The Hales Jewett Theorem

James Bang

November 23, 2022

Abstract

This mathematical essay surveys important results in Ramsey Theory, including the Hales–Jewett Theorem and the Graham–Rothschild Theorem. We present elementary proofs of these theorems, and their numerous corollaries in the field.

0 Introduction to Ramsey Theory

When mathematicians consider "random" objects, we usually visualise an example of an object which has no immediately recognisable structure. However, there are interesting examples of properties in combinatorial structures where it is possible to find a given property in a large enough object of a given kind. This is loosely the study of Ramsey Theory, a branch of combinatorics roughly described as "finding order within randomness". Often we deal with statements of the form:

Let X denote a combinatorial structure (e.g. a graph, or a subset of integers) and a property P defined on this structure. Do sufficiently large objects X contain a substructure $Y \subseteq X$ (e.g. subgraph, or subset) with property P? If so, how large does X have to be?

At the heart of Ramsey Theory are two main results: *Ramsey's Theorem*, and *Van der Waerden's Theorem*. However, while these results demonstrate examples of this "Ramsey Property" P, it is manifested in completely different ways;

- Van der Waerden's theorem requires finding arithmetic progressions on a sufficiently large set of randomly coloured consecutive integers,
- Ramsey's theorem involves finding a complete monochromatic subgraph given a sufficiently large graph whose edges are randomly coloured.

This prompts the question: *Could it be that both theorems follow from a generalised, purely combinatorial result?* The answer is yes – both results follow from the *Graham–Rothschild Theorem*, an entirely combinatorial, yet very deep theorem in the study of Ramsey Theory with far reaching consequences in the study of combinatorics of words.

This essay is split into several parts. In the first two sections, we state and prove Ramsey's Theorem and Van der Waerden's Theorem, to lay the foundations of Ramsey Theory.

We then present the *Hales–Jewett Theorem*, a purely combinatorial generalisation of Van der Waerden's theorem to colourings of a n–dimensional hypercube. This sets the groundwork to introduce the Graham–Rothschild theorem, a further generalisation which also proves Ramsey's Theorem as an immediate corollary.

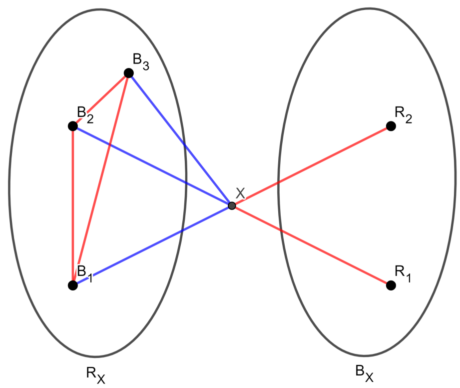
Finally, we state further directions of research into density variants and other generalisations of the theorems, which we will not prove.

1 Ramsey's Theorem

As previously stated, Ramsey's Theorem concerns finding a complete monochromatic subgraph of a given size in a sufficiently large, arbitrarily coloured graph.

Theorem 1.1 [Ramsey's Theorem]. Let $n, m \geq 1$ be integers. There exists $R = R(n, m)$ depending only on m, n, such that any colouring of the complete graph K_R, in red and blue, will contain either a red K_n or a blue K_m as a subgraph.

Proof. We prove that for $n, m \geq 2$, we have $R(n, m) \leq R(n-1, m) + R(n, m-1)$. This, combined with the trivial assertion $R(1, n) = R(n, 1) = 1$, gives a proof that $R(n, m)$ is finite for any $n, m \geq 1$.



To prove the assertion, take a complete graph K_R and a point $x \in K_n$. Separate the remaining vertices into two sets R_x and B_x, such that all edges (x, y) for $y \in R_x$ are red, and similar for B_x. Assuming I choose $R \overset{\text{def}}{=} R(n-1, m) + R(n, m-1)$, we have

$$|R_x| + |B_x| = R - 1 = R(n-1, m) + R(n, m-1) - 1$$

and so we must have either $|R_x| \geq R(n-1, m)$ or $|B_x| \geq R(n, m-1)$.

In the first case, we can find either a blue $K_m \subseteq R_x$, or a red $K_{n-1} \subseteq R_x$ which, combined with $x \notin R_x$, produces a red $K_n \subseteq K_R$ by construction. Hence, we have found either a red K_n or a blue K_m. A similar argument works when we instead have $|B_x| \geq R(n, m-1)$.

Hence, for $R \overset{\text{def}}{=} R(n-1, m) + R(n, m-1)$, any colouring of the complete graph K_R into two colours contains either a red K_n or a blue K_m, and hence $R(n, m) \leq R(n-1, m) + R(n, m-1)$. $\square$

Remark. The bound $R(n, m) \leq R(n-1, m) + R(n, m-1)$ can be used to inductively prove that

$$R(n, m) \leq \binom{n+m-2}{m-1} < 2^{n+m}.$$

Hence, the Ramsey numbers are bounded above by an exponential in n, m. Despite this relatively small upper bound, computing $R(n, m)$ exactly is famously known to currently infeasible, even for small values such as $(n, m) = (5, 5)$.

We can generalise the theorem slightly, to involve multiple colours.

Theorem 1.2 [Ramsey's Theorem, c colours]. Let $c \geq 2$ be a natural number, and $n_1, \ldots, n_c \geq 1$ be integers. There exists $R = R(n_1, \ldots, n_c)$ depending only on the n_i, such that any colouring of the complete graph K_R in one of c colours $1, \ldots, c$ contains a complete K_{n_i} of colour i (for some $i \leq c$).

Proof. The case $c = 2$ is exactly Theorem 1.1.

For $c \geq 3$, consider a complete graph K_R, coloured in the first c colours $1, \ldots, c$. We can view this as a colouring of K_R with $c-1$ colours by not distinguishing the last two colours c and $c-1$.

Suppose we take $R \stackrel{\text{def}}{=} R(n_1, \ldots, n_{c-2}, R(n_{c-1}, n_c))$. By definition, one of the two could happen:

- There exists some $i \leq n-2$, such that there is some monochromatic $K_{n_i} \subseteq K_R$ of colour i; or
- There exists a $K_{R(n_{c-1},n_c)} \subseteq K_R$ whose edges are of colour $c-1$ or c. But then, by the $c = 2$ case, we then obtain either a monochromatic $K_{n_{c-1}}$ of colour $c-1$, or a monochromatic K_{n_c}, of colour c.

In any case, we can find $i \leq c$ such that there exists a monochromatic $K_{n_i} \subseteq K_R$ of colour i. Hence, we have the inequality

$$R(n_1, \ldots, n_c) \leq R(n_1, \ldots, n_{c-2}, R(n_{c-1}, n_c))$$

and by induction we obtain $R(n_1, \ldots, n_c)$ finite for any $c \geq 1$ and $n_1, \ldots, n_c \in \mathbb{N}$. □

Remark. As an alternative proof to Theorem 1.2, it turns out we can apply a similar argument to the one given in Theorem 1.1. This gives the following upper bound:

$$R(n_1, \ldots, n_c) \leq (2-c) + \sum_{i \leq c} R(n_1, \ldots, n_{k-1}, n_k - 1, n_{k+1}, \ldots, n_c) \implies R(n_1, \ldots, n_c) \leq \binom{n_1 + \cdots + n_c}{n_1, \ \ldots, \ n_c}$$

which is much better than the bound from the previous proof.

Remark. Note that in the case $n_i = n$, we have the following equivalent statement. For integers $n \geq m$, denote

$$\tbinom{n}{m} \stackrel{\text{def}}{=} \{f : \{1, \ldots, n\} \to \{1, \ldots, m\} \mid f \text{ strictly increasing}\}$$

and notice that for functions $f \in \binom{n}{m}$ and $g \in \binom{m}{k}$, there is a natural composition $f \circ g$.

Then, Ramsey's Theorem is equivalent to the assertion that there exists $R > n > 0$ such that for any colouring $\Delta : \binom{R}{2} \to \{1, \ldots, c\}$, it is possible to find $f \in \binom{R}{n}$ such that $\Delta \circ f$ is constant on $\binom{n}{2}$.

This can easily be seen, since $f \circ \binom{n}{2} \stackrel{\text{def}}{=} \{f \circ g \, : \, g \in \binom{n}{2}\}$ consists of all strictly increasing functions $\{1, \ldots, n\} \to \{1, 2\}$, so the image of $f \circ g$ is a pair of distinct points. It follows the assertion Δ is constant in $f \circ \binom{n}{2}$ can be interpreted as all "pairs of points" in the image of f being coloured the same by Δ, and hence the image of f produces a complete monochromatic graph of size n.

While we will not immediately use this formulation, we will refer to this later when deducing Ramsey's theorem as a corollary of the Graham–Rothschild Theorem.

Having proved the theorem, we can produce an interesting corollary.

Corollary 1.3 [Schur's Theorem for APs]. For any integer $c \geq 1$, there exists $N = N(c)$, such that for any colouring of the integers $1, 2, \ldots, N$ in c colours, we can find $1 \leq x, y, z \leq N$ of the same colour, such that $x + y = z$.

Proof. Let $\chi : \{1, \ldots, N\} \to \{1, \ldots, c\}$ be the given colouring. Consider the complete graph K_N on N vertices labelled $1, 2, \ldots, N$, where each edge (i, j) is coloured $\chi(|i - j|)$. Then, assuming

$$N > R(\underbrace{3, 3, \ldots, 3}_{r}),$$

we can find a triple $\{i, j, k\}$ where $0 \leq i < j < k \leq N$ and $\chi(|i-j|) = \chi(|j-k|) = \chi(|k-i|)$.

Then, the numbers $x \stackrel{\text{def}}{=} j - i$, $y \stackrel{\text{def}}{=} k - j$ and $z \stackrel{\text{def}}{=} k - i$ are coloured the same and $x + y = z$. □

Now that we've dealt with monochromatic sumsets, we discuss monochromatic arithmetic progressions.

2 Van der Waerden's Theorem

Van der Waerden's theorem allows us to tackle a similar problem, on finding arithmetic progressions. Again, this demonstrates the recurring "Ramsey Property" discussed earlier, on a subset of integers.

Theorem 2 [Van der Waerden's Theorem]. Let $n, c \geq 1$ be integers. There exists $W = W(n, c)$ such that any colouring of the numbers $\{1, 2, \ldots, W\}$ with c colours, contains an arithmetic progression of length n, all whose numbers are the same colour.

