monochromatic in the same color. Adding the generator \( x_d = i_2 - i_1 \) to the list of generators gives a \( d \)-dimensional cube that is monochromatic. \( \square \)

Szemerédi [33] proved the following density version of Hilbert’s coloring theorem.
Lemma 2.2 (Szemerédi). Every subset $S \subset [N]$ of size $|S| \geq 4N^{1-1/2^d}$ contains a Hilbert cube of dimension $d$.

Proof. For any set $T \subset [N]$, define

$$T_i = \{ t \in T : t + i \in T \}.$$ 

We have

$$\sum_{i \in [N]} T_i = \binom{|T|}{2},$$

so there is an $i$ such that $|T_i| \geq \binom{|T|}{2}/N \geq |T|^2/(4N)$ (as long as $|T| \geq 2$). We will use this fact repeatedly.

Applying this fact to $S$, we get an $i_1$ such that (using $4N^{1-1/2^d} \geq (4N)^{1-1/2^d}$)

$$|S_{i_1}| \geq \frac{|S|^2}{4N} \geq \frac{(4N)^{1-1/2^d})^2}{4N} = (4N)^{1-1/2^d-1}.$$ 

We write $S_{i_1,i_2} = (S_{i_1})_{i_2}$, etc. Applying the fact to $S_{i_1}$, we get an $i_2$ such that $|S_{i_1,i_2}| \geq (4N)^{1-1/2^{d-2}}$. Repeating this, we get $i_3, \ldots, i_d$ such that

$$|S_{i_1,\ldots,i_d}| \geq (4N)^{1-1/2^{d-d}} = 1.$$ 

Let $i_0$ be any element of $S_{i_1,\ldots,i_d}$. Then $i_0, i_1, \ldots, i_d$ generate a cube of dimension $d$ contained in $S$. Indeed, $i_0 \in S_{i_1,\ldots,i_d}$ means that $i_0, i_0 + i_d \in S_{i_1,\ldots,i_d-1}$, which then implies that $i_0, i_0 + i_{d-1}, i_0 + i_d, i_0 + i_d + i_{d-1} \in S_{i_1,\ldots,i_{d-2}}$, etc. Altogether we see that $i_0 + \sum_{j \in I} i_j$ is in $S$ for any $I \subset [d]$.

We will actually use the lemma in the following form. Note that $\log = \log_2$.

Corollary 2.3. If $N \geq 2^{\log^2(4/\delta)}$, then any $S \subset [N]$ with $|S| \geq \delta N$ contains a Hilbert cube of dimension at least $\frac{1}{4} \log \log N$.

Proof. We choose $d$ to be any integer between $\frac{1}{4} \log \log N$ and $\frac{1}{2} \log \log N$, so in particular $2^d \leq \log^{1/2} N$. We can rearrange the assumption $N \geq 2^{\log^2(4/\delta)}$ to

$$2^{\log^{1/2} N} \geq 4/\delta,$$

which implies

$$N = 2^{\log N} \geq (4/\delta)^{\log^{1/2} N} \geq (4/\delta)^{2^d}.$$ 

This gives $\delta N \geq 4N^{1-1/2^d}$, so we can apply Lemma 2.2 to get a Hilbert cube of dimension $d \geq \frac{1}{4} \log \log N$. 

\hfill $\square$
2.2 Combinatorial proof of Roth’s Theorem

Roth’s Theorem is the density version of Van der Waerden’s Theorem for arithmetic progressions of length 3.

**Theorem 2.4 (Roth).** For every \( \delta \) there exists \( N_0 \) such that, for all \( N > N_0 \), any \( S \subset [N] \) with \( |S| \geq \delta N \) contains an arithmetic progression of length three.

We give a combinatorial proof of Roth’s theorem, which is due to Szemerédi [33]. Here is an outline of the proof.

- **Density increment:** The main strategy is to show that if \( S \) has no 3-term arithmetic progression, then there is a long arithmetic progression \( P \subset [N] \) such that the density of \( S \) on \( P \) is larger than \( \delta \) by a constant depending on \( \delta \). If \( N \) is large enough, we can keep repeating this argument on subprogressions. Since the density cannot become larger than 1, we must encounter a 3-term arithmetic progression in \( S \).

- **Partitioning into progressions:** To find \( P \), we partition \([N] = A \cup B \) so that \( S \cap B = \emptyset \), and we partition \( A \) into long arithmetic progressions \( P_{1}, \ldots, P_{k} \). Then \( S \) has larger density on \( A \), which implies it has larger density on one of the \( P_{i} \). To be able to partition \( A \) into long arithmetic progressions, we need to choose \( B \) and \( d \) so that \( d \) and \( (B + d) \setminus B \) are small. Indeed, we can partition \( A \) by starting with \( d \) arithmetic progressions covering \([N] \), and then removing \( B \), which gives a new progression starting at every \( b + d \notin B \) for which \( b \in B \).

- **Finding a Hilbert cube:** To find \( B \) and \( d \), we take a large Hilbert cube \( C \subset S \) with generators \( c_{0}, \ldots, c_{k} \). Observe that if \( 2c - s \in 2C - S \), then \( c, 2c - s \) is an arithmetic progression with difference \( c - s \). Thus, if \( S \) has no 3-term arithmetic progression, then \( 2C - S \) is a large set disjoint from \( S \). It may not have the partitioning property that we want, but we can find a subset that does.

- **Decomposing the Hilbert cube:** Decompose \( C \) into the cubes \( C_{i} \) with generators \( c_{0}, \ldots, c_{i} \) so that \( C_{i+1} = C_{i} \cup (C_{i} + c_{i+1}) \). Set \( D_{i} = 2C_{i} - S \), which is disjoint from \( S \). We have a sequence \( D_{0} \subset \cdots \subset D_{k} \subset [N] \), so one of the sets \( D_{i+1} \setminus D_{i} = (D_{i} + 2c_{i+1}) \setminus D_{i} \) must be small; we set \( B = D_{i} \) and \( d = 2c_{i+1} \). Then we have \( B \) disjoint from \( S \) with \( (B + d) \setminus B \) small, so we can partition \( A = [N] \setminus B \) into long arithmetic progressions, and on one of these we find the density increment.

In the actual proof, we have to work out the quantitative details, we have to deal with the complication that \( 2C - S \) might not lie inside \([N] \), and we have to ensure that \( d \) is not too large.

**Proof of Theorem 2.4** We assume that \( S \) contains no 3-term arithmetic progression. To deal with the complications mentioned after the outline, we use a convenient partition of \([N] \). This may lead to small losses in density, for which we roughly compensate with factors \( 1/2 \); we omit the elementary details of checking that these adjustments suffice. We split \([N] \) into four parts \([iM + 1, (i + 1)M] \) for \( i = 0, \ldots, 3 \) (with the last part perhaps slightly smaller). We can assume that \( S \) has density at least \( \delta/2 \) on the second part, since otherwise there is a density increment on one of the other three parts, and we can jump to the last paragraph of the proof. We split the second interval \([M + 1, 2M] \) into about \((\log \log N)/2 \) intervals of length between \( N/(2 \log \log N) \) and \( N/(\log \log N) \), and we pick one such interval, say \( I \), on which \( S \) has density at least \( \delta/2 \).
We assume that $N \geq 2^{2\log^2(4/\delta)}$, which implies $|I| \geq N/(2 \log \log N) \geq 2^{\log^2(4/\delta)}$, so that we can use Corollary 2.3 to find a Hilbert cube
\[ C \subset S \cap I \]
of dimension $k \geq (\log \log |I|)/4 \geq (\log \log N)/8$, and with generators $c_0, \ldots, c_k$, where $c_1, \ldots, c_k \leq |I| \leq N/\log \log N$. We let $C_i$ be the cube generated by $c_0, \ldots, c_i$, and we set
\[ D_i = 2C_i - (S \cap [1, M]), \]
so $D_i \subset [M + 1, 4M]$. As observed in the outline, each $D_i$ is disjoint from $S$. We have a sequence
\[ D_0 \subset D_1 \subset \cdots \subset D_k \subset [N], \]
which implies that there is an $i$ such that $|D_{i+1}\setminus D_i| \leq N/k$. Since $D_{i+1} = D_i \cup (D_i + 2c_{i+1})$, this implies $|(D_i + 2c_{i+1})\setminus D_i| \leq N/k$. Setting $B = D_i$ and $d = 2c_{i+1}$ we have
\[ |(B + d)\setminus B| \leq \frac{N}{k} \leq \frac{8N}{\log \log N}. \]
We also have $d \leq 2N/\log \log N$ and
\[ |B| \geq |D_0| \geq |S \cap [1, M]| \geq (\delta/2)M = (\delta/8)N. \]

We can partition $A = [N]\setminus B$ into
\[ d + |(B + d)\setminus B| \leq 10N/\log \log N \]
progressions with difference $d$, as described in the outline. Thus we can write
\[ A = P_1 \cup \cdots \cup P_\ell, \]
with $\ell \leq 10N/\log \log N$, each $P_i$ an arithmetic progression, and all $P_i$ disjoint.

We have $|B| \geq (\delta/8)N$ and thus $|A| \leq (1 - \delta/8)N$. Let $J \subset [\ell]$ be the set of indices $i$ for which $|P_i| \geq (\delta^2/160)\log \log N$. Then we have
\[ \sum_{i \not\in J} |P_i| \leq \frac{10N}{\log \log N} \cdot \frac{\delta^2 \log \log N}{160} \leq \frac{\delta^2}{16}N. \]
Thus we have
\[ \frac{|S \cap (\bigcup_{i \in J} P_i)|}{|A|} \geq \frac{|S| - \left| \left( \bigcup_{i \not\in J} P_i \right) \right|}{|A|} \geq \frac{\delta N - \frac{\delta^2}{160}N}{(1 - \frac{\delta}{8})N} \geq \left( \frac{\delta - \frac{\delta^2}{16}}{1 - \frac{\delta}{8}} \right) \left( 1 + \frac{\delta}{8} \right) \geq \delta + \frac{\delta^2}{32}. \]
This means that $S$ has density $\delta + \delta^2/32$ on the union of the progressions $P_i$ of length at least $\delta^2 \log \log N)/160$, so there is such a $P_i$ on which $S$ has density at least $\delta + \delta^2/32$.

We now repeat the argument for the set $|S \cap P_i|$ as a subset of $P_i$. After repeating less than $32/\delta^2$ times, we would obtain a density greater than 1, so we must have encountered a 3-term arithmetic progression. Here $N$ should be sufficiently large, which means large enough so that after iterating the function $f(x) = (\delta^2 \log \log x)/160$ up to $32/\delta^2$ times, starting with $N$, we still have a value larger than $2^{2\log^2(4/\delta)}$, so that Corollary 2.3 can be applied. This means that $N_0$ is an exponential tower whose height depends polynomially on $1/\delta$. \qed
2.3 Bounds for Roth’s Theorem

There are several other proofs of Roth’s Theorem. Let \( r_3(N) \) be the maximum size of a set \( S \subset [N] \) that contains no 3-term arithmetic progression. We proved that \( r_3(N) = o(N) \). The original proof of Roth [27] used Fourier analysis, and it showed that

\[
r_3(N) \ll N \cdot \frac{1}{\log \log N}.
\]

This bound is significantly stronger than what we get from our proof of Theorem 2.4, and Fourier analysis seems to be the only method that gives strong quantitative bounds for Roth’s Theorem. Roth’s bound has been improved many times over the years, and the current best bound, due to Bloom [6], is

\[
r_3(N) \ll N \cdot \frac{(\log \log)^4}{\log N}.
\]

On the other hand, Behrend [5] proved a lower bound of the form \( r_3(N) \gg N \cdot 2^{-\frac{1}{2}} \log \log N \), which is still the best known, except for a small improvement of Elkin [11].

