Purely Combinatorial Proofs of Van Der Waerden-Type Theorems

William Gasarch and Clyde Kruskal and Andy Parrish

December 2, 2010

Contents

Preface

Ramsey Theory is a branch of combinatorics that can (very roughly) be characterized by the statement

no matter how you color some combinatorical object there will be a monochromatic part that is orderly

or, to quote Theordore S. Motzkin,

complete disorder is impossible.

There are already two elementary books on Ramsey Theory:

1. *Ramsey Theory* by Graham, Spencer, and Rothchild [13].

2. *Ramsey Theory over the Integers*, Landman and Robertson [19]

These books are elementary in that they mostly do not use advanced techniques. Within Ramsey Theory, there are two theorems that stand out.

1. *Ramsey's theorem*: For all c, m, there exists n, such that, for every c-coloring of K_n there is a monochromatic K_m.

2. *van der Waerden's theorem*: For all c, k, there exists W such that, for every c-coloring of $\{1, \ldots, W\}$ there exists a monochromatic arithmetic sequence of length k.

The books on Ramsey Theory mentioned above contain both of these theorems. By contrast, this book is just about van der Waerden's theorem and its extensions. Given our focus, we can cover more ground.

Our goal is to cover virtually every extension or variant of van der Waerden's theorem that can be proven using purely combinatorial methods. What does it mean to say that a proof is *purely combinatorial*? We take this to mean that no methods from Calculus or Topology are used. This does not mean the proofs are easy; however, it does mean that no prior math is required aside from some simple combinatorics.

In this preface and in the introduction we will state many theorems that can be proven by purely combinatorial methods, and will later prove them. By contrast, we now give two true statements for which currently no purely combinatorial proof is known.

1. For all k there exists a, d such that

   $$a, a+d, a+2d, \ldots, a+(k-1)d$$

   are all primes. This was proven by Green and Tau [14]. They used Fourier Analysis and Topology.

2. Let $W(k,c)$ be the least W such that van der Waerden's theorem holds with this value of W. Then

   $$W(k,c) \leq W(k,c) \leq 2^{2^{c2^{2^{k+9}}}}$$

   This was proven by Gowers[12]. He used techniques from analysis, notably Fourier Analysis. This theorem is important since the combinatorial proofs of van der Warden's theorem yield much larger bounds.

Since we only use purely combinatorial techniques, in terms of background knowledge, a high school student could read this monograph. Well— a high school student who knew some (though not much) combinatorics and was very interested in the topic. Several of the people in the acknowledgements are high school students. Hence we have the following list of people who could read this book:

1. An awesome High School Student.

2. An excellent College Junior math major.

3. A very good first year grad student in math or math-related field.

4. A pretty good third year grad student in math who is working in combinatorics.

5. A mediocre PhD in Combinatorics.

There are a few chapters or sections that require knowledge that is not purely combinatorial. These sections have stars next to them.

Throughout this monograph we use the following conventions.

1. VDW is van der Waerden's Theorem.

2. POLYVDW is the polynomial van der Waerden's Theorem.

3. HJ is the Hales-Jewett Theorem.

4. POLYHJ is the Polynomial Hales-Jewett Theorem.

5. Any of these can be used as a prefix. For example "VDW numbers" will mean "van de Waerden numbers"

The motivation for this monograph is Walters' paper [35]. He gave the first purely combinatorial proofs of POLYVDW. and POLYHJ . However, his techniques can be used to obtain cleaner proofs of the original VDW and HJ . In addition, there are many corollaries of these theorems that, because of his work, now have purely combinatorial proofs.

There is one theorem we will give several proofs of throughout this monograph.

The Square Theorem: For all 2-colorings of $\mathbb{Z} \times \mathbb{Z}$ there exists a square that has all four corners the same color. (We will state this in a different form later.)

In Chapter 2 we prove VDW with the same proof that van der Waerden gave, though expressed as a color-focusing argument. In Chapter 3 we give some applications of VDW, including the square theorem. In Chapter 4 we prove POLYVDW using Walters' proof. In Chapter 3 we give some applications of the POLYVDW.

In Chapter 6 we prove HJ. We prove it two ways. The classical proof yields insanely large upper bounds on the HJ numbers. The proof by Shelah gives sanely large upper bounds. In Chapter 7 we give many applications of HJ, including the square theorem and some lower bounds in communication complexity. In Chapter 8 we prove POLYHJ. In Chapter 9 we give some applications of the POLYHJ including the multidimenional POLYVDW and the generalized POLYVDW theorem. In Chapter 10 we give the proof of Rado's theorem, which is a generalization of VDW. In Chapter 11 we give some appliations of Rado's theorem. In Chapter 12 we briefly look at some theorems that are proven using non-combinatorial methods, including Szemeredi's density theorem for $k = 3$.

Acknowledgements

We would like to thank the members of the *van der Waerden gang* who listened to us as we worked out the ideas and expositions in this monograph. The gang had a somewhat fluid membership; however, the following people deserve thanks: Yosif Berman, Lee Steven Greene, Aki Hogue, Clyde Kruskal, Justin Kruskal, Matt Jordan, Richard Matthew McCutchen, Jefferson Pecht, and Louis Wasserman,

We would also like to thank Hunter Monroe for an incredible job of proofreading.

Chapter 1

Introduction

1.1 What is Van Der Waerden's Theorem?

Imagine that someone colors the the numbers $\{1, \ldots 9\}$ RED and BLUE. Here is an example:

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ R & R & B & B & R & R & B & B & R \end{array}$$

Note that there is a sequence of 3 numbers that are the same color and are equally spaced, namely

$$1, 5, 9.$$

Try 2-coloring $\{1, \ldots, 9\}$ a different way. Say

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ R & B & R & B & B & R & B & B & R \end{array}$$

Note that there is a sequence of 3 numbers that are the same color an equally spaced, namely

$$2, 5, 8.$$

Is there a way to 2-color $\{1, \ldots, 9\}$ and not get such a sequence?

Exercise 1 Show that for all 2-colorings of $\{1, \ldots, 9\}$ there exists a set of three numbers that are equally spaced and the same color.

We want to generalize this.

Def 1.1.1 Let $k \in \mathbb{N}$. An *Arithmetic Sequence of Length* k is a sequence of natural numbers the form

$$a, a+d, a+2d, \ldots, a+(k-1)d$$

where $d \neq 0$. In other words, it is a sequence of k numbers that are equally spaced. We often refer to this as a k-AP. If there is a coloring of the natural numbers (or a finite subset of them) then we will use the term *monochromatic* k*-AP* to mean an arithmetic sequence or length k where all of the elements of it are the same color.

In the above example we looked at 2-coloring $\{1, 2, \ldots, 9\}$ and seeing if there was a monochromatic 3-AP. It turned out that there was always a monochromatic 3-AP. What if you increased the number of colors? What if you increased the length k ?

We now proceed more formally.

Notation 1.1.2 If $m \in \mathbb{N}$ then $[m]$ is $\{1, \ldots, m\}$.

The following was first proven by Van Der Waerden [34].
Van Der Waerden's Theorem: For every $k \geq 1$ and $c \geq 1$ there exists W such that for every c-coloring $COL : [W] \rightarrow [c]$ there exists a monochromatic k-AP. In other words there exists a, d, $d \neq 0$, such that

$$COL(a) = COL(a+d) = \cdots = COL(a+(k-1)d).$$

Def 1.1.3 Let $k, c \in \mathbb{N}$. $W(k, c)$ is the least W that satisfies VDW. $W(k, c)$ is called a *van der Waerden number*.

We will prove van der Waerden's theorem in Section 2.2.3.

1.2 The Polynomial Van Der Waerden Theorem

In van der Waerden's theorem we can think of

$$a, a+d, \ldots, a+(k-1)d$$

as

$$a, a+p_1(d), \ldots, a+p_{k-1}(d)$$

where $p_i(x) = ix$. Why these functions?

Notation 1.2.1 $\mathbb{Z}$ is the set of integers. $\mathbb{Z}[x]$ is the set of polynomials with integer coefficients.

The following remarkable theorem was first proved by Bergelson and Leibman [1].

Polynomial Van Der Waerden Theorem For any natural number c and any polynomials $p_1(x), \ldots, p_k(x) \in \mathbb{Z}[x]$ such that $(\forall i)[p_i(0) = 0]$, there exists W such that, for any c-coloring $COL : [W] \rightarrow [c]$ there exists a, d, $d \neq 0$, such that

$$COL(a) = COL(a+p_1(d)) = COL(a+p_2(d)) = \cdots = COL(a+p_k(d)).$$

Def 1.2.2 Let $p_1, \ldots, p_k \in \mathbb{Z}[x]$ and $c \in \mathbb{N}$. $W(p_1, \ldots, p_k; c)$ is the least W that satisfy POLYVDW . $W(p_1, \ldots, p_k; c)$ is called a *polynomial van der Waerden number*.)

POLYVDW was proved for $k = 1$ by Furstenberg [**?**] and (independently) Sarkozy [27]. The original proof of the full theorem by Bergelson and Leibman [1] used ergodic methods. A later proof by Walters [35] uses purely combinatorial techniques. We will present an expanded version of Walters' proof in Section 2.2.3.

Upon seeing the polynomial van der Waerden Theorem one may wonder, is it true over the reals? Or some other ring? How do you even state this? This requires some discussion. The following is true and is equivalent to POLYVDW.

Polynomial Van Der Waerden Theorem over $\mathbb{Z}$: For any natural number c and any polynomials $p_1(x), \ldots, p_k(x) \in \mathbb{Z}[x]$ such that $(\forall i)[p_i(0) = 0]$, there exists W such that for any c-coloring $COL : [-W, W] \rightarrow [c]$ there exists a $a, d \in \mathbb{Z}$, $d \neq 0$, such that

$$COL(a) = COL(a + p_1(d)) = COL(a + p_2(d)) = \cdots = COL(a + p_k(d)).$$

Note that the domain $\mathbb{Z}$ appears three times— the polynomials have coefficients in $\mathbb{Z}$, the c-coloring is of a subset of $\mathbb{Z}$, and $a, d \in \mathbb{Z}$. What if we replaced $\mathbb{Z}$ by $\mathbb{R}$ or some other integral domain? We would obtain the following:

Generalized Polynomial Van Der Waerden Theorem: Let S be any infinite integral domain. For any natural number c and any polynomials $p_1(x), \ldots, p_k(x) \in S[x]$ such that $(\forall i)[p_i(0) = 0]$, there exists a finite set $S' \subseteq S$ such that for any c-coloring $COL : S' \rightarrow [c]$ there exists $a, d \in S$, $d \neq 0$, such that

$$COL(a) = COL(a + p_1(d)) = COL(a + p_2(d)) = \cdots = COL(a + p_k(d)).$$

This was first proven by Bergelson and Leibman [2]. In that paper they proved POLYHJ (which we will state later) using ergodic techniques, and then derived the Generalized POLYVDW theorem as an easy corollary. Later Walters [35] obtained a purely combinatorial proof of POLYHJ. Putting all of this together one obtains a purely combinatiral proof of Generalized POLYVDW. One of the motivations for this monograph was to do that putting together.

1.3 Hales-Jewett Theorem and Polynomial Hales-Jewett Theorem

HJ [15] is a generalization of van der Waerden's theorem which we will state and prove in Chapter 6. There is also a POLYHJ [2], which we will state and prove in Chapter 8.

The following is a corollary of HJ:

The Square Theorem: For any c there exists W such that for any c-coloring of $[W] \times [W]$ there exists a square with all four corners the same color. Formally: for any c-coloring $COL : [W] \times [W] \rightarrow [c]$ there exists a_1, a_2, d, $d \neq 0$, such that

$$COL(a_1, a_2) = COL(a_1, a_2 + d) = COL(a_1 + d, a_2) = COL(a_1 + d, a_2 + d).$$

Def 1.3.1 Let $c \in \mathbb{N}$. $W_{sq}(c)$ is the least W satisfying the Square Theorem.

One can restate the Square Theorem as saying that you obtain a rectangle similar to the to 1×1 rectangle with all corners the same color. More is known:

The Rectangle Theorem: For any c, for any rectangle R with natural number sides, there exists W such that for any c-coloring of $[W] \times [W]$ there exists a rectangle similar to R with all four corners the same color.

Def 1.3.2 Let R be a rectangle with natural number sides. Let $c \in \mathbb{N}$. $W(R, c)$ is the least W satisfying the Rectangle Theorem.

This can be generalized even further. The result is the Gallai-Witt theorem which we discuss in Section 7.5 The Gallai-Witt theorem is also called (correctly) the multidimensional VDW theorem. There is no publication by Gallai that contains it; however, Rado [23],[24]) credits him. Witt [36] proved it independently.

The following is a corollary of POLYHJ:

The Squared Rectangle Theorem: For any c there exists W such that for any c-coloring of $[W] \times [W]$ there exists $d \in [W]$, $d \neq 0$, such that there is a $d \times d^2$ rectangle with all four corners the same color. Formally: for any c-coloring $COL : [W] \times [W] \to [c]$ there exists a_1, a_2, d, $d \neq 0$, such that

$$COL(a_1, a_2) = COL(a_1, a_2 + d^2) = COL(a_1 + d, a_2) = COL(a_1 + d, a_2 + d^2).$$

Def 1.3.3 Let $c \in \mathbb{N}$. $W_{sq\text{-}rect}(c)$ is the least W satisfying the Squared Rectangle Theorem.

This also has a generalization which we discuss in Section 9.3.

1.4 Summary of Chapters

In Chapter 2 we prove van der Waerden's theorem with the same proof that van der Waerden gave, though expressed as a color-focusing argument. We then give upper and lower bounds on some van der Waerden numbers. In Chapter 3 we give some applications of van der Waerden's theorem including several proofs of the Square Theorem. In Chapter 4 we prove

POLYVDW. We give Walters' proof and then give upper and lower bounds on some POLYVDW numbers. In Chapter 5 we give some applications of POLYVDW.

In Chapter 6 we prove HJ. We prove it two ways. The classical proof yields insanely large upper bounds on the HJ numbers. The proof by Shelah gives sanely large upper bounds. We discuss both upper and lower bounds for the HJ numbers. In Chapter 7 we give many applications of the HJ, including the square theorem, the Gallai-Witt theorem (also known a the multidimensional VDW), and some lower bounds in communication complexity. In Chapter 8 we prove POLYHJ and then give upper and lower bounds on the POLYHJ numbers. In Chapter 9 we give some applications of POLYHJ including the multidimenional POLYVDW and the generalized POLYVDW.

Chapter 2

Van Der Waerden's Theorem

2.1 Introduction

VDW states that given any number of colors, c, and any length, k, there is a (large) number W so that any c-coloring of the numbers $\{1, 2, \ldots, W\}$ contains a monochromatic k-AP. In this chapter we will prove VDW the same way van der Waerden did; however, we will express it in the color-focusing language of Walters [35].

Van der Waerden's theorem: For all $k, c \in \mathbb{N}$ there exists W such that, for all c-colorings $COL : [W] \to [c]$, there exists $a, d \in \mathbb{N}$, $d \neq 0$, such that

$$COL(a) = COL(a + d) = COL(a + 2d) = \cdots = COL(a + (k - 1)d).$$

Def 2.1.1 Let $k, c \in \mathbb{N}$. $W(k, c)$ is the least W that satisfies VDW. $W(k, c)$ is called a *van der Waerden number*.

Example 2.1.2 Consider the 2-coloring of $\{1, 2, \ldots, 15\}$ given below:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R	R	B	B	R	R	B	B	R	R	R	B	B	B	R

1. There are three RED 3-AP: $\{1, 5, 9\}$, $\{1, 6, 11\}$, $\{2, 6, 10\}$, $\{5, 10, 15\}$.

2. There are three BLUE 3-AP: $\{3, 8, 13\}$, $\{4, 8, 12\}$.

3. There are no monochromatic 4-APs.

Before attempting to prove the full theorem, let's look at a few simple base cases.

- $c = 1$ — $W(k, 1) = k$, because the sequence $1, 2, \ldots, k$ forms a k-AP.
- $k = 1$ — $W(1, c) = 1$, because a 1-AP is any single term.
- $k = 2$ — $W(2, c) = c + 1$, because any 2 terms form a 2-AP.

Alright, not bad so far — we have proven the theorem for a countable number of cases. How many more could there be?

2.2 Proof of Van Der Waerden's Theorem

2.2.1 The First Interesting Case: $W(3, 2)$

We show that there exists a W such that any 2-coloring of $[W]$ has a monochromatic 3-AP. There are easier proofs that give smaller values of W (see Section 2.3.6); however, the technique we use generalizes to other c and k.

For this section let $W \in \mathbb{N}$ and let $COL : [W] \to \{R, B\}$. such that there are no monochromatic 3-APs. We will show a bound on W.

Picture breaking up the numbers $\{1, 2, 3, \ldots, W\}$ into blocks of 5 (we can assume 5 divides W).

$$\{1 < 2 < 3 < 4 < 5\} < \{6 < 7 < 8 < 9 < 10\} < \{11, 12, 13, 14, 15\} < \cdots$$

For $i \in \mathbb{N}$ let B_i be the ith block of 5.

In the next lemma we look at what happens *within* a block.

Lemma 2.2.1 *Let $x \in [W-5]$. Restrict W to $\{x+1, x+2, x+3, x+4, x+5\}$. There exists a, d, $d \neq 0$, such that*

$$a, a + d, a + 2d \in \{x + 1, \ldots, x + 5\}$$

and

$$COL(a) = COL(a + d) \neq COL(a + 2d).$$

Proof:

Look at $COL(x+1), COL(x+2), COL(x+3)$. Two of these have to be the same. We can assume the color is R.

Case 1: $COL(x+1) = COL(x+2) = R$**:** Since there are no monochromatic 3-AP's, $COL(x+3) = B$.

Case 2: $COL(x+1) = COL(x+3) = R$**:** Since there are no monochromatic 3-AP's, $COL(x+5) = B$.

Case 3: $COL(x+2) = COL(x+3) = R$**:** Since there are no monochromatic 3-AP's, $COL(x+4) = B$. ▮

The following figure shows what happens within a block of 5 consective natural numbers.

We need to view $COL : [W] \rightarrow \{R, B\}$ differently. Picture breaking up the numbers $\{1, 2, 3, \ldots, W\}$ into blocks of 5 (we can assume 5 divides W).

$$\{1 < 2 < 3 < 4 < 5\} < \{6 < 7 < 8 < 9 < 10\} < \{11, 12, 13, 14, 15\} < \cdots$$

For $i \in \mathbb{N}$ let B_i be the ith block of 5.

COL can be viewed as assigning to each B_i one of the following: $RRRRR$, $RRRRB$, ..., $BBBBB$. That is,

We view COL as a 32-coloring of the blocks. We will use this change of viewpoint over and over again in this monograph!

Lemma 2.2.2 *Assume $W \geq 5 \times 33 = 165$. There exists two blocks B_i and B_j with $1 \leq i < j \leq 33$ that are colored the same.*

Proof: There are only 32 different ways for a block to be colored and there are 33 blocks, hence two of them must be colored the same. ▮

(See Figure **??**.)

Theorem 2.2.3 *Let $W \geq 325$. Let $COL : [W] \rightarrow [2]$ be a 2-coloring of $[W]$. Then there exists $a, d \in [W]$ such that*

$$\{a, a+d, a+2d\} \subseteq [W],$$

$$COL(a) = COL(a+d) = COL(a+2d).$$

Proof: We take the colors to be RED and BLUE.

Assume, by way of contradiction, that there is no monochromatic 3-AP.

View $[W]$ as being 65 blocks of 5. We denote the blocks

$$B_1, B_2, \cdots B_{65}.$$

By Lemma 2.2.2 there exists two blocks B_i, B_j with $1 \le i < j \le 33$ that are colored the same.

1. Block B_{2j-i} exists since $2j - i \le 65$. Let $D = 2j - i$. Note that B_i, B_j, B_{2j-i} is an arithmetic sequence of blocks.

2. B_i and B_j have a R-R-B pattern in them where the three colors are a 3-AP. Let d be the difference of the this arithmetic sequence.

The following figure shows what is going on.

Look at the space that has a question mark. If it is B then we have a monochromatic 3-AP as shown below.

If it is R then we have a monochromatic 3-AP as shown below.

∎

Note 2.2.4 The proof of Theorem 2.2.3 yields $W(3,2) \le 365$. A more careful proof using blocks of 3 can get $W(3,2) \le 42$. A different technique which we will show in Section 2.3.1 shows $W(3,2) \le 27$. A proof by cases shows that $W(3,2) = 9$.

2.2.2 $W(3,512) \implies W(4,2)$

FILL IN- FILL IN

2.2.3 The Full Proof

We will prove a lemma from which van der Waerden's theorem will follow easily.

Lemma 2.2.5 *Fix $k, c \in N$ with $k > 1$. Assume $(\forall c')[W(k-1,c')$ exists$]$. Then, for all r, there exists $U = U(k,c,r)$ such that for all c-colorings $COL : [U] \to [c]$, one of the following statements holds.*

Statement I: $\exists a, d \in \mathbb{N}, d \neq 0$ *such that*

$$\{a, a+d, a+2d, \ldots, a+(k-1)d)\} \subseteq [U]$$

$$COL(a) = COL(a+d) = COL(a+2d) = \cdots = COL(a+(k-1)d).$$

Statement II: $\exists a, d_1, d_2, \ldots, d_r \in \mathbb{N}, d_i \neq 0\ \forall i$, *such that*

$$\{a, a+d_1, a+2d_1, \ldots, a+(k-1)d_1\} \subseteq [U]$$
$$\{a, a+d_2, a+2d_2, \ldots, a+(k-1)d_2\} \subseteq [U]$$
$$\vdots$$
$$\{a, a+d_r, a+2d_r, \ldots, a+(k-1)d_r\} \subseteq [U]$$

$$COL(a+d_1) = COL(a+2d_1) = \cdots = COL(a+(k-1)d_1)$$
$$COL(a+d_2) = COL(a+2d_2) = \cdots = COL(a+(k-1)d_2)$$
$$\vdots$$
$$COL(a+d_r) = COL(a+2d_r) = \cdots = COL(a+(k-1)d_r)$$

With $COL(a+d_i) \neq COL(a+d_j)$ *when* $i \neq j$. *We refer to* a *as the* **anchor**. *(Informally we are saying that if you c-color* $[U]$ *either you will have a monochromatic* k*-AP or you will have many monochromatic* $(k-1)$*-AP's, all of different colors, and different from* a*. Once "many" is more than* c*, then the latter cannot happen, so the former must, and we have van der Waerden's theorem.)*

Proof:

We define $U(k, c, r)$ to be the least number such that this Lemma holds. We will prove $U(k, c, r)$ exists by giving an upper bound on it.

Base Case: If $r = 1$ then $U = U(k, c, 1) \leq 2W(k-1, c)$. Let $COL : [U] \to [c]$ be a c-coloring of U. Consider COL restricted to the last half of U, which is of size $W(k-1, c)$. By the definition of $W(k-1, c)$ there exists $a' \in \{W(k-1, c), \ldots, 2W(k-1, c)\}$ and $d \in [W(k-1, c)]$ such that

$$\{a', a'+d', a'+2d', \ldots, a'+(k-2)d'\} \subseteq \{W(k-1,c), \ldots, 2W(k-1,c)\},$$

$$COL(a') = COL(a'+d') = COL(a'+2d') = \cdots = COL(a'+(k-2)d').$$

Let $a = a' - d$ and $d_1 = d'$. Clearly

$$COL(a+d_1) = COL(a+2d_1) = COL(a+3d_1) = \cdots COL(a+(k-1)d_1)$$

Note that we have a better bound than $d \in [W(k-1,c]$. We easily have $d \in [\lfloor W(k-1,c)/(k-1) \rfloor]$, though all we need is $d \in [W-1]$. Since $a' \geq [W(k-1,c)]$ and $d_1 \in [W-1]$ we have $a = a' - d \geq 1$.

$d < W(k-1,c)$, so $a' - d \geq 1$. Clearly $a' \leq a \leq U$, so $a' \in [U]$.

The first half of $[U]$ will contain the the anchor, hence (2) holds.

Induction Step: By induction, assume $U(k,c,r)$ exists. We will show that $U(k,c,r+1) \leq 2U(k,c,r)W(k-1, c^{U(k,c,r)})$. Let

$$U = 2U(k,c,r)W(k-1, c^{U(k,c,r)}).$$

Let $COL : [U] \rightarrow [c]$ be an arbitrary c-coloring of $[U]$.

We view $[U]$ as being $U(k,c,r)W(k-1,c^{U(k,c,r)})$ numbers followed by $W(k-1,c^{U(k,c,r)})$ blocks of size $U(k,c,r)$. We denote these blocks by

$$B_1, B_2, \ldots, B_{W(k-1,c^{U(k,c,r)})}.$$

The key point here is that **we view a c-coloring of the second half of $[U]$ as a $c^{U(k,c,r)}$-coloring of these blocks.** Let COL^* be this coloring. By the definition of
$W(k-1,c^{U(k,c,r)})$, we get a monochromatic $(k-1)$-AP of blocks. Hence we have A, D' such that (see Figure **??**)

$$COL^*(B_A) = COL^*(B_{A+D'}) = \cdots = COL^*(B_{A+(k-2)D'})$$

Now, consider block B_A. It is colored by COL. It has length $U(k,c,r)$, which tells us that either (1) or (2) from the lemma holds. If (1) holds — we have a monochromatic k-AP — then we are done. If not, then we have the following: $a', d_1, d_2, \ldots, d_r$ with $a \in B_A$, and

$$\begin{gathered}
\{a' + d_1, a' + 2d_1, \ldots, a' + (k-1)d_1 \subseteq B_A \\
\{a' + d_2, a' + 2d_2, \ldots, a' + (k-1)d_2 \subseteq B_A \\
\vdots \\
\{a' + d_r, a' + 2d_r, \ldots, a' + (k-1)d_r \subseteq B_A
\end{gathered}$$

$$\begin{aligned}
COL(a' + d_1) = COL(a' + 2d_1) &= \cdots = COL(a' + (k-1)d_1) \\
COL(a' + d_2) = COL(a' + 2d_2) &= \cdots = COL(a' + (k-1)d_2) \\
&\vdots \quad \vdots \\
COL(a' + d_r) = COL(a' + 2d_r) &= \cdots = COL(a' + (k-1)d_r)
\end{aligned}$$

where $COL(a'+d_i)$ are all different colors, and different from a' (or else there would already be a monochromatic k-AP). How far apart are corresponding elements in adjacent blocks? Since the blocks viewed as points are D' apart, and each block has $U(k,c,r)$ elements in it, correspoinding elements in adjacent blocks are $D = D' \times U(k,c,r)$ numbers apart. Hence

$$\begin{gathered}
COL(a' + d_1) = COL(a' + D + d_1) = \cdots = COL(a' + (k-2)D + d_1) \\
COL(a' + d_2) = COL(a' + D + d_2) = \cdots = COL(a' + (k-2)D + d_2) \\
\vdots \\
COL(a' + d_r) = COL(a' + D + d_r) = \cdots = COL(a' + (k-2)D + d_r)
\end{gathered}$$

We now note that we have only worked with the second half of $[U]$. Since we know that

$$a > \frac{1}{2}U = U(k,c,r)W(k-1, c^{U(k,c,r)})$$

and

$$D \leq \frac{1}{k-1}U(k,c,r)W(k-1, c^{U(k,c,r)}) \leq U(k,c,r)W(k-1, c^{U(k,c,r)})$$

so we find that $a = a' - D > 0$ and thus $a \in [U]$. The number a is going to be our **new anchor**.

So now we have

$$\begin{array}{c} COL(a+(D+d_1)) = COL(a+2(D+d_1)) = \cdots = COL(a+(k-1)(D+d_1)) \\ COL(a+(D+d_2)) = COL(a+2(D+d_2)) = \cdots = COL(a+(k-1)(D+d_2)) \\ \vdots \\ COL(a+(D+d_r)) = COL(a+2(D+d_r)) = \cdots = COL(a+(k-1)(D+d_r)) \end{array}$$

With each sequence a different color.

We need an $(r+1)$st monochromatic set of points. Consider

$$\{a+D, a+2D, \ldots, a+(k-1)D\}.$$

These are correspoinding points in blocks that are colored (by COL^*) the same, hence

$$COL(a+D) = COL(a+2D) = \cdots = COL(a+(k-1)D)).$$

In addition, since

$$(\forall i)[COL(a') \neq COL(a'+d_i)]$$

the color of this new sequence is different from the r sequences above.

Hence we have $r+1$ monochromatic $(k-1)$-AP's, all of different colors,and all with projected first term a. Formally the new parameters are $a, (D+d_1), \ldots, (D+d_r)$, and D. ∎

Theorem 2.2.6 ***(Van der Waerden's theorem)*** $\forall k, c \in \mathbb{N}, \exists W = W(k,c)$ *such that, for all c-colorings* $COL : [W] \rightarrow [c]$, $\exists a, d \in \mathbb{N}, d \neq 0$ *such that*

$$COL(a) = COL(a+d) = COL(a+2d) = \cdots = COL(a+(k-1)d)$$

Proof:

We prove this by induction on k. That is, we show that

- $(\forall c)[W(1,c)$ exists$]$
- $(\forall c)[W(k,c)$ exists$] \implies (\forall c)[W(k+1,c)$ exists $]$

Base Case: $k = 1$ As noted above $W(1, c) = 1$ suffices. In fact, we also know that $W(2, c) = c + 1$ suffices.
Induction Step: Assume $(\forall c)[W(k-1, c)$ exists]. Fix c. Consider what Lemma 2.2.5 says when $r = c$. In any c-coloring of $U = U(k, c, c)$, either there is a monochromatic k-AP or there are c monochromatic $(k-1)$-AP's which are all colored differently, and a number a whose color differs from all of them. Since there are only c colors, this cannot happen, so we must have a monochromatic k-AP. Hence $W(k, c) \le U(k, c, c)$. ▮

Note that the proof of $W(k, c)$ depends on $W(k-1, c')$ where c' is quite large. Formally the proof is an ordering on the following order on (k, c)

$$(1,1) \prec (1,2) \prec \cdots \prec (2,1) \prec (2,2) \prec \cdots \prec (3,1) \prec (3,2) \cdots$$

This is an ω^2 ordering. It is well founded, so induction works.

2.3 The Van Der Waerden Numbers

2.3.1 Upper Bounds on $W(k, c)$

2.3.2 Lower Bounds on $W(k, c)$

Theorem 2.3.1 *For all k, c, $W(k, c) \ge c^{(k-1)/2}\sqrt{k}$.*

Proof:

We will prove this theorem as though we didn't know the result.

Let W be a number to be picked later. We are going to try to c-color $[W]$ such that there are no monochromatic k-AP's. More precisely, we are going to derive a value of W such that we can show that such a coloring exists.

Consider the following experiment: for each $i \in [W]$ randomly pick a color from $[c]$ for i. The distribution is uniform. What is the probability that a monochromatic k-AP was formed?

First pick the color of the sequence. There are c options. Then pick the value of a. There are W options. Then pick the value of d. Once these are determined, the color of the distinct k values in $\{a, a+d, a+2d, \ldots, a+(k-1)d\}$ are determined. There are $W - k$ values left. Hence the number of such colorings is bounded above by $(cW^2c^{W-k})/k$.

Hence the probability that the c-coloring has a monochromatic k-AP is bounded above by

$$\frac{cW^2c^{W-k}}{kc^W} = \frac{W^2}{kc^{k-1}}.$$

We need this to be < 1. Hence we need

$$W^2 < kc^{k-1}.$$

$$W < c^{(k-1)/2}\sqrt{k}.$$

Therefore there is a c-coloring of $[c^{(k-1)/2}\sqrt{k}-1]$ without a monochromatic k-AP. Hence $W(k,c) \geq c^{(k-1)/2}\sqrt{k}$.

∎

FILL IN- PUT IN MORE THAT IS KNOWN.

2.3.3 Lower Bounds on $W(3,c)$

2.3.4 Three-Free Sets

2.3.5 Applications of Three-Free Sets

2.3.6 Some Exact values for $W(k,c)$

$W(3,2)$ is small enough to find it exactly by hand. Our interest is in getting it with as little brute-force as possible.

Theorem 2.3.2

1. $W(3,2) \geq 9$
2. $W(3,2) \leq 14$
3. $W(3,2) = 9$

Proof:
a) The 2-coloring of [8] defined by

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ R & R & B & B & R & R & B & B \end{array}$$

shows that $W(3,2) \geq 9$.

b) Let COL be 2-coloring of [13]. Two of the numbers in the set $\{5,7,9\}$ must have the same color. Let the numbers be $x < y$ and the color be RED. Note that $x+y$ is even and

$$1 \leq 2x - y < (x+y)/2 < 2y - x \leq 13.$$

1. If $2x - y$ is RED then $2x - y, x, y$ is a monochromatic 3-AP, so we are done.

2. If $(x+y)/2$ is RED then $x, (x+y)/2, y$ is a monochromatic 3-AP, so we are done.

3. If $2y - x$ is RED then $x, y, 2y - x$ is a monochromatic 3-AP, so we are done.

4. If $2x-y$, $(x+y)/2$, and $2y-x$ are all BLUE then they form a monochromatic 3-AP so we are done.

c) Let COL be a 2-coloring of [9]. We can assume that $COL(5) = \text{RED}$. At least one of $\{4,6\}$ is BLUE. By symmetry we can assume that $COL(4) = \text{BLUE}$. Then we know the following:

- At least one of $\{3,7\}$ is BLUE.
- At least one of $\{2,8\}$ is BLUE.
- At least one of $\{1,9\}$ is BLUE.
- At least one of $\{2,6\}$ is RED.
- At least one of $\{1,7\}$ is RED.

This can be used to finish a brute force proof that the coloring must have a monochromatic 3-AP. ∎

Very few of the VDW numbers are known. The following table summarizes all that is known.

VDW number	value	reference
$W(2,3)$	9	Folklore and above
$W(3,3)$	27	Chvátal [5]
$W(3,4)$	76	Brown [4]
$W(4,2)$	35	Chvátal [5]
$W(5,2)$	178	Stevens and Shantarum [29]
$W(6,2)$	1132	Kouril [18]

Chapter 3

Applications of Van Der Waerden's Theorem

3.1 The Square Theorem: First Proof

Use VDW.

3.2 The Square Theorem: Second Proof

We will once again prove the square theorem. We will use the same technique used to prove VDW , Color-Focusing.

Proposition 3.2.1 *Fix $c \in \mathbb{N}$. Then, for any c-coloring $COL : [c+1]^2 \to [c], \exists a, b, d \in \mathbb{N}$ such that*

$$COL(a+d, b) = COL(a, b+d)$$

Proof: Fix c, $COL : [c+1]^2 \to [c]$. Consider the points on the diagonal:

$$(1, c+1), (2, c), \ldots, (i, c+2-i), \ldots, (c+1, 1)$$

There are $c+1$ such points, so by pigeonhole principle we have two of them the same color. Let these be $(x, c+2-x)$ and $(y, c+2-y)$, with $x < y$. Then define $a = x$, $b = c+2-y$, and $d = y-x$. Now we have

$$(a, b+d) = (x, (c+2-y) + (y-x)) = (x, c+2-x)$$

$$(a+d,b) = (x+(y-x), c+2-y) = (y, c+2-y)$$

and we have precisely what we wanted — $COL(a+d,b) = COL(a,b+d)$. ∎

Now this is hardly a square, but it's a start. The points of our square will be $(a,b), (a+d,b), (a,b+d)$, and $(a+d,b+d)$, so at this point we're half-way there! But as is typically the case, the second half of the journey is harder than the first.

