van der Waerden's theorem

February 21, 2023

1 t-ary (or base t) representation of natural numbers

This is general formulation of the familiar decimal and binary representation of natural numbers. Many of you already know this.

Definition. Let us be given a $t \geq 2$ positive integer, that we call the *base*, and an n natural number. We say that n is written in t-ary (or base t) representation, if for some k

$$n = \sum_{i=0}^{k} a_i t^i, \text{ where } a_i \in \{0, 1, 2, ..., t-1\} \text{ and } a_k \neq 0.$$

Claim 1 *Every n positive integer has a base t representation, furthermore, k and $a_0, ..., a_k$ are unique.*

We prove by induction on n that n has t-ary representation. The base cases are $0 \leq n \leq t-1$. Then we can take $k = 0$ and $a_0 = n$. So assume that $n \geq t$. Apply division with remainder to obtain

$$n = qt + r, \text{ where } 0 \leq r \leq t-1.$$

By the induction hypothesis, $q < n$, and therefore q has t-ary representation $q = \sum_{i=0}^{\ell} b_i t^i$, with $b_i \in \{0, 1, 2, ..., t-1\}$. Now $k = \ell + 1$, $a_0 = r$, and for $i = 1, 2, ..., k$ choosing $a_i = b_{i-1}$ provides the required t-ary representation for n.

Now we have to show the uniqueness of the representation. Assume that

$$\sum_{i=0}^{k} a_i t^i = n = \sum_{i=0}^{\ell} c_i t^i, \tag{1.1}$$

where $a_k \neq 0$ and $a_\ell \neq 0$. (Occasionally it is convenient to extend the representation with zero "digits" at the higher powers, and it should not cause any confusion. We do so now.) If these two representations are different, then there must be a largest m such that $a_m \neq c_m$. Assume without loss of generality that $a_m < c_m$. We obtain from (1.1)

$$\sum_{i=0}^{m-1} (a_i - c_i) t^i = (c_m - a_m) t^m.$$

However, the last identity cannot be true as

$$\sum_{i=0}^{m-1}(a_i - c_i)t^i \leq \sum_{i=0}^{m-1}(t-1)t^i = (t-1)\frac{t^m-1}{t-1} < t^m \leq (c_m - a_m)t^m.$$

2 Full Ramsey type theorems

If you recall Ramsey theory, in many cases we colored a big underlying set (plane, space, or integers) and obtained a handful of monochromatic objects with some structure, like a triangle congruent to a prescribed one or a solution to some equation. The set Ramsey theorem, however, found a monochromatic substructure that was just different in size. There are just a few of these kind of results, and they are called *full Ramsey type theorems.*

One of the earliest examples was conjectured by Baudet and proved by van der Waerden (1927):

Theorem 1 *For all p, k positive integers there exists a $W(p,k)$ number, such that for all $n \geq W(p,k)$ and every $f : [n] \to [k]$ coloration there exists a monochromatic p-term arithmetic progression.*

Note that the story can be told as k-coloring a sufficiently long arithmetic progression necessarily yields a monochromatic p-term arithmetic progression. For brevity, we write AP for arithmetic progression.

Another full Ramsey type theorem was conjectured by Rota and proved by Graham, Leeb, and Rothschild. This is the q-analogue of the set Ramsey theorem:

Theorem 2 *For all q prime powers and r, k, m positive integers, there exists an $n_0(q, r, k, m)$ number, such that for all $n \geq n_0(q, r, k, m)$ and every c r-coloration of the k-dimensional subspaces of any n-dimensional linear space V over $GF(q)$, there exists an m-dimensional subspace U of V, such that c is constant on the k-dimensional subspaces of U.*

We postpone the proof of Theorem 1 and will skip the proof of Theorem 2.

3 Where van der Waerden differs from Ramsey

One difference is that van der Waerden's Theorem does not allow an infinite version. There is an obvious 2-coloration of $\mathbb{N}$ such that no colorclass contains an infinite arithmetic progression: $1^R, 2^B, 3^B, 4^R, 5^R, 6^R, 7^B, 8^B, 9^B, 10^B, 11^R, ...$

Another difference is that the set Ramsey theorems do not have a "density reason". To clarify what does it mean, consider the Ramsey number $R_3^2(3,3,3)$. This Ramsey number means that 3-coloring the edges of a complete graph with this many vertices, there should be a K_3, whose edges are monochromatic. A "density reason" would be that one-third of the edges of a sufficiently large complete graph must contain a K_3. This is obviously not the case, as a complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$ has roughly half of the number of edges of K_n but does not contain a K_3.

Erdős and Turán conjectured early on that unlike the set Ramsey theorem, van der Waerden's theorem has a density version. Arriving to a proof, however, took several decades.

Let us define

$$r_p(n) = \max\{m : \exists\ 1 \le a_1 < \cdots < a_m \le n \text{ with no } p\text{-term AP}\},$$

i.e. the largest number of integers in $[1, n]$ such that they do not contain a p-term AP. It is easy to see that if $r_p(n) = o(n)$ for some p, then for any fixed k, for a sufficiently large n $r_p(n) < \frac{n}{k}$, and k-coloring $[n]$, one color class will be at least $\frac{n}{k} > r_p(n)$, and hence must contain a p-term AP. Indeed, van der Waerden has a density version!

At least, it is easy to prove that $\lim_{n\to\infty} \frac{r_p(n)}{n}$ exists, so the issue is "only" whether this limit is zero or positive (see below). For $p = 3$, it was solved by Roth (1953, $r_3(n) = O\big(\frac{n}{\log\log n}\big)$) and was cited in his Fields Medal, for $p = 4$ and subsequently for arbitrary p it was solved by Szemerédi (1969, 1975), who received his Abel Prize recently, in part for these results. Szemerédi discovered his Regularity Lemma as a tool for the proof. Fürstenberg (1977) made an ergodic theory proof to the Szemerédi theorem. Tao (2004) made the ergodic theory proof elementary and avoided using AC. An effective upper bound showing the zero limit was found by Gowers (1998), and cited in his Fields Medal. Gowers showed $r_p(n) \le \frac{n}{(\log\log n)^{c_p}}$, with some constant c_p dependent only on p.

First note that $r_p(n) \ge 1$ for $n \ge 1$ and the $r_p(n)$ sequence is *subadditive*, i.e.

$$r_p(n) + r_p(m) \ge r_p(n + m).$$

Indeed, select a subset A of size $r_p(n + m)$ of $[n + m]$ without p-term AP's. Then $A \cap [n]$ and $A \cup \{n + 1, ..., n + m\}$ are also without p-term AP's, and their sizes set lower bound for $r_p(n)$ and $r_p(m)$.

Lemma 3 [Fekete's Lemma] *If an a_n non-negative sequence ($n \ge 1$) is subadditive, i.e. $a_n + a_m \ge a_{n+m}$, then $\lim_{n\to\infty} \frac{a_n}{n}$ exists and equals to $\inf_{n\ge1} \frac{a_n}{n}$.*

Take an arbitrary $\epsilon > 0$ and fix a k such that $0 \le \frac{a_k}{k} - \inf_{n\ge1} \frac{a_n}{n} < \frac{\epsilon}{2}$. Define m such that for all $r = 1, 2, .., k - 1$, the inequality $\frac{a_r}{km} < \frac{\epsilon}{2}$ holds. We are going to show that for any $n \ge km$ the fraction $\frac{a_n}{n}$ is closer to its expected limit, $\inf_{n\ge1} \frac{a_n}{n}$, then ϵ.

Using division by remainder, write $n = sk + r$ with $r = 0, 1, 2, .., k - 1$.

$$\begin{aligned}
\inf_{n\ge1} \frac{a_n}{n} &\le \frac{a_n}{n} \le \frac{sa_k}{sk + r} + \frac{a_r}{sk + r} \\
&\quad \text{(by subadditivity; for } r = 0 \text{ the second term simply does not show!)} \\
&\le \frac{a_k}{k} + \frac{a_r}{km} < \left(\inf_{n\ge1} \frac{a_n}{n} + \frac{\epsilon}{2}\right) + \frac{\epsilon}{2} = \epsilon + \inf_{n\ge1} \frac{a_n}{n}.
\end{aligned}$$

Behrend gave a construction showing that $r_3(n) > n^{1-\frac{c}{\sqrt{\log n}}}$. As $n^{-\frac{c}{\sqrt{\log n}}} = e^{-c\sqrt{\log n}}$ goes to zero very slowly, his construction shows that proving $r_3(n) = o(n)$ must be delicate. To make this construction, one needs a subset $B \subset \{0, 1, ..., n - 1\}$, such that

- B has no 3-term arithmetic progression, and
- $B \ge n^{1-\frac{c}{\sqrt{\log n}}}$ for a fixed $c > 0$ for sufficiently large n.

