COMBINATORIAL PROOFS OF THE POLYNOMIAL VAN DER WAERDEN THEOREM AND THE POLYNOMIAL HALES–JEWETT THEOREM

MARK WALTERS

Abstract

Recently Bergelson and Leibman proved an extension to van der Waerden's theorem involving polynomials. They also generalised the Hales–Jewett theorem in a similar way. The paper presents short and purely combinatorial proofs of those results.

1. *Introduction*

One of the best known results in Ramsey theory is van der Waerden's theorem [**6**].

Van der Waerden theorem. *If $\mathbb{N}$ is finitely coloured then there exist monochromatic arithmetic progressions of arbitrary length.*

Recently Bergelson and Leibman [**1**] proved a remarkable extension of this theorem.

Polynomial van der Waerden theorem. *Suppose that $p_1, p_2, \ldots, p_m$ are polynomials with integer coefficients and no constant term. Then whenever $\mathbb{N}$ is finitely coloured there exist natural numbers a and d such that the point a and all the points $a+p_i(d)$, for $1 \leqslant i \leqslant m$, have the same colour.*

The case with a single polynomial was proved by Furstenberg [**3**] and Sarkozy [**5**] (independently). It is easy to see that this generalises van der Waerden's theorem, which we can recover by setting $p_i(x) = ix$ for $1 \leqslant i \leqslant m$.

Bergelson and Leibman also proved a stronger result. They showed that if $A \subset \mathbb{N}$ has positive upper density, then there exist a and d such that all the points a and $a+p_i(d)$ are contained in A. This immediately implies the polynomial van der Waerden theorem (trivially, at least one colour class must have positive upper density). However, the deduction of this result from the polynomial van der Waerden theorem used similar techniques to those used by Furstenberg in his proof of Szemerédi's theorem (see [**3**]) and it turned out that the main difficulty lay in proving the polynomial van der Waerden theorem. They proved the polynomial van der Waerden theorem using methods from ergodic theory. We prove the same result using a combinatorial technique which was used in the original proof of the van der Waerden theorem, and which is known as *colour focusing*.

Received 5 July 1998.

2000 *Mathematics Subject Classification* 05A18.

Hales and Jewett [4] proved an abstract generalisation of the van der Waerden theorem. They defined a notion of combinatorial space and proved that whenever a 'large-enough'-dimensional combinatorial space is k-coloured there exists a one-dimensional monochromatic subspace. Very recently Bergelson and Leibman [2] generalised this theorem by putting some strong constraints on the structure of the subspace and showed that this new theorem implies the polynomial van der Waerden theorem. Precise statements of these results appear later in this paper. Their proof is significantly harder than that of the polynomial van der Waerden theorem. We shall prove this using colour focusing again. Our proof is much shorter and significantly simpler than that in [2]. Moreover, it directly generalises our proof of the polynomial van der Waerden theorem.

2. *Polynomial van der Waerden theorem*

In this section we present the proof of a special case of the polynomial van der Waerden theorem. We use many of the same ideas as in [1] but in the context of colour focusing, which allows us to avoid the formalism of topological dynamics.

Define an *integral polynomial* to be a polynomial with integer coefficients taking the value zero at zero. For completeness we recall the polynomial van der Waerden theorem from the introduction (stated slightly differently to simplify the proof).

POLYNOMIAL VAN DER WAERDEN THEOREM. *Suppose that $p_1, p_2, \ldots, p_m$ are integral polynomials and that $\mathbb{N}$ is finitely coloured. Then there exist natural numbers a and d such that the set of points $\{a\} \cup \{a+p_i(d) : 1 \leqslant i \leqslant m\}$ is monochromatic.*

(For the remainder of this section we shall implicitly assume that a and d are natural numbers.)

If we let $p_i(n) = in$, $1 \leqslant i < l-1$, then the theorem guarantees that there exist a, d such that the set $\{a, a+d, a+2d, \ldots, a+(l-1)d\}$ is monochromatic, that is, we have recovered the 'normal' van der Waerden theorem. Another example is when we set $p_1(n) = n^2$ as the only polynomial. Then the theorem guarantees the existence of points a and $a+d^2$ with the same colour. It is a little surprising that even this result is not trivial.

The restriction that the polynomials have integer coefficients is unnecessary; the result is also true if the coefficients are rational. This is easy to deduce from the theorem stated above. Indeed suppose that $p_1, p_2, \ldots, p_m$ are polynomials with rational coefficients, and take the value zero at zero. Then define $p_i'(n) = p_i(cn)$, for $1 \leqslant i \leqslant m$, where c is the least common multiple of the denominators of the coefficients of all the polynomials. These polynomials have integer coefficients (the coefficient of the term n^a is multiplied by c^a, so is integral) and take the value zero at zero. Thus we can apply the theorem to these polynomials to obtain a, d with $\{a\} \cup \{a+p_i'(d) : 1 \leqslant i \leqslant m\}$ monochromatic. Now observe that this set is the set $\{a\} \cup \{a+p_i(cd) : 1 \leqslant i \leqslant m\}$. Thus the conclusion of the theorem holds with a and cd. A typical example of where this could occur is for the polynomial $\frac{1}{2}n(n+1)$.