Our next lemma will eventually lead to a third point

Lemma 3.2.2 *Fix c, r. $\exists L = L'(c,r)$ such that, for all c-colorings $COL : [L]^2 \rightarrow [c]$, either*
(1) $\exists a, b, d \in \mathbb{N}$, such that $COL(a,b) = COL(a+d,b) = COL(a,b+d)$

OR

(2) $\exists a, b, d_1, d_2, \ldots, d_r \in \mathbb{N}$, such that $COL(a+d_i,b) = COL(a,b+d_i)$ for each i, $COL(a+d_i,b) \neq COL(a+d_j,b)$ when $i \neq j$, and $COL(a,b) \neq COL(a+d_i,b)$ for every i.

Informally, this says that we either have three vertices of a square the same color, or r pairs of diagonal points focusing on the same point (a,b).

Proof: Define $L'(c,r)$ to be the smallest number such that the Lemma is true. We will show $L'(c,r)$ exists by giving an upper bound.

Base case: $r = 1$ — $L'(c,1) = c+1$ works. (2) is satisfied by proposition 3.2.1.

Inductive case: Assume we know $L'(c,r)$ exists. We will show that $L'(c,r+1) \leq L'(c,r) \times (X+1)$, where $X = c^{[L'(c,r)]^2}$ is the number of ways to c-color $[L'(c,r)]^2$. Let $L = L'(c,r) \times (X+1)$. We will view the elements of $[L]^2$ as a $(X+1) \times (X+1)$ lattice of blocks, each of size $L'(c,r)$. Denote the blocks by

$$B_{1,1}, B_{1,2}, \ldots, B_{X+1,X+1}$$

Each block has $[L'(c,r)]^2$ points, so we view COL as an X-coloring of the blocks. Let $COL^* : [L'(c,r)]^2 \rightarrow [c]$ be this coloring. Then, by Proposition 3.2.1 we have $A, B, D \in \mathbb{N}$ such that $COL^*(B_{A+D,B}) = COL^*(B_{A,B+D})$.

Now look at the block $B_{A+D,B}$, which has size $L'(c,r) \times L'(c,r)$. The Lemma applies, so we have two cases.
Case 1: $\exists a,b,d \in \mathbb{N}$ with $COL(a,b) = COL(a+d,b) = COL(a,b+d)$. This is just what we wanted, so we're done!
Case 2: $\exists a,b,d_1,d_2,\ldots,d_r \in \mathbb{N}$ with $COL(a+d_i,b) = COL(a,b+d_i)$, and the colors different for each i, all different from $COL(a,b)$.

Define $D' = D \times L'(c,r), a' = a - D', b' = b$. Then we get

$$COL(a'+D',b') = COL(a',b'+D')$$

$$COL(a'+D'+d_1,b') = COL(a',b'+D'+d_1)$$
$$COL(a'+D'+d_2,b') = COL(a',b'+D'+d_2)$$
$$\vdots$$
$$COL(a'+D'+d_r,b') = COL(a',b'+D'+d_r)$$

If any of these pairs has the same color as (a',b'), then we get our monochromatic L which satisfies (1). If not, define $d_i' = d_i + D'$ for each i up to r, and $d_{r+1} = D'$ and we have exactly the parameters needed to satisfy (2). ∎

From here, we easily reach the real goal:

Theorem 3.2.3 *Fix c. There exists $L = L(c)$ such that, for any c-coloring $COL : [L]^2 \to [c], \exists a,b,d \in \mathbb{N}$ with*

$$COL(a+d,b) = COL(a,b+d) = COL(a+d,b+d)$$

Note that our 3^{rd} point is $(a+d,b+d)$ instead of (a,b). This is essentially the same, but will make picking up the 4^{th} point slightly cleaner.
Proof: We will show that $L = L'(c,c)$ works. Let $COL : [L]^2 \to [c]$ be any c-coloring of $[L]^2$. To flip things around, we define COL' to be a reversing of COL — that is, we define $COL'(x,y) = COL(L-x+1, L-y+1)$. Now COL' is a c-coloring of $[L]^2$, so we may use Lemma 3.2.2. With $r = c$, (2) requires c pairs of points with different colors, and a $(c+1)^{st}$ point colored different from them all. This means $c+1$ colors, which is more than we have. Thus (1) holds, so we get $a,b,d \in \mathbb{N}$ with

$$COL'(a,b) = COL'(a+d,b) = COL'(a,b+d)$$

Now let $a' = L - a + 1$ and $b' = L - b + 1$. Then our original coloring gives us

$$COL(a' + d, b') = COL(a', b' + d) = COL(a' + d, b' + d)$$

∎

Okay, just one more point and we have a square! The rest of the proof will be use the exact same methods to pick up a fourth point.

Lemma 3.2.4 *Fix c, r. $\exists S = S'(c, r)$ such that, for all c-colorings $COL : [S]^2 \rightarrow [c]$, either*
(1) $\exists a, b, d \in \mathbb{N}$, such that $COL(a, b) = COL(a + d, b) = COL(a, b + d) = COL(a + d, b + d)$

OR

(2) $\exists a, b, d_1, d_2, \ldots, d_r \in \mathbb{N}$, such that $COL(a + d_i, b) = COL(a, b + d_i) = COL(a + d_i, b + d_i)$ for each i, $COL(a + d_i, b) \neq COL(a + d_j, b)$ when $i \neq j$, and $COL(a, b) \neq COL(a + d_i, b)$ for every i.

Informally, this says that we either have all four vertices of a square the same color, or r monochromatic L's focusing on the same point (a, b).

Proof: Define $S'(c, r)$ to be the smallest number such that the Lemma is true. We will show $S'(c, r)$ exists by giving an upper bound.

Base case: $r = 1$ — $S'(c, 1) = L(c)$ works. (2) is satisfied by Theorem 3.2.3

Inductive case: Assume we know $S'(c, r)$ exists. We will show that $S'(c, r+1) \leq S'(c, r) \times L(X)$, where $X = c^{[S'(c,r)]^2}$ is the number of ways to c-color $[S'(c, r)]^2$. Let $S = S'(c, r) \times L(X)$. We will view the elements of $[S]^2$ as a $L(X) \times L(X)$ lattice of blocks, each of size $S'(c, r)$. Denote the blocks by

$$B_{1,1}, B_{1,2}, \ldots, B_{L(X),L(X)}$$

Each block has $[S'(c, r)]^2$ points, so we view COL as an X-coloring of the blocks. Let $COL^* : [S'(c, r)]^2 \rightarrow [c]$ be this coloring. Then, by Theorem 3.2.3 we have $A, B, D \in \mathbb{N}$ such that

$$COL^*(B_{A+D,B}) = COL^*(B_{A,B+D}) = COL^*(B_{A+D,B+D})$$

Now look at the block $B_{A+D,B}$, which has size $S'(c, r) \times S'(c, r)$. The Lemma applies, so we have two cases.

Case 1: $\exists a, b, d \in \mathbb{N}$ with

$$COL(a,b) = COL(a+d,b) = COL(a,b+d) = COL(a+d,b+d)$$

This is just what we wanted, so we're done!
Case 2: $\exists a, b, d_1, d_2, \ldots, d_r \in \mathbb{N}$ with

$$COL(a+d_i,b) = COL(a,b+d_i) = COL(a+d_i,b+d_i)$$

and the colors different for each i, all different from $COL(a,b)$.

Define $D' = D \times L'(c,r), a' = a - D', b' = b$. Then we get

$$COL(a'+D',b') = COL(a',b'+D') = COL(a'+D',b'+D')$$

$$COL(a'+D'+d_1,b') = COL(a',b'+D'+d_1) = COL(a'+D'+d_1,b'+D'+d_1)$$

$$COL(a'+D'+d_2,b') = COL(a',b'+D'+d_2) = COL(a'+D'+d_2,b'+D'+d_2)$$

$$\vdots$$

$$COL(a'+D'+d_r,b') = COL(a',b'+D'+d_r) = COL(a'+D'+d_r,b'+D'+d_r)$$

If any of these L's has the same color as (a',b'), then we get our monochromatic square which satisfies (1). If not, define $d_i' = d_i + D'$ for each i up to r, and $d_{r+1} = D'$ and we have exactly the parameters needed to satisfy (2). ∎

Now, finally, we can state and prove our goal.

Theorem 3.2.5 *Fix c. There exists $S = S(c)$ such that, for any c-coloring $COL : [S]^2 \rightarrow [c], \exists a, b, d \in \mathbb{N}$ such that*

$$COL(a,b) = COL(a+d,b) = COL(a,b+d) = COL(a+d,b+d)$$

Proof: We will show that $S = S'(c,c)$ works. Let $COL : [S]^2 \rightarrow [c]$ be any c-coloring of $[S]^2$. We use Lemma 3.2.4. With $r = c$, (2) requires c monochromatic L's, each with a different color, and a $(c+1)^{st}$ point colored different from them all. This means $c+1$ colors, which is more than we have. Thus (1) holds, so we get $a, b, d \in \mathbb{N}$ with

$$COL(a,b) = COL(a+d,b) = COL(a,b+d) = COL(a+d,b+d)$$

which is precisely what we wanted. ∎

3.3 The Square Theorem: Third Proof

Solynosi proof- better bounds.

3.4 Applications to Number Theory

3.5 Hilbert's Cube Lemma

3.6 The Arithmetic Sequence Game

Chapter 4

The Polynomial Van Der Waerden's Theorem

4.1 Introduction

In this Chapter we state and proof a generalization of van der Waerden's theorem known as *the Polynomial Van Der Waerden's Theorem*. We rewrite van der Waerden's theorem with an eye toward generalizing it.
Van Der Waerden's Theorem: *For all $k, c \in \mathbb{N}$ there exists $W = W(k, c)$ such that, for all c-colorings $COL : [W] \to [c]$, there exists $a, d \in [W]$, such that*

- $\{a\} \cup \{a + id \mid 1 \le i \le k - 1\} \subseteq [W]$,
- $\{a\} \cup \{a + id \mid 1 \le i \le k - 1\}$ *is monochromatic.*

In the proof of Lemma 2.2.5 we needed to take some care to make sure that $a \in [W]$. This needed the fact that $d \le U(k, c, r)$, which was obvious and needed no commentary. For the theorems in this Chapter we will not have the relevant d bounded unless we assume it inductively. Hence we will often have the condition $d \in [W]$ or $d \in [U]$ which will help us show $a \in [W]$. The membership of other elements in $[W]$ will be obvious and not need commentary.

Note that van der Waerden Theorem was really about the set of functions $\{id \mid 1 \le i \le k-1\}$. Why this set of functions? Would other sets of functions work? What about sets of polynomials? The following statement is a natural generalization of van der Waerden's theorem; however, it is not true.

False POLYVDW: *Fix $c \in \mathbb{N}$ and $P \subseteq \mathbb{Z}[x]$ finite. Then there exists $W = W(P, c)$ such that, for all c-colorings $COL : [W] \rightarrow [c]$, there are $a, d \in \mathbb{N}$, $d \neq 0$, such that*

- $\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq [W]$,
- $\{a\} \cup \{a + p(d) \mid p \in P\}$ *is monochromatic.*

The above statement is false since the polynomial $p(x) = 2$ and the coloring

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10\cdots \\ R & R & B & B & R & R & B & B & R & R\cdots \end{array}$$

provides a counterexample. Hence we need a condition to rule out constant functions. The condition $(\forall p \in P)[p(0) = 0]$ suffices.

The Polynomial Van Der Waerden Theorem (POLYVDW) *Fix $c \in \mathbb{N}$ and $P \subseteq \mathbb{Z}[x]$ finite, with $(\forall p \in P)[p(0) = 0]$. Then there exists $W = W(P, c)$ such that, for all c-colorings $COL : [W] \rightarrow [c]$, there are $a, d \in [W]$, such that*

- $\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq [W]$,
- $\{a\} \cup \{a + p(d) \mid p \in P\}$ *is monochromatic.*

(When we apply this theorem to coloring $\{s + 1, \ldots, s + W\}$, we will have $d \in [W]$ and $\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq \{s + 1, \ldots, s + W\}$.)

This was proved for $k = 1$ by Furstenberg [**?**] and (independently) Sarkozy [27]. The original proof of the full theorem by Bergelson and Leibman [1] used ergodic methods. A later proof by Walters [35] uses purely combinatorial techniques. We will present an expanded version of Walters' proof.

Note 4.1.1 Do we need the condition $d \in [W]$? For the classical van der Waerden Theorem $d \in [W]$ was obvious since

$$\{a\} \cup \{a + d, \ldots, a + (k - 1)d\} \subseteq [W] \implies d \in [W].$$

For the polynomial van der Waerden's theorem one could have a polynomial with negative coefficients, hence it would be possible to have

$$\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq [W] \text{ and } d \notin [W].$$

For the final result we do not care where d is; however, in order to prove POLYVDW inductively we will need the condition $d \in [W]$.

Note 4.1.2 The condition $(\forall p \in P)[p(0) = 0]$ is strong enough to make the theorem true. There are pairs (P, c) where $P \subseteq \mathbb{Z}[x]$ (that does not satisfy the condition) and $c \in \mathbb{N}$ such that the theorem is true. Classifying which pairs (P, c) satisfy the theorem is an interesting open problem. We investigate this in Section 4.4.

Note 4.1.3 What happens if instead of polynomials we use some other types of functions? See Section 4.5 for a commentary on that.

Recall that VDW was proven by induction on k and c. The main step was showing that if $(\forall c)[W(k,c)$ exists $]$ then $(\forall c)[W(k+1,c)$ exists $]$. To prove POLYVDW we will do something similar. We will assign to every set of polynomials (that do not have a constant term) a type. The types will be ordered. We will then do an induction on the types of polynomials.

Def 4.1.4 Let $n_e, \ldots, n_1 \in \mathbb{N}$. Let $P \subseteq \mathbb{Z}[x]$. *$P$ is of type $(n_e, \ldots, n_1)$* if the following hold:

1. P is finite.
2. $(\forall p \in P)[p(0) = 0]$
3. The largest degree polynomial in P is of degree $\le e$.
4. For all i, $1 \le i \le e$, There are $\le n_i$ different lead coefficients of the polynomials of degree i. Note that there may be many more than n_i polynomials of degree i.

Note 4.1.5

1. Type $(0, n_e, \ldots, n_1)$ is the same as type $(n_e, \ldots, n_1)$.
2. We have no n_0. This is intentional. All the polynomials $p \in P$ have $p(0) = 0$.
3. By convention P will never have 0 in it. For example, if

$$Q = \{x^2, 4x\}$$

then

$$\{q - 4x : q \in Q\}$$

will be $\{x^2 - 4x\}$. We will just omit the 0.

Example 4.1.6

1. The set $\{x, 2x, 3x, 4x, \ldots, 100x\}$ is of type (100).

2. The set
   $\{x^4+17x^3-65x, x^4+x^3+2x^2-x, x^4+14x^3, -x^4-3x^2+12x, -x^4+78x,$
   $x^3-x^2, x^3+x^2, 3x, 5x, 6x, 7x\}$
   is of type $(2,1,0,4)$

3. The set
   $$\{x^4+b_3x^3+b_2x^2+b_1x \mid -10^{10} \le b_1, b_2, b_3 \le 10^{10}\}$$
   is of type $(1,0,0,0)$.

4. If P is of type $(1,0)$ then there exists $b \in \mathbb{Z}$ and $k \in \mathbb{N}$ such that
   $$P \subseteq \{bx^2-kx, bx^2-(k-1)x, \ldots, bx^2+kx\} \cup \{0\}.$$

5. If P is of type $(1,1)$ then there exists $b_2, b_1 \in \mathbb{Z}$, and $k \in \mathbb{N}$ such that
   $$P \subseteq \{b_2x^2-kx, b_2x^2-(k-1)x, \ldots, b_2x^2+kx\} \cup \{b_1x\} \cup \{0\}.$$

6. If P is of type (f,g,h) then there exists $b_3^{(1)}, \ldots, b_3^{(f)} \in \mathbb{Z}$, $b_2^{(1)}, \ldots, b_2^{(g)} \in \mathbb{Z}$, $b_1^{(1)}, \ldots, b_1^{(h)} \in \mathbb{Z}$, $k_1, k_2 \in \mathbb{N}$, T_1 of type (k_1), and T_2 of type (k_2, k_1) such that
   $$\begin{aligned} P \subseteq \quad & \{b_3^ix^3+p(x) \mid 1 \le i \le f, p \in T_2\} \cup \\ & \{b_2^ix^2+p(x) \mid 1 \le i \le g, p \in T_1\} \cup \\ & \{b_1^ix \mid 1 \le i \le h\} \cup \{0\} \end{aligned}$$

7. Let
   $$P = \{2x^2+3x, x^2+20x, 5x, 8x\}.$$
   Let
   $$Q = \{p(x)-8x \mid p \in P\}.$$
   Then
   $$Q = \{2x^2-5x, x^2+12x, -3x, 0\}.$$
   P is of type $(2,2)$ and Q is of type $(2,1)$. Note that the type 'decreases'.

8. Let
$$P = \{2x^2 + 3x, x^2 + 20x, 5x, 8x\}.$$
Let
$$Q = \{p(x) - 8x \mid p \in P\}.$$
Then
$$Q = \{2x^2 - 5x, x^2 + 12x, -3x, \}.$$
P is of type $(2, 2)$ and Q is of type $(2, 1)$. If we did not have out convention of omitting 0 then the type of Q would have been $(2, 2)$. The type would not have gone "down" (in an ordering to be defined later). This is why we have the convention.

9. Let P be of type $(n_e, \ldots, n_i + 1, 0, \ldots, 0)$. Let bx^i be the leading term of some polynomial of degree i in P (note that we are not saying that $bx^i \in P$). Let
$$Q = \{p(x) - bx^i \mid p \in P\}.$$
There are numbers $n_{i-1}, \ldots, n_1$ such that Q is of type $(n_e, \ldots, n_i, n_{i-1}, \ldots, n_1)$. The type is decreasing in an ordering to be defined later.

Def 4.1.7

1. Let $P \subseteq \mathbb{Z}[x]$ such that $(\forall p \in P)[p(0) = 0]$. POLYVDW($P$) means that the following holds:

   For all $c \in \mathbb{N}$, there exists $W = W(P, c)$ such that for all c-colorings $COL : [W] \to [c]$, there exists $a, d \in [W]$ such that
$$\{a\} \cup \{a + p(d) \mid p \in P\} \text{ is monochromatic.}$$
   (If we use this definition on a coloring of $\{s + 1, \ldots, s + W\}$ then the conclusion would have $a \in \{s + 1, \ldots, s + W\}$ and $d \in [W]$.)

2. Let $n_e, \ldots, n_1 \in \mathbb{N}$. POLYVDW$(n_e, \ldots, n_1)$ means that, for all $P \subseteq \mathbb{Z}[x]$ of type $(n_e, \ldots, n_1)$ POLYVDW(P) holds.

3. Let $(n_e, \ldots, n_i, \omega, \ldots, \omega)$ be the e-tuple that begins with $(n_e, \ldots, n_i)$ and then has $i - 1$ ω's.
$$\text{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega)$$

is the statement

$$\bigwedge_{n_{i-1},\ldots,n_1\in\mathbb{N}} \text{POLYVDW}(n_e,\ldots,n_i,n_{i-1},\ldots,n_1).$$

4. POLYVDW is the statement

$$\bigwedge_{i=1}^{\infty} \text{POLYVDW}(\omega,\ldots,\omega)(\ \omega \text{ occurs } i \text{ times}).$$

   Note that POLYVDW is the complete polynomial van der Waerden theorem.

Example 4.1.8

1. The statement POLYVDW(ω) is equivalent to the ordinary van der Waerden's theorem.
2. To prove POLYVDW$(1,0)$ it will suffice to prove POLYVDW(P) for all P of the form

$$\{bx^2 - kx, bx^2 - (k-1)x, \ldots, bx^2 + kx\}.$$

3. Assume that you know

$$\text{POLYVDW}(n_e,\ldots,n_i,\omega,\ldots,\omega)$$

   and that you want to prove

$$\text{POLYVDW}(n_e,\ldots,n_i+1,0,\ldots,0).$$

   Let P be of type $(n_e,\ldots,n_i+1,0\ldots,0)$. Let bx^i be the first term of some polynomial of degree i in P.

   (a) Let

$$Q = \{p(x) - bx^i \mid p \in P\}.$$

   Then there exists $n_{i-1},\ldots,n_1$, such that Q is of type

$$(n_e,\ldots,n_i,n_{i-1},\ldots,n_1).$$

   Since

$$\text{POLYVDW}(n_e,\ldots,n_i,\omega,\ldots,\omega)$$

   holds by assumption, we can assert that POLYVDW(Q) holds.

(b) Let $U \in \mathbb{N}$. Let

$$Q = \{p(x+u) - p(u) - bx^i \mid p \in P, 0 \le u \le U\}.$$

Note $q(0) = 0$ for all $q \in Q$. Then there exists $n_{i-1}, \ldots, n_1$, such that Q is of type

$$(n_e, \ldots, n_i, n_{i-1}, \ldots, n_1).$$

Since

$$\text{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega)$$

holds by assumption, we can assert that POLYVDW(Q) holds.

We will prove the Polynomial van der Waerden's theorem by an induction on a complicated structure. We state the implications we need to prove and then the ordering.

1. POLYVDW(1) follows from the pigeon hole principle.

2. We will show that, for all $n_e, \ldots, n_i \in \mathbb{N}$,

$$\text{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega) \implies \text{POLYVDW}(n_e, \ldots, n_i{+}1, 0, 0, \ldots, 0).$$

   Note that this includes the case

$$\text{POLYVDW}(n_e, \ldots, n_2, n_1) \implies \text{POLYVDW}(n_e, \ldots, n_2, n_1 + 1).$$

The ordering we use is formally defined as follows:

Def 4.1.9 $(n_e, \ldots, n_1) \preceq (m_{e'}, \ldots, m_1)$ if either

- $e < e'$, or
- $e = e'$ and, for some i, $1 \le i \le e$, $n_e = m_e$, $n_{e-1} = m_{e-1}$, …, $n_{i+1} = m_{i+1}$, but $n_i < m_i$.

This is an ω^ω ordering.

Example 4.1.10 We will use the following ordering on types.

$$(1) \prec (2) \prec (3) \prec \cdots$$

$$(1,0) \prec (1,1) \prec \cdots \prec (2,0) \prec (2,1) \prec \cdots \prec (3,0) \cdots \prec$$

$$(1,0,0) \prec (1,0,1) \prec \cdots \prec (1,1,0) \prec (1,1,1) \prec (1,2,0) \prec (1,2,1) \prec$$

$$(2,0,0) \prec \cdots \prec (3,0,0) \prec \cdots (4,0,0) \cdots .$$

4.2 The Proof of the Polynomial Van Der Waerden Theorem

4.2.1 POLYVDW($\{x^2, x^2+x, \ldots, x^2+kx\}$)

Def 4.2.1 Let $k \in \mathbb{N}$.

$$P_k = \{x^2, x^2+x, \ldots, x^2+kx\}.$$

We show POLYVDW(P_k). This proof contains many of the ideas used in the proof of POLYVDW.

We prove a lemma from which POLYVDW(P_k) will be obvious.

Lemma 4.2.2 *For all $k, c, r \in \mathbb{N}$, there exists $U = U(k, c, r)$ such that for all c-colorings $COL : [U] \to [c]$ one of the following Statements holds.*
Statement I: *There exists $a, d \in [U]$, such that*

- $\{a\} \cup \{a+d^2, a+d^2+d, \ldots, a+d^2+kd\} \subseteq [U]$,
- $\{a\} \cup \{a+d^2, a+d^2+d, \ldots, a+d^2+kd\}$ *is monochromatic.*

Statement II: *There exists $a, d_1, \ldots, d_r \in [U]$ such that the following hold.*

- $\{a+d_1^2, a+d_1^2+d_1, \ldots, a+d_1^2+kd_1\} \subseteq [U]$.
  $\{a+d_2^2, a+d_2^2+d_2, \ldots, a+d_2^2+kd_2\} \subseteq [U]$.
  $$\vdots$$
  $\{a+d_r^2, a+d_r^2+d_r, \ldots, a+d_r^2+kd_r\} \subseteq [U]$.
  (The element a is called the anchor*)*

- $\{a + d_1^2, a + d_1^2 + d_1, \ldots, a + d_1^2 + kd_1\}$ *is monochromatic.*

  $\{a + d_2^2, a + d_2^2 + d_2, \ldots, a + d_2^2 + kd_2\}$ *is monochromatic.*

  $$\vdots$$

  $\{a + d_r^2, a + d_r^2 + d_r, \ldots, a + d_r^2 + kd_r\}$ *is monochromatic.*

  With each monochromatic set being colored differently and differently from a. We refer to a as the **anchor**.

Informal notes:

1. *We are saying that if you c-color $[U]$ either you will have a monochromatic set of the form*

   $$\{a\} \cup \{a + d^2, a + d^2 + d, \ldots, a + d^2 + kd\}$$

   or you will have many *monochromatic sets of the form*

   $$\{a + d^2, a + d^2 + d, \ldots, a + d^2 + kd\},$$

   all of different colors, and different from a. Once "many" is more than c, then the latter cannot happen, so the former must, and we have POLYVDW(P).

2. *If we apply this theorem to a coloring of $\{s + 1, \ldots, s + U\}$ then we either have*

   $$d \in [U] \text{ and } \{a\} \cup \{a + d^2 + d, \ldots, a + d^2 + kd\} \subseteq \{s + 1, \ldots, s + U\}.$$

   or

   $$d_1, \ldots, d_r \in [U] \text{ and, for all } i \text{ with } 1 \le i \le r \text{ such that}$$

   $$\{a\} \cup \{a + d_i^2 + d_i, \ldots, a + d_i^2 + kd_i\} \subseteq \{s + 1, \ldots, s + U\}, \text{ and}$$

   $$\{a+d_i^2+d_i, \ldots, a+d_i^2+kd_i\} \subseteq \{s+1, \ldots, s+U\} \text{ monochromatic for each } i.$$

Proof:

We define $U(k, c, r)$ to be the least number such that this Lemma holds. We will prove $U(k, c, r)$ exists by giving an upper bound on it.

Base Case: $r = 1$. $U(k, c, 1) \le W(k_1, c)^2 + W(k + 1, c)$.

Let COL be any c-coloring of $[W(k+1,c)+W(k+1,c)^2]$. Look at the coloring restricted to the last $W(k+1,c)$ elements. By van der Waerden's theorem applied to the restricted coloring there exists

$$a' \in [(W(k+1,c))^2+1,\ldots,(W(k+1,c))^2+W(k+1,c)]$$

and

$$d' \in [W(k+1,c)]$$

such that

$$\{a', a'+d', a'+2d', \ldots, a'+kd'\} \subseteq \{(W(k+1,c))^2+1, \ldots, (W(k+1,c))^2+W(k+1,c)\}.$$

$$\{a', a'+d', a'+2d', \ldots, a'+kd'\} \text{ is monochromatic .}$$

Let the anchor be $a = a' - (d')^2$ and let $d_1 = d'$.

$$\{a', a'+d', a'+2d', \ldots, a'+kd'\} = \{a+d_1^2, a+d_1^2+d_1, \ldots, a+d_1^2+kd_1\} \text{ is monochromatic.}$$

If a is the same color then Statement I holds. If a is a different color then Statement II holds. There is one more issue– do we have

$$a, d_1 \in [(W(k+1,c))^2 + W(k+1,c)]?$$

FILL IN- NEED FIGURE

Since

$$a' \geq (W(k+1,c))^2 + 1$$

and

$$d' \leq W(k+1,c)$$

we have that

$$a \geq (W(k+1,c))^2 + 1 - (W(k+1,c))^2 = 1.$$

Clearly

$$a < a' \leq W(k+1,c) + (W(k+1,c))^2.$$

Hence

$$a \in [W(k+1,c) + (W(k+1,c))^2].$$

Since $d_1 = d' \in [W(k+1,c)]$ we clearly have

$$d_1 \in [W(k+1,c) + (W(k+1,c))^2].$$

Induction Step: Assume $U(k,c,r)$ exists, and let

$$X = W(k + 2U(k,c,r), c^{U(k,c,r)}).$$

(X stands for eXtremely large.)

We show that

$$U(k,c,r+1) \le (X \times U(k,c,r))^2 + X \times U(k,c,r).$$

Let COL be a c-coloring of

$$[(X \times U(k,c,r))^2 + X \times U(k,c,r)].$$

View this set as $(X \times U(k,c,r))^2$ consecutive elements followed by X blocks of length $U(k,c,r)$. Let the blocks be

$$B_1, B_2, \ldots, B_X.$$

Restrict COL to the blocks. Let $COL^* : [X] \to [c^{U(k,c,r)}]$ be the coloring viewed as a $c^{U(k,c,r)}$-coloring of the blocks. By VDW applied to COL^* and the choice of X there exists $A, D' \in [X]$ such that

- $\{A, A + D', \ldots, A + (k + 2U(k,c,r))D'\} \subseteq [X]$,
- $\{B_A, B_{A+D'}, \ldots, B_{A+(k+2U(k,c,r))D'}\}$ is monochromatic. How far apart are corresponding elements in adjacent blocks? Since the blocks viewed as points are D' apart, and each block has $U(k,c,r)$ elements in it, correspoinding elements in adjacent blocks are $D = D' \times U(k,c,r)$ numbers apart.

Consider the coloring of B_A. Since B_A is of size $U(k,c,r)$ either there exists $a, d \in U(k,c,r)$ such that

- $\{a\} \cup \{a + d^2, a + d^2 + d, \ldots, a + d^2 + kd\} \subseteq B_A$,
- $\{a\} \cup \{a + d^2, a + d^2 + d, \ldots, a + d^2 + kd\}$ is monochromatic

in which case Statement I holds so we are done, or there exists
$a' \in B_A$, $d'_1, \ldots, d'_r \in [U(k,c,r)]$
such that

- $\{a' + {d'_1}^2, a' + {d'_1}^2 + d'_1, \ldots, a' + {d'_1}^2 + kd'_1\} \subseteq B_A$
  $\{a' + {d'_2}^2, a' + {d'_2}^2 + d'_2, \ldots, a' + {d'_2}^2 + kd'_2\} \subseteq B_A$
  $$\vdots$$
  $\{a' + {d'_r}^2, a' + {d'_r}^2 + d'_r, \ldots, a' + {d'_r}^2 + kd'_r\} \subseteq B_A$
- $\{a' + {d'_1}^2, a' + {d'_1}^2 + d'_1, \ldots, a' + {d'_1}^2 + kd'_1\}$ is monochromatic.
  $\{a' + {d'_2}^2, a' + {d'_2}^2 + d'_2, \ldots, a' + {d'_2}^2 + kd'_2\}$ is monochromatic.
  $$\vdots$$
  $\{a' + {d'_r}^2, a' + {d'_r}^2 + d'_r, \ldots, a' + {d'_r}^2 + kd'_r\}$ is monochromatic.

  with each monochromatic set colored differently from the others and from a'.

  Since $\{B_A, B_{A+D}, \ldots, B_{A+(k+2U(k,c,r))D}\}$ is monochromatic we also have that, for all j with $0 \le j \le k + 2U(k,c,r)$,

$$\{a'+{d'_1}^2+jD, a'+{d'_1}^2+d'_1+jD, \ldots, a'+{d'_1}^2+kd'_1+jD \mid 0 \le j \le k+2U(k,c,r)\}$$

is monochromatic

$$\{a'+{d'_2}^2+jD, a'+{d'_2}^2+d'_2+jD, \ldots, a'+{d'_2}^2+kd'_2+jD\} \mid 0 \le j \le k+2U(k,c,r)\}$$

is monochromatic

$$\vdots$$

$$\{a'+{d'_r}^2+jD, a'+{d'_r}^2+d'_r+jD, \ldots, a'+{d'_2}^2+kd'_r+jD\} \mid 0 \le j \le k+2U(k,c,r)\}$$

is monochromatic.

with each monochromatic set colored differently from the others and from a', but the same as their counterpart in B_A.

Let the new anchor be $a = a' - D^2$. Let $d_i = D + d_i'$ for all $1 \le i \le r$, and $d_{r+1} = D$. We first show that these parameters work and then show that $a, d_1, \ldots, d_r \in [U(k,c,r+1)]$.

For $1 \le i \le r$ we need to show that

$$\{a + (D + d_i')^2, a + (D + d_i')^2 + (D + d_i'), \ldots, a + (D + d_i')^2 + k(D + d_i')\}$$

is monochromatic. Let $0 \le j \le k$. Note that

$$a+(D+d_i')^2+j(D+d_i') = (a'-D^2)+(D^2+2Dd_i'+d_i'^2)+(jD+jd_i') = a'+d_i'^2+jd_i'+(j+2d_i')D.$$

Notice that $0 \le j+2d_i' \le k+2U(k,c,r)$. Hence $a+d_i^2+jd_i \in B_{A+(j+2d_i')D'}$, the $(j+2d_i')$th block. Since B_A is the same color as $B_{A+(j+2d_i')D'}$,

$$COL(a + d_i^2) = COL(a + d_i^2 + jd_i).$$

So we have that, for all $0 \le i \le r$, for all j, $0 \le j \le k$, the set

$$\{a + d_i^2, a + d_i^2 + d_i, \ldots, a + d_i^2 + kd_i\}$$

is monochromatic for each i. And, since the original sequences were different colors, so are our new sequences. Finally, if $COL(a) = COL(a+d_i^2)$ for some i, then we have $\{a, a+d_i^2, a+d_i^2+d_i, \ldots, a+d_i^2+kd_i\}$ monochromatic, satisfying Statement I. Otherwise, we satisfy Statement II.

We still need to show that $a, d_1, \ldots, d_r \in [X \times U(k,c,r))^2 + X \times U(k,c,r)]$. This is an easy exercise based on the lower bound on a' (since it came from the later $X \times U(k,c,r)$ coordinates) the inductive upper bound on the d_i's, and the upper bound $D \le U(k,c,r)$.

■

Theorem 4.2.3 *For all k,* POLYVDW(P_k)*.*

Proof: We show $W(P_k,c)$ exists by bounding it. Let $U(k,c,r)$ be the function from Lemma 4.2.2. We show $W(P_k,c) \le U(k,c,c)$. If COL is any c-coloring of $[U(k,c,c)]$ then second case of Lemma 4.2.2 cannot happen. Hence the first case must happen, so there exists $a, d \in [U(k,c,c)]$ such that

- $\{a\} \cup \{a + d^2, a + d^2 + d, \ldots, a + d^2 + kd\} \subseteq [U(k,c,c)]$
- $\{a\} \cup \{a + d^2, a + d^2 + d, \ldots, a + d^2 + kd\}$ is monochromatic.

Therefore $W(P_k, c) \le U(k, c, c)$. ▮

Note 4.2.4 The proof of Theorem 4.2.3 used VDW. Hence it used POLYVDW(ω). The proof can be modified to proof POLYVDW$(1, 0)$. So the proof can be viewed as showing that POLYVDW$(\omega) \implies$ POLYVDW$(1, 0)$.

4.2.2 The Full Proof

We prove a lemma from which the implication

$$\mathrm{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega) \implies \mathrm{POLYVDW}(n_e, \ldots, n_i + 1, 0, 0, \ldots, 0)$$

will be obvious.