For a positive even integer t (to be specified later), write the numbers $0 \le a \le n-1$ in t-ary representation. If $t^k \le n < t^{k+1}$, then $k+1$ digits suffice for every such a. As $k \le \frac{\log n}{\log t} = \log_t n < k+1$, $k+1 \le 1+\frac{\log n}{\log t}$ is an upper bound for the number of digits that we use for any $0 \le a \le n-1$. We add superfluous zero digits in case some number a needs less than $k+1$ digits, and write it as

$$a = a_0 + a_1 t + a_2 t^2 + \ldots + a_k t^k.$$

Define a set A by

$$A = \left\{ a \in \{0, 1, \ldots n-1\} : \forall i \in \{0, 1, \ldots, k\}\ 0 \le a_i < \frac{t}{2} \right\}.$$

Note that as t is even, the number of possible digits that a can use is $t/2$. The significance of the definition of A is that for two elements of A, their sum can be computed in t-ary representation without carry-over, i.e. just by adding up the corresponding digits. Define a function $s : A \to \mathbb{N}$ by $a \mapsto \sum_{i=0}^{k} a_i^2$. As the range of s cannot contain numbers bigger than $(k+1)(t/2)^2 \le \left(1+\frac{\log n}{\log t}\right)\frac{t^2}{4}$, there is a ℓ such that

$$|\{a \in A : s(a) = \ell\}| \ge \frac{|A|}{1+\left(1+\frac{\log n}{\log t}\right)\frac{t^2}{4}}. \tag{3.2}$$

(The $+1$ is in the denominator, since 0 is also in the range of s.) Choose

$$B = \{a \in A : s(a) = \ell\}$$

for such a ℓ. First we verify that B has no 3-term arithmetic progression (this has nothing to do with the choice of ℓ, and works for any ℓ). Assume for contradiction that $a, b, c \in B$ are different numbers and $\frac{a+b}{2} = c$. Then $s(a) = s(b) = s(c) = \ell$ and $c_i = \frac{a_i+b_i}{2}$ (here we use that addition of two elements of A creates no carry-over). We obtain

$$\ell = \sum_i c_i^2 = \sum_i \Big(\frac{a_i+b_i}{2}\Big)^2 = \frac{1}{4}\sum_i a_i^2 + \frac{1}{4}\sum_i b_i^2 + \frac{1}{2}\sum_i a_i b_i = \frac{\ell}{2} + \frac{1}{2}\sum_i a_i b_i,$$

and using the Cauchy–Schwartz inequality,

$$\ell = \sum_i a_i b_i \le \sqrt{\sum_i a_i^2} \cdot \sqrt{\sum_i b_i^2} = \sqrt{\ell} \cdot \sqrt{\ell} = \ell.$$

If the Cauchy–Schwartz inequality holds with equality, then there must be a number λ such that $\lambda a_i = b_i$ for all i. However, $s(a) = s(b)$ implies $\lambda = 1$, therefore a, b, c were not three different numbers, a contradiction.

Estimating $|B|$, first note that

$$|A| \ge \Big(\frac{t}{2}\Big)^k \ge \Big(\frac{t}{2}\Big)^{\frac{\log n}{\log t}-1}$$

and by (3.2)

$$|B| \geq \frac{\left(\frac{t}{2}\right)^{\frac{\log n}{\log t}-1}}{1+\left(1+\frac{\log n}{\log t}\right)\frac{t^2}{4}}.$$

Now we select an even integer t with $t = (1+o(1))e^{\sqrt{\log n}}$. (We may be more specific taking for t the even integer closest to $e^{\sqrt{\log n}}$, if we want it.) Now $\log t = \log\big((1+o(1))e^{\sqrt{\log n}}\big) = \log(1+o(1)) + \sqrt{\log n} = \sqrt{\log n} + o(1)$.

We use the following estimates:

$$\frac{t^2}{4} = \frac{1+o(1)}{4}e^{2\sqrt{\log n}};$$

for sufficiently large n

$$\frac{\log n}{\log t} - 1 \geq \sqrt{\log n} - 2;$$

$$1 + \frac{\log n}{\log t} = \sqrt{\log n} + O(1);$$

and

$$\left(\frac{t}{2}\right)^{\sqrt{\log n}-2} = e^{(\log t - \log 2)(\sqrt{\log n}-2)} = e^{(\sqrt{\log n}-\log 2+o(1))(\sqrt{\log n}-2)}$$

$$= \quad e^{\log n-(2+\log 2+o(1))\sqrt{\log n}+O(1)} = \frac{n}{e^{(2+\log 2+o(1))\sqrt{\log n}}}.$$

We also use the obvious $\sqrt{\log n} = e^{o(\sqrt{\log n})}$.

From here, for sufficiently large n, the construction works with $c = 5$, as we have

$$|B| \geq \frac{\left(\frac{t}{2}\right)^{\sqrt{\log n}-2}}{\frac{1+o(1)}{4}e^{2\sqrt{\log n}} \cdot \left(\sqrt{\log n} + O(1)\right)} \geq \frac{n}{e^{5\sqrt{\log n}}}.$$

It is remarkable that $r_3(n)$ shows up in combinatorial problems that seemingly have nothing to do with arithmetic progressions. Ruzsa and Szemerédi (1976) proved

Theorem 4 *If T is a family of 3-element subsets of n elements, such that no 6 elements carry 3 or more members of T, then $|T| = o(n^2)$.*

We assume that 4 divides n. We make a construction of many 3-element subsets of a $3n$ element set, such that no 6-element subset carries 3 of the 3-element subsets. Take $1 \leq a_1 < \ldots < a_m \leq n/4$ with $m = r_3(n/4)$. Take following lattice points from the plane:

$$V = \{(i,j) : 0 \leq i \leq 2,\ 1 \leq j \leq n\}.$$

Now all the lines $y = a_k x + b$ $(1 \leq b \leq n/2)$ have exactly 3 points in V. There are $\frac{mn}{2}$ such lines, and they define the same number of 3-element subsets. Any two subsets have at most 1 point in common, as their defining lines already have this property. The only way to have 3 such sets on 6 points would be their defining lines making a triangle with the 3 vertices of the triangle in V. Then their slopes (in some order) would show $2a_i = a_j + a_k$, a contradiction with the choice of the slopes.

3.1 Hales–Jewett Theorem

We show a generalization of the van der Waerden theorem, which is purely combinatorial and not number theoretic. Let A be a finite alphabet, $|A| = t$. We are concerned with n-letter words over A. We denote the set of n-letter words over A by A^n.

A key concept will be the *combinatorial line* in A^n. For $\emptyset \neq I \subseteq [n]$ and fixed letters c_j for $j \notin I$, the combinatorial line $L_{I;c_j:j\notin I}$ is defined as set of words

$$L_{I;c_j:j\notin I} = \{w_a \in A^n : a \in A\},$$

where, using the notation $(w_a)_i$ for the i^{th} letter of the word w_a,

$$(w_a)_i = \begin{cases} a, & \text{if } i \in I \\ c_i, & \text{if } i \notin I. \end{cases}$$

So every line has t words, letters in positions not belonging to I are the same constants in every word of the line, and in positions belonging to I they are equal, and in different words of the line they show all possible letters from the alphabet.

For example, if $n = 2$ and $A = \{a, b, c, d\}$, the combinatorial lines are the following:
$\{aa, ab, ac, ad\}; \{ba, bb, bc, bd\}; \{ca, cb, cc, cd\}; \{da, db, dc, dd\};$
$\{aa, ba, ca, da\}; \{ab, bb, cb, db\}; \{ac, bc, cc, dc\}; \{ad, bd, cd, dd\}; \{aa, bb, cc, dd\}.$
The Hales–Jewett theorem is a "purely combinatorial" generalization of the van der Waerden theorem, which does not require any additive structure.

Theorem 5 *For every finite alphabet A with $|A| = t$ and every positive integer r, there is a number $HJ(t, r)$, such that for every $n \geq HJ(t, r)$, for every r-coloration of A^n, there exists a monochromatic combinatorial line.*

There is a related "density result" behind this theorem. We only show how Hales–Jewett implies van der Waerden.

