Van der Waerden's Theorem: Variants and "Applications"

William Gasarch, Clyde Kruskal, Andy Parrish

March 9, 2018

Contents

Preface

0.1 What is Ramsey Theory and why did we write this book?

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

> *No matter how you color some combinatorial object there will be an orderly monochromatic subset.*

Alternatively, to quote Theodore S. Motzkin,

> *Complete disorder is impossible.*

The following two theorems are fundamental to Ramsey Theory and also make the points stated above:

1. *Ramsey's Theorem*: For all c, k, there exists n, such that, for every c-coloring of the edges of the complete graph K_n there is a subset of k vertices such that all of the edges among them are the same color. (That set of k vertices is often called *a homogeneous set* or *a monochromatic* K_k.)

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 progression of length k.

In both cases you are coloring a combinatorial object and get an orderly monochromatic subset. Note that these are both for-all statements: *no matter how you c-color* …. Hence they are saying that complete disorder is impossible. Both theorems have purely combinatorial proofs. We come back to this point later.

There are already two elementary books on Ramsey Theory:

1. *Ramsey Theory* by Graham, Spencer, and Rothschild [36].

2. *Ramsey Theory over the Integers*, Landman and Robertson [51]

Both books are elementary in that they mostly do not use advanced techniques. They contain both Ramsey's Theorem and van der Waerden's Theorem. So, why is our book needed?

We have a more focused goal. Our goal is to give a purely combinatorial proof of the generalized polynomial van der Waerden Theorem, which we describe in Section 1.5. This theorem is not covered in either of those books. Indeed, this theorem was not known when those books were published.

We only cover van der Waerden's Theorem and its variants. Given our focus, we can cover more ground. Our goal is to cover virtually every extension, variant, and application of van der Waerden's Theorem that can be proven using purely combinatorial methods.

0.2 What is a purely combinatorial proof?

In this book we will only use purely combinatorial methods. 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 basic combinatorics. By contrast, we now give two true statements from Ramsey Theory that currently have *no* purely combinatorial proof.

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

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

   are all primes. This was proven by Green and Tao [38]. 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

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

   This was proven by Gowers [35]. He used techniques from analysis, notably Fourier Analysis.

0.3 Who could read this book?

Since we only use purely combinatorial techniques a bright high school or college student who knows basic combinatorics (permutations, combinations, proofs by induction) could read this book. Many of the people in the acknowledgments are high school students. However, some of the material in this book is not well known. Hence even people far more advanced in mathematics would benefit from this book.

0.4 Abbreviations used in this book

Throughout this monograph we use the following conventions.

1. VDW is van der Waerden's Theorem.

2. POLYVDW is the Polynomial van der Waerden 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 der Waerden numbers"

0.5 Features of this book

How is this book different from other books?

1. We will use purely combinatorial methods.

2. We will give both intuition and complete proofs.

3. Our proofs will be unified by the *color focusing method.* This method was first used explicitly in Walter's proof of POLYVDW [90]; however, we will use it to prove VDW, HJ, and POLYHJ. We are quick to note that the proofs are the classical proofs; however, we express them in a unified way.

4. We give a purely combinatorial proof of the Generalized Polynomial van der Waerden Theorem.

5. We will give applications to other branches of mathematics and to theoretical computer science.

6. Consider the following theorem: **The Square Theorem:** For all 2-colorings of $\mathbb{N} \times \mathbb{N}$ there exists a square that has all four corners the same color. We will give several different proofs of this.

0.6 Acknowledgments

We would like to thank all past, present, and future members of the *van der Waerden gang* who listened to us as we worked out the ideas and expositions in this monograph. We list out all past and present members: Yosef Berman, Alexander Briese, Daniel Farias, Bruce Gaffney, Lee Steven Greene, Aki Hogge, Justin Kruskal, Matt Jordan, Richard Matthew McCutchen, Nils Molina, Anand Oza, Jefferson Pecht, James Pinkerton, Rohan Puttagunta, Rafael Setra, Sandow Sinai, Louis Wasserman, Sam Zbarsky, and Thomas Zhang.

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

Chapter 1

Introduction

In this chapter we state, without proof, some of the main theorems in this book. In particular we state

1. VDW Theorem.
2. Multdimensional VDW's Theorem (also called *the Gallai-Witt Theorem*).
3. Rado's Theorem (which could be called *The Equations* VDW *Theorem*).
4. POLYVDW Theorem.
5. The generalized POLYVDW over the reals.

We intentionally omit HJ and POLYHJ. They are great theorems; however, they require some care to state, so we defer their statements until the chapters where they are proved.

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:

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

There is a sequence of three numbers that are the same color and are equally spaced, namely $1, 5, 9$ which are all R.

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

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

There is still a sequence of three numbers that are the same color and equally spaced, namely $2, 5, 8$ which are all B.
Is there a way to 2-color $\{1,\ldots,9\}$ and not get such a sequence?

Exercise 1

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.

2. Show that there is a 2-coloring of $\{1,\ldots,8\}$ such that there is no set of three numbers that are equally spaced and the same color.

We want to generalize this.

Def 1.1.1 An *arithmetic progression of length k* is a sequence of natural numbers of 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.

Notation 1.1.2 We often refer to an arithmetic progression of length k as a k-AP where AP stands for Arithmetic Progression.

Def 1.1.3 Given a coloring of a subset of $\mathbb{N}$ a *monochromatic k-AP* is a k-AP all of whose elements are the same color.

In the above example we looked at 2-colorings of $\{1,2,\ldots,9\}$. It turned out that for all 2-colorings of $\{1,\ldots,9\}$ there is always a monochromatic 3-AP; however, there is a 2-coloring of $\{1,\ldots,8\}$ with no monochromatic 3-AP. What if you increased the number of colors? What if you increased the length k?

We now proceed more formally.

Notation 1.1.4

1. $\mathbb{N}$ is the naturals which we take to be $\{1, 2, 3, \ldots, \}$. (Some books take it to include 0.)
2. If $m \in \mathbb{N}$ then $[m]$ is $\{1, \ldots, m\}$.

The following was first proven by van der Waerden [88].
Van der Waerden's Theorem (henceforth VDW**):**

Theorem 1.1.5 *For every $k \geq 1$ and $c \geq 1$ there exists W such that for every c-coloring $\chi{:}[W] \rightarrow [c]$ there exists a monochromatic k-AP. In other words there exists $a, d \in \mathbb{N}$ such that*

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

Def 1.1.6 Let $k, c \in \mathbb{N}$. The value $W(k, c)$ is the least W that satisfies VDW. It is the *van der Waerden number*; henceforth VDW number.

We will prove van der Waerden's Theorem in Section 2.2.4.

1.2 W

hat is the The Multidimensional van der Waerden Theorem?

Van der Waerden's Theorem is about colorings of the naturals. What if you want to color $\mathbb{N} \times \mathbb{N}$? Or $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$? Do you still get some sort of order? You do!

The Square Theorem:

Theorem 1.2.1 *For all c there exists $G = G(c)$ such that for all c-colorings of $G(c) \times G(c)$ there exists a square that has all four corners the same color.*

The multidimensional VDW is a generalization of this.

Def 1.2.2 Let $k \geq 1$. Let $(a_0, a_1) \in \mathbb{Z} \times \mathbb{Z}$. Let $d \geq 1$. The *$k \times k$ grid with corner (a_0, a_1) and difference d* is the set of points

$$\{(a_0, a_1) + (id, jd) \mid 0 \leq i, j \leq k - 1\}.$$

When the corner and difference of a $k \times k$ grid are unspecified we refer to it as *a $k \times k$ grid.*

The following result is the Gallai-Witt Theorem which we discuss in Section 9.5. There is no publication by Gallai that contains it; however, Rado [68],[69] proved it and credits Gallai. Witt [92] proved it independently. We view it as the 2-dimentional VDW.

Theorem 1.2.3 *For all k, for all c, there exists $G = G(k, c)$ such that for all c-colorings of $G \times G$ there is a regular $k \times k$ grid where all of the lattice points in it are the same color.*

We will prove Theorem 1.2.3 as a corollary of HJ.

Exercise 2 State the above theorem for any number of dimensions.

1.3 What is Rado's Theorem?

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

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

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

We can rewrite this in terms of equations.

For all c, for all c-colorings, there exists W $\chi\colon [W] \rightarrow [c]$, 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

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

We rewrite these equations:

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

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, there exists W $\chi : [W] \to [c]$ *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 1.3.1

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

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

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

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

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

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

4. The definition of when a matrix is *distinct regular* is analogous to when a vector is.

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

The full Rado Theorem gives an exact characterization of which matrices are regular. We state the conditions for a vector to be regular and distinct regular. The full Rado Theorem will be stated in the chapter on Rado's Theorem.

Theorem 1.3.2 *The tuple $(b_1, b_2, \ldots, b_n)$ is regular iff some subset of $\{b_1, \ldots, b_n\}$ sums to 0.*

Theorem 1.3.3 *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.*

1.4 What is the Polynomial van der Waerden Theorem?

The standard VDW is about progressions of length k. The standard literature on the Polynomial van der Waerden Theorem uses progressions of length $k+1$. With this shifted notation the conclusion of VDW states that the progression

$$a, a + p_1(d), \ldots, a + p_k(d)$$

is monochromatic where $p_i(x) = ix$. Why restrict ourselves to these functions? What happens if the p_i are more general functions?

Notation 1.4.1 We denote the integers by $\mathbb{Z}$. We denote the set of polynomials with integer coefficients by $\mathbb{Z}[x]$.

Is the following true?

Potential Theorem 1.4.2 For any polynomials $p_1(x), \ldots, p_k(x) \in \mathbb{Z}[x]$ for any natural number c, there exists W such that, for any c-coloring $\chi{:}[W] \to [c]$ there exists $a, d \in \mathbb{N}$ such that

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

No it is not; there is an easy counterexample:

Exercise 3

1. Show that the above potential theorem is false if $k = 1$, $p_1(x) = 1$, $c = 2$.

2. Find counterexamples to the potential theorem where the polynomial is not constant.

Is there a theorem of interest that avoids these counterexamples? Remarkably, the theorem holds if the polynomials all have constant term 0. This was first proven by Bergelson and Leibman [7].

Polynomial van der Waerden Theorem (Henceforth POLYVDW**)**

Theorem 1.4.3 *For any polynomials* $p_1(x), \ldots, p_k(x) \in \mathbb{Z}[x]$ *such that* $(\forall i)[p_i(0) = 0]$*, for any natural number* c*, there exists* W *such that, for any* c*-coloring* $\chi{:}[W] \rightarrow [c]$ *there exists* $a, d \in \mathbb{N}$ *such that*

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

Def 1.4.4 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 satisfies POLYVDW . $W(p_1, \ldots, p_k; c)$ is called a *polynomial van der Waerden number*, henceforth POLYVDW number.

POLYVDW was proved for $k = 1$ by Fürstenberg [26] and (independently) Sarkozy [73]. The original proof of the full theorem by Bergelson and Leibman [7] used ergodic methods. A later proof by Walters [90] uses purely combinatorial techniques. We will present an expanded version of Walters' proof in Section 2.2.4.

1.5 What is the Generalized POLYVDW?

Upon seeing the Polynomial van der Waerden Theorem one may wonder, is it true over the reals? 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 χ:$[-W, W] \rightarrow [c]$ there exists $a, d \in \mathbb{Z}$, $d \neq 0$, such that

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

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

Polynomial van der Waerden Theorem over $\mathbb{R}$: For any natural number c and any polynomials $p_1(x), \ldots, p_k(x) \in \mathbb{R}[x]$ such that $(\forall i)[p_i(0) = 0]$, there exists a finite subset $\mathbb{R}' \subseteq \mathbb{R}$ such that for any c-coloring χ:$\mathbb{R}' \rightarrow [c]$ there exists $a, d \in \mathbb{R}$, $d \neq 0$, such that

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

Why $\mathbb{R}$? Can we generalize this to other of sets of numbers?

Def 1.5.1 An *integral domain* $\mathbb{D} = (D, +, \times)$ consists of a set D and two operations $+, \times$ such that

1. $+$ and $\times$ are commutative and associative.
2. For all $a, b, c \in D$ $a \times (b + c) = a \times b + a \times c$.
3. There exists $0 \in D$ such that, for all $a \in D$, $a + 0 = a$.
4. There exists $1 \in D$ such that, for all $a \in D$, $a \times 1 = a$.
5. For all $a \in D$ there exists $b \in D$ such that $a + b = 0$.
6. For all $a, b \in D$ if $a \times b = 0$ then either $a = 0$ or $b = 0$.

Convention 1.5.2 Formally the integral domain is the set together with the operations. However, we will use the same symbol for the integral domain as we do for the underlying set. Note that this is already common as we use $\mathbb{R}$ for the reals and for the reals together with $+, \times$.

Notation 1.5.3 If $\mathbb{D}$ is an integral domain then $\mathbb{D}[x]$ is the set of polynomials with coefficients in $\mathbb{D}$.

Example 1.5.4

1. $\mathbb{Z}$ is an integral domain.
2. $\mathbb{R}$ and $\mathbb{Q}$ are integral domain; however, more can be said. Every nonzero element has a multiplicative inverse. Integral domains with this property are called *fields*.

Exercise 4

1. Show that, for all $e \in \mathbb{N}$, the set

$$\{a + b\sqrt{e} \mid a, b \in \mathbb{Z}\}$$

   is an integral domain.

2. Show that if $\mathbb{D}$ is an integral domain then $\mathbb{D}[x]$ is an integral domain.
3. Let $\mathbb{Z}_n$ be the numbers $\{0, 1, \ldots, n-1\}$ with addition and multiplication mod n. Fill in the XXX, YYY, ZZZ in the statements below and prove them.
   - For all $n \geq 2$, $\mathbb{Z}_n$ satisfies conditions XXX for being an integral domain.
   - There exists $n \geq 2$ such that $\mathbb{Z}_n$ does not satisfy condition YYY for being an integral domain.
   - For all $n \geq 2$, $\mathbb{Z}_n$ is an integral domain iff n is ZZZ.

Generalized Polynomial van der Waerden Theorem: Let $\mathbb{D}$ be an infinite integral domain. For any natural number c and any polynomials $p_1(x), \ldots, p_k(x) \in \mathbb{D}[x]$ such that $(\forall i)[p_i(0) = 0]$, there exists a finite subset $\mathbb{D}' \subseteq \mathbb{D}$ such that for any c-coloring $\chi{:}\mathbb{D}' \rightarrow [c]$ there exists $a, d \in \mathbb{D}$, $d \neq 0$, such that

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

This was first proven by Bergelson and Leibman [8]. In that paper they proved the Polynomial Hales-Jewett Theorem (which we will state and prove later) using ergodic techniques, and then derived the Generalized POLYVDW Theorem as an easy corollary. Later Walters [90] obtained a purely combinatorial proof of POLYHJ.

Putting all of this together one obtains a purely combinatorial proof of Generalized POLYVDW. This is the motivation for this book: To present the Generalized POLYVDW in purely combinatorial terms. However, having done that, there was so much more of interest that we just had to include.

Chapter 2

Van der Waerden's Theorem

2.1 Introduction

Traditionally, Baudet is credited with the following conjecture:

For any partition of the natural numbers into two sets, one of the sets will have arbitrarily long arithmetic progressions.

Van der Waerden's paper [88] is titled *Beweis einer Baudetschen Vermutung*, which translates as *Proof of a Conjecture of Baudet*; hence, van der Waerden thought he was solving a conjecture of Baudet. However, Soifer [78] gives compelling evidence that Baudet and Schur deserve joint credit for this conjecture.

As for who proved the conjecture there is no controversy: van der Waerden proved it [88]. The proof we give is essentially his. He has written an account of how he came up with the proof [89] which is reprinted in Soifer's book.

VDW is more general than Baudet's conjecture. VDW guarantees long APs within finite rather than infinite sets of natural numbers, and allows for the natural numbers to be divided up into any finite number of sets (by color) instead of just two.

In this chapter, we will prove VDW the same way van der Waerden did. We will express the proof in the color-focusing language of Walters [90].

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

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

Def 2.1.1 Let $k, c \in \mathbb{N}$. The *van der Waerden number* $W(k, c)$ is the least number W that satisfies van der Waerden's Theorem with parameters k, c.

Before proving the 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 two numbers form a 2-AP.

Not bad— we have proven the theorem for an infinite number of cases. How many more could there be?

Notation 2.1.2 We use VDW(k, c) to mean the statement of VDW with the parameters k and c. Note that the two statements VDW(k, c) *holds* and $W(k, c)$ *exists* are equivalent.

The proof has three key ideas. We prove subcases that illustrate these ideas before proving the full theorem itself.

2.2 Proof of van der Waerden's Theorem

2.2.1 VDW(3, 2) and the first key idea

We show that there exists a W such that any 2-coloring of $[W]$ has a monochromatic 3-AP. By enumeration one can show $W(3, 2) = 9$; however, we prefer to use a technique that generalizes to other values of k and c. The proof will show $W(3, 2) \leq 325$.

For this section let $W \in \mathbb{N}$ and let $\chi : [W] \to \{R, B\}$ (where W will be determined later). Imagine breaking up the numbers $\{1, 2, 3, \ldots, W\}$ into blocks of five. We can assume W is divisible by 5.

$$\{1, 2, 3, 4, 5\}, \{6, 7, 8, 9, 10\}, \ldots, \{W - 4, W - 3, W - 2, W - 1, W\}$$

Let B_i be the ith block. Consider what happens *within* a block. Clearly for any block of five there must be three equally spaced elements for which the first two are the same color. We state this formally so we can refer to it.

Fact 2.2.1 *Let B be a block of five elements. If χ is restricted to B. then there exists $a, d,\ d \neq 0$, such that*

$$a, a+d, a+2d \in B$$

and

$$\chi(a) = \chi(a+d).$$

We need to view $\chi\colon [W] \rightarrow \{R, B\}$ differently. The mapping χ can be viewed as assigning to each block one of the 2^5 possible colorings of five numbers: $RRRRR$, $RRRRB$, ..., $BBBBB$. This is...

The First Key Idea: We view χ as a 32-coloring of the blocks. The following viewpoint will be used over and over again in this book: View a c-coloring of $[W]$ as a c^B coloring of (W/B) blocks of size B.

The following is clear from the pigeonhole principle.

Lemma 2.2.2 *Assume $W \geq 5 \cdot 33 = 165$. There exists two blocks B_i and B_j ($1 \leq i < j \leq 33$) with the same coloring.*

Theorem 2.2.3 *Let $W \geq 325$. Let $\chi\colon [W] \rightarrow [2]$ be a 2-coloring of $[W]$. Then there exists $a, d \in \mathbb{N}$ such that*

$$\chi(a) = \chi(a+d) = \chi(a+2d).$$

Proof: Let the colors be RED and BLUE. Assume, by way of contradiction, that there is no monochromatic 3-AP. View $[W]$ as being 65 blocks of five. By Lemma 2.2.2 there exists two blocks B_i, B_j ($1 \leq i < j \leq 33$) with the same coloring. By Fact 2.2.1, within B_i, there exists a, d such that $\chi(a) = \chi(a+d)$. Since B_i and B_j are the same color and are D apart we have that there exists a, d, D such that, up to recoloring, the following holds.

- $\chi(a) = \chi(a+d) = \chi(a+D) = \chi(a+D+d) =$ RED.
- $\chi(a+2d) = \chi(a+D+2d) =$ BLUE.
- $a + 2D + 2d \in [W]$.

Figure 2.1 represents the situation.

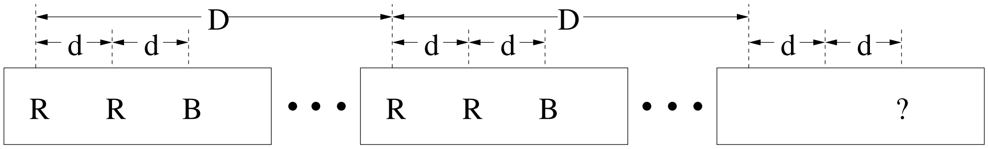

Figure 2.1: Three 5-Blocks

If $\chi(a+2D+2d) =$ BLUE then

$$\chi(a+2d) = \chi(a+D+2d) = \chi(a+2D+2d) = \text{BLUE}.$$

If $\chi(a+2D+2d) =$ RED then

$$\chi(a) = \chi(a+(D+d)) = \chi(a+2(D+d)) = \text{RED}.$$

In either case we get a monochromatic 3-AP, a contradiction. ▮

Exercise 5

1. How many 2-colorings of a 5-block are there that do not have a monochromatic 3-AP? Use the answer to obtain a smaller upper bound on $W(3,2)$ in the proof of Theorem 2.2.3.

2. Use 3-blocks instead of 5-blocks in a proof similar to that of Theorem 2.2.3 to obtain a smaller upper bound on $W(3,2)$ in the proof of Theorem 2.2.3.

3. Show that $W(3,2) = 9$. (Hint: Do not use anything like Theorem 2.2.3.)

4. Find all 3-colorings of [8] that do not have a monochromatic 3-AP.

5. For $n = 10, 11, \ldots$ try to find a 3-coloring of $[n]$ that has no monochromatic 3-AP's by doing the following which is called the Greedy method. (By VDW there will be an n such that this is impossible.)

   - Color the numbers in order.

- If you can color a number RED without forming a monochromatic 3-AP, do so.
- If not, then if you can color that number BLUE without forming a monochromatic 3-AP, do so.
- If not, then if you can color that number GREEN without forming a monochromatic 3-AP, do so.
- If every color would form a monochromatic 3-AP, then stop.

6. (Open-ended) For $n = 10, 11, \ldots$, try to find a 3-coloring of $[n]$ that has no monochromatic 3-AP's. (By VDW there will be an n such that this is impossible.) Do this by whatever means necessary. Get as large an n as you can. Be all you can be!

2.2.2 VDW$(3, 3)$ and the second key idea

We show that there exists a W such that any 3-coloring of $[W]$ has a monochromatic 3-AP. It is known, using a computer program, that $W(3, 3) = 27$ [13]. We use a technique that generalizes to other values of k and c, but does not attain the exact bound.

For this section let $W \in \mathbb{N}$ and let $\chi{:}[W] \to \{R, B, G\}$ (where W is to be determined later). Imagine breaking up the numbers $\{1, 2, 3, \ldots, W\}$ into blocks of seven. We can assume W is divisible by 7.

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

By techniques similar to those used in Section 2.2.1 we obtain that there is some number U such that, for all 3-colorings of $[U]$, up to recoloring, there exists a, d, D such that

- $\chi(a) = \chi(a + d) = \chi(a + D) = \chi(a + D + d) =$ RED.
- $\chi(a + 2d) = \chi(a + D + 2d) =$ BLUE.
- $a + 2D + 2d \in [W]$.

Figure 2.2 represents the situation.

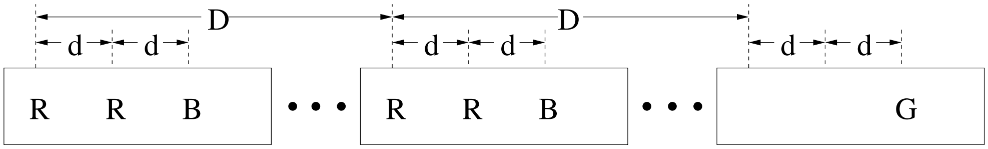

Figure 2.2: Three 7-Blocks

If $\chi(a+2D+2d) = \text{BLUE}$ then

$$\chi(a+2d) = \chi(a+D+2d) = \chi(a+2D+2d) = \text{BLUE}.$$

If $\chi(a+2D+2d) = \text{RED}$ then

$$\chi(a) = \chi(a+(D+d)) = \chi(a+2(D+d)) = \text{RED}.$$

Unfortunately all we can conclude is that $\chi(a+2D+2d) = \text{GREEN}$.
We have sketched a proof of the following:

Lemma 2.2.4 *There exists U such that, up to recoloring, for all 3-colorings of $[U]$ one of the following must occur.*

1. *There exists a monochromatic 3-AP.*
2. *There exists two 3-AP's such that*
   - *One is colored* RED − RED − GREEN.
   - *One is colored* BLUE − BLUE − GREEN.
   - *They have the same third point.*

Let U be as in Lemma 2.2.4. Imagine breaking up the numbers $\{1, 2, 3, \ldots, W\}$ into blocks of U (we can assume W is divisible by U).

The Second Key Idea: We now take $[U]$ to be our block. We view $[W]$ as a sequence of blocks, each of length U. This viewpoint will be used over and over again in this book. First divide $[W]$ into blocks, then later take a block of blocks, and then a block of blocks of blocks, etc.

We resume our discussion. View the 3-coloring of $[W]$ as a 3^U coloring of the blocks. Take W large enough so that there are two blocks B_i, B_j that are the same color and a third block B_k such that B_i, B_j, B_k form an arithmetic progression of blocks.

Figure 2.3 represents the situation.

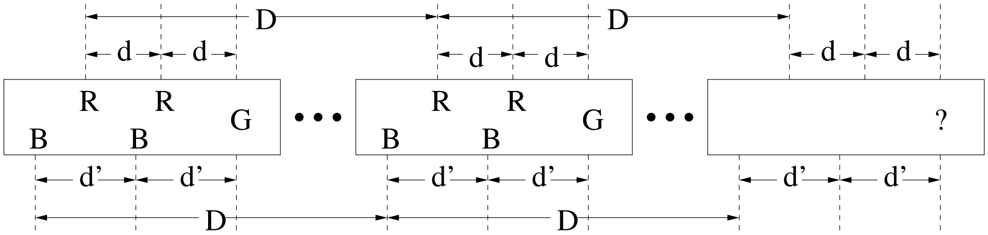

Figure 2.3: Three U-Blocks

We leave the formal proof of Lemma 2.2.4 and the proof of VDW$(3,3)$ to the reader.