Also observe that we gain no generality by insisting that a has the same colour as the points $a+p_i(d)$. The form in which we have stated it is more convenient for our proof.

We will prove a special case first: we show that there must exist a and $a+d^2$ the same colour.

THEOREM 1. *If $\mathbb{N}$ is finitely coloured then there exist a and d such that a and $a+d^2$ have the same colour.*

Proof. Say that a point a_i is *focused at a* if $a_i - a = d_i^2$ for some d_i. Define the points $a_1, a_2, \ldots, a_r$ to be *colour focused at a* if each a_i is focused at a and the a_i have distinct colours. Thus this is exactly the condition needed to force a to have a colour distinct from any of the a_i if there do not exist a and $a+d^2$ the same colour. Observe that this definition is very similar to the one used in the proof of the 'linear' van der Waerden theorem.

INDUCTION HYPOTHESIS. For all $r \leqslant k$ there exists N such that if $[N]$ is k-coloured either there exist r colour focused points $a_1, a_2, \ldots, a_r$ together with their focus a or there exist a and d such that a and $a+d^2$ have the same colour.

The proof is immediate once we have established this result. Indeed, set $r = k$. If the latter case occurs then we are done, and if the former case occurs then the focus a must have the same colour as one of the points a_i, in which case a and a_i are the required points.

The induction is on r. Suppose that N satisfies the hypothesis for $r-1$. We shall show that there is N' satisfying the hypothesis for r. We do not compute the size explicitly; we may simply assume that it is large enough for us to find everything that we need. Without loss of generality we assume that $[N']$ does not contain a and $a+d^2$ the same colour.

Divide $[N']$ into blocks of length N, let $B_s = \{(s-1)N+1, (s-1)N+2, \ldots, sN\}$ and set $l = \lfloor 2\sqrt{N} \rfloor$. Then, since there are only k^N ways of colouring a block, we can apply the linear van der Waerden theorem to the blocks and find an 'arithmetic progression' of blocks $B_s, B_{s+t}, \ldots, B_{s+lt}$ with the blocks all coloured identically. By choice of N there exist $r-1$ colour focused points $a_1, a_2, \ldots, a_{r-1}$ in the block B_s together with their focus a. Say $a_i - a = d_i^2$. As noted above, the colours of $a, a_1, a_2, \ldots, a_{r-1}$ are all distinct.

CLAIM 1. *The points $a_i + 2d_i Nt$ for $1 \leqslant i \leqslant r-1$ together with a are colour focused at $a-(Nt)^2$.*

The focus $a-(Nt)^2$ may be negative; this will not matter: see below.

Proof of Claim 1. First note that

$$a_i + 2d_i Nt - (a-(Nt)^2) = d_i^2 + 2d_i Nt + (Nt)^2 = (d_i + Nt)^2$$

and

$$a - (a-(Nt)^2) = (Nt)^2.$$

Thus the points are focused at $a-(Nt)^2$.

Secondly, observe that the point $a_i + 2d_i Nt$ has the same colour as a_i; since $a_i - a = d_i^2 \leqslant N$ we have $2d_i \leqslant l$ and thus that the block B_{s+2d_i} is coloured identically to the block B_s. Thus the points $a_i + 2d_i Nt$ have distinct colours. Also the point a has a colour distinct from all of these since otherwise the points a and a_i would be the required set. Thus the r points $a_i + 2d_i Nt$ and a are colour focused at $a-(Nt)^2$. $\square$

To finish the proof we remark that $a-(Nt)^2$ might be negative, but that this does not matter. We can bound $a-(Nt)^2$ below, say by $-M$. Since all the properties we have used are translation-invariant, we can do the above construction on blocks $B'_s = M+B_s$ and thus ensure that $a-(Nt)^2$ is positive. □

2.1. *Remarks on the proof of Theorem* 1

Suppose that $d_1, d_2, \ldots, d_{r-1}$ are the $r-1$ 'difference' values in the colour focusing in the block B_s. All that we require in the above proof is that the blocks $B_{s+2d_i t}$ be identically coloured. However, we do not know in advance what the values of the d_i are going to be. On the other hand, we do know that there is a bound on their size, namely $2\sqrt{N}$. Thus we make sure that all possible blocks $B+s+2d_i t$ (over all possible d_i) are identically coloured. These blocks essentially form an arithmetic progression of length $2\sqrt{N}$ (in fact, an arithmetic progression of length $\sqrt{N}$ with common difference a multiple of 2).

The second thing to observe is that we reduce from looking at a quadratic polynomial to a large number of linear ones. These are obtained by taking the differences between the polynomial ($p(n) = n^2$) and 'shifts' of it ($p(n+m)$ for some value m). We already know how to find a and d satisfying the theorem for this new collection of polynomials, and we use this to move up a step in the n^2 case.

3. *The general result*

We use a slightly complicated induction scheme (this is very similar to the one used in [**1**]). We need an ordering on collections of polynomials such that when we form a (generalised) collection of 'difference' polynomials, we already know that we can find a monochromatic set for this new collection.