Assume that the integers in $[0, p^n - 1]$ are colored with k colors, where $n \geq HJ(p, k)$, the threshold in the Hales–Jewett theorem. We will show that this coloration has a p-term monochromatic arithmetic progression. First, write all the integers in $[0, p^n - 1]$ in p-ary representation, like $a = a_0 + a_1 p + a_2 p^2 + ... + a_{n-1} p^{n-1}$, representing missing digits by 0's. Introducing the alphabet $A = \{0, 1, ..., p - 1\}$, there is a bijection

$$a \in [0, p^n - 1] \leftrightarrow (a_0, a_1, ..., a_{n-1}) \in A^n$$

by the p-ary representation of integers, under which the k-coloration of $[0, p^n - 1]$ corresponds to a k-coloration of A^n. By the Hales–Jewett theorem, this k-coloration of A^n has a monochromatic combinatorial line. This monochromatic combinatorial line corresponds to a special type of arithmetic progression, whose elements have the very same color. In this special type of arithmetic progression, certain digits are fixed, and the digits that are not fixed are all the same and run through the set of all digits. Here is an example for $p = 10$, $n = 6$, $I = \{2, 4\}$, $\bar{I} = \{0, 1, 3, 5\}$:

$$120309, 121319, 122329, 123339, 124349, 125359, 126369, 127379, 128389, 129399.$$

Which p-term arithmetic progressions in $[0, p^n - 1]$ arise from combinatorial lines? The answer is among the homework problems.

4 Proof of the van der Waerden theorem

We present the proof of Graham and Rothschild. A basic tool of the proof is the concept of $\sim_\ell$-equivalence classes of m-sequences. The underlying set is length m sequences composed of the entries $0, 1, 2, ..., \ell$. Two such sequences are considered equivalent, if they agree up to the last occurrence of ℓ in any of them. This relation is clearly reflexive, symmetric and transitive, hence is an equivalence relation. In particular, we consider any two ℓ-free sequences $\sim_\ell$-equivalent.

For an example, we list the $\sim_2$-equivalence classes of 2-sequences. The possible entries are 0,1,2, and the sequences are of length 2: $\{00, 01, 10, 11\}; \{02\}; \{12\}; \{22\}; \{20, 21\}$. The $\sim_p$-equivalence classes of 1-sequences will turn important, its classes are: $\{0, 1, ..., p-1\}$ and $\{p\}$.

Consider the following statement $S(\ell, m)$ for positive integers ℓ, m:
"For all $k \geq 1$ integer there exists a $W(\ell, m, k)$ positive integer such that for every $c : [W(\ell, m, k)] \to [k]$ coloring there exists $a, d_1, ..., d_m$ positive integers such that

$$a + \ell \sum_{i=1}^{m} d_i \leq W(\ell, m, k) \tag{4.3}$$

and the map $(x_1, ..., x_m) \mapsto c\Big(a + \sum_{i=1}^m x_i d_i\Big)$ is constant on the $\sim_\ell$-equivalence classes of m-sequences."

Observe that the condition (4.3) just makes sure that the number whose color we talk about, $\sum_{i=1}^m x_i d_i$, is indeed in $[W(\ell, m, k)]$. Also, observe that $S(p, 1)$ in view of what are the $\sim_p$-equivalence classes of 1-sequences, implies the existence of $W(p, k)$ as required in the van der Waerden theorem.

A further observation is that $S(\ell, m)$, formulated about all $f : [W(\ell, m, k)] \to [k]$ colorations, is equivalent to a similar statement about all $g : \{n+1, n+2, ..., n+W(\ell, m, k)\} \to [k]$ colorations, requiring the existence of an $n + 1 \leq b$ integer and $d_1, ..., d_m$ positive integers such that $b + \ell \sum_{i=1}^m d_i \leq n + W(\ell, m, k)$ and the map $(x_1, ..., x_m) \mapsto c\Big(b + \sum_{i=1}^m x_i d_i\Big)$ is constant on the $\sim_\ell$-equivalence classes of m-sequences. A translation by n gives the equivalence of the statements.

We are going to prove by induction that

$$\forall \ell \geq 1 \, \forall m \geq 1 \;\; S(\ell, m).$$

This will prove the van der Waerden theorem as well.

The base case is $S(1, 1)$. We claim that for all $k \geq 1$ the choice $W(1, 1, k) = 2$ suffices with $a = 1$ and $d = 1$. Condition (4.3) holds as $1 + 1 \cdot 1 \leq W(1, 1, k) = 2$. The $\sim_1$-equivalence classes of 1-sequences are $\{0\}; \{1\}$. The coloration is constant on singleton classes for free.

The induction will use two type of inductive steps:

Step 1: $\forall m \geq 1 \Bigg(\Big(S(\ell, 1) \wedge S(\ell, m) \Big) \Rightarrow S(\ell, m+1) \Bigg)$.

Step 2: $\forall \ell \geq 1 \Bigg(\Big(\forall m \geq 1 \;\; S(\ell, m) \Big) \Rightarrow S(\ell+1, 1) \Bigg)$.

It is clear that that repeating these two steps we can reach from $S(1,1)$ any $S(\ell,m)$, so the induction works if these two steps can be verified.

Verification of Step 1.

To prove $S(\ell,m+1)$, we have to show that for all k positive integers $W(\ell,m+1,k)$ exists. We show that

$$\text{the choice } W(\ell,m+1,k) =: \underbrace{W(\ell,m,k)}_{W}\underbrace{W\Big(\ell,1,k^{W(\ell,m,k)}\Big)}_{W'} \tag{4.4}$$

will suffice. The underbraced expressions exists by the hypothesis on $S(\ell,m)$; and by the hypothesis on $S(\ell,m)$ and $S(\ell,1)$. At this point ℓ,m,k are fixed positive integers, and we abbreviate the expressions in (4.4) by W and W'. We have to show that every $c:[WW']\to[k]$ produces what the statement $S(\ell,m+1)$ requires for our particular k. Fix an arbitrary c coloration. First observe that

$$[WW'] = \bigcup_{j=1}^{W'} \underbrace{[(j-1)W+1, jW]}_{W_j},$$

and the interval denoted above by W_j is just the translation of the interval $[1,W]$ by $(j-1)W$. Define a $c':[W']\to k^W$ by $c'(j)=$ the c-coloration of $[WW']$ restricted to W_j. Values of c' are vectors of values of c. By the choice of W' for $S(\ell,1)$ with $k^{W(\ell,m,k)}$ colors, there exists a',d' such that $a'+\ell d'\le W'$ and

$$c'(a') = c'(a'+d') = ... = c'(a'+(\ell-1)d').$$

In other words, there is a length ℓ arithmetic progression of j indices such that the W_j intervals are identically colored by c.

We use the existence of $W(\ell,m,k)$ for the interval $W_{a'}$ instead of $[W]$. As they have the same length, this just requires a to be in $W_{a'}$ instead of $[W]$, and the sums are required not go out from $W_{a'}$, so

$$\text{exists } a,d_1,...,d_m \text{ positive integers, such that } (a'-1)W+1\le a< a+\ell\sum_{i=1}^{m} d_i \le a'W,$$

and the map $(x_1,...,x_m)\mapsto c\Big(a+\sum_{i=1}^m x_i d_i\Big)$ is constant on the $\sim_\ell$-equivalence classes of m-sequences.

Define now $d_{m+1}=d'W$. We have the required inequality

$$a+\ell\sum_{i=1}^{m+1} d_i = \Big(a+\ell\sum_{i=1}^{m} d_i\Big)+\ell d'W \le a'W + ld'W \le W(a'+\ell d') \le WW'.$$

In addition, we need that the map $(x_1,...,x_m,x_{m+1})\mapsto c\Big(a+\sum_{i=1}^{m+1} x_i d_i\Big)$ is constant on the $\sim_\ell$-equivalence classes of $m+1$-sequences. How two $(m+1)$-sequences can be $\sim_\ell$-equivalent? There are two logical possibilities.

Case (a): Their last entries agree to ℓ. Then the equivalence classes are singletons and we have

nothing to prove.
Case (b): The sequences look like $(x_1, ..., x_m, u) \sim_\ell (y_1, ..., y_m, v)$, where $u, v < \ell$, and $(x_1, ..., x_m) \sim_\ell (y_1, ..., y_m)$. Then, because of the hypothesis

$$c\Big(a + \sum_{i=1}^{m} x_i d_i\Big) = c\Big(a + \sum_{i=1}^{m} y_i d_i\Big).$$

In addition, for $0 \le u, v \le \ell - 1$

$$c\left(a + \Big(\sum_{i=1}^{m} x_i d_i\Big) + u d_{m+1}\right) = c\Big(a + \sum_{i=1}^{m} x_i d_i\Big)$$

and

$$c\left(a + \Big(\sum_{i=1}^{m} y_i d_i\Big) + v d_{m+1}\right) = c\Big(a + \sum_{i=1}^{m} y_i d_i\Big),$$

as translation by $ud_{m+1} = ud'W$ and $vd_{m+1} = vd'W$ preserves the c-coloration. Hence the two $\sim_\ell$-equivalent $(m+1)$-sequences are assigned the same color.