Of course, to prove this result we want to use induction on n, as we would like to create an arithmetic progression of length n by "adding one more" to the monochromatic sequence. This motivates the following definition:

Definition. Let $A_1, \ldots, A_N$ be distinct arithmetic progressions of length m, i.e. there exists integers a_i, d_i such that $A_i \stackrel{\text{def}}{=} \{a_i, a_i + d_i, \ldots, a_i + (m-1) \cdot d_i\}$. We call $A_1, \ldots, A_N$ *focused*, if there exists an integer a such that $a_i + m \cdot d_i = a$ for all $1 \leq i \leq N$. We call a the *focus* of the A_i, and say that A_i *focus on* a.

Intuitively, this means the "next" term in the arithmetic sequences are the same common point.

Definition. We call $A_1, \ldots, A_N$ *colour–focused*, if they are focused, each A_i is monochromatic, and no two of the arithmetic sequences have the same colour.

Example. The progressions $(1, 6, 11)$, $(4, 8, 12)$ and $(13, 14, 15)$ are arithmetic progressions, which focus on the point 16.

Also, any monochromatic arithmetic progression A of length m forms a set of colour–focused arithmetic progressions $\{A\}$ of size 1.

Of course, if we colour the integers in c distinct colours, then it is impossible to have a set $\{A_i\}$ of colour–focused arithmetic progressions with size greater than c. When we have such a set of size exactly c, clearly we will be able to extend one of the A_i to length $m+1$ (by considering the colour of the focus point). The goal then would be to construct such a set of progressions.

Figure 1: Here we have three colour–focused arithmetic progressions.

However, what isn't obvious is how to use the inductive hypothesis to construct such a bundle. The key idea to the proof is to instead use the inductive hypothesis on consecutive "blocks" of integers.

Proof. As before, we apply induction on n. Clearly, $W(1, c) = 1$ because the single element in a length–1 arithmetic sequence forms an "arithmetic sequence" of length 1.

So, assume the statement holds for $W(n, c)$, for some $n \geq 1$ and *all* $c \in \mathbb{N}$. We first formalise our discussion above to get our colour–focused bundle.

Lemma 2.2. For each $r \leq c$, there exists N, such that any colouring of $\{1, \ldots, N\}$ with c colours contains either:

- a monochromatic arithmetic progression of length $n+1$; or
- a set of colour–focused arithmetic progressions $\{A_i\}$, each of length n, of size r.

Let's see how this lemma will prove the result. Taking $r = c$, we can find a set of colour–focused arithmetic progressions $A_1, \ldots, A_c$. By considering the colour of the focus point a of these progressions, we can extend exactly one of the A_i into a monochromatic progression of length $n+1$, giving $W(n, c) \leq N_c$.

Proof of lemma. Again, we use induction on r. Let N_r denote the N required for each $r \leq c$.

For $r = 1$, we can use $N_1 \stackrel{\text{def}}{=} W(n, c)$, since we get our one monochromatic arithmetic sequence, and hence a colour–focused bundle size 1.

For $r \geq 1$, let $W \stackrel{\text{def}}{=} W(n, c^{2 \cdot N_r})$. We claim that $N_{r+1} \stackrel{\text{def}}{=} 2N_r \cdot W$ will suffice for the inductive step. We will assume that there is no monochromatic arithmetic progression of length $n+1$ (as otherwise, we would already satisfy the first property of the lemma).

The key idea is to break the sequence $\{1, \ldots, N_{r+1}\}$ into a collection of intervals $\{I_1, \ldots, I_W\}$ where each $I_k \stackrel{\text{def}}{=} (k \cdot 2N_r + 1, \ldots, (k+1) \cdot 2N_r)$ is a tuple of size $2 \cdot N_r$. Clearly there are $c^{2 \cdot N_r}$ possible tuples, and so we consider this as a colouring of the "integers" $\{1, \ldots, W\}$ with $c^{2 \cdot N_r}$ colours.

By the definition of $W \stackrel{\text{def}}{=} W(n, c^{2 \cdot N_r})$, we can find integers $s, t \geq 1$ such that $I_s, I_{s+t}, \ldots, I_{s+(n-1) \cdot t}$ are identically coloured *as tuples*.

Since the interval I_s has $2 \cdot N_r$ number of elements, by the definition of N_r, the first half of I_s (which has N_r numbers) contains all the elements of a colour–focused arithmetic progression $\{A_1, \ldots, A_r\}$; then, the focus point also lies inside I_s due to size arguments.

So, let's focus on the grid G of size $r \times (2 \cdot N_r)$, whose k'th row is $I_{s+k \cdot t}$.

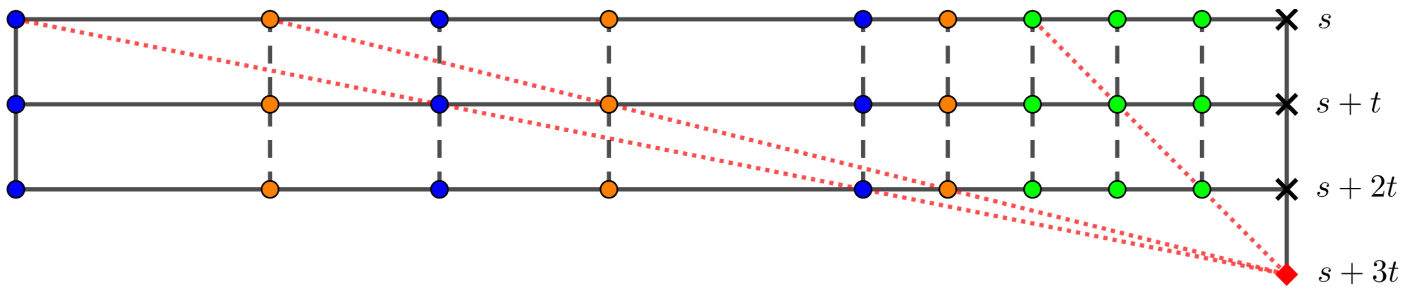


Figure 2: The grid G for $r = 3$. The "diagonal" progressions are in red. The crosses are focus points.

Since each row of G is identically coloured by construction, it follows that there are copies of $\{A_1, \ldots, A_r\}$ in the same spot, in every row. Hence, if we have $A_i \stackrel{\text{def}}{=} \{a_i, a_i + d_i, \ldots, a_i + (n-1)d_i\}$ focusing on a point $a \stackrel{\text{def}}{=} a_i + nd_i$, then the "diagonal" arithmetic progression obtained by

$$A_i' \stackrel{\text{def}}{=} \{a_i, a_i + (d_i + 2t \cdot N_r), \ldots, a_i + (n-1)(d_i + 2t \cdot N_r)\}$$

is also a monochromatic sequence which focus on a point $a' \stackrel{\text{def}}{=} a_i + n(d_i + 2t \cdot N_r)$.

Because we assumed the A_i are colour–focused, it is easy to see that A_i' are also colour–focused, with focus $a' \stackrel{\text{def}}{=} (a_i + n \cdot d_i) + 2ntN_r = a + 2ntN_r$.

However, the set of "focus points" of each row $A \stackrel{\text{def}}{=} \{a, a + 2t \cdot N_r, \ldots, a + (n-1)(2t \cdot N_r)\}$ are also monochromatic (since all rows of G are coloured the same), is of length n, and focuses on $a + 2nt \cdot N_r$. Furthermore, the colour of this arithmetic progression cannot be the same as any of the A_i (and hence A_i'), because otherwise we will get a monochromatic arithmetic progression length n in each row.

Hence, $\{A, A_1', \ldots, A_r'\}$ is colour–focused and of size $r+1$, completing the induction (and hence the proof of Van der Waerden). $\square$

Remark. While the above proof gives a computable upper–bound on $W(n, c)$, in practice it is too large to compute, even for small values of n and c. It turns out we can do far better; the best upper bound established asserts that

$$W(n, c) \leq 2^{2^{n^{2^{2^{c+9}}}}}.$$

Later, we will see several proofs of the Hales–Jewett theorem, a generalisation of Van der Waerden. One of these proofs results in a better bound of $W(n, c)$ than given here.

3 The Hales–Jewett Theorem

The Hales–Jewett Theorem can be viewed as a purely combinatorial generalisation of Van der Waerden's Theorem, stated as a colouring problem on a hypercube.

Theorem 3.1 [**Hales–Jewett**]. Let $n, c \geq 1$ be integers. There exists $H = H(n, c)$ such that any colouring of the H–dimensional hypercube with side length n, with c colours, contains n collinear lattice points all of the same colour.

A common way of interpreting this theorem is that, given a dimension n of a tic–tac–toe board and c players, the game will always have a winner, assuming the dimension H of the board is large enough.

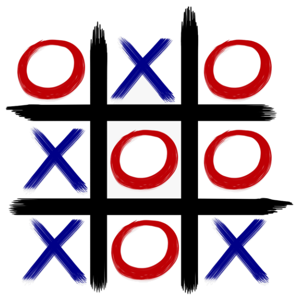


Figure 3: This board (with no winner) shows that $H(3, 2) > 2$.

Before we attempt to prove this result, we need to build a rigorous and workable definition of "n collinear lattice points". It turns out that, in fact, the correct definition we want to use is more restrictive than the geometric equivalent.

Definition. Let A be a set, and denote A^n as the set of n–tuples over A. For integers $m, n \geq 0$, take m elements $\lambda_0, \ldots, \lambda_{m-1}$ to be distinct elements not in A. An *m-parameter word* is an n–tuple

$$f = (f_0, \ldots, f_{n-1}) \in (A \cup \{\lambda_0, \ldots, \lambda_{m-1}\})^n.$$

Notice that f can equivalently be defined as a function $f : \{0, \ldots, n-1\} \to A \cup \{\lambda_0, \ldots, \lambda_{m-1}\}$. Hence, we also refer to f as an *m-parameter map.*

Intuitively, these words can be thought of as a string of length n, where each element of the string is either in A, or are one of m "unknown" parameter elements λ_i.

These words have both the properties of a string (via concatenation) and a function (via function composition), as we define below.

Definition. Let $f \in (A \cup \{\lambda_0, \ldots, \lambda_{m-1}\})^n$ and $g \in (A \cup \{\lambda_0, \ldots, \lambda_{k-1}\})^m$. We define the *composition* $f \circ g \in (A \cup \{\lambda_0, \ldots, \lambda_{k-1}\})^n$, by replacing each occurrence of λ_i in f, by g_i, the i'th component of g.

If $g \equiv a \in A^m$ is a 0–parameter word (i.e. has no λ's), we define $f(a) \overset{\text{def}}{=} f \circ a$.