Let $A = \{p_1, p_2, \ldots, p_m\}$ be a set of integral polynomials. Let D be the maximum degree of these polynomials. For $1 \leqslant i \leqslant D$ let N_i be the number of distinct leading coefficients of the polynomials in A of degree i. Define the *weight vector of A* to be $(N_1, N_2, \ldots, N_D)$. We say that $(N_1, N_2, \ldots, N_D) < (M_1, M_2, \ldots, M_{D'})$ if there exists r such that $N_r < M_r$ and $N_i = M_i$ for $i > r$. This is easily seen to be a well ordering. The induction will be on collections of polynomials in this ordering. We now try to give some intuition about this ordering. Suppose that all the polynomials have the same degree and the same leading coefficient. Then, if we take differences between one polynomial and shifts of another (possibly the same) polynomial, then we always obtain a polynomial of lower degree. This gives some indication that it is the number of distinct leading coefficients that is important, rather than the number of polynomials.

Proof of the polynomial van der Waerden theorem. We do a multiple induction. Let $P = \{p_1, p_2, \ldots, p_m\}$. As mentioned, the induction is on the weight vector of P. However, we need the theorem in its 'compact' form.

Outer induction hypothesis. Suppose that $p_1, p_2, \ldots, p_m$ are integral polynomials. Then for all k there exists N such that whenever $[N]$ is k-coloured there exist a and d such that the set $\{a, a+p_1(d), a+p_2(d), \ldots, a+p_m(d)\}$ is monochromatic.

Suppose that the outer induction hypothesis is true for all collections of polynomials with weight vector less than that of P. We need to show that it is true for P.

We will say that a set A of points a_j for $1 \leqslant j \leqslant m$ is *focused at a* if there exists d such that $a_j - a = p_j(d)$. We will say that sets $A_1, A_2, \ldots, A_r$ are *colour focused at a* if each set A_i is focused at a, each set A_i is monochromatic, and the sets A_i for $1 \leqslant i \leqslant r$ have distinct colours. This is exactly the condition that forces a to have a colour different from that of any of the A_i.

Inner induction hypothesis. For all $r \leqslant k$ there exist N such that if $[N]$ is k-coloured either there exist r colour focused sets $A_1, A_2, \ldots, A_r$ together with their focus a, or there exist a and d such that the set $\{a, a+p_1(d), a+p_2(d), \ldots, a+p_m(d)\}$ is monochromatic.

As before, from this hypothesis the result is immediate. Set $r = k$. If there exist a and d such that the set $\{a, a+p_1(d), a+p_2(d), \ldots, a+p_m(d)\}$ is monochromatic then we are done. Otherwise, there exist $A_1, A_2, \ldots, A_r$ colour focused at a. The focus must have the same colour as one of the sets A_i and then the set $\{a\} \cup A_i$ is the set we require.

The induction is on r. Suppose that the hypothesis is true for $r-1$ and N. We show that there is N' satisfying the hypothesis for r. As before, we do not calculate the bound explicitly (although this is possible). Also, as before, we may assume that there do not exist a and d with the set $\{a\} \cup \{a+p_j(d) : 1 \leqslant j \leqslant m\}$ monochromatic.

Let $d_{\max}$ be the maximum possible d for which points $a, a_1, a_2, \ldots, a_k \in [N]$ exist with $a_i - a = p_i(d)$. (Since all polynomials tend to infinity, this exists.) (We cannot necessarily use $d_{\max} = N$ since the polynomials could all be zero simultaneously for some $d > N$. However, this case is clearly trivial; if we take this d and any a then the set of points $\{a\} \cup \{a+p_i(d) : 1 \leqslant m\} = \{a\}$ is monochromatic. Moreover, if this does not occur then the bound can be replaced by N. Indeed, since the value of each polynomial is a multiple of its argument, if $d > N$ and not all the polynomials are zero at d, then at least one of them is larger than N.)

We may assume that p_1 has minimal degree amongst the p_i. Now define polynomials

$$p_{i,j}(n) = p_j(n+i) - p_j(i) - p_1(n) \quad 0 \leqslant i \leqslant d_{\max}, 1 \leqslant j \leqslant m.$$

(These polynomials were considered in [**1**].) Observe that these are integral polynomials. (The constant $p_j(i)$ cancels the constant term in $p_j(n+i)$ and they obviously have integer coefficients.) This collection is smaller in the ordering defined earlier. Indeed, suppose that p_j either has larger degree than p_1 or a different leading coefficient from that of p_1. Then all the polynomials $p_{i,j}$ for $0 \leqslant i \leqslant d_{\max}$ have the same leading coefficient and the same degree as p_j. If p_j has the same degree and leading coefficient as p_1 then the polynomials $p_{i,j}$ for $1 \leqslant i \leqslant d_{\max}$ all have smaller degree than p_1. Thus the weight vector of this set is the same as for P in its ith components for $i > \deg p_1$ and is one smaller for $i = \deg p_1$. (The components may increase for $i < \deg p_1$.) Thus the weight vector of this set is less than that of P.

We now have to modify these polynomials slightly. We need to do this since later in the proof we are going to divide $[N']$ into blocks of length N and we need to take account of this. Let $q_{i,j}(n) = p_{i,j}(Nn)/N$. These are integral polynomials since the $p_{i,j}$ are. (Recall that integral polynomials take the value zero at zero, and thus the $q_{i,j}$ have integer coefficients.) Also, observe that this collection has the same weight vector as the collection of $p_{i,j}$ since, although the leading coefficients change, the number of distinct leading coefficients of polynomials of a given degree does not. Thus the outer induction hypothesis applies to the $q_{i,j}$.