Verification of Step 2.
To prove $S(\ell+1, 1)$, we have to find for every positive integer k a number $W(\ell+1, 1, k)$ as required. We show that every fixed k

$$\text{the choice } W(\ell + 1, 1, k) =: 2W(\ell, k, k) \tag{4.5}$$

suffices. We have to show that for all $c : [2W(\ell, k, k)] \to [k]$, there exists A and D, such that $A + (\ell+1)D \le 2W(\ell, k, k)$ and the map $c : x \mapsto c(A + xD)$ is constant on $\sim_{\ell+1}$-equivalence classes of 1-sequences. Recall that there are exactly two $\sim_{\ell+1}$-equivalence classes of 1-sequences, namely the singleton $\{(\ell + 1)\}$, and all other 1-sequences, whose only entry is one of 0,1,...,ℓ, are in one class. So we have to verify $c(A) = c(A + d) = c(A + 2D) = \cdots = c(A + \ell D)$.

Restrict c to $[W(\ell, k, k)]$. By the choice of $W(\ell, k, k)$, there exist $a, d_1, ..., d_k$ positive integers, such that $a + \ell \sum_{i=1}^{k} d_i \le W(\ell, k, k)$, and $c : (x_1, ..., x_k) \mapsto c\Big(a + \sum_{i=1}^{k} x_i d_i\Big)$ is constant on the $\sim_\ell$-equivalence classes of k-sequences.

Consider now the colors $c\Big(a + \sum_{i=1}^{s} \ell d_i\Big)$ for $s = 0, 1, ..., k$. For $s = 0$, this color is $c(a)$. As we have only k colors, by the Pigeonhole Principle one color must be repeated, i.e. exist $0 \le s < t \le k$ such that

$$c\Big(a + \sum_{i=1}^{s} \ell d_i\Big) = c\Big(a + \sum_{i=1}^{t} \ell d_i\Big).$$

Take $A = a + \sum_{i=1}^{s} \ell d_i$ and $D = \sum_{i=s+1}^{t} d_i$. We show that this choice suffices. The only element of the singleton class is

$$A + (\ell + 1)D = \left(a + \sum_{i=1}^{s} \ell d_i\right) + (\ell + 1) \sum_{i=s+1}^{t} d_i = \left(a + \sum_{i=1}^{t} \ell d_i\right) + \sum_{i=s+1}^{t} d_i \le 2W(\ell, k, k),$$

as required, and all other terms associated with our sequences are even smaller than $W(\ell,k,k)$. Observe that for $0 \le j_1, j_2 < \ell$, we have

$$c\Big(A + j_1 D\Big) = c\left(a + \sum_{i=1}^{s} \ell d_i + \sum_{i=s+1}^{t} j_1 d_i\right) = c\left(a + \sum_{i=1}^{s} \ell d_i + \sum_{i=s+1}^{t} j_2 d_i\right) = c\Big(A + j_2 D\Big),$$

where the middle equation holds by the choice of $a, d_1, ..., d_k$, with which c was constant on the $\sim_\ell$-equivalence classes of k-sequences, applied to

$$(\underbrace{\ell, \ell, ..., \ell}_{s}, \underbrace{j_1, j_1, ..., j_1}_{t-s}, \underbrace{0, 0, ..., 0}_{k-t}) \sim_\ell (\underbrace{\ell, \ell, ..., \ell}_{s}, \underbrace{j_2, j_2, ..., j_2}_{t-s}, \underbrace{0, 0, ..., 0}_{k-t}).$$

This proves $c(A) = c(A + d) = c(A + 2D) = \cdots = c(A + (\ell - 1)D)$. Furthermore.

$$c\Big(A + 0 \cdot D\Big) = c\Big(a + \sum_{i=1}^{s} \ell d_i\Big) = c\Big(a + \sum_{i=1}^{t} \ell d_i\Big) = c\Big(A + \ell D\Big),$$

where the middle equation holds by the earlier Pigeonhole Principle argument, finishing the proof.

Note that we only need the factor of two in (4.5) to make room to accomodate $A + (\ell + 1)D$, the single term appearing in the singleton class, otherwise all the terms involved in our argument stay in $[W(\ell,k,k)]$.

5 Bounds for the van der Waerden theorem

5.1 Recursive and primitive recursive functions

Consider the following countably infinite sequence of $\mathbb{Z}^+ \to \mathbb{Z}^+$ functions:

$$f_1(x) = x + 1, \quad \text{and for } n \ge 2, \quad f_{n+1}(x) = \underbrace{f_n \circ f_n \circ \cdots \circ f_n}_{x \text{ times}}(2),$$

where $\circ$ indicates the composition of functions. Recall that composition of functions is associative. It is easy to see that

$$\begin{aligned}
f_2(x) &= \underbrace{f_1 \circ f_1 \circ \cdots \circ f_1}_{x \text{ times}}(2) = x + 2, \\
f_3(x) &= \underbrace{f_2 \circ f_2 \circ \cdots \circ f_2}_{x \text{ times}}(2) = 2x + 2, \\
f_4(x) &= \underbrace{f_3 \circ f_3 \circ \cdots \circ f_3}_{x \text{ times}}(2) > 2^x, \\
f_5(x) &= \underbrace{f_4 \circ f_4 \circ \cdots \circ f_4}_{x \text{ times}}(2) > 2^{2^{\cdot^{\cdot^{\cdot^2}}}}, \text{a tower of } x \text{ copies of 2}, \\
f_6(x) &= \underbrace{f_5 \circ f_5 \circ \cdots \circ f_5}_{x \text{ times}}(2) > 2^{2^{\cdot^{\cdot^{\cdot^2}}}}, \text{a tower of } f_6(x-1) \text{ copies of 2}.
\end{aligned}$$

We stop here with the examples.

Recursive or *computable* $\mathbb{N} \to \mathbb{N}$ functions have a technical definition in logic that we do not attempt repeat here. Church's Thesis, however, asserts that computable functions are exactly those that we can compute with a computer program. Here a computer is ideal, has unbounded memory space. The program is supposed to be of finite length, take input, make output, and can use assigment, conditional and loop commands. For every input, the program is supposed to make an output and stop.

The *assignment command* is $x \leftarrow e$, assigning to program variable x the current value of an expression e, which may depend on the input and program variables.

The *conditional command* is

```
IF B THEN
   P
ELSE
   Q
END IF
```

in the conditional command B is a proposional function depending on program variables. If B is true, program segment `P` will be executed, if B is false, program segment `Q` will be executed.

The *loop command* is

```
WHILE  B THEN
   P
END WHILE
```

Here B is a proposional function depending on program variables. If B is true, program segment `P` will be executed, otherwise we leave the loop. If we stay in the loop, the truth value of B will be evaluated again, to see whether we execute `P` again or we leave the loop.

It is easy to write a loop that never stops. Therefore, it is reasonable to ask whether a given code will ever stop. This is the Halting Problem. Turing proved that there is no computer program that would correctly decide after inputting any computer program and its input, whether the computer program would stop with that input.

There is a restricted version of the loop command, the for-loop:

```
FOR  i = 1 TO n
   P
END FOR
```

The for-loop repeats executing the program segment `P` n times. When you enter the for-loop, you know how many times you will do the repetition. n may depend on the input and current values of program variables. Funny things, however, are not allowed! While in `P`, you are not allowed to redefine n or the program variable i. It is easy to see that for-loops can be written as while-loops. But not the other way around. If your code uses only for-loops and no while-loops,

its Halting Problem is trivial, as the computer executes your code line after line, and will stop.

An $f : \mathbb{N} \to \mathbb{N}$ function is called *primitive recursive*, if there is a code for computing f that uses for-loops only, no while-loops. Unless you ran into the Ackermann function, every recursive function with which you worked so far was primitive recursive. We show how f_4 and f_5 can be computed with for-loops and you should be convinced that for every positive integer n, f_n can be evaluated with for-loops:

Code to compute values of f_4:

```
INPUT x
f ← 2
FOR i = 1 TO x
    f ← 2f + 2
END FOR
OUTPUT f
```

The exiting value of f is $f_4(x)$.