**Theorem 2.5 (Behrend).** For all \( N \in \mathbb{N} \) there is a set \( S \subset [N] \) of size \( |S| \geq N \cdot 2^{-\frac{1}{2}} \log \log N \) that does not contain an arithmetic progression of length three.

**Proof.** For \( K, L \) to be chosen later, we intersect \([K]^L\) with a sphere of radius \( r \) to get

\[
S_r = \{(x_i) \in [K]^L : x_1^2 + \cdots + x_L^2 = r^2\}.
\]

The spheres \( S_r \) with \( r \in [K^2L] \) cover \([K]^L\), so, by the pigeonhole principle, there is a radius \( r_0 \) such that

\[
|S_{r_0}| \geq \frac{K^L}{K^2L}.
\]

A line intersects a sphere in at most two points, so there are no three distinct points \( x, y, z \in S_{r_0} \) such that \( x + y = 2z \), i.e., \( S_{r_0} \) contains no 3-term arithmetic progression of vectors.

We define an injective map \( \varphi : [K]^L \to [(2K)^L] \) by

\[
\varphi(x_1, \ldots, x_L) = \sum_{i \in [L]} x_i \cdot (2K)^i.
\]

Note that for \( x, y, z \in S_{r_0} \), we have \( x + y = 2z \) if and only if \( \varphi(x) + \varphi(y) = 2\varphi(z) \), so \( \varphi(S_{r_0}) \subset [(2K)^L] \) does not contain a 3-term arithmetic progression. We set \( S = \varphi(S_{r_0}) \).

The rest is calculation. Given \( N \), we set \( L = \lfloor \sqrt{\log N} \rfloor \) and \( K = \frac{1}{2} 2^L \), so that \( (2K)^L \leq N \) and

\[
(2K)^L \geq 2^{L^2} \geq 2^{\left(\sqrt{\log N - 1}\right)^2} \geq N \cdot 2^{-2\sqrt{\log N}}, \quad K^2 L 2^{L^2} \leq L \cdot 2^{3L} \leq 2^{-4\sqrt{\log N}}.
\]

Then

\[
|S| \geq \frac{K^L}{K^2L} = (2K)^L \cdot (K^2 L 2^{L^2})^{-1} \geq N \cdot 2^{-6\sqrt{\log N}},
\]

which finishes the proof. \( \square \)
2.4 An alternative proof using the regularity lemma

We now give an alternative proof of Roth’s Theorem, which also shows a two-dimensional generalization called the Corners Theorem. The proof relies on the regularity lemma, which we will not prove or even state here. We use the following corollary of the regularity lemma, which is much easier to formulate, and we sketch how it would be proved from the regularity lemma.

**Lemma 2.6** (Triangle removal). For every $\alpha > 0$ there is a $\beta > 0$ such that, if a graph $G$ on $n$ vertices has at least $an^2$ edge-disjoint triangles, then it has at least $\beta n^3$ triangles.

**Proof sketch.** By the regularity lemma, for any $\epsilon > 0$, we can remove at most $\epsilon n^2$ edges from $G$, so that we can partition the vertices of $G$ into not too many subsets $X_1, \ldots, X_k$, with each pair $X_i, X_j$ $\epsilon$-regular. This means that, for any $A \subset X_i$ with $|A| > \epsilon |X_i|$ and $B \subset X_j$ with $|B| > \epsilon |X_j|$, the density of edges between them is close to the density of edges between $X_i, X_j$, i.e.,

$$\left| \frac{|E(A, B)|}{|A||B|} - \frac{|E(X_i, X_j)|}{|X_i||X_j|} \right| < \epsilon.$$  

Moreover, we can assume that each pair $X_i, X_j$ has either density zero, or density above some constant, say $|E(X_i, X_j)|/|X_i||X_j| > \epsilon/2$.

If we take $\epsilon < \alpha$, then the number of removed edges is less than the number of edge-disjoint triangles, so some triangle must be left intact. If the vertices of that triangle are in $X_i, X_j, X_k$, then each of these pairs must have density more than $\epsilon/2$. A calculation with the $\epsilon$-regular property then gives that $X_i, X_j, X_k$ determine at least $f(\epsilon)n^3$ triangles for some polynomial function $f$. We set $\beta = f(\epsilon)$. \qed

**Theorem 2.7** (Corners Theorem). For every $\delta$ there exists $N_0$ such that, for all $N > N_0$, any $S \subset [N]^2$ with $|S| \geq \delta N^2$ contains a corner $\{(s, t), (s + d, t), (s, t + d)\}$ with $d \neq 0$.

**Proof.** We define a tripartite graph $G$ on three vertex sets $X, Y, Z$ consisting of lines:

$$X = \{x = i : i \in [N]\}, \quad Y = \{y = j : j \in [N]\}, \quad Z = \{x + y = k : k \in [2N]\}.$$  

We connect two lines from different sets if their intersection point is in $S$. Then a triangle in $G$ corresponds to lines $x = i, y = j, x + y = k$ such that $(i, j), (k - j, j), (i, k - i) \in S$. This is a corner with $(s, t) = (i, j)$ and $d = k - i - j$, except that if $k = i + j$, then the three lines forming the triangle are concurrent; in that case we call the triangle degenerate. We wish to show that there is a non-degenerate triangle in $G$.

For every point of $S$, the three lines from $X, Y, Z$ intersecting at that point give a degenerate triangle in $G$, so $G$ contains at least $\delta N^2$ degenerate triangles, which are edge-disjoint. Since $G$ has $4N$ vertices, we can apply Lemma 2.6 with $\alpha = \delta/4^2$ to get a $\beta > 0$ such that $G$ has at least $\beta(4N)^3$ triangles. There are at most $N^2$ degenerate triangles in $G$, hence as long as $N$ is sufficiently large so that $\beta(4N)^3 > N^2$, there must be a non-degenerate triangle, which means $S$ contains a corner. \qed

**Corollary 2.8** (Roth). For every $\delta$ there exists $N_0$ such that, for all $N > N_0$, any $S \subset [N]$ with $|S| \geq \delta N$ contains a 3-term arithmetic progression.

**Proof.** Given $S \subset [N]$, define $T = \{(x, y) \in [2N]^2 : x - y \in S\}$. Each line $x - y = s$ for $s \in S$ contains at least $N$ points of $[2N]^2$, so $|S| \geq \delta N$ implies $|T| \geq \delta N^2 = (\delta/4)(2N)^2$. Thus, by Theorem 2.7 $T$ contains a corner $\{(s, t + d), (s, t), (s + d, t)\}$ with $d \neq 0$. That means that $s - t - d, s - t, s - t + d \in S$, and this is a 3-term arithmetic progression with common difference $d$. \qed
Chapter 3

The Density Hales–Jewett Theorem

3.1 Introduction

In this chapter we prove the following theorem.

**Theorem 3.1** (Density Hales–Jewett). For all \( \ell > 1, \delta > 0 \) there is a number \( D_0 = DHJ(\ell, \delta) \) such that, for all \( D > D_0 \), any \( S \subset [\ell]^D \) with \( |S| \geq \delta \ell^D \) contains a Hales–Jewett line.

The theorem was first proved by Furstenberg and Katznelson [14] using ergodic methods. A combinatorial proof was found by Polymath [25], which was later simplified by Dodos, Kanellopoulos, and Tyros [9]. The proof we give here is that of Dodos, Kanellopoulos, and Tyros, with some small changes in the exposition.

For a fixed \( \ell \) we denote the statement of the theorem by \( DHJ(\ell) \). We prove by induction on \( \ell \) that \( DHJ(\ell) \) holds for all \( \ell \geq 2 \). For the base case \( \ell = 2 \), observe that we can think of a vector in \([2]^D\) as a set in \([\ell]^D\), and we can think of a set \( S \subset [2]^D \) as a set system \( S^* \subset 2^D \). A Hales–Jewett line in \([2]^D\) then corresponds to sets \( A, B \in 2^D \) such that \( A \subset B \). Thus \( DHJ(2) \) follows from Sperner’s Theorem [32], which says that if \( S^* \subset 2^D \) with \( |S^*| > \left( \frac{D}{3} \right)^D \), then there are \( A, B \in S^* \) with \( A \subset B \).

To show that \( DHJ(\ell) \) implies \( DHJ(\ell + 1) \), we will use a density increment strategy like in our proof of Theorem 2.4. To motivate this strategy, let us first discuss a set-theoretic interpretation of \( DHJ(3) \).

A point \( x \in [3]^D \) can be thought of as a partition of \([D]\) into three sets, i.e.,

\[(X, Y, Z) \text{ with disjoint } X, Y, Z \in 2^D \text{ such that } X \cup Y \cup Z = [D].\]

For instance, we can let \( X \) be the set of indices of the coordinates where \( x \) has a 1, \( Y \) the ones where \( x \) has a 2, and \( Z \) the ones where \( x \) has a 3. This makes \( Z \) redundant, so we could just as well think of disjoint pairs, i.e.

\[(X, Y) \text{ with disjoint } X, Y \in 2^D.\]

In other words, we identify \([3]^D\) with the disjoint product

\[2^D \times 2^D = \{(X, Y) \in 2^D \times 2^D : X \cap Y = \emptyset \} \text{.}\]

What is a Hales–Jewett line of \([3]^D\) in this interpretation? It is a triple of pairs of the form

\[(X, Y), (X \cup W, Y), (X, Y \cup W) \in 2^D \times 2^D,\]
with $X, Y, W$ disjoint. This should remind us of Theorem 2.7 where we saw corners $(x, y), (x + d, y), (x, y + d)$ — a line in $[3]^2$ is a “corner” in $2^{|O|} \times 2^{|O|}$. This suggests that we might be able to use ideas from proofs of Theorem 2.7. The proof of Theorem 2.7 that we gave used the regularity lemma, and it seems possible to use the same idea to prove $DHJ(3)$, but for larger $\ell$ we would then need the hypergraph regularity lemma. Instead, we will dig up the original proof of Theorem 2.7 due to Ajtai and Szemerédi [1]. We will sketch the Ajtai–Szemerédi proof in the next section, and then we will see how to translate it to a proof of $DHJ(3)$. It will turn out that this argument extends to general $\ell$, which is surprising if one compares it to the situation for arithmetic progressions, because the known proofs of Roth’s Theorem do not extend easily to the general case it all.

3.2 An outline of the proof

3.2.1 An argument of Ajtai and Szemerédi

We sketch the original proof of the Corners Theorem (Theorem 2.7) of Ajtai and Szemerédi [1], which uses a density increment argument similar to the one we used for Theorem 2.4. This proof has the awkward feature that it uses Szemerédi’s Theorem, which seems to be a much deeper statement than the Corners Theorem. Nevertheless, the Ajtai–Szemerédi argument is the inspiration for the proof of Theorem 3.1 of Polymath [25] and Dodos, Kanellopoulos, and Tyros [9].

Given a dense set $S \subset [N]^2$, we wish to show that $S$ contains a corner

$$\{(x, y), (x + d, y), (x, y + d)\}$$

with $d \neq 0$. We take a diagonal $D$ defined by an equation of the form $x + y = t$ on which $S$ is dense. Let $U$ be the set of $x$-coordinates of the points of $S$ on $D$, and let $V$ be the set of $y$-coordinates of points of $S$ on $D$. If $S$ does not contain a corner, then $(U \times V) \cap S = D \cap S$, so $S$ has a relatively low density on $U \times V$.