Again we divide $[N']$ into blocks of size N, setting $B_s = \{(s-1)N+1, (s-1)N+2, \ldots, sN\}$. Then, since there are only k^N ways of colouring a block, we can apply the outer induction to find s, t such that the blocks $B_{s_{i,j}}$ and B_s are all coloured identically and $s_{i,j} - s = q_{i,j}(t)$, provided that we have chosen N' large enough.

By the choice of N and the assumption that there do not exist a and d with the set $\{a\} \cup \{a+p_j(d) : 1 \leqslant j \leqslant m\}$ monochromatic, we know that B_s contains $r-1$ colour focused sets $A_1, A_2, \ldots, A_{r-1}$ together with their focus a such that a has a colour distinct from that of any of the A_i. Suppose that $A_i = \{a_{i,j} : 1 \leqslant j \leqslant m\}$ and that $a_{i,j} - a = p_j(d_i)$.

Claim 2. *The sets $\{a_{i,j} + Nq_{d_i,j}(t) : 1 \leqslant j \leqslant m\}$ for $1 \leqslant i \leqslant r-1$ together with the set $\{a + Nq_{0,j}(t) : 1 \leqslant j \leqslant m\}$ are colour focused at $a - p_1(Nt)$.*

Proof. First, observe that by construction each of the points $a_{i,j} + Nq_{d_i,j}(t)$ has the same colour as the point $a_{i,j}$. Thus the set $\{a_{i,j} + Nq_{d_i,j}(t) : 1 \leqslant j \leqslant m\}$ has the same colour as A_i. Thus these colours are all distinct. Also, each of the points $a + Nq_{0,j}(t)$ has the same colour as a, so the set $\{a + Nq_{0,j}(t) : 1 \leqslant j \leqslant m\}$ is monochromatic. Also, a has a colour distinct from that of the other sets.

Thus we only need to check the focusing:

$$\begin{aligned} a_{i,j} + Nq_{d_i,j}(t) - (a - p_1(Nt)) &= p_j(d_i) + p_{d_i,j}(Nt) + p_1(Nt) \\ &= p_j(d_i) + p_j(Nt + d_i) - p_j(d_i) - p_1(Nt) + p_1(Nt) \\ &= p_j(Nt + d_i) \end{aligned}$$

and

$$\begin{aligned} a + Nq_{0,j}(t) - (a - p_1(Nt)) &= p_{0,j}(Nt) + p_1(Nt) \\ &= p_j(Nt). \end{aligned}$$

Therefore the points are colour focused and the induction is almost complete. □

To finish the proof we note that all the properties that we have used are translation-invariant and that we can bound $a - p_1(Nt)$ below, by $-M$ say. Thus we can apply the above arguments to the blocks $B'_s = M + B_s$ to ensure that the new focus $(a - p_1(Nt))$ is positive. □

There is a multi-dimensional version of van der Waerden's theorem.

Gallai's theorem. *Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_m$ are vectors in $\mathbb{N}^D$ and that $\mathbb{N}^D$ is finitely coloured. Then there exist $a \in \mathbb{N}^D$ and $d \in \mathbb{N}$ such that the set of points $\{a\} \cup \{a + d\lambda_1, a + d\lambda_2, \ldots, a + d\lambda_m\}$ is monochromatic.*

We can prove a polynomial version of Gallai's theorem. Define a D-dimensional integral polynomial $p(n)$ to be a polynomial in n with coefficients in $\mathbb{Z}^D$ which is zero at zero. We can modify the above proof to work in this case if we replace the scalars by vectors and the $[N]$ by $[N] \times [N] \times \ldots \times [N]$. Thus we have the following theorem.

Theorem 2. *Suppose that $p_1, p_2, \ldots, p_m$ are D-dimensional integral polynomials and that $\mathbb{N}^D$ is finitely coloured. Then there exist $a \in \mathbb{N}^D$ and $d \in \mathbb{N}$ such that the set of points $\{a\} \cup \{a + p_i(d) : 1 \leqslant i \leqslant m\}$ is monochromatic.*

We do not prove this in any greater detail since this result is an easy corollary of the polynomial Hales–Jewett theorem proved in the next section. Finally, we state a special case of this theorem.

COROLLARY 1. *Suppose that $p_1, p_2, \ldots, p_m$ are integral polynomials, $u_1, u_2, \ldots u_m$ are vectors in $\mathbb{Z}^D$ and that $\mathbb{N}^D$ is finitely coloured. Then there exist $a \in \mathbb{N}^D$ and $d \in \mathbb{N}$ such that the set of points $\{a\} \cup \{a + u_i p_i(d) : 1 \leqslant i \leqslant m\}$ is monochromatic.*

Note that this result includes Gallai's theorem as a special case.

3.1. *Bounds*

This method does yield explicit bounds (unlike the ergodic theoretic argument, which only proves the existence bounds via a compactness argument) but the bounds are extremely large. Even for van der Waerden's theorem it is well known that this colouring argument yields an Ackermann-type bound. To get the full polynomial result, the induction is much longer. Even for the case of a and $a + d^2$ the bound is very bad, since we need a long arithmetic progression of blocks at each stage of the induction.

