Ramsay Theory
and
Related Topics

Period III, Spring 2015

J. Karhumäki

Contents

The course considers a few "jewels" of combinatorics. The results proved are similar showing

"unavoidable regularities"

or

"impossibility of complete disorder".

More intuitively the above means that

"Any large enough structure contains some (still large) regular substructure".

The results proved are

Ramsey's Theorem *which says that any large enough edge coloured graph contains a large monochromatic subgraph.*

Van der Waerden's Theorem *which says that if* $\mathbb{N}$ *is coloured by a finite number of colors, it contains arbitrarily long monochromatic arithmetic progressions.*

Shirshov's Theorem: *Any long enough word (i.e. a sequence of symbols from a finite alphabet) is either highly periodic (i.e. contains a repetition of high order) or is minimal (in certain sense).*

These results were proved in years 1930, 1927 and 1957, respectively. Amazingly each of the authors were working in different fields than combinatorics!!

Besides the above results we shall prove some related results as well as look after applications. This is partially (I hope) done via seminar presentations.

Related problems were considered in "Combinatorics on Words" under the title *unavoidability*. There it was shown that

"Each long enough word (in fact longer than 3) over a binary alphabet contains a repetition, i.e. a square";

"There exists an infinite word over a binary alphabet which does not contain a cube (in fact even a pattern of the form $uu\text{first}(u)$)";

"There exists an infinite word over a ternary alphabet which does not contain a square".

Consequently, the regularity defined by "squares" is avoidable in ternary alphabets, but not in binary ones, while that of "cubes" is avoidable even in binary alphabets. The above results are from Thue from 1906.

Related literature:

- R.L. Graham, B.L. Rothschild, J.H. Spencer, Ramsey Theory, John Wiley & Sons, 1990.
- L. Lovasz, Combinatorial Problems and Excersices, North-Holland 1979.
- M. Lothaire, Combinatorics on Words, Addison-Wesley, 1983
- M. Lothaire, Algebraic Combinatorics on Words, Cambridge University press, 2002.
- A. de Luca and S. Vanicchio, Finiteness and Regularity in semigroups and Formal Languages, Springer, 1999.

I Simple Examples

Pigeon Hole Principle (PHL). *If n items are put into $n-1$ boxes, at least one box contains at least two items.*

Pigeon Hole Principle, infinite variant. *If infinitely many items are put into a finitely many boxes, at least one box contains infinitely many items.*

The above cases are trivial. However, in the above setting one can find quite nontrivial results.

Example 1. Assume that we want to place n balls into n boxes. Consider two variants of the problem:

(i) each individual ball is placed randomly;

(ii) in each step two boxes are chosen randomly and the ball is placed into the one containing fewer balls.

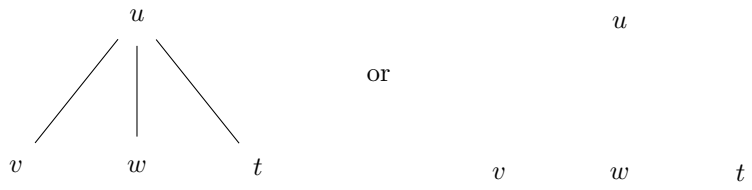
The question asks how many balls are in the box containing the highest number of balls. The answer are as follows.

In (i) the nuber is $(1+O(1))\ln n/\ln\ln n$ with a high probability (i.e. with probability $1-O(1)$).

In (ii) the number is $\ln\ln n/\ln 2+O(1)$ with a high probability.

Consequently, in the former case there are essentially exponentially more balls than in the second case in the box containing the highest number of balls.

Example 2. We claim that every set of at least six persons contains a uniform click of size at least three. Here the *click* means a set of persons where all persons either knows all the others or none of them knows any other. The proof is a simple case analysis. Denote the fact that u and w know each other by an edge between those: u—v. Then without loss of generality we have



In the former case if any two of v, w, t know each other, then we have a click. But this is true also in the other case. The same argument applies to the other case.

Ramsey's Theorem extends this result to larger numbers.

II Ramsey's Theorem

In this section we consider Ramsey's theorem and present several proofs of it (or variants of it).

We start by fixing some terminology:

$$\begin{array}{lcl} N & = & \{1,2,3,\ldots\} \\ |X| & = & \text{cardinality of } X \\ [n] & = & \{1,\ldots,n\} \text{ (or an arbitrary } n \text{ element set)} \\ [X]^k & = & \{Y | Y \subseteq X, |Y| = k\} \\ [X]^{\leq k} & = & \{Y | Y \subseteq X, |Y| \leq k\} \\ [X]^{\omega} & = & \{Y | Y \subseteq X, Y \text{ is finite}\} \\ [[n]]^k & = & [n]^k \\ K_n & : & \text{Complete graph with } n \text{ vertices.} \end{array}$$

r-coloring of S: $\chi : S \to [r]$

color of $s \in S$: $\chi(s)$

$T \subseteq S$ is *monochromatic*: $|\chi(T)| = 1$.