Definition. For parameter words f, g as above, define the *concatenation* $f + g \overset{\text{def}}{=} f_0 \ldots f_{n-1} g'_0, \ldots, g'_{m-1}$ where $g'_0 \ldots g'_{m-1}$ is the word formed by replacing the $(\lambda_0, \ldots, \lambda_{k-1})$'s in g with $(\lambda_m, \ldots, \lambda_{m+k-1})$.

Example. Let $A \overset{\text{def}}{=} \{a, b, c, x, y\}$. We have the following word composition:

$$\begin{array}{rcccccccr} f & \overset{\text{def}}{=} & a & \lambda_0 & b & \lambda_1 & c & \in & (A \cup \{\lambda_0, \lambda_1\})^5 \\ g & \overset{\text{def}}{=} & & x & & \lambda_0 & & \in & (A \cup \{\lambda_0\})^2 \\ \implies f \circ g & = & a & x & b & \lambda_0 & c & \in & (A \cup \{\lambda_0\})^5 \end{array}$$

We also have $f + g = a\,\lambda_0\,b\,\lambda_1\,c\,x\,\lambda_2 \in (A \cup \{\lambda_0, \lambda_1, \lambda_2\})^7$.

With these definitions, we have some groundwork to define a combinatorial line. However, currently we have some redundancies with our definitions: any permutation of the λ_i's will result in a different, yet essentially equivalent, parameter word. To fix this, we simply give the λ_i's an order.

Definition. For integers $n \geq m$, we denote $[A]\binom{n}{m}$ as the set of m–parameter words, such that:

- Each λ_i appears in f; and
- If $i < j$, then the first occurrence of λ_i appears **before** the first occurrence of λ_j.

From this point, we will abuse notation and refer to "m–parameter words" as elements of $[A]\binom{n}{m}$.

Example. If $A = \emptyset$, then $[\emptyset]\binom{n}{k}$ contains words with only λ_i's. This set has the interpretation of equivalence relations on the set $\{0, \ldots, n-1\}$ (where, for a map $f \in [\emptyset]\binom{n}{k}$, we say $i \sim j \iff f(i) = f(j)$).

On the other hand, if $m = 0$, the set $[A]\binom{n}{0}$ contains no parameters λ_i and is hence isomorphic to A^n.

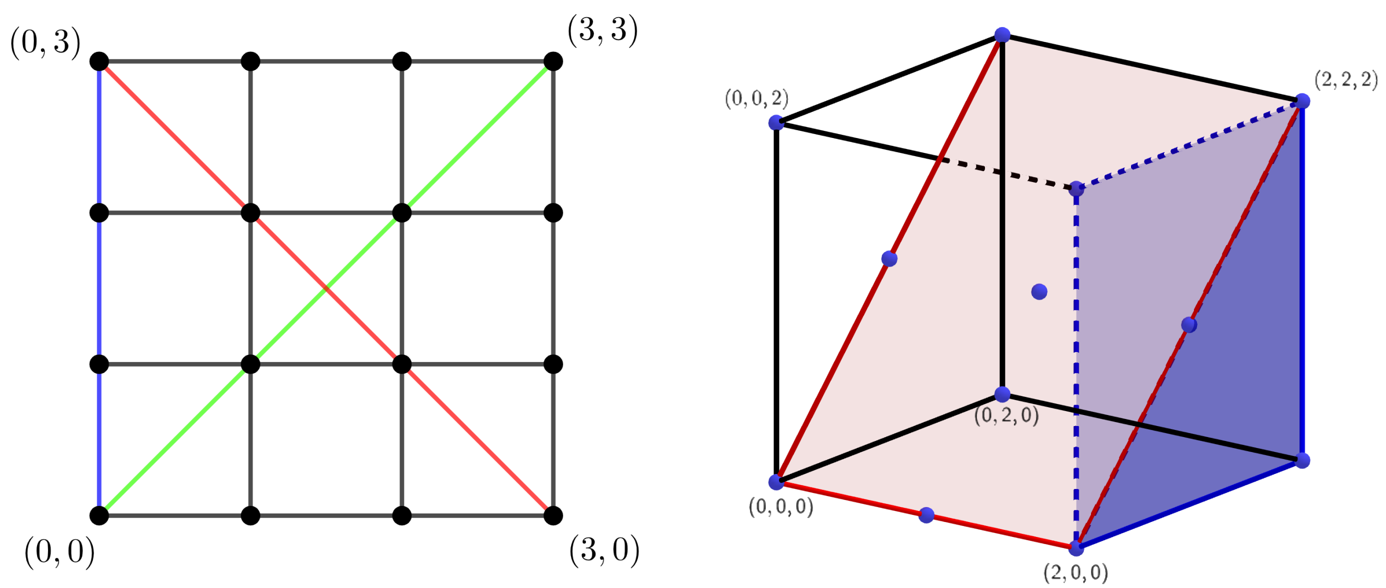
Of course, the point of defining these parameter words/maps is to create a rigorous and workable definition of a "combinatorial line". This is what we define next:

Definition. Let $f \in [A]\binom{n}{d}$. The *combinatorial d–space* generated by f, is the set $L_f \stackrel{\text{def}}{=} \{f(a) : a \in [A]\binom{d}{0}\}$.

If $d = 1$, then we call it a *combinatorial line.*

This definition makes sense, because if we let $A \stackrel{\text{def}}{=} \{0, \ldots, N-1\}$ and "plot" the set L_f on an n–dimensional coordinate space, we recover a set of lattice points of size N^d which lie on a subspace of dimension d. Though, not all *geometric* lines and spaces can be represented in this way.

Hence, given a colouring $\Delta : A^n \to \{0, \ldots, c-1\}$, finding a monochromatic combinatorial line (or d–space) is equivalent to finding $f \in [A]\binom{n}{d}$ such that $\Delta \circ f$ is constant on A^d.



(a) The blue and green lines are combinatorial lines, represented by $0\,\lambda$ and $\lambda\,\lambda$ respectively. The red is not a combinatorial line.

(b) The blue plane is a combinatorial 2–space, given by $2\,\lambda_1\,\lambda_2$. The red plane is not a combinatorial 2–space.

With these new definitions, it is possible to restate the Hales–Jewett Theorem in a formal way. In fact, we can naturally generalise the theorem for higher combinatorial d–spaces, for any $d \geq 1$.

Theorem 3.2 [Hales–Jewett, d–space]. Let $n, d, c \geq 1$ be integers, and A a set of size n. Then, there exists $H = H(n, d, c)$ such that for any colouring $\Delta : A^H \to \{0, \ldots, r-1\}$, then there exists $f \in [A]\binom{H}{d}$ such that $\Delta \circ f$ is constant on A^d.

Proof. Of course, we want to use induction on n and d, though it is unclear which should be applied first. It seems natural to deal with $H(n, d+1, \bullet)$ given information about $H(n, d, \bullet)$, as we want to "add" a parameter, which seems easier than dealing with a new character in A and all the new coloured words resulting from its addition.

Hence, we apply induction on n, followed by d. For $n = 1$, clearly $H(1, d, c) = d$, because we have a cube with side length 1 (of any dimension), so the entire set, consisting of a single point, is a combinatorial d–space as long as $H \geq d$.

For the inductive base case, assume $d = 1$ and $A' = A \cup \{b\}$ has size $n + 1$. We prove a bound for $H(n+1, 1, c)$ by induction on c; clearly, $H(n+1, 1, 1) = 1$ as we can choose $f = \lambda_0$.

For some integer h (to be chosen later), suppose we have a colouring $\Delta : (A \cup \{b\})^h \to \{0, \ldots, c-1\}$. This gets the obvious restricted colouring

$$\Delta_A = \Delta|_{A^h} : A^h \to \{0, \ldots, c-1\}$$

onto the set of words A^h, where $|A| = n$. Hence, as long as we choose $h \overset{\text{def}}{=} H(n, N, c)$ for some integer $N \geq 1$, by induction we can find a combinatorial N–space $f \in [A]\binom{h}{N}$ that is monochromatic under Δ_A, that is, $\Delta_A \circ f$ is constant on A^N. WLOG say $\Delta_A \circ f \equiv 0$. That is, we have

$$\forall (a_0, \ldots, a_{N-1}) \in A^N,\ \Delta(f(a_0, \ldots, a_{N-1})) = 0. \tag{$\dagger$}$$

We now have two cases:

- If there exists $a_1, \ldots, a_{N-1} \in A \cup \{b\}$ for which $\Delta \circ f(b, a_1, \ldots, a_{N-1}) = 0$, then let F be the word defined by
$$F = f(\lambda, a_1, \ldots, a_{N-1}).$$
Clearly, by definition of f being a N–parameter word, F is a 1–parameter word, i.e. $F \in [A]\binom{h}{1}$. Furthermore, by combining the above assertion with ($\dagger$), we see that F is indeed monochromatic under Δ, as desired.

- Otherwise, for any $a_1, \ldots, a_{N-1} \in A \cup \{b\}$, we have $\Delta \circ f(b, a_1, \ldots, a_{N-1}) \neq 0$. We naturally consider the restricted colouring
$$\begin{aligned} \Delta' : (A \cup \{b\})^{N-1} &\to \{1, \ldots, c-1\} \\ (a_1, \ldots, a_{N-1}) &\mapsto \Delta \circ f(b, a_1, \ldots, a_{N-1}). \end{aligned}$$
This motivates us to choose $N \overset{\text{def}}{=} 1 + H(n+1, 1, c-1)$, as then we can use the inductive hypothesis to find a combinatorial line $f^0 \in [A \cup \{b\}]\binom{N-1}{1}$ which is monochromatic under Δ', i.e. $\Delta' \circ f^0$ is constant on $A \cup \{b\}$.

  This produces the obvious 1–parameter map $F^0 \in [A \cup \{b\}]\binom{h}{1}$ defined by
$$F^0(g) \overset{\text{def}}{=} f(b, f^0(g))$$
which is monochromatic under Δ, because $(\Delta \circ F^0)(g) = \Delta \circ f(b, f^0(g)) = \Delta' \circ f^0(g) = \text{const}$.

Hence, we obtain $H(n, 1, c)$ is finite for each $n, c \geq 1$.

Finally, we do the inductive hypothesis. Assume that for some $n, d \geq 1$, the number $H(n, d, c)$ is finite for *any* $c \geq 1$. Motivated by the proof of Van der Waerden's Theorem, we would like to use the finiteness of $H(n, d, \bullet)$ to create a map which is constant on an entire set of words of A.