4. *The polynomial Hales–Jewett theorem*

First we define some notation. Let $Q = [q]^N$. For $a \in Q$, $\gamma \subset [N]$ and $1 \leqslant x \leqslant q$ let $a \oplus x\gamma$ denote the vector b in Q obtained by setting $b_i = x$ if $i \in \gamma$ and $b_i = a_i$ otherwise. A *combinatorial line* is a set of the form $\{a \oplus x\gamma : 1 \leqslant x \leqslant q\}$ (where γ is non-empty – from now on this will be assumed). The set γ is called the set of *active coordinates*. An *m-dimensional combinatorial subspace* is a set of the form $\{a \oplus \bigoplus_{i=1}^m x_i \gamma_i : 1 \leqslant x_i \leqslant q\}$ where $a \in Q$ and $\gamma_1, \gamma_2, \ldots, \gamma_m \subset [N]$ and the γ_i are disjoint.

HALES–JEWETT THEOREM. *For any q and k there exists N such that whenever $Q = [q]^N$ is k-coloured there exist $a \in Q$ and $\gamma \subset [N]$ such that the set of points $\{a \oplus x\gamma : 1 \leqslant x \leqslant q\}$ is monochromatic.*

Note that the identification of the base set with $[q]$ is purely a notational convenience; the theorem is true for any set of cardinality q.

It is easy to deduce an extended (multi-dimensional) version of this theorem from the (one-dimensional) theorem above.

EXTENDED HALES–JEWETT THEOREM. *For any q, k, d there exists N such that whenever $[q]^N$ is k-coloured there exist $a \in Q$ and disjoint $\gamma_1, \gamma_2, \ldots, \gamma_d \subset [N]$ such that the set of points $\{a \oplus \bigoplus_{i=1}^d x_i \gamma_i : 1 \leqslant x_i \leqslant q\}$ is monochromatic.*

Observe that the subspace 'looks like' the original space. Indeed the map $Q(d) \longrightarrow Q(N)$ given by $(x_1, x_2, \ldots, x_d) \longmapsto a \oplus \bigoplus_{i=1}^d x_i \gamma_i$ preserves combinatorial lines and subspaces.

Proof of extended Hales–Jewett theorem. We can identify $[q^d]^N$ with $[q]^{dN}$. Applying the Hales–Jewett theorem to $[q^d]$ we can choose N such that whenever $[q^d]^N$ is k-coloured there exist $a \in [q^d]^N$ and $\gamma \subset [N]$ with $\{a \oplus x\gamma : 1 \leqslant x \leqslant q^d\}$ monochromatic. Under the above identification this corresponds to the set (in $[q]^{dN}$) $\{a \oplus \bigoplus_{i=1}^d x_i \gamma_i : 1 \leqslant x_i \leqslant q\}$ (where γ_i are sets identified to γ) which is thus monochromatic. □

Before stating the polynomial Hales–Jewett theorem in its full generality, we state a special case.

QUADRATIC HALES–JEWETT THEOREM. *For any q, k there exists N such that whenever $[q]^{N\times N}$ is k-coloured there exist $a\in Q$ and $\gamma\subset N$ such that the set of points $\{a\oplus x(\gamma\times\gamma):1\leqslant x\leqslant q\}$ is monochromatic.*

Thus the theorem asserts the existence of a monochromatic combinatorial line whose active coordinates are of a special form; in this case they form a 'square'.

To state the polynomial Hales–Jewett theorem in its full generality we need a little more notation. Let N^d denote the d-fold product $N\times N\times\ldots\times N$. For convenience we will always assume that an i-dimensional subset (for example, γ^i) is in $[N]^i$. We now state the theorem.

POLYNOMIAL HALES–JEWETT THEOREM. *For any q,k,d there exists N such that whenever $Q=Q(N)=[q]^N\times[q]^{N\times N}\times\ldots\times[q]^{N^d}$ is k-coloured there exist $a\in Q$ and $\gamma\subset[N]$ such that the set of points $\{a\oplus x_1\gamma\oplus x_2(\gamma\times\gamma)\oplus\ldots\oplus x_d\gamma^d:1\leqslant x_i\leqslant q\}$ is monochromatic.*

Again, the identification of the coordinates with $[q]$ (rather than just an arbitrary set of size q) is purely a notational convenience. Also, the fact that the same set $[q]$ is used in each term is an unnecessary restriction; the sets do not even need to be the same size.

This theorem is essentially the same as the one in [**2**]. We put it in a more general formulation (although it is easy to deduce it from their results) which is slightly easier to use for applications and leads more naturally to the method of induction that we use. The proof is a colour focusing argument as in the first section. The induction method is very similar to the one used there. It will be convenient, however, to formulate it in a coordinate-free way.