The sets

$$U \times V, \quad U \times V^c, \quad U^c \times [N]$$

partition $[N]^2$ (here the superscript $c$ denotes the complement within $[N]$), so the fact that $S$ has low density on $U \times V$ implies that $S$ has a density increment on $U \times V^c$ or $U^c \times [N]$.

We denote the product on which $S$ has a density increment by $X \times Y$. We wish to partition $X \times Y$ into large sets that are similar to $[M]^2$ for some $M < N$, since then we can repeat the argument. Let us define a grid in $[N]^2$ to be a Cartesian product in which both factors are arithmetic progressions with the same length and the same difference; in other words, a scaled and translated copy of some $[M]^2$.

First we show how a set of the form $X \times [N]$ with a dense $X \subset [N]$ can be partitioned into large grids. By Szemerédi’s Theorem, there is a long arithmetic progression $P_1$ in $X$; we remove that progression, and then find another progression $P_2$; we repeat this to find progressions $P_i$ and remove them until the remaining set is too small to apply Szemerédi’s Theorem. For each $i$, we then partition most of $P_i \times [N]$ into grids of the form $P_i \times \tilde{P}_i$, where each $\tilde{P}_i$ is a translate of $P_i$. As a result, we have partitioned most of $X \times [N]$ into grids of the form $P_i \times \tilde{P}_i$.

To partition $X \times Y$ into grids, we first partition $X \times [N]$ into grids $P_i \times \tilde{P}_i$ as above. Then we consider

$$(X \times Y) \cap (P_i \times \tilde{P}_i) = P_i \times (Y \cap \tilde{P}_i),$$

16
which is a product of an arithmetic progression with a set that is (typically) dense on a translate of that arithmetic progression. Thus we can use the same argument as above to partition $P_i \times (Y \cap P_j)$ into grids.

As a result, after discarding a relatively small number of points, we have partitioned $X \times Y$ into grids that are fairly large. The density increment of $S$ on $X \times Y$ must also occur on one of these subgrids, so we can iterate the argument with a larger $\delta$. Since the density cannot increase to more than 1, we must encounter a corner contained in $S$.

### 3.2.2 Applying the Ajtai–Szemerédi argument to $[3]^D$

We now discuss how we can apply the argument that we sketched in Section 3.2.1 to a dense set $S \subseteq [3]^D$, which as in Section 3.1 we think of as a dense set $S \subseteq 2^{[D]} \otimes 2^{[D]}$. Along the way, we will point out how to generalize the concepts to $[\ell]^D$.

**Finding a good diagonal.** The goal is to find a density increment of $S$ on a set that is equivalent to $2^E \otimes 2^E$ for some $E < D$. The analogue of a diagonal $x + y = t$ is the set of $(X, Y)$ satisfying

$$X \cup Y = T$$

for some fixed $T$. However, not every two pairs $(X_1, Y_1), (X_2, Y_2)$ on such a diagonal are the endpoints of a corner $(X, Y), (X \cup W, Y), (X, Y \cup W)$; this is only the case if we have $X_1 \subset X_2$ or vice versa. Therefore, it is not enough to find a diagonal with many points of $S$, we also have to ensure that this diagonal contains many pairs $(X_1, Y_1), (X_2, Y_2)$ with $X_1 \subset X_2$.

Note that within a diagonal, we may as well ignore the second set $Y$, so we can think of the points on the diagonal as sets $X \in 2^T$. Our goal is then to find a $T$ such that many subsets $X_1, X_2 \in 2^T$ with $X_1 \subset X_2$ come from points in $S$. This is one place where we will use induction, since $DHJ(2)$ (i.e., Sperner’s Theorem) gives us a way to find a pair $X_1 \subset X_2$ in a dense set system. With some more work, we can find a subset $T$ for which there are many such pairs.

Let us translate this part of the argument back into the Hales–Jewett language. The subproduct $2^T \otimes 2^T \subseteq 2^{[D]} \otimes 2^{[D]}$ corresponds to a Hales–Jewett subspace of dimension $|T|$ in $[3]^D$, which we can think of as $[3]^{|T|}$. The diagonal $X \cup Y = T$ then corresponds to $[2]^{|T|} \subseteq [3]^{|T|}$. A pair of points on the diagonal that can be completed to a corner is a pair $X_1, X_2 \in T$ with $X_1 \subset X_2$, which is the same as a line in $[2]^{|T|}$. Therefore, our goal is to use $DHJ(2)$ to find a large subspace $[3]^E \subseteq [3]^D$ such that $S$ contains many lines of $[2]^E$.

In general, given a dense set $S \subseteq [\ell]^D$ we will use $DHJ(\ell - 1)$ to find a large subspace $[\ell]^E$ such that $S$ contains many lines of $[\ell - 1]^E$. Of course, a direct application of $DHJ(\ell - 1)$ only gives a single line, but with some work we can use it to obtain many lines.

**Obtaining a density increment.** Suppose we have a diagonal $X \cup Y = T$ as above, containing many pairs $(X_1, Y_1), (X_2, Y_2) \in S$ with $X_1 \subset X_2$. We can rewrite such a pair as $(X, Y \cup W), (X \cup W, Y)$ for a nonempty set $W$. If $S$ contains no corner, then it follows that the point $(X, Y)$ is not in $S$. We define

$$U = \{X \in 2^D : (X, T \setminus X) \in S\} \quad \text{and} \quad V = \{Y \in 2^D : (T \setminus Y, Y) \in S\},$$

so that many points of $U \otimes V$ are not in $S$. We partition $2^D \otimes 2^D$ into the sets

$$U \otimes V, \quad U \otimes V^c, \quad U^c \otimes 2^D,$$
where the superscript \( c \) denotes the complement within \( 2^{[D]} \). Then \( S \) must have a density increment on \( U \times V^c \) or \( U^c \times 2^{[D]} \). The latter is of the form \( W \times 2^{[D]} \), and the former is an intersection of two sets of the form \( W \times 2^{[D]} \) and \( 2^{[D]} \times W' \), so it suffices to know how to partition sets of the form \( W \times 2^{[D]} \).

Again we translate this into Hales–Jewett language. The set \( W \times 2^{[D]} \) consists of pairs \((X, Y)\) with \( X \in W \) and \( Y \) arbitrary (but disjoint from \( X \)), so it has the property that if we add or remove elements of \( Y \) (while maintaining disjointness from \( X \)), we obtain another element of \( W \times 2^{[D]} \). The corresponding set \( C \subset [3]^D \) has the property that if in any vector of \( C \) we change any 2s to 3s or vice versa, we get another vector in \( C \). Let us define this key property in general.

**Definition 3.2.** A set \( C \subset [\ell]^D \) is \( ij \)-insensitive (for \( i < j \in [\ell] \)) if, whenever in a vector of \( C \) we change some is to js and/or we change some js to is, then the resulting vector is in \( C \).

Thus a set like \( W \times 2^{[D]} \) corresponds to a 23-insensitive set in \( [3]^D \), while \( 2^{[D]} \times W \) corresponds to a 13-insensitive set in \( [3]^D \). In the partition of \( 2^{[D]} \times 2^{[D]} \) into the parts \( U \times V, U \times V^c, U^c \times 2^{[D]} \), each part is a disjoint product, which corresponds in \( [3]^D \) to the intersection of a 23-insensitive set and a 13-insensitive set. This generalizes to the following definition.

**Definition 3.3.** A set \( C \subset [\ell]^D \) is Cartesian if \( C = C_1 \cap \cdots \cap C_{\ell-1} \), where for each \( i \in [\ell-1] \) the set \( C_i \) is \( i\ell \)-insensitive.

In the proof for general \( \ell \), we will partition \( [\ell]^E \) into parts, each of which is a Cartesian set. The properties of the “diagonal” that we found will imply that \( S \) has low density on one of the Cartesian sets in the partition, which then implies that it has a density increment on one of the other Cartesian sets.

**Partitioning into subspaces.** In the last step, we need to partition a set of the form \( W \times 2^{[D]} \) into subsets that are of the form \( 2^{[E]} \times 2^{[E]} \) for some \( E < D \). As in the Ajtai–Szemerédi proof, this can then be used to partition \( U \times V^c \) or \( U^c \times 2^{[D]} \).

Here we again use induction, but now we need a multidimensional version of \( DHJ(2) \), which will follow from \( DHJ(2) \) itself. In set language, it tells us that a dense subset of \( 2^{[D]} \) contains a “Hilbert cube” of sets, i.e., a subset of the form

\[
\{ S_0 \cup \bigcup_{i \in I} S_i : I \subset [E] \}
\]

for disjoint sets \( S_0, \ldots, S_E \in 2^{[D]} \). Such a set system is equivalent to \( 2^{[E]} \). In Hales–Jewett language, it is a subspace of \([2]^{D}\).

For general \( \ell \), we find a density increment for \( S \) on a Cartesian set, and we partition that Cartesian set into subspaces of sufficiently large dimension. Then \( S \) must have a density increment on one of those subspaces, which is equivalent to some \([\ell]^{F}\) for a sufficiently large \( F \). Then we can repeat the argument in that \([\ell]^{F}\), so, since the density cannot keep increasing, we must encounter a Hales–Jewett line that is contained in \( S \).
3.3 Preparation for the proof

In this section we prove some results that play an important role in the proof of Theorem 3.1 but that are somewhat independent from the main argument.

3.3.1 Coloring Hales–Jewett lines

The proof will use a variant of the Hales–Jewett theorem, in which we we color the lines instead of the points. It was first proved by Graham and Rothschild [16] in a much more general form, but here we prove only the specific case that we need.

**Theorem 3.4.** For all \( \ell, m \) there is a number \( GR(\ell, m) \) such that for all \( D \geq GR(\ell, m) \) the following holds. If the lines in \( [\ell]^D \) are 2-colored, then there is an \( m \)-dimensional Hales–Jewett space \( V \subset [\ell]^D \) such that all lines in \( V \) have the same color.

The proof requires two lemmas. We define a map \( \mathcal{X} \) from the set of lines \( L \) in \( [\ell]^D \) to \( 2^{|D|} \) by letting \( \mathcal{X}(L) \) be the set of active coordinates of \( L \); more formally, if \( L \) is represented by the word \( (L_i)_{i \in [D]} \in ([\ell] \cup \{\ast\})^D \), then

\[
\mathcal{X}(L) = \{ i \in [D] : L_i = \ast \}.
\]

**Lemma 3.5.** For any \( \ell, m \) there exists a number \( N = N(\ell, m) \) such that the following holds. For any 2-coloring \( c \) of \( [\ell]^N \) there is an \( m \)-dimensional subspace \( W \subset [\ell]^N \) such that the color of a line in \( W \) depends only on the positions of its active coordinates, i.e., for any lines \( L, L' \subset W \) we have that \( \mathcal{X}(L) = \mathcal{X}(L') \) implies \( c(L) = c(L') \).

**Proof.** We set \( \lambda_1 = (\ell + 1)^{m-1} \) and \( a_1 = HJ(2^{\lambda_1}, \ell) \), and then for \( i \leq m \) we recursively define

\[
\lambda_i = (\ell + 1)^{m-i+\sum_{j=1}^{i-1} a_j}, \quad a_i = HJ(2^{\lambda_i}, \ell).
\]

We write \( N = \sum_{i \in [m]} a_i \) and

\[
[\ell]^N = A_1 \times \cdots \times A_m,
\]

where \( A_i = [\ell]^{a_i} \).