Exercise 6

1. Use the ideas in this section to produce a rigorous proof of VDW$(3,3)$. Obtain an actual number that bounds $W(3,3)$.

2. Use the ideas in this section to produce a rigorous proof of VDW$(3,4)$.

3. Use the ideas in this section to produce a rigorous proof that, for all c, $vdw(3,c)$ holds.

2.2.3 VDW (4,2): and the third key idea

We show that there exists a W such that any 2-coloring of $[W]$ has a monochromatic 4-AP. It is known, using a computer program, that $W(4,2) = 35$ [13]. We use a technique that generalizes to other values of k and c, but does not attain the exact bound.

For this section let $W \in \mathbb{N}$ and let χ:$[W] \rightarrow \{R, B\}$ (W to be determined later). Imagine breaking up the numbers $\{1,2,3,\ldots,W\}$ into blocks of length $2W(3,2)$ (we can assume W is divisible by $2W(3,2)$).

$$\{1,2,3,4,5,\ldots,2W(3,2)\}, \{2W(3,2)+1,\ldots,4W(3,2)\}, \{4W(3,2)+1,\ldots,6W(3,2)\}, \cdots$$

We will use VDW$(3,c)$ for rather large values of c to prove VDW$(4,2)$. This is...

The Third Key Idea: To prove VDW$(k,2)$ we will use VDW$(k-1,c)$ for an enormous value of c. Formally, this is an ω^2 induction. We will discuss and use inductions on complicated orderings later in this book.

We leave the following easy lemma to the reader.

Lemma 2.2.5 *Let $c \in \mathbb{N}$. Let χ:$[2W(3,c)] \rightarrow [c]$. There exists $a, d \in \mathbb{N}$ such that*

- $\chi(a) = \chi(a+d) = \chi(a+2d)$, *and*
- $a+3d \in [2W(3,c)]$ *so* $\chi(a+3d)$ *is defined, though we make no claims of its value.*

Theorem 2.2.6 *Let $W \geq 4W(3,2) \times W(3, 2^{2W(3,2)})$. Let χ:$[W] \rightarrow [2]$ be a 2-coloring of $[W]$. Then there exists $a, d \in \mathbb{N}$ such that*

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

Proof: Let the colors be RED and BLUE. Assume, by way of contradiction, that there is no monochromatic 4-AP. View $[W]$ as being $2W(3, 2^{2W(3,2)})$ blocks of size $2W(3,2)$. We view the 2-coloring of $[W]$ as a $2^{2W(3,2)}$-coloring of the blocks. We will use VDW$(3, 2^{2W(3,2)})$ on the block-coloring and VDW$(3,2)$ on the coloring of each block. By Lemma 2.2.5 applied to *both* the coloring of the blocks and the coloring within a block, and symmetry, we have the following: There exists $a, d, D \in \mathbb{N}$ such that

- $\chi(a) = \chi(a+d) = \chi(a+2d) = \text{RED}$,
  $\chi(a+D) = \chi(a+D+d) = \chi(a+D+2d) = \text{RED}$,
  $\chi(a+2D) = \chi(a+2D+d) = \chi(a+2D+2d) = \text{RED}$.
- $\chi(a+3d) = \chi(a+D+3d) = \chi(a+2D+3d) = \text{BLUE}$.
- $a+3D+3d \in [W]$.

Figure 2.4 represents the situation.

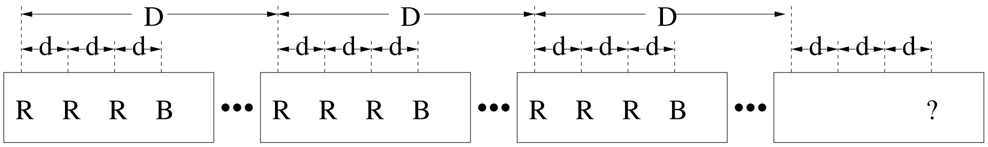

Figure 2.4: Four $2W(3, 2)$-Blocks

If $\chi(a + 3D + 3d) =$ BLUE then

$$\chi(a + 3d) = \chi(a + 3d + D) = \chi(a + 2D + 3d) = \chi(a + 3D + 3d) = \text{BLUE}.$$

If $\chi(a + 3D + 3d) =$ RED then

$$\chi(a) = \chi(a + (D + d)) = \chi(a + 2(D + d)) = \chi(a + 3(D + d)) = \text{RED}.$$

In either case we get a monochromatic 4-AP, a contradiction. ▮

Exercise 7

1. Use the proof of Theorem 2.2.6 to obtain an actual bound on $W(4, 2)$.
2. Fix k. Assume that, for all c, VDW$(k-1, c)$ is true. Prove VDW$(k, 2)$.
3. Fix k. Assume that, for all c, VDW$(k-1, c)$ is true. Prove VDW$(k, 3)$.
4. Prove the full VDW.

2.2.4 The full proof

Now that you know the Key Ideas you have all of the intuitions for the proof. We formalize them here. The method we use here, color focusing, will occur again and again in this book.

We will prove a lemma from which van der Waerden's Theorem will follow easily. Informally, the lemma states the following: if you c-color a large enough $[U]$, then *either* there will be a monochromatic k-AP *or* there will be an arbitrarily large number of monochromatic $(k-1)$-AP's, all of different colors. Once there are $c + 1$ such $(k-1)$-AP's the latter cannot happen, so the former must.

Lemma 2.2.7 *Fix $k, c \in N$ with $k \geq 3$. Assume $(\forall c')[\mathrm{VDW}(k-1, c')]$holds]. Then, for all r, there exists $U = U(r)$*[1] *such that for all c-colorings χ:$[U] \rightarrow [c]$, one of the following statements holds.*

Statement I: *There are $a, d \in \mathbb{N}$ such that*

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

Statement II: *There exists an anchor $a \in \mathbb{N}$ and numbers $d_1, d_2, \ldots, d_r \in \mathbb{N}$, such that*

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

$$\chi(a+d_2) = \chi(a+2d_2) = \cdots = \chi(a+(k-1)d_2)$$

$$\vdots$$

$$\chi(a+d_r) = \chi(a+2d_r) = \cdots = \chi(a+(k-1)d_r)$$

and, for all $i \neq j$, $\chi(a+d_i) \neq \chi(a+d_j)$.

Proof:

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

Base Case: $r = 1$. We show that $U(1) \leq 2W(k-1, c)$. Let χ:$[2W(k-1, c)] \rightarrow [c]$. Apply $\mathrm{VDW}(k-1, c)$ to the *last half of* $[U(1)]$ to obtain $a', d \in \mathbb{N}$ such that

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

and

$$a' - d \in [U(1)].$$

Figure 2.5 represents the situation.

[1] Formally U depends on k, c, r; however, we suppress the dependence on k and c for ease of notation.

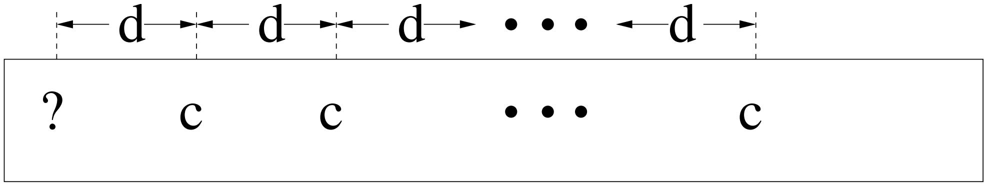

Figure 2.5: Base Case

Let $a = a' - d$. If $\chi(a) = \chi(a')$ then $a' - d, a', a' + d, \ldots, a' + (k-2)d$ is a monochromatic k-AP that satisfies Statement I. If $\chi(a) \neq \chi(a')$ then a, d satisfy Statement II.

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

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

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

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

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

Just one of these block looks like Figure 2.6. Figure 2.7 represents the situation we have with $W(k-1, c^{U(r)})$ blocks.

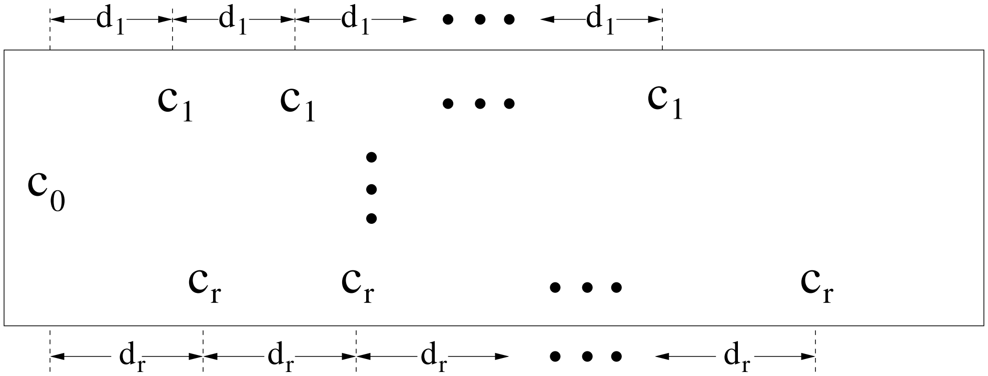

Figure 2.6: One Block of Size $U(r)$

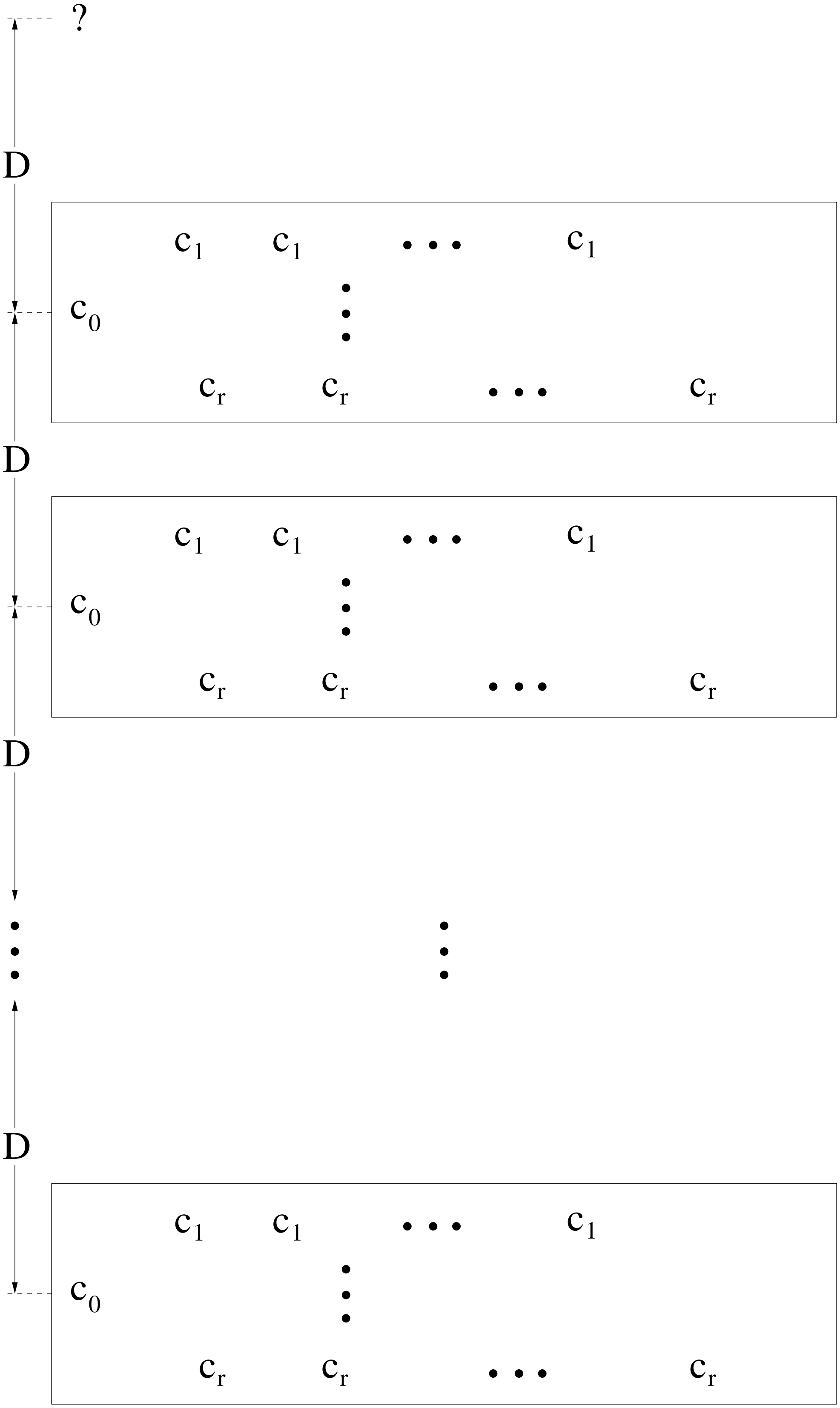

Figure 2.7: Many Blocks of $U(r)$

We view a c-coloring of the second half of $[U]$ as a $c^{U(r)}$-coloring of these blocks.

Let χ^* be this coloring. By the definition of $W(k-1, c^{U(r)})$, we get a monochromatic $(k-1)$-AP of blocks. Hence we have A, D' such that

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

Figure 2.7 represents the situation.

Now, consider block B_A. It is colored by χ. It has length $U(r)$, which tells us that either Statement I or II from the lemma holds. If Statement I 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

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

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

where $\chi(a'+d_i)$ 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 $U(r)$ elements in it, corresponding elements in adjacent blocks are $D = D' \times U(r)$ apart. Hence

$$\begin{aligned}
\chi(a'+d_1) = \chi(a'+D+d_1) = \cdots &= \chi(a'+(k-2)D+d_1) \\
\chi(a'+d_2) = \chi(a'+D+d_2) = \cdots &= \chi(a'+(k-2)D+d_2)
\end{aligned}$$

$$\vdots$$

$$\chi(a'+d_r)=\chi(a'+D+d_r)=\cdots=\chi(a'+(k-2)D+d_r)$$

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

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

and

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

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

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

Where each progression uses different color.

We need an $(r+1)^{\text{st}}$ monochromatic set of points. Consider

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

These are corresponding points in blocks which have the same color under χ^*, hence

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

In addition, since

$$(\forall i)[\chi(a')\ne\chi(a'+d_i)]$$

the color of this new progression is different from the r progression 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 2.2.8 ***(Van der Waerden's Theorem)*** $\forall k, c \in \mathbb{N}, \exists W = W(k,c)$ *such that, for all c-colorings* $\chi{:}[W] \rightarrow [c]$, $\exists a, d \in \mathbb{N}, d \neq 0$ *such that*

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

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.

Recall that VDW(k,c) means that Van der Waerden's Theorem holds with parameters k, c.

Induction Step: Assume $(\forall c)[$VDW$(k-1,c)$ holds$]$. Fix c. Consider what Lemma 2.2.7 says when $r = c$. In any c-coloring of $U = U(c)$, 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 $W(k,c) \leq U(c)$ and hence exists. ▮

Note that the proof of VDW(k,c) depends on VDW$(k-1,c')$ where c' is quite large. Formally the proof is an induction on the following order on $\mathbb{N} \times \mathbb{N}$.

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

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

Exercise 8 We describe several games related to VDW. For all of them there are two parameters n and k, and two players: I and II. Initially $[n]$ is uncolored. Players alternate coloring a previously uncolored element with Player I going first (duh). Player I uses RED, Player II uses BLUE.

1. AP game: the first player to get a k-AP in his color wins. (Variant: Play to the end and whoever has the most k-AP's wins. It can be a tie.)

2. AVOID-AP game: The first player to get a k-AP in his color loses. (Variant: Play to the end and whoever has the fewest k-AP's wins. It can be a tie.)

3. Maker-Breaker Game: The first player wins if he gets a k-AP in his color, the second player wins if the first player does not.

4. A variant of all of them: Have two more parameters a and b. Player I colors a numbers in his turn, and Player II colors b numbers in his turn.

Here are questions about these games.

1. Fix k. Show that there exists n such that, no matter how the players play, the game cannot be a draw.

2. (Open-ended) For various choices of the parameters, determine which player wins, assuming they both play perfectly. (Hint: only use small values of the parameters.)

3. (With a computer program) For various choices of the parameters determine which player is more likely to win if they both play randomly.

4. (With a computer program) For various choices of the parameters, determine the probability that a perfect player wins while playing against a random opponent.

5. (Fun) Play the game with your younger relatives to get them interested in math.

Note 2.2.9 József Beck [3, 4, 5, 6] has studied many games of the type in Exercise 8, as well as other types of combinatorial games.

Exercise 9 Consider the following games played between EMPTIER and FILLER. We denote EMPTIER by E and FILLER by F. There is one parameter: an ordered set $(X, \preceq)$.

- FILLER fills a box with a finite number of balls. Each ball has an element of X on it. (Many balls may have the same element of X on them.)

- In every round E makes the first move. E's move consists of taking a ball from the box. F then counters by replacing the ball with a finite number of balls that have a smaller element of X on it (with respect to the order $\preceq$). (For example, if $X = \mathbb{N}$ a ball labeled 1000 could be replaced with 999,999,999 balls of rank 999 and 888,876,234,012 balls of rank 8.)
- If the box is ever empty then E wins. If the box is always nonempty (i.e., the game goes on forever) then F wins.

For each of the following $(X, \preceq)$ determine which player wins. Prove your result. Note what kind of induction you need to use.

1. $(X, \preceq)$ is $\mathbb{N}$ with its usual ordering.
2. $(X, \preceq)$ is $\mathbb{Q}$ with its usual ordering.
3. $(X, \preceq)$ is $\mathbb{Z}$ with its usual ordering.
4. $X = \{0, 1\} \times \mathbb{N}$ with the ordering
$$(0,0) \preceq (0,1) \preceq (0,2) \preceq \cdots \preceq (1,0) \preceq (1,1) \preceq (1,2) \preceq (1,3) \preceq \cdots$$
5. $X = \mathbb{N} \times \mathbb{N} \times \mathbb{N}$ with the ordering $(a, b, c) \preceq (d, e, f)$ if either (1) $a < c$ or (2) $a = c$ and $b < d$, or (3) $a = d$ and $b = e$ but $c < f$.

Find a property P so that this statement is true:
"E can win the $(X, \preceq)$-game if and only if $(X, \preceq)$ has property P."

Note 2.2.10 Raymond Smullyan [77] came up with the game in Exercise 9 Similar games are discussed in an article by Martin Gardner [29].

Exercise 10 The notation $(x)_b$ means n is a base b number. If we use this notation then n is written as a sequence of elements from $\{0, 1, 2, \ldots, b-1\}$. Consider the following method to generate a sequence. We are initially given b and a base b number n. We set $G(0) = (n)_b$. Intuitively we will increase the base by 1 and decrease the number by 1. Formally, for all $n \geq 1$,

$$[G(n) = (G(n-1))_{b+n} - 1.$$

For example, if $b = 10$ and $G(0) = (1928)_{10}$ then we obtain the following sequence. (We write the base 10 numbers in parenthesis on each line so we can see how the numbers grow.) (we write

- $G(0) = (1928)_{10}$. (This is 1928 in base 10.)
- $G(1) = (1928)_{11} - 1 = (1927)_{11}$. (This is 2449 in base 10.)
- $G(2) = (1927)_{12} - 1 = (1926)_{12}$. (This is 3054 in base 10.)
- $G(3) = (1926)_{13} - 1 = (1925)_{13}$. (This is 3749 in base 10.)

1. (Use a computer program.) For $1 \le i \le 10$ determine the first 10 elements of the sequence that begins with $G(i)$ in base 5.
2. Prove that, for all i, the G-sequence that begins with $G(0) = i$ is eventually 0.
3. (Informal) What kind of induction did you use in that proof?
4. (Informal) Let $T(n)$ be the least t such that the G-sequence beginning with $G(0) = t$ terminates. How fast does $T(n)$ grow?

Exercise 11 (This exercise is an extension of the last one.) Let $b \in \mathbb{N}$. A numbers is in *hereditary base b form* if it is written as a weighted sum of powers of b (the weights are $\le b-1$). and the exponents are written as a weight sum of powers of b (the weights are $\le b-1$), and the exponents of that, etc. For example, 649782 in base 3 hereditary notation we do in steps:

$$649782 = 3^{12} + 2 \times 3^{10} + 3^5 = 3^{3^2+3^1} + 2 \times 3^{3^2+3^0} + 3^{3^1+2\times 3^0}.$$

The notation $[x]_b$ means n is written in hereditary base b. If we use this notation then n is written in hereditary base b form. Consider the following method to generate a sequence. We are initially given b and a number n written in hereditary base b form. We set $H(0) = [n]_b$. Intuitively we will increase the hereditary base by 1 and decrease the number by 1. Formally, for all $n \ge 1$,

$$[H(n) = [H(n-1)]_{b+n} - 1.$$

For example, if $b = 3$ and $H(0) = 649782$ then

- $H(0) = 3^{3^2+3^1} + 2 \times 3^{3^2+3^0} + 3^{3^1+2\times 3^0}$ (This is 64972 in base 10.)

- $H(1) = 4^{4^2+4^1} + 2 \times 4^{4^2+4^0} + 4^{4^1+2\times 4^0}) - 1 =$
  $4^{4^2+4^1} + 2 \times 4^{4^2+4^0} + +3 \times 4^{4^1+4^0} + 3 \times 4^{4^1} + 3 \times 4^{3\times 4^0} + 3 \times 4^{2\times 4^0} + 3 \times 4^{4^0}$.
  (This is 1133871370239 in base 10.)

1. (Use a computer program.) For $1 \le i \le 5$ determine the first 5 elements of the sequence that begins with $H(i)$ in base 2.

2. Prove that, for all i, the H-sequence that begins with $H(0) = i$ is eventually 0.

3. (Informal) What kind of induction did you use in that proof?

4. (Informal) Let $S(n)$ be the least t such that the H-sequence beginning with $H(0) = t$ terminates. How fast does $S(n)$ grow?

Note 2.2.11 Goodstein [33] first proved for all i, the H-sequence that begins with $H(0) = i$ is eventually 0. Kirby and Paris [44], Cichon [14], and Caicedo [12] have studied the growth rate of $S(n)$.

Definition: Let $e, d_1, \ldots, d_k \in \mathbb{N}$. The *cube on* $(e, d_1, \ldots, d_k)$, denoted $C(e, d_1, \ldots, d_k)$, is the set $\{e + b_1 d_1 + \cdots + b_k d_k \mid b_1, \ldots, b_k \in \{0, 1\}\}$. To emphasize the number of d_i's, this is also called a *k-cube*.

Exercise 12

Consider the following statement which we call Hilbert's Cube Lemma (HCL):

For all k, c there exists $H = H(k, c)$ such that, for all c-colorings of $[H]$, there exists a monochromatic cube.

1. Prove HCL from VDW. Express a bound for $H(k, c)$ in terms of VDW numbers.

2. Prove HCL without using VDW.

3. Compare the bounds you get in each proof.

Note 2.2.12 Hilbert [41] first proved HCL. This seems to be the first Ramseyian theorem ever proven. This theorem could have launched Ramsey Theory; however, Hilbert saw it as a minor lemma en route to a theorem. Nobody else picked up on it as being a new type of mathematics.

Exercise 13 Prove the following: For all e, k there exists U such that for any sequence of naturals

$$A = \{a_1 < a_2 < \ldots < a_U\}$$

with $(\forall i)[a_{i+1} - a_i \le e]$, there exists an arithmetic sequence of length k within A.

2.3 The van der Waerden numbers

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

Our bounds are not primitive recursive!

We discuss the upper bounds on $W(k, c)$ provided by the proof of Theorem 2.2.8. We actually discuss a rather large set of functions that encompass most of mathematics, the primitive recursive functions.

Def 2.3.1 A function $f(x_1, \ldots, x_n)$ is *primitive recursive* if either:

1. f is the *Zero* function, i.e. $f(x_1, \ldots, x_n) = 0$;

2. f is the *Successor* function, i.e. $f(x_1, \ldots, x_n) = x_i + 1$;

3. f is the *Projection* function, i.e. $f(x_1, \ldots, x_n) = x_i$;

4. f is defined by the *Composition* of (previously defined) primitive recursive functions, i.e. if $g_1(x_1, \ldots, x_n)$, $g_2(x_1, \ldots, x_n)$, …, $g_k(x_1, \ldots, x_n)$ are primitive recursive and $h(x_1, \ldots, x_k)$ is primitive recursive, then

   $$f(x_1, \ldots, x_n) = h(g_1(x_1, \ldots, x_n), \ldots, g_k(x_1, \ldots, x_n))$$

   is primitive recursive. This rule for deriving a primitive recursive function is called *the Composition rule.*

5. f is defined by *Recursion* of two primitive recursive functions, i.e. if $g(x_1, \ldots, x_{n-1})$ and $h(x_1, \ldots, x_{n+1})$ are primitive recursive then the following function is also primitive recursive

   $$\begin{aligned} f(x_1, \ldots, x_{n-1}, 0) &= g(x_1, \ldots, x_{n-1}) \\ f(x_1, \ldots, x_{n-1}, m+1) &= h(x_1, \ldots, x_{n-1}, m, f(x_1, \ldots, x_{n-1}, m)) \end{aligned}$$

This rule for deriving a primitive recursive function is called *the Recursion rule.*

Virtually every $\mathbb{N}$-valued function you can think of is primitive recursive. Essentially every computer program computes a primitive recursive function.