An indication that this is a sensible formulation for the theorem is given by the easy deduction of the polynomial van der Waerden theorem. Suppose that we have a k-colouring of $\mathbb{N}$. Let d be the maximum degree of any of the polynomials, and let m be the maximum modulus of their coefficients. We will use $\{-m,\ldots,m\}^N$ as the base set. Let $Q(N)$ be the corresponding set. Map $Q(N)$ into $\mathbb{N}$ by sending $q\in Q$ to the sum of its coordinate values plus a constant M to ensure that all the numbers are positive ($M=m^N+1$ will do). This induces a k-colouring of Q. By the theorem we have a, γ such that the set $\{a\oplus\bigoplus_{i=1}^d x_i\gamma^i:-m\leqslant x_i\leqslant m\}$ is monochromatic. This maps back to a pair b, c such that the set $\{M+b+\sum_{i=1}^d x_i c^i:-m\leqslant x_i\leqslant m\}$ is monochromatic. By the choice of m and d this implies that the points $M+b$ and $M+b+p(c)$, for any p in the set of polynomials, are the same colour.

We now prove the polynomial Hales–Jewett theorem. First we define the space that we will work in. For any sets $q_1,q_2,\ldots,q_d$ let $Q=Q(N)=q_1^N\times q_2^{N\times N}\times\ldots\times q_d^{N^d}$. Call a space of this form a *Hales–Jewett space*. Define a *polynomial* $p(\gamma)$ to be a formal sum $\bigoplus_{i=1}^d c_i\gamma^i$ where $c_i\in q_i$. Thus for any $a\in Q$ and any $\gamma\subset[N]$ the vector $a\oplus p(\gamma)$ is defined.

Now, because the q_i are arbitrary sets we have no notion of a zero polynomial or natural definition of degree. However, we can define the degree of one polynomial relative to another. Suppose that p_1, p_2 are polynomials with coefficients c_i and d_i respectively. Define the *degree of p_2 relative to p_1* to be the maximal i such that $c_i\neq d_i$. Define the *leading coefficient of p_2 relative to p_1* to be d_i, where i is the degree of p_2 relative to p_1.

Suppose that $P = \{p_0, p_1, \dots p_m\}$ is a collection of polynomials. Define the *weight vector* $(w_1, w_2, \dots, w_d)$ *of* P *relative to* p_i by setting w_j equal to the number of distinct leading coefficients (relative to p_i) of the polynomials of degree j (relative to p_i). (This is an analogous definition to that in the proof of the polynomial van der Waerden theorem.) We use the same ordering as before; that is, $(v_1, v_2, \dots, v_d) < (w_1, w_2, \dots, w_d)$ if there exists i with $v_i \leqslant w_i$ and $v_j = w_j$ for all $j > i$. Define the *weight vector of* P to be the minimum over i of the weight vector of P relative to p_i (the minimum is in the above ordering).

Outer induction hypothesis. Let $Q = Q(N) = q_1^N \times q_2^{N\times N} \times \dots \times q_d^{N^d}$. Suppose that $p_0, p_1, \dots, p_m$ are polynomials. Then for all k there exists N such that whenever $Q(N)$ is k-coloured there exist $a \in Q$ and $\gamma \subset N$ such that the set $\{a \oplus p_0(\gamma), a \oplus p_1(\gamma), \dots, a \oplus p_m(\gamma)\}$ is monochromatic.

The induction is on the weight vector of the polynomials. Let $P = \{p_0, p_1, \dots, p_m\}$. Thus we assume that the hypothesis is true for any collection of polynomials with weight vector less than that of P. We will assume that the weight vector of P is minimal with respect to p_0. We will further assume that p_1 has minimal degree with respect to p_0.

We say that a pair (a', γ) is *focused at* a if $a = a' \oplus p_0(\gamma)$. We say that r pairs $(a_1, \gamma_1), (a_2, \gamma_2), \dots, (a_r, \gamma_r)$ are *colour focused at* a if each pair (a_i, γ_i) is focused at a, each of the sets $A_i = \{a_i \oplus p_1(\gamma_i), a_i \oplus p_2(\gamma_i), \dots, a_i \oplus p_m(\gamma_i)\}$ is monochromatic, and all these sets have distinct colours. As before, this is exactly the condition that forces a to have a colour different from that of any of the A_i.

Inner induction hypothesis. For all $r \leqslant k$ there exists N such that if $Q(N)$ is k-coloured then either there exist r pairs $(a_1, \gamma_1), (a_2, \gamma_2), \dots, a_r, \gamma_r)$ colour focused at a, or there exist $a \in Q$ and $\gamma \subset [N]$ such that the set $\{a \oplus p_0(\gamma), a \oplus p_1(\gamma), \dots, a \oplus p_m(\gamma)\}$ is monochromatic.

As usual, the result is immediate from here. Indeed, set $r = k$. If the latter case occurs then we are done. If the former occurs then the focus a has the same colour as one of the sets A_i and $A_i \cup \{a\}$ forms the required monochromatic set.

The induction is on r. Suppose that the hypothesis is true for $r-1$ and N. We shall show that there is an N' such that $Q(N+N')$ satisfies the hypothesis for r. As before, we do not calculate the bound explicitly.

Suppose that we have a k-colouring of $Q(N+N')$. We may assume that we do not have $a \in Q$ and $\gamma \subset [N]$ with $\{a \oplus p_0(\gamma), a \oplus p_1(\gamma), \dots, a \oplus p_m(\gamma)\}$ monochromatic.