The following *arrow notation* is important. We write

$$n \longrightarrow (l)$$

if for any 2-coloring of $[n]^2$ there exists $T \subseteq [n]$ and $|T| = l$ such that $[T]^2$ is monochromatic. If the above holds we can identify $[T]^2$ with K_l, i.e. we can say that $[T]^2$ is a complete graph of l edges.

Example 3. Example 2 of I can be reformulated as

$$6 \longrightarrow (3).$$

Indeed, $[n]$ denotes the vertices of the graph (i.e. persons) and $[n]^2$ the edges (i.e. the relations "know each other").

The above *arrow notation* extends to: We write

$$n \longrightarrow (l_1, \ldots, l_r) \tag{1}$$

if for any r-coloring of $[n]^2$ there exists $i \in [r]$ and $T \subseteq [n]$ such that $|T| = l_i$ and T is (monochromatic and) colored by i.

More intuitively, if we write (1) then

- $[n]^2$ is r-colored,
- $[n]^2$ contains one of the graphs K_{l_i} as monochromatic, i.e. $[n]^2$ cannot avoid all monochromatic graphs $K_{l_1}, \ldots, K_{l_r}$.

In the case $l_1 = l_2 = \cdots = l_r$ we write instead of (1) simply

$$n \longrightarrow (l)_r.$$

In particular:

$$n \longrightarrow (l,l) \iff n \longrightarrow (l)_2 \iff n \longrightarrow (l).$$

Example 4. The Formula

$$10 \longrightarrow (4,3)$$

can be interpreted as: Any group of 10 persons contain either a subgroup of four persons who know mutually each other or a subgroup of three persons who do not know pairwise each other.

Simple facts:

(i) $10 \longrightarrow (4,3) \implies 10 \longrightarrow (3,3)$ "reducing a subgraph"

(ii) $10 \longrightarrow (4,3) \implies 11 \longrightarrow (4,3)$ "increasing the graph"

(iii) $10 \longrightarrow (4,3) \implies 10 \longrightarrow (3,4)$ "changing the coloring"

(iv) $10 \longrightarrow (4,3) \implies 10 \longrightarrow (4,3,2)$ "adding K_2 to a subgraph"

Parts (i)–(iii) are obvious, and part (iv) is easy to conclude. Note also that all conditions (i)–(iv) extend immediately to the general case, as illustrated by explanations.

Example 5.

$$l \longrightarrow (l,2) \quad \text{is true (by (iv))}$$
$$l-1 \longrightarrow (l,2) \quad \text{is not true.}$$

The important *Ramsey's numbers* $R(l_1, \ldots, l_r)$ are defined as follows:

$$R(l_1, \ldots, l_r) = \mu n[n \longrightarrow (l_1, \ldots, l_r)],$$

where $\mu n[\ ]$ denotes the smallest n such that $[\ ]$ is true.

We use the abbreviations:

$$R(\underbrace{l,\ldots,l}_{r \text{ copies}}) = R(l;r)$$

$$R(l;2) = R(l,l) = R(l).$$

The conditions (i)–(iv) can be reformulated:

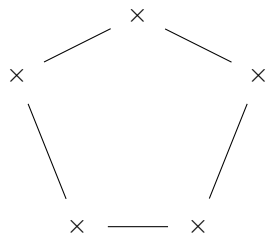
$$\begin{array}{lcl} R(l_1,\ldots,l_r) & & \text{is increasing with respect to all arguments} \\ R(l_1,\ldots,l_r) & & \text{is symmetric} \\ R(l_1,\ldots,l_r,2) & = & R(l_1,\ldots,l_r). \end{array}$$

Example 6. The known values (in 1990) of the Ramsey's numbers $R(k,l)$ are as follows:

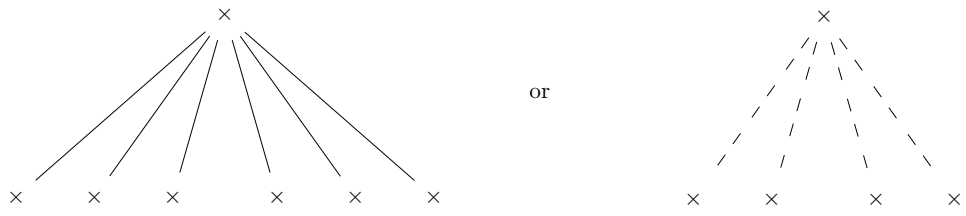
k / l	3	4	5	6	7	8	9
3	6	9	14	18	23	28	36
4		18	25	35/41	49/61	56/84	69/115
5			43/49	58/87	80/143	101/216	121/316
6				102/165	111/298	127/495	169/780

And, of course, $R(l,2) = l$. The value $R(4,4)$ was discovered in 1955, and already $R(5,5)$ is expected to be beyond the capabilities of current computers!