Example 2.3.2

1. $f(x) = x + 5$. This is primitive recursive by composing Successor 5 times.

2. $f(x, y) = x + y$. This is primitive recursive by defining addition recursively.

$$\begin{aligned} f(x,0) &= x \\ f(x,y+1) &= f(x,y)+1 \end{aligned}$$

3. $f(x, y) = xy$. This is primitive recursive by defining multiplication recursively.

$$\begin{aligned} f(x,0) &= x \\ f(x,y+1) &= f(x,y)+y \end{aligned}$$

4. $$f(x) = \begin{cases} 0 & \text{if } x \text{ is prime} \\ 1 & \text{if } x \text{ is not prime} \end{cases} \tag{2.1}$$

   $f(x)$ is primitive recursive; however, this is difficult to show so we omit it. The reader is invited to try to prove it.

5. Let $f(x)$ be the xth prime. $f(x)$ is primitive recursive; however, this is difficult to show so we omit it. The reader is invited to try to prove it.

6. If you replace *prime* with *square* or *power of 14* or *Fibonacci prime* or *perfect number* in the last two example then the resulting function is primitive recursive. Several of these are difficult.

Exercise 14

1. Show that $f(x, y) = x^y$ is primitive recursive.

2. Show that there exists a 0-1 valued function that is *not* primitive recursive. (Hint: Use diagonalization.)

Notice the pattern here. Successor is primitive recursive. Iterated successor is addition, so that is primitive recursive. Iterated addition is multiplication, so that is primitive recursive. Iterated multiplication is exponentiation, so that is primitive recursive. What is iterated exponentiation? It is often called *the tower function*, which we denote by TOW. $TOW(x, y)$ is the result of taking x to the x to the x ... to the x, where you do this y times. That is quite large. But it does not stop there. What happens if you iterate TOW? This is called the WOW function.

The bound we proved on $W(k, c)$ is not primitive recursive! The reason is that the bound was quite large. Is there some natural function that is similar to our bound? Yes, Ackermann's function.

Def 2.3.3 *Ackermann's function* is the function defined by

$$\begin{aligned} A(0, y) &= y + 1 \\ A(x + 1, 0) &= A(x, 1) \\ A(x + 1, y + 1) &= A(x, A(x + 1, y)) \end{aligned}$$

It is easy to see that Ackermann's function is computable; however, it cannot be primitive recursive because it grows too fast. Ackermann's function grows fast by using a recursion where the depth of the recursion is itself an argument to the function. Primitive recursive functions have a constant bound on the depth of the recursion. While this is not a proof that Ackermann's function is not primitive recursive, it is an intuition.

Primitive recursive bounds are found!

By the mid 1980's the best known bounds on $W(k, c)$ were still those from Theorem 2.2.8. In particular, they were not primitive recursive. It was thought that one of the following would happen:

1. A *logician* would prove that $W(k, c)$ *was not primitive recursive*. In 1977 Paris and Harrington showed that the *Large Ramsey numbers* were not primitive recursive [61]. In fact,they showed something much stronger: the Large Ramsey Theorem could not be proven in Peano Arithmetic. There were similar results by Kanamori and McAloon [43, 59]. The proof of VDW was in Peano Arithmetic; however, perhaps a refinement of their techniques would lead to $W(k, c)$ not being primitive recursive.

2. A *combinatorialist* would prove that $W(k, c)$ *is primitive recursive* with a different proof of VDW.

What happened? In 1988 Shelah, *a logician*, proved that $W(k, c)$ *is primitive recursive* with a different proof of VDW! His proof is purely combinatorial. He actually found a proof of the Hales-Jewett Theorem which yielded better upper bounds on the Hales-Jewett numbers which, in turn, gave better upper bounds on the VDW numbers. We will give his proof in Chapter8.

Real bounds are found!

Shelah's upper bounds were still quite large. They do not have an easy description like $W(k, c) \leq 2^{2^{kc}}$. Better bounds were found by following a research plan that was first formulated by ErdHos.

Imagine that you 12-color $\mathbb{N}$. There must be a monochromatic 84-AP. Imagine that I tell you, informally, that RED appears the most times. Then you might think that there is a RED 84-AP. Erdős conjectured that this was indeed the case. We state this rigorously in our timeline below.

Def 2.3.4 If $A \subseteq \mathbb{N}$, then the *upper density of A* is $\limsup_{n\to\infty} \frac{|A\cap[n]|}{n}$.

We give a history of how better bounds on $W(k, c)$ were found. We also include some side roads about POLYVDW.

1. In 1927 van der Waerden proves VDW. His proof yields bounds on $W(k, c)$ that are not primitive recursive.

2. In 1936 Erdős and Turan [24] conjectured that every set of positive upper density has a 3-AP. A proof of this would yield a proof of VDW$(3, c)$ that is very different from the original proof (and from Shelah's proof). They have often been credited with conjecturing that every set of positive upper density has a k-AP with this paper as the reference; however,

Soifer [78] gives compelling evidence that the conjecture for k-AP was made by Erdős in 1957 (next item).

3. In 1953 Roth [72] (see also [58]) proved that, for every $\delta > 0$, for every $N \geq 2^{2^{O(\delta^{-1}}}$, for every $A \subseteq [N]$ with $|A| \geq \delta N$, A has a 3-AP. The proof used Fourier Analysis. This result *did* lead to better bounds on $W(3, c)$, namely $W(3, c) \leq 2^{2^{O(c)}}$ (see Exercise 17).

4. In 1957 Erdős [22] conjectured that every set of positive upper density has a k-AP. A proof of this would yield a proof of VDW that is very different from the original proof (and from Shelah's proof). Such a proof might lead to better bounds on $W(k, c)$. We will call this *The Conjecture.*

5. In 1974 Szemerédi [83] proved the $k = 4$ case of The Conjecture with a purely combinatorial proof. This result *did not* lead to better bounds on $W(4, c)$. Even though it is purely combinatorial, it is rather difficult.

6. In 1975 Szemerédi [84] proved The Conjecture with a purely combinatorial proof. His result *did not* lead to better bounds on $W(k, c)$ because the proof used VDW. Even though it is purely combinatorial, it is rather difficult. In order to prove this he first proved *Szemerédi's Regularity Lemma* which has been very useful in a variety of fields [47, 48, 79].

7. In 1977 Fürstenberg [26] proved The Conjecture with ergodic methods. His proof *did not appear* to lead to *any* bounds on $W(k, c)$ since it was nonconstructive. Avigad and Towsner [2] (see also [31, 32, 45, 1]) have shown that, in principle, one can extract bounds from the proof; however, these bounds are no better than the classic bounds and may be worse.

8. In 1988 Shelah [76] obtained a new proof of VDW that yielded primitive recursive bounds on $W(k, c)$. The bounds are still quite large and cannot be written down. The proof is purely combinatorial and does not use any of the techniques related to The Conjecture.

9. In 1996 Bergelson and Leibman [7] used ergodic techniques to prove the following generalization of The Conjecture:

Let $p_1, \ldots, p_k \in \mathbb{Z}[x]$ such that $(\forall i)[p_i(0) = 0]$. If A is a set of positive upper density then there exists $a, d \in \mathbb{N}$ such that $a, a + p_1(d), a + p_2(d), \ldots, a + p_k(d) \in A$.

An easy corollary is POLYVDW which we restate here:

For any polynomials $p_1(x), \ldots, p_k(x) \in \mathbb{Z}[x]$ such that $(\forall i)[p_i(0) = 0]$, for any natural number c, there exists $W = W(p_1, \ldots, p_k; c)$ such that, for any c-coloring $\chi{:}[W] \rightarrow [c]$ there exists $a, d \in \mathbb{N}$ such that $\chi(a) = \chi(a + p_1(d)) = \chi(a + p_2(d)) = \cdots = \chi(a + p_k(d))$.

Their proof *did not appear* to lead to *any* bounds on W since it was nonconstructive. Towsner [87] showed that, in principle, one can extract bounds from the proof; however, these bounds are no better than the classic bounds and may be worse.

10. In 2000 Walters [90] obtained a proof of POLYVDW that yielded bounds $W(p_1, \ldots, p_k; c)$. these bounds were not primitive recursive.

11. In 2001 Gowers [34, 35] proved The Conjecture using Fourier methods. His proof *did* yield better bounds on $W(k, c)$. In particular he obtains

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

12. In 2002 Shelah [75] obtained a proof of POLYVDW that yielded primitive recursive bounds on $W(p_1, \ldots, p_k; c)$.

13. In 2006 Graham and Solymosi [37] obtained a purely combinatorial proof that $W(3, c) \leq 2^{2^{2^{O(c)}}}$.

This is currently the best-known bound on $W(k, c)$. Note that this bound can actually be written down, unlike the original proof or Shelah's proof.

Exercise 15 Let $p(k, c)$ be the bound on $W(k, c)$ that comes out of the original proof of VDW (the proof we gave of Theorem 2.2.8). Give upper and lower bounds on $p(k, c)$ using Ackermann's function.

Exercise 16 Let $n \in \mathbb{N}$. Think of n as being large. Only use elementary methods in solving this problem. (Also, you cannot use any of the results in the list above.)

1. Show that if $A \subseteq [n]$ and $|A| > 2n/3$ then A has a 3-AP.
2. (Open-ended) Find the smallest α you can such that you can prove, using elementary methods and a computer program, that if $|A| > \alpha n$ then A has a 3-AP.
3. (Open-ended) Fix k. Find the smallest α you can such that you can prove, using elementary methods and a computer program, that if $|A| > \alpha n$ then A has a k-AP.

Exercise 17

1. Use Roth's theorem to show that $W(3,c) \le 2^{2^{O(c)}}$.
2. Bourgain [10] showed that $A \subseteq [n]$ and $|A| \ge \Omega(n\sqrt{\frac{\log\log n}{\log n}})$ then A has a 3-AP. Use this to obtain a bound on $W(3,c)$ that is better than the one just found above.

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

Theorem 2.3.5 *For all k, c, $W(k,c) \ge \sqrt{k-1}c^{(k-1)/2}$.*

Proof:

We will derive the result as we prove the theorem. 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 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 is formed?

First pick the color of the progression. There are c options. Then pick the value of the first point a. There are at most W options. Then pick the value of the difference d. There are at most $W/(k-1)$ options. 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 $cW^2c^{W-k}/(k-1)$.

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

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

If this number is less than 1, then the probability of no monochromatic k-AP must be positive, so such a coloring must exist. Hence we need

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

meaning

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

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

Note that the proof of Theorem 2.3.5 is nonconstructive in that a coloring is not produced. Is there a constructive proof for this bound? This depends on how you define *constructive*. Are there better lower bounds? Yes. Gasarch and Haeupler [30] have a survey of lower bounds on van der Waerden numbers that also clarifies the issues surrounding non-constructive proofs.

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

2.3.4 Three-free sets

2.3.5 Some exact values for $W(k,c)$

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 [13]
$W(3,4)$	76	Brown [11]
$W(4,2)$	35	Chvátal [13]
$W(4,3)$	293	Kouril (personal communication)
$W(5,2)$	178	Stevens and Shantarum [80]
$W(6,2)$	1132	Kouril [50]

The results above *were not* obtained by brute force search over all possible colorings. That would take too much time. We briefly discuss how Kouril and Paul [50] obtained $W(6,2) = 1132$. We note that Kouril and Franco [49] obtained $W(6,2) \le 1131$.

Def 2.3.6 Let $c, n, L \in \mathbb{N}$. Let χ be a c-coloring of $[n]$. Let $b_1, \ldots, b_L \in [c]$. We say that χ *has the pattern* $b_1 b_2 \cdots b_L$ if there exists an x such that $\chi(x)\chi(x+1)\cdots\chi(x+L-1) = b_1 b_2 \cdots b_L$.

Def 2.3.7 A 2-coloring of $[n]$ with no monochromatic 6-AP is called a *good coloring.*

The first step in proving that $W(6,2) = 1132$ is to find some unavoidable patterns.

1. They showed that any good coloring of [240] has either the pattern 0000 or 1111. By symmetry they assume its 0000.
2. They found extensions of 0000 that were unavoidable. They stopped when they had 2,537,546 unavoidable patters of length either 28 or 29. Several clever heuristics were used; however, there approach did yield *all* such unavoidable patterns of these lengths.
3. Which patterns can be in the middle of a good coloring of [240]? For each pattern p they formulated as a SAT question: Is there a good coloring of [240] with p in the middle portion?
4. Using a SAT-solver they found all solutions to this formula. There were only 111. Hence there are patterns $p_1 l, \ldots, p_{111}$ such that any good coloring of [240] has one of these patterns as the middle. Hence the same is true for any good colorings of [326].
5. Using a SAT-solver see for which of the 111 patterns p is there a good coloring of [326] with p in the middle. There are only 52 such patterns. This yields 648,005 good colorings of [326].
6. For each of the 648,005 possible good coloring of [326] p, for all $1 \le i \le 806$ find all good colorings where the pattern p is the coloring for the numbers between i and $i+325$. There are only 3552 of them. Note that these are all possible good colorings of [1031].

7. Test if any of these 3552 good colorings of [1031] can be extended to good colorings of [1032]. They cannot.

The above sequence (when actually carried out) shows that there is a good coloring of [1331] but not of [1332]. Hence $W(6,2) = 1132$.

Exercise 18

1. Use techniques similar to that above to obtain some of the van der Warden Numbers in Table For $W(2,3)$ do not use a computer. For the rest use it sparingly.

2. (Open) Find some new van der Waerden Numbers.

3. (Open) Let $1 \le i < j$. For small values of i, j, c find the least U such that, for any c-coloring of $[U]$, there exists a, d such that $a, a+id, a+jd$ are the same color.

Chapter 3

The Square Theorem

Chapter 4

Coloring and Equations: Rado's Theorem

4.1 Introduction

VDW Theorem with $k = 4$ can be stated as follows:

For all c, for all c-colorings χ:$\mathbb{N} \rightarrow [c]$, 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 χ:$\mathbb{N} \rightarrow [c]$, there exists distinct 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:

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

Let A be the matrix:

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

VDW for $k = 4$ can be rewritten as

For all c, for all c-colorings χ:$\mathbb{N} \rightarrow [c]$ *there exist distinct* e_1, e_2, e_3, e_4 *such that*

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

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

where $\vec{e} = (e_1, e_2, e_3, e_4)$.

What other matrices have this property? We drop the requirement that the e_i be distinct for now.

Def 4.1.1

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* χ:$[R] \rightarrow [c]$ *there exist* $e_1, \ldots, e_n \in [R]$ *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 the following holds: *For all c, there exists* $R = R(A; c)$ *such that for all c-colorings* χ:$[R] \rightarrow [c]$ *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.)

In Section 4.2 we will prove the *Single Equation Rado Theorem* which gives an exact condition for single equations to be regular. In Section 4.3 we will prove the *Full Rado Theorem* which gives an exact condition for matrices to be regular.

Exercise 19 Write down the matrix that represents the theorem VDW$(5,2)$.

Exercise 20 Write a computer program that will, given k, c, output the matrix that corresponds to VDW(k,c).

Exercise 21

a) Find a number X such that

- For every 2-coloring of $[X]$. there exists monochromatic e_1, e_2, e_3 such that $e_1 + e_2 = e_3$. (The e_i need not be distinct.)
- There is a 2-coloring of $[X-1]$ such that there is no monochromatic e_1, e_2, e_3 (not necessarily distinct) with $e_1 + e_2 = e_3$.

b) Do Part a of with the extra condition that e_1, e_2, e_3 are all distinct.

4.2 The Single Equation Rado Theorem

We will prove two theorems, Theorem 4.2.7 and Theorem 4.2.8 that when combined yield *The Single Equation Rado Theorem.* This theorem was first proven by Rado [68, 69]; however, see the historical notes at the end of chapter 5 for more details.

Notation 4.2.1 $\mathbb{Z}_{\neq 0}$ is $\mathbb{Z} \setminus \{0\}$. Hence $\mathbb{Z}^n_{\neq 0}$ is the set of n-vectors of nonzero integers.

Theorem 4.2.2 $(b_1, \ldots, b_n) \in \mathbb{Z}^n_{\neq 0}$ *is regular if and only if some nonempty subset of* $\{b_1, \ldots, b_n\}$ *sums to 0.*

In the proof of Theorem 4.2.2 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 4.2.3 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 χ:$\mathbb{N} \rightarrow [c]$ there exists $e_1, \ldots, e_n$, all distinct, such that*

$$\chi(e_1) = \cdots = \chi(e_n), \text{ and}$$

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

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 essentially the only exceptions.

We will prove the following

Theorem 4.2.4 *If $(b_1, b_2, \ldots, b_n) \in \mathbb{Z}^n_{\neq 0}$ is regular and there exists $\lambda_1, \ldots, \lambda_n \neq \vec{0}$, all distinct, 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:

4.2.1 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:

We define a 17-coloring χ:$\mathbb{N} \rightarrow \{0, \ldots, 16\}$.

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

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:

We define a 16-coloring $\chi : \mathbb{N} \to [16]$. Assume that $n = 17^i n'$ where 17 does not divide n'.

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

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, $e_3 = 17^k e_3'$.
4. $e_1' \equiv e_2' \equiv e_3' \equiv e \pmod{17}$.

If

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

then

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

To use what we know, we must cancel some 17's and exploit the congruence of e_1', e_2', e_3' mod 17. We must explore several cases. All congruences below are mod 17.

1. $i < j$ and $i < k$: We have

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

   Divide by 17^i to get

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

   We take this equation mod 17 to be left with

   $$2e_1' \equiv 2e \equiv 0.$$

   Since 2 has an inverse mod 17 we have $e = 0$. This contradicts that $e \neq 0$.

2. $i = j < k$: We have

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

   Divide by 17^i to get

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

   We take this equation mod 17 to be left with

   $$2e_1' + 5e_2' \equiv 7e \equiv 0.$$

   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 cases, including those above.

   (a) $i < j$ and $i < k$: $2e \equiv 0$.

   (b) $j < i$ and $j < k$: $5e \equiv 0$.

   (c) $k < i$ and $k < j$: $-e \equiv 0$.

   (d) $i = j < k$: $2e + 5e \equiv 0$.

   (e) $i = k < j$: $2e - e \equiv 0$.

   (f) $j = k < i$: $5e - e \equiv 0$.

   (g) $i = j = k$: $2e + 5e - e \equiv 0$.

There are $7 = 2^3 - 1$ cases, each corresponding to a non-empty combination of the coefficients. The key is that every combination was relatively prime to 17.

We now use the ideas in the above example to prove a theorem.

Def 4.2.5 Fix a prime p. For every $n \in \mathbb{N}$, write $n = n'p^k$ where $p \not| n'$. Let χ_p:$\mathbb{N} \rightarrow [p-1]$ be the coloring so that $\chi_p(n) \equiv n' \pmod{p}$. Note that the color is never 0.

Theorem 4.2.6 *Let* $(b_1, \ldots, b_n) \in \mathbb{Z}^n_{\neq 0}$. *If* $(b_1, \ldots, b_n)$ *is regular than there exists* $I \subseteq [n]$, $I \neq \emptyset$ *such that* $\sum_{i \in I} b_i = 0$.

Proof:

We will use χ_p from Definition 4.2.5, for some prime p to be selected later.

By regularity, we know that there exists $x_1, \ldots, x_n \in \mathbb{N}$ and $e \in [p-1]$ such that

$$\chi_p(x_1) = \cdots = \chi_p(x_n) = e$$

and

$$b_1x_1 + \cdots + b_nx_n = \vec{0}. \tag{4.1}$$

By dividing $\vec{x}$ by p as needed, we may assume that some x_i is not divisible by p.

Then we know

$$\begin{aligned} x_1 &= e_1 p^{k_1} \\ x_2 &= e_2 p^{k_2} \\ &\vdots \\ x_n &= e_n p^{k_n}, \end{aligned}$$

where $e_1 \equiv e_2 \equiv \cdots \equiv e_n \equiv e \pmod{p}$. Note that $e \neq 0$.

$I_0 = \{i \mid k_i = 0\}$, so that $i \in I_0$ if any only if $p \not| \, x_i$. Since we assumed that some x_i is not divisible by p, I_0 is nonempty.

Reducing the Equation 4.1 modulo p, we get

$$\sum_{i \notin I_0} b_i e_i p^{k_i} + \sum_{i \in I_0} b_i e_i \equiv 0 \pmod{p}.$$

Since, for all $i \notin I_0$, $k_i \geq 1$, and, for all i, $e_i \equiv e \pmod{p}$ we have

$$e \sum_{i \in I_0}^{m} b_i \equiv 0 \pmod{p},$$

Since e is relatively prime to p, this implies that

$$\sum_{i=1}^{m} b_i \equiv 0 \pmod{p}.$$

We would like to say that $\sum_{i=1}^m b_i$ is *equal* to 0, regardless of p. To get this, we can pick $p > |\sum_{i=1}^m b_i|$.

However, we do not know ahead of time which subset of the b_i's we will be looking at (recall that we renumbered the indices for convenience). Thus we should actually pick p so that

$$p > \max_{J \subseteq [n]} \left\{ \left| \sum_{i \in J} b_i \right| \right\}.$$

In particular, it suffices to pick $p \geq 1 + \sum_{i=1}^n |b_i|$. ∎

Exercise 22 In the proof of Theorem 4.2.6 we took $p > \left|\sum_{i \in J} b_i\right|$ for every nonempty $J \in [n]$. In fact, this is a large overestimate in some cases. Using the same argument as above, find a function $f(b_1, \ldots, b_n)$ which is (much) smaller than this bound infinitely often, but still has the same property: For all $(b_1, \ldots, b_n)$ with no subset adding to zero, and all primes $p \geq f(b_1, \ldots, b_n)$, χ_p has no monochromatic solutions to $\sum_{i=1}^n b_i x_i = 0$.

Theorem 4.2.7 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n$. If all nonempty subsets of $\{b_1, \ldots, b_n\}$ have a non-zero sum then $(b_1, \ldots, b_n)$ is not regular.*

Proof: For every $I \subseteq \{1, \ldots, n\}$ let $S_I = \sum_{i \in I} b_i$. By hypothesis none of these are 0. Hence there exists a number c that is not a multiple of any S_I. By Theorem 4.2.6 there is a $(c-1)$-coloring of $\mathbb{N}$ with no monochromatic solutions. ▮

4.2.2 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 4.2.7 then some nonempty subset of $\{b_1, \ldots, b_n\}$ sums to 0. It turns out this is enough.

Theorem 4.2.8 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n_{\neq 0}$. Assume a nonempty subset of $\{b_1, \ldots, b_n\}$ sums to 0. Then $(b_1, \ldots, b_n)$ is regular.*

Before proving this theorem we do an example. Consider the equation

$$5e_1 + 6e_2 - 11e_3 + 7e_4 - 2e_5 = 0.$$

Note that the first three coefficients add to 0: $5+6-11 = 0$. We are thinking about colorings. 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 χ:$\mathbb{N} \rightarrow [c]$ 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

$$5e_1 + 6e_2 - 11e_3 + 7e_4 - 2e_5 =$$

$$5(a + x_1 d) + 6(a + x_2 d) - 11(a + x_3 d) + 7(a + x_4 d) - 2(a + x_5 d) =$$

$$(5+6-11)a+d(5x_1+6x_2-11x_3)+(7-2)a+d(7x_4-2x_5)=$$

$$(5+6-11)a+d(5x_1+6x_2-11x_3+7x_4-2x_5)+5a.$$

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 such that

$$5x_1+6x_2-11x_3+7x_4-2x_5=0.$$

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 4.2.9 *For all k, s, c, there exists $U = EW(k,s,c)$*[1] *such that for every c-coloring χ:$[U] \to [c]$ 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,

$$EW(k,s,1)=\max\{k,s\}.$$

We assume $EW(k,s,c-1)$ exists and show that $EW(k,s,c)$ exists. We will show that

$$EW(k,s,c) \le W((k-1)sEW(k,s,c-1)+1,c).$$

Let χ be a c-coloring of $[W((k-1)sEW(k,s,c-1)+1,c)]$. By the definition of W there exists a, d such that

$$\chi(a)=\chi(a+d)=\cdots=\chi(a+(k-1)sEW(k,s,c-1)d).$$

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.

[1] The name EW in Lemma 4.2.9 stands for *Extended van der Waerden*

Case 2: If $2sd$ is RED then, since $a, a+2d, a+4d, \ldots, a+2(k-1)d$ are all RED, we are done.

$\vdots$

Case EW(k,s,c-1): If $EW(k,s,c-1)sd$ is RED then, since a,$a+EW(k,s,c-1)d$,$a+2EW(k,s,c-1)d$,. . ., $a+(k-1)EW(k,s,c-1)d$ are all RED, we are done.

Case EW(k,s,c-1)+1: None of the above cases happen. Hence $sd, 2sd, 3sd, \ldots, EW(k,s,c-1)sd$ are all *not* RED.