Lemma 4.2.5 *Let* $n_e, \ldots, n_i \in \mathbb{N}$. *Assume that* POLYVDW$(n_e, \ldots, n_i, \omega, \ldots, \omega)$ *holds. For all* $P \subseteq \mathbb{Z}[x]$ *of type* $(n_e, \ldots, n_i + 1, 0, \ldots, 0)$, *for all* $c \in \mathbb{N}$, *for all* r, *there exists* $U = U(P, c, r)$ *such that for all* c*-colorings* $COL : [U] \to [c]$ *one of the following Statements holds.*
Statement I: *there exists* $a, d \in [U]$, *such that*

- $\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq [U]$.
- $\{a\} \cup \{a + p(d) \mid p \in P\}$ *is monochromatic.*

Statement II: *there exists* $a, d_1, \ldots, d_r \in [U]$ *such that the following hold.*

- $\{a + p(d_1) \mid p \in P\} \subseteq [U]$
  $\{a + p(d_2) \mid p \in P\} \subseteq [U]$
  $\vdots$
  $\{a + p(d_r) \mid p \in P\} \subseteq [U]$
  (The number a is called **the anchor***)*
- $\{a + p(d_1) \mid p \in P\}$ *is monochromatic*
  $\{a + p(d_2) \mid p \in P\}$ *is monochromatic*
  $\vdots$
  $\{a + p(d_r) \mid p \in P\}$ *is monochromatic*

With each monochromatic set being colored differently and differently from a.

Informal notes:

1. *We are saying that if you c-color* $[U]$ *either you will have a monochromatic set of the form*

$$\{a\} \cup \{a + p(d) \mid p \in P\}$$

or you will have many monochromatic sets of the form

$$\{a + p(d) \mid p \in P\},$$

all of different colors, and different from a. *Once "many" is more than* c, *then the latter cannot happen, so the former must, and we have*

$$\text{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega) \implies \text{POLYVDW}(n_e, \ldots, n_i{+}1, 0, \ldots, 0).$$

2. *If we apply this theorem to a coloring of* $\{s + 1, \ldots, s + U\}$ *then we either have*

$$d \in [U] \textit{ and } \{a\} \cup \{a + p(d) \mid p \in P\} \subseteq \{s + 1, \ldots, s + U\}$$

or

$$d_1, \ldots, d_r \in [U] \textit{ and, for all } i \textit{ with } 1 \le i \le r$$

$$\{a\} \cup \{a + p(d_i) \mid p \in P\} \subseteq \{s + 1, \ldots, s + U\}.$$

Proof: We define $U(P, c, r)$ to be the least number such that this Lemma holds. We will prove $U(P, c, r)$ exists by giving an upper bound on it. In particular, for each r, we will bound $U(P, c, r)$. We will prove this theorem by induction on r.

One of the fine points of this proof will be that we are careful to make sure that $a \in [U]$. The fact that we have inductively bounded the d_i's will help that.

Fix $P \subseteq \mathbb{Z}[x]$ of type $(n_e, \ldots, n_i + 1, 0, \ldots, 0)$. Fix $c \in \mathbb{N}$. We can assume P actually has $n_i + 1$ lead coefficents for degree i polynomials (else P is of smaller type and hence POLYVDW(P, c) already holds and the lemma is true). In particular there exists some polynomial of degree i in P. Let bx^i

be the first term of some polynomial of degree i in P. We will assume that $b > 0$. The proof when $b < 0$ is very similar.
Base Case: $r = 1$. Let

$$Q = \{p(x) - bx^i \mid p \in P\}.$$

It is easy to show that there exists $n_{i-1}, \ldots, n_1$ such that Q is of type $(n_e, \ldots, n_i, n_{i-1}, \ldots, n_1)$, and that $(\forall q \in Q)[q(0) = 0]$. By the assumption that POLYVDW$(n_e, \ldots, n_i, \omega, \ldots, \omega)$ is true, POLYVDW(Q) is true. Hence $W(Q, c)$ exists.

We show that

$$U(P, c, 1) \leq bW(Q, c)^i + W(Q, c).$$

Let COL be any c-coloring of $[bW(Q, c)^i + W(Q, c)]$. Look at the coloring restricted to the last $W(Q, c)$ elements. By POLYVDW(Q) applied to the restricted coloring there exists $a' \in \{bW(Q, c)^i + 1, \ldots, bW(Q, c)^i + W(Q, c)\}$ and $d' \in [W(Q, c)]$ such that

$$\{a'\} \cup \{a' + q(d') \mid q \in Q\} \subseteq \{bW(Q, c)^i + 1, \ldots, bW(Q, c)^i + W(Q, c)\}$$

$$\{a'\} \cup \{a' + q(d') \mid q \in Q\} \text{ is monochromatic .}$$

(Note- we will only need that $\{a' + q(d') \mid q \in Q\}$ is monochromatic.)

Let the new anchor be $a = a' - b(d')^i$. Let $d_1 = d'$. (We will use $b > 0$ later to show that $a \in [U(P, c, 1) \leq bW(Q, c)^i + W(Q, c)]$.)

Then

$$\begin{aligned}
\{a' + q(d') \mid q \in Q\} &= \{a' + p(d') - b(d')^i \mid p \in P\} \\
&= \{(a' - b(d_1)^i) + p(d_1) \mid p \in P\} \\
&= \{a + p(d_1) \mid p \in P\} \text{ is monochromatic.}
\end{aligned}$$

If a is the same color then Statement I holds. If a is a different color then Statement II holds. There is one more issue– do we have $a, d \in [U(P, c, 1)]$?

Since

$$a' \geq bW(Q, c)^i + 1$$

and

$$d' \leq W(Q, c) \text{ (Recall that POLYVDW has the restriction } d \in [W].)$$

we have that

$$a = a' - b(d')^i \geq bW(Q,c)^i + 1 - bd(d')^i \geq bW(Q,c)^i + 1 - bW(Q,c)^i = 1$$

Clearly

$$a < a' \leq bW(Q,c)^i + W(Q,c)$$

Hence

$$a \in [bW(Q,c)^i + W(Q,c)].$$

Since $d_1 = d' \in [W(Q,c)]$ we clearly have

$$d_1 \in [bW(Q,c)^i + W(Q,c)].$$

Induction Step: Assume $U(P,c,r)$ exists. Let

$$Q = \{p(x+u) - p(u) - bx^i \mid p \in P, 0 \leq u \leq U(P,c,r)\}.$$

Note that

$$\{p(x) - bx^i \mid p \in P\} \subseteq Q.$$

Clearly $(\forall q \in Q)[q(0) = 0]$. It is an easy exercise to show that, there exists $n_i, \ldots, n_1$ such that Q is of type $(n_e, \ldots, n_{i+1}, n_i, \ldots, n_1)$.

Now, let

$$Q' = \left\{ \frac{q(x \times U(P,c,r))}{U(P,c,r)} \mid q \in Q \right\}$$

Since every $q \in Q$ is an integer polynomial with $q(0) = 0$, it follows that $U(P,c,p)$ divides $q(xU(P,c,r))$, so we have $Q' \subseteq \mathbb{Z}[x]$. Moreover, it's clear that Q' has the same type as Q.

Since POLYVDW$(n_e, \ldots, n_i, \omega, \ldots, \omega)$ holds, we have POLYVDW(Q').

Hence $(\forall c')[W(Q',c')$ exists]. We show that

$$U(P,c,r+1) \leq b\left(U(P,c,r)W(Q',c^{U(P,c,r)})\right)^i + U(P,c,r)W(Q',c^{U(P,c,r)}).$$

(Note that we are using $b > 0$) here.)

Let COL be a c-coloring of

$$\left[b\left(U(P,c,r)W(Q',c^{U(P,c,r)})\right)^i + U(P,c,r)W(Q',c^{U(P,c,r)})\right].$$

View this set as $b\left(U(P,c,r)W(Q',c^{U(P,c,r)})\right)^i$ elements followed by $W(Q',c^{U(P,c,r)})$ blocks of size $U(P,c,r)$ each. Restrict COL to the blocks. Now **view the restricted c-coloring of numbers as a $c^{U(P,c,r)}$-coloring of blocks.** Call this coloring COL^*. Let the blocks be

$$B_1, B_2, \ldots, B_{W(Q',c^{U(P,c,r)})}.$$

By the definition of $W(Q',c^{U(P,c,r)})$ applied to COL^*, and the assumption that POLYVDW$(n_e,\ldots,n_i,\omega,\ldots,\omega)$ holds, there exists $A, D' \in [W(Q',c^{U(P,c,r)})]$ such that

$$\{B_{A+q'(D')} \mid q' \in Q'\} \text{ is monochromatic.}$$

Note that we are saying that the blocks are the same color. Let $D = D' \times U(P,c,r)$ be the distance between corresponding elements of the blocks. Because each block is length $U(P,c,r)$, if we have an element $x \in B_A$, then in block $B_{A+q'(D')}$ we have a point x', where

CHECK NORMAL VDW WITH THIS POINT ABOUT BLOCKS

FILL IN - NEED FIGURE

$$\begin{aligned}
x' &= x + q'(D')U(P,c,r)\\
&= x + q'\left(\frac{D}{U(P,c,r)}\right)U(P,c,r)\\
&= x + q(D) \text{ for some } q \in Q, \text{ by definition of } Q'
\end{aligned}$$

This will be very convenient.

Consider the coloring of B_A. Since B_A is of size $U(P,c,r)$ one of the following holds.

I) There exists $a \in B_A$ and $d \in [U(P,c,r)]$ such that

- $\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq B_A$
- $\{a\} \cup \{a + p(d) \mid p \in P\}$ is monochromatic (so we are done).

II) There exists $a' \in B_A$ (so $a' \geq bW(Q',c^{U(P,c,r)})^i + 1$) and $d'_1,\ldots,d'_r \in [U(P,c,r)]$ such that

- $\{a' + p(d'_1) \mid p \in P\} \subseteq B_A$

  $\{a' + p(d'_2) \mid p \in P\} \subseteq B_A$

$\vdots$

$\{a' + p(d'_r) \mid p \in P\} \subseteq B_A$

- $\{a' + p(d'_1) \mid p \in P\}$ is monochromatic

  $\{a' + p(d'_2) \mid p \in P\}$ is monochromatic

  $\vdots$

  $\{a' + p(d'_r) \mid p \in P\}$ is monochromatic

  with each monochromatic set being colored differently from each other and from a'.

Since $\{B_{A+q'(D')} \mid q' \in Q'\}$ is monochromatic, and since we know that $x \in B_A$ corresponds to $x + q(D) \in B_{A+q'(D')}$, we discover that, for all $q \in Q$,

$$\begin{array}{c} \{a' + p(d'_1) + q(D) \mid p \in P\} \text{ is monochromatic} \\ \{a' + p(d'_2) + q(D) \mid p \in P\} \text{ is monochromatic} \\ \vdots \\ \{a' + p(d'_r) + q(D) \mid p \in P\} \text{ is monochromatic.} \end{array}$$

with each monochromatic set being colored differently from each other, and from a', but the same as their counterpart in B_A.

Our new **anchor** is $a = a' - bD^i$. Note that since

$$a' \geq bW(Q', c^{U(P,c,r)})^i + 1$$

and

$$D \leq W(Q', c^{U(P,c,r)})$$

we have

$$a = a' - bD^i \geq bW(Q', c^{U(P,c,r)})^i + 1 - bW(Q', c^{U(P,c,r)})^i = 1$$

Clearly $a \leq a' \leq bW(Q', c^{U(P,c,r)} + U(P,c,r)W(Q', c^{U(P,c,r)})$. Hence

$$a \in [bW(Q', c^{U(P,c,r)})^i + U(P,c,r)W(Q', c^{U(P,c,r)})].$$

Since

$$\{B_{A+q'(D')} \mid q' \in Q'\}$$

is monochromatic (viewing the coloring on blocks) we know that

$$\{a' + q(D) \mid q \in Q\}$$

is monochromatic (viewing the coloring on numbers). Remember that the following is a subset of Q:

$$\{p(x) - bx^i \mid p \in P\}.$$

Hence the following set is monochromatic:

$$\begin{aligned}\{a' + p(D) - bD^i \mid p \in P\} &= \{a + bD^i + p(D) - bD^i \mid p \in P\} \\ &= \{a + p(D) \mid p \in P\}.\end{aligned}$$

If a is the same color then Statement I holds and we are done. If a is a different color then we have one value of d, namely $d_{r+1} = D$. We seek r additional ones to show that Statement II holds.

For each i we want to find a new d_i that works with the new anchor a. Consider the monochromatic set $\{a' + p(d_i') \mid p \in P\}$. We will take each element of it and shift it $q(D)$ elements for some $q \in Q$. The resulting set is still monochromatic. We will pick $q \in Q$ carefully so that the resulting set, together with the new anchor a and the new values $d_i = d_i' + D$ work.

CHECK VDW AND QVDW FOR THIS POINT

For each $p \in P$ we want to find a $q \in Q$ such that $a + p(d_i' + D)$ is of the form $a' + p(d_i') + q(D)$, and hence the color is the same as $a' + p(d_i')$.

$$\begin{aligned}a' + p(d_i') + q(D) &= a + p(d_i' + D) \\ a' + p(d_i') + q(D) - a &= p(d_i' + D) \\ bD^i + p(d_i') + q(D) &= p(d_i' + D) \\ q(D) &= p(d_i' + D) - p(d_i') - bD^i\end{aligned}$$

Take $q(x) = p(x + d_i') - p(d_i') - bD^i$. Note that $d_i' \leq U(Q, c, r)$ so that $q \in Q$.

— Put bounds on d_i in here.

FILL IN- CHECK THIS

Let $d_i = d_i' + D$ for $1 \leq i \leq r$, and $d_{r+1} = D$.

We have seen that

$$\{a + p(d_1) \mid p \in P\} \text{ is monochromatic}$$

$$\vdots$$

$$\{a + p(d_r) \mid p \in P\} \text{ is monochromatic}$$

AND

$$\{a + p(d_{r+1}) \mid p \in P\} \text{ is monochromatic}$$

The first r are guaranteed to be different colors by the inductive assumption. The $(r+1)^{st}$ is yet another color, because it shares a color with the anchor of our original sequences, which we assumed had its own color. So here we see that the Lemma is satisfied with parameters $a, d_1, \ldots, d_r, d_{r+1}$. ∎

Lemma 4.2.6 *For all $n_e, \ldots, n_i$*

$$\text{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega) \implies \text{POLYVDW}(n_e, \ldots, n_i{+}1, 0, 0, \ldots, 0).$$

Proof: Assume POLYVDW$(n_e, \ldots, n_i, \omega, \ldots, \omega)$. Let P be of type POLYVDW$(n_e, \ldots, n_i + 1, 0, 0, \ldots, 0)$. Apply Lemma 4.2.5 to P with $r = c$. Statement II cannot hold, so statement I must, and we are done. ∎

We can now prove the Polynomial van der Waerden theorem.

Theorem 4.2.7 *For all $P \subseteq \mathbb{Z}[x]$ finite, such that $(\forall p \in P)[p(0) = 0]$, for all $c \in \mathbb{N}$, there exists $W = W(P, c)$ such that for all c-colorings $COL : [W] \to [c]$, there exists $a, d \in [W]$ such that*

- $\{a\} \cup \{a + p(d) \mid p \in P\} \subseteq [W]$,
- $\{a\} \cup \{a + p(d) \mid p \in P\}$ *is monochromatic.*

Proof:

We use the ordering from Definition 4.1.9. The least element of this set is (0). POLYVDW(0) is the base case. The only sets of polynomials of type (0) are $\emptyset$. For each of these sets, the Polynomial van der Waerden theorem requires only one point to be monochromatic (the anchor), so of course POLYVDW (0) holds.

Lemma 4.2.5 is the induction step.

This proves the theorem.

∎

Note 4.2.8

1. Our proof of POLYVDW did not use van der Waerden's Theorem. The base case for POLYVDW was POLYVDW(0) which is trivial.

2. Let $p(x) = x^2 - x$ and $P = \{p(x)\}$. Note that $p(1) = 0$. The statement POLYVDW$(P, 1870)$ is true but stupid: if COL is an 1870-coloring of $[1]$ then let $a = 0$ and $d = 1$. Then $a, a + p(d)$ are the same color since they are the same point. Hence POLYVDW$(P, 1870)$ holds but is stupid. The proof of POLYVDW we gave can be modified to obtain a d so that not only is $d \neq 0$ but

$$\{a\} \cup \{a + p(d) \mid p \in P\}$$

   has all distinct elements. Once this is done POLYVDW$(P, 1870)$ is true in a way that is not stupid.

4.3 Bounds on the Polynomial Van Der Waerden Numbers

4.3.1 Upper Bounds

4.3.2 Upper Bounds via Alternative Proofs

4.3.3 Lower Bounds

Theorem 4.3.1 *Let $P \subseteq \mathbb{Z}[x]$ be a set of $k-1$ polynomials with 0 constant term. Assume that there is no positive integer for which any pair assumes the same value. For all c,* POLYVDW$(P, c) \geq c^{(k-1)/2}$.

Proof: We will prove this theorem as though we didn't know the result.

Let W be a number to be picked later. We are going to try to c-color $[W]$ such that there is no a, d with $\{a\} \cup \{a + p(d) : p \in P\}$ monochromatic. More precisely, we are going to derive a value of W such that we can show that such a coloring exists.

Consider the following experiment: for each $i \in [W]$ randomly pick a color from $[c]$ for i. The distribution is uniform. What is the probability that an a, d exist such that $\{a\} \cup \{a + p(d) : p \in P\}$ is monochromatic?

The number of colorings is c^W. We now find the number of colorings that have such an a, d.

First pick the color of the sequence. There are c options. Then pick the value of a. There are W options. Then pick the value of d. Note that we are using the version of the POLYVDW where $d \in [W]$, so there are W options. Once these are determined, the color of the distinct k values in $\{a\} \cup \{a + p(d) : p \in P\}$ are determined (they are distinct because of the assumption in the premise of this theorem.) There are $W - k$ values left. Hence the number of such colorings is bounded above by cW^2c^{W-k}.

The probability that the c-coloring has a monochromatic k-AP is bounded above by

$$\frac{cW^2c^{W-k}}{c^W} = \frac{W^2}{c^{k-1}}.$$

We need this to be < 1. Hence we need

$$W^2 < c^{k-1}.$$

$$W < c^{(k-1)/2}.$$

Therefore there is a c-coloring of $[c^{(k-1)/2} - 1]$ without a monochromatic k-AP. Hence POLYVDW$(P, c) \geq c^{(k-1)/2}$. ∎

We actually obtained a better bound in Theorem 2.3.1 when dealing with ths ordinary VDW. This is because we knew more about the actual polynomials involved. Below we obtain better bounds for particular sets of polynomials.

Theorem 4.3.2 *Let $c, k \in \mathbb{N}$. Let $P = \{x, x^2, \ldots, x^k\}$.* POLYVDW$(P, c) \geq$

Proof: We will prove this theorem as though we didn't know the result.

Let W be a number to be picked later. We are going to try to c-color $[W]$ such that there is no a, d with $\{a\} \cup \{a + d^j : 1 \leq j \leq k\}$ monochromatic. More precisely, we are going to derive a value of W such that we can show that such a coloring exists.

Consider the following experiment: for each $i \in [W]$ randomly pick a color from $[c]$ for i. The distribution is uniform. What is the probability that an a, d exist such that $\{a\} \cup \{a + p(d) : p \in P\}$ is monochromatic?

The number of colorings is c^W. We now find the number of colorings that have such an a, d.

First pick the color of the sequence. There are c options. Then pick the value of a. There are W options. Then pick the value of d. Note that we need to have $a + d^k \in [W]$. Hence $d \leq W^{1/k}$, so there are $W^{1/k}$ options. Once these are determined, the color of the distinct k values in $\{a\} \cup \{a + d^j : 1 \leq j \leq k\}$ are determined There are $W - k$ values left. Hence the number of such colorings is bounded above by $cW^{1+1/k}c^{W-k}$.

The probability that the c-coloring has a monochromatic k-AP is bounded above by

$$\frac{cW^{1+1/k}c^{W-k}}{c^W} = \frac{W^{1+1/k}}{c^{k-1}}.$$

We need this to be < 1. Hence we need

$$W^{1+1/k} < c^{k-1}.$$

$$W < c^{(1-\epsilon)k} \text{ where } \epsilon = \tfrac{2}{k+1} .$$

Therefore there is a c-coloring of $[c^{(1-\epsilon)k} - 1]$ without such an a, d. Hence $\mathrm{POLYVDW}(P, c) \geq c^{(1-\epsilon)k} - 1$ ▮

Better bounds are known. See [30] and [22]

FILL IT IN- ADD MORE REFS AND POSSIBLY PROOFS

4.4 What if we use Polynomials with a Constant term?

4.5 What if we do not use Polynomials?

The POLYVDW was motivated by replacing $d, 2d, \ldots, (k-1)d$ with polynomials in d. Would other functions work? Would exponential functions work? For which choice of $b, c \in \mathbb{N}$ is the following true:

for every c-coloring COL of $\mathbb{N}$ there exists $a, d \in \mathbb{N}$ such that with

$$COL(a) = COL(a + b^d)$$

Alas, this is never true.

Theorem 4.5.1 *Fix $b \in \mathbb{N}$. Let p be the smallest prime number which is not a factor of b, Then there is a p-coloring $COL : \mathbb{N} \rightarrow [p]$ such that, $\forall a, d \in \mathbb{N}, COL(a) \neq COL(a + b^d)$.*

Proof: Fix $b, p \in \mathbb{N}$ with p the smallest prime non-factor of b. Now define the p-coloring $COL : \mathbb{N} \rightarrow [p]$ such that $COL(n) = n'$, where n' is the reduction of n modulo p with $n' \in [p]$. Most importantly, $COL(n) \equiv n \pmod{p}$. Thus, $COL(a) = COL(b)$ if and only if $p \mid (b - a)$.

Now let $a, d \in \mathbb{N}$, and consider $COL(a)$ and $COL(a + b^d)$. Well, since p is prime and $p \nmid b$, we know that $p \nmid b^d$. This guarantees that $COL(a) \neq COL(a + b^d)$, which is what was to be shown. ▮

It is an open question to determine if Theorem 4.5.1 is tight. Also, it is open to investigate other functions.

Chapter 5

Applications of the Polynomial Van Der Waerden Theorem

FILL IN- VDW where the d has to be a square, or some other poly,

Chapter 6

The Hales-Jewett Theorem

6.1 Introduction

HJ feels very much like VDW, despite living in a very different domain. In the case of HJ, we replace $[W]$ with a hypercube, and the arithmetic sequences with monochromatic lines, but it will feel very similar. Here's the cast of players in HJ:

- The hypercube — Given $c, t, N \in \mathbb{N}$, we will color the elements of the N-dimensional hypercube of length t — namely $[t]^N$.

  When $t = 26$, we can look at $[t]^N$ as strings of letters. For example, PUPPY and TIGER are points in $[26]^5$.

- The lines — In $[t]^N$, a *line* is a collection of points $P_1, P_2, \ldots, P_t$ such that $\exists \lambda \subseteq [N], \lambda \neq \emptyset$ satisfying

  $(\forall s \in \lambda)(\forall i)[P_i^s = s \text{ and } \forall s \notin \lambda, \forall i, j, P_i^s = P_j^s]$(See Example below)

  where P_i^s denotes the s^{th} component of the point P_i. We call λ the "moving" coordinates, and the rest are static.

  Example 6.1.1 The following form a line in $[26]^9$, with $\lambda = \{2, 3, 5, 8\}$:

  GAABARDAA

  GBBBBRDBA

  $\vdots$

  GZZBZRDZA

- The line$^-$. A *line*$^-$ is the first $t-1$ points of a line in $[t]^N$. The line$^-$ corresponding to the previous example is

$$\begin{array}{c}\text{GAABARDAA}\\ \text{GBBBBRDBA}\\ \vdots\\ \text{GYYBYRDYA}\end{array}$$

  Given a line L, we will refer to L^- as *the line*$^-$ *corresponding to* L.

- Completion — the would-be t^{th} point of a line$^-$. The completion of our line$^-$ is the point GZZBZRDZ. If more than one point would complete the line, we choose the least such point, according to a lexicographical ordering of $[t]^N$.

  Note 6.1.2 When $t \le 2$, a line$^-$ may have more than one completion, since in that case a line$^-$ is a single point. For example, $\{$BAA$\}$ is a line$^-$ in $[2]^3$. Its completions are BAB, BBA, and BBB, depending on our choice of moving coordinates. However, when $t \ge 3$, a line$^-$ will have at least 2 points, which establishes the set of moving coordinates, and thus the completion of the line. This means, when $t \ge 3$, every line$^-$ has a unique, predetermined t^{th} point. The definition's use of the "least" t^{th} point only matters when $t \le 2$

We are now ready to present HJ .

Hales-Jewett theorem $\forall t, c, \exists N = HJ(t,c)$ such that, for all c-colorings $COL : [t]^N \to [c], \exists L \subseteq [t]^N, L$ a monochromatic line.

There are some easy base cases:

Fact 6.1.3

- $c = 1$ — $HJ(t,1) = 1$*; Any 1-coloring of* $[t]^1 = [t]$ *easily has a monochramtic line. For example, if we 1-color* $[4]$ *we have that (1), (2), (3), (4) are all* RED *and they form a line.*

- $t = 1$ — $HJ(1, c) = 1$*; when* $t = 1$ *there is only a single point.*

There is also a slightly harder base case:

Proposition 6.1.4 $HJ(2, c) = c + 1$

Proof:

Let $COL : [2]^{c+1} \rightarrow [c]$ be a c-coloring of $[2]^{c+1}$. Consider the following elements of $[2]^{c+1}$

$$\begin{array}{cccccc}
1 & 1 & 1 & \cdots & 1 & 1 \\
1 & 1 & 1 & \cdots & 1 & 2 \\
1 & 1 & 1 & \cdots & 2 & 2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1 & 2 & 2 & \cdots & 2 & 2 \\
2 & 2 & 2 & \cdots & 2 & 2
\end{array}$$

Since there are $c+1$ elements and only c colors, two of these elements are the same color. We can assume they are of the form

$1^i 2^j$ where $i + j = c + 1$

$1^{i'} 2^{j'}$ where $i' + j' = c + 1$

These two elements form a monochromatic line. (For example

$$\begin{array}{ccccccccc}
1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 \\
1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2
\end{array}$$

form a monochromatic line with $\lambda = \{5, 6\}$.) ∎

We will give two proof of the HJ: (1) The original proof due to Hales and Jewett [15], presented as a color-focusing argument, and (2) a proof due to Shelah [28] and yields much better upper bounds on the Hales-Jewett Numbers.

6.2 Proof of the Hales-Jewett Theorem

We prove a lemma from which the theorem will easily follow.

Lemma 6.2.1 *Fix* $t, c \in \mathbb{N}$*. Assume* $(\forall c')[HJ(t-1, c')$ *exists*]*. Then, for all* r*, there exists* $N = U(t, c, r)$ *such that, for all c-colorings* $COL : [t]^N \rightarrow [c]$ *one of the following statements holds.*

Statement I: *There exists a monochromatic line* $L \subseteq [t]^N$.

Statement II: *There exist* r *monochromatic lines*$^-$ $L_1^-, L_2^-, \ldots, L_r^- \subseteq [t]^N$, *and a point* $Q \in [t]^N$, *such that each* L_i^- *has a different color,* Q *is yet another color, and* Q *is the completion of every* L_i^-. *(Informally, we say that if you c-color* $[t]^N$ *then you will either have a monochromatic line, or many monochromatic line*$^-$ *structures, each of a different color. Once "many" becomes more than c, we must have a monochromatic line.)*

Proof:

We define $U(t,c,r)$ to be the least number such that this Lemma holds. We will prove $U(t,c,r)$ exists by giving it an upper bound on it.

Base Case: If $r = 1$ then $U(t,c,1) \le HJ(t-1,c)$ suffices (actually $U(t,c,1) = HJ(t-1,c)$). We take any c-coloring of $[t]^{HJ(t-1,c)}$, and restrict the domain to a c-coloring of $[t-1]^{HJ(t-1,c)}$ to find a monochromatic line, which it has by definition of HJ. This becomes a monochromatic line$^-$ in $[t]^{HJ(t-1,c)}$, so we are done.

Induction Step: By induction, assume $U(t,c,r)$ exists. Let

$X = c^{t^{U(t,c,r)}}$. This is the number of ways to c-color $[t]^{U(t,c,r)}$.

(X stands for eXtremely large.)

We will show that

$$U(t,c,r+1) \le HJ(t-1,X) + U(t,c,r).$$

Let $N = HJ(t-1,X) + U(t,c,r)$. Now we view $[t]^N$ as

$$[t]^{HJ(t-1,X)} \times [t]^{U(t,c,r)}.$$

Define $S = \{\chi \mid \chi : [t]^{U(t,c,r)} \to [c]\}$. Note that $|S| = X$. How convenient.

Let $COL : [t]^N \to [c]$ be our c-coloring. We define, for each $\sigma \in [t-1]^{HJ(t-1,X)}$,

$$COL'(\sigma) : [t-1]^{HJ(t-1,X)} \to S.$$

Note 6.2.2 At this point, it is essential to realize that COL' is an X-coloring of $[t-1]^{HJ(t-1,X)}$. With every vector in $[t-1]^{HJ(t-1,X)}$, we associate some $\chi \in S$. Although χ is itself a *coloring*, here we treat it as a *color*.

For example, $COL(\sigma)$ might be the following 3-coloring of $[2]^3$

$$COL'(\sigma)(0,0,0) = 1$$

$$COL'(\sigma)(0,0,1) = 1$$
$$COL'(\sigma)(0,1,0) = 3$$
$$COL'(\sigma)(0,1,1) = 2$$
$$COL'(\sigma)(1,0,0) = 1$$
$$COL'(\sigma)(1,0,1) = 3$$
$$COL'(\sigma)(1,1,0) = 2$$
$$COL'(\sigma)(1,1,1) = 2$$

Given $\sigma \in [t-1]^{HJ(t-1,X)}$, $COL'(\sigma)$ will be a c-coloring of $[t]^{U(t,c,r)}$. Accordingly, we define COL' by telling the color of $COL'(\sigma)(\tau)$ for $\tau \in [t]^{U(t,c,r)}$. From here, our choice is clear — we associate to σ the c-coloring $COL'(\sigma) : [t]^{U(t,c,r)} \to [c]$ defined by

$$COL'(\sigma)(\tau) = COL(\sigma\tau).$$

Here $\sigma\tau$ is the vector in $[t]^N$ which is the concatenation of σ and τ.

We treat COL' as an X-coloring of $[t-1]^{HJ(t-1,X)}$. By definition of $HJ(t-1,X)$, we are guaranteed a monochromatic line, L, where $L \subseteq [t-1]^{HJ(t-1,X)} \subset [t]^{HJ(t-1,X)}$. Let $L = \{P_1, P_2, \ldots, P_{t-1}\}$. So we have

$$COL'(P_1) = COL'(P_2) = \cdots = COL'(P_{t-1}) = \chi$$

L is a line in $[t-1]^{U(t,c,r)}$, but it is only a line$^-$ in $[t]^{HJ(t-1,X)}$. Let P_t be its completion.

Of course, χ itself is a c-coloring of $[t]^{U(t,c,r)}$. By definition of $U(t,c,r)$, we get one of two things:

Case 1: If χ gives a monochromatic line $L' = \{Q_1, Q_2, \ldots, Q_t\}$, then our monochromatic line in $[t]^N$ is

$$\{P_1Q_1, P_1Q_2, \ldots, P_1Q_t\}$$

and we are done. (Note that $\{P_2Q_1, P_2Q_2, \ldots, P_2Q_t\}$ also would have worked, as would $\{P_3Q_1, P_3Q_2, \ldots, P_3Q_t\}$ etc.)

Case 2: We have $L_1^-, L_2^-, \ldots, L_r^-$, each a monochromatic line$^-$ in $[t]^{U(t,c,r)}$, and each with the same completion $Q \in [t]^{U(t,c,r)}$. Note that Q must have an $(r+1)^{st}$ color, or else we would be in case 1. Let Q_i^j denote the j^{th} point on L_i^-. We now have all the components needed to piece together $r+1$ monochromatic line$^-$ structures:

$$\{P_1Q_1^1, P_2Q_1^2, \ldots, P_{t-1}Q_1^{t-1}\}$$
$$\{P_1Q_2^1, P_2Q_2^2, \ldots, P_{t-1}Q_2^{t-1}\}$$
$$\vdots$$
$$\{P_1Q_r^1, P_2Q_r^2, \ldots, P_{t-1}Q_r^{t-1}\}$$
$$AND$$
$$\{P_1Q, P_2Q, \ldots, P_{t-1}Q\}$$

We already know the first r have different colors.

Case 2.1: The line$^-$ $\{P_1Q, P_2Q, \ldots, P_{t-1}Q\}$ is the same color as the sequence $\{P_1Q_i^1, P_2Q_i^2, \ldots, P_{t-1}Q_i^{t-1}\}$ for some i. Then the line given by

$$\{P_1Q_i^1, P_1Q_i^2, \ldots, P_1Q_i^{t-1}, P_1Q\}$$

is monochromatic, so we are done, satisfying Statement I.

Case 2.2: The line$^-$ structures listed are all monochromatic and different colors. Note that P_tQ is the completion for all of them, so Statement II is satisfyied. ∎

We now restate and prove the HJ:

Theorem 6.2.3 ***Hales-Jewett theorem*** *$\forall t, c, \exists N = HJ(t,c)$ such that, for all c-colorings $COL : [t]^N \rightarrow [c], \exists L \subseteq [t]^N, L$ a monochromatic line.*

Proof:

We prove this by induction on t. We show that

- $(\forall c)[HJ(1,c)$ exists$]$
- $(\forall c)[HJ(t-1,c)$ exists$] \implies (\forall c)[HJ(t,c)$ exists$]$

Base Case: $t = 1$ As noted in Fact 6.1.3 $HJ(1, c) = 1$ exists.
Induction Step: Assume $(\forall c)[HJ(t-1, c)$ exists $]$. Fix c. Consider Lemma 6.2.1 when $r = c$. In any c-coloring of $[t]^{U(t,c,c)}$, either there is a monochromatic line, or there are c monochromatic line$^-$ structures which are all colored differently, and share a completion Q colored differently. Since there are only c colors, this cannot happen, and we must have a monochromatic line. Hence $HJ(t, c) \leq U(t, c, c)$. ∎

6.3 Shelah's Proof of the Hales-Jewett Theorem

6.4 Bounds on the Hales-Jewett Numbers

6.4.1 Upper Bounds on the Hales-Jewett Numbers

6.4.2 Lower Bounds on the Hales-Jewett Numbers

Chapter 7

Applications of the Hales-Jewett Theorem

7.1 Positional Games

7.2 VDW and Variants

7.3 Comm. Comp. of XXX

7.4 The Square Theorem: Fourth Proof

Use HJ directly.

7.5 The Gallai-Witt Theorem (Multidim VDW)

Theorem 7.5.1 *Let $c, M \in \mathbb{N}$. Let $COL^* : \mathbb{N} \times \mathbb{N} \rightarrow [c]$. There exists a, d, D such that all of the following are the same color.*

$$\{(a + iD, d + jD) \mid -M \leq i, j \leq M\}.$$

7.6 The Canonical VDW

We first recall the following version of van der Waerden's theorem.