Given a 2-coloring of \( [\ell]^N \), we find a line in each \( A_i \) by starting with \( A_m \) and working backwards until \( A_1 \). We color \( A_m \) by giving \( x \in A_m \) a color that encodes all colors of all lines in \( A_1 \times \cdots \times A_{m-1} \times x \). There are \( \lambda_m \) such lines, so this requires \( 2^{\lambda_m} \) colors. Since \( a_m = HJ(2^{\lambda_m}, \ell) \), there is a monochromatic line \( L_m \subset A_m \). This means that for each \( x \in L_m \), the lines in \( A_1 \times \cdots \times A_{m-1} \times x \) are colored in the same way.

Recursively we do the following. We color \( A_i \) by giving \( x \in A_i \) a color that encodes all colors of all lines in

\[
A_1 \times \cdots \times A_{i-1} \times x \times L_{i+1} \times \cdots \times L_m.
\]

There are \( \lambda_i \) such lines, so this requires \( 2^{\lambda_i} \) colors. Since \( a_i = HJ(2^{\lambda_i}, \ell) \), there is a monochromatic line \( L_i \subset A_i \).

When we are done, we set \( W = L_1 \times \cdots \times L_m \), which is an \( m \)-dimensional subspace of \( [\ell]^N \). Consider distinct lines \( L, L' \subset W \) with \( \mathcal{X}(L) = \mathcal{X}(L') \). Then \( L, L' \) must have fixed coordinates in the coordinates corresponding to some \( L_i \subset A_i \), where \( L \) has a value \( x_i \in L_i \) and \( L' \) has a value \( x'_i \in L_i \). By the choice of \( L_i \), all lines in \( L_1 \times \cdots \times L_{i-1} \times x_i \times L_{i+1} \times \cdots \times L_m \) and in \( L_1 \times \cdots \times L_{i-1} \times x'_i \times L_{i+1} \times \cdots \times L_m \) are colored in the same way, which implies that \( L \) and \( L' \) have the same color. □
Lemma 3.6 (Finite Unions Theorem). For any $m$ there exists a number $F = FU(m)$ such that for any 2-coloring of $2^F$ there are disjoint subsets $X_1, \ldots, X_m \subseteq 2^F$ such that every union $\bigcup_{i \in I} X_i$ for a nonempty $I \subseteq [m]$ has the same color.

Proof. We prove the stronger statement that for any $p, q$ there is an $FU(p, q)$, such that if $2^{FU(p, q)}$ is colored red and blue, then there are $X_1, \ldots, X_p$ with all unions red, or there are $Y_1, \ldots, Y_q$ with all unions blue. For $p = q = m$ this is the statement of the theorem. We prove it by induction on $p + q$, with the statement being trivial if $p = 1$ or $q = 1$.

Set $F = FU(p, q) = MHJ(2, 2, FU(p - 1, q))$, where the function $MHJ$ comes from the Multidimensional Hales–Jewett Theorem (Corollary 1.7). Given a 2-coloring of $2^F$, we get a 2-coloring of $[2]^F$, so Corollary 1.7 gives a monochromatic $FU(p - 1, q)$-dimensional subspace of $[2]^F$. This subspace corresponds in $2^F$ to disjoint sets $S_0, S_1, \ldots, S_{FU(p - 1, q)}$ such that

$$S_0 \cup \bigcup_{i \in I} S_i$$

has the same color $c_1$ for every $I \subseteq [FU(p - 1, q)]$.

Suppose $c_1$ is red. We can consider the set of all unions of $S_1, \ldots, S_{FU(p - 1, q)}$ as an instance of $2^{FU(p - 1, q)}$, so by induction, either there are $q$ sets $S_{i_1}, \ldots, S_{i_q}$ such that all unions are blue, and we are done, or there are $p - 1$ sets $S_i, \ldots, S_{i_{p - 1}}$ such that all unions are red. In the second case, the $p$ sets $S_0, S_{i_1}, \ldots, S_{i_{p - 1}}$ all have all unions red, so we are also done. If $c_1$ is blue instead of red, we observe that $FU(p - 1, q) = FU(p, q - 1)$ and we argue symmetrically.

Proof of Theorem 3.4. Set $GR(\ell, m) = N(\ell, FU(m))$ and take $D \geq GR(\ell, m)$. Given a 2-coloring $c$ of the lines in $[\ell]^D$, Lemma 3.5 gives an $FU(m)$-dimensional subspace $W \subseteq [\ell]^D$, which we can think of as $[\ell]^{FU(m)}$, in which $X \subseteq W$ implies $c(X) = c(L)$. Thus we can define a coloring $c^*$ on $[\ell]^{FU(m)}$ by setting $c^*(X) = c(L)$, where $X \subseteq [\ell]^{FU(m)}$ and $L$ is any line in $[\ell]^{FU(m)}$ such that $X \subseteq L$. Lemma 3.6 then gives disjoint sets $X_1, \ldots, X_m \subseteq [\ell]^{FU(m)}$ with all unions the same color. These correspond to disjoint sets $Y_1, \ldots, Y_m$ of coordinates in $[\ell]^D$ (each coordinate of $[\ell]^{FU(m)}$ corresponds to some $* _i$ in the word for $W$, which corresponds to one or more coordinates in $[\ell]^D$), which define an $m$-dimensional subspace $V \subseteq W$ (the word for $W$ has $* _i$ in the coordinates of $Y_i$), and the unions of the $Y_i$ correspond to all lines in $W$. By construction of $c^*$, all these lines have the same color.

3.3.2 The multidimensional density Hales–Jewett theorem

We now prove that $DHJ(k)$ implies a multidimensional version of itself. This fact will play an important role in the proof of $DHJ(k + 1)$.

Lemma 3.7. Assume that $DHJ(k)$ holds. For all $n, \eta$ there is a number $MDHJ(k, n, \eta)$ such that for $E \geq MDHJ(k, n, \eta)$ the following statement holds. If $X \subset [k]^E$ with $|X| \geq \eta k^E$, then there is a subspace $V \subset [k]^E$ of dimension $n$ such that $V \subset X$.

Proof. We use induction on $m$. For $n = 1$, the statement is $DHJ(k)$. Take $\beta = DHJ(k, \eta / 2)$ and $\alpha \geq E(n - 1, \eta / (2(k + 1)^\beta))$, and set $E = \alpha + \beta$. Consider $X \subset [k]^\beta = [k]^\alpha \times [k]^\beta$ with $|X| \geq \eta k^{\alpha + \beta}$. For $x \in [k]^\beta$, we write $X_x = \{ y \in [k]^\beta : x \times y \in X \} \subset [k]^\beta$. We have

$$\sum_{x \in [k]^\alpha} |X_x| = |X| \geq \eta k^{\alpha + \beta}.$$
Lemma 3.8. For all $k$, $M$ and $\eta_1 > \eta_2$ there is a number $E(k, m, \eta_1, \eta_2)$ such that the following holds for all $E > E(k, M, \eta_1, \eta_2)$. If $X \subset [k]^E$ with $|X| \geq \eta_1 k^E$, then there is an $\alpha < E$ and an $M$-dimensional subspace $V \subset [k]^\alpha$ such that

$$|\{y \in [k]^{E-\alpha} : x \times y \in X\}| \geq (\eta_1 - \eta_2)k^{E-\alpha}$$

for all $x \in V$.

Proof. We use a density increment argument. Whatever $\alpha$ is, we write $X_\alpha = \{y \in [k]^{E-\alpha} : x \times y \in X\}$.

If $\alpha = M$ and $V = [k]^M$ satisfy the requirements, then we are done; otherwise there is an $x \in [k]^M$ such that $|X_x| < (\eta_1 - \eta_2)k^{E-\alpha}$, which implies that there is an $x_1 \in [k]^M$ with $|X_{x_1}| \geq (\eta_1 + \eta_2/k^M)k^{E-\alpha}$. Now consider $\alpha = 2M$ and $V = x_1 \times [k]^M$; again we are done unless there is an $x_2 \in x_1 \times [k]^M$ such that $|X_{x_2}| \geq (\eta_1 + 2\eta_2/k^M)k^{E-\alpha}$. Since $E > M k^M/\eta_2$, we can repeat this $M/k^M/\eta_2$ times, but the density cannot increase above 1, so we must encounter a $V$ with the claimed property. □

Given an $n$-dimensional subspace $W \subset [k+1]^E$ we write $\rho(W)$ for its restriction, which is the subset of $W$ where the active coordinates take only the values $1, \ldots, k$ (and not $k+1$). Then $\rho(W)$ can be viewed as a copy of $[k]^n$ inside $W$. Note that the fixed coordinates in $\rho(W)$ may still take the value $k+1$. The following lemma is a variant of Lemma 3.7 where the given set $X$ is dense in $[k+1]^E$, and we want to find a subspace $W$ that itself need not be contained in $X$, but so that we do have $\rho(W) \subset X$. It does not follow directly from Lemma 3.7 since $X$ need not be dense in $[k+1]^E$.

Lemma 3.9. Assume that $\mathcal{D\cup J}(k)$ holds. For all $n, \eta$ there is a number $MD\mathcal{HJ}^*(k, n, \eta)$ such that for $E \geq MD\mathcal{HJ}^*(k, n, \eta)$ the following holds. If $X \subset [k+1]^E$ with $|X| \geq \eta(k+1)^E$, then there is a subspace $W \subset [k+1]^E$ of dimension $n$ such that $\rho(W) \subset X$.

Proof. Set $M = MD\mathcal{HJ}(k, n, \eta/2)$ and $E = E(k, M, \eta, \eta/2)$. By Lemma 3.8 there exist an $\alpha < E$ and an $M$-dimensional subspace $V_1 \subset [k]^\alpha$ such that, with $\beta = E - \alpha$, we have $|X_\alpha| \geq (\eta/2)k^\beta$ for every $x \in V_1$. Then we get

$$|X \cap (\rho(V_1) \times [k]^\beta)| = \sum_{x \in \rho(V_1)} |X_x| \geq (\eta/2)k^\beta \cdot |\rho(V_1)| = (\eta/2) \rho(V_1) \times [k]^\beta|.$$

In other words, $X$ is $(\eta/2)$-dense on $\rho(V_1) \times [k]^\beta$. The sets $\rho(V_1) \times y$ for $y \in [k]^\beta$ partition $\rho(V_1) \times [k]^\beta$, so there is a $y_0 \in [k]^\beta$ such that $X$ is $(\eta/2)$-dense on $\rho(V_1) \times y_0$, i.e.,

$$|X \cap (\rho(V_1) \times y_0)| \geq (\eta/2)\rho(V_1) \times y_0|.$$

By Lemma 3.7 since $M = MD\mathcal{HJ}(k, n, \eta/2)$ and $\rho(V_1) \times y_0$ is an $M$-dimensional space, there is an $n$-dimensional subspace $V_2 \subset \rho(V_1) \times y_0$ such that $V_2 \subset X$. Let $W$ be the $n$-dimensional space in $[k+1]^E$ that contains $V_2$, so that $\rho(W) = V_2 \subset X$. □
3.4 The proof of the density Hales–Jewett theorem

In this section we prove Theorem 3.1 in the three steps which were outlined in Section 3.2.2 for the case \( \ell = 3 \). As mentioned in Section 3.1 we use induction on \( \ell \), and the base case \( \mathcal{DHJ}(2) \) follows from Sperner’s Theorem. We assume throughout this section that \( \mathcal{DHJ}(\ell - 1) \) holds, and we will deduce \( \mathcal{DHJ}(\ell) \). We assume that \( S \subset [\ell]^D \) satisfies \( |S| \geq \delta^D \), with \( D > D_0 \) for some \( D_0 \) specified below, and that \( S \) contains no Hales–Jewett line. The goal is to find a density increment for \( S \), specifically a Hales–Jewett subspace \( V \subset [\ell]^D \) such that \( |S \cap V| \geq (\delta + f(\delta))\ell^D \), where \( V \) has at least a certain dimension \( d \), and \( f(\delta) \) is an increasing function of \( \delta \).