Consider the c-coloring χ':$[EW(k,s,c-1)] \rightarrow [c]$ defined by

$$\chi'(x) = \chi(xsd).$$

The key is that *none* of these will be colored RED, so there are only $c-1$ colors (even though the colors are taken from $[c]$). 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

$$\chi(A) = \chi(A+D) = \cdots = \chi(A+(k-1)D) = \chi(sD).$$

∎

NEED FIGURE CHANGE: insert a picture illustrating this proof.

We state without proof an easy corollary of Lemma 4.2.9 which we will use in the next theorem.

Corollary 4.2.10 *For all $x_1, \ldots, x_m \in \mathbb{Z}$, s, c, there exists $U = EW(x_1, \ldots, x_m; s, c)$ such that for every c-coloring χ:$[U] \rightarrow [c]$ there exists a, d such that*

$$\chi(a + x_1 d) = \cdots = \chi(a + x_m d) = \chi(sd).$$

Back to the Proof

We now restate and prove the main theorem of this section.

Theorem 4.2.11 *Let $(b_1, \ldots, b_n) \in \mathbb{Z}^n_{\neq 0}$. If there is a nonempty subset of $\{b_1, \ldots, b_n\}$ that sums to 0 then $(b_1, \ldots, b_n)$ is regular.*

Proof: By renumbering, we can assume that there is an $m \le n$ such that

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

We need to find a number R such that any c-coloring of $[R]$ gives positive integers $e_1, \ldots, e_n$ all the same color with $b_1e_1 + \cdots + b_ne_n = 0$. We will show that $R = EW(x_1, \ldots, x_m; s, c)$ from Lemma 4.2.10 works, for values of $x_1, \ldots, x_m$ and s to be found later.

Let χ be a c-coloring of $[R]$. By choice of R, there exists a, d such that

$$\chi(a + x_1d) = \chi(a + x_2d) = \cdots = \chi(a + x_md) = \chi(sd).$$

We will let

$$\begin{aligned} e_1 &= a + x_1d, \\ e_2 &= a + x_2d, \\ &\vdots \\ e_m &= a + x_md, \end{aligned}$$

and

$$e_{m+1} = \cdots = e_n = sd.$$

Then

$$\sum_{i=1}^{n} b_ie_i = \sum_{i=1}^{m} b_ie_i + \sum_{i=m+1}^{n} b_ie_i = \sum_{i=1}^{m} b_i(a + x_id) + \sum_{i=m+1}^{n} b_isd.$$

This is equal to

$$a\sum_{i=1}^{m} b_i + d\sum_{i=1}^{m} b_ix_i + sd\sum_{i=m+1}^{n} b_i.$$

Here we recall that $\sum_{i=1}^{m} b_i = 0$, so the first term drops out. Hence we need $x_1, \ldots, x_m \in \mathbb{Z}_{\neq 0}$ and $s \in \mathbb{N}$ such that

$$d\sum_{i=1}^{m} b_i x_i + sd \sum_{i=m+1}^{n} b_i = 0.$$

Divide by d to get

$$\sum_{i=1}^{m} b_i x_i + s \sum_{i=m+1}^{n} b_i = 0.$$

Let $\sum_{i=m+1}^{n} b_i = B$. We rewrite this as

$$\sum_{i=1}^{m} b_i x_i + sB = 0.$$

We can take

$$s = m \cdot |\mathrm{lcm}\{b_1, \ldots, b_m\}|$$

$$\begin{aligned} x_1 &= -\frac{sB}{mb_1} \\ x_2 &= -\frac{sB}{mb_2} \\ &\vdots \\ x_m &= -\frac{sB}{mb_m}. \end{aligned}$$

∎

Exercise 23 Schur's Theorem is as follows: For all c there exists $s = s(c)$ such that for all c-colorings of $[s]$ there exists monochromatic x, y, z such that $x+y = z$. Schur's Theorem follows from the Single Equation Rado Theorem. Let $s(c)$ be the smallest value of n such that, for any c-coloring of $[n]$ there exists x, y, z the same color such that $x + y = z$.

a) Use Rado's Theorem to obtain upper bounds on $s(c)$. Try to determine this bound explicitly for some small values of c.

b) Find a statement of Ramsey's Theorem for graphs.

c) Use Ramsey's Theorem to prove Schur's Theorem.

d) Use this proof to obtain upper bounds on $s(c)$. Try to determine this bound explicitly for some small values of c.

e) (Speculative) Which proof gave better bounds?

f) (Open-ended) Obtain better bounds for $s(c)$.

4.2.3 If ... then $(b_1, \ldots, b_n)$ is distinct-regular

We will prove the following theorem due to Rado [69, 68].

Theorem 4.2.12 *If $(b_1, b_2, \ldots, b_n)$ is regular and there exists $\lambda_1, \ldots, \lambda_n$ distinct 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 Full Rado Theorem.

The following definitions are used in the third part of the lemma.

Def 4.2.13 Let $n \in \mathbb{N}$.

1. A set $G \subseteq \mathbb{N}^n$ is *homogeneous* 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 4.2.14

1. Let $G = \{(a, a+d, \ldots, a+(k-1)d) \mid a, d \in \mathbb{N}\}$ be the set of k-APs in $\mathbb{N}$. G is homogeneous. By VDW, G is also regular.

2. 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 homogeneous. G is regular if and only if $(b_1, \ldots, b_n)$ is.

3. Let A be an $m \times n$ matrix. Let $G = \{\vec{e} \mid A\vec{e} = \vec{0}\}$. G is homogeneous. G is regular if and only if M is.

Lemma 4.2.15

1. *For all $(b_1, \ldots, b_n) \in \mathbb{Z}^n$ regular, for all $c, M \in \mathbb{N}$, there exists $L = L(b_1, \ldots, b_n; c, M)$ with the following property. For any c-coloring χ: $[L] \to [c]$ there exists $e_1, \ldots, e_n, d \in [L]$ such that the following hold.*

   (a) $b_1e_1 + \cdots + b_ne_n = 0$.

   (b) All of these numbers have the same color:

$$\begin{array}{ccccccc} e_1 - Md, & \ldots, & e_1 - d, & e_1, & e_1 + d, & \ldots, & e_1 + Md \\ e_2 - Md, & \ldots, & e_2 - d, & e_2, & e_2 + d, & \ldots, & e_2 + Md \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ e_n - Md, & \ldots, & e_n - d, & e_n, & e_n + d, & \ldots, & e_n + Md. \end{array}$$

2. *For all $(b_1, \ldots, b_n) \in \mathbb{Z}^n$ regular, for all $c, M, s \in \mathbb{N}$, there exists $L_2 = L_2(b_1, \ldots, b_n; c, M, s)$ with the following property. For any c-coloring $\chi\colon [L_2] \to [c]$ there exists $e_1, \ldots, e_n, d \in [L_2]$ such that the following hold.*

   (a) $b_1e_1 + \cdots + b_ne_n = 0$.

   (b) All of these numbers have the same color:

$$\begin{array}{ccccccc} e_1 - Md, & \ldots, & e_1 - d, & e_1, & e_1 + d, & \ldots, & e_1 + Md \\ e_2 - Md, & \ldots, & e_2 - d, & e_2, & e_2 + d, & \ldots, & e_2 + Md \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ e_n - Md, & \ldots, & e_n - d, & e_n, & e_n + d, & \ldots, & e_n + Md \\ & & & sd. & & & \end{array}$$

3. *For all $n \in \mathbb{N}$, for all $G \subseteq \mathbb{N}^n$, G regular and homogeneous, for all $c, M, s \in \mathbb{N}$ there exists $L_3 = L_3(G; c, M, s)$ with the following property. For any c-coloring $\chi\colon [L_3] \to [c]$ there exists $e_1, \ldots, e_n, d \in [L_3]$ such that the following hold.*

   (a) $(e_1, \ldots, e_n) \in G$.

(b) All of these numbers have the same color:

$$\begin{array}{cccccccc}
e_1 - Md, & \ldots, & e_1 - d, & e_1, & e_1 + d, & \ldots, & e_1 + Md \\
e_2 - Md, & \ldots, & e_2 - d, & e_2, & e_2 + d, & \ldots, & e_2 + Md \\
\vdots & & \vdots & \vdots & \vdots & & \vdots \\
e_n - Md, & \ldots, & e_n - d, & e_n, & e_n + d, & \ldots, & e_n + Md \\
 & & & sd. & & &
\end{array}$$

Proof: (Part 1)
Since $b_1, \ldots, b_n$ is regular, by Definition 4.1.1 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 will choose the desired number L later. Throughout the proof we will add conditions to L. The first one is that R divides L.

Let χ:$[L] \to [c]$ be a coloring.

We want to show that the conclusion of the theorem holds for χ.

We define a new coloring χ^*:$[L/R] \to [c]^R$ as follows:

$$\chi^*(n) = (\chi(n), \chi(2n), \chi(3n), \ldots, \chi(Rn)).$$

In order to find an arithmetic progression, we will pick L so that $L/R \geq W(2X+1, c^R)$. We will determine X later.

Apply (a slight variant of) VDW to the c^R-coloring χ to obtain the following: There exists a, D (but not our desired d) such that

$$\chi^*(a - XD) = \chi^*(a - (X-1)D) = \cdots = \chi^*(a) = \cdots = \chi^*(a + XD).$$

Since we know

$$\chi^*(n) = (\chi(n), \chi(2n), \ldots, \chi(Rn)),$$

this gives us

$$\begin{array}{ccccccccccc}
\chi(a - XD) & = & \chi(a - (X-1)D) & = & \cdots & = & \chi(a) & = & \cdots & = & \chi(a + XD) \\
\chi(2(a - XD)) & = & \chi(2(a - (X-1)D)) & = & \cdots & = & \chi(2a) & = & \cdots & = & \chi(2(a + XD)) \\
\chi(3(a - XD)) & = & \chi(3(a - (X-1)D)) & = & \cdots & = & \chi(3a) & = & \cdots & = & \chi(3(a + XD)) \\
\vdots & = & \vdots & = & \cdots & = & \vdots & = & \cdots & = & \vdots \\
\chi(R(a - XD)) & = & \chi(R(a - (X-1)D)) & = & \cdots & = & \chi(Ra) & = & \cdots & = & \chi(R(a + XD)).
\end{array}$$

We need a subset of these that are all the same color. Consider the coloring χ^{**}:$[R] \rightarrow [c]$ 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:

$$\begin{array}{ccccc}
(a-XD)f_1, & (a-(X-1)D)f_1, & \cdots, & af_1, & \cdots, & (a+XD)f_1 \\
(a-XD)f_2, & (a-(X-1)D)f_2, & \cdots, & af_2, & \cdots, & (a+XD)f_2 \\
(a-XD)f_3, & (a-(X-1)D)f_3, & \cdots, & af_3, & \cdots, & (a+XD)f_3 \\
\vdots & \vdots & & \vdots & & \vdots \\
(a-XD)f_n, & (a-(X-1)D)f_n, & \cdots, & af_n, & \cdots, & (a+XD)f_n.
\end{array}$$

For all i, $1 \le i \le n$ let $e_i = af_i$. We rewrite the above:

$$\begin{array}{cccccc}
e_1 - f_1XD, & e_1 - f_1(X-1)D, & \cdots, & e_1, & \cdots, & e_1 + f_1XD \\
e_2 - f_2XD, & e_2 - f_2(X-1)D, & \cdots, & e_2, & \cdots, & e_2 + f_2XD \\
e_3 - f_3XD, & e_3 - f_3(X-1)D, & \cdots, & e_3, & \cdots, & e_3 + f_3XD \\
\vdots & \vdots & & \vdots & & \vdots \\
e_n - f_nXD, & e_n - f_n(X-1)D, & \cdots, & e_n, & \cdots, & e_n + f_nXD.
\end{array}$$

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 just 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, but we haven't chosen X or d yet. We know that, for all i, $f_i \le R$. Clearly $d = f_1f_2 \cdots f_nD \le R^nD$ is a sensible choice, so we use that.

We need, for every $1 \le i \le n$,

$$\left\{ \left(\prod_{j=1}^{n} f_i\right) D, 2\left(\prod_{j=1}^{n} f_i\right) D, \ldots, M\left(\prod_{j=1}^{n} f_i\right) D \right\} \subseteq \{f_iD, 2f_iD, \ldots, Xf_iD\}.$$

Equivalently, we need

$$\left\{ \left(\prod_{j=1}^{n} f_i\right), 2\left(\prod_{j=1}^{n} f_i\right), \ldots, M\left(\prod_{j=1}^{n} f_i\right) \right\} \subseteq \{f_i, 2f_i, \ldots, Xf_i\}.$$

Taking $X = MR^{n-1}$ will suffice.

Since we have $X = R^{n-1}M$, we now know our bound for L:

$$L = R \cdot W(2R^{n-1}M + 1, c^R), \text{ where } R = R(b_1, \ldots, b_n; c).$$

(Part 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 $\sum_{i=1}^{n} b_ie_i = 0$ such that $e_{\min} = \min\{e_1, \ldots, e_n\} > M$. Let $e_{\max} = \max\{e_1, \ldots, e_n\} > M$. Let $L_2 = L_2(b_1, \ldots, b_n; 1, M, s) = \max\{e_{\max} + M, s\}$. Let $\chi{:}[L_2] \to [1]$. We claim that $e_1, \ldots, e_n, 1$ work. Note that, for all $i \in [n]$ and $j \in \{-M, \ldots, M\}$, we have $e_i + j \times 1 \in [L_2]$. Also note that $s \times 1 \in [L_2]$. Thus, taking $d = 1$, we have our solution.

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 4.2.9.

Induction Step: We want to show that $L_2(b_1, \ldots, b_n; c, M, s)$ exists. We show that there is M' so that, if you c-color $[L]$ (where $L = L(b_1, \ldots, b_n; c, M')$ from part 1), then there exists the required $e_1, \ldots, e_n, d$. The M' will depend on L_2 for $c-1$ colors. 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.

$$\begin{array}{ccccccc}
E_1 - M'D, & \ldots, & E_1 - D, & E_1, & E_1 + D, & \ldots, & E_1 + M'D \\
E_2 - M'D, & \ldots, & E_2 - D, & E_2, & E_2 + D, & \ldots, & E_2 + M'D \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
E_n - M'D, & \ldots, & E_n - D, & E_n, & E_n + D, & \ldots, & E_n + M'D.
\end{array}$$

There are now several cases.

Case 1: If sD is RED then we are done so long as $M' \ge M$. Use $d = D$.

Case 2: If $2sD$ is RED then we are done so long as $M' \ge 2M$. Use $d = 2D$.

$\vdots$

Case X: If XsD is RED then so long as $M' \ge MX$ we are done. Use $d = XD$.

Case X+1: None of the above cases hold. Hence

$$sD, 2sD, \ldots, XsD$$

are all *not* RED. Hence the coloring restricted to this set is a $c-1$ coloring. Let $X = L_2(b_1, \ldots, b_n; c-1, M, s)$, and $M' = MX$. Consider the $(c-1)$-coloring χ^* of $[M']$ 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 under χ^*:

$$\begin{array}{cccccc}
e_1 - Md, & e_1 - (M-1)d, & \ldots, & e_1, & \ldots, & e_1 + Md \\
e_2 - Md, & e_2 - (M-1)d, & \ldots, & e_2, & \ldots, & e_2 + Md \\
\vdots & \vdots & & \vdots & & \vdots \\
e_n - Md, & e_n - (M-1)d, & \ldots, & e_n, & \ldots, & e_n + Md
\end{array}$$

$$sd.$$

By the definition of χ^*, the following have the same color via χ:

$$\begin{array}{ccccccc}
(e_1 - Md)sD, & (e_1 - (M-1)d)sD, & \ldots, & e_1sD, & \ldots, & (e_1 + Md)sD \\
(e_2 - Md)sD, & (e_2 - (M-1)d)sD, & \ldots, & e_2sD, & \ldots, & (e_2 + Md)sD \\
\vdots & \vdots & & \vdots & & \vdots \\
(e_n - Md)sD, & (e_n - (M-1)d)sD, & \ldots, & e_nsD, & \ldots, & (e_n + Md)sD
\end{array}$$

$$sdsD.$$

By taking the vector $(e_1sD, \ldots, e_nsD)$ and common difference sdD, we obtain the result.

(Part 3)
In both of the above parts, the only property of the set

$$\left\{ (x_1, \ldots, x_n) \;\middle|\; \sum_{i=1}^{n} b_i x_i = 0 \right\}$$

that we used is that it was homogeneous 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 4.2.16 *If $(b_1, \ldots, b_n)$ 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 part 1 of Lemma 4.2.15. Let χ be a c-coloring of $[L]$. We know that there exists $e_1, \ldots, e_n, d \in [L]$ such that the following occur.

1. $b_1e_1 + \cdots + b_ne_n = 0$.

2. The following are the same color:

$$\begin{array}{ccccccc}
e_1 - Md, & \ldots, & e_1 - d, & e_1, & e_1 + d, & \ldots, & e_1 + Md \\
e_2 - Md, & \ldots, & e_2 - d, & e_2, & e_2 + d, & \ldots, & e_2 + Md \\
\vdots & \vdots & & \vdots & \vdots & & \vdots \\
e_n - Md, & \ldots, & e_n - d, & e_n, & e_n + d, & \ldots, & e_n + Md.
\end{array}$$

Let $A \in \mathbb{Z}$ be a constant to be picked later. Note that

$$\sum_{i=1}^{n} b_i(e_i + Ad\lambda_i) = \left(\sum_{i=1}^{n} b_i e_i\right) + \left(Ad\sum_{i=1}^{n} b_i\lambda_i\right) = 0.$$

Thus $(e_1+Ad\lambda_1, \ldots, e_n+Ad\lambda_n)$ is a solution. For it to be monochromatic, 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, $|A\lambda_i| \le M$.

Since $\lambda_i \neq \lambda_j$, there is at most 1 value of A which makes $e_i + Ad\lambda_i = e_j + Ad\lambda_j$ — viewing this condition as a linear equation in A. Therefore, there are at most $\binom{n}{2}$ values of A which 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 $M = \binom{n}{2}\lambda$. Any choice of A with $|A| \le \binom{n}{2}$ will satisfy condition 2. There are more than $\binom{n}{2}$ values of A that satisfy this, hence we can find a value of A one that satisfies items 1 and 2. ▮

Exercise 24 (Open-ended)

a) Consider the equation $10x_1 + 13x_2 - 40x_3 = 0$. By Theorem 4.2.6 there is a 40-coloring of $\mathbb{N}$ such that there is no monochromatic solution. Exercise 22 gives a 6-coloring with the same property, but we do not know whether it is best. Find the value of c such that

- There is a c-coloring of $\mathbb{N}$ such that $10x_1 + 13x_2 - 40x_3 = 0$ has no monochromatic solution.
- For every $c-1$-coloring of $\mathbb{N}$ there is a monochromatic solution to $10x_1 + 13x_2 - 40x_3 = 0$.

b) We define $(b_1, \ldots, b_n)$ be be *c-regular* if, for every c-coloring of $\mathbb{N}$, there is a monochromatic solution to $\sum_{i=1}^{n} b_i x_i = 0$. Find some condition X such that, for all $(b_1, \ldots, b_n)$ and c, $(b_1, \ldots, b_n)$ is c-regular iff X.

c) Define c-distinct-regular in the analogous way. Repeat the problem above with that notion of c-distinct regular.

4.3 The Full Rado Theorem

Recall what it means for a matrix to be regular:

Def 4.3.1 A matrix A of integers is *regular* if if the following holds: *For all c, there exists $R = R(A;c)$ such that for all c-colorings $\chi\colon[R] \to [c]$ there exists $\vec{e} = (e_1, \ldots, e_n)$ such that*

$$\begin{aligned} \chi(e_1) = \cdots = \chi(e_n), \\ A\vec{e} = \vec{0}. \end{aligned}$$

Def 4.3.2 A matrix A with columns $\vec{a}_1, \ldots, \vec{a}_n$ satisfies the *columns condition* if the set $[n]$ can be partitioned into sets $I_0, \ldots, I_k$ such that

$$\sum_{i \in I_0} \vec{a}_i = \vec{0},$$

and for all $j \in \{1, \ldots, k\}$, $\left(\sum_{i \in I_j} \vec{a}_i\right)$ can be written as a linear combination of the vectors $\{\vec{a}_i \mid i \in I_0 \cup \cdots \cup I_{j-1}\}$.

Example 4.3.3

1. We began the chapter with this matrix corresponding to van der Waerden's Theorem (for 4-APs):

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

The columns add to $\vec{0}$, so we may take $I_0 = \{1, 2, 3, 4\}$.

2. Let

$$A = \begin{pmatrix} 1 & -1 & 0 & 1 & -2 & 0 \\ 2 & 0 & -2 & 1 & 0 & -2 \\ 3 & -3 & 0 & 1 & -4 & 4 \end{pmatrix}.$$

One partition of columns is $I_0 = \{1, 2, 3\}, I_1 = \{4, 5\}, I_2 = \{6\}$. The relevant relations between columns are

$$\begin{aligned} \vec{a}_1 + \vec{a}_2 + \vec{a}_3 &= \vec{0} \\ \vec{a}_4 + \vec{a}_5 &= \tfrac{1}{2}\vec{a}_1 + \tfrac{3}{2}\vec{a}_2 \\ \vec{a}_6 &= -2\vec{a}_2 - 2\vec{a}_4. \end{aligned}$$

We will prove the following:
The Full Rado Theorem:

Theorem 4.3.4 *A is regular if and only if A satisfies the columns condition.*

First we will show the "only if" part. This will use the same coloring approach as used in Section 4.2.1:

4.3.1 If A is regular then ...

For the proof of the Full Rado Theorem we will need to look at matrix equations modulo prime powers and deduce things about the matrix equations without the modulus. The following lemma will be helpful.

Lemma 4.3.5 *Let I be a finite index set.*

1. *Let $\vec{b}, \{\vec{a}_i\}_{i \in I}$ be integer vectors of the same length. There exists a p_0 such that, for all $p \geq p_0$, for all $j \geq 1$, the following holds:*

   If $p^j\vec{b}$ is a linear combination of $\{\vec{a}_i\}_{i\in I}$ modulo p^{j+1} (with coefficients in $\{0,\ldots,p^{j+1}-1\}$), then $\vec{b}$ is a linear combination of $\{\vec{a}_i\}_{i\in I}$ with coefficients in $\mathbb{Q}$.

2. *Let $\{\vec{a}_i\}_{i\in I}$ be integer vectors of the same length. There exists a p_0 such that, for all $p \geq p_0$, for all $j \geq 1$, for all $I' \subseteq I$ the following holds.*

   If $\vec{b} = \sum_{i\in I'} \vec{a}_i$ and if $p^j\vec{b}$ is a linear combination of $\{\vec{a}_i\}_{i\in I-I'}$ modulo p^{j+1} with coefficients in $\{0,\ldots,p^{j+1}-1\}$. then $\vec{b}$ is a linear combination of $\{\vec{a}_i\}_{i\in I}$ with coefficients in $\mathbb{Q}$. (This follows from part 1 by taking the maximum p over all possible $I' \subseteq I$.)

Proof:

Assume, by way of contradiction, that for all p there exists $(f_i)_{i\in I_0}$ such that

$$p^j\vec{b} \equiv \sum_{i\in I} f_i\vec{a}_i \pmod{p^{j+1}}$$

but $\vec{b}$ is not a linear combination of $\{\vec{a}_i\}_{i\in I}$ with coefficients in $\mathbb{Q}$.

Since $\vec{b}$ and all of the $\vec{a}_i$ are vectors over the integers, this means that $\vec{b}$ is not in the span of $\{\vec{a}_i\}_{i\in I}$. By the Gram-Schmidt orthogonalization process

(See Exercise 25) there exists an integer vector $\vec{v}$ that is orthogonal to all of the $\vec{a}_i$ but not to $\vec{b}$. Let $\vec{b} \cdot \vec{v} = t$. We pick $p > |t|$.

Note that

$$p^j \vec{b} \cdot \vec{v} \equiv \sum_{i \in I} f_i \vec{a}_i \cdot \vec{v} \pmod{p^{j+1}}$$

$$p^j t \equiv 0 \pmod{p^{j+1}}$$

Since $p > |t|$ this is a contradiction. ▮

Theorem 4.3.6 *Let A be a matrix. If A is regular then A satisfies the columns condition.*

Proof:

We will use χ_p from Definition 4.2.5, for some prime p to be selected later.

By regularity, we know that there exists $x_1, \ldots, x_n \in \mathbb{N}$ and $e \in [p-1]$ such that

$$\chi_p(x_1) = \cdots = \chi_p(x_n) = e$$

and

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

Rewrite this as

$$x_1\vec{a}_1 + \cdots + x_n\vec{a}_n = \vec{0}. \tag{4.2}$$

By dividing $\vec{x}$ by p as needed, we may assume that some x_i is not divisible by p.

Then we know

$$\begin{array}{rcl} x_1 & = & e_1 p^{k_1} \\ x_2 & = & e_2 p^{k_2} \\ & \vdots & \\ x_n & = & e_n p^{k_n}, \end{array}$$

where $e_1 \equiv e_2 \equiv \cdots \equiv e_n \equiv e \pmod{p}$. Note that $e \neq 0$.

For each $j \geq 0$, let $I_j = \{i \mid k_i = j\}$, so that $i \in I_j$ if any only if $p^j | x_i$, but $p^{j+1} \not| \, x_i$.

Since we assumed that some x_i is not divisible by p, I_0 is nonempty.