Code to compute values of f_5:

```
INPUT x
g ← 2
FOR j = 1 TO x
    f ← 2
    FOR i = 1 TO g
        f ← 2f + 2
    END FOR
g ← f
END FOR
OUTPUT g
```

The exiting value of f after repeated inner loops is $f_4(2), f_4 \circ f_4(2), ...$, and the exiting value of g after the completion of the outer loop is $f_5(x) = \underbrace{f_4 \circ f_4 \circ \cdots \circ f_4}_{x \text{ times}}(2)$, as required.

First we verify some properties of the sequence of functions above.

Claim 2 *(a) For all $i \geq 1$ and $x \geq 1$, we have $f_i(x) < f_{i+1}(x)$.*
(b) For all $i \geq 1$ and $x \geq 1$, we have $f_i(x) > x$.
(c) For all $i \geq 1$, $f_i(x)$ is strictly monotone increasing, and therefore has infinite limit,
(d) For all $i \geq 3$, $f_i(x) = o(f_{i+1}(x))$.

Proof. We prove (a) by induction on i. For $i = 1$, $f_1(x) = x + 1 < x + 2 = f_2(x)$. Using (a) for a certain i as hypothesis, we obtain for any x

$$f_{i+2}(x) = \underbrace{f_{i+1} \circ f_{i+1} \circ \cdots \circ f_{i+2}}_{x \text{ times}}(2) > \underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x \text{ times}}(2) = f_{i+1}(x),$$

as required. Part (b) follows from $x < x + 1 = f_1(x)$ and $f_1(x) \leq f_i(x)$ from Part (a). We do Part (c) by induction on i. For $i = 1$, the function $f_1(x) = x + 1$ is monotone increasing. Assume that

$f_i(x)$ is monotone increasing. Then, for any $x \geq 1$,

$$f_{i+1}(x) = \underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x \text{ times}}(2) < f_i\Big(\underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x \text{ times}}(2)\Big) = f_{i+1}(x+1),$$

where in the middle inequality we used (b). We do Part (d) by induction on i. The base case $i = 3$ follows from inspection of $f_3(x)$ and $f_4(x)$. Assume that we know Part (d) for a number i. Now

$$\begin{aligned} f_{i+1}(x) &= \underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x \text{ times}}(2) = f_i\Big(\underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x-1 \text{ times}}(2)\Big) \\ &\quad (\text{with} \quad z = \underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x-1 \text{ times}}(2) \to \infty \quad \text{as} \quad x \to \infty) \\ &= f_i(z) = o(f_{i+1}(z)) = o\left(f_{i+1}\Big(\underbrace{f_i \circ f_i \circ \cdots \circ f_i}_{x-1 \text{ times}}(2)\Big) \right) \\ &= o\Big(\underbrace{f_{i+1} \circ f_{i+1} \circ \cdots \circ f_{i+1}}_{x \text{ times}}(2)\Big) = o\Big(f_{i+2}(x)\Big). \end{aligned}$$

The following theorem characterizes primitive recursive functions:

Theorem 6 *An $f : \mathbb{N} \to \mathbb{N}$ function primitive recursive if and only if it is recursive, and there exists an m, such that $f(x) = O(f_m(x))$.*

In view of this theorem, if we find a function, which is recursive, but grows faster than every $f_i(x)$ function in our sequence, then it is not primitive recursive. Cantor's diagonal method gives such a function: $g(n) = f_n(n)$. Clearly we can compute the values of g. On the other hand, for all $x > i$, $f_i(x) < f_x(x) = g(x)$ by Part (a) of the Claim. Assume that $g(x)$ is primitive recursive, i.e., there is a positive integer m, such that $g(x) = O(f_m(x))$. Then, for $x > m + 1$, $f_{m+1}(x) < g(x) = O(f_m(x))$, contradicting Part (d) of Claim 2.

5.2 Upper bounds for the van der Waerden theorem

For 60 years, the known upper bounds for the van der Waerden numbers $W(p, 2)$ were not primitive recursive. It was contemplated whether the $W(p, 2)$ sequence is not primitive recursive.

However, in 1987, Shelah gave an entirely different proof, verifying that $W(p, 2) \leq u_{p+2}$, where $u_1 = 2$ and $u_{n+1} = 2^{2^{\cdot^{\cdot^{2}}}}$, a tower of u_n copies of 2. Induction gives that $W(p, 2) \leq u_{p+2} \leq f_6(p+2)$, so Shelah's bound is primitive recursive. Gowers' result gave the upper bound

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

which gives $W(p, 2) \leq f_5(p + 9)$.

5.3 The upper bound that Section 4 provided for $W(p,2)$ is not primitive recursive

Let us use the following notation. $W(\ell,m,k)$ is the *least* integer, for which the statement in quotation mark holds, while $W_1(\ell,m,k)$ is the number, recursively defined in the proof, for which the statement holds. Obviously $W(\ell,m,k) \le W_1(\ell,m,k)$. The recursive definition went in the following way:

$$W_1(1,1,x) = 2, \tag{5.6}$$

$$W_1(\ell,m+1,x) = W_1(\ell,m,x)\cdot W_1\Big(\ell,1,x^{W_1(\ell,m,x)}\Big), \tag{5.7}$$

$$W_1(\ell+1,1,x) = 2W_1(\ell,x,x). \tag{5.8}$$

First we are going to show that for all $\ell \ge 2$, and all $x \ge 2$,

$$W_1(\ell,1,x) \ge f_\ell(x). \tag{5.9}$$

First observe that $W_1(2,x,x) \ge 1+2x$ by the condition (4.3). By (5.8), $W_1(2,1,x) = 2W_1(2,x,x) \ge 2(1+2x) > f_2(x)$, which provides the $\ell = 2$ base case for the induction proof to (5.9). In the inductive step, we use Claim 2 all the time without giving reference to it:

$$\begin{aligned}
& W_1(\ell+1,1,x) = 2W_1(\ell,x,x) \quad \text{by (5.8)} \\
= \;& 2W_1(\ell,x-1,x)W_1\Big(\ell,1,x^{W_1(\ell,x-1,x)}\Big) \quad \text{by (5.7)} \\
\ge \;& f_\ell\Big(x^{W_1(\ell,x-1,x)}\Big) \ge f_\ell\Big(W_1(\ell,x-1,x)\Big) \\
= \;& f_\ell\left(W_1(\ell,x-2,x)W_1\Big(\ell,1,x^{W_1(\ell,x-2,x)}\Big)\right) \quad \text{by (5.7)} \\
\ge \;& f_\ell\circ f_\ell\left(x^{W_1(\ell,x-2,x)}\right) \ge f_\ell\circ f_\ell\left(W_1(\ell,x-2,x)\right) \\
\ge \;& \underbrace{f_\ell\circ f_\ell\circ\cdots\circ f_\ell}_{x-1 \text{ times}}\Big(W_1(\ell,1,x)\Big) \ge \underbrace{f_\ell\circ f_\ell\circ\cdots\circ f_\ell}_{x \text{ times}}(x) \\
\ge \;& \underbrace{f_\ell\circ f_\ell\circ\cdots\circ f_\ell}_{x \text{ times}}(2) = f_{\ell+1}(x).
\end{aligned}$$

For the next step, observe that for $\ell \ge 4$, we have $W(\ell,1,2) \ge \ell$. Indeed, color differently the numbers up to $\lfloor \ell/2 \rfloor$ and above it. Now we have $\ell \ge 4$:

$$2^{W_1(\ell,1,2)} > W_1(\ell,1,2) \ge W(\ell,1,2) \ge \ell.$$

Now, for $\ell \ge 4$, we have

$$\begin{aligned}
W_1(\ell+1,1,2) &\ge W_1(\ell,2,2) \quad \text{by (5.8)} \\
&= W_1(\ell,1,2)\cdot W_1\Big(\ell,1,2^{W_1(\ell,1,2)}\Big) \quad \text{by (5.7)} \\
&\ge f_\ell\Big(2^{W_1(\ell,1,2)}\Big) \ge f_\ell(\ell).
\end{aligned}$$

So $W_1(\ell+1,1,2)$ is bigger than the function $g(\ell) = f_\ell(\ell)$, which was not primitive recursive.