In the three subsections that follow, we will work through several claims, and there will be several parameters to keep track of. Let us give the definitions of all these parameters here, so that they are easy to find, although the reader is of course advised to ignore these formulas until they turn up in the proof (the function \( GR \) comes from Theorem 3.4, the function \( E \) comes from Lemma 3.8) and the function \( F \) will be specified at the end of Section 3.4.3.

\[
\delta_1 = \delta / (8\ell^{DHJ(\ell-1, \delta/4)}) , \quad \delta_2 = \delta \cdot \delta_1 / 10 , \quad \delta_3 = \delta_2 / \ell , \quad m = \max \left\{ DHJ(\ell - 1, \delta/4), \log_{(\ell - 1)/\ell}(\delta_2), F(\ell, \delta_3, d) \right\}, \quad D_0 = E(\ell - 1, GR(m), \delta, \delta^2_3).
\]

Some terminology. We use the term \((\ell - 1)\)-space to refer to a Hales–Jewett space in \([\ell]^D\) whose active coordinates run only from 1 to \( \ell - 1 \) (but whose fixed coordinates may still take the value \( \ell \)). Sometimes we may use the term \( \ell \)-space in \([\ell]^D\) to refer to a space, just to distinguish it from an \((\ell - 1)\)-space. An \( \ell \)-space \( V \subset [\ell]^D \) contains one maximal \((\ell - 1)\)-space, which we denote by \( \rho(V) \). Similarly, we talk about \((\ell - 1)\)-lines and the \((\ell - 1)\)-line \( \rho(L) \) corresponding to an \( \ell \)-line \( L \).

### 3.4.1 Finding a subspace where \( S \) has many \((\ell - 1)\)-lines

Our goal in this subsection is to find a subspace \( V \subset [\ell]^D \) such that \( S \) contains a dense subset of all \((\ell - 1)\)-lines in \( \rho(V) \). Our first claim is a direct application of Lemma 3.8 using the assumption that \( D > D_0 = E(\ell - 1, GR(m), \delta, \delta^2_3) \).

**Claim.** There is an \( \alpha < D \) and a \( GR(m) \)-dimensional subspace \( V_1 \subset [\ell]^\alpha \) such that for all \( x \in V_1 \) we have

\[
|\{y \in [\ell]^{D-\alpha} : x \times y \in S\}| \geq (\delta - \delta^2_3)\ell^{D-\alpha}.
\]

From now on we write \( \beta = D - \alpha \) and \([\ell]^D = [\ell]^{\alpha} \times [\ell]^\beta\). For \( x \in [\ell]^{\alpha} \) we write

\[
S_x = \{y \in [\ell]^\beta : x \times y \in S\},
\]

so by the claim we have \( |S_x| \geq (\delta - \delta^2_3)\ell^\beta \). Similarly, for an \((\ell - 1)\)-line \( L \) in \([\ell]^\alpha\), we define

\[
S_L = \{y \in [\ell]^\beta : L \times y \subset S\}.
\]
Claim. There is an \(m\)-dimensional subspace \(V_2 \subset V_1\) such that for every \((\ell - 1)\)-line \(L\) in \(\rho(V_2)\) we have \(|S_L| \geq 2\delta_1\ell^B\).

Proof. We use Theorem \ref{thm:3.4} We color an \((\ell - 1)\)-line \(L\) in \(\rho(V_1)\) red if
\[
|S_L| \geq 2\delta_1\ell^B,
\]
and blue otherwise. By Theorem \ref{thm:3.4} applied to \(\rho(V_1)\), which has dimension \(GR(m)\), \(V_1\) has an \(m\)-dimensional subspace \(V_2\) such that the \((\ell - 1)\)-lines in \(\rho(V_2)\) are either all red or all blue. We show that \(\rho(V_2)\) has at least one red line, which then implies that all lines in \(\rho(V_2)\) are red.

Since \(m \geq DHJ(\ell - 1, \delta/4)\), we can pick an \((\ell - 1)\)-subspace \(W \subset \rho(V_2)\) that has dimension exactly \(n = DHJ(\ell - 1, \delta/4)\). Since \(W \subset V_1\), we have \(|S_L| \geq (\delta - \delta^2\ell^B \geq (\delta/2)\ell^B\) for all \(x \in W\). Thus
\[
\sum_{y \in \ell^B} |S \cap (W \times y)| = |S \cap (W \times [\ell^B])| = \sum_{x \in W} |S_x| \geq (\delta/2)\ell^B \cdot \ell^n.
\]

Let \(Y\) be the set of \(y \in [\ell^B]\) such that \(|S \cap (W \times y)| \geq (\delta/4)\ell^n\). Then we have
\[
\sum_{y \in Y} |S \cap (W \times y)| \geq \sum_{y \in [\ell^B]} |S \cap (W \times y)| - \sum_{y \notin Y} |S \cap (W \times y)|
\]
\[
\geq (\delta/2)\ell^B \cdot \ell^n - (\delta/4)\ell^B \cdot \ell^n = (\delta/4)\ell^B \cdot \ell^n.
\]

Since for every \(y\) we have \(|S \cap (W \times y)| \leq |W| = \ell^n\), we get \(|Y| \geq (\delta/4)\ell^B\).

For any \(y \in Y\), since \(W \times y\) is an \((\ell - 1)\)-space of dimension \(n = DHJ(\ell - 1, \delta/4)\) on which \(S\) is \((\delta/4)\)-dense, there is an \((\ell - 1)\)-line \(L_y \subset W\) such that \(L_y \times y\) is contained in \(S\). The number of \((\ell - 1)\)-lines in \(W\) is less than \(\ell^n\), so there is an \((\ell - 1)\)-line \(L_0\) in \(W\) that occurs as \(L_y\) for at least \((\text{using } \delta_1 = \delta/(8\ell^n))\)
\[
|Y|/\ell^n \geq 2\delta_1\ell^B
\]
different \(y \in Y\). This means that \(|S_{L_0}| \geq 2\delta_1\ell^B\), so \(L_0\) is a red line contained in \(\rho(V_2)\). By the choice of \(V_2\), this implies that all lines in \(\rho(V_2)\) are red, i.e., they satisfy \(|S_L| \geq 2\delta_1\ell^B\). \(\square\)

Claim. There is a set \(Y_1 \subset [\ell^B]\) with \(|Y_1| > \delta_1\ell^B\) such that for all \(y \in Y_1\) we have \(L \times y \subset S\) for a \(\delta_1\)-dense subset of all \((\ell - 1)\)-lines in \(\rho(V_2)\).

Proof. Let us write \(\mathcal{L}\) for the set of \((\ell - 1)\)-lines in \(\rho(V_2)\). We have \(|S_L| \geq 2\delta_1\ell^B\) for every \((\ell - 1)\)-line \(L\) in \(V_2\), so we have
\[
\sum_{y \in [\ell^B]} |\{L \in \mathcal{L} : y \in S_l\}| = \sum_{L \in \mathcal{L}} |S_L| \geq 2\delta_1\ell^B \cdot |\mathcal{L}|.
\]

Note that \(y \in S_l\) if and only if \(L \times y \subset S\). Let \(Y_1\) be the set of \(y \in [\ell^B]\) such that \(|\{L \in \mathcal{L} : L \times y \subset S\}| \geq \delta_1|\mathcal{L}|\). Then we have
\[
\sum_{y \in Y_1} |\{L \in \mathcal{L} : L \times y \subset S\}| \geq \sum_{y \in [\ell^B]} |\{L \in \mathcal{L} : L \times y \subset S\}| - \sum_{y \notin Y_1} |\{L \in \mathcal{L} : L \times y \subset S\}|
\]
\[
> 2\delta_1\ell^B|\mathcal{L}| - \delta_1|\mathcal{L}| \cdot \ell^B = \delta_1\ell^B|\mathcal{L}|.
\]

Since each term in this sum has size at most \(|\mathcal{L}|\), it follows that \(|Y_1| > \delta_1\ell^B\). \(\square\)
Claim. There is a set $Y_2 \subset [\ell]^\beta$ with $|Y_2| > (1 - 3\delta_2)\ell^\beta$ such that for all $y \in Y_2$ we have $|S \cap (V_2 \times y)| \geq (\delta - \delta_2)\ell^m$.

Proof. If for any $y \in [\ell]^\beta$ we have $|S \cap (V_2 \times y)| \geq (\delta + \delta_2^2)\ell^m$, then we have a density increment on an $m$-dimensional subspace and we are done since $m \geq d$, so we can assume that

$$|S \cap (V_2 \times y)| < (\delta + \delta_2^2)\ell^m$$

for all $y \in [\ell]^\beta$. Since $V_2 \subset V_1$, we also have $|S_x| \geq (\delta - \delta_2^2)\ell^m$ for all $x \in V_2$, which implies

$$\sum_{y \in [\ell]^\beta} |S \cap (V_2 \times y)| = \sum_{x \in V_2} |S_x| \geq (\delta - \delta_2^2)\ell^{m+\beta}$$

Set

$$Y_2 = \{y \in [\ell]^\beta : |S \cap (V_2 \times y)| \geq (\delta - \delta_2)\ell^m\}.$$

We have

$$|Y_2| \cdot (\delta + \delta_2^2)\ell^m > \sum_{y \in Y_2} |S \cap (V_2 \times y)| \geq \sum_{y \in [\ell]^\beta} |S \cap (V_2 \times y)| - \sum_{y \notin Y_2} |S \cap (V_2 \times y)|$$

$$\geq (\delta - \delta_2^2)\ell^{m+\beta} - (\delta - \delta_2)\ell^m(\ell^\beta - |Y_2|) = (\delta_2^2 - \delta_2)\ell^{m+\beta} + (\delta - \delta_2)|Y_2|\ell^m.$$

This implies

$$|Y_2| > \frac{\delta_2^2 - \delta_2}{\delta_2 + \delta_2^2}\ell^\beta > (1 - 3\delta_2)\ell^\beta.$$

Hence $Y_2$ is as claimed. \qed

Claim. There is an $m$-dimensional subspace $V \subset [\ell]^D$ such that $S$ contains a $\delta_1$-dense subset of all $(\ell - 1)$-lines in $\rho(V)$, and $S$ has density at least $\delta - \delta_2$ on $V$.