VDW For every $k \ge 1$ and $c \ge 1$ for every c-coloring $COL : [\mathbb{N}] \rightarrow [c]$ there exists a monochromatic k-AP. In other words there exists a, d such that

$$COL(a) = COL(a+d) = \cdots = COL(a+(k-1)d).$$

What if we use an infinite number of colors instead of a finite number of colors? Then the theorem is false as the coloring $COL(x) = x$ shows. However, in this case, we may get something else.

Def 7.6.1 Let $k \in \mathbb{N}$. Let COL be a coloring of $\mathbb{N}$ (which may use a finite or infinite number of colors). A *rainbox k-AP* is an arithmetic sequence $a, a+d, a+2d, \ldots, a+(k-1)d$ such that all of these are colored *differently.*

The following is the *Canonical van der Waerden's theorem.* Erdos and Graham [7] claim that it follows from Szemerëdi's theorem on density. Later Prömel and Rödl [21] obtained a proof that used the Gallai-Witt theorem.

Theorem 7.6.2 *Let $k \in \mathbb{N}$. Let $COL : \mathbb{N} \rightarrow \mathbb{N}$ be a coloring of the naturals. One of the following two must occur.*

- *There exists a monochromatic k-AP.*
- *There exists a rainbox k-AP.*

Proof:

Let COL^* be the following *finite* coloring of $\mathbb{N} \times \mathbb{N}$. Given (a, d) look at the following sequence

$$(COL(a), COL(a+d), COL(a+2d), \ldots, COL(a+(k-1)d)).$$

This coloring partitions the numbers $\{0, \ldots, k-1\}$ in terms of which coordinates are colored the same. For example, if $k = 4$ and the coloring was (R, B, R, G) then the partition is $\{\{0, 2\}, \{1\}, \{3\}\}$. We map (a, d) to the partition induced on $\{0, \ldots, k-1\}$ by the coloring. There are only a finite number of such partitions. (The Stirling numbers of the second kind are $S(k, L)$ are the number of ways to partition k numbers into L nonempty sets. The Bell numbers are $B_k = \sum_{L=1}^{k} S(k, L)$. The actual number is colors is B_k.)

Example 7.6.3

1. Let $k = 10$ and assume

   $$(COL(a), COL(a{+}d), \ldots, COL(a{+}(9d)) = (R, Y, B, I, V, Y, R, B, B, R).$$

   Then (a, d) maps to $\{\{0, 6, 9\}, \{1, 5\}, \{2, 7, 8\}, \{3\}, \{4\}, \}$.

2. Let $k = 6$ and assume

   $$(COL(a), COL(a + d) \ldots, COL(a + (5d)) = (R, Y, B, I, V, Y).$$

   Then (a, d) maps to $\{\{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}\}$.

Let M be a constant to be picked later. By Theorem 7.5.1 There exists a, d, D such that all of the following are the same COL^*

$$\{(a + iD, d + jD) \mid -M \le i, j \le M\}.$$

There are two cases.

Case 1: $COL^*(a, d)$ is the partition of every element into its own class. This means that there is a rainbow k-AP and we are done.

Case 2: There exists x, y such that $COL^*(a, d)$ is the partition that puts $a{+}xd$ and $a{+}yd$ in the same class. More simply, $COL(a{+}xd) = COL(a{+}yd)$. Since for all $-M \le i, j \le M$,

$$COL^*(a, d) = COL^*(a + iD, d + jD).$$

we have that, for all $-M \le i, j \le M$,

$$COL(a + iD + x(d + jD)) = COL(a + iD + y(d + jD)).$$

Assume that $COL(a + xd) = COL(a + yd) =$ RED. Note that we do not know $COL(a + iD + x(d + jD))$ or $COL(a + iD + y(d + jD))$, but we do know that they are the same.

We want to find the (i, j) with $-M \le i, j \le M$ such that
$COL^*(a + iD, d + jD)$ affects $COL(a + xd)$.

Note that
if

$$a + xd = a + iD + x(d + jD)$$

then

$$xd = iD + xd + xjD$$

$$0 = iD + xjD$$

$$0 = i + xj$$

$$i = -xj.$$

Hence we have that

$$a + xd = (a - xj)D + x(d + jD).$$

So what does this tell us? For all $-M \le i, j \le M$,

$$COL(a + iD + x(d + jD)) = COL(a + iD + y(d + jD)).$$

Let $i = -xj$ and you get

$$COL(a - xjD + x(d + jD)) = COL(a - xjD + y(d + jD)).$$

$$\text{RED} = COL(a + xd) = COL(a + yd + j(yD - xD)).$$

This holds for $-M \le j \le M$. Looking at $j = 0, 1, \ldots, k-1$, and letting $A = a + yd$ and $D' = yD - xD$, we get

$$COL(A) = COL(A{+}D') = COL(A{+}2D') = \cdots = COL(A{+}(k{-}1)D') = \text{RED}.$$

This yields an monochromatic k-AP.

What value do we need for M? We want $j = 0, 1, \ldots, k-1$. We want $i = -xj$. We know that $x \le k-1$. Hence it suffices to take $M = (k-1)^2$. ∎

Note 7.6.4 We used the two-dimensional VDW to prove the one-dimensional canonical VDW. For all d there is a d-dimensional canonical VDW, and it is proven using the $d+1$-dimensional VDW. The actual statement is somewhat complicated. The interested reader can see [21].

7.7 Comm. Comp. of $\sum_{i=1}^{k} x_i \equiv 0 \pmod{2^n - 1}$

Chapter 8

The Polynomial Hales-Jewett Theorem

8.1 Introduction

Much as VDW has a generalization to POLYVDW, so does HJ. To get there, we must first generalize a few definitions, and create some we had no need for in the original version.

Recall that, in HJ, we colored elements of $[t]^N$ and looked for monochromatic lines. Of course, we used the ground set $[t]$ only for convenience — we used none of the numerical properties. In that spirit, we may replace $[t]$ with any alphabet Σ of t letters.

Let $\Sigma = (\Sigma_d, \ldots, \Sigma_1)$ be a list of alphabets, and n a natural number. A *Hales-Jewett space* has the form

$$S_\Sigma(n) = \Sigma_d^{n^d} \times \Sigma_{d-1}^{n^{d-1}} \times \cdots \times \Sigma_1^n$$

We view an element $A \in S_\Sigma(n)$ as a collection of structures: a vector with coordinates from Σ_1, a square with coordinates from Σ_2, a cube with coordinates from Σ_3, and so on. In the ase of $d = 1$, and $\Sigma = [t]$, this is precisely the space colored in the ordinary HJ.

We define a set of formal polynomials over Σ by

$$\Sigma[\gamma] = \left\{ a_d\gamma^d + \cdots + a_1\gamma \mid a_i \in \Sigma_i \right\}$$

Note that every polynomial has exactly d terms — omitting a term is not permitted. This differs from POLYVDW where we allowed any polynomials.

For example, x^3 is a valid polynomial when dealing with POLYVDW. The closest we can come to this in POLYHJ is $1x^3 + 0x^2 + 0x$. Note that $1x^3 + 0x^2 + 0x$ is not equivalent to x^3. In fact, the term x^3 has no meaning since the coefficients come out of a finite alphabet. Note that although the coefficients may suggest meaning to the reader, they will have no numerical significance in the context of HJ .

Let $A \in S_\Sigma(n)$, $p \in \Sigma[\gamma]$ of the form $p(\gamma) = a_d\gamma^d + \cdots + a_1\gamma$, and $\lambda \subseteq [n]$. Then we define $A + p(\lambda) \in S_\Sigma(n)$ as follows. Take the line from A, and replace the coordinates in λ by a_1. Similarly, replace the coordinates from the square in $\lambda^2 = \lambda \times \lambda$ with a_2, and so on.

Example 8.1.1 Let $\Sigma_1 = \{a, b, c, d\}, \Sigma_2 = [9]$, and $\Sigma = (\Sigma_2, \Sigma_1)$. Let $A \in S_\Sigma(3)$ be

$$A = \begin{pmatrix} 3 & 1 & 2 \\ 8 & 8 & 9 \\ 4 & 5 & 3 \end{pmatrix} \quad (a \;\; d \;\; c)$$

Note that A consists of a 3×3 block and a 1×3 block together, but they have no mathematical significance as a matrix or a vector.

Now, let $p \in \Sigma[\gamma]$ be given by $p(\gamma) = 1\gamma^2 + b\gamma$, and let $\lambda = \{1, 2\}$. Then

$$A + p(\lambda) = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 9 \\ 4 & 5 & 3 \end{pmatrix} \quad (b \;\; b \;\; c)$$

Now, we can restate HJ in this language.

Theorem 8.1.2 ***Hales-Jewett theorem***
For every c, every finite alphabet Σ, there is some N such that, for any c-coloring $COL : S_\Sigma(N) \to [c]$, there is a point $A \in S_\Sigma(N), \lambda \subseteq [N]$, with $\lambda \neq \emptyset$ such that the set $\{A + \sigma\lambda \mid \sigma \in \Sigma\}$ is monochromatic.

From this terminology, we see a very natural generalization to a polynomial version of the theorem.

Theorem 8.1.3 ***Polynomial Hales-Jewett theorem***
For every c, every list of finite alphabets $\Sigma = (\Sigma_d, \ldots, \Sigma_1)$, and every collection $P \subseteq \Sigma[\gamma]$, there is a number $N = HJ(\Sigma, P, c)$ with the following property. For any c-coloring $COL : S_\Sigma(N) \to [c]$, there is a point $A \in S_\Sigma(N), \lambda \subseteq [N]$ with $\lambda \neq \emptyset$, such that the set $\{A + p(\lambda) \mid p \in P\}$ is monochromatic.

Example 8.1.4 Let $d = 2, \Sigma_2 = \{0, \ldots, 9\}, \Sigma_1 = \{a, \ldots, z\}$. Let

$$P = \{1\gamma^2 + a\gamma, 1\gamma^2 + b\gamma, 2\gamma^2 + c\gamma\}.$$

If $N = 3$ and $\lambda = \{2, 3\}$, then the following would be an appropriate monochromatic set:

$$\begin{pmatrix} 5 & 2 & 4 \\ 7 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \; (f \quad a \quad a)$$

$$\begin{pmatrix} 5 & 2 & 4 \\ 7 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \; (f \quad b \quad b)$$

$$\begin{pmatrix} 5 & 2 & 4 \\ 7 & 2 & 2 \\ 1 & 2 & 2 \end{pmatrix} \; (f \quad c \quad c)$$

POLYHJ was first proven by Bergelson and Leibman [2] using Ergodic methods. We present the proof by Walters [35] that uses purely combinatorial techniques.

8.2 Defining Types of Sets of Polynomials

Note that, in POLYVDW, we needed to assume that $p(0) = 0$ for every $p \in P$. We have no such statement here, because we have no notion of a constant term for a polynomial in $\Sigma[\gamma]$.

To prove this theorem, we will do induction on the "type" of the set of polynomials P, as in the POLYVDW. However, each polynomial necessarily has degree d, which makes the notion of type used previously rather unhelpful. In order to get the induction to work, we need to introduce a *relative* notion of degree, and tweak the definition of type.

Def 8.2.1 Let Σ be a list of finite languages, and $p, q \in \Sigma[\gamma]$. Then we say the degree of p relative to q is the degree of the highest term on which they differ. Formally, let $p(\gamma) = a_d\gamma^d + \cdots + a_1\gamma^1$, and $q(\gamma) = b_d\gamma^d + \cdots + b_1\gamma^1$. Let k be the largest index such that $a_k \neq b_k$ (or 0 if $p = q$). Then p has degree k with respect to q. We also say that p has leading coefficient a_k with respect to q

Note that the definition is symmetric: the degree of p relative to q is the same as the degree of q relative to p.

Example 8.2.2 Define

$$f(\gamma) = a\gamma^3 + 3\gamma^2 + \heartsuit\gamma$$
$$g(\gamma) = a\gamma^3 + 3\gamma^2 + \diamondsuit\gamma$$
$$h(\gamma) = b\gamma^3 + 3\gamma^2 + \heartsuit\gamma$$

The we see the following:

- f has degree 1 relative to g.
- f has leading coefficient $\heartsuit$ relative to g, and g has leading coefficient $\diamondsuit$ relative to f.
- h has degree 3 relative to both f and g.
- h has leading coefficient b relative to f and g, which each have leading coefficient a relative to h.

With this definition, we can now define the type of a set of polynomials relative to q virtually the same as we did for POLYVDW.

Def 8.2.3 Let Σ be a list of d finite alphabets, and $P \subseteq \Sigma[\gamma]$, $q \in \Sigma[\gamma]$. For each index k, let $P_k \subseteq P$ be the subset of polynomials with degree k relative to q. Let n_k be the number of leading coefficients relative to q of polynomials in P_k. Then the type of P relative to q is vector $(n_d, \ldots, n_1)$. We give these type vectors the same ordering as before, as seen in Definition 4.1.9.

For each $p_i \in P$, let t_i be the type of P relative to p_i. Then we say P has [absolute] type $t = \min t_i$.

Example 8.2.4 Let $P = \{p_1, p_2, p_3, p_4, p_5\}$, where

$$p_1 = a\gamma^3 + 6\gamma^2 + \diamondsuit\gamma$$
$$p_2 = a\gamma^3 + 6\gamma^2 + \heartsuit\gamma$$
$$p_3 = a\gamma^3 + 7\gamma^2 + \heartsuit\gamma$$
$$p_4 = b\gamma^3 + 6\gamma^2 + \diamondsuit\gamma$$
$$p_5 = b\gamma^3 + 6\gamma^2 + \heartsuit\gamma$$

Let $Q = Q - \{p_5\}$. We see that:

- The type of P relative to p_1 and p_2 is $(1,1,1)$.
- The type of P relative to p_3 is $(1,1,0)$.
- The type of P relative to p_4 and p_5 is $(1,0,1)$.
- The [absolute] type of P is $(1,0,1)$, minimized by p_4 and p_5.
- P and $P-\{p_5\}$ have the same type relative to p_1, p_2, and p_3, the type remains unchanged.
- The type of $P-\{p_5\}$ relative to p_4 is $(1,0,0)$ — lower than the type of P.

The next proposition states that this last point always happens — the type of a set always decreases when you remove the polynomial which minimizes it.

Proposition 8.2.5 *Let P be a set of polynomials, such that $p \in P$ minimizes its type. Then $P-\{p\}$ has lower type.*

Proof: Let P have type $(n_d, \ldots, n_1)$, and let p minimize the type of P. Choose $q \in P$ to have minimal degree with respect to p, and call that degree k. Define $Q = P - \{p\}$. For polynomials in Q of degree greater than k relative to p, the degree is unchanged relative to q. Since the leading coefficients are also unchanged, the first $d-k$ coefficients of the type vector are identical for P and Q.

Now, let $Q_k \subseteq Q$ be the set of polynomials with degree $\leq k$ relative to p. By definition of the type vector, there are [exactly] n_k different leading coefficients of degree k polynomials in this set. Moreover, there are no polynomials of lower degree relative to p, since q was chosen to minimize k. Now, q has one of the n_k leading coefficients relative to p. Thus, relative to q, Q_k has $n_k - 1$ leading coefficients of degree k, with the remaining polynomials reducing in degree, because they agree with q on that coefficient. Thus, the type of Q relative to q is $(n_d, \ldots, n_{k+1}, n_k - 1, n'_{k-1}, \ldots, n'_1)$, for some values of $n'_{k-1}, \ldots, n'_1$. This type is lower than that of P, so the minimum type of Q is lower as well. ∎

Remark: We picked k to be the minimal degree of a polynomial relative to p. This means that the type of P is $(n_d, \ldots, n_k, 0, \ldots, 0)$. If there were any polynomials of degree $< k$, we would have picked one of those rather than q.

8.3 How to View and Alphabet

Now, in proving the HJ, it was important to view Σ^{n+m} as $\Sigma^n \times \Sigma^m$. We will need something similar for the polynomial version.

Proposition 8.3.1 *Let $n, m \in \mathbb{N}$, and Σ be a list of finite alphabets. Then there is a finite list of alphabets Σ' so that $S_\Sigma(n+m) \cong S_\Sigma(n) \times S_{\Sigma'}(m)$, where Σ' is independent of m.*

The proof of this is rather messy, but is done by manipulating the definition of $S_\Sigma(n+m)$. Rather than prove it in general here, we show the case when $\Sigma = (\Sigma_2, \Sigma_1)$.

$$\begin{aligned} S_\Sigma(n+m) &= \Sigma_2^{(n+m)^2} \times \Sigma_1^{n+m} \cong \Sigma_2^{n^2} \times \Sigma_2^{2nm} \times \Sigma_2^{m^2} \times \Sigma_1^n \times \Sigma_1^m \\ &\cong \left(\Sigma_2^{n^2} \times \Sigma_1^n\right) \times \left(\Sigma_2^{m^2} \times [\Sigma_2^{2n} \times \Sigma_1]^m\right) \end{aligned}$$

By setting $\Sigma' = (\Sigma_2, \Sigma_2^{2n} \times \Sigma_1)$, this comes out to be $S_\Sigma(n) \times S_{\Sigma'}(m)$, as desired. We view the transformation from $S_\Sigma(n+m)$ to $S_\Sigma(n) \times S_{\Sigma'}(m)$ as follows:

- Cut the line of length $n+m$ into two lines — one of length n, and one of length m. The former belongs to $S_\Sigma(n)$, and the latter to $S_{\Sigma'}(m)$.
- Cut the $(n+m) \times (n+m)$ block into four blocks. One is an $n \times n$ square, which belongs to $S_\Sigma(n)$. Another is an $m \times m$ square, which lives in the 2-dimensional portion of $S_{\Sigma'}(m)$. Leftover are blocks of size $n \times m$ and $m \times n$. We view these as "thick" lines of length m, with each entry representing n entries of the original space. In this way, we attach these pieces of the square to the line in $S_{\Sigma'}(m)$.
- Similarly, the k-dimensional block of $S_\Sigma(n+m)$ will be cut into 2^k pieces. One goes to the k-dimensional portion of $S_\Sigma(n)$ and another to the k-dimensional portion of $S_{\Sigma'}(m)$. The rest go to lower-dimensional portions of $S_{\Sigma'}(m)$.

Looking the other direction, let A' be a point in $S_{\Sigma'}(m)$.

- The d-dimensional part of A' comes from the d-dimensional portion of the point in the original space ($S_\Sigma(n+m)$).

- The $(d-1)$-dimensional part has one piece which is "truly" $(d-1)$-dimensional, and the rest of the pieces originally lived in d dimensions.

- The k-dimensional part of A' has one piece which is "truly" k-dimensional, and the other pieces are from higher dimensions.

How does viewing $S_\Sigma(n+m)$ like this affect polynomials? Let $\lambda \subseteq \{1, \ldots, n\}$, and $\kappa \subseteq \{n+1, \ldots, n+m\}$. Consider a polynomial $p(\gamma) = 1\gamma^2 + 2\gamma$. Then, given a point $A \in S_\Sigma(n+m)$, $A + p(\lambda \cup \kappa)$ involves putting a 1 at every point in $(\lambda \cup \kappa)^2$, and a 2 in $\lambda \cup \kappa$. That is, we must put a 1 everywhere in $\lambda \times \lambda, \lambda \times \kappa, \kappa \times \lambda$, and $\kappa \times \kappa$, and a 2 in λ and κ. We may now [nearly] view p as two polynomials: one in $\Sigma[\gamma]$, and the other in $\Sigma'[\gamma]$. The first is just p, since the alphabet has not changed. For the other, we need to know ahead of time what λ will be, to correctly place the 1's in $\lambda \times \kappa$ and $\kappa \times \lambda$, since we have control over all entries in $[n] \times \kappa$ and $\kappa \times [n]$. For this, we define $p|_\lambda$, the restriction of p to the entries of λ, by

$$p|_\lambda(\gamma) = 1\gamma^2 + (2, a_1, \ldots, a_{2n})\gamma$$

Here $a_i = 1$ if $i \in \lambda$ or $i + n \in \lambda$. For all other a_i, we have the freedom to prescribe any entries from Σ_2. For now we will use x as an unspecified symbol from Σ_2 to highlight where the choice lies.

So how do these polynomials work together? Let $A \in S_\Sigma(2+3)$ be all 0's:

$$A = \left(\begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad (0 \;\; 0 \mid 0 \;\; 0 \;\; 0)$$

Now, let $\lambda = \{1\}$, and $\kappa = \{3, 4\}$, and let (B, C) be the decomposition of A as an element of $S_\Sigma(2) \times S_{\Sigma'}(3)$. Then we get

$$A + p(\lambda \cup \kappa) = \left(\begin{array}{cc|ccc} 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad (2 \;\; 0 \mid 2 \;\; 2 \;\; 0)$$

$$A' = (B + p(\lambda), C + p|_\lambda(\kappa)) = \left(\begin{array}{cc|ccc} 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & x & x & 0 \\ \hline 1 & x & 1 & 1 & 0 \\ 1 & x & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \quad (2 \;\; 0 \mid 2 \;\; 2 \;\; 0)$$

Note 8.3.2 A' is a close approximation of $A+p(\lambda\cup\kappa)$ — it agrees on $(\lambda\cup\kappa)^2$ and on $\lambda \cup \kappa$ as required by p. It only differs where x appears, because we could not predict what entries A would have there. Fortunately, in proving the theorem, we will only be interested in controlling the entries of $(\lambda \cup \kappa)^2$ and $(\lambda\cup\kappa)$ and ensuring the rest does not change. Therefore, if we are given a set of polynomials $P \subseteq \Sigma[\gamma]$, we may decompose each $p \in P$ as $(p, p|_\lambda)$ as above, and prescribe constant values for the x's. In proving POLYHJ , if we have a sequence

$$\{(B + p(\lambda), C + p|_\lambda(\kappa)) \mid p \in P\}$$

we will fix the x's so that it is equal to

$$\{(B, C) + p(\lambda \cup \kappa) \mid p \in P\}$$

To do this, we will have a fixed polynomial p_0 over Σ which will dictate all these choices. Formally, let $p, p_0 \in \Sigma[\gamma]$ be polynomials, with $p(\gamma) = a_d\gamma^d + \cdots + a_1\gamma$, and $p_0(\gamma) = b_d\gamma^d + \cdots + b_1\gamma$. Then $p|_\lambda(\gamma) = c_d\gamma^d + \cdots + c_1\gamma$ has the following structure:

- $c_d = a_d$
- c_{d-1} is a list of symbols. One of these is a_{d-1}. The rest come from a_d and b_d, but which goes where depends on λ. These coefficients are for the d-dimensional piece of the polynomial. We can therefore define c_{d-1} as $(a_{d-1}, f(a_d, b_d, \lambda))$.

- c_k is a list of symbols. One of these is a_k. The rest are divided up based on which dimension they represent. The coefficients representing dimension j come from a_j or b_j, depending on λ. Thus, we can write

$$c_k = (a_k, f(a_d, \ldots, a_{k+1}, b_d, \ldots, b_{k+1}, \lambda))$$

- If $a_d = b_d, \ldots, a_{k+1} = b_{k+1}$, then λ has no on the k^{th} coefficient, so we can write it as

$$c_k = (a_k, g(a_d, \ldots, a_{k+1}))$$

Def 8.3.3 Just as in the proof of the POLYVDW, we define $POLYHJ(n_d, \ldots, n_1)$ to be the statement that the POLYHJ holds for all sets of polynomials of type $(n_d, \ldots, n_1)$. As in Definition 4.1.7, we also define $POLYHJ(n_d, \ldots, n_k, \omega, \ldots, \omega)$ to be the analogous statement.

8.4 The Proof

We are now ready to prove a lemma from which the theorem will become trivial.

Lemma 8.4.1 *Assume $POLYHJ(n_d, \ldots, n_k, \omega, \ldots, \omega)$ holds. Fix a finite list of alphabets Σ and let $P \subseteq \Sigma[\gamma]$ have type $(n_d, \ldots, n_k + 1, 0, \ldots, 0)$, minimized by $p_0 \in P$. Then, for all numbers $c, r > 0$, there is a number $U = U(\Sigma, P, c, r)$ with the following property. For all c-colorings $COL : S_\Sigma(U) \to [c]$, one of the following Statements holds:*

Statement I: *There is a point $A \in S_\Sigma(U)$, $\lambda \subseteq [U], \lambda \neq \emptyset$, where $\{A + p(\lambda) \mid p \in P\}$ is monochromatic, or*

Statement II: *There are points $A_1, \ldots, A_r, A' \in S_\Sigma(U)$, and $\lambda_1, \ldots, \lambda_r \subseteq [U]$ with each $\lambda_i \neq \emptyset$ so that each set $\{A_i + p(\lambda_i) \mid p \in P, p \neq p_0\}$ is monochromatic, each with its own color, and each different from A'. Additionally, $A' = A_i + p_0(\lambda_i)$ as points for each $i \leq r$. We call A' the completion point of the sequences.*

Proof: By induction on r:

Base case $(r = 1)$ — Recall that $P - \{p_0\}$ has lower type than P. Thus, Poly HJ holds for $P - \{p_0\}$. Let $U = HJ(\Sigma, P - \{p_0\}, c)$. Take any c-coloring of $S_\Sigma(U)$. By the definition of this number, there is some $A_1 = A \in S_\Sigma(U)$, and $\lambda_1 = \lambda \subseteq [U]$ with $\lambda \neq \emptyset$ so that $\{A_1 + p(\lambda_1) \mid p \in P - \{p_0\}\}$ is monochromatic. If the completion point is the same color, then Statement I holds. If not, Statement II holds.

Inductive case — Assume the lemma holds for r. We show that $U(\Sigma, P, c, r+1)$ exists by giving an upper bound. In particular,

$$U(\Sigma, P, c, r+1) \leq U + HJ = U(\Sigma, P, c, r) + HJ(\Sigma', Q, X)$$

where Q will be given shortly, and $X = c^{|S_\Sigma(U)|}$ is the number of c-colorings of $S_\Sigma(U)$. How convenient. By Proposition 8.3.1, $S_\Sigma(U + HJ) \cong S_\Sigma(U) \times S_{\Sigma'}(HJ)$, for some list of finite alphabets Σ' independent of the value of HJ. Then let

$$Q = \{p|_\lambda \in \Sigma'[\gamma] \,:\, p \in P - \{p_0\}, \lambda \subseteq [U]\}$$

where the free choice of entries is prescribed by p_0. This will ensure that

$$(A + p(\lambda), B + p|_\lambda(\kappa)) = (A, B) + p(\lambda \cup \kappa)$$

for any choice of p, λ, κ.

Claim: Q has type $(n_d, \ldots, n_k, n'_{k-1}, \ldots, n'_1)$ for some choice of $n'_{k-1}, \ldots, n'_1$.

Proof: By Proposition 8.2.5, $P - \{p_0\}$ has lower type than P, attained by some p_1 of degree k relative to p_0. Thus $P - \{p_0\}$ has type $(n_d, \ldots, n_k, m'_{k-1}, \ldots, m'_1)$ for some choice of $m'_{k-1}, \ldots, m'_1$. We will use this to show that the type of Q relative to $p_1|_\emptyset$ is low enough. In particular, we will show two things:

1. If p has degree $\ell \geq k$ relative to p_1, then $p|_\lambda$ has degree ℓ relative to $p_1|_\emptyset$ for every λ.

2. If p and q both have degree $\ell \geq k$ relative to p_1, and they have the same leading coefficient, then $p|_\lambda$ and $q|_\kappa$ have the same leading coefficient, for all λ and κ.

The two things together will guarantee that the number of distinct leading coefficients in Q will agree with the number in $P - \{p_0\}$, for all degrees $\geq k$, which is exactly what we want.

To see (1), write $p(\gamma) = a_d\gamma^d + \cdots + a_1\gamma$, $p_1(\gamma) = b_d\gamma^d + \cdots + b_1\gamma$, and $p_0(\gamma) = c_d\gamma^d + \cdots + c_1\gamma$. Since p_1 has degree k relative to p_0, the two polynomials agree on $b_d, \ldots, b_{k+1}$. Similarly, since p has degree $\ell \geq k$ relative to p_1, we get that $a_d = b_d = c_d, \ldots, a_{\ell+1} = b_{\ell+1} = c_{\ell+1}$. By Note 8.3.2, we see that the j^{th} coefficient of $p|_\lambda$ and $p_1|_\emptyset$ are given by $(a_j, f(a_d, \ldots, a_{j+1}))$ and $(b_j, f(b_d, \ldots, b_{j+1}))$ when $j \geq \ell$. Since these are identical when $j > \ell$, and different when $j = \ell$, we see that $p|_\lambda$ has degree ℓ relative to $p_1|_\emptyset$.

To see (2), let p and q have the same degree $\ell \geq k$ relative to p_1 (and thus relative to p_0), and also the same leading coefficient. Let $p(\gamma) = a_d\gamma^d + \cdots + a_1\gamma$, $q(\gamma) = b_d\gamma^d + \cdots + b_1\gamma$, and $p_0(\gamma) = c_d\gamma^d + \cdots + c_1\gamma$. As before, we see that $a_d = b_d = c_d, \ldots, a_{\ell+1} = b_{\ell+1} = c_{\ell+1}$. Fixing $\lambda, \kappa \subseteq [U]$, consider $p|_\lambda$ and $q|_\kappa$. By Note 8.3.2, for $j \geq \ell$, the j^{th} coefficient of these are given by $(a_\ell, f(a_d, \ldots, a_{\ell+1}))$ and $(b_\ell, f(b_d, \ldots, b_{\ell+1}))$ respectively. Since $a_j = b_j$ for $j \geq \ell$, these coeffcents are identical. Thus, the two polynomials share a common leading coefficient relative to $p_1|_\emptyset$. ∎

This gives Q a lower type than P, which will allow us to use $POLYHJ$ as assumed.

Now, Let $COL : S_\Sigma(U + HJ) \to [c]$ be a c-coloring. Then we view COL as a c-coloring of $S_\Sigma(U) \times S_{\Sigma'}(HJ)$. As such, for each $\sigma \in S_{\Sigma'}(HJ)$, define $COL^*(\sigma) : S_{\Sigma'} \to [c]$ as the coloring of $S_\Sigma(U)$ induced by COL — for $\tau \in S_\Sigma(U)$, the map is defined so that $COL^*(\sigma)(\tau) = COL(\tau, \sigma)$. This makes COL^* a map from $S_{\Sigma'}$ to the c-colorings of $S_\Sigma(U)$.

The crucial observation here is there are X possible c-colorings of $S_\Sigma(U)$, so COL^* serves as an X-coloring of $S_{\Sigma'}(HJ)$. Thus, by choice of HJ, there is some point $B \in S_{\Sigma'}(HJ)$, and $\Lambda \subseteq [HJ]$ with $\Lambda \neq \emptyset$ so that

$$\{B + q(\Lambda) \mid q \in Q\} = \{B + p|_\lambda(\Lambda) \mid p \in P - \{p_0\}, \lambda \subseteq [U]\}$$

is monochromatic. This means that each point induces the same coloring χ on $S_\Sigma(U)$.

Now χ is a c-coloring of $S_\Sigma(U)$, so the choice of U allows us to use the inductive hypothesis on χ. Thus, either Statement I or II hold.

Case 1: There is a point $A \in S_\Sigma(U), \lambda \subseteq [U], \lambda \neq \emptyset$, so that $\{A + p(\lambda) \mid p \in P\}$ is monochromatic under χ. Then fix any $q \in Q$. Define $C = B + q(\Lambda)$. Since C induces the coloring χ on $S_\Sigma(U)$, we see that $\{(A+p(\lambda), C) \mid p \in P\}$ is monochromatix under COL. Moreover, viewing λ as a subset of $[U+HJ]$, these points are actually $(A + C) + p(\lambda)$, so we satisfy Statement I.

Case 2: There is are points $A_1, \ldots, A_r, A' \in S_\Sigma(U), \lambda_1, \ldots, \lambda_r \subseteq [U]$ with each $\lambda_i \neq \emptyset$ with the following properties:

$$\{A_1 + p(\lambda_1) \mid p \in P - \{p_0\}\} \text{ is monochromatic under } \chi$$

$$\vdots$$

$$\{A_r + p(\lambda_r) \mid p \in P - \{p_0\}\} \text{ is monochromatic under } \chi$$

and each of these sets has a different color, all different from $\chi(A')$. We also have $A' = A_i + p(\lambda_i)$ for all $i \leq r$

Since each $B+q(\Lambda)$ induces χ on $S_\Sigma(U)$, this gives us very many monochromatic points. For each i, this set is monochromatic under COL:

$$\{(A_i + p(\lambda_i), B + q) \mid p \in P - \{p_0\}, q \in Q\}$$

In particular, the following r sets of points are monochromatic, so that each set has its own color:

$$\{(A_1+p(\lambda_1), B+p|_{\lambda_1}(\Lambda)) \mid p \in P-\{p_0\}\} = \{(A_1, B)+p(\lambda_1 \cup \Lambda) \mid p \in P-\{p_0\}\}$$

$$\vdots$$

$$\{(A_r+p(\lambda_r), B+p|_{\lambda_r}(\Lambda)) \mid p \in P-\{p_0\}\} = \{(A_r, B)+p(\lambda_r \cup \Lambda) \mid p \in P-\{p_0\}\}$$

Let $B' = B + p_0|_\emptyset$. Then we see that the final point of each of these sequences is given by

$$(A_i + p_0(\lambda_i), B + p_0|_{\lambda_i}(\Lambda)) = (A', B') + p_0(\Lambda)$$

This realization gives us the following choice for the $(r+1)^{\text{st}}$ sequence:

$$\{(A', B' + p|_\emptyset(\Lambda)) \mid p \in P - \{p_0\}\} = \{(A', B') + p(\Lambda) \mid p \in P - \{p_0\}\}$$

Since each $B' + p|_\emptyset(\Lambda)$ induces χ on $S_\Sigma(U)$, each of these has the color $\chi(A')$, so this set is monochromatic. It is also immediate that its completion point is the same as the other r: $(A', B') + p_0(\Lambda)$.

If the completion point has the same color as the i^{th} sequence, then that sequence with its completion satisfies Statement I. If not, then Statement II holds. Either way, we have the goal. ∎

Theorem 8.4.2 ***Polynomial Hales-Jewett theorem***
For every c*, every lits of finite alphabets* $\Sigma = (\Sigma_d, \ldots, \Sigma_1)$*, and every collection* $P \subseteq \Sigma[\gamma]$*, there is a number* $N = HJ(\Sigma, P, c)$ *with the following property. For any* c*-coloring* $COL : S_\Sigma(N) \rightarrow [c]$*, there is a point* $A \in S_\Sigma(N), \lambda \subseteq [N]$ *with* $\lambda \neq \emptyset$*, such that the set* $\{A + p(\lambda) \mid p \in P\}$ *is monochromatic.*

Proof: By induction on the type of P. Note that, as in the proof of the POLYVDW, types are well-ordered, so induction is a correct approach.

Base case: Let P have type $(0, \ldots, 0)$, so that $P = \{p\}$ is a single polynomial (p has degree 0 relative to itself). Set $N = 1$, and let $COL : S_\Sigma(1) \rightarrow [c]$ be any c-coloring. Then, for any $A \in S_\Sigma(1)$, we have $\{A + p(\{1\})\}$ monochromatic, since it is just one point.