5.4 Lower bounds for the van der Waerden theorem

We show Berlekamp's construction:

Theorem 7 *For a prime number p, $W(p+1,2) > p2^p$.*

We will use three facts from algebra, the first without proof:

Lemma 8 *In every finite field, the multiplicative group of the non-zero elements is cyclic.*

Lemma 9 *For a prime p, $GF(2^p)$ does not have any subfield other than its prime field, $\{0,1\}$.*

If there was a field K such that $GF(2) \subsetneq K \subsetneq GF(2^p)$, then K were a linear space over $GF(2)$ and $GF(2^p)$ were a linear space over K. As the cardinality of a linear space is cardinality of the field raised to the dimension, $|K| = 2^d$ and $2^p = |GF(2^p)| = |K|^{d'}$. Hence $dd' = p$, a contradiction unless $K = GF(2)$ or $K = GF(2^p)$.

Lemma 10 *No element of $GF(2^p) \setminus \{0,1\}$ is a root of any polynomial over $GF(2)$ of degree less than p.*

For contradiction, assume that there is a polynomial $q(x)$ over $GF(2)$, such that $deg(q(x)) < p$ and $q(s) = 0$ for some $q \in GF(2^p) \setminus \{0,1\}$. We can assume without loss of generality that $q(x)$ has the smallest possible degree r among such polynomials. The leading coefficient of $q(x)$ is non-zero, therefore it is 1. Clearly $1 < r$, as linear polynomials over $GF(2)$ already have a root in $GF(2)$, and cannot have a second root in the bigger field.

Consider the expressions

$$K = \Big\{\sum_{i=0}^{r-1} \alpha_i s^i : \alpha_i \in \{0,1\}\Big\} \subseteq GF(2^p). \tag{5.10}$$

K is closed for addition. K is closed for the additive inverse, as for every element of K, it's additive inverse is itself.
It takes a little more work to show that K is closed for multiplication. We claim that for every $p(x) = \sum_{i=0}^{m} \beta_i x^i$ polynomial with $\beta_i \in \{0,1\}$, there is an $\alpha_i \in \{0,1\}$ for $i = 0, 1, ..., r-1$, such that $p(s) = \sum_{i=0}^{r-1} \alpha_i s^i$. There is division with remainder for polynomials over any field. Write

$$p(x) = s(x)q(x) + t(x), \tag{5.11}$$

where $\deg(t(x)) \leq \deg(q(x)) - 1 = r - 1$. As $q(s) = 0$, plugging in s into the variable in (5.11), we get $p(s) = t(s)$, and the looked for α_i numbers are the coefficients of $t(x)$.
K has at most 2^r distinct elements by definition. We show that K has exactly 2^r distinct elements. If not, then there are $p_1(x)$ and $p_2(x)$ polynomials over $GF(2)$, such that

$$\deg(p_2(x)) \leq \deg(p_1(x)) \leq r-1, \text{and } p_1(s) = p_2(s),$$

providing two equal expressions in (5.10). Using division of polynomials with remainder, $p_1(x) = a(x)p_2(x) + b(x)$, with $\deg(b(x)) \leq \deg(p_2(x)) - 1 \leq r - 2$. The existence of this $b(x)$ contradicts the choice of $q(x)$ and the minimality of r.

The set $K \setminus \{0\}$ is also closed for multiplication, as non-zero elements have a non-zero product (as we are in the field $GF(2^p)$). Fixed any $a \in K \setminus \{0\}$. The map $x \mapsto ax$ from $K \setminus \{0\}$ into $K \setminus \{0\}$ is an injection, as $ab = ac$ implies $a(b - c) = 0$ and $b = c$. As $K \setminus \{0\}$ is finite, this map is also a surjection. Hence $ax = 1$ has a solution in $K \setminus \{0\}$, i.e. multiplicative inverse of a exists within $K \setminus \{0\}$.

So K is a subfield of 2^r elements, with $1 < r < p$, contradicting the previous lemma.

After the lemmata, we return to the proof of the theorem. To simplify Berlekamp's construction, we show only a somewhat weaker result: $W(p + 1, 2) > p(2^p - 1)$. Let α be a generator of the cyclic group of non-zero elements in $GF(2^p)$. Let $\mathbf{b}_0, ..., \mathbf{b}_{p-1}$ be basis in $GF(2^p)$, when it is viewed as a linear space over $GF(2)$. Therefore, for $j = 0, 1, .., p(2^p - 1) - 1$, there are $\lambda_{ij} \in \{0, 1\}$, such that

$$\alpha^j = \sum_{i=0}^{p-1} \lambda_{ij} \mathbf{b}_i.$$

The 2-coloration of $[0, p(2^p - 1) - 1]$ will be

$$c(j) = \lambda_{0j} \in \{0, 1\}, \text{ which comes from the basis representation of } \alpha^j.$$

Assume now that there is a $(p+1)$-term monochromatic arithmetic progression in $[0, p(2^p-1)-1]$, say $a, a + d, ..., a + pd$. As $a + pd \leq p(2^p - 1) - 1$, we have $0 < d < 2^p - 1$, and therefore $\alpha^d \neq 1$. Being monochromatic, the coordinates of $\alpha^a, \alpha^{a+d}, ..., \alpha^{a+pd}$ in $\mathbf{b}_0$ are all 0 or are all 1. If they are all zero, then these vectors lie in a $(p-1)$-dimensional subspace generated by $\mathbf{b}_1, ..., \mathbf{b}_{p-1}$, and therefore they are linearly dependent, even more, *the first p of these vectors* are linearly dependent. In other words, there are coefficients $\mu_i \in \{0, 1\}$ making a non-trivial linear combination

$$\sum_{i=0}^{p-1} \mu_i \alpha^{a+id} = 0.$$

Hence $\sum_{i=0}^{p-1} \mu_i (\alpha^d)^i = 0$, and $\alpha^d \notin \{0, 1\}$ is a root of a polynomial of degree less than p, namely $\sum_{i=0}^{p-1} \mu_i x^i$, a contradiction.
If the colors of the arithemetic progression are all 1, then all coefficients $\lambda_{0j} = 1$. Hence the following p vectors:

$$\alpha^{a+d} - \alpha^a, ..., \alpha^{a+pd} - \alpha^a$$

all lie in a $(p - 1)$-dimensional subspace generated by $\mathbf{b}_1, ..., \mathbf{b}_{p-1}$. Therefore these p vectors are linearly dependent and there are coefficients $\mu_i \in \{0, 1\}$ making a non-trivial linear combination

$$\sum_{i=1}^{p} \mu_i \Big(\alpha^{a+id} - \alpha^a \Big) = 0.$$

Dividing with α^a, we have $\sum_{i=1}^{p} \mu_i\big((\alpha^d)^i - 1\big) = 0$, and $\alpha^d \notin \{0,1\}$ is a root of the polynomial $\sum_{i=1}^{p} \mu_i(x^i - 1)$. However, $x^i - 1 = (x-1)(x^{i-1} + x^{i-2} + ... + x + 1)$ over $GF(2)$ as well. Hence $\alpha^d \notin \{0,1\}$ is a root of the polynomial $\sum_{i=1}^{p} \mu_i(x^{i-1} + x^{i-2} + ... + x + 1)$, of degree less than p, a contradiction again.
Note that the proof gave slightly more than what we promised: no $p+1$ term AP in color 1 and no p term AP in color 0.

Using the Lovász Local Lemma, one can prove $W(p+1,k) > \frac{k^{p+1}}{(p+1)ke} - \frac{1}{p+1}$. (See the homework problems.) For $k = 2$ it gives $W(p+1,2) > \frac{2^{p+1}}{2(p+1)e} - \frac{1}{p+1}$, which is weaker than Berlekamp's bound, but it does not require p to be prime. Also, Berlekamp's bound does not apply to more than 2 colors.

6 Some applications of the van der Waerden theorem

For hundreds of years, quadratic residues have been studied in number theory. Schur conjectured that for any sufficiently large prime p, one finds k consecutive quadratic residues and quadratic non-residues. Brauer (1928) proved the following generalization of van der Waerden's theorem and used it to settle the conjecture of Schur.