Reducing the Equation 4.2 modulo p, we get

$$\sum_{i\notin I_0} e_i\vec{a}_i + \sum_{i\in I_0} e_i\vec{a}_i \equiv \vec{0} \pmod{p}.$$

Since $k_i \geq 1$ for all $i \notin I_0$, and $e_i \equiv e \pmod{p}$ for all i, we have

$$\sum_{i\in I_0} e\vec{a}_i \equiv \vec{0} \pmod{p}.$$

Since e is relatively prime to p we can divide it out to obtain

$$\sum_{i\in I_0} \vec{a}_i \equiv \vec{0} \pmod{p}. \tag{4.3}$$

Recall that we would like to show $\sum_{i\in I_0} \vec{a}_i$ is *equal to* $\vec{0}$, regardless of p.

We will show that $I_0, I_1, \ldots$ is the partition we seek. Note that only a finite number of the I_j will be nonempty.

Now consider I_j for $j > 0$. We will reduce Equation 4.2 modulo p^{j+1}. All of the terms $x_i\vec{a}_i$ where $i \in I_{j+1} \cup I_{j+2} + \cdots$ will vanish. The terms $x_i\vec{a}_i$ where $i \in I_j$ are the most interesting to us. Note that $x_i = e_ip^j$ where $e_i \equiv e \pmod{p}$. Hence there is an $A \in \mathbb{N}$ such that $e_i = e + Ap$. Therefore

$$x_i = e_ip^j = (e+Ap)p^j = ep^j + Ap^{j+1} \equiv ep^j \pmod{p^{j+1}}.$$

With this in mind we obtain

$$\sum_{i\in I_j} e\vec{a}_ip^j + \sum_{i\in I_{j-1}} e_i\vec{a}_ip^{j-1} + \cdots + \sum_{i\in I_0} e_i\vec{a}_i \equiv \vec{0} \pmod{p^{j+1}}.$$

Rearranging this and multiplying by $e^{-1} \pmod{p^{j+1}}$, we get

$$p^j\sum_{i\in I_j}\vec{a}_i \equiv -\sum_{i\in I_{j-1}} e_ie^{-1}\vec{a}_ip^{j-1} - \cdots - \sum_{i\in I_0} e_ie^{-1}\vec{a}_i \pmod{p^{j+1}}.$$

In other words,

$$p^j\sum_{i\in I_j}\vec{a}_i \text{ is a linear combination of } \{\vec{a}_i\}_{I_0\cup\cdots\cup I_{j-1}} \text{ modulo } p^{j+1}. \tag{4.4}$$

Recall that we would like to show that $\sum_{i\in I_j}\vec{a}_i$ is a (rational) linear combination of $\{\vec{a}_i\}_{i\in I_0\cup\cdots\cup I_{j-1}}$ regardless of p.

We claim that Equation 4.3 and Equation 4.4 can each only hold for finitely many primes p, so that picking a prime large enough should give a contradiction. However, the reader must keep in mind that we don't know a priori what the sets I_j will be.

For Equation 4.3 take p to be any prime that is larger than the sum of the absolute values of all of the elements of A. For Equation 4.4 use Lemma 4.3.5.

▮

Exercise 25 (Gram-Schmidt orthogonalization process)

a) Given two vectors $\vec{a}, \vec{b}$, define

$$\vec{c} = \vec{b} - \left(\frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}} \right) \vec{a}$$

Show that $\{\vec{a}, \vec{c}\}$ have the same span as $\{\vec{a}, \vec{b}\}$, and $\vec{c}$ is orthogonal to $\vec{a}$.

b) Given vectors $\vec{a}_1, \ldots, \vec{a}_n$, show how to find $\vec{c}_1, \ldots, \vec{c}_n$ orthogonal with the same span as the original vectors.

c) Use this process to show that, given vectors $\vec{a}_1, \ldots, \vec{a}_n$, and $\vec{b}$ not in their span, there is a vector $\vec{v}$ orthogonal to $\vec{a}_1, \ldots, \vec{a}_n$ but not orthogonal to $\vec{b}$. Hint: consider $\vec{a}_{n+1} = \vec{b}$.

d) Notice that if $\vec{a}_1, \ldots, \vec{a}_n, \vec{b}$ were all integer vectors, then $\vec{v}$ can be taken as an integer vector (after perhaps multiplying by a scalar).

4.3.2 If . . . then A is regular

Def 4.3.7 We say that $S \subseteq \mathbb{Z}$ is an (m, p, c)-set if there are values $y_0, \ldots, y_m > 0$ so that S consists of the values

$$\begin{array}{ccccccccccc}
cy_0 & + & \lambda_1 y_1 & + & \lambda_2 y_2 & + & \cdots & + & \lambda_{m-1} y_{m-1} & + & \lambda_m y_m \\
 & & cy_1 & + & \lambda_2 y_2 & + & \cdots & + & \lambda_{m-1} y_{m-1} & + & \lambda_m y_m \\
 & & & & & & & & & & \vdots \\
 & & & & & & & & cy_{m-1} & + & \lambda_m y_m \\
 & & & & & & & & & & cy_m
\end{array}$$

for every choice of $|\lambda_1| \leq p, \ldots, |\lambda_m| \leq p$.

The values $y_0, \ldots, y_m$ are said to generate S.

Note 4.3.8

- Unfortunately, in this section c refers both to the prescribed coefficient in (m, p, c)-sets and the number of colors in a coloring. We will suppress the number of colors, which is usually irrelevant.
- In this section, p is no longer restricted to being prime.

Example 4.3.9 A $(1, p, c)$-set has the form

$$S = \{cy - pz, \dots, cy - z, cy, cy + z, \dots, cy + pz, cz\}$$

for some choice of y and z. Writing $a = cy$ and $d = z$, this gives us the now-familiar

$$S = \{a - pd, \dots, a - d, a, a + d, \dots, a + pd, cd\},$$

where the value a is restrained to be in $c\mathbb{N}$. Since $c\mathbb{N}$ is a homogeneous set, and trivially regular, part 3 of Lemma 4.2.15 (with $s = c$) tells us that there is an L so that any finite coloring of $[L]$ yields a monochromatic set of this type. That is, the collection of all $(1, p, c)$-sets is regular (for a fixed p and c). This can also be obtained more simply from Lemma 4.2.9.

Note 4.3.10 The definitions of a set G being homogeneous and regular require that $G \subseteq \mathbb{N}^n$ — the elements must be lists, not sets. When we say the collection of all $(1, p, c)$-sets has these properties, we really mean that, if one were to order them (say in increasing order) then the result would be regular. We will ignore this distinction from now on.

Theorem 4.3.11 *For every m, p, c, the collection of (m, p, c)-sets is regular.*

Proof: We will show this by induction on m. The case $m = 1$ was Example 4.3.9.

Suppose we know the collection of (m, p, c)-sets is regular. Since we may scale the $y_1, \dots, y_m$, we see that this collection is also homogeneous. Thus we may apply part 3 of Lemma 4.2.2 with $M = p$ and $s = c$.

Let $y_0, \dots, y_m$ be the generators for the monochromatic (m, p, c)-set found, and let $y_{m+1} = d$ be the common difference given by the application of Lemma 4.2.9. Then the reader may check that $y_0, \dots, y_m, y_{m+1}$ generate a monochromatic $(m + 1, p, c)$-set. ▮

Now we restate and prove our claim.

Theorem 4.3.12 *If A satisfies the columns condition, then it is regular.*

Proof: As usual, let $\vec{a}_1, \ldots, \vec{a}_n$ be the columns of A, and $[n] = I_0 \cup \cdots \cup I_k$ be the partition of indices given by the columns condition. We rewrite these conditions as follows:

$$t_j \sum_{i \in I_j} \vec{a}_i + \sum_{i \in I_0 \cup \cdots \cup I_{j-1}} b_{ij} \vec{a}_i = \vec{0}$$

for some nonzero integers $t_0 = 1, t_1, \ldots, t_k$, and integers $\{b_{ij} \mid i \in I_\ell \text{ for some } \ell < j\}$. By scaling each equation, we may assume $t_0 = \cdots = t_k = c$, which would be the least common multiple of the original $t_0, \ldots, t_k$.

For completion, set

$$b_{ij} = t_j \text{ for all } i \in I_j, \text{ and}$$

$$b_{ij} = 0 \text{ for } i \in I_{j+1} \cup \cdots \cup I_k.$$

Finally, define the matrix $B = (b_{ij})$, so that we may compactly say $AB = (0)$. Note that B is, in a certain sense, lower triangular with diagonal c. That is, if $i \in I_j$, then $b_{ij} = c$ and $b_{i\ell} = 0$ for $\ell > j$.

Since $AB = (0)$, we see that $AB\vec{y} = \vec{0}$ for any $\vec{y}$. This suggests we search for any solution vector of the form $B\vec{y}$. Let $p = \max |b_{ij}|$, and $m = k$, the final index of our partition $I_0 \cup \cdots \cup I_k$.

Let $(y_0, \ldots, y_m)$ be the generators of any (m, p, c)-set S, using the values we have chosen for m, p, c. By definition of an (m, p, c)-set, and by the noted property of the matrix B, S contains all coordinates of the vector $B\vec{y}$. Thus, since the collection of (m, p, c)-sets is regular, so is A. ▮

Example 4.3.13 When finitely coloring $\mathbb{N}$, we would like to find a monochromatic solution to the following system of equations:

$$\begin{array}{rcrcrcrcrcrcl}
x_1 & - & x_2 & & & + & x_4 & - & 2x_5 & & & = & 0 \\
2x_1 & & & - & 2x_3 & + & x_4 & & & - & 2x_6 & = & 0 \\
3x_1 & - & 3x_2 & & & + & x_4 & - & 4x_5 & + & 4x_6 & = & 0.
\end{array}$$

This corresponds to the following matrix:

$$A = \begin{pmatrix} 1 & -1 & 0 & 1 & -2 & 0 \\ 2 & 0 & -2 & 1 & 0 & -2 \\ 3 & -3 & 0 & 1 & -4 & 4 \end{pmatrix}.$$

As we saw in Example 4.3.3, we may partition the columns as $I_0 = \{1, 2, 3\}, I_1 = \{4, 5\}, I_2 = \{6\}$. The needed relations between columns are

$$\begin{aligned} \vec{a}_1 + \vec{a}_2 + \vec{a}_3 &= \vec{0} \\ \vec{a}_4 + \vec{a}_5 &= \tfrac{1}{2}(\vec{a}_1 + 3\vec{a}_2) \\ \vec{a}_6 &= -2(\vec{a}_2 + \vec{a}_4). \end{aligned}$$

After scaling the equations by 2, the corresponding matrix B is

$$B = \begin{pmatrix} 2 & -1 & 0 \\ 2 & -3 & 4 \\ 2 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Note that $AB = (0)_{3\times 3}$.

Here we have $c = 2$, $p = 4$, $m = 2$. If y_0, y_1, y_2 generate a $(2, 4, 1)$-set S, then we see it includes the following values:

$$2y_0 - y_1$$

$$2y_0 - 3y_1 + 4y_2$$

$$2y_0$$

$$2y_1 + 4y_2$$

$$2y_1$$

$$2y_2.$$

These are exactly the coordinates of $B\vec{y}$. Thus, a monochromatic $(2, 4, 1)$-set gives us a monochromatic vector of the form $\vec{x} = B\vec{y}$, which is a monochromatic solution to $A\vec{x} = \vec{0}$.

Finally, without any more work, we see that any regular matrix A with a solution of distinct entries is distinct-regular.

Theorem 4.3.14 *If A is a regular matrix, and $\vec{\lambda}$ is an integer vector with distinct entries so that $A\vec{\lambda} = \vec{0}$, then A is distinct-regular.*

We omit the proof, as it is identical to the proof of Theorem 4.2.16, the analogous theorem which tells that a regular sequence $(b_1, \ldots, b_n)$ with a solution of distinct values $b_1x_1 + \cdots b_nx_n = 0$ is distinct-regular.

Exercise 26 Prove Theorem 4.3.14.

Exercise 27 We prove a subcase of Theorem 4.3.12 that is interesting in its own right, referred to as *Folkman's Theorem* in the literature. Its statement is part 2 of this exercise.

a) Let $n(k, c)$ be the least n (if it exists) such that the following happens: for all c-colorings χ of $[n]$ there exists

$$a_1 < a_2 < \cdots < a_k$$

such that
(1) $a_1 + a_2 + \cdots + a_k \le n$,
(2) for all $I \subseteq [k]$ $\chi(\sum_{i \in I} a_i)$ depends only on $\max\{i \in I\}$.
Use the following to give two proofs that $n(k, c)$ exists:

(a) Show that $n(1, c)$, $n(2, c)$, and $n(k, 1)$ all exist.

(b) Show that, for all $k \ge 3$ for all $c \ge 2$, $n(k + 1, c) \le n(k, c)W(k + 2, c^{n(k,c)})$. (This yields one proof that, for all $k, c \ge 1$, $n(k, c)$ exists.)

(c) Show that, for all $k \ge 3$ for all $c \ge 2$, $n(k+1, c) \le 2W(n(k, c), c)$. (This yields another proof that, for all $k, c \ge 1$, $n(k, c)$ exists.)

b) Show that for all k and c there exists $F = F(k, c)$ such that, for all c-colorings χ of $[n]$ there exists

$$a_1 < a_2 < \cdots < a_k$$

such that
(1) $a_1 + a_2 + \cdots + a_k \le n$,
(2) for all $I, J \subseteq [k]$ $\chi(\sum_{i \in I} a_i) = \chi(\sum_{j \in J} a_j)$. (That is, all sums of subsets are the same color.)

Exercise 28 (Open-ended) Obtain better bounds for $F(k, c)$ then the ones you get from Exercise 27

Chapter 5

Applications II

5.1 Fermat's Last Theorem Mod p Fails for Almost all p

In 1637 Fermat wrote in the margins of *Arithmetica*, a book on Number Theory by Diophantus, the following (translated from Latin)

> *To divide a cube into two cubes, a fourth power, or in general any power whatever above the second into two powers of the same denomination, is impossible, and I have assuredly found a proof of this, but the margin is too narrow to contain it.*

In modern terminology he was stating that the equation $x^n + y^n = z^n$ has no solution where $x, y, z, n \in \mathbb{N}$, $x, y, z \geq 1$, and $n \geq 3$. Many mathematicians have tried to prove this and did not succeed. This open question was known as *Fermat's Last Theorem*.

In 1993 Andrew Wiles announced that he had a proof of Fermat's Last Theorem. This proof had a gap; however, with the help of Richard Taylor, the gap was fixed and a correct proof was published in 1994 [91],[86]. This lead to Tom Lehrer adding the following verse to his song *That's Mathematics* [56].

Andrew Wiles gently smiles,
Does his thing and voila!
QED, we agree, and we all shout hurrah!
As he confirms what Fermat
Jotted down in that margin,
Which could've used some enlargin'.

This was the most famous open problem in mathematics. Why?

1. You could explain it to a layperson.

2. The story of Fermat writing it in the margin is interesting.

3. There was steady progress on it. By the year 1900 Fermat's Last Theorem was proven true for all $n \le 100$ [21]. By the year 1993, with the help of computers, Fermat's Last Theorem was proven for all $n \le 4,000,000$.

4. Fermat's Last Theorem connected to many other fields of mathematics including algebraic geometry.

5. There was a cash prize offered for its solution. To quote [81]

> Interest in FLT rocketed when a German doctor and amateur mathematician called Paul Wolfskehl offered a huge cash prize in 1908 for its solution. Many people, mostly amateur mathematicians, sent in their potential solutions to their nearest universities. Over a *thousand* proofs were received in this period, mail got so heavy that some institutions resorted to printing form sheets saying "Thank you for your 'proof' of FLT. The first mistake is on page XX"! The hyperinflation of the 1920's considerably reduced the value of the prize but it was still worth $50,000 when Andrew Wiles collected it in a grand ceremony in June 1997 in the Great Hall at Gottingen University.

This problem inspired many mathematicians, including Wiles. In that respect, it is almost a shame that it has been solved.

Here is an approach to proving Fermat's Last Theorem that was first proposed by Sophie Germain. (See [52] and [70] for more information.) We first give a trivial theorem and then we give her sophisticated theorem.

Theorem 5.1.1 *Fix $n \ge 3$, a prime.*

1. *Fix a prime p. Let $A_p = \{1^n, 2^n, \ldots, (p-1)^n\}$, all reduced modulo p. If there is no $a, b, c \in A_p$ such that $a+b \equiv c \pmod{p}$ then in any solution of $x^n + y^n = z^n$ one of x, y, z is $\equiv 0 \pmod{p}$.*

2. *Assume that for an infinite number of primes p there are not $a, b, c \in A_p$ such that $a + b \equiv c \pmod{p}$. Then $x^n + y^n = z^n$ has no solution.*

Proof:
a) If $x^n + y^n = z^n$ then $x^n + y^n \equiv z^n \pmod{p}$. Since there is no $a, b, c \in A_p$ such that $a + b = c$, at least one of x^n, y^n, z^n is $\equiv 0 \pmod{p}$. Hence at least one of x, y, z is divisible by p.

b) Let $p_1, p_2, p_3, \ldots$ be the infinite set of primes. If $x^n + y^n = z^n$, then each p_i must divide at least one of x, y, z, meaning one the three must be divisible by an infinite subset of $\{p_1, p_2, p_3, \ldots\}$. This is not possible.
▮

Exercise 29

a) For $1 \le a \le 28$ compute $a^7 \pmod{29}$.

b) Show that if $x^7 + y^7 = z^7$ then one of x, y, z is divisible by 29.

Note that in Theorem 5.1.1 we needed to look at *all* $a, b, c \in A_p$. Also note that if there was a d such that $d, d+1 \in A_p$ then by letting $a = 1$, $b = d$, and $c = d + 1$ we would have $a, b, c \in A_p$ with $a + b \equiv c \pmod{d}$. Sophie Germain showed that, in some cases, this is the *only* case we need to check.

Theorem 5.1.2 *Fix $n \ge 3$, a prime.*

1. *Fix a prime p. Let $A_p = \{1^n, 2^n, \ldots, (p-1)^n\}$, all reduced modulo p. Suppose $n \notin A_p$, and there is no a such that a and $a + 1$ are both in A_p. Then in any solution of $x^n + y^n = z^n$, one of x, y, z is divisible by p.*

2. *Assume there is an infinite number of primes p such that $n \notin A_p$ and there is no a with a and $a+1$ in A_p. Then $x^n+y^n = z^n$ has no solution.*

Using either Theorem 5.1.1 or 5.1.2 looks promising. Alas, neither approach can work. Libri [57] showed that for $n = 3, 4$, for all but a finite number of primes, that $x^n + y^n \equiv z^n \pmod{p}$ has a solution. He stated that this was true for all n, but did not prove it in general. Pellet [62] proved that for all n, for almost all p, $x^n + y^n \equiv z^n \pmod{p}$ has a solution. Pellet's proof did not give any bounds on how large p has to be. Cornacchia [15], Dickson [19], and Hurwitz [42] obtained that, for all n, for all

$$p \geq (n-1)^2(n-2)^2 + 6n - 2.$$

$x^n + y^n \equiv z^n \pmod{p}$ has a solution. Pepin [63, 64] also did some work on this problem. The proofs of Pellet, Cornacchia, Dickson, Hurwitz, and Pepin are not combinatorial. The papers are not in English; however, they are available. (See Exercise 30.)

Schur [74] later gave a proof using Rado's Theorem (Theorem 4.2.2). It is more accurate to say that he proved what is now a corollary to Rado's Theorem in order to get a proof. Schur obtained bounds for p, but they are not as good as those of Cornacchia, Dickson, and Hurwitz.

We will need some number theory first. The results are standard and can be found in any basic number theory text.

Def 5.1.3 Let p be a prime, $A \subseteq \{0, 1, \ldots, p-1\}$, and $b \in [p-1]$. Then

$$bA = \{bx \mid x \in A\}.$$

Lemma 5.1.4 *Let p be a prime. There exists a number $g \in [p-1]$ such that, doing arithmetic mod p,*

$$\{1, \ldots, p-1\} = \{g^0, g^1, \ldots, g^{p-2}\}.$$

In addition, $g^{p-1} \equiv 1 \pmod{p}$.

Such a g is called a *generator of* $\mathbb{Z}_p^*$, the multiplicative group of non-zero integers mod p.

Lemma 5.1.5 *Let p be a prime, $b, c \in [p-1]$, $n \in \mathbb{N}$ and*

$$H_n = \{x^n \pmod{p} \mid x \in [p-1]\}.$$

1. *H_n has $\frac{p-1}{\gcd(n,p-1)}$ elements.*

2. *Either $bH_n = cH_n$ or $bH_n \cap cH_n = \emptyset$.*

3. *$|H_n| = |bH_n|$*

4. *Let $h = \frac{p-1}{|H_n|} = \gcd(n, p-1) \le n$. There exists $b_1, \ldots, b_h$ such that the collection $\{b_i A\}$ form a partition of $[p-1]$.*

Def 5.1.6 Let $b_1, \ldots, b_m \in \mathbb{Z}$ and $c \in \mathbb{N}$. Let $R = R(b_1, \ldots, b_m; c)$ be the least number (if it exists) such that, for all c-colorings of $[R]$, there exists a monochromatic $e_1, \ldots, e_m$ such that $\sum_{i=1}^{m} e_i b_i = 0$.

Note that, by either Theorem 4.2.2 or Exercise 23 from Chapter 4, $R(1,1,-1;n)$ exists for all n.

Theorem 5.1.7 *For all $n \ge 1$ and all primes $p \ge R(1,1,-1;n)$. there exists $x, y, z \not\equiv 0 \pmod{p}$ such that*

$$x^n + y^n \equiv z^n \pmod{p}.$$

Note 5.1.8 If $\gcd(n, p-1) = 1$ then, by Lemma 5.1.5, $|H_n| = p-1$. Hence all numbers in $\{1, \ldots, p-1\}$ are nth powers mod p. In this case Theorem 5.1.7 is easy: just take any three numbers $a, b, c \in [p-1]$ such that $a + b \equiv c \pmod{p}$ and note that there is an x, y, z such that $a \equiv x^n \pmod{p}$, $b \equiv y^n \pmod{p}$, and $c \equiv z^n \pmod{p}$.

Proof: Fix n.

Let p be a prime such that $p \ge R(1,1,-1;n)$. All congruences are modulo p.

Let $H_n = \{x^n \pmod{p} \mid x \in [p-1]\}$. Let $h = \gcd(n, p-1) \le n$. By Lemma 5.1.5 there exists $b_1, \ldots, b_h$ such that the collection $b_i A$ form a partition of $[p-1]$. Note that h is independent of p.

Here is the key step: we now define a coloring χ:$[p-1] \to [h]$ by

$$\chi(x) = i \text{ such that } x \in b_i H_n.$$

By the definition of $R(1,1,-1;n)$, there exists $e_1, e_2, e_1 + e_2$ of the same color. Let the color be i. Let $b = b_i$. Hence $e_1, e_2, e_1 + e_2 \in bH_n$. Let $x, y, z \in [p-1]$ be such that $e_1 \equiv bx^n$, $e_2 \equiv by^n$, $(e_1 + e_2) \equiv bz^n$. Note that

$$bx^n + by^n \equiv e_1 + e_2 \equiv bz^n.$$

Hence

$$x^n + y^n \equiv z^n.$$

■

Exercise 30 Consider the following papers: Libri [57], Pellet [62], Pepin [63, 64], Dickson [19], or Cornacchia [15]. All of these papers are at `http://www.cs.umd.edu/$sim$gasarch/res`.

a) For those that are in languages other than English, translate them to English.

b) Write a survey article, with proofs, of these results. Include history.

d) We are particularly interested in Pellet's proof of the following: For all n there exists a prime p_0 such that, for all primes $p \ge p_0$ there exists $x, y, z \not\equiv 0 \pmod{p}$ such that $x^n + y^n \equiv z^n \pmod{p}$. This proof did not give bounds on p_0.

5.2 nth powers and Non-nth Powers for $n \ge 2$

Theorem 5.1.2 raises the question of when there are consecutive nth powers mod p. In this section we show that, for all k, for all $n \ge 2$, for almost all primes p, there are k consecutive nth powers mod p and k consecutive non-nth powers mod p. We will obtain Theorem 5.1.7 as an easy corollary.

5.3 Quadratic Residues and Non-Residues

Exercise 31 Show that $x^2 \equiv (p-x)^2 \pmod{p}$.

We use this exercise so that when looking at squares mod p we need only look at $1^2, 2^2, \ldots, (\frac{p-1}{2})^2$.

Consider the integers mod 13. Here are the squares mod 13, excluding 0. $1^2 \equiv 1$, $2^2 \equiv 4$, $3^2 \equiv 9$, $4^2 \equiv 3$, $5^2 \equiv 12$, $6^2 \equiv 10$.

Look at the squares in increasing order: $\{1, 3, 4, 9, 10, 12\}$. There is a sequence of two consecutive squares: 3, 4. Is there a p so that we can get three consecutive squares? Four? More? Now look at the non-squares in increasing order: $\{2, 5, 6, 7, 8, 11\}$. There is a sequence of four consecutive non-squares: 5, 6, 7, 8. Is there a p so that we can get five consecutive non-squares? Six? More?