To do this, let $h, h' \in \mathbb{N}$ to be defined later. Notice that any colouring $\Delta : A^{h+h'} \to \{0, \ldots, c-1\}$ induces the colouring

$$\begin{aligned} \Delta' : A^h &\to \{0, \ldots, c-1\}^{n^{h'}} \\ (a_0, \ldots, a_{h-1}) &\mapsto \Big(\Delta(a_0, \ldots, a_{h-1}, b_0, \ldots, b_{h'-1}) : (b_0, \ldots, b_{h'-1}) \in A^{h'} \Big). \end{aligned}$$

Like we did before, we can view this as a "colouring" of A^h in one of $c^{n^{h'}}$ colourings, one for each word in $\{0, \ldots, c-1\}^{n^{h'}}$. This motivates us to choose $h \stackrel{\text{def}}{=} H(n, d, c^{n^{h'}})$, as then we can find a combinatorial d–space $f^1 \in [A]\binom{h}{d}$ which is monochromatic under Δ'.

What does this mean? Explicitly, there exists some tuple $t \in \{0, \ldots, c-1\}^{n^{h'}}$ (indexed by elements of $A^{h'}$) such that for any $a_0, \ldots, a_{d-1} \in A$, we have

$$\forall\, \mathbf{b} \stackrel{\text{def}}{=} (b_0, \ldots, b_{h'-1}) \in A^{h'},\ \Delta(f^1(a_0, \ldots, a_{d-1}),\ \mathbf{b}) = t_{\mathbf{b}}.$$

Now that we've sorted the first h entries in $A^{h+h'}$, we turn our attention to the last h'. Fix a word $\mathbf{a} \in A^d$. Similar to the above, we can define another induced map by

$$\begin{aligned} \Delta'' : A^{h'} &\to \{0, \ldots, c-1\} \\ (b_0, \ldots, b_{h'-1}) &\mapsto \Delta(f^1(\mathbf{a}), b_0, \ldots, b_{h'-1}). \end{aligned}$$

This time, we are motivated to choose $h' \stackrel{\text{def}}{=} H(n, 1, c)$, as then we can find a monochromatic combinatorial line $f^0 \in [A]\binom{h'}{1}$ such that $\Delta'' \circ f^0$ is constant on A. We can now "stick together" the two monochromatic spaces we have found, by defining a new *concatenation* map $f \in [A]\binom{h+h'}{d+1}$ with $d+1$ parameters, by:

$$f(a_0, \ldots, a_d) \stackrel{\text{def}}{=} f^1(a_0, \ldots, a_{d-1}) + f^0(a_d).$$

This is monochromatic under Δ, because

$$\begin{aligned} (\Delta \circ (f^1 + f^0))(a_0, \ldots, a_d) &= \Delta(f^1(a_0, \ldots, a_{d-1}), f^0(a_d)) \\ &= t_{f^0(a_d)} \\ &= \Delta(f^1(\mathbf{a}), f^0(a_d)) \\ &= \Delta''(f^0(a_d)) = (\Delta'' \circ f^0)a_d \end{aligned}$$

which is constant for any $a_0, \ldots, a_d$ by property of f^0, as desired.

Hence, by induction, not only have we shown $H(n, d, c)$ are finite for all $n, d, c \geq 1$, but we have also obtained the bounds:

1. $H(1, d, c) = d$;
2. $H(n+1, 1, c) \leq H(n, H(n+1, 1, c-1) + 1, c)$ for $c \geq 2$, and $H(n+1, 1, 1) = 1$;
3. $H(n, d+1, c) \leq H(n, 1, c) + H(n, d, c^{n^{H(n,1,c)}})$. □

Remark. The bound proved for $H(n, d, c)$ appears similar to the Ackermann function, and is so large to the point of not being primitive recursive[1]. As we will see in the next section, we will be able to give another proof of the Hales–Jewett Theorem, which gets a primitive recursive (yet still enormous) upper bound for H.

[1] Recall that a *primitive recursive function* is, roughly speaking, a function that can be computed with a finite number of "for" loops on a computer program.

Having proved this theorem, we now focus on some immediate corollaries. As promised, we will deduce the Van der Waerden's theorem from Hales–Jewett.

Proof of Van der Waerden's Theorem. Let $H \stackrel{\text{def}}{=} H(n, c)$ and $W(n, c) \stackrel{\text{def}}{=} H \cdot c$. It suffices to show that any colouring $\chi : \{0, \dots, W-1\} \to \{0, \dots, c-1\}$ contains a monochromatic arithmetic progression of length n.

Let $A \stackrel{\text{def}}{=} \{0, 1, \dots, n-1\}$. Notice that there is a natural "projection" mapping given by:

$$\begin{aligned} \pi : A^H &\to \{0, \dots, W-1\} \\ (a_1, \dots, a_H) &\mapsto a_1 + \dots + a_H. \end{aligned}$$

Using this projection, we can define a colouring on A^H, by:

$$\begin{aligned} \Delta : A^H &\to \{0, \dots, c-1\} \\ \mathbf{a} &\mapsto \chi \circ \pi(\mathbf{a}). \end{aligned}$$

By definition of $H \stackrel{\text{def}}{=} H(n, c)$, there exists a combinatorial line $f \in [A]\binom{H}{1}$ which is monochromatic under Δ, i.e. $\Delta \circ f \stackrel{\text{def}}{=} \chi \circ \pi \circ f = \text{const}$, say r.

Since f (as a word) consists of elements of A and an unknown parameter λ, and π sums the coordinates of f, we can let $P > 0$ be the number of λ's in f and Q the sum of non–λ elements of f. Hence, we get $\pi \circ f(k) = P \cdot k + Q$, and so

$$\forall\, k \in \{0, \dots, n-1\} : \ \chi(P \cdot k + Q) = r$$

and we have obtained our monochromatic arithmetic sequence of length n. □

Finally, we can prove a slightly generalised version of the Van der Waerden's theorem, which allows any finite set of points in $\mathbb{Z}^d$.

Theorem 3.3 [Gallai–Witt]. Let $A, B \subset \mathbb{Z}^d$ be finite sets. Say A is a *homothetic copy* of B, if there exists $u \in \mathbb{Z}^d$ and $\alpha \in \mathbb{N}$ such that

$$A = u + \alpha \cdot V \stackrel{\text{def}}{=} \{u + \alpha \cdot \mathbf{b} : \mathbf{b} \in B\}.$$

Suppose we have a colouring $\chi : \mathbb{Z}^d \to \{0, \dots, c-1\}$. Then, any finite set $A \subset \mathbb{Z}^d$ has a monochromatic, homothetic copy in $\mathbb{Z}^d$.

Proof. A similar argument to Van der Waerden's theorem works. Let $A = \{v_0, \dots, v_{n-1}\}$ be a set of vectors of dimension d, and $H \stackrel{\text{def}}{=} H(n, c)$. Consider the projection map

$$\begin{aligned} \pi : A^H &\to \mathbb{Z}^d \\ (\mathbf{a}_1, \dots, \mathbf{a}_H) &\mapsto \mathbf{a}_1 + \dots + \mathbf{a}_H. \end{aligned}$$

This defines a colouring $\Delta : A^H \to \{0, \dots, c-1\}$ by $\Delta \stackrel{\text{def}}{=} \chi \circ \pi$.

By definition of $H \stackrel{\text{def}}{=} H(n, c)$, there exists a combinatorial line $f \in [A]\binom{H}{1}$ which is monochromatic under Δ, or $\Delta \circ f \equiv \text{const}$; that is, for $0 \le i < n$, we have $\pi \circ f(v_i)$ is monochromatic under χ.

However, $\pi \circ f(v_i)$ is the sum of coordinates of $f(v_i)$. So, if we let $\mathbf{u}$ be the sum of non–λ elements of f, and $\alpha > 0$ the number of λ's in f, then we get $\pi \circ f(v_i) = \mathbf{u} + \alpha \cdot v_i$. Hence:

$$\pi \circ f(A) = \{\mathbf{u} + \alpha \cdot v_i : \ 0 \le i < n\} = \mathbf{u} + \alpha \cdot V$$

and hence we have found our homothetic, monochromatic copy of V. □

3.1 Shelah's Proof of Hales–Jewett Theorem

As discussed before, the previous proof of the Hales–Jewett Theorem has the unfortunate consequence of producing a non–primitive recursive upper bound for $H(n, c)$. The following proof offers an alternative method which does produce such an upper bound.

Theorem 3.1 [Hales–Jewett]. Let $n, c \geq 1$ be integers. There exists $H = H(n, c)$ such that any colouring of the H–dimensional hypercube with side length n, with c colours, contains n collinear lattice points all of the same colour.

Proof [Shelah]. We proceed by induction on n. Observe that for $n = 1$, $H(1, c) = 1$ for any $c \geq 1$, as there is one point in the hypercube which is its own "collinear" line.

Assume the statement holds for $(n-1, c)$ for some $n \geq 2$ and $c \geq 1$, and define $h \overset{\text{def}}{=} H(n-1, c)$. We show that there exists a choice of positive integers $N_1 < \cdots < N_h$, such that if $N \overset{\text{def}}{=} N_1 + \cdots + N_h$, then any colouring $\Delta : A^N \to \{0, \ldots, c-1\}$ will contain a monochromatic combinatorial line. As discussed previously, it suffices to find a 1–parameter word $f \in [A]\binom{N}{1}$ for which $\Delta \circ f$ is constant on A.

For the rest of this proof, WLOG assume $A = \{0, \ldots, n-1\}$.

Definition. We say two strings $a, b \in A^n$ are *neighbours*, if they differ in exactly one coordinate.

We further say a, b are *binary neighbours*, if they are neighbours and the coordinate i for which they differ satisfy $\{a_i, b_i\} = \{0, 1\}$.

The crux of the entire proof comes down to the following lemma.