Proof. From the previous two claims we have $|Y_2| > (1 - 3\delta_2)\ell^\beta$ and $|Y_2| > \delta_1\ell^\beta$. By our choice of $\delta_2$ we have $\delta_1 = 10\delta_2/\delta > 3\delta_2$, so $Y_1 \cap Y_2$ is nonempty. We pick any $y \in Y_1 \cap Y_2$ and set $V = V_2 \times y$, so that by the previous two claims $V$ has the desired properties. \qed

### 3.4.2 Obtaining a density increment

By the last claim of Section 3.4.1 we have an $m$-dimensional subspace $V \subset [\ell]^D$ on which $S$ has density at least $\delta - \delta_2$, and such that $S$ contains a $\delta_1$-dense subset of all $(\ell - 1)$-lines in $\rho(V)$. We identify $V$ with $[\ell]^m$, and we let $T \subset [\ell]^m$ be the subset corresponding to $S \cap V$, so $T$ contains no $\ell$-lines, we have $|T| \geq (\delta - \delta_2)\ell^m$, and $T$ contains a $\delta_1$-dense subset of the $(\ell - 1)$-lines in $[\ell]^{\ell}$. Let us make the definition of insensitivity slightly more formal. For $x \in [\ell]^m$ and $i \in [\ell - 1]$, let $\sigma_i(x)$ be the point obtained by replacing every $\ell$ in $x$ by $i$. Then $X \subset [\ell]^m$ is $i\ell$-insensitive if we have

$$x \in X \text{ if and only if } \sigma_i(x) \in X.$$

To see that this corresponds to the definition in Section 3.2.2 observe that $x, y \in [\ell]^m$ can be obtained from each other by swapping some $is$ and $\ell$s if and only if $\sigma_i(x) = \sigma_i(y)$. Note that with this definition it is clear that if $X$ is $i\ell$-insensitive, then so is $X^c = [\ell]^m \setminus X$.

Finally, recall that $C \subset [\ell]^m$ is Cartesian if there are $i\ell$-insensitive sets $C_i$ for $i \in [\ell - 1]$ such that $C = C_1 \cap \cdots \cap C_{\ell - 1}$. Note that some of the $C_i$ may be equal to $[\ell]^m$, or in other words, they could be omitted from the intersection.
Claim. There is a Cartesian set $C \subseteq [\ell]^m$ with $|C| \geq \delta_2 \ell^m$ and $|T \cap C| \geq (\delta + \delta_2)|C|$.

Proof. Let $E$ be the set of all endpoints in $[\ell]^m$ of the $(\ell - 1)$-lines in $T \cap [\ell - 1]^m$, and set

$$P = (T \cap [\ell - 1]^m) \cup E.$$ 

Note that $E$ is in bijection with the set of $(\ell - 1)$-lines in $T \cap [\ell - 1]^m$, so we have

$$|E| \geq \delta_1(\ell^m - (\ell - 1)^m) \geq (\delta_1/2)\ell^m.$$ 

Also note that, since $T$ contains no $\ell$-line, $T$ is disjoint from $E$.

Define

$$P_i = \{x \in [\ell]^m : \sigma_i(x) \in T\}.$$ 

Then $P_i$ is $i\ell$-insensitive by the definition above, since by definition of $P_i$ we have $x \in P_i$ if and only if $\sigma_i(x) \in T$, which because of $\sigma_i(\sigma_i(x)) = \sigma_i(x)$ is equivalent to $\sigma_i(\sigma_i(x)) \in T$, which by definition of $P_i$ means $\sigma_i(x) \in P_i$. We also have

$$P = P_1 \cap \cdots \cap P_{\ell-1}.$$ 

Indeed, if $x \in T \cap [\ell - 1]^m$, then $\sigma_i(x) = x \in T$ for every $i$, so $x \in P_i$ for every $i$; if $x \in E$, then $x$ is an endpoint of a line in $T \cap [\ell - 1]^m$, and each $\sigma_i(x)$ lies on that line, so $x \in P_i$ for every $i$. Conversely, if $x \in P_1 \cap \cdots \cap P_{\ell-1}$, then either $x \in [\ell - 1]^m$, in which case $x = \sigma_i(x) \in T$, or $x \notin [\ell - 1]^m$, in which case $\sigma_i(x) \in T$ for every $i$, implying that $x$ is the endpoint of a line in $[\ell - 1]^m$. To summarize, $P$ is a Cartesian set.

We have $|P| \geq |E| \geq (\delta_1/2)\ell^m$, so (writing $P^c = [\ell]^m \setminus P$)

$$|P^c| \leq (1 - \delta_1/2)\ell^m.$$ 

On the other hand, we have $T \cap E = \emptyset$ and thus $T \cap P \subseteq [\ell - 1]^m$, so using the assumption $m \geq \log((\ell - 1)/\ell)(\ell_2)$ we get

$$|T \cap P| \leq ((\ell - 1)/\ell)^m \cdot \ell^m \leq \delta_2 \ell^m.$$ 

Since $|T| \geq (\delta - \delta_2)\ell^m$, we get

$$|T \cap P^c| \geq (\delta - 2\delta_2)\ell^m.$$ 

Define

$$Q_i = P_1 \cap \cdots \cap P_{i-1} \cap P_i^c$$

for $i \in [\ell]$. The sets $Q_1, \ldots, Q_\ell$ partition $P^c$. Since taking complements preserves insensitivity, $P_i^c$ is $i\ell$-insensitive, so $Q_i$ is a Cartesian set for each $i$.

Let $I \subseteq [\ell]$ be the set of indices $i$ for which $|Q_i| \geq \delta_3 \ell^m$. Then, since we chose $\delta_3 = \delta_2/\ell$,

$$\sum_{i \notin I} |Q_i| < \ell \cdot \delta_3 \ell^m = \delta_2 \ell^m.$$ 

Hence, using $\delta_2 = \delta \cdot \delta_1/10$, we have

$$\frac{|(T \cap P^c) \cap \bigcup_{i \in I} Q_i|}{|P^c|} \geq \frac{|T \cap P^c| - \bigcup_{i \in I} |Q_i|}{|P^c|} \geq \frac{(\delta - 2\delta_2) - \delta_2}{(1 - \delta_1/2)} \geq (\delta - 3\delta_2)(1 + \delta_1/2) \geq \delta + \delta_2.$$ 

Hence there exists $i_0 \in I$ such that $|T \cap Q_{i_0}| \geq (\delta + \delta_2)|Q_{i_0}|$.

We set $C = Q_{i_0}$, so $C$ is Cartesian, since we observed that every $Q_i$ is Cartesian. By the choice of $i_0$ we have $|C| \geq \delta_3 \ell^m$ and $|T \cap C| \geq (\delta + \delta_2)|C|$. \hfill \qed
3.4.3 Partitioning into subspaces

It remains to show how to partition a Cartesian set into subspaces of a certain dimension. First we do this for insensitive sets.

**Claim.** For all $\ell, n, \eta$ there is a number $N(\ell, n, \eta)$ such that for $N > N(\ell, n, \eta)$ the following holds. Given an $i\ell$-insensitive set $X \subseteq [\ell]^N$ with $|X| \geq \eta \ell^N$, there exists a family $\mathcal{V}$ of pairwise disjoint $n$-dimensional subspaces contained in $X$ such that $|X \setminus (\bigcup_{V \in \mathcal{V}} V)| < \eta \ell^N$.

**Proof.** Set $\beta = MDHJ^*(\ell, n, \eta/2)$ (where the function $MDHJ^*$ comes from Lemma 3.9). $N = \beta \cdot 2(\ell + n)^{\beta / \eta^2} \ell^{\beta^2 - n / \eta}$, and $\alpha = N - \beta$. We have

$$\sum_{x \in \ell^\alpha} |X_x| = |X| \geq \eta \ell^\alpha + \beta.$$  

Set

$$A_1 = \{ x \in \ell^\alpha : |X_x| \geq (\eta/2) \ell^\beta \},$$

so we have $|A_1| \geq (\eta/2) \ell^\alpha$.

For $x \in A_1$, by Lemma 3.9 and the choice of $\beta$, there is an $n$-dimensional subspace $V_x \subseteq [\ell]^\beta$ such that $\rho(V_x) \subset X$. Then $x \times \rho(V_x) \subset X$, so the fact that $X$ is $i\ell$-insensitive implies that $x \times V_x \subset X$. Since there are less than $(\ell + n)^\beta$ distinct $n$-dimensional subspaces in $[\ell]^\beta$, it follows that some $V_1$ occurs as $V_x$ at least $\eta/(2(\ell + n)^\beta) \ell^\alpha$ times, or in other words the set

$$A_2 = \{ x \in \ell^\alpha : x \times V_1 \subset X \}$$

satisfies $|A_2| \geq (\eta/(2(\ell + n)^\beta)) \ell^\alpha$. Note that $A_2$ is also $i\ell$-insensitive, because the fact that $X$ is $i\ell$-insensitive implies that $x \times V_1 \subset X$ if and only if $a(x) \times V_1 \subset X$.

Set

$$\mathcal{V}_1 = \{ x \times V_1 : x \in A_2 \}.$$

This is a family of pairwise disjoint $n$-dimensional subspaces contained in $X$, with

$$|\bigcup_{V \in \mathcal{V}_1} V| = |A_1| \cdot \ell^\alpha \geq \frac{\eta}{2(\ell + n)^\beta} \ell^\alpha \cdot \ell^\alpha = \frac{\eta}{2(\ell + n)^\beta} \ell^{\beta^2 - n} \ell^N.$$

If $|X \setminus (\bigcup_{V \in \mathcal{V}_1} V)| < \eta \ell^N$, we are done, so we can assume that

$$X_1 = X \setminus (\bigcup_{V \in \mathcal{V}_1} V),$$

satisfies $|X_1| \geq \eta \ell^N$. The set $X_1$ need not be $i\ell$-insensitive, but it has the property that for every $y \in [\ell]^\beta$ the set

$$X_1^y = \{ x \in [\ell]^{\alpha} : x \times y \in X_1 \}$$

is $i\ell$-insensitive. We can apply the argument above to $X_1^y \subseteq [\ell]^{N-\beta}$ for each $y \in [\ell]^\beta$, to obtain a family $\mathcal{V}_2$ of pairwise disjoint $n$-dimensional subspaces contained in $X_1$ with

$$|\bigcup_{V \in \mathcal{V}_1 \cup \mathcal{V}_2} V| \geq 2 \cdot \frac{\eta}{2(\ell + n)^\beta} \ell^{\beta^2 - n} \ell^N.$$

Note that all subspaces in $\mathcal{V}_1 \cup \mathcal{V}_2$ are pairwise disjoint.

By the choice of $N$, we could repeat this $2(\ell + n)^\beta \ell^{\beta^2 - n} / \eta$ times to get families $\mathcal{V}_3, \mathcal{V}_4, \ldots$, with the analogous properties. Since the density of the union of the subspaces cannot increase above 1, we must obtain $|X \setminus (\bigcup_{V \in \mathcal{V}_1} V)| < \eta \ell^N$. Then we set $\mathcal{V} = \bigcup_{V \in \mathcal{V}_1} V$.  

We can now define the function $F$ that we used to define $m$. Set $\theta = \delta_3/(\ell - 1)$. Define $N_r$ for $r \in [\ell - 1]$ by $N_{\ell - 1} = d$ and $N_r = N(\ell, N_{r+1}, \theta)$, where the function $N$ comes from the claim above. Then we set $F(\ell, \delta_3, d) = N_1$.

**Claim.** Given the Cartesian set $C \subset [\ell]^m$ with $|C| \geq \delta_3 \ell^m$, there exists a family $V$ of pairwise disjoint $d$-dimensional subspaces contained in $C$ such that $|C \setminus (\cup_{V \in V} V)| < \delta_3 \ell^m$.

**Proof.** Let $C_1, \ldots, C_{r-1}$ such that $C = C_1 \cap \ldots \cap C_{r-1}$, with $C_i$ an $i\ell$-insensitive set. We prove by induction on $r$ that for all $r \in [\ell - 1]$ there exists a family $V_r$ of pairwise disjoint $N_r$-dimensional subspaces contained in $C_1 \cap \ldots \cap C_r$ such that

$$|(C_1 \cap \ldots \cap C_r) \setminus (\cup_{V \in V} V)| < r \theta \ell^m.$$  \hfill (3.1)

For $r = 1$, this follows from the previous claim, with $\eta = \theta$. For $r = \ell - 1$ we obtain the current claim, since $(\ell - 1)\theta = \delta_3$.