Inductive case: Suppose we know $POLYHJ(n_d, \ldots, n_k, \omega, \ldots, \omega)$. Let P have type $(n_d, \ldots, n_k + 1, 0, \ldots, 0)$. Let $N = U(\Sigma, P, c, c)$ as guaranteed by the lemma above. Let $COL : S_\Sigma(U) \rightarrow [c]$ be a c-coloring. Statement II cannot hold, since it requires $c+1$ different colors. Thus, Statement I holds, which was the goal. ∎

8.5 Bounds on the Polynomial Hales-Jewett Numbers

8.5.1 Upper Bounds

8.5.2 Lower Bounds

Chapter 9

Applications of Polynomial Hales-Jewett Theorem

9.1 The Polynomial Van Der Waerden Theorem

9.2 The Poly Van Der Waerden Theorem Over a Communative Ring

9.3 The Multidim Poly Van Der Waerden Theorem

Chapter 10

Coloring and Equations: Rado's Theorem

10.1 Introduction

VDW theorem with $k = 4$ is can be rewritten as follows:

For all c, for all c-colorings $\chi : N \rightarrow [c]$, there exists a, d such that

$$\chi(a) = \chi(a+d) = \chi(a+2d) = \chi(a+3d),$$

We rewrite this in terms of equations.

For all c, for all c-colorings $\chi : N \rightarrow [c]$, there exists e_1, e_2, e_3, e_4 such that

$$\chi(e_1) = \chi(e_2) = \chi(e_3) = \chi(e_4)$$

and

$$\begin{aligned} e_2 - e_1 = &\ e_3 - e_2 \\ e_2 - e_1 = &\ e_4 - e_3 \end{aligned}$$

We rewrite these equations:

$$\begin{aligned} 0e_4 - e_3 + 2e_2 - e_1 = &\ 0 \\ -e_4 + e_3 + e_2 - e_1 = &\ 0 \end{aligned}$$

Let A be the matrix:

$$\begin{pmatrix} 0 & -1 & 2 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix}$$

VDW for $k = 4$ can be rewritten as

For all c, for all c-colorings $\chi : \mathbb{N} \rightarrow [c]$ there exists $\vec{e} = e_1, \ldots, e_n$ such that

$$\chi(e_1) = \cdots = \chi(e_n),$$

$$A\vec{e} = \vec{0}.$$

What other matrices have this property?

Def 10.1.1

1. $(b_1, \ldots, b_n) \in \mathbb{Z}^n$ is *regular* if the following holds: *For all c, there exists $R = R(b_1, \ldots, b_n; c)$ such that for all c-colorings $\chi : [R] \rightarrow [c]$ there exists $e_1, \ldots, e_n \in [R]$ such that*

$$\chi(e_1) = \cdots = \chi(e_n),$$

$$\sum_{i=1}^{n} b_i e_i = 0.$$

2. A matrix A of integers is *regular* if if the following holds: *For all c, for all c-colorings $\chi : \mathbb{N} \rightarrow [c]$ there exists $\vec{e} = e_1, \ldots, e_n$ such that*

$$\chi(e_1) = \cdots = \chi(e_n),$$

$$A\vec{e} = \vec{0}.$$

(Note that the definition of a regular matrix subsumes that of a regular vector.)

We will prove the *Abridged Rado Theorem* which gives an exact condition for single equations to be regular, and then the *Full Rado Theorem* which gives an exact condition for matrices to be regular.

10.2 The Abridged Rado's Theorem

We will prove two theorems, Theorems 10.2.5 and Theorem 10.2.6 that when combined yield *The Abridged Rado's Theorem.*

Theorem 10.2.1 *$(b_1, \ldots, b_n)$ is regular iff there exists some nonempty subset of $\{b_1, \ldots, b_n\}$ that sums to 0.*

In the proof of Theorem 10.2.8 we will be given $(b_1, \ldots, b_n)$ such that some subset sums to 0, a c-coloring of $\mathbb{N}$, and we will find $(e_1, \ldots, e_n)$ such that they are all colored the same and $\sum_{i=1}^{n} b_i e_i = 0$. However, many of the e_i's are the same. What if we want all of the e_i's to be different?

Def 10.2.2 A vector $(b_1, \ldots, b_n) \in \mathbb{Z}^n$ is *distinct-regular* if the following holds: *For all c, for all c-colorings $\chi : \mathbb{N} \to [c]$ there exists $e_1, \ldots, e_n$, all distinct, such that*

$$\chi(e_1) = \cdots = \chi(e_n),$$

$$\sum_{i=1}^{n} b_i e_i = 0.$$

Is it possible that all regular $(b_1, \ldots, b_n)$ are also distinct regular? NO, consider $(1, -1)$ or any $(b, -b)$. These are clearly regular but not distinct-regular. We will see that these are the only exceptions.

We will prove the following

Theorem 10.2.3 *If $(b_1, b_2, \ldots, b_n)$ is regular and there exists $(\lambda_1, \ldots, \lambda_n) \neq \vec{0}$ such that $\sum_{i=1}^{n} \lambda_i b_i = 0$ then $(b_1, \ldots, b_n)$ is distinct-regular.*

10.2.1 If ... then $(b_1, \ldots, b_n)$ is not Regular

We show that $(2, 5, -1)$ is not regular. We find a 17-coloring (actually 16-coloring) that demonstrates this. Our first attempt at finding a 17-coloring will not quite work, but our second one will.

First Attempt

$$\chi(n) \text{ is the number between 0 and 16 that is } \equiv n \pmod{17}.$$

Assume $\chi(e_1) = \chi(e_2) = \chi(e_3)$. We will try to show that

$$2e_1 + 5e_2 - e_3 \neq 0.$$

Assume, by way of contradiction, that

$$2e_1 + 5e_2 - e_3 = 0.$$

Let e be such that $e_1 \equiv e_2 \equiv e_3 \equiv e \pmod{17}$ and $0 \le e \le 16$. Then

$$0 = 2e_1 + 5e_2 - e_3 \equiv 2e + 5e - e \equiv 6e \pmod{17}.$$

Hence $6e \equiv 0 \pmod{17}$. Since 6 has an inverse mod 17, we obtain $e \equiv 0 \pmod{17}$.

We have *not* arrived at a contradiction. We have just established that if

$$\chi(e_1) = \chi(e_2) = \chi(e_3)$$

and

$$2e_1 + 5e_2 - e_3 = 0.$$

Then $\chi(e_1) = \chi(e_2) = \chi(e_3) = 0$.

Hence we will do a similar coloring but do something else when $n \equiv 0 \pmod{17}$.

Second Attempt

Given n let i, n' be such that 17^i divides n, 17^{i+1} does not divide n, and $n = 17^i n'$.

We define the coloring as follows:

$$\chi(n) \text{ is the number between 1 and 16 that is } \equiv n' \pmod{17}.$$

NOTE- $\chi(n)$ will never be 0. Hence this is really a 16-coloring.

Assume

$$\chi(e_1) = \chi(e_2) = \chi(e_3).$$

We show that

$$2e_1 + 5e_2 - e_3 \neq 0.$$

Let $i, j, k, e_1', e_2', e_3', e$ be such that

1. 17^i divides e_1, 17^{i+1} does not divide e_1, $e_1 = 17^i e_1'$.
2. 17^j divides e_2, 17^{j+1} does not divide e_2, $e_2 = 17^j e_2'$.
3. 17^k divides e_3, 17^{k+1} does not divide e_3, $e_3 = 17^j e_3'$.
4. $e_1' \equiv e_2' \equiv e_3' \equiv e \pmod{17}$

If

$$2e_1 + 5e_2 - e_3 = 0$$

then

$$2 \times 17^i e_1' + 5 \times 17^j e_2' - 17^k e_3' = 0.$$

Every mathematical bone in my body wants to cancel some of the 17's. There are cases. All $\equiv$ are mod 17.

1. $i < j \le k$ or $i < k \le j$.

   $$2 \times 17^i e_1' + 5 \times 17^j e_2' - 17^k e_3' = 0.$$

   Divide by 17^i.

   $$2 \times e_1' + 5 \times 17^{j-i} e_2' - 17^{k-i} e_3' = 0.$$

   We take this equation mod 17.

   $$2e_1' \equiv 2e \equiv 0.$$

   Since 2 has an inverse mod 17 we have $e = 0$. This contradicts that $e \neq 0$.

2. $i = j < k$.

   $$2 \times 17^i e_1' + 5 \times 17^i e_2' - 17^k e_3' = 0.$$

   Divide by 17^i.

   $$2 \times e_1' + 5 \times 17^{j-i} e_2' - 17^{k-i} e_3' = 0.$$

We take this equation mod 17.

$$2e_1' + 5e_2' \equiv 7e \equiv 0.$$

Since 7 has an inverse mod 17 we have $e = 0$. This contradicts that $e \neq 0$.

3. Rather than go through all of the cases in detail, we say what results in all caes, including those above.

   (a) $i < j \leq k$ or $i < k \leq j$: $2e \equiv 0$.

   (b) $i = j < k$: $2e + 5e \equiv 0$.

   (c) $i = k < j$: $2e - e \equiv 0$.

   (d) $i = j = k$: $2e + 5e - 3e \equiv 0$.

   (e) $j < i \leq k$ or $j < k \leq i$: $5e \equiv 0$.

   (f) $j = k < i$: $2e - e \equiv 0$.

   (g) $k < i = j$: $-e \equiv 0$.

There were 7 cases. Each corresponded to a combination of the coefficients. The key is that every combination was relatively prime to 17. The reader should be able to prove the following two theorems.

Theorem 10.2.4 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n$. If there exists c that is relatively prime to every nonempty subsum of $\{b_1, \ldots, b_n\}$ then there is a $c-1$-coloring of $\mathbb{N}$ that shows $(b_1, \ldots, b_n)$ is not regular.*

Theorem 10.2.5 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n$. If all subsets of $\{b_1, \ldots, b_n\}$ (except the empty set) have a non-zero sum then $(b_1, \ldots, b_n)$ is not regular.*

10.2.2 If ... then $(b_1, \ldots, b_n)$ is Regular

Motivation

So when is $b_1, \ldots, b_n$ regular? If $(b_1, \ldots, b_n)$ does not satisfy the premise of Theorem 10.2.5 then some nontempty subset of $\{b_1, \ldots, b_n\}$ sums to 0.

Theorem 10.2.6 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n$. Assume there exists a nonempty subset of $\{b_1, \ldots, b_n\}$ that sums to 0. Then $(b_1, \ldots, b_n)$ is regular.*

Before proving this theorem we talk about how to go about it. Lets use

$$5e_1 + 6e_2 - 11e_3 + 7e_4 - 2e_5 = 0$$

as an example. Note that the first three coefficients add to 0: $5+6-11=0$. We are thinking about colorings. OH, we can use the following version of van der Waerden's theorem!

Van der Waerden's Theorem: For all $x_1, \ldots, x_k \in \mathbb{Z}$, for all c, for all c-colorings $\chi : \mathbb{N} \to [c]$ there exists a, d such that

$$\chi(a) = \chi(a + x_1 d) = \chi(a + x_2 d) = \cdots = \chi(a + x_k d).$$

We use the $k = 5$ case. Is there a choice of x_1, x_2, x_3, x_4, x_5 that will give us our theorem?

Say that $e_i = a + x_i d$. Then

$$\begin{aligned} 5e_1 + 6e_2 - 11e_3 + 7e_4 - 2e_5 = {} & 5(a + x_1 d) + 6(a + x_2 d) - 11(a + x_3 d) + 7(a + x_4 d) - 2(a + x_5 d) \\ = {} & (5 + 6 - 11)a + d(5x_1 + 6x_2 - 11x_3) + (7 - 2)a + d(7x_4 - 2x_5). \\ = {} & (5 + 6 - 11)a + d(5x_1 + 6x_2 - 11x_3 + 7x_4 - 2x_5) + 5a. \end{aligned}$$

GOOD NEWS: The first a has coefficient $(5 + 6 - 11) = 0$.

GOOD NEWS: We can pick x_1, x_2, x_3, x_4, x_5 to make the $5x_1 + 6x_2 - 11x_3 + 7x_4 - 2x_5 = 0$.

BAD NEWS: The $5a$ looks hard to get rid of.

It would be really great if we did not have that '$5a$' term.

Hence we need a variant of van der Waerden's theorem.

Variant of VDW

Lemma 10.2.7 *For all k, s, c, there exists $U = U(k, s, c)$ such that for every c-coloring $\chi : [U] \to [c]$ there exists a, d such that*

$$\chi(a) = \chi(a + d) = \cdots = \chi(a + (k-1)d) = \chi(sd)$$

Proof: We prove this by induction on c. Clearly, for all k, s,

$$U(k, s, 1) = \max\{k, s\}.$$

We assume $U(k, s, c-1)$ exists and show that $U(k, s, c)$ exists. We will show that

$$U(k, s, c) \leq W((k-1)sU(k, s, c-1) + 1, c).$$

Let χ be a coloring of $[W((k-1)sU(k, s, c-1) + 1, c)]$. By the definition of W there exists a, d such that

$$\chi(a) = \chi(a + d) = \cdots = \chi(a + (k-1)sU(k, s, c-1)d).$$

Assume the color is RED. There are several cases.

Case 1: If sd is RED then since $a, a + d, \ldots, a + (k-1)d$ are all RED, we are done.

Case 2: If $2sd$ is REDthen since. $a, a + 2d, a + 4d, \ldots, a + 2(k-1)d$ are all RED, we are done.

$\vdots$

Case U(k,s,c-1): If $U(k, s, c-1)sd$ is REDthen since
a,$a + U(k, s, c-1)d$,$a + 2U(k, s, c-1)d$,$\ldots$, $a + (k-1)U(k, s, c-1)d$
are all RED, we are done.

Case U(k,s,c-1)sd+1: None of the above cases happen. Hence
$sd, 2sd, 3sd, \ldots, U(k, s, c-1)sd$
are all NOT RED.
Consider the coloring $\chi' : [U(k, s, c-1)] \to [c-1]$ defined by

$$\chi'(x) = \chi(xsd).$$

The KEY is that NONE of these will be colored REDso there are only $c-1$ colors. By the inductive hypothesis there exists a', d' such that

$$\chi'(a') = \chi'(a' + d') = \cdots = \chi'(a' + (k-1)d') = \chi'(sd')$$

so

$$\chi(a'sd) = \chi(a'sd + d'sd) = \cdots = \chi(a'sd + (k-1)d'sd) = \chi(sd'sd)$$

Let $A = a'sd$ and $D = d'sd$. Then

$$\chi(A) = \chi(A + D) = \cdots = \chi(A + (k-1)D = \chi(sD).$$

▮

FILL IN - NEED FIGURE

Back to the Proof

We now restate and prove the main theorem of this section.

Theorem 10.2.8 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n$. If there exists a nonempty subset of $\{b_1, \ldots, b_n\}$ that sums to 0 then $(b_1, \ldots, b_n)$ is regular.*

Proof: The cases of $n = 1$ and $n = 2$ are easy and left to the reader. Hence we assume $n \geq 3$. If any of the b_i's are 0 then we can omit the term with that b_i. So we can assume that $(\forall i)[b_i \neq 0]$.

By renumbering we can assume that there is an $m \leq n$ such that

$$\sum_{i=1}^{m} b_i = 0.$$

We need to find a number R such that if χ is a c-coloring of R then blah-de-blah. Instead we will let χ be a c-coloring of $\mathbb{N}$. We leave it to the reader to extract out a finite R from our proof.

Let χ be a c-coloring of $\mathbb{N}$. We will determine $x_1, \ldots, x_m \in \mathbb{Z} - \{0\}$ and $s \in \mathbb{N}$ later. By Lemma 10.2.7 there exists a, d such that

$$\chi(a) = \chi(a + x_1 d) = \chi(a + x_2 d) = \cdots = \chi(x + x_m d) = \chi(sd).$$

We will let

$$\begin{aligned} e_1 &= a + x_1 d, \\ e_2 &= a + x_2 d, \\ &\vdots \\ e_m &= a + x_m d, \end{aligned}$$

and

$$e_{m+1} = \cdots = e_n = sd$$

Then

$$\sum_{i=1}^{n} b_i e_i = \sum_{i=1}^{m} b_i e_i + \sum_{i=m+1}^{n} b_i e_i = \sum_{i=1}^{m} b_i(a + x_i d) + \sum_{i=m+1}^{n} b_i sd.$$

This is equal to

$$a\sum_{i=1}^{m} b_i + d\sum_{i=1}^{m} b_i x_i + sd\sum_{i=m+1}^{n} b_i$$

KEY: $\sum_{i=1}^{m} b_i = 0$ so the first term drops out.

KEY: All of the remaining terms have a factor of d. If we want to set this to 0 we can cancel the d's. Hence we need $x_1, \ldots, x_n \in \mathbb{Z} - \{0\}$ and $s \in \mathbb{N}$ such that the following happens.

$$\sum_{i=1}^{m} b_i x_i + s\sum_{i=m+1}^{n} b_i = 0.$$

Let $\sum_{i=m+1}^{n} b_i = B$. Then we rewrite this as

$$\sum_{i=1}^{m} b_i x_i + sB = 0.$$

We can take

$$s = |mb_1 \cdots b_m|$$

$$x_1 = -\frac{sB}{mb_1}$$
$$x_2 = -\frac{sB}{mb_2}$$
$$\vdots$$
$$x_m = -\frac{sB}{mb_m}.$$

∎

10.2.3 If ... then $(b_1, \ldots, b_n)$ is Distinct-Regular

We will prove the following

Theorem 10.2.9 *If $(b_1, b_2, \ldots, b_n)$ is regular and there exists $(\lambda_1, \ldots, \lambda_n)$ such that $\sum_{i=1}^{n} \lambda_i b_i = 0$ then $(b_1, \ldots, b_n)$ is distinct-regular.*

To prove this we need a Key Lemma:

Key Lemma

The lemma is in three parts. The first one we use to characterize which vectors are distinct-regular. The second and third are used in a later section when we prove the matrix Rado Theorem.

The following definitions are used in the third part of the lemma.

Def 10.2.10 Let $n \in \mathbb{N}$.

1. A set $G \subseteq \mathbb{N}^n$ is *homogenous* if, for all $\alpha \in \mathbb{N}$,
$$(e_1, \ldots, e_n) \in G \implies (\alpha e_1, \ldots, \alpha e_n) \in G.$$
2. A set $G \subseteq \mathbb{N}^n$ is *regular* if, for all c, there exists $R = R(G; c)$ such that the following holds: For all c-colorings $\chi : [R] \to [c]$ there exists $\vec{e} = (e_1, \ldots, e_n) \in G$ such that all of the e_i's are colored the same.

Example 10.2.11

1. Let $b_1, \ldots, b_n \in \mathbb{Z}$. Let $G = \{(e_1, \ldots, e_n) \mid \sum_{i=1}^{n} b_i e_i = 0\}$. G is homogenous. $(b_1, \ldots, b_n)$ is regular iff G is regular.
2. Let M be an $n \times m$ matrix. Let $G = \{\vec{e} \mid M\vec{e} = \vec{0}\}$. G is homogenous.

Lemma 10.2.12

1. *For all $(b_1, \ldots, b_n) \in \mathbb{Z}^n$ regular, $c, M \in \mathbb{N}$ there exists $L = L(b_1, \ldots, b_n; c, M)$ such that the following holds. For any c-coloring $\chi : [L] \to [c]$. Then there exists $e_1, \ldots, e_n, d \in [L]$ such that the following occurs.*
   (a) $b_1 e_1 + \cdots + b_n e_n = 0$

(b) *The following are the same color:*

$$\begin{array}{cccccccc} e_1 - Md, & e_1 - M(d-1) & ,\ldots, & e_1 - d, & e_1, & e_1 + d & ,\ldots, & e_1 + Md \\ e_2 - Md, & e_2 - M(d-1) & ,\ldots, & e_2 - d, & e_2, & e_2 + d & ,\ldots, & e_2 + Md \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ e_n - Md, & e_n - M(d-1) & ,\ldots, & e_n - d, & e_n, & e_n + d & ,\ldots, & e_n + Md \end{array}$$

2. *For all* $(b_1, \ldots, b_n) \in \mathbb{Z}^n$ *regular,* $c, M, s \in \mathbb{N}$ *there exists* $L_2 = L_2(b_1, \ldots, b_n; c, M, s)$ *such that the following holds. For any c-coloring* $\chi : [L_2] \to [c]$*. Then there exists* $e_1, \ldots, e_n, d \in [L]$ *such that the following occurs.*

   (a) $b_1e_1 + \cdots + b_ne_n = 0$

   (b) *The following are the same color:*

$$\begin{array}{cccccccc} e_1 - Md, & e_1 - M(d-1) & ,\ldots, & e_1 - d, & e_1, & e_1 + d & ,\ldots, & e_1 + Md \\ e_2 - Md, & e_2 - M(d-1) & ,\ldots, & e_2 - d, & e_2, & e_2 + d & ,\ldots, & e_2 + Md \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ e_n - Md, & e_n - M(d-1) & ,\ldots, & e_n - d, & e_n, & e_n + d & ,\ldots, & e_n + Md \end{array}$$

   (c) $sd \in [L_2]$ *is the same color as the numbers in the last item.*

3. *For all* $n \in \mathbb{N}$*, for all* $G \subseteq \mathbb{N}^n$*, G regular and homogenous, for all* $c, M, s \in \mathbb{N}$ *there exists* $L_3 = L_3(G; c, M, s)$ *such that the following holds. For any c-coloring* $\chi : [L_3] \to [c]$*. Then there exists* $e_1, \ldots, e_n, d \in [L_3]$ *such that the following occurs.*

   (a) $(e_1, \ldots, e_n) \in G$.

   (b) *The following are the same color:*

$$\begin{array}{cccccccc} e_1 - Md, & e_1 - M(d-1) & ,\ldots, & e_1 - d, & e_1, & e_1 + d & ,\ldots, & e_1 + Md \\ e_2 - Md, & e_2 - M(d-1) & ,\ldots, & e_2 - d, & e_2, & e_2 + d & ,\ldots, & e_2 + Md \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ e_n - Md, & e_n - M(d-1) & ,\ldots, & e_n - d, & e_n, & e_n + d & ,\ldots, & \end{array}$$

   (c) $sd \in [L_3]$ *is the same color as the numbers in the last item.*

Proof:
1) Since $b_1, \ldots, b_n$ is regular, by Theorem 10.2.8 (or Definition 10.1.1) there exists $R = R(b_1, \ldots, b_n, c)$ such that for any c-coloring of $[R]$ there exists $e_1, \ldots, e_n$ such that (1) all of the e_i's are the same color, and (2) $\sum_{i=1}^{n} b_i e_i = 0$.

We determine L later; however, we note *conditions* on L when they arise. Let $\chi : [L] \to [c]$.

Condition 1 on L**:** R divides L.

We define $\chi^* : [L/R] \to [c]^R$ as follows.

$$\chi^*(n) = \chi(n)\chi(2n)\chi(3n)\cdots\chi(Rn).$$

(This is concatenation, not multiplication.)

Condition 2 on L**:** $L/R \geq W(X, c^R)$ where we will determine X later. We rewrite this $L \geq RW(X, c^R)$.

Apply (a slight variant of) VDW to χ with length-of-sequence set to X (to be determined later) and number-of-colors set to c^R to obtain the following: There exists a, D (we use D instead of d since this value will not be the final d we need for our conclusion) such that

$$\chi^*(a-XD) = \chi^*(a-(X-1)D) = \cdots = \chi^*(a) = \cdots = \chi^*(a+(X-1)D) = \chi^*(a+XD).$$

Note that

$$\begin{aligned}
\chi^*(a - XD) = & \quad \chi(a - XD)\chi(2(a - XD))\cdots\chi(R(a - XD)) \\
\chi^*(a - (X - 1)D) = & \quad \chi(a - (X - 1)D)\chi(2(a - (X - 1)D))\cdots\chi(R(a - (X - 1)D)) \\
\vdots = & \quad \vdots \\
\chi^*(a - D) = & \quad \chi(a - D)\chi(2(a - D))\cdots\chi(R(a - D)) \\
\chi^*(a) = & \quad \chi(a)\chi(2a)\cdots\chi(Ra) \\
\chi^*(a + D) = & \quad \chi(a + D)\chi(2(a + D))\cdots\chi(R(a + D)) \\
\vdots = & \quad \vdots \\
\chi^*(a + (X - 1)D) = & \quad \chi(a + (X - 1)D)\chi(2(a + (X - 1)D))\cdots\chi(R(a + (X - 1)D)) \\
\chi^*(a + XD) = & \quad \chi(a + XD)\chi(2(a + XD))\cdots\chi(R(a + XD))
\end{aligned}$$

Since

$$\chi^*(a-XD) = \chi^*(a-(X-1)D) = \cdots = \chi^*(a) = \cdots = \chi^*(a+(X-1)D) = \chi^*(a+XD)$$

we have

$$\begin{array}{rllllllll}
\chi(a-XD) = & \chi(a-(X-1)D) = & \cdots = & \chi(a) = & \cdots = & \chi(a+XD) \\
\chi(2(a-XD)) = & \chi(2(a-(X-1)D)) = & \cdots = & \chi(2a) = & \cdots = & \chi(2(a+XD)) \\
\chi(3(a-XD)) = & \chi(3(a-(X-1)D)) = & \cdots = & \chi(3a) = & \cdots = & \chi(3(a+XD)) \\
\vdots = & \vdots \quad = & \cdots = & \vdots \quad = & \cdots = & \vdots \\
\chi(R(a-XD)) = & \chi(R(a-(X-1)D)) = & \cdots = & \chi(Ra) = & \cdots = & \chi(R(a+XD))
\end{array}$$

Condition 3 on L: $L \geq R(a+XD)$. Since $a, D \leq W(X, c^R)$ this means $L \geq R(W(X, c^R) + XW(X, c^R))$. We simplify this by using the slightly worse bound $L \geq 2RXW(X, c^R)$. This bound implies Condition 2,hence we only need Condition's 1 and 3.

We need a subset of these that are all the same color. Consider the coloring $\chi^{**} : [R] \rightarrow [c]$ defined by

$$\chi^{**}(n) = \chi(na).$$

By the definition of R there exists $f_1, \ldots, f_n$ such that

1. $\sum_{i=1}^{n} b_i f_i = 0$. Hence $\sum_{i=1}^{n} b_i(af_i) = a \sum_{i=1}^{n} b_i f_i = 0$

2. $$\chi^{**}(f_1) = \chi^{**}(f_2) = \cdots = \chi^{**}(f_n).$$

   By the definition of χ^{**} we have

   $$\chi(af_1) = \chi(af_2) = \cdots = \chi(af_n).$$

Note that we have that the following are *all* the same color.

$$\begin{array}{llllll}
f_1(a-XD), & f_1(a-(X-1)D) & ,\cdots, & f_1 a & ,\cdots, & f_1(a+XD) \\
f_2(a-XD), & f_2(a-(X-1)D) & ,\cdots, & f_2 a & ,\cdots, & f_2(a+XD) \\
f_3(a-XD), & f_3(a-(X-1)D) & ,\cdots, & f_3 a & ,\cdots, & f_3(a+XD) \\
\vdots & \vdots & ,\cdots, & \vdots & ,\cdots, & \vdots \\
f_n(a-XD), & f_n(a-(X-1)D) & ,\cdots, & f_n a & ,\cdots, & f_n(a+XD)
\end{array}$$

For all i, $1 \leq i \leq n$ let $e_i = af_i$. We rewrite the above.

$$\begin{array}{llllll}
e_1 - f_1XD, & e_1 - f_1(X-1)D) & ,\cdots, & e_1 & ,\cdots, & e_1 + f_1XD) \\
e_2 - f_2XD, & e_2 - f_2(X-1)D) & ,\cdots, & e_2 & ,\cdots, & e_2 + f_2XD) \\
e_3 - f_3XD, & e_3 - f_3(X-1)D) & ,\cdots, & e_3 & ,\cdots, & e_3 + f_3XD) \\
\vdots & \vdots & ,\cdots, & \vdots & ,\cdots, & \vdots \\
e_n - f_nXD, & e_n - f_n(X-1)D) & ,\cdots, & e_n & ,\cdots, & e_n + f_nXD)
\end{array}$$

We are almost there- we have our $e_1, \ldots, e_n$ that are the same color, and lots of additive terms from them are also that color. We need a value of d such that

$$\{d, 2d, 3d, \ldots, Md\} \subseteq \{f_1D, 2f_1D, 3f_1D, \ldots, Xf_1D\}$$
$$\{d, 2d, 3d, \ldots, Md\} \subseteq \{f_2D, 2f_2D, 3f_2D, \ldots, Xf_2D\}$$
$$\vdots$$
$$\{d, 2d, 3d, \ldots, Md\} \subseteq \{f_nD, 2f_nD, 3f_nD, \ldots, Xf_nD\}$$

We have no control over D. We have complete control over X. We know that, for all i, $f_i \le R$. Let $X = 3R^nM$ ($X = 2R^nM + 1$ would suffice but we take the worse bound for easy manipulation.) Let $d = f_1f_2\cdots f_nD$. For $1 \le i \le M$ we have

$$id = if_1f_2\cdots f_nD$$

We need

$$\begin{array}{ll}
\{f_1f_2\cdots f_n, 2f_1f_2\cdots f_n, 3f_1f_2\cdots f_n, \ldots, Mf_1f_2\cdots f_n\} & \subseteq \{f_1, 2f_1, 3f_1, \ldots, R^{n-1}Mf_1\} \\
\{f_1f_2\cdots f_n, 2f_1f_2\cdots f_n, 3f_1f_2\cdots f_n, \ldots, Mf_1f_2\cdots f_n\} & \subseteq \{f_2, 2f_2, 3f_2, \ldots, R^{n-1}Mf_2\} \\
 & \vdots \\
\{f_1f_2\cdots f_n, 2f_1f_2\cdots f_n, 3f_1f_2\cdots f_n, \ldots, Mf_1f_2\cdots f_n\} & \subseteq \{f_n, 2f_n, 3f_n, \ldots, R^{n-1}Mf_n\}
\end{array}$$

Since $i \le M$ and $f_1\cdots f_n \le R^n$, we have

$$if_1f_2\cdots f_n \le MR^n$$

Hence we have what we need.

Since $X = 3R^nM$ we can now determine L. By condition 1 and 3 on L we can take $L = 2RXW(3R^nM, c^R)$ where $R = R(b_1, \ldots, b_n; c)$.

2) We prove this by induction on c.

Base Case: For $c = 1$ this is easy; however, we find the actual bound anyway. The only issue here is to make sure that the objects we want to color are actually in $L(b_1, \ldots, b_n; 1, M, s)$. Let $(e_1, \ldots, e_n) \in \mathbb{N}^n$ be a solution to $\sum_{i=1}^n b_i e_i = 0$ such that $e = \min\{e_1, \ldots, e_n\} > M$. Let $L_2 = L_2(b_1, \ldots, b_n; 1, M, s) = \max\{e + M, s\}$. Let $\chi : [L_2] \to [1]$. We claim that $e_1, \ldots, e_n, 1$ work. Note that, for all i, $1 \le i \le n$, for all j $-M \le j \le M$, $e_i + j \times 1 \in [L_2]$. Also note that $s \times 1 \in [L_2]$.

Induction Hypothesis: We assume the theorem is true for $c-1$ colors. In particular, for any M', $L_2(b_1, \ldots, b_n; c-1, M', s)$ exists. This proof will be similar to the proof of Lemma 10.2.7.

Induction Step: We want to show that $L_2(b_1, \ldots, b_n; c, M, s)$ exists. We show that there exists M' such that if you color $L = L(b_1, \ldots, b_n; c, M')$ (note that this is L not L_2) there exists the required $e_1, \ldots, e_n, d$. The M' will depend on L_2. Let χ be a c-coloring of $[L]$. By part 1 there exists $E_1, \ldots, E_n, D$ such that $\sum_{i=1}^n b_i E_i = 0$ and the following are all the same color, which we will call RED.

$$\begin{array}{cccccccc}
E_1 - M'D, & E_1 - M'(D-1) & ,\ldots, & E_1 - D, & E_1, & E_1 + D & ,\ldots, E_1 + M'D \\
E_2 - M'D, & E_2 - M'(D-1) & ,\ldots, & E_2 - D, & E_2, & E_2 + D & ,\ldots, E_2 + M'D \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
E_n - M'D, & E_n - M'(D-1) & ,\ldots, & E_n - D, & E_n, & E_n + D & ,\ldots, E_n + M'D
\end{array}$$

For $1 \le i \le M'$ consiDer

$$\begin{array}{cccccccc}
E_1 - iD, & E_1 - i(D-1) & ,\ldots, & E_1 - D, & E_1, & E_1 + D & ,\ldots, E_1 + iD \\
E_2 - iD, & E_2 - i(D-1) & ,\ldots, & E_2 - D, & E_2, & E_2 + D & ,\ldots, E_2 + iD \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
E_n - iD, & E_n - i(D-1) & ,\ldots, & E_n - D, & E_n, & E_n + D & ,\ldots, E_n + iD
\end{array}$$

There are now several cases.

Case 1: If sD is REDthen we are done so long as $M' \ge M$. Use $d = D$.

Case 2: If $2sD$ is REDthen we are done so long as $M' \ge 2M$. Use $d = 2D$.

$\vdots$

Case M'sD: If $M'sD$ is RED then so long as $M' \ge M^2$ we are done. Use $d = M'D$.

FILL IN - CHECK THIS CAREFULLY

Case M'sD+1: None of the above cases hold. Hence

$sD, 2sD, \ldots, M'sD$

are all NOT RED. Hence the coloring restricted to this set is a $c-1$ coloring. Let $M' = L_2(b_1, \ldots, b_n; c-1, M, s)$. Consider the coloring $\chi^*[M'] \rightarrow [c-1]$ defined by

$$\chi^*(x) = \chi(xsD).$$

By the induction hypothesis and the definition of M' there exists $e_1, \ldots, e_n, d$ such that $\sum_{i=1}^{n} b_i e_i = 0$ and all of the following are the same color via χ^*:

1.
$$\begin{array}{ccccccc} e_1 - Md, & e_1 - M(d-1) & , \ldots, & e_1 - d, & e_1, & e_1 + d & , \ldots, e_1 + Md \\ e_2 - Md, & e_2 - M(d-1) & , \ldots, & e_2 - d, & e_2, & e_2 + d & , \ldots, e_2 + Md \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\vdots \\ e_n - Md, & e_n - M(d-1) & , \ldots, & e_n - d, & e_n, & e_n + d & , \ldots, e_n + Md \end{array}$$

2. sd

By the definition of χ^* the following have the same color via χ:

1.
$$\begin{array}{ccccccc} e_1 s - Mds, & e_1 s - M(d-1)s & , \ldots, & e_1 s - ds, & e_1 s, & e_1 s + ds, & \ldots, e_1 s + Mds \\ e_2 s - Mds, & e_2 s - M(d-1)s & , \ldots, & e_2 s - ds, & e_2 s, & e_2 s + ds, & \ldots, e_2 s + Mds \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\vdots \\ e_n s - Mds, & e_n s - M(d-1)s & , \ldots, & e_n s - ds, & e_n s, & e_n s + ds, & \ldots, e_n s + Mds \end{array}$$

2. $s^2 d$

By setting e_i to $e_i s$ and d to ds we obtain the result.