Lemma 3.4. For some choice of $N_1 < \cdots < N_h$, there exists 1–parameter words $g^1, \ldots, g^h$ with $g^i \in [A]\binom{N_i}{1}$, such that their concatenation $g \overset{\text{def}}{=} g^1 + \cdots + g^h \in [A]\binom{N}{h}$ satisfies for any binary neighbours $a, b \in A^h$,

$$\Delta(g(a)) = \Delta(g(b)).$$

Before we prove this lemma, let's first understand why it suffices to prove this. For simplicity, denote $A_0 \overset{\text{def}}{=} A \backslash \{0\}$. Notice that, by restriction, we can define a new colouring

$$\begin{aligned} \Delta' : A^h &\to \{0, \ldots, c-1\} \\ a &\mapsto \Delta(g(a)) \end{aligned}$$

where g is the concatenation word found above. Since $|A_0| = n - 1$ and $h \overset{\text{def}}{=} H(n-1, c)$, by further restricting Δ' to $(A_0)^h$ we can find a 1–parameter word $p \in [A_0]\binom{h}{1}$ for which $p(A_0)$ is monochromatic under Δ'. That is, we have

$$\Delta'(p(1)) = \cdots = \Delta'(p(n-1)).$$

However, by definition, $p(0)$ is obtained by replacing all occurrences of λ in p by 0, and similar for $p(1)$. This means we can reach $p(1)$ from $p(0)$ by flipping some 0's to 1's, and so there exists a sequence of strings $a_i \in A^h$ for which $p(0) = a_0, a_1, \ldots, a_k = p(1)$ and each $\{a_i, a_{i+1}\}$ are binary equivalent. Hence, Lemma 3.4 applies and we obtain

$$\Delta(g(p(0))) = \Delta(g(p(1))) \implies \Delta'(p(0)) = \Delta'(p(1)).$$

Hence, $\Delta' \circ p = \Delta \circ g \circ p$ is constant on the entire of A, giving the monochromatic line $f \overset{\text{def}}{=} g \circ p \in [A]\binom{N}{1}$.

It hence suffices to prove Lemma 3.4. We will construct the g^i inductively (using backwards induction). Suppose we have defined, for some $1 \leq i \leq n$, the words $g^{i+1}, \ldots, g^h$ (if $i = h$, then we have not yet defined anything), and we would like to define $g^i \in [A]\binom{N_i}{1}$ of length N_i.

Suppose a, b are binary neighbours, that is,

$$a = a_1 \dots a_{i-1}\ 0\ a_{i+1} \dots a_h \implies g(a) = g^1(a_1) \dots g^{i-1}(a_{i-1})\ g^i(0)\ g^{i+1}(a_{i+1}) \dots g^h(a_h)$$
$$b = a_1 \dots a_{i-1}\ 1\ a_{i+1} \dots a_h \implies g(b) = g^1(a_1) \dots g^{i-1}(a_{i-1})\ g^i(1)\ g^{i+1}(a_{i+1}) \dots g^h(a_h).$$

We would like Δ to send both sequences to the same colour. Of course, by our inductive hypothesis, we want to be able to construct g^i independently of any g^j for any $j < i$.

The idea is to consider g^i of the form

$$g^i \stackrel{\text{def}}{=} \underbrace{00\dots0}_{s}\underbrace{\lambda\lambda\dots\lambda}_{t-s}\underbrace{11\dots1}_{N_i-k}$$

for some numbers $1 \le s < t \le N_i$. In this case, if we define $W_k \stackrel{\text{def}}{=} \underbrace{00\dots0}_{k}\underbrace{11\dots1}_{N_i-k}$, then we have $g^i(0) = W_t$ and $g^i(1) = W_s$. This motivates the following construction: for some integer $M = M_i$, and for each number $0 \le k \le N_i$, consider the following induced colouring:

$$\Delta_k : A^{M+h-i} \to \{0, \dots, c-1\}$$
$$(x_1, \dots, x_M, y_{i+1}, \dots, y_h) \mapsto \Delta(x_1, \dots, x_M, W_k, g^{i+1}(y_{i+1}), \dots, g^h(y_h)).$$

Because Δ takes in strings of length N, we can see that $M = N_1 + \dots + N_{i-1}$ for $i \ge 1$, and 0 else.

Then, by the previous discussion, if we can find numbers $s < t$ for which $\Delta_s = \Delta_t$, then we have

$$\begin{aligned}
\Delta(g(a)) &= \Delta(g^1(a_1) \dots g^{i-1}(a_{i-1}),\ g^i(0),\ g^{i+1}(a_{i+1}) \dots g^h(a_h)) \\
&= \Delta(g^1(a_1) \dots g^{i-1}(a_{i-1}),\ W_t,\ g^{i+1}(a_{i+1}), \dots g^h(a_h)) \\
&= \Delta_t(g^1(a_1), \dots g^{i-1}(a_{i-1}),\ a_{i+1} \dots a_h) \\
&= \Delta_s(g^1(a_1), \dots g^{i-1}(a_{i-1}),\ a_{i+1} \dots a_h) \\
&= \Delta(g^1(a_1) \dots g^{i-1}(a_{i-1}),\ W_s,\ g^{i+1}(a_{i+1}) \dots g^h(a_h)) \\
&= \Delta(g^1(a_1) \dots g^{i-1}(a_{i-1}),\ g^i(1),\ g^{i+1}(a_{i+1}) \dots g^h(a_h)) \\
&= \Delta(g(b))
\end{aligned}$$

and the proof would be completed.

Hence, it remains to show that we can find such s, t. Notice that we have $N_i + 1$ such colouring maps Δ_k, and the total number of possible colourings $d : A^{M+h-i} \to \{0, \dots, c-1\}$ is

$$c^{\#\text{ words}} = c^{n^{M+h-i}} < c^{n^{M+h}}.$$

By the pigeonhole principle, it suffices to let $N_0 \stackrel{\text{def}}{=} c^{n^h}$, and $N_i = c^{n^{h+N_1+\dots+N_{i-1}}}$ for $i \ge 1$. In this case, as there are more Δ_k's than possible colourings of A^{M+h-i}, it follows that some two, say Δ_s and Δ_t, are the same, as required.

In conclusion, we have shown that, if $h \stackrel{\text{def}}{=} H(n-1, c)$ and we choose $N_1 < \dots < N_h$ by

$$N_1 \stackrel{\text{def}}{=} c^{n^h} \quad \text{and} \quad N_i \stackrel{\text{def}}{=} c^{n^{h+N_1+\dots+N_{i-1}}},$$

then any colouring $\Delta : A^N \to \{0, \dots, c-1\}$ will contain a monochromatic combinatorial line, and hence $H(n, c) \le N$, as required. □

Remark. It is an exercise for the reader to prove the following result on the growth rate of $H(n, c)$. Define $h_1(n) \stackrel{\text{def}}{=} 2 \cdot n$ and $h_i(n) \stackrel{\text{def}}{=} \underbrace{h_{i-1}(\dots(h_{i-1}(1)))}_{n}$. Then $H(n, c)$ is order at most h_4; explicitly, we have

$$H(n, c) \le \frac{1}{n \cdot c} h_4(n + c + 2).$$

It is possible to prove a primitive recursive bound for the general, dimensional version $H(n, d, c)$, though we will not do so here.

3.2 A $\star$–version of Hales–Jewett

So far, we have the Hales–Jewett theorem for fixed–length sequences of A. Could we generalise this theorem to *variable–length* words? Of course, we don't necessarily know which words to admit or how composition works, so we'll need to define those first.

Definition. Let $A \stackrel{\text{def}}{=} \{0, 1, \ldots, n-1\}$ be a set, and $\lambda_0, \ldots, \lambda_{m-1}, \star$ be distinct elements not in A. The set of *variable m–parameter words* $[A]^\star\binom{n}{m}$ are the m–parameter words $f \in [A \cup \{\star\}]\binom{n}{m}$, where $\star$ either does not appear, or the set of appearances forms a suffix array. That is, for some $n_0 \leq n$ and $g \in [A]\binom{n_0}{m}$:

$$f_i = \begin{cases} g_i & \text{if } i \leq n_0 \\ \star & \text{if } i > n_0. \end{cases}$$

This can be interpreted as a collection of variable–length m–parameter words, because for each $n_0 \leq n$ we can "fill up" the end of every $g \in [A]\binom{n_0}{m}$ with stars to get an element in $[A]^\star\binom{n}{m}$.

Definition. For each $0 \leq i < n$, we denote $[A]^i\binom{n}{m}$ as the set of words in $[A]^\star\binom{n}{m}$ with exactly i stars.

It is immediate from the definition that $\{[A]^i\binom{n}{0} : i < n\}$ partition $[A]^\star\binom{n}{0}$ by the number of stars.

However, we haven't yet defined how to compose two of these, so this is what we'll do next.

Definition. For variable length words $f \in [A]^\star\binom{n}{m}$ and $g \in [A]^\star\binom{m}{k}$, define the *composition* $f \circ g \in [A]^\star\binom{n}{k}$:

$$(f \circ g)_i = \begin{cases} \star & \text{if } \exists j < i : (f \circ g)_j = \star; \\ f_i & \text{if } \forall j < i : (f \circ g)_j \neq \star, \text{ and } f_i \in A \cup \{\star\}; \\ g_j & \text{if } \forall j < i : (f \circ g)_j \neq \star, \text{ and } f_i = \lambda_j. \end{cases}$$

Intuitively, we take the composition $f \circ g$ as variable length words over $A \cup \{\star\}$ in the usual way, and "delete" its elements from the right until there are no more stars. Then, we just add stars on the right until it has length n.

It is an exercise for the reader to verify that $f \circ g$ is an element of $[A]^\star\binom{n}{k}$, i.e. it has k parameters.

Example. Let $A \stackrel{\text{def}}{=} \{a, b, c, x, y\}$.

$$\begin{array}{rcccccccccccc} f & \stackrel{\text{def}}{=} & a & \lambda_0 & b & \lambda_1 & c & \lambda_2 & d & \star & \star & \in & [A \cup \{\star\}]\binom{9}{3} \\ g & \stackrel{\text{def}}{=} & & x & & \lambda_0 & & \star & & & & \in & [A \cup \{\star\}]\binom{3}{1} \\ \implies f \circ g & = & a & x & b & \lambda_0 & c & \star & d & \star & \star & \in & [A \cup \{\star\}]\binom{9}{1} \\ \implies f \circ g & = & a & x & b & \lambda_0 & c & \star & \star & \star & \star & \in & [A]^\star\binom{9}{1} \end{array}$$

With these stated, it is possible to state the $\star$–version of Hales–Jewett.

Theorem 3.5 [Hales–Jewett, $\star$–version]. Let $n, d, c \geq 1$ be integers, and A a set of size n. Then, there exists $H = H^\star(n, d, c)$ such that for any colouring $\Delta : [A]^\star\binom{H}{0} \to \{0, \ldots, r-1\}$, there exists $f \in [A]\binom{H}{d}$ such that $\Delta \circ f$ is constant on $[A]^\star\binom{d}{0}$.

This version will be important when proving the Graham–Rothschild Theorem, in the next section.

Proof. Define positive integers $\{h_i : 0 \leq i \leq cd\}$ inductively by $h_{cd} = cd$, and for $0 \leq j < cd$:

$$h_j \stackrel{\text{def}}{=} j + H(n, h_{j+1} - j, c).$$

Let $h \stackrel{\text{def}}{=} h_0$. We claim that $H^\star(n, d, c) \leq h$; that is, for any colouring $\Delta : [A]^\star\binom{h}{0} \to \{0, \ldots, c-1\}$, we want to find $f \in [A]\binom{h}{d}$ such that $\Delta \circ f$ is constant.