We now look at the squares mod 23. $1^2 \equiv 1$, $2^2 \equiv 4$, $3^2 \equiv 9$, $4^2 \equiv 16$, $5^2 \equiv 2$, $6^2 \equiv 13$, $7^2 \equiv 3$, $8^2 \equiv 18$, $9^2 \equiv 12$, $10^2 \equiv 8$, $11^2 \equiv 6$. Look at the squares in increasing order: $\{1, 2, 3, 4, 6, 8, 9, 12, 13, 16, 18\}$. There is a sequence of four consecutive squares: 1, 2, 3, 4. There is a sequence of four consecutive non-squares: 19, 20, 21, 22.

Are the following statements true?

- *For every k there exists a prime p such that there are k consecutive squares mod p.*
- *For every k there exists a prime p such that there are k consecutive non-squares mod p.*

We will prove both of these. What will we use? a variant of VDW! (Why else would we talk about it?) Recall the following lemma that we used to prove the single equation Rado Theorem. It is this lemma that we can apply.
Lemma 4.2.9 *For all k, s, c, there exists $U = EW(k,s,c)$ such that for every c-coloring χ:$[U] \rightarrow [c]$ there exists a, d such that*

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

Def 5.3.1 Let p be a prime.

1. A number $x \in [p-1]$ is a *quadratic residue mod p* if there exists $y \in \{1, \ldots, p-1\}$ such that $y^2 \equiv x \pmod{p}$. If no such y exists then x is called a *quadratic non-residue mod p*. In both cases the "mod p" may be omitted if it is understood. Likewise, we may omit the word "quadratic" when we are only dealing with squares. The number 0 is considered to be neither a residue nor a non-residue.
2. If x and y are either both residues or both non-residues then we say they have *the same status.*

We will use the results in the following exercises implicitly throughout the rest of this chapter.

Exercise 32

a) Show that if x is a residue then x^{-1} is a residue.

b) Show that if x is a non-residue residue then x^{-1} is a non-residue.

c) Show that if x, y are residues then xy is a residue.

Exercise 33 The point of this exercise is to guide you through a proof that the product of two elements of the same status is a residue. Assume throughout that p is a prime. All arithmetic is done mod p.

a) Show that if $a \in [p-1]$ then $a^2 \pmod{p} \in [p-1]$.

b) Let SQ_p be the function from $[p-1]$ to $[p-1]$ that maps of a to $a^2 \pmod{p}$. Show that for every b in the image, there are exactly two a's such that $SQ_p(a) = b$.

c) Show that there are exactly $\frac{p-1}{2}$ residues and $\frac{p-1}{2}$ non-residues.

d) Let $a \in [p-1]$. Show that $\{ab \mid b \in [p-1]\} = [p-1]$. That is, show that $a \cdot [p-1] = [p-1]$.

e) Show that if a and b have different status then ab is not a residue.

f) Show that if a and b have the same status then ab is a residue.

Exercise 34 The point of this exercise is to guide you through a *different proof* that the product of two elements of the same status is a residue. Assume throughout that p is a prime. All arithmetic is done mod p. For this problem we use -1 to represent $p-1$.

a) Show that the only roots of $x^2 - 1 = 0$ are 1 and -1. (Note: this is false when p is not prime.)

b) Show that for all $a \in [p-1]$, $a^{(p-1)/2} \in \{-1, 1\}$.

c) Show that if a is a residue then $a^{(p-1)/2} = 1$.

d) Show that the polynomial $x^{(p-1)/2} - 1 = 0$ has exactly $(p-1)/2$ roots.

e) Show that if a is a non-residue then $a^{(p-1)/2} = -1$.

f) Show that if a and b have the same status then ab is a residue.

Theorem 5.3.2 *Let $k \in \mathbb{N}$. If $p \geq EW(k,1,2)$ and p is prime then there exists k consecutive residues mod p. (Recall EW from Lemma 4.2.9.)*

Proof:

Let p be a prime such that $p \geq EW(k, 1, 2)$. All of the arithmetic in this proof is mod p.

Let χ be the following 2-coloring of $[p-1]$:

$$\chi(x) = \begin{cases} 0 & \text{if } x \text{ is not a residue} \\ 1 & \text{if } x \text{ is a residue} \end{cases} \tag{5.1}$$

There exists a, d such that $a, a+d, a+2d, \ldots, a+(k-1)d, d$ all have the same status. Hence all of the following are residues:

$$ad^{-1}, (a+d)d^{-1}, (a+2d)d^{-1}, \ldots, (a+(k-1)d)d^{-1}$$

which is

$$ad^{-1}, ad^{-1}+1, ad^{-1}+2, \ldots, ad^{-1}+k-1.$$

▮

Def 5.3.3

1. Let $\text{con}_2(k)$ be the least prime such that, for all primes $p \geq \text{con}_2(k)$, there are k consecutive residues.
2. Let $\text{connon}_2(k)$ be the least prime such that, for all primes $p \geq \text{connon}_2(k)$, there are k consecutive non-residues.

Theorem 5.3.2 states that $\text{con}_2(k) \leq EW(k, 1, 2)$. By the proof of Lemma 4.2.9 $EW(k, 1, 2) \leq W(k^2, 2)$. This bound is large. We will consider the question of better bounds in Section 5.4.1.

Theorem 5.3.4 *For all k, for all primes $p \geq \text{con}_2((k-1)k!)$, there are k consecutive non-residues mod p. Hence $\text{connon}_2(k) \leq \text{con}_2((k-1)k!)$.*

Proof:

Fix k. Since $p \geq \text{con}_2((k-1)k!)$ there is a consecutive sequence of residues. Let

$$b, b+1, b+2, \ldots, b+(k-1)k!.$$

be that sequence.

Let n be the least a non-residue in $[p-1]$.

Case 1: $n \le k!$. Then the numbers $b, b+n, b+2n, \ldots, b+(k-1)n$ are all greater than b and less than $b+(k-1)k!$ when reduced modulo p.

Hence

$$b, b+n, b+2n, \ldots, b+(k-1)n.$$

are all residues. Multiply them all by the non-residue n^{-1} to obtain the following consecutive sequence of non-residues:

$$bn^{-1}, bn^{-1}+1, bn^{-1}+2, \ldots, bn^{-1}+(k-1).$$

Case 2: $n > k!$. Hence the numbers

$$1, 2, 3, \ldots, n-1$$

are all residues. (Note that for this case we are not using the sequence $b, b+1, \ldots, b+(k-1)k!$ at all.)

Consider the following subset of the above sequence

$$n-k!, n-\frac{k!}{2}, n-\frac{k!}{3}, \ldots, n-\frac{k!}{k}.$$

Since $1, 2, 3, \ldots, k$ are residues the following are residues:

$$1 \cdot \left(n-\frac{k!}{1}\right), 2 \cdot \left(n-\frac{k!}{2}\right), 3 \cdot \left(n-\frac{k!}{3}\right), \ldots, k \cdot \left(n-\frac{k!}{k}\right).$$

$$n-k!, 2n-k!, 3n-k!, \ldots, kn-k!.$$

We multiply by non residue n^{-1} to obtain the following sequence of consecutive non residues:

$$1-n^{-1}k!, 2-n^{-1}k!, 3-n^{-1}k!, \ldots, k-n^{-1}k!.$$

▮

Exercise 35 Show that, for all $p \equiv 1 \mod 4$, there are k consecutive residues modulo p if and only if there are k consecutive non-residues modulo p.

5.4 nth Residues and nth-Non-Residues for $n \geq 3$

Def 5.4.1 Let p be a prime and g be a generator of $\mathbb{Z}_p^*$. Let h divide $p-1$, with $h > 1$. Then $\text{logmod}_{p,g,h}$:$[p-1] \rightarrow \mathbb{Z}_h$ is defined by

$$\text{logmod}_{p,g,h}(g^a) = a \pmod{h}.$$

When p, g and h are understood we use logmod.

Note 5.4.2 Since $g^{p-1} \equiv 1 \pmod{p}$, the exponent b satisfying $g^b \equiv a \pmod{p}$ is only determined mod $p-1$.

Theorem 5.4.3 *For all $n \geq 3$, for all k, for all primes $p \geq EW(k,1,h)$ there are k consecutive nth roots mod p.*

As noted before, this theorem is trivial when $\gcd(n, p-1) = 1$, since everything is an nth power mod p in this case.

Proof: Let p be a prime larger than $EW(k,1,n)$. All of the arithmetic in this proof is mod p unless otherwise specified.

Let g be the generator of $\mathbb{Z}_p^*$, which exists by Lemma 5.1.4. Let logmod = $\text{logmod}_{p,g,h}$, where $h = gcd(n, p-1)$. Let $\ell = \frac{n}{h}$ The key observation is that, since logmod maps to $\mathbb{Z}_h$, we may treat it as an h-coloring.

By Lemma 4.2.9 there exists a, d such that

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

all have the same color. Let that color be i. Hence there exists $b_0, \ldots, b_{k-1}$ such that

$$a = g^{b_0h+i}, a+d = g^{b_1h+i}, a+2d = g^{b_2h+i}, \ldots, a+(k-1)d = g^{b_{k-1}h+i}.$$

Note also that there exists m such that $d = g^{mh+i}$. Hence $d^{-1} = g^{-mh-i}$. Note that

$$ad^{-1}, (a+d)d^{-1}, (a+2d)d^{-1}, \ldots, (a+(k-1)d)d^{-1}$$

is both

$$g^{(b_0-m)h}, g^{(b_1-m)h}, \ldots, g^{(b_{k-1}-m)h}$$

and

$$ad^{-1}, ad^{-1}+1, \ldots, ad^{-1}+k-1.$$

Hence we have k consecutive hth powers. Finally, rewriting

$$g^{rh} = \left(g^{\ell^{-1}r}\right)^{h\ell} = \left(g^{\ell^{-1}r}\right)^{n},$$

where ℓ^{-1} is the inverse of ℓ mod $p-1$, we see that these are really consecutive nth powers as well.

▮

Def 5.4.4 Let $\mathrm{con}_n(k)$ be the least prime such that, for all primes $p \geq \mathrm{con}_n(k)$, there are k consecutive nth powers. Let $\mathrm{connon}_n(k)$ be the least prime such that, for all primes $p \geq \mathrm{connon}_n(k)$ with $\gcd(n, p-1) > 1$, there are k consecutive non-nth powers.

Theorem 5.4.5 *For all n, k,* $\mathrm{connon}_n \leq \mathrm{con}_n((k-1)k!)$.

Proof: This proof is virtually identical to the proof of Theorem 5.3.4. ▮

Exercise 36 Prove that for all n there is an infinite set of primes p such that all numbers mod p are nth powers.

Exercise 37

a) Prove Theorem 5.4.5.

b) Prove the following slightly stronger version of Theorem 5.4.5: For all n, k, for all primes $p \geq \mathrm{con}_n((k-1)k!)$ with $\gcd(n, p-1) > 1$, for all generators g of $\mathbb{Z}_p$, there are k consecutive numbers that have the same $\mathrm{logmod}_{p,g,n}$ value.

5.4.1 Better Bounds on $\mathrm{con}_n(k)$ and $\mathrm{connon}_n(k)$

By Theorems 5.4.3 and 5.4.5

$$\mathrm{con}_n(k) \leq EW(k,1,n)$$

and

$$\mathrm{connon}_n(k) \leq \mathrm{con}_n((k-1)k!) \leq EW((k-1)k!,1,n).$$

What more can be said about these two functions? The first two results below are due to Peralta [65]. The rest are due to Davenport [16, 17]. Both Peralta and Davenport used number theory, not combinatorics.

1. $\mathrm{con}_2(k) \leq 3k2^{k-1}$
2. $\mathrm{connon}_2(k) \leq 3k2^{k-1}$.
3. $\mathrm{con}_n(2) \leq O(n^2)$
4. $\mathrm{connon}_n(2) \leq O(n^2)$
5. $\mathrm{con}_n(3) \leq 100n^{12}$
6. $\mathrm{connon}_n(3) \leq 100n^{12}$

We have not been able to find anything more on bounds for $\mathrm{con}_n(k)$ or $\mathrm{connon}_n(k)$. However, the following are known:

1. Lehmer and Lehmer [53] proved that $\mathrm{con}_2(2) = 7$. See Exercise 38.
2. Dunton [20] proved that $\mathrm{con}_3(2) = 79$. See Exercise 40.
3. Lehmer, Lehmer, Mills, and Selfridge [55] proved that $\mathrm{con}_3(3) = 293$.
4. Mills and Bierstedt [9] proved that $\mathrm{con}_4(2) = 43$.
5. Lehmer and Lehmer and Mills [54] showed that $\mathrm{con}_5(2) = 103$.
6. Lehmer and Lehmer and Mills [54] showed that if $\mathrm{con}_6(2) = 281$.

Exercise 38 In this exercise you will prove $\mathrm{con}_2(2) = 7$. Let $p \notin \{2,3,5\}$. Throughout we abbreviate *consecutive run of two square* by CR2S.

a) Show that if 2 or 5 is a square mod p then there is a CR2S.

b) Show that if 2 and 5 are not squares mod p then there is a CR2S.

c) Where did your proof use that $p \notin \{2, 3, 5\}$?

d) Characterize exactly which primes have a CR2S.

e) Determine $\mathrm{con}_2(2)$.

Exercise 39 Let p be a prime of the form $p = 3k + 1$. Let g be a generator of $\mathbb{Z}_p^*$. Let logmod $=$ $\mathrm{logmod}_{p,g,3}$.

a) Show that $\mathrm{logmod}(xy) \equiv \mathrm{logmod}(x) + \mathrm{logmod}(y) \pmod 3$.

b) Show that if $x, y \in [p-1]$, $\mathrm{logmod}(x), \mathrm{logmod}(y) \neq 0$, and $\mathrm{logmod}(x) \neq \mathrm{logmod}(y)$ then $\mathrm{logmod}(xy) = 0$.

c) Show that, for all $x, y \in [p-1]$. one of x, y, xy, xy^2 is a cube mod p (hence its logmod value is 0).

Exercise 40 In this exercise you will prove that $\mathrm{con}_3(2) \leq 79$. Let p be a prime of the form $p = 3k+1$ such that $p \geq 79$. All arithmetic in this exercise is done mod p. Throughout we abbreviate *consecutive run of two cubes* by CR2C. Hint: Several of the problems in this exercise use Exercise 39.

a) Show that if $\mathrm{logmod}(2) = 0$ or $\mathrm{logmod}(3) = 0$ or $\mathrm{logmod}(7) = 0$ or $\mathrm{logmod}(28) = 0$ then there is a CR2C.

b) Show that if $\mathrm{logmod}(2) \neq 0$ and $\mathrm{logmod}(7) \neq 0$ and $\mathrm{logmod}(28) \neq 0$ then $\mathrm{logmod}(14) = 0$ and $\mathrm{logmod}(2) \neq \mathrm{logmod}(7)$.

   Note: In parts c,d,e,f of this problem we assume $\mathrm{logmod}(2) \neq 0$, $\mathrm{logmod}(3) \neq 0$, $\mathrm{logmod}(7) \neq 0$, $\mathrm{logmod}(28) \neq 0$, $\mathrm{logmod}(14) = 0$, and $\mathrm{logmod}(2) \neq \mathrm{logmod}(7)$. Since clearly if $\mathrm{logmod}(13) = 0$ we have a CR2C we will also assume $\mathrm{logmod}(13) \neq 0$. Let **FACT 1** be the set of facts we are assuming at this point.

c) Show that if $\mathrm{logmod}(2) \neq \mathrm{logmod}(13)$ then there is a CR2C.

   Note: In parts d,e,f of this problem we assume **FACT 2:** $\mathrm{logmod}(2) = \mathrm{logmod}(13)$.

d) Show that if logmod(3) = logmod(7) then there is a CR2C.

   Note: In part e,f of this problem we assume **FACT 3:** logmod(3) = logmod(7).

e) Assume **FACT 1,2,3**. For which x, y such that logmod(x) $\neq 0$ and logmod(y) $\neq 0$ do we know that logmod(x) $\neq$ logmod(y)? For which x, y such that logmod(x) $\neq 0$ and logmod(y) $\neq 0$ do we know that logmod(x) = logmod(y)? For which x do we now know logmod(x) = 0?

f) Show the following.

   (a) If logmod(11) = 0 then there is a CR2C.

   (b) If logmod(11) = logmod(7) then there is a CR2C.

   (c) If logmod(11) $\neq 0$ and logmod(11) $\neq$ logmod(7) then there is a CR2C.

g) Where in the above steps did you use the fact that $p \geq 79$?

h) For all primes p, $2 \leq p \leq 79$, determine which ones have CR2C. (To save time use Note 5.1.8.)

i) Determine $\text{con}_3(2)$ exactly.

Exercise 41 Research problem. Find upper and lower bounds on $\text{con}_n(k)$ for small values of k and n.

Exercise 42 Write a survey paper summarizing everything that is known about $\text{con}_n(k)$. You can use the website `http://www.cs.umd.edu/~gasarch/res` as a starting point.

5.5 Historical Notes

In chapters 4 and 5 we presented Rado's Theorem and then some applications to Number Theory. Historically the applications and the theorem came hand in hand.

The first theorem involving equations and colorings was Schur's Theorem which we presented in Exercise 23. Schur proved this theorem to obtain Theorem 5.1.7. There had been prior proves of it. Schur's proof is combinatorial rather than number theoretic.

Rado, a student of Schur, later [68, 69] proved the full Rado Theorem, which we stated and prove in Section 4.3.

Schur conjectured that, for all k, for almost all primes p, there are k consecutive quadratic residues. Bauer, a student of Schur, proved this using VDW (not Extended VDW, Lemma 4.2.9). His proof essentially proved extended VDW in this context. Schur then proved the Extended VDW. Schur also, building on Bauer's proof, showed that for all n, k, for almost all primes p, there are k consecutive nth powers. There is some confusion about whose result this is: Schur claims it is Bauer's and Bauer claims it is Schur's.

Chapter 6

The Polynomial van der Waerden's Theorem

6.1 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 χ:$[W] \to [c]$, there exists $a, d \in [W]$, such that the following set is monochromatic:*

$$\{a\} \cup \{a + id \mid 1 \leq i \leq k - 1\}.$$

Note that van der Waerden's Theorem was really about the set of functions $\{id \mid 1 \leq i \leq k - 1\}$. 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 χ:$[W] \to [c]$, there are $a, d \in \mathbb{N}$, $d \neq 0$, such that the following set is monochromatic:*

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

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 χ:$[W] \rightarrow [c]$, there are $a, d \in [W]$, such that the following set is monochromatic:*

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

This was proved for $k = 1$ by Fürstenberg [26] and (independently) Sarkozy [73]. The original proof of the full theorem by Bergelson and Leibman [7] used ergodic methods. A later proof by Walters [90] uses purely combinatorial techniques. We will present an expanded version of Walters' proof.

Note 6.1.1 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 6.1.2 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 6.4. What happens if instead of polynomials we use some other types of functions? This is also an interesting open question. See Section 6.5 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 6.1.3 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 6.1.4

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 6.1.5

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
$$P \subseteq \{bx^2+ix \mid -k \le i \le k\}.$$

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 (n_3, n_2, n_1) then there exists $b_3^{(1)}, \ldots, b_3^{(n_3)} \in \mathbb{Z}$, $b_2^{(1)}, \ldots, b_2^{(n_2)} \in \mathbb{Z}$, $b_1^{(1)}, \ldots, b_1^{(n_1)} \in \mathbb{Z}$, $k_1, k_2 \in \mathbb{N}$, T_1 of type (k_1), and T_2 of type (k_2, k_1) such that
$$\begin{aligned} P \subseteq\ & \{b_3^ix^3+p(x) \mid 1 \le i \le f, p \in T_2\} \cup \\ & \{b_2^ix^2+p(x) \mid 1 \le i \le g, p \in T_1\} \cup \\ & \{b_1^ix \mid 1 \le i \le h\} \end{aligned}$$

7. Let
$$P = \{2x^2+3x, x^2+20x, 5x, 8x\}.$$
Let
$$Q = \{p(x)-8x \mid p \in P\}.$$
Then
$$Q = \{2x^2-5x, x^2+12x, -3x, \}.$$
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.

8. 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 6.1.6

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 χ:$[W] \rightarrow [c]$, 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 6.1.7

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 will prove the following:

1. POLYVDW(1) (this will easily follow from the pigeon hole principle).

2. 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 6.1.8 $(n_e, \ldots, n_1) \preceq (m_{e'}, \ldots, m_1)$ 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 6.1.9 We will use the following ordering on types.

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

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

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

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

6.2 The Proof of the Polynomial van der Waerden Theorem

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

Def 6.2.1 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 6.2.2 *Fix k, c throughout. For all r there exists $U = U(r)$ such that for all c-colorings χ:$[U] \rightarrow [c]$ 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 $U(r)$ to be the least number such that this Lemma holds. We will prove $U(r)$ exists by giving an upper bound on it.

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

Let χ 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

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

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

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

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

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 $U(r)$ exists, and let

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

(X stands for eXtremely large.)
We show that

$$U(r+1) \leq (X \times U(r))^2 + X \times U(r).$$

Let χ be a c-coloring of

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

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

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

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

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

Consider the coloring of B_A. Since B_A is of size $U(r)$ either there exists $a, d \in U(r)$ 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
$a' \in B_A$, $d_1', \ldots, d_r' \in [U(r)]$
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 $\{B_A, B_{A+D}, \ldots, B_{A+(k+2U(r))D}\}$ is monochromatic we also have that, for all j with $0 \le j \le k + 2U(r)$,

NEED FIGURE

$$\{a' + {d_1'}^2 + jD, a' + {d_1'}^2 + d_1' + jD, \ldots, a' + {d_1'}^2 + kd_1' + jD \mid 0 \le j \le k + 2U(r)\}$$

is monochromatic

$$\{a' + {d_2'}^2 + jD, a' + {d_2'}^2 + d_2' + jD, \ldots, a' + {d_2'}^2 + kd_2' + jD\} \mid 0 \le j \le k + 2U(r)\}$$

is monochromatic

$$\vdots$$

$$\{a' + {d_r'}^2 + jD, a' + {d_r'}^2 + d_r' + jD, \ldots, a' + {d_2'}^2 + kd_r' + jD\} \mid 0 \le j \le k + 2U(r)\}$$

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 $a, d_1, \ldots, d_r \in [U(r+1)]$.

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 $0 \le j + 2d_i' \le k + 2U(r)$. 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 $B_{A+(j+2d_i')D'}$,

$$\chi(a + d_i^2) = \chi(a + d_i^2 + jd_i).$$

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

▮

Theorem 6.2.3 *For all k,* POLYVDW(P_k).

Proof: We show $W(P_k, c)$ exists by bounding it. Let $U(r)$ be the function from Lemma 6.2.2. We show $W(P_k, c) \le U(c)$. If χ is any c-coloring of $[U(c)]$ then second case of Lemma 6.2.2 cannot happen. Hence the first case must happen, so there exists $a, d \in [U(c)]$ such that

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

Therefore $W(P_k, c) \le U(c)$. ▮

Note 6.2.4 The proof of Theorem 6.2.3 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$(1,0)$.

6.2.2 The Full Proof

We prove a lemma from which the implication

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

will be obvious.

Lemma 6.2.5 *Let $n_e, \ldots, n_i \in \mathbb{N}$. Assume that* $\mathrm{POLYVDW}(n_e, \ldots, n_i, \omega, \ldots, \omega)$ *holds. Let $P \subseteq \mathbb{Z}[x]$ of type $(n_e, \ldots, n_i + 1, 0, \ldots, 0)$. Let $c \in \mathbb{N}$. We regard these as fixed. For all r, there exists $U = U(r)$*[1] *such that for all c-colorings $\chi{:}[U] \to [c]$ 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

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

[1] Formally U depends on P, c, r; however, we suppress the dependence on P and c for notational ease.

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*

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

or

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

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

Proof: We define $U(r)$ to be the least number such that this Lemma holds. We will prove $U(r)$ exists by giving an upper bound on it. In particular, for each r, we will bound $U(r)$. 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 coefficients 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. We assume that x^i be the first term of some polynomial of degree i in P (the proof for bx^i with $b \in \mathbb{Z}$ is similar).

Base Case: $r=1$. Let

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

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 $(\forall q \in Q)[q(0)=0]$. Since POLYVDW$(n_e,\ldots,n_i,\omega,\ldots,\omega)$ is true, POLYVDW(Q) is true. Hence $W(Q,c)$ exists.

We show that

$$U(1) \le W(Q,c)^i + W(Q,c).$$

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

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

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

(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 $a \in [U(1) \le W(Q,c)^i + W(Q,c)]$.)

Then

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

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 $a, d \in [U(1)]$?

Since

$$a' \ge W(Q,c)^i + 1$$

and

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

we have that

$$a = a' - b(d')^i \ge W(Q,c)^i + 1 - d(d')^i \ge W(Q,c)^i + 1 - W(Q,c)^i = 1$$

Clearly

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

Hence

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

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

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

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

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

Note that

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

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

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

Since every $q \in Q$ is an integer polynomial with $q(0) = 0$, it follows that $U(r)$ divides $q(xU(r))$, 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

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

Let χ be a c-coloring of

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

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

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

By the definition of $W(Q', c^{U(r)})$ applied to χ^*, and the assumption that POLYVDW$(n_e, \ldots, n_i, \omega, \ldots, \omega)$ holds, there exists $A, D' \in [W(Q', c^{U(r)})]$ 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 $D = D' \times U(r)$ be the distance between corresponding elements of the blocks.