3) In both of the above parts the only property of the set

$$\{(x_1, \ldots, x_n) \mid \sum_{i=1}^{n} b_i x_i = 0\}$$

that we used is that it was homogenous and regular. Hence all of the proofs go through without any change and we obtain this part of the lemma. ∎

Back to our Story

Theorem 10.2.13 *If $(b_1, b_2, \ldots, b_n)$ is regular and there exists $(\lambda_1, \ldots, \lambda_n)$ such that $\sum_{i=1}^{n} \lambda_i b_i = 0$ and all of the λ_i are distinct then $(b_1, \ldots, b_n)$ is distinct-regular.*

Proof: Let M be a parameter to be picked later. Let $L = L(b_1, \ldots, b_n; c, M)$ from Lemma 10.2.12.1. Let COL be a c-coloring of $]L]$. We show that there exists $e_1, \ldots, e_n, d \in [L]$ such that the following occurs.

1. $b_1 e_1 + \cdots + b_n e_n = 0$
2. The following are the same color:

$$\begin{array}{cccccccc} e_1 - Md, & e_1 - M(d-1) & ,\ldots, & e_1 - d, & e_1, & e_1 + d & ,\ldots, & e_1 + Md \\ e_2 - Md, & e_2 - M(d-1) & ,\ldots, & e_2 - d, & e_2, & e_2 + d & ,\ldots, & e_2 + Md \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ e_n - Md, & e_n - M(d-1) & ,\ldots, & e_n - d, & e_n, & e_n + d & ,\ldots, & e_n + Md \end{array}$$

Let $A \in \mathbb{Z}$ be a constant to be picked later. Note that

$$\sum_{i=1}^{n} b_i(e_i + Ad\lambda_i) = (\sum_{i=1}^{n} b_i e_i) + (Ad \sum_{i=1}^{n} b_i \lambda_i) = 0.$$

So we need M to be such that there exists an A with

1. $e_1 + Ad\lambda_1$, …, $e_n + Ad\lambda_n$ are all distinct, and
2. For all i, $|A\lambda_i| \le M$

There are at most $\binom{n}{2}$ values of A that make item 1 false. In order to satisfy item 2 we need, for all i, $|A| \le M/\lambda_i$. Let $\lambda = \max\{|\lambda_1|, \ldots, |\lambda_n|\}$. We let $M = 2\binom{n}{2}\lambda$. In order to satisfy item 2 we need, for all i,

$$|A| \le 2\binom{n}{2}\lambda/|\lambda_i|.$$

It will suffice to take $|A| \le 2\binom{n}{2}$. There are clearly more than $\binom{n}{2}$ values of A that satisfy this, hence we can find one that satisfies item 1.

∎

10.3 The Full Rado's Theorem

Recall when a matrix is regular:

Def 10.3.1 A matrix A of integers is *regular* if if the following holds: *For all c, for all c-colorings* $\chi : \mathbb{N} \rightarrow [c]$ *there exists* $\vec{e} = e_1, \ldots, e_n$ *such that*

$$\chi(e_1) = \cdots = \chi(e_n),$$

$$A\vec{e} = \vec{0}.$$

Def 10.3.2 A matrix A satisfies the *columns condition* if the columns can be ordered $\vec{c}_1, \ldots, \vec{c}_n$ and the set $\{1, \ldots, n\}$ can be partitioned into nonempty contigous sets $I_1, \ldots, I_k$ such that

$$\sum_{i \in I_1} \vec{c}_i = \vec{0},$$

For all j, $2 \leq j \leq k$, $\sum_{i \in I_j} \vec{c}_i$ can be written as a linear combination of the vectors $\{c_i\}_{i \in I_1 \cup \cdots \cup I_{j-1}}$.

We will prove the following:
The Full Rado's Theorem:

Theorem 10.3.3 *A is regular iff A satisfies the columns condition.*

FILL IN THE PROOF

10.4 Coloring $\mathbb{R}$*

(This section was co-written with Steven Fenner.)
Do you think the following is TRUE or FALSE?
For any $\aleph_0$*-coloring of the reals,* $\chi : \mathbb{R} \rightarrow \mathbb{N}$ *there exist distinct* e_1, e_2, e_3, e_4 *such that*

$$\chi(e_1) = \chi(e_2) = \chi(e_3) = \chi(e_4),$$

$$e_1 + e_2 = e_3 + e_4.$$

It turns out that this question is equivalent to the negation of CH. Komjáth [17] claims that Erdős proved this result. The prove we give is due to Davies [6].

Def 10.4.1 The *Continuum Hypothesis* (CH) is the statement that there is no order of infinity between that of $\mathbb{N}$ and $\mathbb{R}$. It is known to be independent of Zermelo-Frankel Set Theory with Choice (ZFC).

Def 10.4.2 ω_1 is the first uncountable ordinal. ω_2 is the second uncountable ordinal.

Fact 10.4.3

1. *If CH is true, then there is a bijection between* $\mathbb{R}$ *and* ω_1*. This has the followng counter-intuitive consequence: there is a way to list the reals:*

$$x_0, x_1, x_2, \ldots, x_\alpha, \ldots$$

   as $\alpha \in \omega_1$ *such that, for all* $\alpha \in \omega_1$*, the set* $\{x_\beta \mid \beta < \alpha\}$ *is countable.*

2. *If CH is false, then there is an injection from* ω_2 *to* $\mathbb{R}$*. This has the consequence that there is a list of distinct reals:*

$$x_0, x_1, x_2, \ldots, x_\alpha, \ldots, x_{\omega_1}, x_{\omega_1+1}, \ldots, x_\beta, \ldots$$

   where $\alpha \in \omega_1$ *and* $\beta \in [\omega_1, \omega_2)$*.*

10.4.1 CH $\implies$ FALSE

Def 10.4.4 Let $X \subseteq \mathbb{R}$. Then $CL(X)$ is the smallest set $Y \supseteq X$ of reals such that

$$a, b, c \in Y \implies a + b - c \in Y.$$

Note 10.4.5 $X \subseteq CL(X)$ since we can take $b = c$.

Lemma 10.4.6

1. *If* X *is countable then* $CL(X)$ *is countable.*

2. *If* $X_1 \subseteq X_2$ *then* $CL(X_1) \subseteq CL(X_2)$*.*

Proof:
1) Assume X is countable. $CL(X)$ can be defined with an ω-induction (that is, an induction just through ω).

$$\begin{aligned} C_0 &= X \\ C_{n+1} &= C_n \cup \{a+b-c \mid a,b,c \in C_n\} \end{aligned}$$

One can easily show that $CL(X) = \cup_{i=0}^{\infty} C_i$ and that this set is countable.
2) This is an easy exercise. ▮

Theorem 10.4.7 *Assume CH is true. There exists an $\aleph_0$-coloring of $\mathbb{R}$ such that there are no distinct e_1, e_2, e_3, e_4 such that*

$$\chi(e_1) = \chi(e_2) = \chi(e_3) = \chi(e_4),$$

$$e_1 + e_2 = e_3 + e_4.$$

Proof: Since we are assuming CH is true, we have, by Fact 10.4.3.1, there is a bijection between $\mathbb{R}$ and ω_1. If $\alpha \in \omega_1$ then x_α is the real associated to it. We can picture the reals as being listed out via

$$x_0, x_1, x_2, x_3, \ldots, x_\alpha, \ldots$$

where $\alpha < \omega_1$.

Note that every number has only countably many numbers less than it in this ordering.

For $\alpha < \omega_1$ let

$$X_\alpha = \{x_\beta \mid \beta < \alpha\}.$$

Note the following:

1. For all α, X_α is countable.
2. $X_0 \subset X_1 \subset X_2 \subset X_3 \subset \cdots \subset X_\alpha \subset \cdots$
3. $\bigcup_{\alpha<\omega_1} X_\alpha = \mathbb{R}$.

We define another increasing sequence of sets Y_α by letting

$$Y_\alpha = CL(X_\alpha).$$

Note the following:

1. For all α, Y_α is countable. This is from Lemma 10.4.6.1.
2. $Y_0 \subset Y_1 \subset Y_2 \subset Y_3 \subset \cdots \subset Y_\alpha \subset \cdots$. This is from Lemma 10.4.6.2.
3. $\bigcup_{\alpha<\omega_1} Y_\alpha = \mathbb{R}$.

We now define our last sequence of sets:
For all $\alpha < \omega_1$,

$$Z_\alpha = Y_\alpha - \left(\bigcup_{\beta<\alpha} Y_\beta \right).$$

Note the following:

1. Each Z_α is finite or countable.
2. The Z_α form a partition of $\mathbb{R}$.

We will now define an $\aleph_0$-coloring of $\mathbb{R}$. For each Z_α, which is countable, assign colors from ω to Z_α's elements in some way so that no two elements of Z_α have the same color.

Assume, by way of contradiction, that there are distinct e_1, e_2, e_3, e_4 such that

$$\chi(e_1) = \chi(e_2) = \chi(e_3) = \chi(e_4)$$

and

$$e_1 + e_2 = e_3 + e_4.$$

Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be such that $e_i \in Z_{\alpha_i}$. Since all of the elements in any Z_α are colored differently, all of the α_i's are different. We will assume $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4$. The other cases are similar. Note that

$$e_4 = e_1 + e_2 - e_3.$$

and

$$e_1, e_2, e_3 \in Z_{\alpha_1} \cup Z_{\alpha_2} \cup Z_{\alpha_3} \subseteq Y_{\alpha_1} \cup Y_{\alpha_2} \cup Y_{\alpha_3} = Y_{\alpha_3}.$$

Since $Y_{\alpha_3} = CL(X_{\alpha_3})$ and $e_1, e_2, e_3 \in Y_{\alpha_3}$, we have $e_4 \in Y_{\alpha_3}$. Hence $e_4 \notin Z_{\alpha_4}$. This is a contradiction. ∎

What was it about the equation

$$e_1 + e_2 = e_3 + e_4$$

that made the proof of Theorem 10.4.7 work? Absolutely nothing:

Theorem 10.4.8 *Let $n \geq 2$. Let $a_1, \ldots, a_n \in \mathbb{R}$ be nonzero. Assume CH is true. There exists an $\aleph_0$-coloring of $\mathbb{R}$ such that there are no distinct $e_1, \ldots, e_n$ such that*

$$\chi(e_1) = \cdots = \chi(e_n),$$

$$\sum_{i=1}^{n} a_i e_i = 0.$$

Proof sketch: Since this prove is similar to the last one we just sketch it.

Def 10.4.9 Let $X \subseteq R$. $CL(X)$ is the smallest superset of X such that the following holds:

For all $m \in \{1, \ldots, n\}$ and for all $e_1, \ldots, e_{m-1}, e_{m+1}, \ldots, e_n$,

$$e_1, \ldots, e_{m-1}, e_{m+1}, \ldots, e_n \in CL(X) \implies -(1/a_m) \sum_{i \in \{1,\ldots,n\} - \{m\}} a_i e_i \in CL(X).$$

Let X_α, Y_α, Z_α be defined as in Theorem 10.4.7 using this new defintion of CL. Let χ be defined as in Theorem 10.4.7.

Assume, by way of contradiction, that there are distinct $e_1, \ldots, e_n$ such that

$$\chi(e_1) = \cdots = \chi(e_n)$$

and

$$\sum_{i=1}^{n} a_i e_i = 0.$$

Let $\alpha_1, \ldots, \alpha_n$ be such that $e_i \in Z_{\alpha_i}$. Since all of the elements in any Z_α are colored differently, all of the α_i's are different. We will assume $\alpha_1 < \alpha_2 < \cdots < \alpha_n$. The other cases are similar. Note that

$$e_n = -(1/a_n) \sum_{i=1}^{n-1} a_i e_i \in CL(X)$$

and

$$e_1, \ldots, e_{n-1} \in Z_{\alpha_1} \cup \cdots \cup Z_{\alpha_{n-1}} \subseteq Y_{\alpha_{n-1}}.$$

Since $Y_{\alpha_{n-1}} = CL(X_{\alpha_{n-1}})$ and $e_1, \ldots, e_{n-1} \in Y_{\alpha_{n-1}}$, we have $e_n \in Y_{\alpha_{n-1}}$. Hence $e_n \notin Z_{\alpha_n}$. This is a contradiction. ∎

FILL IN -LOOK UP PAPER THIS CAME FROM TO GET MORE

10.4.2 $\neg$ CH $\implies$ TRUE

Theorem 10.4.10 *Assume CH is false. Let χ be an $\aleph_0$-coloring of $\mathbb{R}$. There exist distinct e_1, e_2, e_3, e_4 such that*

$$\chi(e_1) = \chi(e_2) = \chi(e_3) = \chi(e_4),$$

$$e_1 + e_2 = e_3 + e_4.$$

Proof: By Fact 10.4.3 there is an injection of ω_2 into $\mathbb{R}$. If $\alpha \in \omega_2$, then x_α is the real associated to it.

Let χ be an $\aleph_0$-coloring of $\mathbb{R}$. We show that there exist distinct e_1, e_2, e_3, e_4 of the same color such that $e_1 + e_2 = e_3 + e_4$.

We define a map F from ω_2 to $\omega_1 \times \omega_1 \times \omega_1 \times \omega$.

1. Let $\beta \in \omega_2$.
2. Define a map from ω_1 to ω by
$$\alpha \mapsto \chi(x_\alpha + x_\beta).$$
3. Let $\alpha_1, \alpha_2, \alpha_3 \in \omega_1$ be distinct elements of ω_1, and $i \in \omega$, such that $\alpha_1, \alpha_2, \alpha_3$ all map to i. Such $\alpha_1, \alpha_2, \alpha_3, i$ clearly exist since $\aleph_0 + \aleph_0 = \aleph_0 < \aleph_1$. (There are $\aleph_1$ many elements that map to the same element of ω, but we do not need that.)
4. Map β to $(\alpha_1, \alpha_2, \alpha_3, i)$.

Since F maps a set of cardinality $\aleph_2$ to a set of cardinality $\aleph_1$, there exists some element that is mapped to twice by F (actually there is an element that is mapped to $\aleph_2$ times, but we do not need this). Let $\alpha_1, \alpha_2, \alpha_3, \beta, \beta', i$ be such that $\beta \neq \beta'$ and

$$F(\beta) = F(\beta') = (\alpha_1, \alpha_2, \alpha_3, i).$$

Choose distinct
$$\alpha, \alpha' \in \{\alpha_1, \alpha_2, \alpha_3\}$$
such that
$$x_\alpha - x_{\alpha'} \notin \{x_\beta - x_{\beta'}, x_{\beta'} - x_\beta\}.$$
We can do this because there are at least two positive values for $x_\alpha - x_{\alpha'}$.

Since $F(\beta) = (\alpha_1, \alpha_2, \alpha_3, i)$, we have

$$\chi(x_\alpha + x_\beta) = \chi(x_{\alpha'} + x_\beta) = i.$$

Since $F(\beta') = (\alpha_1, \alpha_2, \alpha_3, i)$, we have

$$\chi(x_\alpha + x_{\beta'}) = \chi(x_{\alpha'} + x_{\beta'}) = i.$$

Let

$$\begin{aligned} e_1 &= x_\alpha + x_\beta \\ e_2 &= x_{\alpha'} + x_{\beta'} \\ e_3 &= x_{\alpha'} + x_\beta \\ e_4 &= x_\alpha + x_{\beta'}. \end{aligned}$$

Then

$$\chi(e_1) = \chi(e_2) = \chi(e_3) = \chi(e_4)$$

and

$$e_1 + e_2 = e_3 + e_4.$$

Since $x_\alpha \neq x_{\alpha'}$ and $x_\beta \neq x_{\beta'}$, we have $\{e_1, e_2\} \cap \{e_3, e_4\} = \emptyset$.

Moreover, the equation $e_1 = e_2$ is equivalent to

$$x_\alpha - x_{\alpha'} = x_{\beta'} - x_\beta,$$

which is ruled out by our choice of α, α', and so $e_1 \neq e_2$.

Similarly, $e_3 \neq e_4$.

Thus e_1, e_2, e_3, e_4 are all distinct. ∎

Remark. All the results above hold practically verbatim with $\mathbb{R}$ replaced by $\mathbb{R}^k$, for any fixed integer $k \geq 1$. In this more geometrical context, e_1, e_2, e_3, e_4 are vectors in k-dimensional Euclidean space, and the equation $e_1 + e_2 = e_3 + e_4$ says that e_1, e_2, e_3, e_4 are the vertices of a parallelogram (whose area may be zero).

10.4.3 More is Known!

To state the generalization of this theorem we need a definition.

Def 10.4.11 An equation $E(e_1, \ldots, e_n)$ (e.g., $e_1 + e_2 = e_3 + e_4$) is *regular* if the following holds: *for all colorings* $\chi : \mathbb{R} \to \mathbb{N}$ *there exists* $\vec{e} = (e_1, \ldots, e_n)$ *such that*

$$\chi(e_1) = \cdots = \chi(e_n),$$
$$E(e_1, \ldots, e_n),$$

and $e_1, \ldots, e_n$ *are all distinct.*

If we combine Theorems 10.4.7 and 10.4.10 we obtain the following.

Theorem 10.4.12 $e_1 + e_2 = e_3 + e_4$ *is regular iff* $2^{\aleph_0} > \aleph_1$.

Jacob Fox [8] has generalized this to prove the following.

Theorem 10.4.13 *Let* $s \in \mathbb{N}$. *The equation*

$$e_1 + se_2 = e_3 + \cdots + e_{s+3} \tag{10.1}$$

is regular iff $2^{\aleph_0} > \aleph_s$.

Fox's result also holds in higher dimensional Euclidean space, where it relates to the vertices of $(s + 1)$-dimensional parallelepipeds. Subtracting $(s + 1)e_2$ from both sides of (10.1) and rearranging, we get

$$e_1 - e_2 = (e_3 - e_2) + \cdots + (e_{s+3} - e_2),$$

which says that e_1 and e_2 are opposite corners of some $(s + 1)$-dimensional parallelepiped P where $e_3, \ldots, e_{s+3}$ are the corners of P adjacent to e_2. Of course, there are other vertices of P besides these, and Fox's proof actually shows that if $2^{\aleph_0} > \aleph_s$ then *all* the 2^{s+1} vertices of some such P must have the same color.

Chapter 11

Applications of Rado's Theorem

The title of this chapter is a cheat. We will not be applying Rado's Theorem. We give one application of Schur's theorem (Theorem **??**) and an application of the Lemma used to prove Rado's Theorem- Lemma 10.2.7

BILL- the theorem used to prove RADO- the thing with a,a+d,.., a+(k-1)d AND d itself bing the same color, we apply to Number Theory- QR tuff.

Chapter 12

Advanced Topics*

12.1 Every Set of Positive Upper Density has a 3-AP

12.1.1 Combinatorial Proof

Notation 12.1.1 Let $[n] = \{1, \ldots, n\}$. If $k \in \mathbb{N}$ then k-AP means an arithmetic sequence of size k.

Consider the following statement:

If $A \subseteq [n]$ and $|A|$ is 'big' then A must have a 3-AP.

This statement, made rigorous, is true. In particular, the following is true and easy:

Let $n \geq 3$. If $A \subseteq [n]$ and $|A| \geq 0.7n$ then A must have a 3-AP.

Can we lower the constant 0.7? We can lower it as far as we like if we allow n to start later:

Roth [13, 25, 26] proved the following using analytic means.

$(\forall \lambda > 0)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(\forall A \subseteq [n])[|A| \geq \lambda n \implies A \text{ has a 3-AP}]$.

The analogous theorem for 4-APs was later proven by Szemeredi [13, 31] by a combinatorial proof. Szemeredi [32] later (with a much harder proof) generalized from 4 to any k.

We prove the $k = 3$ case using the combinatorial techniques of Szemeredi. Our proof is essentially the same as in the book *Ramsey Theory* by Graham, Rothchild, and Spencer [13].

More is known. A summary of what else is known will be presented in the next section.

Def 12.1.2 Let $sz(n)$ be the least number such that, for all $A \subseteq [n]$, if $|A| \ge sz(n)$ then A has a 3-AP. Note that if $A \subseteq [a, a+n-1]$ and $|A| \ge sz(n)$ then A has a 3-AP. Note also that if $A \subseteq \{a, 2a, 3a, \ldots, na\}$ and $|A| \ge sz(n)$ then A has a 3-AP. More generally, if A is a subset of any equally spaced set of size n, and $|A| \ge sz(n)$, then A has a 3-AP.

We will need the following Definition and Lemma.

Def 12.1.3 Let $k, e, d_1, \ldots, d_k \in \mathbb{N}$. The *cube on* $(e, d_1, \ldots, d_k)$, denoted $C(e, d_1, \ldots, d_k)$, is the set $\{e + b_1d_1 + \cdots + b_kd_k \mid b_1, \ldots, b_k \in \{0, 1\}\}$. A k-*cube* is a cube with k d's.

Lemma 12.1.4 *Let I be an interval of $[1, n]$ of length L. If $|B| \subseteq I$ then there is a cube $(e, d_1, \ldots, d_k)$ contained in B with $k = \Omega(\log\log|B|)$ and $(\forall i)[d_i \le L]$.*

Proof:

The following procedure produces the desired cube.

1. Let $B_1 = B$ and $\beta_1 = |B_1|$.

2. Let D_1 be all $\binom{\beta_1}{2}$ positive differences of elements of B_1. Since $B_1 \subseteq [n]$ all of the differences are in $[n]$. Hence some difference must occur $\binom{\beta_1}{2}/n \sim \beta_1^2/2n$ times. Let that difference be d_1. Note that $d_1 \le L$.

3. Let $B_2 = \{x \in B_1 : x + d_1 \in B_1\}$. Note that $|B_2| \ge \beta_1^2/2n$. Let $|B_2| = \beta_2$. Note the trivial fact that

   $x \in B_1 \implies x + d_1 \in B$.

4. Let D_2 be all $\binom{\beta_2}{2}$ positive differences of elements of B_2. Since $B_2 \subseteq [n]$ all of the differences are in $[n]$. Hence some difference must occur $\binom{\beta_1}{2}/n \sim \beta_2^2/2n$ times. Let that difference be d_2. Note that $d_2 \le L$.

5. Let $B_3 = \{x \in B_2 : x + d_2 \in B_2\}$. Note that $|B_3| \ge \beta_2^2/2n$. Let $|B_3| = \beta_3$. Note that

   $x \in B_3 \implies x + d_2 \in B$

   $x \in B_3 \implies x \in B_2 \implies x + d_1 \in B$

   $x \in B_3 \implies x + d_2 \in B_2 \implies x + d_1 + d_2 \in B$

6. Keep repeating this procedure until $B_{k+2} = \emptyset$. (We leave the deatils of the definition to the reader.) Note that if $i \le k+1$ then

   $x \in B_i \implies x + b_1 d_1 + \cdots + b_{i-1} d_{i-1} \in B$ for any $b_1, \ldots, b_{i-1} \in \{0, 1\}$.

7. Let e be any element of B_{k+1}. Note that we have $e + b_1 d_1 + \cdots + b_k d_k \in B$ for any $b_1, \ldots, b_k \in \{0, 1\}$.

We leave it as an exercise to formally show that $C(e, d_1, \ldots, d_k)$ is contained in B and that $k = \Omega(\log\log |B|)$. ∎

The next lemma states that if A is 'big' and 3-free then it is somewhat uniform. There cannot be sparse intervals of A. The intuition is that if A has a sparse interval then the rest of A has to be dense to make up for it, and it might have to be so dense that it has a 3-AP.

Lemma 12.1.5 *Let $n, n_0 \in \mathbb{N}; \lambda, \lambda_0 \in (0,1)$. Assume $\lambda < \lambda_0$ and $(\forall m \ge n_0)[sz(m) \le \lambda_0 m]$. Let $A \subseteq [n]$ be a 3-free set such that $|A| \ge \lambda n$.*

1. *Let a, b be such that $a < b$, $a > n_0$, and $n - b > n_0$. Then $\lambda_0(b-a) - n(\lambda_0 - \lambda) \le |A \cap [a,b]|$.*

2. *Let a be such that $n - a > n_0$. Then $\lambda_0 a - n(\lambda_0 - \lambda) \le |A \cap [1,a]|$.*

Proof:

1) Since A is 3-free and $a \ge n_0$ and $n - b \ge n_0$ we have $|A \cap [1, a-1]| < \lambda_0(a-1) < \lambda_0 a$ and $|A \cap [b+1, n]| < \lambda_0(n-b)$. Hence

$$\begin{aligned}
\lambda n \le |A| = &\ |A \cap [1, a-1]| + |A \cap [a,b]| + |A \cap [b+1, n]| \\
\lambda n \le &\ \lambda_0 a + |A \cap [a,b]| + \lambda_0(n-b) \\
\lambda n - \lambda_0 n + \lambda_0 b - \lambda_0 a \le &\ |A \cap [a,b]| \\
\lambda_0(b-a) - n(\lambda_0 - \lambda) \le &\ |A \cap [a,b]|.
\end{aligned}$$

2) Since A is 3-free and $n-a > n_0$ we have $|A\cap[a+1,n]| \le \lambda_0(n-a)$. Hence

$$\begin{aligned}
\lambda n \le |A| = &\ |A\cap[1,a]| + |A\cap[a+1,n]| \\
\lambda n \le &\ |A\cap[1,a]| + \lambda_0(n-a) \\
\lambda n - \lambda_0 n + \lambda_0 a \le &\ |A\cap[1,a]| \\
\lambda_0 a - (\lambda_0-\lambda)n \le &\ |A\cap[1,a]|.
\end{aligned}$$

∎

Lemma 12.1.6 *Let $n,n_0\in\mathbb{N}$ and $\lambda,\lambda_0\in(0,1)$. Assume that $\lambda<\lambda_0$ and that $(\forall m\ge n_0)[sz(m)\le\lambda_0 m]$. Assume that $\frac{n}{2}\ge n_0$. Let $a,L\in\mathbb{N}$ such that $a\le\frac{n}{2}$, $L<\frac{n}{2}-a$, and $a\ge n_0$. Let $A\subseteq[n]$ be a 3-free set such that $|A|\ge\lambda n$.*

1. *There is an interval $I\subseteq[a,\frac{n}{2}]$ of length $\le L$ such that*

$$|A\cap I| \ge \left\lfloor \frac{2L}{n-2a}(\lambda_0(\frac{n}{2}-a)-n(\lambda_0-\lambda))\right\rfloor.$$

2. *Let α be such that $0<\alpha<\frac{1}{2}$. If $a=\alpha n$ and $\sqrt{n} << \frac{n}{2}-\alpha n$ then there is an interval $I\subseteq[a,\frac{n}{2}]$ of length $\le O(\sqrt{n})$ such that*

$$|A\cap I| \ge \left\lfloor \frac{2\sqrt{n}}{(1-2\alpha)}(\lambda_0(\frac{1}{2}-(\lambda_0-\lambda)-\alpha))\right\rfloor = \Omega(\sqrt{n}).$$

Proof: By Lemma 12.1.5 with $b=\frac{n}{2}$, $|A\cap[a,\frac{n}{2}]|\ge\lambda_0(\frac{n}{2}-a-n(\lambda_0-\lambda)$. Divide $[a,\frac{n}{2}]$ into $\lceil\frac{n-2a}{2L}\rceil$ intervals of size $\le L$. There must exist an interval I such that

$$|A\cap I| \ge \left\lfloor \frac{2L}{n-2a}(\lambda_0(\frac{n}{2}-a)-n(\lambda_0-\lambda))\right\rfloor.$$

If $L=\lceil\sqrt{n}\rceil$ and $a=\alpha n$ then

$$\begin{aligned}
|A\cap I| &\ge \left\lfloor \tfrac{2L}{n-2a}(\lambda_0(\tfrac{n}{2}-a)-n(\lambda_0-\lambda))\right\rfloor \\
&\ge \left\lfloor \tfrac{2\sqrt{n}}{n(1-2\alpha)}(\lambda_0(\tfrac{n}{2}-\alpha n)-n(\lambda_0-\lambda)))\right\rfloor \\
&\ge \left\lfloor \tfrac{2\sqrt{n}}{(1-2\alpha)}(\lambda_0(\tfrac{1}{2}-\alpha)-(\lambda_0-\lambda))\right\rfloor = \Omega(\sqrt{n}).
\end{aligned}$$

∎

Theorem 12.1.7 *For all λ, $0 < \lambda < 1$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $sz(n) \leq \lambda n$.*

Proof:

Let $S(\lambda)$ be the statement

there exists n_0 such that, for all $n \geq n_0$, $sz(n) \leq \lambda n$.

It is a trivial exercise to show that $S(0.7)$ is true.

Let

$$C = \{\lambda \mid S(\lambda)\}.$$

C is closed upwards. Since $0.7 \in C$ we know $C \neq \emptyset$. Assume, by way of contradiction, that $C \neq (0,1)$. Then there exists $\lambda < \lambda_0$ such that $\lambda \notin C$ and $\lambda_0 \in C$. We can take $\lambda_0 - \lambda$ to be as small as we like. Let n_0 be such that $S(\lambda_0)$ is true via n_0. Let $n \geq n_0$ and let $A \subseteq [n]$ such that $|A| \geq \lambda n$ but A is 3-free. At the end we will fix values for the parameters that (a) allow the proof to go through, and (b) imply $|A| < \lambda n$, a contradiction.

PLAN : We will obtain a $T \subseteq \overline{A}$ that will help us. We will soon see what properties T needs to help us. Consider the bit string in $\{0,1\}^n$ that represents $T \subseteq [n]$. Say its first 30 bits looks like this:

$$T(0)T(1)T(2)T(3)\cdots T(29) = 000111111100001110010111100000$$

The set A lives in the blocks of 0's of T (henceforth 0-blocks). We will bound $|A|$ by looking at A on the 'small' and on the 'large' 0-blocks of T. Assume there are t 1-blocks. Then there are $t+1$ 0-blocks. We call a 0-block *small* if it has $< n_0$ elements, and *big* otherwise. Assume there are t^{small} small 0-blocks and t^{big} big 0-blocks. Note that $t^{\text{small}} + t^{\text{big}} = t+1$ so $t^{\text{small}}, t^{\text{big}} \leq t+1$. Let the small 0-blocks be $B_1^{\text{small}}, \ldots, B_{t^{\text{small}}}^{\text{small}}$, let their union be B^{small}, let the big 0-blocks be $B_1^{\text{big}}, \ldots, B_{t^{\text{big}}}^{\text{big}}$, and let their union be B^{big}. It is easy to see that

$$|A \cap B^{\text{small}}| \leq t^{\text{small}} n_0 \leq (t+1)n_0.$$

Since each B_i^{big} is bigger than n_0 we must have, for all i, $|A \cap B_i^{\text{big}}| < \lambda_0 |B_i^{\text{big}}|$ (else $A \cap B_i^{\text{big}}$ has a 3-AP and hence A does). It is easy to see that

$$|A \cap B^{\text{big}}| = \sum_{i=1}^{t^{\text{big}}} |A \cap B_i^{\text{big}}| \leq \sum_{i=1}^{t^{\text{big}}} \lambda_0 |B_i^{\text{big}}| \leq \lambda_0 \sum_{i=1}^{t^{\text{big}}} |B_i^{\text{big}}| \leq \lambda_0(n - |T|).$$

Since A can only live in the (big and small) 0-blocks of T we have

$$|A| = |A \cap B^{\text{small}}| + |A \cap B^{\text{big}}| \le (t+1)n_0 + \lambda_0(n - |T|).$$

In order to use this inequality to bound $|A|$ we will need T to be big and t to be small, so we want T to be a big set that has few blocks.

If only it was that simple. Actually we can now reveal the

REAL PLAN: The real plan is similar to the easy version given above. We obtain a set $T \subseteq \overline{A}$ and *a parameter d*. A *1-block* is a maximal AP with difference d that is contained in T (that is, if $FIRST$ and $LAST$ are the first and last elements of the 1-block then $FIRST - d \notin T$ and $LAST + d \notin T$). A *0-block* is a maximal AP with difference d that is contained in $\overline{T}$. Partition T into 1-blocks. Assume there are t of them.

Let $[n]$ be partitioned into $N^0 \cup \cdots \cup N^{d-1}$ where $N_j = \{x \mid x \le n \wedge x \equiv j \pmod{d}\}$.

Fix j, $0 \le j \le d-1$. Consider the bit string in $\{0,1\}^{\lfloor n/d \rfloor}$ that represents $T \cap N_j$ Say the first 30 bits of $T \cap N_j$ look like

$$T(j)T(d+j)T(2d+j)T(3d+j)\cdots T(29d+j) = 000111111100001110010111111100$$

During PLAN we had an intuitive notion of what a 0-block or 1-block was. Note that if we restrict to N_j then that intuitive notion is still valid. For example the first block of 1's in the above example represents $T(3d+j)$, $T(4d+j)$, $T(5d+j)$, $\ldots$, $T(9d+j)$ which is a 1-block as defined formally.

Each 1-block is contained in a particular N_j. Let t_j be the number of 1-blocks that are contained in N_j. Note that $\sum_{j=0}^{d-1} t_j = t$. The number of 0-blocks that are in N_j is at most $t_j + 1$.

Let j be such that $0 \le j \le d-1$. By reasoning similar to that in the above PLAN we obtain

$$|A \cap N_j| \le (t_j + 1)n_0 + \lambda_0(N_j - |T|).$$

We sum both sides over all $j = 0$ to $d-1$ to obtain

$$|A| \le (t+d)n_0 + \lambda_0(n - |T|)$$

In order to use this inequality to bound $|A|$ we need T to be big and t, d to be small. Hence we want a big set T which when looked at mod d, for some small d, decomposes into a small number of blocks.

What is a 1-block within N_j? For example, lets look at $d = 3$ and the bits sequence for T is

$$\begin{array}{ccccccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17; \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0. \end{array}$$

Note that T looked at on $N_2 \cup T$ has bit sequence

$$\begin{array}{cccccc} 2 & 5 & 8 & 11 & 14 & 17; \\ 0 & 1 & 1 & 1 & 1 & 0. \end{array}$$

The numbers $5, 8, 11, 14$ are all in T and form a 1-block in the N_2 part. Note that they also from an arithmetic progession with spacing $d = 3$. Also note that this is a maximal arithmetic sequence with spacing $d = 3$ since $0 \notin T$ and $17 \notin T$. More generally *1-blocks of T within N_j are maximal arithmetic progessions with spacing d.* With that in mind we can restate the kind of set T that we want.

We want a set $T \subseteq \overline{A}$ and a parameter d such that

1. T is big (so that $\lambda_0(n - |T|)$ is small),
2. d is small (see next item), and
3. the number of maximal arithmetic sequences of length d within T, which is the parameter t above, is small (so that $(t + d)n_0$ is small).

How do we obtain a big subset of $\overline{A}$? We will obtain many pairs $x, y \in A$ such that $2y - x \leq n$. Note that since $x, y, 2y - x$ is a 3-AP and $x, y \in A$ we must have $2y - x \in \overline{A}$.

Let α, $0 < \alpha < \frac{1}{2}$, be a parmater to be determined later. (For those keeping track, the parameters to be determined later are now λ_0, λ, n, and α. The parameter n_0 depends on λ_0 so is not included in this list.)