Let $Q' = Q'(N') = Q(N+N')/Q(N)$. Then Q' is a product of sets of the form $q^{N^s N'^t}$ with $t \geqslant 1$. The set $q^{N^s N'^t}$ can be identified with $(q_i^{N^s})^{N'^t}$. Collecting the terms obtained in this way, we can identify Q' with a larger Hales–Jewett space on N'. For example, in two dimensions,

$$\begin{aligned} Q(N+N') &= q_1^{N+N'} \times q_2^{(N+N')\times(N+N')} \\ &= q_q^{N+N'} \times q_2^{N\times N} \times q_2^{N\times N'} \times q_2^{N'\times N} \times q_2^{N'\times N'}. \end{aligned}$$

Thus

$$\begin{aligned} Q'(N') &= q_1^{N'} \times q_2^{N\times N'} \times q_2^{N'\times N} \times q_2^{N'\times N'} \\ &\equiv (q_1 \times q_2^N \times q_2^N)^{N'} \times q_2^{N'\times N'} \end{aligned}$$

which is a Hales–Jewett space on N'. We have to be slightly careful when defining polynomials; we have to make sure that they are polynomials on the identified space.

For example suppose that we identify $[q]^{N\times N'}$ with $[q^N]^{N'}$ and that $\delta\subset N$. Then $p(\gamma)=c(\delta\times\gamma)$ is not a polynomial. This is because we have not defined a coordinate in q^N, but only in q^δ. To obtain a polynomial we have to define what happens to $\{i\}\times\gamma$ for every $i\in[N]$.

It is convenient to define some polynomials on Q' in two parts. Suppose that $\delta\subset N$. One part of the polynomial will define its effect on terms of the form $q_i^{\delta^s N'^t}$ (that is, not on all of $q_i^{N^s N'^t}$) and the other part its effect on $q_i^{N^s\backslash\delta^s N'^t}$. Let Q'_δ denote the product of terms of the first form and Q'^c_δ the product of those of the second.

For $\gamma\subset N'$ let $q_\delta(\gamma)=p_0(\delta\cup\gamma)-p_0(\delta)$ where the minus denotes that we ignore the contribution on $Q(N)$ and view this as a polynomial on Q'_δ. Let $q^c_\delta(\gamma)$ be such that $q_\delta(\gamma)+q^c_\delta(\gamma)=q_N(\gamma)$ (so q^c_δ is a polynomial on Q'^c_δ).

Now define

$$q_{j,\delta}(\gamma)=p_j(\delta\cup\gamma)-p_j(\delta)+q^c_\delta(\gamma)\quad 1\leqslant j\leqslant m,\ \delta\subset N.$$

Observe that these are polynomials; the q^c term completes the definition.

Claim 3. *The weight vector of this collection of polynomials is less than the weight vector of P.*

Proof. We show that the weight vector of this collection of polynomials with respect to $q_{1,\varnothing}$ is smaller than the weight vector of P. This clearly implies the result. (Recall that the weight vector is the smallest that can be obtained.)

First consider the coefficient of γ^r in $q_{j,\delta}$ for $r>\deg p_j$ (relative to p_0). Since the coefficients of p_j are equal to the coefficients of p_0 for terms of degree at least r and these are the only coefficients that affect the coefficient of γ^r in $q_{j,\delta}$ we see that this coefficient is independent of j. For $j=0$ the polynomial is just $q_\delta(\gamma)+q^c_\delta(\gamma)=q_N(\gamma)$, which is independent of δ. Thus the coefficient of γ^r in $q_{j,\delta}$ is independent of j and δ. In particular, it is the same as the coefficient of γ^r in $q_{1,\varnothing}$. Thus the degree of $q_{j,\delta}$ relative to $q_{1,\varnothing}$ is at most the degree of p_j relative to p_0.

Next we consider the coefficient of γ^r in $q_{j,\delta}$ for $r=\deg p_j$. The above argument shows that only the leading coefficient of p_j affects this coefficient. Thus it is only the $(\delta\cup\gamma)^r$ term in $p_j(\delta\cup\gamma)$ that has any effect. Moreover, if we expand $(\delta\cup\gamma)^r$ the only term that affects this coefficient is the γ^r term. In particular, it is independent of δ. Thus the coefficient of γ^r in $q_{j,\delta}$ depends only upon the leading coefficient of p_j (and different leading coefficients do give rise to different coefficients).

Thus all the polynomials p_j with degree r and leading coefficient c, where $r>\deg p_1$ or c is not equal to the leading coefficient of p_1, give rise to a collection of polynomials of degree r and all with same leading coefficient. The collection of polynomials with the same degree and leading coefficient as p_1 give rise to a collection of polynomials of degree less than that of p_1 (since we are measuring degree relative to $q_{1,\varnothing}$). (This collection may have many different leading coefficients.) Thus the weight vector of the new collection is the same in coordinates greater than the degree of p_1 and one smaller for the $\deg p_1$ coordinate. (It may have increased for the smaller coordinates.) Therefore, the weight vector of the new collection is smaller than the weight vector of P. □

Colour Q' by the complete colouring of $Q(N)$ it induces. By the outer induction hypothesis applied to this colouring we can find (b,γ) such that all the points $b\oplus q_{j,\delta}(\gamma)$ have the same colour (that is, they induce the same complete colouring of Q).

The inner induction hypothesis implies that in this colouring of $Q(N)$ we have $r-1$ pairs (a_i, δ_i) colour focused at a.