Because each block is length $U(r)$, 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

NEED FIGURE

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

This will be very convenient.

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

I) There exists $a \in B_A$ and $d \in [U(r)]$ 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 $a' \geq W(Q', c^{U(r)})^i + 1$) and $d'_1, \ldots, d'_r \in [U(r)]$ 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 $q \in Q$,

$$\begin{gathered}\{a' + p(d_1') + q(D) \mid p \in P\} \text{ is monochromatic} \\ \{a' + p(d_2') + q(D) \mid p \in P\} \text{ is monochromatic} \\ \vdots \\ \{a' + p(d_r') + q(D) \mid p \in P\} \text{ is monochromatic.}\end{gathered}$$

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 $a = a' - D^i$. Note that since

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

and

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

we have

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

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

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

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:

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

Hence the following set is monochromatic:

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

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 $a' + p(d_i')$.

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

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

— Put bounds on d_i in here.

BILL - CHECK THIS

Let $d_i = d_i' + D$ for $1 \le i \le r$, 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 6.2.6 *For all $n_e, \ldots, n_i$*

POLYVDW$(n_e, \ldots, n_i, \omega, \ldots, \omega) \implies$ 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 6.2.5 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 6.2.7 *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 χ:$[W] \rightarrow [c]$, 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 6.1.8. 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 6.2.5 is the induction step.

This proves the theorem.

▮

Note 6.2.8

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$(P, 2012)$ is true but stupid: if χ is an 2012-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$(P, 2012)$ holds. 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$(P, 2012)$ is true in a way that is not stupid.

BILL- ADD STUFF ON CAROLYN NUMBERS

6.3 Bounds on the Polynomial van der Waerden numbers

6.3.1 Upper Bounds

6.3.2 Upper Bounds via Alternative Proofs

6.3.3 Lower Bounds

Theorem 6.3.1 *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,* $\mathrm{POLYVDW}(P, c) \ge c^{(k-1)/2}$.

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 2.3.5 when dealing with the ordinary VDW. This is because we knew more about the actual polynomials involved. Below we obtain better bounds for particular sets of polynomials.

Theorem 6.3.2 *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 POLYVDW$(P, c) \geq c^{(1-\epsilon)k} - 1$ ▮

Better bounds are known. See [82] and [67]
BILL- FILL IT IN- ADD MORE REFS AND POSSIBLY PROOFS

6.4 What if we use Polynomials with a Constant term?

6.5 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 χ of $\mathbb{N}$ there exists $a, d \in \mathbb{N}$ such that with

$$\chi(a) = \chi(a + b^d)$$

Alas, this is never true.

Theorem 6.5.1 *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 $\chi:\mathbb{N} \rightarrow [p]$ such that, $\forall a, d \in \mathbb{N}, \chi(a) \neq \chi(a + b^d)$.*

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

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

It is an open question to determine if Theorem 6.5.1 is tight. Also, it is open to investigate other functions.

Chapter 7

An Applications of the Poly van der Waerden Theorem

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

Chapter 8

The Hales-Jewett Theorem

8.1 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 progressions 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

  $(\forall s \in \lambda)(\forall i)[P_i^s = s$ 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 8.1.1 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

$$\begin{gathered}\text{GAABARDAA}\\ \text{GBBBBRDBA}\\ \vdots\\ \text{GYYBYRDYA}\end{gathered}$$

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 8.1.2 When $t \le 2$, a line$^-$ may have more than one completion, since in that case a line$^-$ is a single point. For example, $\{\text{BAA}\}$ 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 χ:$[t]^N \rightarrow [c], \exists L \subseteq [t]^N, L$ a monochromatic line.

There are some easy base cases:

Fact 8.1.3

- $c = 1$*:* $HJ(t,1) = 1$*. Any 1-coloring of* $[t]^1 = [t]$ *easily has a monochromatic 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 8.1.4 $HJ(2, c) = c + 1$

Proof:

Let χ:$[2]^{c+1} \rightarrow [c]$ 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 $\lambda = \{5, 6\}$.) $\blacksquare$

We will give two proof of the HJ: (1) The original proof due to Hales and Jewett [39], presented as a color-focusing argument, and (2) a proof due to Shelah [76] and yields much better upper bounds on the Hales-Jewett numbers.

8.2 Proof of the Hales-Jewett Theorem

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

Lemma 8.2.1 *Fix $t, c \in \mathbb{N}$. Assume $(\forall c')[HJ(t-1,c')$ exists$]$. Then, for all r, there exists $U = U(r)$* [1] *such that, for all c-colorings $\chi:[t]^U \rightarrow [c]$ one of the following statements holds.*

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

Statement II: *There exist r monochromatic lines$^-$ $L_1^-, L_2^-, \ldots, L_r^- \subseteq [t]^U$, and a point $Q \in [t]^U$, 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 $[t]^U$ 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 $U(r)$ to be the least number such that this Lemma holds. We will prove $U(r)$ exists by giving an upper bound for it.

Base Case: If $r = 1$ then $U(1) \leq HJ(t-1,c)$ suffices (actually $U(1) = HJ(t-1,c)$). 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 $U(r)$ exists. Let

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

(X stands for eXtremely large.)

We will show that

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

BILL- CHECK ON THIS- HERE WE HAVE $U(r+1)$ BOUNDED, ELSEWHERE WE HAVE IT EXACT. IS IT SUPPOSED TO BE THE LEAST NUMBER OR A NUMBER?

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

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

[1] Formally U depends on k, c, r; however, we suppress the dependence on k and c for ease of notation.

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

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

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

Note 8.2.2 At this point, it is essential to realize that χ' 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, $\chi'(\sigma)$ might be the following 3-coloring of $[2]^3$

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

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

$$(\sigma)(0,1,0) = 3$$

$$(\sigma)(0,1,1) = 2$$

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

$$(\sigma)(1,0,1) = 3$$

$$(\sigma)(1,1,0) = 2$$

$$(\sigma)(1,1,1) = 2$$

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

$$(\sigma)(\tau) = \chi(\sigma\tau).$$

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

We treat $\chi'(\sigma)$ 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

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

L is a line in $[t-1]^{U(r)}$, 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 $[t]^{U(r)}$. By definition of $U(r)$, 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 $[t]^{U(r)}$, and each with the same completion $Q \in [t]^{U(r)}$. 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 satisfied. ∎

We now restate and prove HJ:

Theorem 8.2.3 ***Hales-Jewett Theorem*** *$\forall t, c, \exists N = HJ(t,c)$ such that, for all c-colorings $\chi{:}[t]^N \rightarrow [c], \exists L \subseteq [t]^N, L$ 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 8.1.3 $HJ(1,c) = 1$ exists.
Induction Step: Assume $(\forall c)[HJ(t{-}1,c)$ exists $]$. Fix c. Consider Lemma 8.2.1 when $r = c$. In any c-coloring of $[t]^{U(c)}$, 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 $HJ(t,c) \leq U(t)$. ▮

8.3 Shelah's Proof of the Hales-Jewett Theorem

8.4 Bounds on the Hales-Jewett numbers

8.4.1 Upper Bounds on the Hales-Jewett numbers

8.4.2 Lower Bounds on the Hales-Jewett numbers

Chapter 9

Applications of the Hales-Jewett Theorem

9.1 Positional Games

9.2 VDW and Variants

9.3 Comm. Comp. of χ

9.4 The Square Theorem: Fourth Proof

Use HJ directly.

9.5 The Gallai-Witt Theorem (Multidim VDW)

Theorem 9.5.1 *Let $c, M \in \mathbb{N}$. Let $\chi^*: \mathbb{N} \times \mathbb{N} \rightarrow [c]$. There exists a, d, D such that all of the following are the same color.*

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

9.6 The Canonical VDW

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

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

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

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

Def 9.6.1 Let $k \in \mathbb{N}$. Let χ be a coloring of $\mathbb{N}$ (which may use a finite or infinite number of colors). A *rainbow k-AP* is an arithmetic progression $a, a+d, a+2d, \ldots, a+(k-1)d$ such that all of these are colored *differently.*

The following is the *Canonical van der Waerden's Theorem.* Erdős and Graham [23] claim that it follows from Szemerëdi's Theorem on density. Later Prömel and Rödl [66] obtained a proof that used the Gallai-Witt Theorem.

Theorem 9.6.2 *Let $k \in \mathbb{N}$. Let $\chi:\mathbb{N} \rightarrow \mathbb{N}$ be a coloring of the naturals. One of the following two must occur.*

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

Proof:

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

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

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 9.6.3

1. Let $k = 10$ and assume

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

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

2. Let $k = 6$ and assume

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

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

Let M be a constant to be picked later. By Theorem 9.5.1 There exists a, d, D such that all of the following are the same χ^*

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

There are two cases.
Case 1: $\chi^*(a, d)$ 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 $\chi^*(a, d)$ is the partition that puts $a + xd$ and $a + yd$ in the same class. More simply, $\chi(a + xd) = \chi(a + yd)$. Since for all $-M \leq i, j \leq M$,

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

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

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

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

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

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 $-M \le i, j \le M$,

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

Let $i = -xj$ and you get

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

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

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

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

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 9.6.4 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 [66].

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

Chapter 10

The Polynomial Hales-Jewett Theorem

10.1 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 case 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 10.1.1 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 10.1.2 ***Hales-Jewett Theorem***
For every c, every finite alphabet Σ, there is some N such that, for any c-coloring $\chi{:}S_\Sigma(N) \rightarrow [c]$, there is a point $A \in S_\Sigma(N), \lambda \subseteq [N]$, with $\lambda \neq \emptyset$ such that the set $\{A + \sigma\lambda \mid \sigma \in \Sigma\}$ is monochromatic.

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

Theorem 10.1.3 ***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 $\chi{:}S_\Sigma(N) \rightarrow [c]$, 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.

Example 10.1.4 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:

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

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

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

POLYHJ was first proven by Bergelson and Leibman [8] using ergodic methods. We present the proof by Walters [90] that uses purely combinatorial techniques.

10.2 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 10.2.1 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 10.2.2 Define

$$\begin{aligned} f(\gamma) &= a\gamma^3 + 3\gamma^2 + \heartsuit\gamma \\ g(\gamma) &= a\gamma^3 + 3\gamma^2 + \diamondsuit\gamma \\ h(\gamma) &= b\gamma^3 + 3\gamma^2 + \heartsuit\gamma \end{aligned}$$

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 10.2.3 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 6.1.8.

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

$$\begin{aligned} p_1 &= a\gamma^3 + 6\gamma^2 + \diamondsuit\gamma \\ p_2 &= a\gamma^3 + 6\gamma^2 + \heartsuit\gamma \\ p_3 &= a\gamma^3 + 7\gamma^2 + \heartsuit\gamma \\ p_4 &= b\gamma^3 + 6\gamma^2 + \diamondsuit\gamma \\ p_5 &= b\gamma^3 + 6\gamma^2 + \heartsuit\gamma \end{aligned}$$

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 10.2.5 *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.

10.3 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 10.3.1 *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 $\lambda \subseteq \{1,\dots,n\}$, and $\kappa \subseteq \{n+1,\dots,n+m\}$. 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

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

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 10.3.2 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 10.3.3 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 6.1.6, we also define $POLYHJ(n_d, \ldots, n_k, \omega, \ldots, \omega)$ to be the analogous statement.

10.4 The Proof

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

Lemma 10.4.1 *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 $\chi : S_\Sigma(U) \to [c]$, 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 10.3.1, $S_\Sigma(U + HJ) \cong S_\Sigma(U) \times S_{\Sigma'}(HJ)$, 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 10.2.5, $P - \{p_0\}$ 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 10.3.2, 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 10.3.2, 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 coefficients 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 $\chi{:}S_\Sigma(U+HJ) \to [c]$ be a c-coloring. Then we view χ as a c-coloring of $S_\Sigma(U) \times S_{\Sigma'}(HJ)$. As such, for each $\sigma \in S_{\Sigma'}(HJ)$, define $\chi^*(\sigma){:}S_{\Sigma'} \to [c]$ as the coloring of $S_\Sigma(U)$ induced by χ — for $\tau \in S_\Sigma(U)$, the map is defined so that $\chi^*(\sigma)(\tau) = \chi(\tau, \sigma)$. This makes χ^* 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 χ^* 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 monochromatic under χ. 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 $A_1, \dots, A_r, A' \in S_\Sigma(U), \lambda_1, \dots, \lambda_r \subseteq [U]$ 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 χ:

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

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 10.4.2 ***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 $\chi{:}S_\Sigma(N) \rightarrow [c]$, 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 $\chi{:}S_\Sigma(1) \rightarrow [c]$ 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 $\chi{:}S_\Sigma(U) \rightarrow [c]$ be a c-coloring. Statement II cannot hold, since it requires $c + 1$ different colors. Thus, Statement I holds, which was the goal. ▮

10.5 Bounds on the Polynomial Hales-Jewett numbers

10.5.1 Upper Bounds

10.5.2 Lower Bounds

Chapter 11

Applications of Polynomial Hales-Jewett Theorem

11.1 The Polynomial van der Waerden Theorem

11.2 The Poly van der Waerden Theorem Over a Commutative Ring

11.3 The Multidim Poly van der Waerden Theorem

Chapter 12

Advanced Topics*

12.1 Every Set of Positive Upper Density has a 3-AP

12.1.1 Combinatorial Proof

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 [36, 71, 72] 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 [36, 83] by a combinatorial proof. Szemeredi [84] 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, Rothschild, and Spencer [36].

More is known. A summary of what else is known will be presented in the next section.

Def 12.1.1 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 12.1.2 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 12.1.3 *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 details of the definition to the reader.) Note that if $i \le k+1$ then

   $x \in B_i \implies x + b_1d_1 + \cdots + b_{i-1}d_{i-1} \in B$ 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 $e+b_1d_1+\cdots+b_kd_k \in B$ 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 12.1.4 *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 12.1.5 *Let $n, n_0 \in \mathbb{N}$ and $\lambda, \lambda_0 \in (0,1)$. Assume that $\lambda < \lambda_0$ and that $(\forall m \geq n_0)[sz(m) \leq \lambda_0 m]$. Assume that $\frac{n}{2} \geq n_0$. Let $a, L \in \mathbb{N}$ such that $a \leq \frac{n}{2}$, $L < \frac{n}{2} - a$, and $a \geq n_0$. Let $A \subseteq [n]$ be a 3-free set such that $|A| \geq \lambda n$.*

1. *There is an interval $I \subseteq [a, \frac{n}{2}]$ of length $\leq L$ such that*

$$|A \cap I| \geq \left\lfloor \frac{2L}{n-2a}(\lambda_0(\frac{n}{2} - a) - n(\lambda_0 - \lambda)) \right\rfloor .$$

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 $\leq O(\sqrt{n})$ such that*

$$|A \cap I| \geq \left\lfloor \frac{2\sqrt{n}}{(1-2\alpha)}(\lambda_0(\frac{1}{2} - (\lambda_0 - \lambda) - \alpha)) \right\rfloor = \Omega(\sqrt{n}).$$

Proof: By Lemma 12.1.4 with $b = \frac{n}{2}$, $|A \cap [a, \frac{n}{2}]| \geq \lambda_0(\frac{n}{2} - a - n(\lambda_0 - \lambda)$. Divide $[a, \frac{n}{2}]$ into $\lceil \frac{n-2a}{2L} \rceil$ intervals of size $\leq L$. There must exist an interval I such that

$$|A \cap I| \geq \left\lfloor \frac{2L}{n-2a}(\lambda_0(\frac{n}{2} - a) - n(\lambda_0 - \lambda)) \right\rfloor .$$

If $L = \lceil \sqrt{n} \rceil$ and $a = \alpha n$ then

$$\begin{aligned} |A \cap I| &\geq \left\lfloor \tfrac{2L}{n-2a}(\lambda_0(\tfrac{n}{2} - a) - n(\lambda_0 - \lambda)) \right\rfloor \\ &\geq \left\lfloor \tfrac{2\sqrt{n}}{n(1-2\alpha)}(\lambda_0(\tfrac{n}{2} - \alpha n) - n(\lambda_0 - \lambda))) \right\rfloor \\ &\geq \left\lfloor \tfrac{2\sqrt{n}}{(1-2\alpha)}(\lambda_0(\tfrac{1}{2} - \alpha) - (\lambda_0 - \lambda)) \right\rfloor = \Omega(\sqrt{n}). \end{aligned}$$

▮

Theorem 12.1.6 *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

$$|A| = |A \cap B^{\text{small}}| + |A \cap B^{\text{big}}| \leq (t+1)n_0 + \lambda_0(n - |T|).$$

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

$$T(j)T(d{+}j)T(2d{+}j)T(3d{+}j)\cdots T(29d{+}j) = 00011111110000111001011111100$$

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 $T(3d+j)$, $T(4d+j)$, $T(5d+j)$, …, $T(9d+j)$ 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

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.

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 progression with spacing $d = 3$. Also note that this is a maximal arithmetic progression with spacing $d = 3$ since $0 \notin T$ and $17 \notin T$. More generally *1-blocks of T within N_j are maximal arithmetic progressions 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 progressions 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 $2y - x \le n$. 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 parameter 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 12.1.5.2.b to $n, n_0, a = \alpha n$. Hence we need the following conditions.

$$\begin{aligned} \alpha n &\ge n_0 \\ \tfrac{n}{2} &\ge n_0 \\ \tfrac{n}{2} - \alpha n &\ge \sqrt{n} \end{aligned}$$

Assuming these conditions hold, we proceed. By Lemma 12.1.5.b there is an interval $I \subseteq [\alpha n, \frac{n}{2}]$ of length $O(\sqrt{n})$ such that

$$|A \cap I| \ge \left\lfloor \frac{2\sqrt{n}}{(1 - 2\alpha)}(\lambda_0(\frac{1}{2} - \alpha) - (\lambda_0 - \lambda)) \right\rfloor = \Omega(\sqrt{n}).$$

By Lemma 12.1.3 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 $d \ge \sqrt{n}$.

For i such that $1 \le i \le k$ we define the following.

1. Define $C_0 = \{e\}$ and, for $1 \le i \le k$, 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 necessarily 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 $|A \cap [1, e]| \ge |A \cap [1, \alpha n]|$.

We want to use Lemma 12.1.4.2 on $A \cap [1, \alpha n]$. Hence we need the condition

$$n - \alpha n \ge n_0.$$

By Lemma 12.1.4

$$|T_0| \ge |A \cap [1, \alpha n]| \ge \lambda_0 \alpha n - n(\lambda_0 - \lambda) = n(\lambda_0 \alpha - (\lambda_0 - \lambda)).$$

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 $|T_k| \le n$ (a rather generous estimate) there must exist an i such that $|T_{i+1} - T_i| \le \frac{n}{k}$. 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 $\le t$ 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 $t = |T_{i+1} - T_i| \le n/k$.

To recap, we have

$$|A| \le (t+d)n_0 + \lambda_0(n - |T|)$$

with $t \le \frac{n}{k} = O(\frac{n}{\log\log n})$, $d = O(\sqrt{n})$, and $|T| \ge n(\lambda_0\alpha - (\lambda_0 - \lambda))$. Hence we have

$$|A| \le O((\frac{n}{\log\log n} + \sqrt{n})n_0) + n\lambda_0(1 - \lambda + \lambda_0 - \lambda_0\alpha).$$

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.

$$\begin{array}{rl}
\alpha n & \ge n_0 \\
\frac{n}{2} & \ge n_0 \\
\frac{n}{2} - \alpha n & \ge \sqrt{n} \\
n - \alpha n & \ge n_0 \\
\lambda_0\alpha & > \lambda_0 - \lambda \\
\lambda_0(1 - \lambda + \lambda_0 - \lambda_0\alpha) & < \lambda
\end{array}$$

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

$$\begin{array}{rl}
10^{-1}\lambda_0 & > 10^{-1}\lambda_0^2 \\
1 & > \lambda_0
\end{array}$$

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 inequalities hold. ▮

12.1.2 Analytic Proof

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 [36, 71, 72] 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 [36, 83] by a combinatorial proof. Szemeredi [84] 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 [35, 34]

Def 12.1.7 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 12.1.8 *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 12.1.9 *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 12.1.8

$$\frac{1}{n} \sum_{x,y,z \in [n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)}$$

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

$$\#(\{(x, y, z) : (x+z = 2y) \wedge x \neq y \wedge y \neq z \wedge x \neq z\}) = \frac{1}{n} \sum_{x,y,z \in [n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)} - O$$

We are not done yet. Note that $(5,10,15)$ may show up as $(15,10,5)$. Every triple appears at most twice. Hence

$\#(\{(x,y,z) : (x+z=2y) \wedge x \neq y \wedge y \neq z \wedge x \neq z\})$
$\le \quad 2\#(\{(x,y,z) : (x<y<z) \wedge (x+z=2y) \wedge x \neq y \wedge y \neq z \wedge x \neq z\})$.

Therefore

$$\frac{1}{2n} \sum_{x,y,z\in[n]} A(x)B(y)C(z) \sum_{r=1}^{n} \omega^{-r(x-2y+z)} - O(n) \le \text{ the number of 3-AP's with } x \in A,\ y \in B,\ z \in C\ .$$

▮

We will need to re-express this sum. For that we will use Fourier Analysis.

Fourier Analysis

Def 12.1.10 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 following facts.

Lemma 12.1.11 *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 progression P with difference r^{-1} (mod n) such that $\#(A\cap P)$ is large.

We need a lemma before we can proof the claim.

Lemma 12.1.12 *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 constraints 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,

$x_j - MIN \le x_j - (\lambda - \frac{m-1}{m}(\lambda+\epsilon))n$

Using the bound on x_j from constraint (2) we obtain

$$\begin{aligned} x_j - MIN &\le (\lambda+\epsilon)\tfrac{n}{m} - (\lambda - \tfrac{m-1}{m}(\lambda+\epsilon))n \\ &\le ((\lambda+\epsilon)\tfrac{1}{m} - (\lambda - \tfrac{m-1}{m}(\lambda+\epsilon)))n \\ &\le ((\lambda+\epsilon)\tfrac{1}{m} - \lambda + \tfrac{m-1}{m}(\lambda+\epsilon))n \\ &\le \epsilon n \end{aligned}$$

Note that

$$\begin{aligned}
|\sum_{j=1}^{m} x_j\omega^{s_j}| = \; & |\sum_{j=1}^{m}(x_j - MIN)\omega^{s_j} + \sum_{j=1}^{m} MIN\omega^{s_j}| \\
\le \; & |\sum_{j=1}^{m}(x_j - MIN)\omega^{s_j}| + |\sum_{j=1}^{m} MIN\omega^{s_j}| \\
\le \; & \sum_{j=1}^{m} |(x_j - MIN)||\omega^{s_j}| + MIN|\sum_{j=1}^{m} \omega^{s_j}| \\
\le \; & \sum_{j=1}^{m} \epsilon n + MIN|\sum_{j=1}^{m} \omega^{s_j}| \\
\le \; & \epsilon mn + MIN|\sum_{j=1}^{m} \omega^{s_j}| \\
\le \; & \epsilon mn + (\lambda+\epsilon)\frac{n}{m}|\sum_{j=1}^{m} \omega^{s_j}|
\end{aligned}$$

▮

Lemma 12.1.13 *Let $A \subseteq [n]$, $r \in [n]$, and $0 < \alpha < 1$. If $|\hat{A}(r)| \ge \alpha n$ and $|A| \ge \lambda n$ then there exists $m \in \mathbb{N}$, $0 < \epsilon < 1$, and an arithmetic progression P within $\mathbb{Z}_n$, of length $\frac{n}{m} \pm O(1)$ such that $\#(A \cap P) \ge (\lambda + \epsilon)\frac{n}{m}$. 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 $1 \le j \le m$ let

$$P_j = \{s \in [n] : a_j \le rs \pmod{n} < a_{j+1}\}.$$

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 progression 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.

$$\begin{aligned}
\hat{A}(r) &= \textstyle\sum_{s\in[n]} A(s)\omega^{-rs} \\
&= \textstyle\sum_{j=1}^{m} \sum_{s\in P_j} A(s)\omega^{-rs} \\
&\sim \textstyle\sum_{j=1}^{m} \sum_{s\in P_j} A(s)\omega^{-rs_j} \\
&= \textstyle\sum_{j=1}^{m} \omega^{-rs_j} \sum_{s\in P_j} A(s) \\
&= \textstyle\sum_{j=1}^{m} \omega^{-rs_j} \#(A\cap P_j) \\
&= \textstyle\sum_{j=1}^{m} \#(A\cap P_j)\omega^{-rs_j} \\
\alpha n \le |\hat{A}(r)| &= \textstyle|\sum_{j=1}^{m} \#(A\cap P_j)\omega^{-rs_j}|
\end{aligned}$$

We will not use ϵ. We intend to use Lemma 12.1.12; therefore we have the constraint $(\lambda - \frac{m-1}{m}(\lambda+\epsilon)) \ge 0$.