In a sense, the set $[A]^\star\binom{h}{0}$ is just a disjoint union of elements with k stars (for each $0 \leq k \leq h$). The Hales–Jewett theorem tells us information about each individual $[A]^k\binom{h}{0}$; hence, to prove a statement about the whole set, we want to combine this information. This is set out in the next lemma.

Lemma 3.6. For each $j \le cd$, there exists a word $f^j \in [A]\binom{h}{h_{j+1}}$ such that the function $\Delta \circ f^j$ is constant on $[A]^k\binom{h_{j+1}}{0}$ (for each $k \le j$).

Proof. We use induction on j. For $j = 0$, note by the recurrence we have $h \stackrel{\text{def}}{=} h_0 \stackrel{\text{def}}{=} H(n, h_1, r)$ and so by definition there exists $f^0 \in [A]\binom{h}{h_1}$ such that $\Delta \circ f^0$ is constant on $A^{h_1} \equiv [A]^0\binom{h_1}{0}$, which is what we need.

Suppose we have the result for some $0 \le j < mr$. That is, we have a h_{j+1}–parameter word $f^j \in [A]\binom{h}{h_{j+1}}$ for which $\Delta \circ f^j$ is constant on $[A]^k\binom{h_{j+1}}{0}$, for each $k \le j$.

The idea is to consider a map f^{j+1} of the form

$$f^{j+1} = f^j(F, \lambda_N, \ldots, \lambda_{N+j}) \tag{$\diamondsuit$}$$

where F is some smaller N–parameter map (for some N). This way, by construction, f^{j+1} is already constant on the set of words with a fixed $k \le j$ stars, and it suffices to consider words with $j + 1$ stars.

This motivates us to consider an induced map:

$$\begin{aligned} \Delta^{j+1} : [A]^{j+1}\binom{h_{j+1}}{0} &\to \{0, \ldots, c-1\} \\ g &\mapsto \Delta \circ f^j(g). \end{aligned}$$

Since by definition, the last $j + 1$ elements of any $f \in [A]^{j+1}\binom{h_{j+1}}{0}$ is a $\star$, we can consider Δ^{j+1} as a colouring $\Delta_0^{j+1} : [A]\binom{h_{j+1}-(j+1)}{0} \to \{0, \ldots, c-1\}$ by removing the last $j + 1$ stars.

However, since we have $h_{j+1} - (j + 1) \stackrel{\text{def}}{=} H(n, h_{j+2} - (j + 1), c)$, we can find $F \in [A]\binom{h_{j+1}-(j+1)}{h_{j+2}-(j+1)}$ for which $\Delta_0^{j+1} \circ F \equiv$ const on $A^{h_{j+2}-(j+1)}$.

It turns out that this F satisfies the requirements in ($\diamondsuit$), where we let $N \stackrel{\text{def}}{=} h_{j+2} - (j + 1)$. Consider the $N + (j + 1) = h_{j+2}$–parameter word $(F, \lambda_N, \ldots, \lambda_{N+j})$. Then, for any $g \in [A]^\star\binom{h_{j+2}}{0}$ with $j + 1$ stars, $(F, \lambda_N, \ldots, \lambda_{N+j}) \circ g$ also has $j + 1$ stars, and so

$$\Delta \circ f^{j+1} = \Delta^{j+1}(F, \lambda_N, \ldots, \lambda_{N+j}) = \Delta_0^{j+1} \circ F \equiv \text{const}$$

on the set $[A]^\star\binom{h_{j+2}}{0}$, as desired. Hence, this f^{i+1} satisfies the inductive step.

Finally, we return to the main proof. For $j \stackrel{\text{def}}{=} cd$, we obtain a word $f^{cd} \in [A]\binom{h}{cd}$ such that for any $k \le cd$, the value of $\Delta \circ f^{cd}$ is constant on $[A]^k\binom{cd}{0}$. This induces an r–colouring of the integers

$$\begin{aligned} \Delta' : \{0, \ldots, cd\} &\to \{0, \ldots, c-1\} \\ i &\mapsto \Delta(f^{cd} \circ g) \quad \text{for any } g \in [A]^i\binom{cd}{0}. \end{aligned}$$

Hence, there exists integers $0 \le i_0 < \cdots < i_d \le cd$ which are monochromatic under Δ', say $= r$.

Pick some $a \in A$, and define a map $T \in [A]^\star\binom{cd}{d}$ by

$$T_i = \begin{cases} a & \text{if } i < cd - i_d; \\ \lambda_j & \text{if } \exists j,\ (cd - i_{j+1}) \le i < (cd - i_j); \\ \star & \text{if } i > (cd - i_0). \end{cases}$$

Then, for any $g \in [A]^\star\binom{d}{0}$, notice that by the way we compose things, $T \circ g$ has exactly i_k number of stars, for some $k \le d$. It follows that $f \stackrel{\text{def}}{=} f^{cd} \circ T$ is monochromatic, since for any $g \in [A]^\star\binom{d}{0}$,

$$\Delta \circ f(g) = \Delta' \circ T(g) = r.$$

Clearly, f has d parameters and is a variable–length word. Hence, this $f \in [A]^\star\binom{h}{d}$ satisfies the requirements of the theorem. $\square$

4 The Graham–Rothschild Theorem

Finally, we reach the most general of these theorems: the Graham–Rothschild Theorem. This theorem allows us to not only colour points, but the *k–parameter words* themselves. The theorem also generalises Hales–Jewett, to allow a group G acting on A.

This is particularly useful in graph theory for the case $k = 2$, where we want to colour *edges*, which are essentially functions that "pick out" a set of two distinct vertices from the vertex set V. It is not hard to see that the set of such functions can be represented by the set $[V]\binom{n}{2}$.

However, to state the Graham–Rothschild Theorem, we first need some additional definitions.

Definition. Let A be a set, and G a group acting on A. For an integer $m \geq 0$, take m elements $\lambda_0, \ldots, \lambda_{m-1}$ to be distinct elements not in A. A *generalised m–parameter word* is a function

$$f : \{0, \ldots, n-1\} \to A \cup (G \times \{\lambda_0, \ldots, \lambda_{m-1}\})$$

such that the following properties hold:

1. For each $0 \leq i < m$, the element (e, λ_i) appears in f [where e is the identity element on G];
2. If $i < j$, then the first occurrence of (g, λ_i) appears **before** the first occurrence of (g', λ_j), for any group elements $g, g' \in G$.

This can be seen as a slight generalisation of the n–parameter word we have seen before; we essentially allow each λ_i to be "indexed" by a group element $g \in G$. The same definitions for concatenation of these generalised words still hold.

Of course, the content of this generalisation lies on what happens when we compose two such parameter words together.

Definition. Given $f \in [A, G]\binom{n}{m}$ and $g \in [A, G]\binom{m}{k}$, we define the *composition* $f \circ g \in [A, G]\binom{n}{k}$ by:

$$(f \circ g)_i \stackrel{\text{def}}{=} \begin{cases} f_i & \text{if } f_i \in A; \\ a \cdot b & \text{if } f_i = (a, \lambda_j) \text{ and } g_j = b \in A; \\ (a \cdot b, \lambda_\ell) & \text{if } f_i = (a, \lambda_j) \text{ and } g_j = (b, \lambda_\ell). \end{cases}$$

What does this mean? The way to think about the composition is to extend the group action to the larger set $A \cup (\cup_{i<m} \{g \cdot \lambda_i : g \in G\})$ by left multiplication. When viewed in this way, the usual "composition" analogy (of plugging things into the λ_i) still holds.

Hence, we write $g \cdot \lambda \stackrel{\text{def}}{=} (g, \lambda)$ for any $g \in G$ and $\lambda \in \{\lambda_0, \ldots\}$ for notational convenience.

Example. Let $G \stackrel{\text{def}}{=} C_4^\times = \{\pm 1, \pm i\}$ and $A \stackrel{\text{def}}{=} \{a, b, c, x, y\}$. We have the following composition:

$$\begin{array}{rcccccccl} f & \stackrel{\text{def}}{=} & a & 1 \cdot \lambda_0 & b & i \cdot \lambda_1 & c & \in & [A, G]\binom{5}{2} \\ g & \stackrel{\text{def}}{=} & & x & & 1 \cdot \lambda_0 & & \in & [A, G]\binom{2}{1} \\ \implies f \circ g & = & a & x & b & i \cdot \lambda_0 & c & \in & [A, G]\binom{5}{1} \end{array}$$

With these definitions, it is now possible to state the Graham–Rothschild Theorem.

Theorem 4.1 [Graham–Rothschild Theorem]. Let $n, d, c \geq 1$ and $k \geq 0$ be integers, A a set of size n, and a finite group G acting on A.

There exists $n \stackrel{\text{def}}{=} GR(n, |G|, k, d, c)$ such that for any colouring $\Delta : [A, G]\binom{n}{k} \to \{0, \ldots, c-1\}$, there exists $f \in [A, G]\binom{n}{d}$, such that $\Delta \circ f$ is constant on $[A, G]\binom{d}{k}$.

Proof. We proceed by induction on k.

For $k = 0$, we can ignore the group G as follows. Let $H \stackrel{\text{def}}{=} H(n, d, c)$, and $\Delta' : A^H \to \{0, \ldots, c-1\}$ be the restriction of Δ onto A^H. By Hales–Jewett, we can find $f \in [A]\binom{H}{d}$ for which $\Delta' \circ f \equiv \text{const}$ on A^d. Then, simply view f as an element of $[A, G]\binom{H}{d}$ by replacing each λ_i in f with $e \cdot \lambda_i$.

It follows that $\Delta \circ f$ is constant on $[A, G]\binom{d}{0} \equiv A^d$, and hence f is the required monochromatic combinatorial d–space. Hence, $GR(n, |G|, 0, d, c) \leq H(n, d, c)$.

Now, assume for some $k > 0$, $GR(n, |G|, k-1, d, c)$ is finite for all $n, d, c \geq 1$ and finite group G. Let $x \stackrel{\text{def}}{=} HJ^\star(n, d, c)$ and $n_0, \ldots, n_x \in \mathbb{N}$ by $n_x = x + k$ and

$$n_j \stackrel{\text{def}}{=} (j+1) + GR(n + |G|, |G|, k-1, n_{j+1} - (j+1), c^{n^j}).$$

We claim that $GR(n, |G|, k, d, c) \leq n \stackrel{\text{def}}{=} n_0$.