We now prove the statement for $r + 1$. By induction, the statement for $C_1 \cap \ldots \cap C_r$ gives a family $V_r$ of pairwise disjoint $N_r$-dimensional subspaces contained in $C_1 \cap \ldots \cap C_r$ satisfying (3.1). Note that we have $|V_r| \leq \ell^{m-N_r}$.

Let

$$U = \{ V \in V_r : |V \cap C_{r+1}| \geq \theta |V| \}.$$ 

The set $V \cap C_{r+1}$ is $i(r + 1)$-insensitive as a subset of $V$. For every $V \in U$ we apply the previous claim with $\eta = \theta$ to $V \cap C_{r+1}$ as a subset of $V$, using the fact that $N_r = N(\ell, N_{r+1}, \theta)$. Since $V \cap C_{r+1} = V \cap (C_1 \cap \ldots \cap C_{r+1})$, we get a family $W_V$ of $N_{r+1}$-dimensional subspaces such that

$$|(V \cap (C_1 \cap \ldots \cap C_{r+1})) \setminus (\cup_{W \in W_V} W)| < \theta \ell^{N_r}.$$  \hfill (3.2)

Then we set

$$V_{r+1} = \{ W : W \in W_V, V \in U \}.$$ 

We prove that $V_{r+1}$ is as required. By (3.1), the number of elements of $C_1 \cap \ldots \cap C_{r+1}$ that are missed by all $V \in V_r$ is less than $r \theta \ell^m$. On the other hand, for every $V \in V_r$, either we have $V \notin U$ and then $|V \cap (C_1 \cap \ldots \cap C_{r+1})| < \theta \ell^{N_r}$, or $V \in U$ and then by (3.2) all but $\theta \ell^{N_r}$ elements of $V \cap (C_1 \cap \ldots \cap C_{r+1})$ are covered by $V_{r+1}$. Since $|V_r| \leq \ell^{m-N_r}$, it follows that the number of elements of $C_1 \cap \ldots \cap C_{r+1}$ missed in this way is at most $\ell^{m-N_r} \cdot \theta \ell^{N_r} = \theta \ell^m$. Altogether, the subspaces in $V_{r+1}$ miss at most $r \theta \ell^m + \theta \ell^m = (r + 1)\theta \ell^m$ elements of $C_1 \cap \ldots \cap C_{r+1}$, as required. \hfill \square

### 3.4.4 Completing the proof

We can now wrap up the proof. From Sections 3.4.1 and 3.4.2 we obtained a Cartesian set $C \subset [\ell]^m$ with $|C| \geq \delta_3 \ell^m$ and $|T \cap C| \geq (\delta + \delta_2)|C|$, unless we have a density increment $\delta_3^2$ on a $d$-dimensional subspace (in the proof of the fourth claim in Section 3.4.1). By the second claim in Section 3.4.3, $C$ can partitioned into disjoint $d$-dimensional subspaces, with a remainder of size at most $\delta_3 \ell^m$. Since $\delta_3 < \delta_2/2$, $T$ has density at least $\delta + \delta_2/2$ on the union of these subspaces, so there must be a $d$-dimensional subspace $W'$ for which $|T \cap W'| \geq (\delta + \delta_2/2)|W'|$. In the original setting, $V$ corresponds to a subspace $W \subset [\ell]^D$ such that $|S \cap W| \geq (\delta + \delta_2/2)|W|$. Thus we have a density increment $\delta_2/2$ (or $\delta_3^2$) on a $d$-dimensional subspace, and if $d$ is sufficiently large we can repeat this argument. Since the density cannot keep increasing, we must encounter a Hales–Jewett lines contained in $S$. This finishes the proof.
3.5 Consequences

**Corollary 3.10** (Szemerédi’s Theorem). For all $\delta, \ell$ there is a number $N_0$ such that for all $N > N_0$ the following holds. If $S \subset [N]$ and $|S| \geq \delta N$, then $S$ contains an arithmetic progression of length $\ell$.

**Proof.** We set $D = DHJ(\ell, \delta/2)$ and $N_0 = \ell^D$. Given $N > \ell^D$ and $S \subset [N]$ with $|S| \geq \delta N$, we can find an interval of length $\ell^D$ on which has $S$ has density at least $\delta/2$, for instance as follows. Either the first half or the second half of $[N]$ contains $(\delta/2)N$ elements of $S$, and that half can be covered by disjoint intervals of length $\ell^D$. Then $S$ has density at least $\delta/2$ on the union of these disjoint intervals, which implies that $S$ has density at least $\delta/2$ on one of the intervals. After translating and renaming, we can assume that $S \subset [\ell^D]$ and $S$ has density at least $\delta/2$.

We define a map $\varphi : [\ell]^D \to [\ell^D]$ by

$$\varphi(x_1, \ldots, x_D) = 1 + \sum_{i \in [D]} (x_i - 1)\ell^{i-1},$$

which is a bijection. Therefore $\varphi^{-1}(S)$ is a subset of $[\ell]^D$ with density at least $\delta/2$, so by Theorem 3.1 and the choice of $D$, $\varphi^{-1}(S)$ contains a Hales–Jewett line $L$ of length $\ell$. Then $\varphi(L)$ is an arithmetic progression of length $\ell$ contained in $S$. ∎
Chapter 4

Arithmetic progressions over finite fields

4.1 Introduction

It is natural to look for variants of Roth’s Theorem in other groups. For a group $G$, write $r_3(G)$ for the maximum size of a subset of $G$ that does not contain a three-term arithmetic progression (a solution of $x - 2y + z = 0$). This is slightly different from Roth’s Theorem, since there we consider subsets of a specific subset $[N]$ of the group $\mathbb{Z}$, but in fact Roth’s Theorem is closely related to the group $\mathbb{Z}/N\mathbb{Z}$. Roth’s proof $[27]$ actually showed that $r_3(\mathbb{Z}/N\mathbb{Z}) = o(N)$, and this implies Roth’s Theorem in the way that we stated it (given a dense subset of $[N]$, we can embed it into $\mathbb{Z}/2N\mathbb{Z}$, so that the modulo arithmetic plays no role).

Another finite group for which this question is interesting is the vector space $\mathbb{F}_p^D$ (where $p$ is a prime, so that $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is a field). In particular, bounding $r_3(\mathbb{F}_p^D)$ has become known as the capset problem. Note that in $\mathbb{F}_3^D$, three points form a three-term arithmetic progression if and only if they form a line. Thus, a set without a three-term arithmetic progression is a set that does not contain a line (such a set is sometimes called a “cap” or a “capset”, hence the name for the problem). Also note that we can compare $\mathbb{F}_3^D$ to $[3]^D$; then every Hales–Jewett line corresponds to a line over $\mathbb{F}_3$, but there are far more $\mathbb{F}_3$-lines than Hales–Jewett lines. This suggests that a density theorem for $\mathbb{F}_3$ lines, i.e. for $r_3(\mathbb{F}_3^D)$, could be quantitatively stronger than the Density Hales-Jewett Theorem.

Brown and Buhler $[7]$ proved that $r_3(\mathbb{F}_p^D) = o(3^D)$, Meshulam $[23]$ improved that to $r_3(\mathbb{F}_p^D) = O(3^D/D)$ (which is $O(N/\log N)$ if we write $N = 3^D$), and Bateman and Katz $[3]$ obtained $r_3(\mathbb{F}_3^D) = O(3^D/D^{1+\epsilon})$. Meshulam’s proof uses Fourier analysis in a similar way to Roth’s original proof of Roth’s Theorem $[27]$. From below, Edel $[10]$ proved $r_3(\mathbb{F}_3^D) \geq (2/324)^D$.

Recently, Ellenberg and Gijswijt $[12]$, based on work of Croot, Lev, and Pach $[8]$, gave a simpler proof of a much better bound. They proved that

$$r_3(\mathbb{F}_3^D) = O((3 - \alpha)^D)$$

for some $\alpha > 0$, specifically $\alpha \geq 0.24$. Their proof uses basic properties of polynomials and some linear algebra, and can be extended to other finite groups. Tao $[34]$ elegantly reformulated the proof of $[12]$ in terms of the slice rank of a tensor.

In this chapter we present Tao’s proof. First we will introduce the slice rank and apply it to a problem from extremal set theory, where some of the details are easier, and then we will use it to prove the result of Ellenberg and Gijswijt.
4.2 The slice rank method

As a warmup for the slice rank method, we present a classic example of a proof using the linear algebra method. This proof uses the rank of a matrix, and we present it in a (somewhat unusual) way that will make the generalization to tensors more natural. The theorem is due to Berlekamp [4].

**Theorem 4.1** (Oddtown theorem). Let $X$ be a finite set and let $S \subset 2^X$ be a set system. If $|S|$ is odd for every $S \in S$, and $|S \cap T|$ is even for all distinct $S, T \in S$, then $|S| \leq |X|$.

**Proof.** Let $F : S \times S \to \mathbb{F}_2$ be the function defined for $S, T \in S$ by

$$F(S, T) = |S \cap T| \mod 2.$$ 

By assumption, for $S, T \in S$ we have $F(S, T) = 1$ if $S = T$ and $F(S, T) = 0$ if $S \neq T$.

We can think of $F$ as a matrix in which an entry is nonzero if and only if it is on the diagonal. Thus the rank of this matrix is even. On the other hand, we can write

$$F(S, T) = \sum_{x \in X} f_x(S)f_x(T),$$

where $f_x : S \to \mathbb{F}_2$ is the function that is 1 if $x \in S$ and 0 if $x \notin S$. We can think of $f_x$ as a vector of length $|S|$, so we can write the matrix of $F$ as a sum of $|X|$ matrices, each of which is a product of the form $v v^T$ for a vector $v$. Since a matrix of the form $v v^T$ has rank one, and since matrix rank is subadditive, this implies that the rank of the matrix of $F$, which we found to be $|S|$, is at most $|X|$. $\square$

In general, a $k$-tensor is a function $X^k \to \mathbb{F}$. Traditionally, the tensor rank of a tensor $T$ is the minimum $R$ for which we can write $T = \sum_{i=1}^R f_i$, with each $f_i$ of the form $g_1(x_1) \cdots g_k(x_k)$. In other words, tensors of the form $g_1(x_1) \cdots g_k(x_k)$ have rank 1, and the rank of a tensor is the minimum number of rank one tensors that we can decompose it in. Then a 2-tensor is a matrix, and its tensor rank equals its matrix rank (since the rank of a matrix equals the minimum number of rank one matrices, which have the form $uv^T$, that the matrix can be decomposed into).

Here we will use a different notion of rank for tensors, called the slice rank. A tensor has slice rank one if it has the form $gh$, where $g$ and $h$ depend on disjoint nonempty subsets of the variables. The slice rank of a general tensor is the minimum number of slice rank one tensors that we can decompose it in. We will only use the slice rank for $k = 3$, so for concreteness we define it formally only in that case, but the generalization would be straightforward.

**Definition 4.2.** Let $\mathbb{F}$ be a field and let $X$ be any finite set. A function $F : X \times X \times X \to \mathbb{F}$ has slice rank 1 if we can write

$$F(x, y, z) = g(x)h(y, z), \text{ or } F(x, y, z) = g(y)h(x, z), \text{ or } F(x, y, z) = g(z)h(x, y).$$

For a function $F : X \times X \times X \to \mathbb{F}$, the slice rank $SR(F)$ is the minimum number $R$ such that we can write $F = \sum_{i=1}^R F_i$, where each $F_i$ has slice rank 1.