Example 3 (revisited). As we saw $R(3,3) \leq 6$. By the graph below here is indeed equality:



Example 7. $R(4,3) \leq 10$. Indeed,



Now on the left part we can apply the previously proved fact $R(3,3) = 6$, and on the right part it is clear that there must be either a click of 3 connected by dash-line or click of 4 connected by solid line. Hence the bound follows. Note that the above does not give any lower bound nor prove the optimality yet.

We defined the Ramsey's numbers, but did not prove yet that they exist.

Theorem 1. *$R(l,k)$ exists for each $k, l \geq 2$.*

Proof. By double induction. So we assume:

(i) $R(l,2) = R(2,l) = l$ (which is true!)

(ii) $R(l,k-1)$ and $R(l-1,k)$ are defined.

Claim: $R(l,k-1) + R(l-1,k) \longrightarrow (l,k)$ (and thus $R(l,k)$ is defined).
Proof of Claim: Set $n = R(l,k-1) + R(l-1,k)$. Further let χ be 2-coloring of $[n]^2$ and $x \in [n]$.

We define

$$\begin{aligned} I_x &= \{y \in [n] \mid \chi(x,y) = 1\}, \\ I\!I_x &= \{y \in [n] \mid \chi(x,y) = 2\} = [n] - I_x - \{x\}. \end{aligned}$$

Now,

$$|I_x| + |I\!I_x| = n - 1$$

so that

$$|I_x| \geq R(l-1,k)$$

or

$$|I\!\!I_x| \geq R(l, k-1).$$

Consider the former alternative (the other is symmetric). By i.h. there exists

$$T \subseteq I_x,\ |T| = k \text{ such that } [T]^2 \text{ is colored by } 2$$

or

$$S \subseteq I_x,\ |S| = l-1 \text{ such that } [S]^2 \text{ is colored by } 1.$$

In the former case we are done, T is a required substructure. But so are we also in the latter case, $S \cup \{x\}$ is a required substructure. □

Another proof for theorem 1 in the case $l = k$. We show directly that

$$2^{2l-1} - 1 \longrightarrow (l).$$

Fix S_1 such that

$$|S_1| \geq 2^{2l-1} - 1$$

and 2-coloring

$$\chi : [S_1]^2 \to \{1, 2\}.$$

For $i = 1, \ldots, 2l-1$ we define sets S_i and elements $x_i \in S_i$:

(i) When S_i is defined, choose $x_i \in S_i$;

(ii) When x_i is chosen, set

$$T_j = \{u \in S_i \mid \chi(x_i, u) = j\} \text{ for } j = 1, 2,$$

and choose

$$S_{i+1} = \text{larger of the sets } T_1 \text{ and } T_2.$$

Now,

$$|S_{i+1}| \geq (|S_i| - 1)/2 \text{ (since } T_1 + T_2 = |S_i| - 1),$$

and so by the choice of S_1, we have

$$|S_{2l-1}| \geq 2^1 - 1 = 1.$$

Consequently, the points $x_1, \ldots, x_{2l-1}$ are defined and pairwise disjoint.

Next we define a new coloring

$$\chi^* : \{x_1, \ldots, x_{2l-1}\} \to \{1, 2\}$$

by

$$\chi^*(x_i) = j \text{ where } \chi(x_i, y) = j, \ \forall y \in S_{i+1}.$$

χ^* divides the set $\{x_1, \ldots, x_{2l-1}\}$ into two disjoint parts. Hence one of those contains at least l points, say

$$\{x_{i_1}, \ldots, x_{i_l}\}.$$

So these are colored by the same color, say r. Next consider indices $1 \leq s < t \leq l$. Then