Assume, by way of contradiction, that $(\forall j)[|A\cap P_j| \le (\lambda+\epsilon)\frac{n}{m}$. Applying Lemma 12.1.12 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 12.1.14 *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 12.1.15 *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 12.1.14 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 12.1.11 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 12.1.5 we can take $\#(B), \#(C) \ge n\lambda/4$. We already have $\#(A) \ge \lambda n$. 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

$$\sum_{i=1}^{n-1} |\hat{B}(-2r)||\hat{C}(r)| \le (\sum_{i=1}^{n-1} |\hat{B}(-2r)|^2)^{1/2})(\sum_{i=1}^{n-1} |\hat{C}(r)|^2)^{1/2}).$$

Hence

$$\frac{\lambda^3 n^2}{16} < |\frac{1}{n}\max_{1\le r\le n-1} |\hat{A}(r)|(\sum_{i=1}^{n-1} |\hat{B}(-2r)|^2)^{1/2})(\sum_{i=1}^{n-1} |\hat{C}(r)|^2)^{1/2}).$$

By Parsaval's inequality and the definition of B and C we have

$$\sum_{i=1}^{n-1} |\hat{B}(-2r)|^2)^{1/2} \le n\#(B) = \frac{\lambda n^2}{3}$$

and

$$\sum_{i=1}^{n-1} |\hat{C}(r)|^2)^{1/2} \le n\#(C) = \frac{\lambda n^2}{3}$$

Hence

$$\frac{\lambda^3 n^2}{16} < (\max_{1\le r\le n-1} |\hat{A}(r)|)\frac{1}{n}\frac{\lambda n^2}{3} = (\max_{1\le r\le n-1} |\hat{A}(r)|)\frac{\lambda n}{3}.$$

Therefore
$|\hat{A}(r) \ge \frac{3\lambda^2 n}{16}$. ▮

12.1.3 What more is known?

The following is known.

Theorem 12.1.16 *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 [40] and Szemeredi [85]

Theorem 12.1.17 *There exists c such that $sz(n) = \Omega(n\frac{1}{(\log n)^c})$. (Szemeredi estimates $c \leq 1/20$).*

Bourgain [10] improved this further to obtain the following.

Theorem 12.1.18 *$sz(n) = \Omega(n\sqrt{\frac{\log\log n}{\log n}})$.*

12.2 Ergodic proofs of van der Waerden's Theorem

Van der Waerden [88] 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 progression of length k.*

Fürstenberg [27] 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 [36] who in turn followed the approach of [28].

12.2.1 Definitions from Topology

Def 12.2.2 X is a *metric space* if there exists a function $d{:}X \times X \rightarrow \mathbb{R}^{\geq 0}$ (called a metric) with the following properties.

1. $d(x, y) = 0$ iff $x = y$
2. $d(x, y) = d(y, x)$,

3. $d(x,y) \le d(x,z) + d(z,y)$ (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 $(\forall \epsilon > 0)(\exists i)(\forall j)[j \ge i \implies x_j \in B(x,\epsilon)]$.

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 $f{:}\mathbb{Z} \to [c]$.

2. Let $d{:}\mathrm{BISEQ} \times \mathrm{BISEQ} \to \mathbb{R}^{\geq 0}$ 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

$$\begin{aligned} \text{orbit}(x) = &\ \{T^{(i)}(x) \mid i \in \mathbb{N}\} \\ \text{dorbit}(x) = &\ \{T^{(i)}(x) \mid i \in \mathbb{Z}\} \text{ (dorbit stands for for double-orbit)} \end{aligned}$$

Def 12.2.9 Let X be a metric space, $T{:}X \rightarrow X$ be a bijection, and $x \in X$.

1. $$\text{CLDOT}(x) = \text{cl}(\{\ldots, T^{(-3)}(x), T^{(-2)}(x), \ldots, T^{(2)}(x), T^{(3)}(x), \ldots)$$

   CLDOT(x) stands for *Closure of Double-Orbit of* x.
2. x is *homogeneous* if

   $$(\forall y \in \text{CLDOT}(x))[\text{CLDOT}(x) = \text{CLDOT}(y)].$$
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] Munkres [60] 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 [60] page 178.

1. Let $f \in \mathrm{BISEQ}$ be defined by

$$f(x) = \begin{cases} \mathrm{RED} & \text{if } |x| \text{ is a square;} \\ \mathrm{BLUE} & \text{otherwise.} \end{cases} \tag{12.2}$$

The set $\{T^{(i)}(f) \mid i \in \mathbb{Z}\}$ has one limit point. It is the function

$$(\forall x \in \mathbb{Z})[g(x) = \mathrm{BLUE}].$$

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

$$\mathrm{CLDOT}(f) = \{T^{(i)}(f) \mid i \in \mathbb{Z}\} \cup \{g\}.$$

2. Let $f \in \mathrm{BISEQ}$ be defined by

$$f(x) = \begin{cases} \mathrm{RED} & \text{if } x \ge 0 \text{ and } x \text{ is a square or } x \le 0 \text{ and } x \text{ is not a square;} \\ \mathrm{BLUE} & \text{otherwise.} \end{cases} \tag{12.3}$$

The set $\{T^{(i)}(f) \mid i \in \mathbb{Z}\}$ has two limit points. They are

$$(\forall x \in \mathbb{Z})[g(x) = \mathrm{BLUE}]$$

and

$$(\forall x \in \mathbb{Z})[h(x) = \mathrm{RED}].$$

This is because their are arbitrarily long runs of REDs and arbitrarily long runs of BLUEs.

$$\mathrm{CLDOT}(f) = \{T^{(i)}(f) \mid i \in \mathbb{Z}\} \cup \{g, h\}.$$

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 $f_j \in$ 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 $f \in$ 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 $A \subseteq$ BISEQ be infinite. Let $f_1, f_2, f_3, \ldots \in A$. We construct $f \in$ 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.

$$\begin{aligned} f(a_n) &= \text{least color in } [c] \text{ that occurs infinitely often in } \{f_i(a_n) \mid i \in I_{n-1}\} \\ I_n &= \{i \mid (\forall j)[1 \leq j \leq n \implies f_i(a_j) = f(a_j)]\} \end{aligned}$$

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 $T{:}X \rightarrow X$ 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 $f,g \in$ 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

$$\begin{aligned}
d(T^{(n)}(x), y) &< \epsilon \\
d(T^{(2n)}(x), y) &< \epsilon \\
d(T^{(3n)}(x), y) &< \epsilon \\
&\vdots \\
d(T^{(rn)}(x), y) &< \epsilon \\
d(T^{((r+1)n)}(x), y) &< \epsilon
\end{aligned}$$

▮

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 $y \preceq z$, $z \in \text{CLDOT}(y)$.
Proof: Since $y \preceq z$ we have $\text{CLDOT}(z) \subseteq \text{CLDOT}(y)$. Note that

$$z \in \text{CLDOT}(z) \subseteq \text{CLDOT}(y).$$

End of Proof of Claim 1
Claim 2: For every $y \in C$ $x \in \text{CLDOT}(y)$.
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 $(\forall i)[\text{CLDOT}(y) \supseteq \text{CLDOT}(y_i)]$. Hence $(\forall i)[y_i \in \text{CLDOT}(y)]$. Since $\lim_i y_i = x$, $(\forall i)[y_i \in \text{CLDOT}(y)]$, and $\text{CLDOT}(y)$ is closed under limit points, $x \in \text{CLDOT}(y)$.
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 $y \in \text{CLDOT}(x)$. We show $\text{CLDOT}(y) = \text{CLDOT}(x)$.

Since $y \in \text{CLDOT}(x)$, $\text{CLDOT}(y) \subseteq \text{CLDOT}(x)$ by Lemma 12.2.28.

Since x is maximal $\text{CLDOT}(x) \subseteq \text{CLDOT}(y)$.

Hence $\text{CLDOT}(x) = \text{CLDOT}(y)$.
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 progression of length k.*

Proof:

Let BISEQ and T be as in Example 12.2.6.2.

Let $f \in \text{BISEQ}$. Let $Y = \text{CLDOT}(f)$. 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 $\text{CLDOT}(g)$

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 progression of length k.

▮

12.3 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, χ:$\mathbb{R} \rightarrow \mathbb{N}$ 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 [46] claims that Erdős proved this result. The prove we give is due to Davies [18].

Def 12.3.1 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 12.3.2 ω_1 is the first uncountable ordinal. ω_2 is the second uncountable ordinal.

Fact 12.3.3

1. *If CH is true, then there is a bijection between $\mathbb{R}$ and ω_1. This has the following 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 $\beta \in [\omega_1, \omega_2)$.

12.3.1 CH $\implies$ FALSE

Def 12.3.4 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 12.3.5 $X \subseteq CL(X)$ since we can take $b = c$.

Lemma 12.3.6

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 12.3.7 *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 12.3.3.1, 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 12.3.6.1.
2. $Y_0 \subset Y_1 \subset Y_2 \subset Y_3 \subset \cdots \subset Y_\alpha \subset \cdots$. This is from Lemma 12.3.6.2.
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 12.3.7 work? Absolutely nothing:

Theorem 12.3.8 *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 12.3.9 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 12.3.7 using this new definition of *CL*. Let χ be defined as in Theorem 12.3.7.

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. ▮

BILL- FILL IN -LOOK UP PAPER THIS CAME FROM TO GET MORE

12.3.2 $\neg$ CH $\implies$ TRUE

Theorem 12.3.10 *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 12.3.3 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).

12.3.3 More is Known!

To state the generalization of this theorem we need a definition.

Def 12.3.11 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* $\chi{:}\mathbb{R} \rightarrow \mathbb{N}$ *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 12.3.7 and 12.3.10 we obtain the following.

Theorem 12.3.12 $e_1 + e_2 = e_3 + e_4$ *is regular iff* $2^{\aleph_0} > \aleph_1$.

Jacob Fox [25] has generalized this to prove the following.

Theorem 12.3.13 *Let* $s \in \mathbb{N}$. *The equation*

$$e_1 + se_2 = e_3 + \cdots + e_{s+3} \tag{12.7}$$

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 (12.7) 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 13

The Generalized Polynomial van der Waerden Theorem

Bibliography

[1] J. Avigad. Metamathematics of ergodic theory. *Annals of Pure and Applied Logic*, 157:64–76, 2009. `http://www.andrew.cmu.edu/user/avigad/`.

[2] J. Avigad and H. Towsner. Metastability in the Furstenberg-Zimmer tower. *Fundamenta Mathematicae*, 210:243–268, 2010. `http://www.andrew.cmu.edu/user/avigad/`.

[3] J. Beck. Van der Waerden and Ramsey type games. *Combinatorica*, 1, 1981. `http://www.springer.com/new+%26+forthcoming+titles+%28default%29/journal/493`.

[4] J. Beck. Positional games and the second moment method. *Combinatorica*, 22, 2002. `http://www.springer.com/new+%26+forthcoming+titles+%28default%29/journal/493`.

[5] J. Beck. *Combinatorial games: Tic-Tac-Toe Theory*. Cambridge Press, 2008.

[6] J. Beck. Surplus of graphs and the local Lovasz lemma. In *Building bridges between math and computer science*, pages 47–103, New York, Heidelberg, Berlin, 2008. Bolyai society mathematical studies number 19.

[7] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *Journal of the American Mathematical Society*, 9:725–753, 1996. `http://www.math.ohio-state.edu/~vitaly/` or `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[8] 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.

[9] R. G. Bierstedy and W. H. Mills. On the bound for a pair of consecutive quartic residues of a prime. *Proceedings of the American Mathematical Society*, 14:628–632, 1963. http://www.jstor.org/stable/2033874 or http://www.cs.umd.edu/~gasarch/res/.

[10] J. Bourgain. On triples in arithmetic progression. *Geometric and Functional Analysis*, 9:968–984, 1999. http://www.springer.com/new+%26+forthcoming+titles+%28default%29/journal/493.

[11] T. Brown. Some new van der Waerden numbers (preliminary report). *Notices of the American Mathematical Society*, page 432, 1974.

[12] A. Caicedo. Goodstein's function. *Revista Columbiana de Matematicas*, 2:381–391, 2007. The article is in English and its here: http://www.scm.org.co/Articulos/878.pdf.

[13] 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. Math Reviews 266891.

[14] E. Cichon. A short proof of two recently discovered independence results using recursive theoretic methods. *Proceedings of the American Mathematical Society*, 87:704–706, 1983. http://www.jstor.org/stable/2043364.

[15] G. Cornacchia. Sulla congruenza $x^n + y^n + z^n \equiv 0 \pmod{p}$. *Giornale di matematiche di Battaglini*, pages 219–268, 1909. http://www.cs.umd.edu/~gasarch/res/.

[16] H. Davenport. On the distribution of quadratic residues mod p. *Journal of the London Mathematical Society*, 6:49–54, 1932. http://jlms.oxfordjournals.org/. This is part 1.

[17] H. Davenport. On the distribution of quadratic residues mod p. *Journal of the London Mathematical Society*, 8:46–52, 1933. http://jlms.oxfordjournals.org/. This is part 2.

[18] R. O. Davies. Partioning the plane into denumerably many sets without repeated differences. *Proceedings of the Cambridge Philosophical Society*, 72:179–183, 1972.

[19] L. E. Dickson. Lower limit for the number of sets of solutions of $x^n + y^n + z^n \equiv 0 \pmod{p}$. *Journal für die reine und angewandte Mathematik*, 135:181–189, 1909. `http://www.cs.umd.edu/~gasarch/res/`.

[20] M. Dunton. Bounds for pairs of cubic residues. *Proceedings of the American Mathematical Society*, 16:330–332, 1965. Online at `http://www.jstor.org/stable/2033874` or `http://www.cs.umd.edu/~gasarch/res/`.

[21] H. Edwards. *Fermat's Last Ttheorem: A genetic introduction to algebraic number theory.* Springer, New York, 2000.

[22] P. Erdős. Some unsolved problems. *Michigan Mathematical Journal*, 4:291–300, 1957. `http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.mmj`.

[23] P. Erdős and R. Graham. *Old and New Problems and results in combinatorial Number Theory.* Academic Press, 1980. http://www.renyi.hu/perdos/Erdos.html.

[24] P. Erdős and P. Turán. On some sequences of integers. *Journal of the London Mathematical Society*, 11(2):261–264, 1936. `http://jlms.oxfordjournals.org/`.

[25] J. Fox. An infinite color analogue of Rado's theorem. *Journal of Combinatorial Theory, Series A*, 114:1456–1469, 2007. `http://math.mit.edu/~fox/~publications.html`.

[26] H. Fürstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi's on arithmetic progressions. *Journal d'Analyse Mathematique*, 31:204–256, 1977. `http://www.cs.umd.edu/~gasarch/vdw/furstenbergsz.pdf`.

[27] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, 1981.

[28] H. Furstenberg and B. Weiss. Topological dynamics and combinatorial number theory. *Journal d'Analyse Mathematique*, 34:61–85, 1978.

[29] M. Gardner. Bulgarian solitaire and other seemingly endless tasks. In *The Colossal Book of mathematics*, pages 455–470, 2001.

[30] W. Gasarch and B. Haeupler. Lower Bounds on van der Waerden Numbers: Randomized-and Deterministic-Constructive. *Electronic Journal of Combinatorics*, 18(P64):1, 2011. `http://www.combinatorics.org`.

[31] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame Journal of Formal Logic*, 49:431–446, 2008. `http://folk.uio.no/philipge/`.

[32] J.-Y. Girard. *Proof theory and logic complexity (Vol I).* Elsevier, 1990.

[33] R. Goodstein. On the restricted ordinal theorem. *Journal of Symbolic Logic*, 9:33–41, 1944. `http://www.jstor.org/action/showPublication?journalCode=jsymboliclogic`.

[34] 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.springer.com/new+%26+forthcoming+titles+%28default%29/journal/493` or `http://www.dpmms.cam.ac.uk/~wtg10/papers.html`.

[35] 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.springer.com/new+%26+forthcoming+titles+%28default%29/journal/493`.

[36] R. Graham, B. Rothschild, and J. Spencer. *Ramsey Theory*. Wiley, New York, 1990.

[37] R. Graham and J. Solymosi. Monochromatic equilateral right triangles on the integer grid. *Topics in Discrete Mathematics, Algorithms and Combinatorics*, 26, 2006. `www.math.ucsd.edu/~/ron/06\_03\_righttriangles.pdf` or `www.cs.umd.edu/~/vdw/graham-solymosi.pdf`.

[38] 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.

[39] A. Hales and R. Jewett. Regularity and positional games. *Transactions of the American Math Society*, 106, 1963.

[40] D. Heath-Brown. Integer sets containing no arithmetic progressions. *Proceedings of the London Mathematical Society*, 35(2):385–394, 1987. http://www.cs.umd.edu/~gasarch/vdw/heathbrown.pdf.

[41] D. Hilbert. Uber die irreduzibilitat ganzer rationalen funktionen mit ganzzahligen koeffizienten. *Journal fur die reine und angewandte Mathematik.*, 110:104–129, 1892.

[42] A. Hurwitz. U¨ber die kongruenz $ax^e + by^e + cz^e \equiv 0 \pmod{p}$. *Journal für die reine und angewandte Mathematik*, 136:272–292, 1909. http://www.cs.umd.edu/~gasarch/res/.

[43] A. Kanamori and K. McAloon. On Gödel incompleteness and finite combinatorics. *Annals of Pure and Applied Logic*, 33(1):23–41, 1987. http://math.bu.edu/people/aki/k.pdf.

[44] L. Kirby and J. Paris. Accessible independent results for Peano arithmetic. *Bulletin of the London Mathematical Society*, 14:285–293, 1982. http://blms.oxfordjournals.org/content/by/year.

[45] U. Kohlenbach. *Applied proof theory: Proof interpretations and their use in mathematics*. Springer, New York, 2008.

[46] P. Komjáth. Partitions of vector spaces. *Periodica Mathematica Hungarica*, 28:187–193, 1994.

[47] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications to graph theory. In *Theoretical aspects of computer science*, Lecture Notes in Computer Science, 2002. http://www.springerlink.com/content/vn0t7870ctnb6man/.

[48] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications to graph theory, 1996. Article at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.2310.

[49] M. Kouril and J.Franco. Resolution tunnels for improving SAT solver performance. In *Eighth International Conference on theory and application of satisfiability testing*, volume 8, pages 143–157, 2005.

[50] M. Kouril and J. Paul. The van der Waerden number w(2,6) is 1132. *Experimental Mathematics*, 17:53–61, 2008. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`.

[51] B. Landman and A. Robertson. *Ramsey Theory on the integers*. AMS, Providence, 2004.

[52] R. Laubenbacher and D. Pengelley. *Voice ce que jái trouve* sophie germain's grand plan to prove Fermat's Last Theorem. *Historica Mathematica*, 37:641–692, 2010. `http://www.math.nmsu.edu/~history/germain.html`. The article is in English.

[53] D. H. Lehmer and E. Lehmer. On runs of residues. *Proceedings of the American Mathematical Society*, 13:102–106, 1962. `http://www.jstor.org/stable/2033781` or `http://www.cs.umd.edu/~gasarch/res/`.

[54] D. H. Lehmer, E. Lehmer, and W. Mills. Pairs of consecutive power residues. *Canadian Journal of Mathematics*, 15:172–177, 1963. `http://cms.math.ca/cjm/` or `http://www.cs.umd.edu/~gasarch/res/`.

[55] D. H. Lehmer, E. Lehmer, W. Mills, and J.L.Selfridge. Machine proof of a theorem on cubic residues. *Mathematics of Computation*, 16:407–415, 1962. `http://cms.math.ca/cjm/`.

[56] T. Lehrer. That's mathematics, 1995. Full version is on *Dr. Demento Basement Tapes Number 4* from 1995 or on You-Tube. Full version is not in *The Remains of Tome Lehrer* which is supposed to be a boxed set of all of his works.

[57] G. Libri. Mem̈oire sur la theörie des nombres. *Journal für die reine und angewandte Mathematik*, 9:54–81, 1832. `http://www.cs.umd.edu/~gasarch/res/`. The article is in French.

[58] N. Lyall. Roth's Theorem: the Fourier analytic approach. Unpublished manuscript. `http://www.math.uga.edu/~lyall/REU/index.html`.

[59] K. McAloon. Diagonal methods and strong cuts in models of arithmetic. In *Logic Colloquium; 1978*, pages 171–181, 1978.

[60] J. Munkres. *Topology: A first course.* Prentice-Hall, 1975.

[61] J. Paris and L. Harrington. A mathematical incompleteness in Peano arithmetic. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 1133–1142. North-Holland, Amsterdam, 1977.

[62] A. E. Pellet. Meḿorire sur la theórie algeb́rique des eq́uations. *Bulletin de la societe Mathematique de France*, 15:61–103, 1887. `http://www.cs.umd.edu/~gasarch/res/`.

[63] P. Pepin. Et́ude sur la theórie des reśidues cubiques. *Journal de Mathématiques Pures et Appliquées*, 2:313–324, 1876. `http://www.cs.umd.edu/~gasarch/res/`.

[64] P. Pepin. Sur divers tentatives de deḿonstration du theóreḿe de Fermat. *Comptes Rendus de l'Académie des Sciences Paris*, 91:366–367, 1880. `http://www.cs.umd.edu/~gasarch/res/`.

[65] R. Peralta. On the distribution of quadratic residues and nonresidues modulo a prime number. *Mathematics of Computation*, 58:433–440, 1992. `http://www.jstor.org/stable/2153045` or `http://www.cs.umd.edu/~gasarch/res/`.

[66] H. J. Prömel and V. Rödl. 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`.

[67] Radhakrishnan and Srinivasan. Improved bounds and algorithms for hypergraph 2-coloring. *Random Structures and Algorithms*, 16:4–32, 2000.

[68] R. Rado. Studien zur Kombinatorik. *Mathematische Zeitschrift*, 36:424–480, 1933. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`. Includes Gallai's theorem and credits him.

[69] R. Rado. Notes on combinatorial analysis. *Proceedings of the London Mathematical Society*, 48:122–160, 1943. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`. Includes Gallai's theorem and credits him.

[70] L. Riddle. Sophie Germain and Fermat's Last Theorem. http://www.agnesscott.edu/lriddle/women/germain-FLT/SGandFLT.htm.

[71] K. Roth. Sur quelques ensembles d' entiers. *C.R. Acad. Sci Paris*, 234:388–3901, 1952.

[72] K. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, 28:104–109, 1953. http://jlms.oxfordjournals.org/. For exposition see http://www.math.missouri.edu/~iosevich/expositorypapers.html.

[73] 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.

[74] I. Schur. Uber die kongruenz of $x^m + y^m \equiv z^m \pmod{p}$. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 25:114–116, 1916.

[75] Shelah. A partition theorem. *Scientiae Math Japonicae*, 56:413–438, 2002. Paper 679 at the Shelah Archive: http://shelah.logic.at/short600.html.

[76] S. Shelah. Primitive recursive bounds for van der Waerden numbers. *Journal of the American Mathematical Society*, 1:683–697, 1988. http://www.jstor.org/view/08940347/di963031/96p0024f/0.

[77] R. Smullyan. Trees and ball games. *Annals of the New York Academy of Sciences*, 321:86–90, 1979.

[78] A. Soifer. *The mathematical coloring book: mathematics of coloring and the colorful life of its creators.* Springer-Verlag, New York, Heidelberg, Berlin, 2009.

[79] A. Sperotto and M. Pelilo. Szemerédi's regularity lemma and its applications to pairwise clustering and segmentation. In *Energy minimization methods in computer science and pattern recognition*, volume 4679 of *Lecture Notes in Computer Science*, New York, 2007. Springer. Article at http://www.springerlink.com/content/u11535301j31g123/.

[80] R. Stevens and R. Shantaram. Computer-generated van de Waerden partitions. *Mathematics of Computation*, 32:635–636, 1978. http://www.jstor.org/action/showPublication?journalCode=mathcomp.

[81] D. Surendan. Fermat's last theorem (website). http://www.uz.ac.zw/science/maths/zimaths/flt.htm.

[82] 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.

[83] E. Szeméredi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Sci. Hung.*, 20:89–104, 1974. http://www.cs.umd.edu/~gasarch/vdw.

[84] E. Szeméredi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:299–345, 1975. http://www.cs.umd.edu/~gasarch/vdw/szdensity.pdf.

[85] E. Szemerédi. Integer sets containing no arithmetic progressions. *Acta Math. Sci. Hung.*, 56:155–158, 1990. http://www.cs.umd.edu/~gasarch/vdw/szlog.pdf.

[86] R. Taylor and A. Wiles. Ring-theoretic properties of certain Hecke algebras. *Annals of Mathematics*, 141:553–572, 1995. http://math.stanford.edu/~lekheng/flt/taylor-wiles.pdf.

[87] H. Towsner. Metastability in the Furstenberg-Zimmer tower II, 2009. http://arxiv.org/abs/0909.5668.

[88] B. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Arch. Wisk.*, 15:212–216, 1927. This article is in Dutch and I cannot find it online.

[89] B. van der Waerden. How the proof of Baudet's conjecuture was found. In L. Mirsky, editor, *Studies in Pure Math*, pages 251–260. Academic Press, 1971.

[90] 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.

[91] A. Wiles. Modular elliptic curves and Fermat's Last Theorem. *Annals of Mathematics*, 141:443–551, 1995. `http://math.stanford.edu/~lekheng/flt/wiles.pdf`. The paper itself has a typo and says it is in volume 142. This is not correct. It is in volume 141.

[92] E. Witt. Ein kombinatorischer satz de elementargeometrie. *Mathematische Nachrichten*, 6:261–262, 1951. `http://www.cs.umd.edu/~gasarch/vdw/vdw.html`. Contains Gallai-Witt Theorem, though Gallai had it first so it is now called Gallai's theorem.