CLAIM 4. *The $r-1$ pairs $((a_i, b \oplus q^c_{\delta_i}(\gamma)), \delta_i \cup \gamma)$ together with $((a, b \oplus q^c_{\varnothing}(\gamma)), \gamma)$ are colour focused at $(a, b \oplus q_N(\gamma))$.*

Proof. First observe that

$$\begin{aligned}(a_i, b \oplus q^c_{\delta_i}(\gamma)) \oplus p_0(\delta_i \cup \gamma) &= (a_i \oplus p_0(\delta_i), b \oplus q^c_{\delta_i}(\gamma) \oplus q_{\delta_i}(\gamma)) \\ &= (a, b \oplus q_N(\gamma))\end{aligned}$$

and

$$\begin{aligned}(a, b \oplus q^c_{\varnothing}(\gamma)) \oplus p_0(\gamma) &= (a, b \oplus q^c_{\varnothing}(\gamma) \oplus q_{\varnothing}(\gamma)) \\ &= (a, b \oplus q_N(\gamma)).\end{aligned}$$

Thus the pairs are focused.

We also have

$$\begin{aligned}(a_i, b \oplus q^c_{\delta_i}(\gamma)) \oplus p_j(\delta_i \cup \gamma) &= (a \oplus p_j(\delta_i), b \oplus q_{\delta_i}(\gamma) \oplus p_j(\delta_i \cup \gamma) - p_j(\delta_j)). \\ &= (a \oplus p_j(\delta_i),\ b \oplus q_{j,\delta_i}(\gamma)).\end{aligned}$$

For any j, δ the colouring that $b \oplus q_{j,\delta}(\gamma)$ induces on Q is the same. Thus the point has the same colour as $a \oplus p_j(\delta_i)$ in the induced colouring. Therefore the sets are monochromatic and have distinct colours.

Finally, we have

$$\begin{aligned}(a, b \oplus q^c_{\varnothing}(\gamma)) \oplus p_j(\gamma) &= (a, b \oplus q^c_{\varnothing}(\gamma) \oplus p_j(\varnothing \cup \gamma) - p_j(\varnothing)) \\ &= (a, b \oplus q_{j,\varnothing}(\gamma)).\end{aligned}$$

Thus each of these points has the same colour as a in the induced colouring. Therefore the set is monochromatic and has colour distinct from that of the other $r-1$ sets. Therefore the sets are colour focused, and we have done. □

To conclude this section we use this result to deduce an extended version.

EXTENDED POLYNOMIAL HALES–JEWETT THEOREM. *For any q, m, k, d there exists N such that whenever $Q = Q(N) = [q]^N \times [q]^{N \times N} \times \ldots \times [q]^{N^d}$ is k coloured there exists $a \in Q$ and $\gamma_1, \gamma_2, \ldots, \gamma_m \subset N$ such that the set of points*

$$\{a \oplus \bigoplus_{i \leqslant m} x_i \gamma_i \oplus \bigoplus_{i_1, i_2 \leqslant m} x_{i_1, i_2}(\gamma_{i_1} \times \gamma_{i_2}) \ldots \oplus \bigoplus_{i_1, i_2, \ldots, i_d \leqslant m} x_{i_1, i_2, \ldots, i_d}(\gamma_{i_1} \times \gamma_{i_2} \times \ldots \times \gamma_{i_d}) : \quad 1 \leqslant x_{i_1, i_2, \ldots, i_r} \leqslant q\}$$

is monochromatic.

Thus this theorem allows us to vary the 'cross terms' (such as $\gamma_i \times \gamma_j$) as well as the normal 'square' terms. This subspace is isomorphic to $Q(m)$.

Proof of extended polynomial Hales–Jewett theorem. The proof is the same as for the linear case but we have to identify more copies. For example in two dimensions we identify $[q^{r^2}]^{N \times N}$ with $[q]^{Nr \times Nr}$. Then a square combinatorial line in the former corresponds to an r-dimensional subspace in the latter. □

5. *Further results*

Polynomial extensions of many Ramsey theorems can be proved using the results in this paper. This is done in [7]. The results proved there include polynomial versions of the Carlson–Simpson theorem and the Graham–Rothschild theorem on parameter sets.

References

1. V. Bergelson and A. Leibman, 'Polynomial extensions of van der Waerden and Szemerédi theorems', *J. Amer. Math. Soc.* 9 (1996) 725–753.
2. V. Bergelson and A. Leibman, 'Set polynomials and polynomial extension of the Hales–Jewett theorem', preprint, 1997.
3. H. Furstenberg, 'Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions', *J. Anal. Math.* 31 (1977) 204–256.
4. A. W. Hales and R. I. Jewett, 'Regularity and positional games', *Trans. Amer. Math. Soc.* 106 (1963) 222–229.
5. A. Sarkozy, 'On difference sets of integers I', *Acta. Acad. Sci. Hungar.* 31 (1978) 125–149.
6. B. L. van der Waerden, 'Beweis einer Baudetschen Vermutung', *Nieuw. Arch. Wisk.* 15 (1927) 212–216.
7. M. J. Walters, 'Some extensions of the polynomial Hales–Jewett theorem', in preparation.

Department of Pure Mathematics and Mathematical Statistics
Cambridge University
16 Mill Lane
Cambridge CB2 1SB