We want to apply Lemma 12.1.6.2.b to $n, n_0, a = \alpha n$. Hence we need the following conditions.

$$\begin{aligned} \alpha n &\geq n_0 \\ \tfrac{n}{2} &\geq n_0 \\ \tfrac{n}{2} - \alpha n &\geq \sqrt{n} \end{aligned}$$

Assuming these conditions hold, we proceed. By Lemma 12.1.6.b there is an interval $I \subseteq [\alpha n, \frac{n}{2}]$ of length $O(\sqrt{n})$ such that

$$|A \cap I| \geq \left\lfloor \frac{2\sqrt{n}}{(1 - 2\alpha)}(\lambda_0(\frac{1}{2} - \alpha) - (\lambda_0 - \lambda)) \right\rfloor = \Omega(\sqrt{n}).$$

By Lemma 12.1.4 there is a cube $C(e, d_1, \ldots, d_k)$ contained in $|A \cap I|$ with $k = \Omega(\log\log|A\cap I|) = \Omega(\log\log\sqrt{n}) = \Omega(\log\log n)$ and $d \geq \sqrt{n}$.

For i such that $1 \leq i \leq k$ we define the following.

1. Define $C_0 = \{e\}$ and, for $1 \leq i \leq k$, define $C_i = C(e, d_1, \ldots, d_i)$.
2. T_i is the third terms of AP's with the first term in $A \cap [1, e-1]$ and the second term in C_i. Formally $T_i = \{2m - x \mid x \in A \cap [1, e-1] \wedge m \in C_i\}$.

Note that, for all i, $T_i \cap A = \emptyset$. Hence we look for a large T_i that can be decomposed into a small number of blocks. We will end up using $d = 2d_{i+1}$.

Note that $T_0 \subseteq T_1 \subseteq T_2 \subseteq \cdots \subseteq T_k$. Hence to obtain a large T_i it suffices to show that T_0 is large and then any of the T_i will be large (though not neccesarily consist of a small number of blocks).

Since $C_0 = \{e\}$ we have

$T_0 = \{2m - x \mid x \in A \cap [1, e-1] \wedge m \in C_0\} = \{2e - x \mid x \in A \cap [1, e-1]\}$.

Clearly there is a bijection from $A \cap [1, e-1]$ to T_0, hence $|T_0| = |A \cap [1, e-1]|$. Since $e \in [\alpha n, \frac{n}{2}]$ we have $|A \cap [1, e]| \geq |A \cap [1, \alpha n]|$.

We want to use Lemma 12.1.5.2 on $A \cap [1, \alpha n]$. Hence we need the condition

$$n - \alpha n \geq n_0.$$

By Lemma 12.1.5

$$|T_0| \geq |A \cap [1, \alpha n]| \geq \lambda_0 \alpha n - n(\lambda_0 - \lambda) = n(\lambda_0 \alpha - (\lambda_0 - \lambda)).$$

In order for this to be useful we need the following condition

$$\begin{aligned} \lambda - \lambda_0 + \lambda_0 \alpha &> 0 \\ \lambda_0 \alpha &> \lambda_0 - \lambda \end{aligned}$$

We now show that some T_i has a small number of blocks. Since $|T_k| \leq n$ (a rather generous estimate) there must exist an i such that $|T_{i+1} - T_i| \leq \frac{n}{k}$. Let $t = \frac{n}{k}$ (t will end up bounding the number of 1-blocks).

Partition T_i into maximal AP's with difference $2d_{i+1}$. We call these maximal AP's 1-blocks. We will show that there are $\leq t$ 1-blocks by showing a bijection between the blocks and $T_{i+1} - T_i$.

If $z \in T_i$ then $z = 2m - x$ where $x \in A \cap [1, \alpha n - 1]$ and $m \in C_i$. By the definitions of C_i and C_{i+1} we know $m + d_{i+1} \in C_{i+1}$. Hence $2(m + d_{i+1}) - x \in T_{i+1}$. Note that $2(m + d_{i+1}) - x = z + 2d_{i+1}$. In short we have

$$z \in T_i \implies z + 2d_{i+1} \in T_{i+1}.$$

NEED PICTURE

We can now state the bijection. Let $z_1, \ldots, z_m$ be a block in T_i. We know that $z_m + 2d_{i+1} \notin T_i$ since if it was the block would have been extended to include it. However, since $z_m \in T_i$ we know $z_m + 2d_{i+1} \in T_{i+1}$. Hence $z_m + 2d_{i+1} \in T_{i+1} - T_i$. This is the bijection: map a block to what would be the next element if it was extended. This is clearly a bijection. Hence the number of 1-blocks is at most $t = |T_{i+1} - T_i| \leq n/k$.

To recap, we have

$$|A| \leq (t + d)n_0 + \lambda_0(n - |T|)$$

with $t \leq \frac{n}{k} = O(\frac{n}{\log\log n})$, $d = O(\sqrt{n})$, and $|T| \geq n(\lambda_0\alpha - (\lambda_0 - \lambda))$. Hence we have

$$|A| \leq O((\frac{n}{\log\log n} + \sqrt{n})n_0) + n\lambda_0(1 - \lambda + \lambda_0 - \lambda_0\alpha).$$

We want this to be $< \lambda n$. The term $O((\frac{n}{\log\log n} + \sqrt{n})n_0)$ can be ignored since for n large enough this is less than any fraction of n. For the second term we need

$$\lambda_0(1 - \lambda + \lambda_0 - \lambda_0\alpha) < \lambda$$

We now gather together all of the conditions and see how to satisfy them all at the same time.

$$\begin{aligned}
\alpha n &\geq n_0\\
\tfrac{n}{2} &\geq n_0\\
\tfrac{n}{2} - \alpha n &\geq \sqrt{n}\\
n - \alpha n &\geq n_0\\
\lambda_0\alpha &> \lambda_0 - \lambda\\
\lambda_0(1 - \lambda + \lambda_0 - \lambda_0\alpha) &< \lambda
\end{aligned}$$

We first choose λ and λ_0 such that $\lambda_0 - \lambda < 10^{-1}\lambda_0^2$. This is possible by first picking an initial (λ', λ_0') pair and then picking (λ, λ_0) such that $\lambda' < \lambda < \lambda_0 < \lambda_0'$ and $\lambda_0 - \lambda < 10^{-1}(\lambda')^2 < 10^{-1}\lambda_0^2$. The choice of λ_0 determines n_0. We then chose $\alpha = 10^{-1}$. The last two conditions are satisfied:

$\lambda_0\alpha > \lambda_0 - \lambda$ becomes

$$\begin{aligned} 10^{-1}\lambda_0 &> 10^{-1}\lambda_0^2 \\ 1 &> \lambda_0 \end{aligned}$$

which is clearly true.

$\lambda_0(1 - \lambda + \lambda_0 - \lambda_0\alpha) < \lambda$ becomes

$$\begin{aligned} \lambda_0(1 - 10^{-1}\lambda_0^2 - 10^{-1}\lambda_0) &< \lambda \\ \lambda_0 - 10^{-1}\lambda_0^3 - 10^{-1}\lambda_0^2 &< \lambda \\ \lambda_0 - \lambda - 10^{-1}\lambda_0^3 - 10^{-1}\lambda_0^2 &< 0 \\ 10^{-1}\lambda_0^2 - 10^{-1}\lambda_0^3 - 10^{-1}\lambda_0^2 &< 0 \\ -10^{-1}\lambda_0^3 &< 0 \end{aligned}$$

which is clearly true.

Once λ, λ_0, n_0 are picked, you can easily pick n large enough to make the other inqualities hold. ∎

12.1.2 Analytic Proof

Notation 12.1.8 Let $[n] = \{1, \ldots, n\}$. If $k \in \mathbb{N}$ then k-AP means an arithmetic sequence of size k.

Consider the following statement:

If $A \subseteq [n]$ and $\#(A)$ is 'big' then A must have a 3-AP.

This statement, made rigorous, is true. In particular, the following is true and easy:

Let $n \geq 3$. If $A \subseteq [n]$ and $\#(A) \geq 0.7n$ then A must have a 3-AP.

Can we lower the constant 0.7? We can lower it as far as we like if we allow n to start later:

Roth [13, 25, 26] proved the following using analytic means.

$(\forall \lambda > 0)(\exists n_0 \in \mathbb{N})(\forall n \geq n_0)(\forall A \subseteq [n])[\#(A) \geq \lambda n \implies A \text{ has a 3-AP}]$.

The analogous theorem for 4-APs was later proven by Szemeredi [13, 31] by a combinatorial proof. Szemeredi [32] later (with a much harder proof) generalized from 4 to any k.

We prove the $k = 3$ case using the analytic techniques of Roth; however, we rely heavily on Gowers [12, 11]

Def 12.1.9 Let $sz(n)$ be the least number such that, for all $A \subseteq [n]$, if $\#(A) \geq sz(n)$ then A has a 3-AP. Note that if $A \subseteq [a, a+n-1]$ and $\#(A) \geq sz(n)$ then A has a 3-AP. Note also that if $A \subseteq \{a, 2a, 3a, \ldots, na\}$ and $\#(A) \geq sz(n)$ then A has a 3-AP. More generally, if A is a subset of any equally spaced set of size n, and $\#(A) \geq sz(n)$, then A has a 3-AP.

Throughout this section the following hold.

1. $n \in \mathbb{N}$ is a fixed large prime.
2. $\mathbb{Z}_n = \{1, \ldots, n\}$ with modular arithmetic.
3. $\omega = e^{2\pi i/n}$.
4. If a is a complex number then $|a|$ is its length.
5. If A is a set then $|A|$ is its cardinality.

Counting 3-AP's

Lemma 12.1.10 *Let $A, B, C \subseteq [n]$. The number of $(x, y, z) \in A \times B \times C$ such that $x + z \equiv 2y \pmod{n}$ is*

$$\frac{1}{n} \sum_{x,y,z \in [n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)}.$$

Proof:

We break the sum into two parts:
Part 1:

$$\frac{1}{n} \sum_{x,y,z \in [n], x+z \equiv 2y \pmod{n}} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)}.$$

Note that we can replace $\omega^{-r(x-2y+z)}$ with $\omega^0 = 1$. We can then replace $\sum_{r=1}^{n} 1$ with n. Hence we have

$$\frac{1}{n} \sum_{x,y,z \in [n], x+z \equiv 2y \pmod{n}} A(x)B(y)C(z)n = \sum_{x,y,z \in [n], x+z \equiv 2y \pmod{n}} A(x)B(y)C(z)$$

This is the number of $(x,y,z) \in A \times B \times C$ such that $x+z \equiv 2y \pmod n$.
Part 2:

$$\frac{1}{n} \sum_{x,y,z\in[n], x+z\not\equiv 2y \pmod n} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)}.$$

We break this sum up depending on what the (nonzero) value of $w = x+z-2y \pmod n$. Let

$$S_u = \sum_{x,y,z\in[n], x-2y+z=2} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-ru}.$$

Since $u \neq 0$, $\sum_{r=1}^{n} \omega^{-ru} = \sum_{r=1}^{n} \omega^{-r} = 0$. Hence $S_u = 0$.
Note that

$$\frac{1}{n} \sum_{x,y,z\in[n], x+z\not\equiv 2y \pmod n} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)} = \frac{1}{n} \sum_{u=1}^{n-1} S_u = 0$$

The lemma follows from Part 1 and Part 2. ▮

Lemma 12.1.11 *Let $A \subseteq [n]$. Let $B = C = A \cap [n/3, 2n/3]$. The number of $(x,y,z) \in A \times B \times C$ such that x,y,z forms a 3-AP is at least*

$$\frac{1}{2n} \sum_{x,y,z\in[n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)} - O(n).$$

Proof: By Lemma 12.1.10

$$\frac{1}{n} \sum_{x,y,z\in[n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)}$$

is the number of $(x,y,z) \in A \times B \times C$ such that $x+z \equiv 2y \pmod n$. This counts three types of triples:

- Those that have $x=y=z$. There are $n/3$ of them.
- Those that have $x+z = 2y+n$. There are $O(1)$ of them.

- Those that have $x \neq y$, $y \neq z$, $x \neq z$, and $x+z=2y$.

Hence

$$\#(\{(x,y,z) : (x+z=2y) \wedge x \neq y \wedge y \neq z \wedge x \neq z\}) = \frac{1}{n} \sum_{x,y,z \in [n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)} - O(n).$$

We are not done yet. Note that $(5,10,15)$ may show up as $(15,10,5)$. Every triple appears at most twice. Hence
$\#(\{(x,y,z) : (x+z=2y) \wedge x \neq y \wedge y \neq z \wedge x \neq z\})$
$\leq \quad 2\#(\{(x,y,z) : (x<y<z) \wedge (x+z=2y) \wedge x \neq y \wedge y \neq z \wedge x \neq z\})$.
Therefore

$$\frac{1}{2n} \sum_{x,y,z \in [n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)} - O(n) \leq \text{ the number of 3-AP's with } x \in A,\ y \in B,\ z \in C\ .$$

▮

We will need to re-express this sum. For that we will use Fourier Analysis.

Fourier Analysis

Def 12.1.12 If $f : \mathbb{Z}_n \rightarrow \mathbb{N}$ then $\hat{f} : \mathbb{Z}_n \rightarrow \mathbb{C}$ is

$$\hat{f}(r) = \sum_{s \in [n]} f(s)\omega^{-rs}.$$

$\hat{f}$ is called the *Fourier Transform* of f.

What does $\hat{f}$ tell us? We look at the case where f is the characteristic function of a set $A \subseteq [n]$. Henceforth we will use $A(x)$ instead of $f(x)$.

We will need the followng facts.

Lemma 12.1.13 *Let $A \subseteq \{1, \ldots, n\}$.*

1. $\hat{A}(n) = \#(A)$.
2. $\max_{r \in [n]} |\hat{A}(r)| = \#(A)$.

3. $A(s) = \frac{1}{n}\sum_{r=1}^{n} \hat{A}(r)\omega^{-rs}$. *DO WE NEED THIS?*
4. $\sum_{r=1}^{n} |\hat{A}(r)|^2 = n\#(A)$.
5. $\sum_{s=1}^{n} A(s) = \frac{1}{n}\sum_{r=1}^{n} \hat{A}(r)$.

Proof:

Note that $\omega^n = 1$. Hence

$$\hat{A}(n) = \sum_{s\in[n]} A(s)\omega^{-ns} = \sum_{s\in[n]} A(s) = \#(A).$$

Also note that

$$|\hat{A}(r)| = |\sum_{s\in[n]} A(s)\omega^{-rs}| \le \sum_{s\in[n]} |A(s)\omega^{-rs}| \le \sum_{s\in[n]} |A(s)||\omega^{-rs}| \le \sum_{s\in[n]} |A(s)| = \#(A).$$

∎

Informal Claim: If $\hat{A}(r)$ is large then there is an arithmetic sequence P with difference $r^{-1} \pmod n$ such that $\#(A \cap P)$ is large.

We need a lemma before we can proof the claim.

Lemma 12.1.14 *Let $n, m \in \mathbb{N}$, $s_1, \ldots, s_m$, and $0 < \lambda, \alpha, \epsilon < 1$ be given (no order on $\lambda, \alpha, \epsilon$ is implied). Assume that $(\lambda - \frac{m-1}{m}(\lambda + \epsilon)) \ge 0$. Let $f(x_1, \ldots, x_m) = |\sum_{j=1}^{m} x_j\omega^{s_j}|$. The maximum value that $f(x_1, \ldots, x_m)$ can achieve subject to the following two constraints (1) $\sum_{j=1}^{m} x_j \ge \lambda n$, and (2) $(\forall j)[0 \le x_i \le (\lambda + \epsilon)\frac{n}{m}]$ is bounded above by $\epsilon mn + (\lambda + \epsilon)\frac{n}{m}|\sum_{j=1}^{m} \omega^{s_j}|$*

Proof:

Assume that the maximum value of f, subject to the constraints, is achieved at $(x_1, \ldots, x_m)$. Let MIN be the minimum value that any variable x_i takes on (there may be several variables that take this value). What is the smallest that MIN could be? By the contraints this would occur when all but one of the variables is $(\lambda + \epsilon)\frac{n}{m}$ and the remaining variable has value MIN. Since $\sum_{x_i} \ge \lambda n$ we have

$MIN + (m-1)(\lambda + \epsilon)\frac{n}{m} \ge \lambda n$

$MIN + \frac{m-1}{m}(\lambda + \epsilon)n \ge \lambda n$

$MIN \ge \lambda n - \frac{m-1}{m}(\lambda + \epsilon)n$

$MIN \ge (\lambda - \frac{m-1}{m}(\lambda + \epsilon))n$

Hence note that, for all j,
$x_j - MIN \leq x_j - (\lambda - \frac{m-1}{m}(\lambda+\epsilon))n$
Using the bound on x_j from constraint (2) we obtain

$$\begin{aligned}
x_j - MIN &\leq (\lambda+\epsilon)\tfrac{n}{m} - (\lambda - \tfrac{m-1}{m}(\lambda+\epsilon))n \\
&\leq ((\lambda+\epsilon)\tfrac{1}{m} - (\lambda - \tfrac{m-1}{m}(\lambda+\epsilon)))n \\
&\leq ((\lambda+\epsilon)\tfrac{1}{m} - \lambda + \tfrac{m-1}{m}(\lambda+\epsilon))n \\
&\leq \epsilon n
\end{aligned}$$

Note that

$$\begin{aligned}
|\textstyle\sum_{j=1}^m x_j\omega^{s_j}| = &\ |\textstyle\sum_{j=1}^m (x_j - MIN)\omega^{s_j} + \sum_{j=1}^m MIN\omega^{s_j}| \\
\leq &\ |\textstyle\sum_{j=1}^m (x_j - MIN)\omega^{s_j}| + |\sum_{j=1}^m MIN\omega^{s_j}| \\
\leq &\ \textstyle\sum_{j=1}^m |(x_j - MIN)||\omega^{s_j}| + MIN|\sum_{j=1}^m \omega^{s_j}| \\
\leq &\ \textstyle\sum_{j=1}^m \epsilon n + MIN|\sum_{j=1}^m \omega^{s_j}| \\
\leq &\ \epsilon mn + MIN|\textstyle\sum_{j=1}^m \omega^{s_j}| \\
\leq &\ \epsilon mn + (\lambda+\epsilon)\tfrac{n}{m}|\textstyle\sum_{j=1}^m \omega^{s_j}|
\end{aligned}$$

▮

Lemma 12.1.15 *Let $A \subseteq [n]$, $r \in [n]$, and $0 < \alpha < 1$. If $|\hat{A}(r)| \geq \alpha n$ and $|A| \geq \lambda n$ then there exists $m \in \mathbb{N}$, $0 < \epsilon < 1$, and an arithmetic sequence P within $\mathbb{Z}_n$, of length $\frac{n}{m} \pm O(1)$ such that $\#(A \cap P) \geq (\lambda+\epsilon)\frac{n}{m}$. The parameters ϵ and m will depend on λ and α but not n.*

Proof: Let m and ϵ be parameters to be picked later. We will note constraints on them as we go along. (Note that ϵ will not be used for a while.)

Let $1 = a_1 < a_2 < \cdots < a_{m+1} = n$ be picked so that
$a_2 - a_1 = a_3 - a_2 = \cdots = a_m - a_{m-1}$ and $a_{m+1} - a_m$ is as close to $a_2 - a_1$ as possible.

For $1 \leq j \leq m$ let

$$P_j = \{s \in [n] : a_j \leq rs \pmod{n} < a_{j+1}\}.$$

Let us look at the elements of P_j. Let r^{-1} be the inverse of r mod n.

1. s such that $a_j \equiv rs \pmod{n}$, that is, $s \equiv a_j r^{-1} \pmod{n}$.

2. s such that $a_j + 1 \equiv rs \pmod{n}$, that is $s \equiv (a_j+1)r^{-1} \equiv a_j r^{-1} + r^{-1} \pmod{n}$.

3. s such that $a_j + 2 \equiv rs \pmod{n}$, that is $s \equiv (a_j+2)r^{-1} \equiv a_j r^{-1} + 2r^{-1} \pmod{n}$.

4. $\vdots$

Hence P_j is an arithmetic sequence within $\mathbb{Z}_n$ which has difference r^{-1}. Also note that $P_1, \ldots, P_m$ form a partition of $\mathbb{Z}_n$ into m parts of size $\frac{n}{m} + O(1)$ each.

Recall that

$$\hat{A}(r) = \sum_{s \in [n]} A(s)\omega^{-rs}.$$

Lets look at $s \in P_j$. We have that $a_j \le rs \pmod{n} < a_{j+1}$. Therefore the values of $\{\omega^{rs} : s \in P_j\}$ are all very close together. We will pick $s_j \in P_j$ carefully. In particular we will constrain m so that it is possible to pick $s_j \in P_j$ such that $\sum_{j=1}^{m} \omega^{-rs_j} = 0$. For $s \in P_j$ we will approximate ω^{-rs} by ω^{-rs_j}. We skip the details of how good the approximation is.

We break up the sum over s via P_j.

$$\begin{array}{rl}
\hat{A}(r) = & \sum_{s\in[n]} A(s)\omega^{-rs} \\
= & \sum_{j=1}^{m} \sum_{s\in P_j} A(s)\omega^{-rs} \\
\sim & \sum_{j=1}^{m} \sum_{s\in P_j} A(s)\omega^{-rs_j} \\
= & \sum_{j=1}^{m} \omega^{-rs_j} \sum_{s\in P_j} A(s) \\
= & \sum_{j=1}^{m} \omega^{-rs_j} \#(A \cap P_j) \\
= & \sum_{j=1}^{m} \#(A \cap P_j)\omega^{-rs_j} \\
\alpha n \le |\hat{A}(r)| = & |\sum_{j=1}^{m} \#(A \cap P_j)\omega^{-rs_j}|
\end{array}$$

We will not use ϵ. We intend to use Lemma 12.1.14; therefore we have the contraint $(\lambda - \frac{m-1}{m}(\lambda+\epsilon)) \ge 0$.

Assume, by way of contradiction, that $(\forall j)[|A \cap P_j| \le (\lambda+\epsilon)\frac{n}{m}$. Applying Lemma 12.1.14 we obtain

$$|\sum_{j=1}^{m} \#(A \cap P_j)\omega^{-rs_j}| \le \epsilon mn + (\lambda+\epsilon)\frac{n}{m}|\sum_{j=1}^{m} \omega^{-rs_j}| = \epsilon mn.$$

Hence we have

$\alpha n \le \epsilon mn$
$\alpha \le \epsilon m$.
In order to get a contradiction we pick ϵ and m such that $\alpha > \epsilon m$.
Having done that we now have that $(\exists j)[|A \cap P_j| \ge (\lambda+\epsilon)\frac{n}{m}]$.
We now list all of the constraints introduced and say how to satisfy them.

1. m is such that there exists $s_1 \in P_1, \ldots, s_m \in P_m$ such that $\sum_{j=1}^{m} \omega^{-rs_j} = 0$, and
2. $(\lambda - \frac{m-1}{m}(\lambda+\epsilon)) \ge 0$.
3. $\epsilon m < \alpha$.

First pick m to satisfy item 1. Then pick ϵ small enough to satisfy items 2,3. ▮

Lemma 12.1.16 *Let $A, B, C \subseteq [n]$. The number of 3-AP's $(x, y, z) \in A \times B \times C$ is bounded below by*

$$\frac{1}{2n}\sum_{r=1}^{n} \hat{A}(r)\hat{B}(-2r)\hat{C}(r) - O(n).$$

Proof:

The number of 3-AP's is bounded below by

$$\frac{1}{2n}\sum_{x,y,z\in[n]} A(x)B(y)C(z)\sum_{r=1}^{n}\omega^{-r(x-2y+z)} - O(n) =$$

We look at the inner sum.

$$\sum_{x,y,z\in[n]} A(x)B(y)C(z)\sum_{r=1}^{n}\omega^{-r(x-2y+z)} =$$

$$\sum_{r=1}^{n}\sum_{x,y,z\in[n]} A(x)\omega^{-rx}B(y)\omega^{2yr}C(z)\omega^{-rz} =$$

$$\sum_{r=1}^{n}\sum_{x\in[n]} A(x)\omega^{-rx}\sum_{y\in[n]} B(y)\omega^{2yr}\sum_{z\in\mathbb{Z}_r} C(z)\omega^{-rz} =$$

$$\sum_{r=1}^{n} \hat{A}(r)\hat{B}(-2r)\hat{C}(r).$$

The Lemma follows. ∎

Main Theorem

Theorem 12.1.17 *For all λ, $0 < \lambda < 1$, there exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$, $sz(n) \le \lambda n$.*

Proof:

Let $S(\lambda)$ be the statement

there exists n_0 such that, for all $n \ge n_0$, $sz(n) \le \lambda n$.

It is a trivial exercise to show that $S(0.7)$ is true.

Let

$$C = \{\lambda : S(\lambda)\}.$$

C is closed upwards. Since $0.7 \in C$ we know $C \neq \emptyset$. Assume, by way of contradiction, that $C \neq (0,1)$. Then there exists $\lambda < \lambda_0$ such that $\lambda \notin C$ and $\lambda_0 \in C$. We can take $\lambda_0 - \lambda$ to be as small as we like. Let n_0 be such that $S(\lambda_0)$ is true via n_0. Let $n \ge n_0$ and let $A \subseteq [n]$ such that $\#(A) \ge \lambda n$ but A is 3-free.

Let $B = C = A \cap [n/3, 2n/3]$.

By Lemma 12.1.16 the number of 3-AP's of A is bounded below by

$$\frac{1}{2n}\sum_{r=1}^{n} \hat{A}(r)\hat{B}(-2r)\hat{C}(r) - O(n).$$

We will show that either this is positive or there exists a set $P \subseteq [n]$ that is an AP of length XXX and has density larger than λ. Hence P will have a 3-AP.

By Lemma 12.1.13 we have $\hat{A}(n) = \#(A)$, $\hat{B}(n) = \#(B)$, and $\hat{C}(n) = \#(C)$. Hence

$$\frac{1}{2n}\hat{A}(n)\hat{B}(n)\hat{C}(n) + \frac{1}{2n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r) - O(n) =$$

$$\frac{1}{2n}\#(A)\#(B)\#(C) + \frac{1}{2n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r) - O(n).$$

By Lemma 12.1.6 we can take $\#(B), \#(C) \geq n\lambda/4$. We already have $\#(A) \geq \lambda n$. This makes the lead term $\Omega(n^3)$; hence we can omit the $O(n)$ term. More precisely we have that the number of 3-AP's in A is bounded below by

$$\frac{\lambda^3 n^2}{32} + \frac{1}{2n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r)).$$

We are assuming that this quantity is ≤ 0.

$$\frac{\lambda^3 n^2}{32} + \frac{1}{2n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r)) < 0.$$

$$\frac{\lambda^3 n^2}{16} + \frac{1}{n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r)) < 0.$$

$$\frac{\lambda^3 n^2}{16} < -\frac{1}{n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r)).$$

Since the left hand side is positive we have

$$\begin{array}{rl} \frac{\lambda^3 n^2}{16} < & |\frac{1}{n}\sum_{r=1}^{n-1} \hat{A}(r)\hat{B}(-2r)\hat{C}(r)| \\ < & \frac{1}{n}(\max r\hat{A}(r))\sum_{r=1}^{n-1} |\hat{B}(-2r)||\hat{C}(r)| \end{array}$$

By the Cauchy Schwartz inequality we know that

$$\sum_{i=1}^{n-1} |\hat{B}(-2r)||\hat{C}(r)| \leq (\sum_{i=1}^{n-1} |\hat{B}(-2r)|^2)^{1/2})(\sum_{i=1}^{n-1} |\hat{C}(r)|^2)^{1/2}).$$

Hence

$$\frac{\lambda^3 n^2}{16} < |\frac{1}{n}\max_{1\leq r\leq n-1} |\hat{A}(r)|(\sum_{i=1}^{n-1} |\hat{B}(-2r)|^2)^{1/2})(\sum_{i=1}^{n-1} |\hat{C}(r)|^2)^{1/2}).$$

By Parsaval's inequality and the definition of B and C we have

$$\sum_{i=1}^{n-1} |\hat{B}(-2r)|^2)^{1/2} \leq n\#(B) = \frac{\lambda n^2}{3}$$

and

$$\sum_{i=1}^{n-1} |\hat{C}(r)|^2)^{1/2} \leq n\#(C) = \frac{\lambda n^2}{3}$$

Hence

$$\frac{\lambda^3 n^2}{16} < (\max_{1\leq r\leq n-1} |\hat{A}(r)|)\frac{1}{n}\frac{\lambda n^2}{3} = (\max_{1\leq r\leq n-1} |\hat{A}(r)|)\frac{\lambda n}{3}.$$

Therefore
$|\hat{A}(r) \geq \frac{3\lambda^2 n}{16}$. ∎

12.1.3 What more is known?

The following is known.

Theorem 12.1.18 *For every $\lambda > 0$ there exists n_0 such that for all $n \geq n_0$, $sz(n) \leq \lambda n$.*

This has been improved by Heath-Brown [16] and Szemeredi [33]

Theorem 12.1.19 *There exists c such that $sz(n) = \Omega(n\frac{1}{(\log n)^c})$. (Szemeredi estimates $c \leq 1/20$).*

Bourgain [3] improved this further to obtain the following.

Theorem 12.1.20 $sz(n) = \Omega(n\sqrt{\frac{\log\log n}{\log n}})$.

12.2 Ergodic Proofs of Van Der Waerden's Theorem

Van Der Waerden [34] proved the following combinatorial theorem in a combinatorial way

Theorem 12.2.1 *For all $c \in \mathbb{N}$, $k \in \mathbb{N}$, any c-coloring of $\mathbb{Z}$ will have a monochromatic arithmetic sequence of length k.*

Furstenberg [9] later proved it using topological methods. We give a detailed treatment of this proof using as much intuition and as little Topology as needed. We follow the approach of [13] who in turn followed the approach of [10].

12.2.1 Definitions from Topology

Def 12.2.2 X is a *metric space* if there exists a function $d : X \times X \to \mathbb{R}^{\geq 0}$ (called a metric) with the following properties.

1. $d(x, y) = 0$ iff $x = y$
2. $d(x, y) = d(y, x)$,
3. $d(x, y) \leq d(x, z) + d(z, y)$ (this is called the triangle inequality).

Def 12.2.3 Let X, Y be metric spaces with metrics d_X and d_Y.

1. If $x \in X$ and $\epsilon > 0$ then $B(x, \epsilon) = \{y \mid d_X(x, y) < \epsilon\}$. Sets of this form are called *balls*.
2. Let $A \subseteq X$ and $x \in X$. x is a *limit point of A* if

$$(\forall \epsilon > 0)(\exists y \in A)[d(x, y) < \epsilon].$$

3. If $x_1, x_2, \ldots \in X$ then $\lim_i x_i = x$ means $(\forall \epsilon > 0)(\exists i)(\forall j)[j \geq i \implies x_j \in B(x, \epsilon)]$.
4. Let $T : X \to Y$.

   (a) T is *continuous* if for all $x, x_1, x_2, \ldots \in X$

$$\lim_i x_i = x \implies \lim_i T(x_i) = T(x).$$

   (b) T is *uniformly continuous* if

$$(\forall \epsilon)(\exists \delta)(\forall x, y \in X)[d_X(x,y) < \delta \implies d_Y(T(x), T(y)) < \epsilon].$$

5. T is *bi-continuous* if T is a bijection, T is continuous, and T^- is continuous.

6. T is *bi-unif-continuous* if T is a bijection, T is uniformly continuous, and T^- is uniformly continuous.

7. If $A \subseteq X$ then

   (a) A' is the set of all limit points of A.

   (b) $\mathrm{cl}(A) = A \cup A'$. (This is called the *closure of* A).

8. A set $A \subseteq X$ is *closed under limit points* if every limit point of A is in A.

Fact 12.2.4 *If X is a metric space and $A \subseteq X$ then $\mathrm{cl}(A)$ is closed under limit points. That is, if x is a limit point of $\mathrm{cl}(A)$ then $x \in \mathrm{cl}(A)$. Hence $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$.*

Note 12.2.5 The intention in defining the closure of a set A is to obtain the smallest set that contains A that is also closed under limit points. In a general topological space the closure of a set A is the intersection of all closed sets that contain A. Alternatively one can define the closure to be $A \cup A' \cup A'' \cup \cdots$. That $\cdots$ is not quite what is seems- it may need to go into transfinite ordinals (you do not need to know what transfinite ordinals are for this section). Fortunately we are looking at metric spaces where $\mathrm{cl}(A) = A \cup A'$ suffices. More precisely, our definition agrees with the standard one in a metric space.

Example 12.2.6

1. $[0,1]$ with $d(x,y) = |x-y|$ (the usual definition of distance).
   (a) If $A = (\frac{1}{2}, \frac{3}{4})$ then $\mathrm{cl}(A) = [\frac{1}{2}, \frac{3}{4}]$.
   (b) If $A = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$ then $\mathrm{cl}(A) = A \cup \{0\}$.
   (c) $\mathrm{cl}(\mathbb{Q}) = \mathbb{R}$.
   (d) Fix $c \in \mathbb{N}$. Let BISEQ be the set of all c-colorings of $\mathbb{Z}$. (It is called BISEQ since it is a bi-sequence of colors. A bi-sequence is a sequence in two directions.) We represent elements of BISEQ by $f : \mathbb{Z} \rightarrow [c]$.

2. Let d : BISEQ $\times$ BISEQ $\rightarrow \mathbb{R}^{\geq 0}$ be defined as follows.

$$d(f,g) = \begin{cases} 0 & \text{if } f = g, \\ \frac{1}{1+i} & \text{if } f \neq g \text{ and } i \text{ is least number s.t. } f(i) \neq g(i) \text{ or } f(-i) \neq g(-i). \end{cases} \tag{12.1}$$

   One can easily verify that $d(f,g)$ is a metric. We will use this in the future alot so the reader is urged to verify it.

3. The function T is defined by $T(f) = g$ where $g(i) = f(i+1)$. One can easily verify that T is bi-unif-continuous. We will use this in the future alot so the reader is urged to verify it.

Notation 12.2.7 Let $T : X \rightarrow X$ be a bijection. Let $n \in \mathbb{N}$.

1. $T^{(n)}(x) = T(T(\cdots T(x)\cdots))$ means that you apply T to x n times.
2. $T^{(-n)}(x) = T^{-}(T^{-}(\cdots T^{-}(x)\cdots))$ means that you apply T^{-} to x n times.

Def 12.2.8 If X is a metric space and $T : X \rightarrow X$ then

$$\begin{aligned} \mathrm{orbit}(x) = & \quad \{T^{(i)}(x) \mid i \in \mathbb{N}\} \\ \mathrm{dorbit}(x) = & \quad \{T^{(i)}(x) \mid i \in \mathbb{Z}\} \text{ (dorbit stands for for double-orbit)} \end{aligned}$$

Def 12.2.9 Let X be a metric space, $T : X \rightarrow X$ be a bijection, and $x \in X$.

1. $$\mathrm{CLDOT}(x) = \mathrm{cl}(\{\ldots, T^{(-3)}(x), T^{(-2)}(x), \ldots, T^{(2)}(x), T^{(3)}(x), \ldots)$$

   CLDOT(x) stands for *Closure of Double-Orbit of* x.

2. x is *homogeneous* if

   $$(\forall y \in \mathrm{CLDOT}(x))[\mathrm{CLDOT}(x) = \mathrm{CLDOT}(y)].$$

3. X is *limit point compact*[1] if every infinite subset of X has a limit point in X.

Example 12.2.10 Let BISEQ and T be as in Example 12.2.6.2. Even though BISEQ is formally the functions from $\mathbb{Z}$ to $[c]$ we will use colors as the co-domain.