In the proofs that follow, we will use the slice rank in a similar way to how we used the matrix rank in the proof of Theorem 4.1. Given some set $S$ whose size we want to bound, we construct a diagonal tensor that has $|S|$ nonzero entries on its diagonal. Then the slice rank of the tensor is $|S|$, and an upper bound on the slice rank gives an upper bound on $|S|$. However, first we have to carefully prove the natural fact that the slice rank of a diagonal tensor equals the number of nonzero entries on its diagonal.
Lemma 4.3. If $F : X \times X \times X \to \mathbb{F}$ has the property that $F(x, y, z) \neq 0$ if and only if $x = y = z$, then $\mathcal{SR}(F) = |X|$.

Proof. Define $\delta(a, b)$ to be 1 if $a = b$ and 0 if $a \neq b$. Then we can write

$$F(x, y, z) = \sum_{x_0 \in X} \delta(x, x_0) F(x_0, y, z),$$

which shows that $\mathcal{SR}(F) \leq |X|$.

Suppose that $\mathcal{SR}(F) = \ell < |X|$. Then we can write

$$F(x, y, z) = \sum_{i=1}^{j} f_i(x)g_i(y, z) + \sum_{i=j+1}^{k} f_i(y)g_i(x, z) + \sum_{i=k+1}^{\ell} f_i(z)g_i(x, y).$$

Let $V$ be the vector space of all functions $v : X \to \mathbb{F}$ such that for all $1 \leq i \leq j$ we have

$$\sum_{x \in X} f_i(x)v(x) = 0.$$

Since $V$ is the solution set of $j$ homogeneous equations within the $|X|$-dimensional space of all functions on $X$, it follows that $V$ has dimension at least $|X| - j$.

Let $v \in V$ be a function with maximal support $S_v = \{x \in X : v(x) \neq 0\}$, so that

$$|S_v| \geq |X| - j.$$

Indeed, if we had $|S_v| < |X| - j = \dim(V)$, then there would be a nonzero $w \in V$ that vanishes on $S_v$ (since the number of equations $v(x) = 0$ for $x \in S_v$ is less than the dimension of $V$), so that $v + w \in V$ would have larger support than $v$.

We have

$$\sum_{x \in X} v(x)F(x, y, z) = \sum_{i=1}^{j} g_i(y, z) \left( \sum_{x \in X} f_i(x)v(x) \right) + \sum_{i=j+1}^{k} f_i(y) \left( \sum_{x \in X} v(x)g_i(x, z) \right) + \sum_{i=k+1}^{\ell} f_i(z) \left( \sum_{x \in X} v(x)g_i(x, y) \right)$$

$$= \sum_{i=j+1}^{k} f_i(y)g_i(z) + \sum_{i=k+1}^{\ell} f_i(z)h_i(y)$$

for certain functions $h_i$. This shows that the 2-tensor

$$G(y, z) = \sum_{x \in X} v(x)F(x, y, z),$$

which corresponds to a matrix, has matrix rank at most $\ell - j < |X| - j$. Since $F(x, y, z) \neq 0$ if and only if $x = y = z$, we have $G(y, z) = 0$ whenever $y \neq z$. Moreover, we have $G(y, y) = v(y)F(y, y, y)$, which is nonzero whenever $y \in S_v$. Thus $G$ corresponds to an $|X| \times |X|$ diagonal matrix with $|S_v|$ nonzero entries on its diagonal, so it has matrix rank $|S_v|$. Since $|S_v| \geq |X| - j$, this is a contradiction. $\square$

31
4.3 Sunflowers

A $k$-sunflower is a family of $k$ sets such that any two of the sets have the same intersection. In other words, any element is either contained in all the sets of the sunflower, or in at most one of the sets. Erdős and Szemerédi [13] conjectured that if a set system $\mathcal{S} \subset 2^{[N]}$ contains no 3-sunflower, then $|\mathcal{S}| < (2 - \beta)^N$ for some $\beta > 0$. Prior to the work of Ellenberg and Gijswijt [12], Alon, Shpilka, and Umans [2] observed that a bound for the capset problem like that in [12] would imply the conjecture of Erdős and Szemerédi. The result of Ellenberg and Gijswijt [12] thus proved this conjecture, and the resulting bound comes to $\beta \geq 0.06$.

Naslund and Sawin [24] then observed that the slice rank method of Tao [34] can be applied directly to this problem, and gives a better bound. We give the proof of Naslund and Sawin, which is also a natural introduction to Tao’s proof of the Ellenberg–Gijswijt bound.

**Theorem 4.4 (Naslund–Sawin).** If $\mathcal{S} \subset 2^{[N]}$ is a set system without a 3-sunflower, then $|\mathcal{S}| \leq 3(N + 1)C^N$, with $C = 3/2^{2/3} \leq 1.89$.

**Proof.** We can identify $2^{[N]}$ with $\{0, 1\}^N$, so that we can view $\mathcal{S}$ as either a set system or a set of binary vectors. Then $\mathcal{S}$ has the property that for any three distinct vectors $x, y, z \in \mathcal{S}$, the sum $x + y + z$ has a coordinate equal to 2, since otherwise the sets corresponding to $x, y, z$ would form a 3-sunflower. Let $\mathcal{S}_m$ be the subset of sets $x \in \mathcal{S}$ that have exactly $m$ elements. Each $\mathcal{S}_m$ has the property that for any $x, y, z \in \mathcal{S}_m$ that are not all equal, the sum $x + y + z$ has a coordinate equal to 2.

Define a function $F : \mathcal{S}_m^3 \to \mathbb{Q}$ by

$$F(x, y, z) = \prod_{i=1}^N (2 - (x_i + y_i + z_i)).$$

Because of the property of $\mathcal{S}_m$ just mentioned, for $x, y, z \in \mathcal{S}_m$ we have $F(x, y, z) = 0$ if and only if $x = y = z$. By Lemma 4.3 this implies $\text{SR}(F) = |\mathcal{S}_m|$.

On the other hand, expanding $F$ gives (using the notation $x^I = x_1^{i_1} \cdots x_N^{i_N}$ and $|I| = \sum_{i \in I} i$)

$$F(x, y, z) = \sum_{I, J, K \subset \{0, 1\}^N, |I| + |J| + |K| \leq 1} c_{IJK} x^I y^J z^K = \sum_{I \subset \{0, 1\}^N} x^I f_I(x, y, z) + \sum_{J \subset \{0, 1\}^N} y^J g_J(x, y, z) + \sum_{K \subset \{0, 1\}^N} z^K h_K(x, y)$$

for some functions $f_I, g_J, h_K$. Here we used the pigeonhole principle to observe that if $I + J + K \leq \{1, \ldots, 1\}$, then we have either $|I| \leq N/3$, or $|J| \leq N/3$, or $|K| \leq N/3$. Since each term in this sum is a slice, it follows that the slice rank of $F$ satisfies

$$|\mathcal{S}_m| = \text{SR}(F) \leq 3 \sum_{k \leq N/3} \binom{N}{k}.$$

It remains to compute a numerical upper bound. By the binomial theorem we have $(1 + x)^N = \sum_{k \leq N} \binom{N}{k} x^k$, so for any $0 < x < 1$ we have

$$x^{-N/3}(1 + x)^N = \sum_{k \leq N} \binom{N}{k} x^{k-N/3} > \sum_{k \leq N/3} \binom{N}{k} x^{k-N/3} > \sum_{k \leq N/3} \binom{N}{k}.$$

We can choose $x \in (0, 1)$ to minimize $f(x) = x^{-1/3}(1 + x)$ using basic calculus. We have $f'(x) = \frac{1}{3} x^{-4/3}(2x - 1)$, so $f$ has a minimum at $x = 1/2$, with the value $f(1/2) = 3/2^{2/3}$. Hence

$$|\mathcal{S}| = \sum_{m=0}^N |\mathcal{S}_m| \leq 3(N + 1) \sum_{k \leq N/3} \binom{N}{k} \leq 3(N + 1)(3/2^{2/3})^N,$$

which completes the proof. □

32
4.4 Arithmetic progressions in $\mathbb{F}_3^D$

We now give Tao’s proof [34] of the capset bound of Ellenberg and Gijswijt [12].

**Theorem 4.5 (Ellenberg–Gijswijt).** If $S \subset \mathbb{F}_3^D$ contains no arithmetic progression of length three, then $|S| \leq 3C^D$ for some $C \leq 2.76$.

**Proof.** An arithmetic progression in $\mathbb{F}_3^D$ corresponds to distinct $x, y, z$ such that $x - 2y + z = 0$, which over $\mathbb{F}_3$ is equivalent to $x + y + z = 0$, so $S$ contains no such solution with $x, y, z$ distinct. Moreover, if $x, y, z$ are not distinct but not all equal, say $x = y \neq z$, then $x + y + z = 0$ is not possible (since then $2x + z = 0$, so $z = 3x + z = x$). Thus, for $x, y, z \in S$, we have $x + y + z = 0$ if and only if $x = y = z$.

We define $F : S \times S \times S \to \mathbb{F}_3$ by

$$F(x, y, z) = \prod_{i=1}^{D}(1 - (x_i + y_i + z_i)^2).$$

Recall that in $\mathbb{F}_3$ we have $1 - x^2 \neq 0$ if and only if $x = 0$. Thus $F(x, y, z) \neq 0$ if and only if $x + y + z = 0$, which for $x, y, z \in S$ is equivalent to $x = y = z$. Hence $F$ is a diagonal tensor, so by Lemma 4.3 we have $SR(F) = |S|$.

On the other hand, $F(x, y, z)$ expands to

$$\sum_{IJK \in \{0,1,2\}^D} c_{ijk} x^iy^jz^K = \sum_{I \in \{0,1,2\}^D} x^I f_I(x, z) + \sum_{J \in \{0,1,2\}^D} y^J g_J(x, z) + \sum_{K \in \{0,1,2\}^D} z^K h_K(x, y)$$

for some functions $f_I, g_J, h_K$. This implies that

$$SR(F) \leq 3 \sum_{a+b+c=D \atop b+2c \leq 2D/3} \frac{D!}{a!b!c!}.$$

By the multinomial theorem we have

$$(1 + x + x^2)^D = \sum_{a+b+c=D} \frac{D!}{a!b!c!} x^{b+2c}.$$

Therefore, for $0 < x < 1$ we have

$$x^{-2D/3}(1 + x + x^2)^D = \sum_{a+b+c=D \atop b+2c \leq 2D/3} \frac{D!}{a!b!c!} x^{b+2c-2D/3} \geq \sum_{a+b+c=D \atop b+2c \leq 2D/3} \frac{D!}{a!b!c!} x^{b+2c-2D/3} > \sum_{a+b+c=D \atop b+2c \leq 2D/3} \frac{D!}{a!b!c!} x^{b+2c-2D/3}.$$

To minimize $f(x) = x^{-2/3}(1 + x + x^2)$, we calculate that $f'(x) = \frac{1}{3}x^{-5/3}(4x^2 + x - 2)$, so $f$ has a minimum at $x_0 = (\sqrt{33} - 1)/8$ with the value $f(x_0) \leq 2.76$. This proves the theorem. $\square$
Bibliography


[22] Imre Leader, Ramsey Theory, notes from a course at Cambridge University, 2000. Available online at https://www.dpmms.cam.ac.uk/par31/notes/


[29] Issai Schur, Über die Kongruenz $x^m + y^m \equiv z^m (mod p)$, Jahresbericht der Deutschen Mathematiker-Vereinigung 25, 114–117, 1916.