To prove this, we look at the similarities between this theorem and the star version of Hales–Jewett. We can see there is a natural map from elements in $[A, G]\binom{n}{k} \to [A]^\star\binom{n}{0}$, that "ignores" all of the parameters; i.e. replacing all the $g \cdot \lambda_i$'s (and everything after them) with stars. This motivates the following definition.

Definition. Given a word $f \in [A, G]\binom{n}{d}$, define its *prefix* $p(f)$ as the longest word $f_0 \ldots f_{\ell-1}$ which does not contain any parameters $g \cdot \lambda_i$'s (for any $g \in G$, $0 \leq i < d$). The *prefix length* is defined as $\ell(f) \stackrel{\text{def}}{=} \ell$.

Naturally, we would like to try following similar steps to the proof in the star version.

Lemma 4.2. For each $0 \leq j \leq x$, there exists $f^j \in [A, G]\binom{n}{n_j}$ such that for any $g \in [A, G]\binom{n_j}{k}$ with prefix length $\ell(g) = \ell < j$, the composition $(\Delta \circ f^j)(g)$ depends only on $p(\ell)$.

Notice the similarities with Lemma 3.6 – it allows us to combine the properties of each of the separate "classes" (separated by length) in the sets. However, this lemma is notably weaker than Lemma 3.6, as we get $\Delta \circ f^j$ depends on the prefix, not just its length.

Of course, the real content of this lemma is when we use it for $j = x$. This produces an n_x–parameter word $F \stackrel{\text{def}}{=} f^{n_x} \in [A, G]\binom{n}{n_x}$, for which $(\Delta \circ F)(g)$ only depends on $p(g)$ (no restriction on the size of $\ell(g)$).

Proof of Lemma 4.2. We use induction on j. For $j = 0$, as long as $n_0 \geq k$ we can choose an arbitrary $f^0 \in [A, G]\binom{n_0}{n_0}$, which must work as there is no $g \in [A, G]\binom{n_0}{k}$ with $\ell(g) < 0$, leaving nothing to check.

Assume the lemma is true for some $0 \leq j < x$, and we want to prove it for $j + 1$. That is, there exists a word $f^j \in [A, G]\binom{n}{n_j}$ which satisfies the given properties.

In particular, we want to look at mappings of the form $f^{j+1} \stackrel{\text{def}}{=} f^j(T)$ for a word T of the shape $T \stackrel{\text{def}}{=} [(e \cdot \lambda_0) \ldots (e \cdot \lambda_j)]\, T'$, since by the properties of f^j, this already satisfies the given properties for $g \in [A, G]\binom{n_{j+1}}{k}$ with prefix length $\ell(g) < j$. This allows us to focus on just those g with $\ell(g) = j$.

Notice that, for generic variables, we can view the set $[A, G]\binom{N}{k}$ as $[A \cup G \times \{\lambda_0\}, G]\binom{N}{k-1}$ by simply declaring any element $g \cdot \lambda_0 \in G \times \{\lambda_0\}$ to be an element of A. Of course, this removes one parameter from the set.

This, together with the earlier observation, motivates us to consider the following induced colouring:

$$\begin{aligned} \Delta^j : [A \cup G \times \{\lambda_0\}, G]\binom{n_j - (j+1)}{k-1} &\to \{0, \ldots, c-1\}^{n^j} \\ g &\mapsto \left(\Delta \circ f^j(\mathbf{a} + e \cdot \lambda_0 + g) : \mathbf{a} \in A^j\right). \end{aligned}$$

which is well-defined because the total length of $\mathbf{a}, \{e \cdot \lambda_0\}, g$ is n_j. Then, as long as we define

$$n_j - (j+1) = GR(n + |G|, |G|, k-1, n_{j+1} - (j+1), c^{n^j}),$$

we can find $f' \in [A \cup G \times \{\lambda_0\}, G]\binom{n_j-(j+1)}{n_{j+1}-(j+1)}$ for which $\Delta^j \circ f'$ is monochromatic on $[A, G]\binom{n_{j+1}-(j+1)}{k}$. That is, there exists a constant $t \in \{0, \dots, c-1\}^{n^j}$ (indexed by subsets of A^j) for which

$$\forall \mathbf{a} = (a_0, \dots, a_{j-1}) \in A^j,\ \forall g \in [A, G]\binom{n_{j+1}-(j+1)}{k} \implies \Delta \circ f^j(\mathbf{a} + e \cdot \lambda_0 + f' \circ g)) = t_{\mathbf{a}}. \qquad (\spadesuit)$$

We will use f' to reconstruct a word $T \in [A, G]\binom{n_j}{n_{j+1}}$ by the following:

$$T \stackrel{\text{def}}{=} [(e \cdot \lambda_0) \ \dots \ (e \cdot \lambda_j)] + f'.$$

Then, for any $g \in [A, G]\binom{n_{j+1}}{k}$ with length $\ell(g) = j$, say $g = g_0\, g_1 \dots g_{j-1}\, (e \cdot \lambda_0)\, g'$, the composition is

$$T \circ g = [g_0\, g_1 \ \dots \ g_{j-1}\, (e \cdot \lambda_0)] + f''$$

for some other word f'', and elements $g_0, \dots, g_{j-1} \in A$. Hence the value of $\Delta \circ f^j(T \circ g) = t_{g_0 \dots g_{j-1}}$ only depends on $g_0, \dots, g_{j-1}$ by ($\spadesuit$).

Hence, $f^{j+1} \stackrel{\text{def}}{=} f^j \circ T$ satisfies the requirements of the induction. $\square$

We return to the proof of the main theorem.

As discussed before, we have found a map $F \stackrel{\text{def}}{=} f^x \in [A, G]\binom{n}{n_x}$, for which $(\Delta \circ F)(g)$ depends only on the prefix $p(g)$, for $g \in [A, G]\binom{n_x}{k}$.

Explicitly, there exists a function $C : \{p(f) :\ f \in [A, G]\binom{n_x}{k}\} \to \{0, \dots, c-1\}$ such that $(\Delta \circ F)(g) = C(p(g))$ for any $g \in [A, G]\binom{n_x}{k}$.

This induces a colouring

$$\begin{aligned} \Delta' : [A, G]\tbinom{n_x}{k} &\to \{0, \dots, c-1\} \\ g &\mapsto (\Delta \circ F)g \equiv C(p(g)). \end{aligned}$$

However, notice that the map

$$\begin{aligned} \pi : \{p(f) :\ f \in [A, G]\tbinom{n_x}{k}\} &\to [A]^\star\tbinom{n_x-k}{0} \\ p(f) &\mapsto p(f) + \star + \dots + \star \end{aligned}$$

that adds enough stars to $p(f)$ to get the correct length, is actually a bijection – it's clearly injective, and any element in $[A]^\star\binom{n_x-k}{0}$ is in the image of π by removing all the stars and adding λ_i's appropriately to the end.

Hence, by choosing $n_x \stackrel{\text{def}}{=} x + k$, we can induce a colouring

$$\begin{aligned} \Delta_0' : [A]^\star\tbinom{x}{0} &\to \{0, \dots, c-1\} \\ g &\mapsto \Delta'(\pi(g')) \end{aligned}$$

where $g' \in [A, G]\binom{x+k}{k}$ is any word whose prefix is equal to g (modulo some stars at the end).

At this point, we have a colouring Δ_0' of $[A]^\star\binom{x}{0}$. This motivates us to define $x \stackrel{\text{def}}{=} H^\star(n, d, c)$; hence, by definition, there exists a map $P \in [A]^\star\binom{x}{d}$, for which $\Delta_0' \circ P = \text{const}$ in $[A]^\star\binom{d}{0}$.

Finally, we need to produce a map $F^0 \in [A, G]\binom{x+k}{d}$ using P. We can do this as follows: fix an element $a \in A$, and define F^0 by:

$$F_i^0 = \begin{cases} P_i & \text{if } P_i \in A \\ e \cdot \lambda_j & \text{if } \exists j :\ P_i = \lambda_j \\ a & \text{if } P_i = \star \text{ or } i \geq x, \end{cases}$$

that is, F^0 is the result of "embedding" P into $[A]^\star\binom{x+k}{k}$ by turning λ_i's into $e \cdot \lambda_i$ and filling the end with random elements $a \in A$ to match the length.

We claim that for any $g \in [A,G]\binom{d}{k}$ and $g' \in [A]^\star\binom{d}{0}$ such that $p(g) = g'$ (up to adding stars), we have $P \circ g'$ and $F^0 \circ g$ have the same prefix string. The easiest way to see this is to just write out the strings: define

$$
\begin{aligned}
g &= g_0\, g_1\, \ldots\, g_{k-1}\, e \cdot \lambda_0\, \ldots \\
g' &= g_0\, g_1\, \ldots\, g_{k-1}\, \star \star \cdots \\
P &= a_0 \ldots a_{\ell_1 - 1}\, \lambda_0\, a_{\ell_1+1} \ldots a_{\ell_2-1}\, \lambda_1\, \cdots a_{\ell_{k+1}-1}\, \lambda_k\, a_{\ell_{k+1}+1}\, \ldots .
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
P \circ g' &= a_0 \ldots a_{\ell_1-1}\, g_0\, a_{\ell_1+1} \ldots a_{\ell_2-1}\, g_1\, \cdots a_{\ell_{k+1}-1}\, \lambda_0\, a_{\ell_{k+1}+1}\, \ldots \\
F^0 \circ g &= a_0 \ldots a_{\ell_1-1}\, g_0\, a_{\ell_1+1} \ldots a_{\ell_2-1}\, g_1\, \cdots a_{\ell_{k+1}-1} \star \star \star \ldots
\end{aligned}
$$

so it is indeed clear that the prefix strings are the same. Hence, by the definition of Δ_0', we get:

$$\text{const } = \Delta_0' \circ P(g') = \Delta' \circ F^0(g) = (\Delta' \circ F^0)g$$

which implies $\Delta' \circ F^0 \stackrel{\text{def}}{=} \Delta \circ F \circ F^0$ is constant in $[A,G]\binom{d}{k}$.

Hence, we may choose $f \stackrel{\text{def}}{=} F \circ F^0 \in [A,G]\binom{n}{d}$, which satisfies all conditions of the theorem. □

Remark. The structure of the proof, as shown above, comes down to the following steps:

- Finding an appropriate disjoint partition $\{S_{\mathbf{a}} : \mathbf{a} \in I\}$ of $[A,G]\binom{x}{k}$ [for some set I];
- Finding $F \in [A,G]\binom{n}{x}$, such that $\Delta \circ F$ is constant on each $S_{\mathbf{a}}$ ($\mathbf{a} \in I$);
  - This induces a colouring $\Delta' : I \to \{0, \ldots, c-1\}$ by $\Delta'(\mathbf{a}) \stackrel{\text{def}}{=} \Delta \circ F(g)$ for any $g \in S_{\mathbf{a}}$.
- Finding $F^0 \in [A,G]\binom{x}{d}$, such that $\Delta' \circ F^0$ is constant on $[A,G]\binom{d}{k}$.