1. Let $f \in$ BISEQ be defined by

   $$f(x) = \begin{cases} \text{RED} & \text{if } |x| \text{ is a square;} \\ \text{BLUE} & \text{otherwise.} \end{cases} \tag{12.2}$$

   The set $\{T^{(i)}(f) \mid i \in \mathbb{Z}\}$ has one limit point. It is the function

   $$(\forall x \in \mathbb{Z})[g(x) = \text{BLUE}].$$

   This is because their are arbitrarily long runs of non-squares. For any M there is an $i \in \mathbb{Z}$ such that $T^{(i)}(f)$ and g agree on $\{-M, \ldots, M\}$. Note that

   $$d(T^{(i)}(f), g) \le \frac{1}{M+1}.$$

   Hence

   $$\mathrm{CLDOT}(f) = \{T^{(i)}(f) \mid i \in \mathbb{Z}\} \cup \{g\}.$$

[1] Munkres [20] is the first one to name this concept "limit point compact"; however, the concept has been around for a long time under a variety of names. Originally, what we call "limit point compact" was just called "compact". Since then the concept we call limit point compact has gone by a number of names: Bolzano-Weierstrass property, Frechet Space are two of them. This short history lesson is from Munkres [20] page 178.

2. Let $f \in \text{BISEQ}$ be defined by

$$f(x) = \begin{cases} \text{RED} & \text{if } x \geq 0 \text{ and } x \text{ is a square or } x \leq 0 \text{ and } x \text{ is not a square;} \\ \text{BLUE} & \text{otherwise.} \end{cases} \tag{12.3}$$

The set $\{T^{(i)}(f) \mid i \in \mathbb{Z}\}$ has two limit points. They are

$$(\forall x \in \mathbb{Z})[g(x) = \text{BLUE}]$$

and

$$(\forall x \in \mathbb{Z})[h(x) = \text{RED}].$$

This is because their are arbitrarily long runs of REDs and arbitrarily long runs of BLUEs.

$$\text{CLDOT}(f) = \{T^{(i)}(f) \mid i \in \mathbb{Z}\} \cup \{g, h\}.$$

3. We now construct an example of an f such that the number of limit points of $\{T^{(i)}(f) \mid i \in \mathbb{Z}\}$ is infinite. Let $f_j \in \text{BISEQ}$ be defined by

$$f_j(x) = \begin{cases} \text{RED} & \text{if } x \geq 0 \text{ and } x \text{ is a } j\text{th power} \\ \text{BLUE} & \text{otherwise.} \end{cases} \tag{12.4}$$

Let $I_k = \{2^k, \ldots, 2^{k+1}-1\}$. Let $a_1, a_2, a_3, \ldots$ be a list of natural numbers so that every single natural number occurs infinitely often. Let $f \in \text{BISEQ}$ be defined as follows.

$$f(x) = \begin{cases} f_j(x) & \text{if } x \geq 1,\ x \in I_k \text{ and } j = a_k; \\ \text{BLUE} & \text{if } x \leq 0. \end{cases} \tag{12.5}$$

For every j there are arbitrarily long segments of f that agree with some translation of f_j. Hence every point f_j is a limit point of $\{T^{(i)}f \mid i \in \mathbb{Z}\}$.

Example 12.2.11 We show that BISEQ is limit point compact. Let $A \subseteq \text{BISEQ}$ be infinite. Let $f_1, f_2, f_3, \ldots \in A$. We construct $f \in \text{BISEQ}$ to be a limit point of $f_1, f_2, \ldots$. Let $a_1, a_2, a_3, \ldots$ be an enumeration of the integers.

$$\begin{aligned} I_0 &= \mathbb{N} \\ f(a_1) &= \text{least color in } [c] \text{ that occurs infinitely often in } \{f_i(a_1) \mid i \in I_0\} \\ I_1 &= \{i \mid f_i(a_1) = f(a_1)\} \end{aligned}$$

Assume that $f(a_1), I_1, f(a_2), I_2, \ldots, f(a_{n-1}), I_{n-1}$ are all defined and that I_{n-1} is infinite.

$$\begin{array}{rl} f(a_n) = & \text{least color in } [c] \text{ that occurs infinitely often in } \{f_i(a_n) \mid i \in I_{n-1}\} \\ I_n = & \{i \mid (\forall j)[1 \le j \le n \implies f_i(a_j) = f(a_j)]\} \end{array}$$

Note that I_n is infinite.

Note 12.2.12 The argument above that BISEQ is limit point compact is a common technique that is often called a *compactness argument*.

Lemma 12.2.13 *If X is limit point compact, $Y \subseteq X$, and Y is closed under limit points then Y is limit point compact.*

Proof: Let $A \subseteq Y$ be an infinite set. Since X is limit point compact A has a limit point $x \in X$. Since Y is closed under limit points, $x \in Y$. Hence every infinite subset of Y has a limit point in Y, so Y is limit point compact. ▮

Def 12.2.14 Let X be a metric space and $T : X \to X$ be continuous. Let $x \in X$.

1. The point x is *recurrent for T* if

   $$(\forall \epsilon)(\exists n)[d(T^{(n)}(x), x) < \epsilon].$$

   Intuition: If x is recurrent for T then the orbit of x comes close to x infinitely often. Note that this may be very irregular.

2. Let $\epsilon > 0$, $r \in \mathbb{N}$, and $w \in X$. w is *(ϵ, r)-recurrent for T* if

   $$(\exists n \in \mathbb{N})[d(T^{(n)}(w), w) < \epsilon \wedge d(T^{(2n)}(w), w) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(w), w) < \epsilon.]$$

   Intuition: If w is (ϵ, r)-recurrent for T then the orbit of w comes within ϵ of w r times on a regular basis.

Example 12.2.15

1. If $T(x) = x$ then all points are recurrent (this is trivial).
2. Let $T : \mathbb{R} \rightarrow \mathbb{R}$ be defined by $T(x) = -x$. Then, for all $x \in \mathbb{R}$, $T(T(x)) = x$ so all points are recurrent.
3. Let $\alpha \in [0,1]$. Let $T : [0,1] \rightarrow [0,1]$ be defined by $T(x) = x + \alpha \pmod 1$.
   (a) If $\alpha = 0$ or $\alpha = 1$ then all points are trivially recurrent.
   (b) If $\alpha \in \mathbb{Q}$, $\alpha = \frac{p}{q}$ then it is easy to show that all points are recurrent for the trivial reason that $T^{(q)}(x) = x + q(\frac{p}{q}) \pmod 1 = x$.
   (c) If $\alpha \notin \mathbb{Q}$ then T is recurrent. This requires a real proof.

12.2.2 A Theorem in Topology

Def 12.2.16 Let X be a metric space and $T : X \rightarrow X$ be a bijection. (X, T) is *homogeneous* if, for every $x \in X$,

$$X = \mathrm{CLDOT}(x).$$

Example 12.2.17

Let $X = [0, 1]$, $d(x, y) = |x - y|$, and $T(x) = x + \alpha \pmod 1$.

1. If $\alpha \in \mathbb{Q}$ then (X, T) is not homogeneous.
2. If $\alpha \notin \mathbb{Q}$ then (X, T) is homogeneous.
3. Let $f, g \in \mathrm{BISEQ}$, so $f : \mathbb{Z} \rightarrow \{1, 2\}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x \equiv 1 \pmod 2; \\ 2 & \text{if } x \equiv 0 \pmod 2 \end{cases} \tag{12.6}$$

   and

$$g(x) = 3 - f(x).$$

   Let $T : \mathrm{BISEQ} \rightarrow \mathrm{BISEQ}$ be defined by

$$T(h)(x) = h(x + 1).$$

Let $X = \mathrm{CLDOT}(f)$. Note that

$$X = \{f, g\} = \mathrm{CLDOT}(f) = \mathrm{CLDOT}(g).$$

Hence (X, T) is homogeneous.

4. All of the examples in Example 2.9 are not homogeneous.

The ultimate goal of this section is to show the following.

Theorem 12.2.18 *Let X be a metric space and $T : X \rightarrow X$ be bi-unif-continuous. Assume (X, T) is homogeneous. Then for every $r \in \mathbb{N}$, for every $\epsilon > 0$, T has an (ϵ, r)-recurrent point.*

Important Convention for the Rest of this Section:

1. X is a metric space.
2. T is bi-unif-continuous.
3. (X, T) is homogeneous.

We show the following by a multiple induction.

1. A_r: $(\forall \epsilon > 0)(\exists x, y \in X, n \in \mathbb{N})$
   $d(T^{(n)}(x), y) < \epsilon \wedge d(T^{(2n)}(x), y) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(x), y) < \epsilon.$

   Intuition: There exists two points x, y such that the orbit of x comes very close to y on a regular basis r times.

2. B_r: $(\forall \epsilon > 0)(\forall z \in X)(\exists x \in X, n \in \mathbb{N})$
   $d(T^{(n)}(x), z) < \epsilon \wedge d(T^{(2n)}(x), z) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(x), z) < \epsilon.$

   Intuition: For any z there is an x such that the orbit of x comes very close to z on a regular basis r times.

3. C_r: $(\forall \epsilon > 0)(\forall z \in X)(\exists x \in X)(\exists n \in \mathbb{N})(\exists \epsilon' > 0)$
   $T^{(n)}(B(x, \epsilon')) \subseteq B(z, \epsilon) \wedge T^{(2n)}(B(x, \epsilon')) \subseteq B(z, \epsilon) \wedge \cdots \wedge T^{(rn)}(B(x, \epsilon'))) \subseteq B(z, \epsilon).$

   Intuition: For any z there is an x such that the orbit of a small ball around x comes very close to z on a regular basis r times.

4. D_r: $(\forall \epsilon > 0)(\exists w \in X, n \in \mathbb{N})$
   $d(T^{(n)}(w), w) < \epsilon \wedge d(T^{(2n)}(w), w) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(w), w) < \epsilon.$

   Intuition: There is a point w such that the orbit of w comes close to w on a regular basis r times. In other words, for all ϵ, there is a w that is (ϵ, r)-recurrent.

Lemma 12.2.19 $(\forall \epsilon > 0)(\exists M \in \mathbb{N})(\forall x, y \in X)$

$$\min\{d(x, T^{(-M)}(y)), d(x, T^{(-M+1)}(y)), \ldots, d(x, T^{(M)}(y))\} < \epsilon$$

Proof:
Intuition: Since (X, T) is homogeneous, if $x, y \in X$ then x is close to some point in the double-orbit of y (using T).

Assume, by way of contradiction, that $(\exists \epsilon > 0)(\forall M \in \mathbb{N})(\exists x_M, y_M \in X)$

$$\min\{d(x_M, T^{(-M)}(y_M)), d(x_M, T^{(-M+1)}(y_M)), \ldots, d(x_M, T^{(M)}(y_M))\} \geq \epsilon$$

Let $x = \lim_{M\to\infty} x_M$ and $y = \lim_{M\to\infty} y_M$. Since (X, T) is homogeneous (so it is the closure of a set) and Fact 12.2.4, $x, y \in X$. Since (X, T) is homogeneous

$$X = \{T^{(i)}(y) \mid i \in \mathbb{Z}\} \cup \{T^{(i)}(y) \mid i \in \mathbb{Z}\}'.$$

Since $x \in X$

$$(\exists^\infty i \in \mathbb{Z})[d(x, T^{(i)}(y)) < \epsilon/4].$$

We don't need the $\exists^\infty$, all we need is to have one such I. Let $I \in \mathbb{Z}$ be such that

$$d(x, T^{(I)}(y)) < \epsilon/4$$

Since $T^{(I)}$ is continuous, $\lim_M y_M = y$, and $\lim_M x_M = x$ there exists $M > |I|$ such that

$$d(T^{(I)}(y), T^{(I)}(y_M)) < \epsilon/4 \wedge d(x_M, x) < \epsilon/4.$$

Hence

$$d(x_M, T^{(I)}(y_M)) \leq d(x_M, x) + d(x, T^{(I)}(y)) + d(T^{(I)}(y), T^{(I)}(y_M)) \leq \epsilon/4 + \epsilon/4 + \epsilon/4 < \epsilon.$$

Hence $d(x_M, T^{(I)}(y_M)) < \epsilon$. This violates the definition of x_M, y_M. ∎

Note 12.2.20 The above lemma only used that T is continuous, not that T is bi-unif-continuous.

$A_r \implies B_r$

Lemma 12.2.21 *A_r:* $(\forall \epsilon > 0)(\exists x, y \in X, n \in \mathbb{N})$
$d(T^{(n)}(x), y) < \epsilon \wedge d(T^{(2n)}(x), y) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(x), y) < \epsilon$
$\implies$
B_r: $(\forall \epsilon > 0)(\forall z \in X)(\exists x \in X, n \in \mathbb{N})$
$d(T^{(n)}(x), z) < \epsilon \wedge d(T^{(2n)}(x), z) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(x), z) < \epsilon.$

Proof:
Intuition: By A_r there is an x, y such that the orbit of x will get close to y regularly. Let $z \in X$. Since (X, T) is homogeneous the orbit of y comes close to z. Hence z is close to $T^{(s)}(y)$ and y is close to $T^{(in)}(x)$, so z is close to $T^{(in+s)}(x) = T^{(in)}(T^{(s)}(x))$. So z is close to $T^{(s)}(x)$ on a regular basis.
Note: The proof merely pins down the intuition. If you understand the intuition you may want to skip the proof.

Let $\epsilon > 0$.

1. Let M be from Lemma 12.2.19 with parameter $\epsilon/3$.

2. Since T is bi-unif-continuous we have that for $s \in \mathbb{Z}$, $|s| \le M$, $T^{(s)}$ is unif-cont. Hence there exists ϵ' such that

   $$(\forall a, b \in X)[d(a, b) < \epsilon' \implies (\forall s \in \mathbb{Z}, |s| \le M)[d(T^{(s)}(a), T^{(s)}(b)) < \epsilon/3].$$

3. Let $x, y \in X$, $n \in \mathbb{N}$ come from A_r with ϵ' as parameter. Note that

   $$d(T^{(in)}(x), y) < \epsilon' \text{ for } 1 \le i \le r.$$

Let $z \in X$. Let y be from item 3 above. By the choice of M there exists s, $|s| \le M$, such that

$$d(T^{(s)}(y), z) < \epsilon/3.$$

Since x, y, n satisfy A_r with ϵ' we have

$$d(T^{(in)}(x), y) < \epsilon' \text{ for } 1 \le i \le r.$$

By the definition of ϵ' we have

$$d(T^{(in+s)}(x), T^{(s)}(y)) < \epsilon/3 \text{ for } 1 \le i \le r.$$

Note that

$d(T^{(in)}(T^{(s)}(x), z)) \le d(T^{(in)}(T^{(s)}(x)), T^{(s)}(y)) + d(T^{(s)}(y), z) \le \epsilon/3 + \epsilon/3 < \epsilon.$

▮

$B_r \implies C_r$

Lemma 12.2.22 *B_r:* $(\forall \epsilon > 0)(\forall z \in X)(\exists x \in X, n \in \mathbb{N})$
$d(T^{(n)}(x), z) < \epsilon \wedge d(T^{(2n)}(x), z) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(x), z) < \epsilon$
$\implies$
C_r: $(\forall \epsilon > 0)(\forall z \in X)(\exists x \in X, n \in \mathbb{N}, \epsilon' > 0)$
$T^{(n)}B(x, \epsilon') \subseteq B(z, \epsilon) \wedge T^{(2n)}(B(x, \epsilon') \subseteq B(z, \epsilon) \wedge \cdots \wedge T^{(rn)}(B(x, \epsilon') \subseteq B(z, \epsilon)$.

Proof:
Intuition: Since the orbit of x is close to z on a regular basis, balls around the orbits of x should also be close to z on the same regular basis.

Let $\epsilon > 0$ and $z \in X$ be given. Use B_r with $\epsilon/3$ to obtain the following:

$$(\exists x \in X, n \in \mathbb{N})[d(T^{(n)}(x), z) < \epsilon/3 \wedge d(T^{(2n)}(x), z) < \epsilon/3 \wedge \cdots \wedge d(T^{(rn)}(x), z) < \epsilon/3].$$

By uniform continuity of $T^{(in)}$ for $1 \le i \le r$ we obtain ϵ' such that

$$(\forall a, b \in X)[d(a, b) < \epsilon' \implies (\forall i \le r)[d(T^{(in)}(a), T^{(in)}(b)) < \epsilon^2]$$

We use these values of x and ϵ'.

Let $w \in T^{(in)}(B(x, \epsilon'))$. We show that $w \in B(z, \epsilon)$ by showing $d(w, z) < \epsilon$.

Since $w \in T^{(in)}(B(x, \epsilon'))$ we have $w = T^{(in)}(w')$ for $w' \in B(x, \epsilon')$. Since

$$d(x, w') < \epsilon'$$

we have, by the definition of ϵ',

$$d(T^{(in)}(x), T^{(in)}(w')) < \epsilon/3.$$

$$d(z, w) = d(z, T^{(in)}(w')) \le d(z, T^{(in)}(x)) + d(T^{(in)}(x), T^{(in)}(w')) \le \epsilon/3 + \epsilon/3 < \epsilon.$$

Hence $w \in B(z\epsilon)$. ▮

Note 12.2.23 The above proof used only that T is unif-continuous, not bi-unif-continuous. In fact, the proof does not use that T is a bijection.

$C_r \implies D_r$

Lemma 12.2.24 C_r: $(\forall \epsilon > 0)(\forall z \in X)(\exists x \in X, n \in \mathbb{N}, \epsilon' > 0)$
$T^{(n)}B(x, \epsilon') \subseteq B(z, \epsilon) \wedge T^{(2n)}(B(x, \epsilon') \subseteq B(z, \epsilon) \wedge \cdots \wedge T^{(rn)}(B(x, \epsilon') \subseteq B(z, \epsilon)$
$\implies$
D_r: $(\forall \epsilon > 0)(\exists w \in X, n \in \mathbb{N})$
$d(T^{(n)}(w), w) < \epsilon \wedge d(T^{(2n)}(w), w) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(w), y) < \epsilon$.

Proof:
Intuition: We use the premise iteratively. Start with a point z_0. Some z_1 has a ball around its orbit close to z_0. Some z_2 has a ball around its orbit close to z_1. Etc. Finally there will be two z_i's that are close: in fact the a ball around the orbit of one is close to the other. This will show the conclusion.

Let $z_0 \in X$. Apply C_r with $\epsilon_0 = \epsilon/2$ and z_0 to obtain z_1, ϵ_1, n_1 such that

$$T^{(in_1)}(B(z_1, \epsilon_1)) \subseteq B(z_0, \epsilon_0) \text{ for } 1 \le i \le r.$$

Apply C_r with ϵ_1 and z_1 to obtain z_2, ϵ_2, n_2 such that

$$T^{(in_2)}(B(z_2, \epsilon_2)) \subseteq B(z_1, \epsilon_1) \text{ for } 1 \le i \le r.$$

Apply C_r with ϵ_2 and z_2 to obtain z_3, ϵ_3, n_3 such that

$$T^{(in_3)}(B(z_3, \epsilon_3)) \subseteq B(z_2, \epsilon_2) \text{ for } 1 \le i \le r.$$

Keep doing this to obtain $z_0, z_1, z_2, \ldots$.

One can easily show that, for all $t < s$, for all i $1 \le i \le r$,

$$T^{(i(n_s+n_{s+1}+\cdots+n_{s+t}))}(B(z_s, \epsilon_s)) \subseteq B(z_t, \epsilon_t)$$

Since X is closed $z_0, z_1, \ldots$ has a limit point. Hence

$$d(z_s, z_t) < \epsilon_0.$$

Using these s, t and letting $n_s + \cdots + n_{s+t} = n$ we obtain

$$T^{(in)}(B(z_s, \epsilon_s)) \subseteq B(z_t, \epsilon_t)$$

Hence

$$d(T^{(in)}(z_s), z_t) < \epsilon_t.$$

Let $w = z_s$. Hence, for $1 \le i \le r$

$$d(T^{(in)}(w), w) \le d(T^{(in)}(z_s), z_s) \le d(T^{(in)}(z_s), z_t) + d(z_t, z_s) < \epsilon_t + \epsilon_0 < \epsilon.$$

∎

$D_r \implies A_{r+1}$

Lemma 12.2.25 *D_r: $(\forall \epsilon > 0)(\exists w \in X, n \in \mathbb{N})$*
$d(T^{(n)}(w), w) < \epsilon \wedge d(T^{(2n)}(w), w) < \epsilon \wedge \cdots \wedge d(T^{(rn)}(w), y) < \epsilon.$
$\implies$
A_{r+1}: $(\forall \epsilon > 0)(\exists x, y \in X, n \in \mathbb{N})$
$d(T^{(n)}(x), y) < \epsilon \wedge d(T^{(2n)}(x), y) < \epsilon \wedge \ , \ldots , \ d(T^{((r+1)n)}(x), y) < \epsilon.$

Proof:

By D_r and $(\forall x)[d(x, x) = 0]$ we have that there exists a $w \in X$ and $n \in \mathbb{N}$ such that the following hold.

$$\begin{aligned} d(w, w) &< \epsilon \\ d(T^{(n)}(w), w) &< \epsilon \\ d(T^{(2n)}(w), w) &< \epsilon \\ &\vdots \\ d(T^{(rn)}(w), w) &< \epsilon \end{aligned}$$

We rewrite the above equations.

$$\begin{aligned} d(T^{(n)}(T^{(-n)}(w)), w) &< \epsilon \\ d(T^{(2n)}(T^{(-n)}(w)), w) &< \epsilon \\ d(T^{(3n)}(T^{(-n)}(w)), w) &< \epsilon \\ &\vdots \\ d(T^{(rn)}(T^{(-n)}(w)), w) &< \epsilon \\ d(T^{((r+1)n)}(T^{(-n)}(w)), w) &< \epsilon \end{aligned}$$

Let $x = T^{(-n)}(w)$ and $y = w$ to obtain

$$\begin{array}{rl} d(T^{(n)}(x),y) & <\epsilon \\ d(T^{(2n)}(x),y) & <\epsilon \\ d(T^{(3n)}(x),y) & <\epsilon \\ & \vdots \\ d(T^{(rn)}(x),y) & <\epsilon \\ d(T^{((r+1)n)}(x),y) & <\epsilon \end{array}$$

▮

Theorem 12.2.26 *Assume that*

1. *X is a metric space,*
2. *T is bi-unif-continuous.*
3. *(X,T) is homogeneous.*

For every $r \in \mathbb{N}$, $\epsilon > 0$, there exists $w \in X$, $n \in \mathbb{N}$ such that w is (ϵ, r)-recurrent.

Proof:

Recall that A_1 states

$$(\forall \epsilon)(\exists x, y \in X)(\exists n)[d(T^{(n)}(x), y) < \epsilon].$$

Let $x \in X$ be arbitrary and $y = T(y)$. Note that

$$d(T^{(1)}(x), y) = d(T(x), T(x)) = 0 < \epsilon.$$

Hence A_1 is satisfied.

By Lemmas 12.2.21, 12.2.22, 12.2.24, and 12.2.25 we have $(\forall r \in \mathbb{N})[D_r]$. This is the conclusion we seek. ▮

12.2.3 Another Theorem in Topology

Recall the following well known theorem, called **Zorn's Lemma**.

Lemma 12.2.27 *Let $(X, \preceq)$ be a partial order. If every chain has an upper bound then there exists a maximal element.*

Proof: See Appendix TO BE WRITTEN ▮

Lemma 12.2.28 *Let X be a metric space, $T : X \rightarrow X$ be bi-continuous, and $x \in X$. If $y \in \mathrm{CLDOT}(x)$ then $\mathrm{CLDOT}(y) \subseteq \mathrm{CLDOT}(x)$.*

Proof: Let $y \in \mathrm{CLDOT}(x)$. Then there exists $i_1, i_2, i_3, \ldots \in \mathbb{Z}$ such that

$$T^{(i_1)}(x), T^{(i_2)}(x), T^{(i_3)}(x), \ldots \rightarrow y.$$

Let $j \in Z$. Since $T^{(j)}$ is continues

$$T^{(i_1+j)}(x), T^{(i_2+j)}(x), T^{(i_3+j)}(x), \ldots \rightarrow T^{(j)}y.$$

Hence, for all $j \in \mathbb{Z}$,

$$T^{(j)}(y) \in \mathrm{cl}\{T^{(i_k+j)}(x) \mid k \in \mathbb{N}\} \subseteq \mathrm{cl}\{T^{(i)}(x) \mid i \in \mathbb{Z}\} = \mathrm{CLDOT}(x).$$

Therefore

$$\{T^{(j)}(y) \mid j \in \mathbb{Z}\} \subseteq \mathrm{CLDOT}(x).$$

By taking cl of both sides we obtain

$$\mathrm{CLDOT}(y) \subseteq \mathrm{CLDOT}(x).$$

▮

Theorem 12.2.29 *Let X be a limit point compact metric space. Let $T : X \rightarrow X$ be a bijection. Then there exists a homogeneous point $x \in X$.*

Proof:
We define the following order on X.

$$x \preceq y \text{ iff } \mathrm{CLDOT}(x) \supseteq \mathrm{CLDOT}(y).$$

This is clearly a partial ordering. We show that this ordering satisfies the premise of Zorn's lemma.

Let C be a chain. If C is finite then clearly it has an upper bound. Hence we assume that C is infinite. Since X is limit point compact there exists x, a limit point of C.

Claim 1: For every $y, z \in C$ such that $y \preceq z$, $z \in \mathrm{CLDOT}(y)$.
Proof: Since $y \preceq z$ we have $\mathrm{CLDOT}(z) \subseteq \mathrm{CLDOT}(y)$. Note that

$$z \in \mathrm{CLDOT}(z) \subseteq \mathrm{CLDOT}(y).$$

End of Proof of Claim 1
Claim 2: For every $y \in C$ $x \in \mathrm{CLDOT}(y)$.
Proof: Let $y_1, y_2, y_3, \ldots$ be such that

1. $y = y_1$,
2. $y_1, y_2, y_3, \ldots \in C$,
3. $y_1 \preceq y_2 \preceq y_3 \preceq \cdots$, and
4. $\lim_i y_i = x$.

Since $y \prec y_2 \prec y_3 \prec \cdots$ we have $(\forall i)[\mathrm{CLDOT}(y) \supseteq \mathrm{CLDOT}(y_i)]$. Hence $(\forall i)[y_i \in \mathrm{CLDOT}(y)]$. Since $\lim_i y_i = x$, $(\forall i)[y_i \in \mathrm{CLDOT}(y)]$, and $\mathrm{CLDOT}(y)$ is closed under limit points, $x \in \mathrm{CLDOT}(y)$.
End of Proof of Claim 2

By Zorn's lemma there exists a maximal element under the ordering $\preceq$. Let this element be x.
Claim 3: x is homogeneous.
Proof: Let $y \in \mathrm{CLDOT}(x)$. We show $\mathrm{CLDOT}(y) = \mathrm{CLDOT}(x)$.

Since $y \in \mathrm{CLDOT}(x)$, $\mathrm{CLDOT}(y) \subseteq \mathrm{CLDOT}(x)$ by Lemma 12.2.28.

Since x is maximal $\mathrm{CLDOT}(x) \subseteq \mathrm{CLDOT}(y)$.

Hence $\mathrm{CLDOT}(x) = \mathrm{CLDOT}(y)$.
End of Proof of Claim 3 ∎

12.2.4 VDW Finally

Theorem 12.2.30 *For all c, for all k, for every c-coloring of $\mathbb{Z}$ there exists a monochromatic arithmetic sequence of length k.*

Proof:

Let BISEQ and T be as in Example 12.2.6.2.

Let $f \in \mathrm{BISEQ}$. Let $Y = \mathrm{CLDOT}(f)$. Since BISEQ is limit point compact and Y is closed under limit points, by Lemma 12.2.13 Y is limit point compact. By Theorem 12.2.29 there exists $g \in X$ such that $\mathrm{CLDOT}(g)$

is homogeneous. Let $X = \mathrm{CLDOT}(g)$. The premise of Theorem 12.2.26 is satisfied with X and T. Hence we take the following special case.

There exists $h \in X$, $n \in \mathbb{N}$ such that h is $(\frac{1}{4}, k)$-recurrent. Hence there exists n such that

$$d(h, T^{(n)}(h)), d(h, T^{(2n)}(h)), \ldots, d(h, T^{(rn)}(h)) < \frac{1}{4}.$$

Since for all i, $1 \le i \le r$, $d(h, T^{(in)}(h)) < \frac{1}{4} < \frac{1}{2}$ we have that

$$h(0) = h(n) = h(2n) = \cdots = h(kn).$$

Hence h has an AP of length k. We need to show that f has an AP of length k.

Let $\epsilon = \frac{1}{2(kn+1)}$. Since $h \in \mathrm{CLDOT}(g)$ there exists $j \in \mathbb{Z}$ such that

$$d(h, T^{(j)}(g)) < \epsilon.$$

Let ϵ' be such that

$$(\forall a, b \in X)[d(a, b) < \epsilon' \implies d(T^{(j)}(a), T^{(j)}(b)) < \epsilon].$$

Since $g \in \mathrm{CLDOT}(f)$ there exists $i \in \mathbb{Z}$ such that $d(g, T^{(i)}(f)) < \epsilon'$. By the definition of ϵ' we have

$$d(T^{(j)}(g), T^{(i+j)}(f)) < \epsilon.$$

Hence we have

$$d(h, T^{(i+j)}(f)) \le d(h, T^{(j)}(g)) + d(T^{(j)}(g), T^{(i+j)}f) < 2\epsilon \le \frac{1}{kn+1}.$$

Hence we have that h and $T^{(i+j)}(f)$ agree on $\{0, \ldots, kn\}$. In particular
$h(0) = f(i + j)$.
$h(n) = f(i + j + n)$.
$h(2n) = f(i + j + 2n)$.
$\vdots$
$h(kn) = f(i + j + kn)$.
Since

$$h(0) = h(n) = \cdots = h(kn)$$

we have

$$f(i+j) = f(i+j+n) = f(i+j+2n) = \cdots = f(i+j+kn).$$

Thus f has a monochromatic arithmetic sequence of length k.

$\blacksquare$

Chapter 13

Appendix: Compactness Theorems

Bibliography

[1] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *Journal of the American Mathematical Society*, pages 725–753, 1996. `http://www.math.ohio-state.edu/~vitaly/` or `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[2] V. Bergelson and A. Leibman. Set-polynomials and polynomial extension of the Hales-Jewett theorem. *Annals of Mathematics*, 150:33–75, 1999. `http://www.math.ohio-state.edu/~vitaly/` or `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[3] J. Bourgain. On triples in arithmetic progression. *Geometric and Functional Analysis*, 9:968–984, 1999.

[4] T. Brown. Some new van der Waerden numbers (preliminary report). *Notices of the American Mathematical Society*, page 432, 1974.

[5] V. Chvátal. Some unknown van der Waerden numbers. In R. G. et al, editor, *Combinatorial Structures and their applications*, pages 31–33. Gordon and Breach, 1969. Proceedings of the Calgary international conference.

[6] R. O. Davies. Partioning the plane into denumerably many sets without repeated differences. *Proceedings of the Cambridge Philosophical Society*, 72:179–183, 1972.

[7] P. Erödos and R. Graham. *Old and New Problems and results in Combinatorial Number Theory*. Academic Press, 1980. book 28 in a series called *L Enseignement Math*. This book seems to be out of print.

[8] J. Fox. An infinite color analogue of Rado's theorem. `www.princeton.edu/~jacobfox/~publications.html`.

[9] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, 1981.

[10] H. Furstenberg and B. Weiss. Topological dynamics and combinatorial number theory. *Journal d'Analyse Mathematique*, 34:61–85, 1978.

[11] W. Gowers. A new proof for Szemerédi's theorem for arithmetic progressions of length four. *Geometric and Functional Analysis*, 8:529–551, 1998. `http://www.springerlink.com` or `http://www.dpmms.cam.ac.uk/~wtg10/papers.html`.

[12] W. Gowers. A new proof of Szemerédi's theorem. *Geometric and Functional Analysis*, 11:465–588, 2001. `http://www.dpmms.cam.ac.uk/~wtg10/papers/html` or `http://www.springerlink.com`.

[13] R. Graham, B. Rothchild, and J. Spencer. *Ramsey Theory*. Wiley, 1990.

[14] B. Green and T. Tao. The primes contain arbitrarily long arithmetic sequence. *Annals of Mathematics*, 167:481–547, 2008. `http://annals.math.princeton.edu/issues/2008/March2008/GreenTao.pdf` or `http://arxiv.org/abs/math/0508063`.

[15] A. Hales and R. Jewitt. Regularity and positional games. *Transactions of the American Mathematical Society*, 1963.

[16] D. Heath-Brown. Integer sets containing no arithmetic progressions. *Proceedings of the London Mathematical Society*, 35(2):385–394, 1987.

[17] P. Komjáth. Partitions of vector spaces. *Periodica Mathematica Hungarica*, 28:187–193, 1994.

[18] M. Kouril. The van der Waerden number w(2,6). *Experimental Mathematics*. to appear.

[19] B. Landmann and A. Robertson. *Ramsey Theory over the integers*. AMS, 2003.

[20] J. Munkres. *Topology: A first course*. Prentice-Hall, 1975.

[21] H. J. Prömel and V. Prömel. An elementary proof of the canonizing version of Gallai-Witt's theorem. *Journal of Combinatorial Theory, Series A*, 42:144–149, 1986. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[22] Radhakrishnan and Srinivasan. Improved bounds and algorithms for hypergraph 2-coloring. *Random Structures and Algorithms*, 16:4–32, 2000.

[23] R. Rado. Studien zur Kombinatorik. *Mathematische Zeitschrift*, pages 424–480, 1933. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[24] R. Rado. Notes on combinatorial analysis. *Proceedings of the London Mathematical Society*, pages 122–160, 1943. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[25] K. Roth. Sur quelques ensembles d' entiers. *C.R. Acad. Sci Paris*, 234:388–3901, 1952.

[26] K. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, 28:104–109, 1953. `http://jlms.oxfordjournals.org/`.

[27] A. Sárközy. On difference sets of sequences of integers I. *Acta Math. Sci. Hung.*, 31:125–149, 1977. `http://www.cs.umd.edu/~gasarch/vdw/sarkozyONE.pdf`.

[28] S. Shelah. Primitive recursive bounds for van der Waerden numbers. *Journal of the American Mathematical Society*, pages 683–697, 1988. `http://www.jstor.org/view/08940347/di963031/96p0024f/0`.

[29] R. Stevens and R. Shantaram. Computer-generated van de Waerden partitions. *Mathematics of Computation*, 32:635–636, 1978.

[30] Z. Szabó. An application of lovász' local lemma–a new lower bound for the van der waerden number. *Random Structures and Algorithms*, 1:343–360, 1990.

[31] E. Szeméredi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Sci. Hung.*, 20:89–104, 1969.

[32] E. Szeméredi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:299–345, 1975.

[33] E. Szemerédi. Integer sets containing no arithmetic progressions. *Acta Math. Sci. Hung.*, 56:155–158, 1990.

[34] B. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Arch. Wisk.*, 15:212–216, 1927.

[35] M. Walters. Combinatorial proofs of the polynomial van der Waerden theorem and the polynomial Hales-Jewett theorem. *Journal of the London Mathematical Society*, 61:1–12, 2000. `http://jlms.oxfordjournals.org/cgi/reprint/61/1/1` or `http://jlms.oxfordjournals.org/` or or `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[36] Witt. Ein kombinatorischer satz de elementargeometrie. *Mathematische Nachrichten*, pages 261–262, 1951. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.