Theorem 11 *For all p, k positive integers there exists a $B(p,k)$ number, such that for all $n \geq B(p,k)$ and every $f : [n] \to [k]$ coloration there exists a monochromatic p-term arithmetic progression, and in addition, the difference d of the arithmetic progression has the same color.*

The proof goes by induction on k. If $k = 1$, $B(p,1) =: p$ suffices. For $k > 1$, we assume that we already have a bound $B(p,k-1)$. We show that the choice $B(p,k) =: W(p \cdot B(p,k-1) + 1, k)$ suffices. Take an arbitrary coloration $c : [B(p,k)] \to [k]$. By the choice for $B(p,k)$, there exists an arithmetic progression P of length $p \cdot B(p,k-1) + 1$, which is monochromatic, say in color k. Let a denote the first term and d' denote the difference of this arithmetic progression. As the last term of P is still in $[B(p,k)]$, we have the inequality

$$a + pB(p,k-1)d' \leq B(p,k). \tag{6.12}$$

There are two possible cases:
Case 1:
For some $1 \leq j \leq B(p,k-1)$, $c(jd') = k$. Consider a new p-term arithmetic progression

$$a, a + jd', a + 2(jd'), ..., a + (p-1)(jd').$$

By the choice of j,

$$(p-1)(jd') \leq \big(pB(p,k-1)\big)d',$$

and therefore the new arithmetic progression is a subset of P. Therefore every term of the new arithmetic progression has color k, and so has its difference, jd'.
Case 2:
For all $1 \leq j \leq B(p,k-1)$, $c(jd') < k$. As

$$jd' \leq B(p,k-1)d' \leq B(p,k),$$

the element jd' received a color from the coloration c. Define a $k-1$-coloration $c' : [B(p, k-1)] \to [k-1]$ by the rule $c'(j) = c(jd')$ and apply the induction hypothesis. There is a monochromatic p-term arithmetic progression with first term b and difference f with respect to the coloration c', and f has the same color as the members of the progression:

$$c'(f) = c'(b) = c'(b+f) = c'(b+2f) = \cdots = c'\big(b+(p-1)f\big).$$

Now $d'b, d'(b+f), ..., d'(b+(p-1)f)$ and its difference $d'f$ have the same color in c:

$$c(d'f) = c(d'b) = c\big(d'(b+f)\big) = c\big(d'(b+2f)\big) = \cdots = c\big(d'(b+(p-1)f)\big).$$

To prove Schur's conjecture, we have to go through basic properties of quadratic residues. Let p denote an odd prime number and fix it for rest of this section. Given an integer a, consider whether the congruence

$$x^2 \equiv a \mod p \tag{6.13}$$

has an integer solution. If a is a multiple of p, then $x = p$ is a solution; this is considered a trivial case. The existence of the solution depends just on the congruence class of a modulo p. $a \not\equiv 0$ is called a *quadratic residue* modulo p, if a solution exists, and a *quadratic non-residue* modulo p, if a solution does not exist.

The following facts are easy to see directly from the definition:
Product of quadratic residues is a quadratic residue.
Indeed, $x^2 \equiv a \mod p$ and $y^2 \equiv b \mod p$ imply that $(xy)^2 \equiv x^2y^2 \equiv ab \mod p$.
Every quadratic residue is the square of at least two different residue classes.
Indeed, if $x^2 \equiv a \mod p$, then also $(p-x)^2 \equiv x^2 \equiv a \mod p$, and $x \not\equiv p-x \mod p$, as p was odd.
There is a quadratic non-residue.
Indeed, squaring the non-zero residue classes is not a bijection, therefore some non-zero residue class is not a value under squaring.
Product of a quadratic residue with a quadratic non-residue is a quadratic non-residue.
Take an arbitrary a that is a quadratic residue, such that $x^2 \equiv a \mod p$, and an arbitrary h that is a quadratic non-residue. Assume for contradiction that ha is a quadratic residue. Then $ha \equiv y^2 \mod p$ for some y. There is an x^*, the multiplicative inverse of x, such that $x^*x \equiv 1 \mod p$. Now we have $(x^*y)^2 \equiv (x^*)^2y^2 \equiv (x^*)^2ah \equiv (x^*)^2x^2h \equiv h \mod p$. h is a quadratic residue, a contradiction.
Every quadratic residue is the square of exactly two residue classes.
Indeed, if for integers x and a, $x^2 \equiv a^2 \mod p$, then $(x-a)(x+a) = x^2 - a^2 \equiv 0 \mod p$. Hence p divides $x-a$ or p divides $x+a$. Either $a \equiv x \mod p$ or $a \equiv -x \mod p$. There are no more possibilities, and these two possibilities are distinct, as p is odd.
Exactly half of the non-zero residue classes are quadratic residues, and exactly half of the non-zero residue classes are quadratic non-residues.
Squaring the non-zero residue classes is a two-to-one map to itself. Therefore exactly half of elements of this set do not appear as a value.
Product of two quadratic non-residues is a quadratic residue.
Let h be a quadratic non-residue. Multiplication by h maps the non-zero residue classes to the non-zero residue classes bijectively, as h has a multiplicative inverse. We already know that h

times a quadratic residue is a quadratic non-residue. Therefore, multiplication by h should map quadratic non-residues to quadratic residues, as the number quadratic residues and quadratic non-residues are the same.
Number 1 *is a quadratic residue.*
$1 \cdot 1 \equiv 1 \bmod p$.
Take the mod p *multiplicative inverse* a^* *of any* $a \not\equiv 0$ *mod* p. *Either* a, a^* *are both quadratic residues, or are both quadratic non-residues.*
If one of them was a quadratic residue, the other is a quadratic non-residue, then their product would be a quadratic non-residue.

Euler's Lemma allows to decide whether a number is a quadratic residue or not. A useful application of it is that for $p = 4k + 3$, -1 is a quadratic non-residue mod p. The Lemma easily follows from the observations above.

Lemma 12 Euler's Lemma. *For* $a \not\equiv 0$ *modulo* p, *where* p *is an odd prime,*

$$a^{\frac{p-1}{2}} = \begin{cases} 1 \mod p, & \text{if } a \text{ is quadratic residue modulo } p, \\ -1 \mod p, & \text{if } a \text{ is quadratic non-residue modulo } p. \end{cases}$$

Proof. Observe that $\left(a^{\frac{p-1}{2}}\right)^2 \equiv a^{p-1} \equiv 1 \bmod p$ by Fermat's Little Theorem. $x^2 \equiv 1 \bmod p$ has exactly two solutions, $x \equiv \pm 1 \bmod p$. Therefore, for every $a \not\equiv 0$, $a^{\frac{p-1}{2}} \equiv 1$ or $a^{\frac{p-1}{2}} \equiv -1 \bmod p$.

Next, observe that if a is a quadratic residue, then $a^{\frac{p-1}{2}} \equiv 1 \bmod p$. Indeed, if $a \equiv b^2 \bmod p$, then $a^{\frac{p-1}{2}} \equiv b^{p-1} \equiv 1 \bmod p$ by Fermat's Little Theorem.

Consider the polynomial $x^{\frac{p-1}{2}} - 1$ over $GF(p)$. This polynomial cannot have more different roots than its degree. In fact, it has exactly that many, as each of the $\frac{p-1}{2}$ quadratic residues are roots. Therefore, for every quadratic non-residue c, $c^{\frac{p-1}{2}} - 1 \not\equiv 0 \bmod p$. This fact forces that $c^{\frac{p-1}{2}}$ has the only other possible value mod p, namely -1.

The distribution of quadratic residues and quadratic non-residues is erratic. For many aspects, it mimics randomness. Schur conjectured the following theorem:

Theorem 13 *For every* k, *there exists an* $n(k)$, *such that for every* $p > n(k)$ *prime there are* k *consecutive integers that are quadratic residues and* k *consecutive integers that are quadratic non-residues modulo* p.

Brauer proved his generalization of the van der Waerden theorem to settle this conjecture. For $k = 1$, we just need a quadratic residue and a quadratic non-residue, and we do have them for $p \geq 3$. From now on we assume $k \geq 2$. First we show that $n(k) = B(k, 2)$ suffices to find k consecutive quadratic residues. Color $x \in [1, p-1]$ by ± 1 according to $a^{\frac{p-1}{2}} \bmod p$ as in Euler's Lemma, i.e., the coloration is with colors "quadratic residue" and "quadratic non-residue". By the choice of $n(k)$, there is an arithmetic progression $a, a+d, ..., a+(k-1)d$, such that the terms of this progression and its difference d have the same color. Let d^* denote the multiplicative inverse of $d \bmod p$. By the rules for products, d and d^* are both quadratic residues or non-residues, as their product is the congruence class 1, a quadratic residue. Multiplying the terms of the

arithmetic progression with d^* we obtain quadratic residues, and evaluating the products we see $ad^*, ad^*+1, ..., ad^*+k-1$. These are consecutive residues, as they cannot wrap around 0, as 0 is not among them.

To find k consecutive non-residues, we need a much bigger threshold: $n(k) = B\big(k!k+1, 2\big)$. Assume that $p > n(k)$ is a prime. Then, from the first part of the proof, there exists $k!k+1$ consecutive quadratic residues. Denote them by $b, b+1, ..., b+k!k$. Let d denote the smallest quadratic non-residue in $[1, p-1]$.