Then, by construction, $f \stackrel{\text{def}}{=} F \circ F^0$ satisfies the inductive step.

This general proof structure appears to be present in many of the proofs we have presented above. We first consider a known theorem or an inductive hypothesis P, which finds you a monochromatic substructure on a subset of the desired space. We then inductively apply the theorem P to get an equivalence class of monochromatic substructures, which can be indexed nicely by some set I. Finally, apply known theorems to the set I itself to finish the proof.

Now that we've finally proved this result, we can find some easy corollaries. In particular, we present a proof of Ramsey's Theorem.

Proof of Ramsey's Theorem. Let $\binom{n}{k}$ be defined as the set of strictly increasing mappings from $\{0, \ldots, k-1\} \to \{0, \ldots, m-1\}$, as before. There is a map $\Phi : [A]\binom{m}{k} \to \binom{m}{k}$, which assigns $(\Phi \circ f)_i$ as the minimum coordinate for which λ_i occurs.

Given any colouring of the edges $\Delta : \binom{n}{2} \to \{0, \ldots, c-1\}$, we can extend this to a colouring

$$
\begin{aligned}
\Delta_\Phi : [A]\tbinom{n}{2} &\to \{0, \ldots, c-1\} \\
f &\mapsto \Delta(\Phi(f)).
\end{aligned}
$$

By the Graham–Rothschild Theorem, there exists n for which there is some $f \in [A]\binom{n}{d}$ such that $\Delta_\Phi \circ f$ is constant on $[A]\binom{d}{2}$. This means that $\Phi \circ f$ is monochromatic under Δ.

Notice f has d parameters; let p_i denote the position of first occurrence of λ_i for $0 \le i < d$. It is easy to see that for $g \in \binom{n}{2}$, the image of $\Phi \circ (f \circ g) \in \binom{n}{2}$ is of the form $(0,1) \mapsto (p_i, p_j)$ for some $i < j$; furthermore, all such maps are attained. It follows that the subgraph generated by $\{p_0, \ldots, p_{d-1}\}$ is a monochromatic K_d. □

5 Further directions of research

In this section, we briefly review further directions of research on these theorems, and present important results without proof.

5.1 A Density Version

In all of the theorems we proved, we proved statements of the form:

If we colour all the points of some combinatorial structure X in one of c colours, then there exists a monochromatic substructure $Y \subseteq X$ with some desired property.

An obvious potential generalisation would be: instead of colouring all the points of X in a certain number of colours, we could instead pick some *portion* of points in X. These are called *density versions*.

Theorem 5.1 [Hales–Jewett, density version]. Let $n \geq 1$ be an integer and $0 < \delta < 1$. There exists $H = H(n, \delta)$ such that any subset $B \subseteq A^H$ with $|B| \geq \delta \cdot n^H$ contains n collinear lattice points all of the same colour.

Clearly, Theorem 5.1 implies the regular version, because if we choose $\delta < 1/c$ then at least one of the monochromatic sets have density at least δ.

In a sense, this theorem says that there isn't anything special about having to partition the set of points in the hypercube; we just need to choose enough points to obtain the result.

Following a similar proof to Theorem 3.3 deduces the famous Szemerédi's Theorem, the density version of Van der Waerden.

Theorem 5.2 [Szemerédi's Theorem]. Let $n \geq 1$ be an integer, and $0 < \delta < 1$. There exists $S = S(n, \delta)$, such that any subset of $\{1, \ldots, S\}$ of size $\geq \delta \cdot S$ contains an arithmetic progression of length n.

5.2 A Polynomial Version

Another possible direction of the generalisation of Van der Waerden's Theorem is to consider *polynomials* instead of linear differences.

Theorem 5.3 [Van der Waerden's Theorem, Polynomial Version].

Let $c, K, L \geq 1$ be natural numbers. Suppose $P : \mathbb{Z}^K \to \mathbb{Z}^L$ be a polynomial mapping such that $P(0) = 0$. For any finite set $F \subset \mathbb{Z}^K$ and any colouring $\chi : \mathbb{Z}^L \to \{0, \ldots, c-1\}$, there exists $\mathbf{a} \in \mathbb{Z}^L$ and an integer $d \neq 0$, such that the following set is monochromatic:

$$\mathbf{a} + P(n \cdot F) \stackrel{\text{def}}{=} \{\mathbf{a} + P(d \cdot \mathbf{f}) : \mathbf{f} \in F\}.$$

Note that the usual Van der Waerden's Theorem follows from $K = L = 1$, $F \stackrel{\text{def}}{=} \{1, \ldots, n\}$ and $P(X) = X$. However, there are more interesting corollaries; for any $n \geq 1$, we can find a, d such that there exists monochromatic configurations of the form

$$\{a, a + d, a + d^2, \ldots, a + d^n\}.$$

It is completely not obvious how any of our methods will generalise to prove this result (or even this special case), and we will not prove this here.

Perhaps more interesting is how we could try generalising this theorem to the setting of Hales–Jewett. It turns out that while we *can* give a generalisation, we must first give a new formulation.

Theorem 3.1' [Hales–Jewett Theorem]. Given a set S, denote $\mathcal{F}(S)$ the collection of finite subsets of S.

For any integers $n, c \geq 1$, there exists $H' = H'(n, c)$ such that for any set $|S| \geq H'$ and any colouring $\chi : \mathcal{F}(S)^n \to \{0, \ldots, c-1\}$, there exists $\emptyset \neq \gamma \in \mathcal{F}(S)$ and $\alpha_1, \ldots, \alpha_n \in \mathcal{F}(S)$ such that:

- $\gamma \cap \alpha_1 = \cdots = \gamma \cap \alpha_n = 0$;
- The sets $\{(\alpha_1, \ldots, \alpha_n)\} \cup \{(\alpha_1, \ldots,\ \alpha_i \cup \gamma,\ \ldots, \alpha_n) : 1 \leq i \leq n\}$ all have the same colour.

Of course, the first question is: why is this statement equivalent to the original? Suppose we set $S \stackrel{\text{def}}{=} \{A_1, \ldots, A_N\}$. It turns out there is a natural bijection

$$\begin{aligned} \Phi : \{0, \ldots, n-1\}^{|S|} &\to \mathcal{F}'(S)^n \\ (a_1, \ldots, a_{|S|}) &\mapsto (\{A_i : j = a_i\} : 1 \leq j \leq n) \end{aligned}$$

where $\mathcal{F}'(S)^n \subset \mathcal{F}(S)^n$ is the set of subsets whose total lengths of its constituent sets is $\leq |S|$. It is not too hard to see from here that the two theorems are equivalent; depending on the value of n, we can find isomorphic copies of one colouring in the other.

Finally, we can state the polynomial generalisation of this theorem, which occurs naturally as an extension to a higher dimensional space.

Theorem 5.4 [Hales–Jewett Theorem, Polynomial Version].

For integers $n, d, c \geq 1$, there exists $H_0 = H_0(n, d, c)$ such that for any set $|S| \geq H_0$ and any colouring $\chi : \mathcal{F}(S^d)^n \to \{0, \ldots, c-1\}$, there exists $\emptyset \neq \gamma \in \mathcal{F}(S)$ and $\alpha_1, \ldots, \alpha_n \in \mathcal{F}(S^d)$ such that:

- $\gamma^d \cap \alpha_1 = \cdots = \gamma^d \cap \alpha_n = 0$;
- The sets $\{(\alpha_1, \ldots, \alpha_n)\} \cup \{(\alpha_1, \ldots,\ \alpha_i \cup \gamma^d,\ \ldots, \alpha_n) : 1 \leq i \leq n\}$ all have the same colour.

Here, $\gamma^d \stackrel{\text{def}}{=} (\gamma, \ldots, \gamma)$, and all unions and intersections are done component–wise.

In a sense, this is called a "polynomial" generalisation because the terms that appear in the proof are *set polynomials* and can roughly be understood as a polynomial whose coefficients form a set. We won't delve into the details as to what this means.

Finally, we will give an example application of the polynomial version.

Corollary 5.5. Let $d \geq 1$ be an integer. For any colouring of $\mathbb{Z}$ with c colours, there exists an arbitrarily long monochromatic arithmetic progression whose difference is of the form n^d.

Proof. Let $\chi : \mathbb{Z} \to \{0, \ldots, c-1\}$ be the colouring. Introduce another colouring

$$\begin{aligned} \chi' : \mathcal{F}(S^d) &\to \{0, \ldots, c-1\} \\ \chi'(a) &\stackrel{\text{def}}{=} \chi(|a_1| + 2 \cdot |a_2| + \cdots + n \cdot |a_n|). \end{aligned}$$

Then, by Theorem 5.4 we can find $a_1, \ldots, a_n$ and γ as defined in the theorem.

Then, it is easy to observe that

$$\chi'(a_1, \ldots, a_i \cup \gamma^d, \ldots, a_n) = \chi(|a_1| + \cdots + i \cdot |a_i \cup \gamma^d| + \cdots + n \cdot |a_n|) = \chi(T + i \cdot |\gamma|^d)$$

where $T \stackrel{\text{def}}{=} |a_1| + \cdots + n \cdot |a_n|$. It follows that $\{T + i \cdot |\gamma|^d : 1 \leq i \leq n\}$ is the monochromatic arithmetic sequence of common difference a d–th power. $\square$

References

[1] Imre Leader, *Ramsey Theory*, University of Cambridge, 2010.

[2] Nilli, A. (1990). Shelah's Proof of the Hales-Jewett Theorem. In: Nešetřil, J., Rödl, V. (eds) Mathematics of Ramsey Theory. Algorithms and Combinatorics, vol 5. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-72905-8_10

[3] Prömel, H.J., Voigt, B. (1990). Graham-Rothschild Parameter Sets. In: Nešetřil, J., Rödl, V. (eds) Mathematics of Ramsey Theory. Algorithms and Combinatorics, vol 5. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-72905-8_9

[4] Bergelson, V., Leibman, A. (1999) Set-polynomials and polynomial extension of the Hales-Jewett Theorem. https://arxiv.org/abs/math/9907201

Acknowledgements

I would like to thank Professor Imre Leader for providing me with the materials and emotional support to write this mathematical essay.