Case: $d \leq k!$

There should be a multiple of d not very far after b: there is a $0 \leq c \leq k!$ such that d divides $b+c$. Observe $b+c+(k-1)d \leq b+k!+k!(k-1) = b+k!k$. Set $A = \{b+c, b+c+d, ..., b+c+(k-1)d\}$, by the inequality above $A \subseteq \{b, b+1, ..., b+k!k\}$. Therefore every element of A is a quadratic residue. Since d was a quadratic non-residue, its multiplicative inverse, d^*, is also a quadratic non-residue. Multiplying the elements of A with d^* we obtain the following set of consecutive quadratic non-residues: $\{\frac{b+c}{d}, \frac{b+c}{d}+1, ..., \frac{b+c}{d}+k-1\}$.

Case: $d > k!$

Divide d with $k!$ and find remainder c, i.e. $d = k!m + c$ with $0 \leq c < k!$, and by the Case we are in, $m \geq 1$ and $c < d$. Therefore, for every $1 < j \leq k$ we have

$$c - d \equiv 0 \mod j \text{ and } c - d + jd \equiv 0 \mod j.$$

Fix an $1 < j \leq k$. Then $\frac{c-d}{j} + d$ is an integer, which is at least 1. If not, $\frac{c-d}{j} + d \leq 0$ gives a contradiction. By $c < d$, $\frac{c-d}{j} + d < d$. As d was the smallest quadratic non-residue, $\frac{c-d}{j} + d$ must be a quadratic residue.

Observe that every j with $1 < j \leq k$ is quadratic residue, as $d > k! \geq k$. Therefore the product of quadratic residues $\Big(\frac{c-d}{j} + d\Big)j = c + (j-1)d$ are quadratic residues for all $1 < j \leq k$. For the missing $j = 1$ term, $c < k! < d$, and therefore c is a quadratic residue as well. Hence $c, c+d, c+2d, ..., c+(k-1)d$ are all quadratic residues but their difference d is quadratic non-residue. Multiplying them by the quadratic non-residue d^* we find k consecutive quadratic non-residues.

7 Generalizations of the van der Waerden theorem for systems of linear equations

This generalization is due to Rado (1933). An $m \times n$ integer matrix A is called *partition regular*, if for every r-coloration of positive integers, there exists a column vector $\mathbf{x}$ composed of positive integers of the same color, such that $A\mathbf{x} = \mathbf{0}$. (Compactness arguments show, that if A is partition regular, then exists a number $N = N(A, r)$, such that every r-coloration of $[N]$ already exhibits the required monochromatic solution.)

An $m \times n$ integer matrix $A = [\mathbf{a}_1, ..., \mathbf{a}_n]$ has the *columns property* if its columns can be partitioned into sets $\{I_1, ..., I_t\}$, such that for all i

$$\sum_{\mathbf{v} \in I_i} \mathbf{v} \in Span(\cup_{\ell=1}^{i-1} I_\ell),$$

where the span is understood with rational coefficients. In particular, the span of the emptyset is the zero vector, and therefore for $i = 1$, $\sum_{\mathbf{v} \in I_1} \mathbf{v} = \mathbf{0}$.

Theorem 14 [Rado] *Assume that A is a matrix with integer entries. The matrix A has the columns property if and only if it is partition regular.*

We skip the proof of Theorem 14. Assume now that A has only one row. In this case the column property is equivalent to the existence of an $\mathbf{x}$ vector with 0-1 components, such that $A\mathbf{x} = \mathbf{0}$, but $\mathbf{x} \neq \mathbf{0}$ (see the homework problems). This criterion applies Schur theorem: for all k there exists an $N(k)$, such that any k-coloration of $[N(k)]$ admits a monochromatic solution to the equation $x + y - z = 0$. We have $A = [1, 1, -1]$ and indeed, $1 + 0 = 1$ is a 0-1 solution of the equation. However, the equation $x + y - 2z = 0$, for which $A = [1, 1, -2]$, has a non-trivial 0-1 solution $1 + 1 - 2 \cdot 1 = 0$, which implies the existence of a monochromatic solution. However, there are trivial monochromatic solutions, for example $x = y = z = 5$, which is monochromatic in any coloration, being a single entry, making an arithmetic progression of difference 0. Therefore partition regularity does not prove in itself van der Waerden's theorem for 3-term arithmetic progressions.

Brauer's theorem, however, can be proved using a multirow matrix A, which in turn proves van der Waerden's theorem. I show it for 5-term arithmetic progressions, the general case is similar. We want to find a monochromatic positive integer solution to

$$\begin{aligned} u_1 + d &= u_2 \\ u_2 + d &= u_3 \\ u_3 + d &= u_4 \\ u_4 + d &= u_5. \end{aligned}$$

Set

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

and $\mathbf{x}^T = [d, u_1, u_2, u_3, u_4, u_5]$, and we want to solve $A\mathbf{x} = \mathbf{0}$ in monochromatic positive integers. Rado's theorem applies, as $I_1 = \{\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6\}$ and $I_2 = \{\mathbf{a}_1\}$ shows that A has the columns property through $\mathbf{a}_1 = 5\mathbf{a}_2 + 4\mathbf{a}_3 + 3\mathbf{a}_4 + 2\mathbf{a}_5 + 1\mathbf{a}_6$. Now d is a positive integer, so we cannot get a constant arithmetic progression.

Another application of Rado's method is Folkman's theorem, which was indeed discovered multiple times.

Theorem 15 *For all $m \geq 1$ and $r \geq 1$, there is an $N(m, r)$, such that for every r-coloration of $[N(m, r)]$, there are $x_1 < x_2 < \cdots < x_m$ positive integers, such that $x_1 + x_2 + \ldots + x_m \leq N(m, r)$, and all $\sum_{i=1}^{m} \epsilon_i x_i$ sums, where $\epsilon_i = 0$ or 1, but not all zero, these $2^m - 1$ sums all have the same color.*

For $m = 2$, Folkman theorem gives back Schur's theorem. I give the proof for $m = 3$ using Rado's theorem. We want to find a monochromatic positive integer solution to

$$\begin{aligned} u_{12} &= u_1 + u_2 \\ u_{23} &= u_2 + u_3 \\ u_{13} &= u_1 + u_3 \\ u_{123} &= u_1 + u_2 + u_3. \end{aligned}$$

We have a problem, however, like with proving van der Waerden's theorem: we may find a solution, where u_1, u_2, u_3 not all distinct, as required. We add 2 more equations (and require two more numbers be of the same color!):

$$\begin{aligned} u_2 - u_1 &= y_1 \\ u_3 - u_2 &= y_2. \end{aligned}$$

Now take

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

and $\mathbf{x}^T = [u_{12}, u_{23}, u_{13}, u_{123}, u_1, u_2, u_3, y_1, y_2]$ and we want to solve $A\mathbf{x} = \mathbf{0}$ in monochromatic positive integers. Rado's theorem applies, as $I_1 = \{\mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_7, \mathbf{a}_9\}$, $I_2 = \{\mathbf{a}_1, \mathbf{a}_6, \mathbf{a}_8\}$, $I_3 = \{\mathbf{a}_5\}$.

8 Homework Problems

vw1) Count the number of combinatorial lines in A^n.
vw2) If you are familiar with the Lovász Local Lemma, derive a lower bound for $W(p+1, k)$.
vw3) Assume now that A has only one row. Show that in this case the column property is equivalent to the existence of an $\mathbf{x}$ vector with 0-1 components, such that $A\mathbf{x} = \mathbf{0}$, but $\mathbf{x} \neq \mathbf{0}$.
vw4) Prove Folkman's theorem, either as it is stated, or in a weaker version, when the x_i terms may coincide.
vw5) For which $W(t+1, 2)$ numbers can we obtain better bounds from Berlekamp's theorem than from the Lovász Local Lemma? (In other words, how close t should be to a prime number?)
ww6) Write the number $i \in [0, p^n - 1]$ as $i = a_0 + a_1 p + a_2 p^2 + ... + a_{n-1} p^{n-1}$ in a base p representation, where $a_i \in \{0, 1, 2, ..., p-1\}$. Now a p-term arithmetic progressions $\{a, a+d, a+2d, ..., a+(p-1)d\}$ in $[0, p^n - 1]$ arises from a combinatorial line among length n words over the alphabet $\{0, 1, 2, ..., p-1\}$ if and only if (i) the base p representation of d uses only 0 and 1 digits, and (ii) at any position, where d uses 1, the base p representation of a uses 0.