$$x_{i_t} \in S_{i_t} \subseteq S_{i_s+1}.$$

On the other hand, by the choice of indices and the definition of χ^* we have

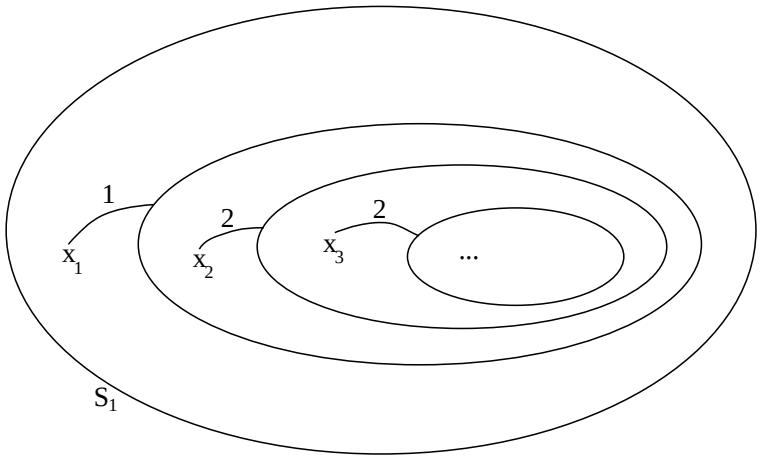
$$\chi^*(x_{i_s}) = \chi(x_{i_s}, y) \quad \forall y \in S_{i_s+1}.$$

Therefore

$$\chi(x_{i_s}, x_{i_t}) = r \quad \forall s, t,$$

showing that we have found a monochromatic subgraph of l elements. □

The idea behind our proof is so called *Induced coloring method*:



Both of the above proofs can be modified for the general case:

Theorem 2. *All Ramsey's numbers $R(l_1, \ldots, l_r)$ exist.*

Proof. (Extension of that of Theorem 1 for r-colorings.) We have to show

$$n \longrightarrow (l_1, \ldots, l_n)$$

for large enough n.

Method I. We set

$$n = 2 + \sum_{i=1}^{r} (R(l_1, \ldots, l_i - 1, \ldots, l_r) - 1).$$

Then the earlier arguments work.

Method II. We prove directly

$$r^{(l-1)r+1} - 1 \longrightarrow (l; r).$$

Now, as in the case $r = 2$, we construct points x_i and sets S_i. Here the sets T_j, for $j = 1, \ldots, r$, are constructed as intermediate steps, and S_{j+1} is chosen to be the largest of these. Then

$$\begin{aligned}
|S_{i+1}| &\geq \left\lfloor \frac{|S_i|}{r} \right\rfloor, \\
|S_1| &\geq r^{r(l-1)+1} - 1, \\
|S_2| &\geq \left\lfloor r^{r(l-1)} - \frac{1}{r} \right\rfloor = r^{r(l-1)} - 1, \\
&\cdots \\
|S_{r(l-1)+1}| &\geq r^1 - 1.
\end{aligned}$$

So the points $x_1, \ldots, x_{r(l-1)+1}$ are defined. For this set we define the induced coloring χ^*. That derives a monochromatic subsequence $x_{i_1}, \ldots, x_{i_l}$ of length l, which, in turn, defines, by the construction, monochromatic complete subgraph of $[n]^2$ of size l. □

Next we extend Theorems 1 and 2 to colorings of k-element subsets, that is we consider colorings of $[n]^k$. We write

$$n \longrightarrow (l_1, \ldots, l_r)^k,$$

if for any r-coloring of $[n]^k$, there exists an index i, $1 \leq i \leq r$, and subset $T \subseteq [n]^k$ such that $|T| = l_i$ and $[T]^k$ is colored by i.

In the case $l = l_1 = \cdots = l_r$ we simply write

$$n \longrightarrow (l)^k_r$$

and we can say that any r-coloring of $[n]^k$ yields a monochromatic $[l]^k$. Further in the case $r = 2$ we have

$$n \longrightarrow (l)^k \iff n \longrightarrow (l)^k_2 \iff n \longrightarrow (l,l)^k.$$

The *generalized Ramsey numbers* are defined as:

$$\begin{aligned} R_k(l_1,\ldots,l_r) &= \mu n[n \longrightarrow (l_1,\ldots,l_r)^k], \\ R_k(l;r) &= \mu n[n \longrightarrow (l)^k_r], \\ R_k(l) &= \mu n[n \longrightarrow (l)^k] \qquad (2-\text{colorings}). \end{aligned}$$

Of course the conditions (i)–(iv) in page 5 extend here, too.

Theorem 3. *Generalized Ramsey numbers are defined.*

Proof. By induction on k using the method of induced colorings.

$k = 1$: Trivial (by PHP): If $\sum_{i=1}^{r}(l_i - 1) + 1$ balls are placed into r boxes (e.g. colored by r colors), then at least one box contains at least l balls. Consequently

$$1 + \sum_{i=1}^{r}(l_i - 1) \longrightarrow (l_1,\ldots,l_r)^1.$$

$k = 2$: Was proved for clarity (although it is part of the induction step).

Induction step: Assume that the theorem is true for $(k-1)$-element sets. By the monotonicity properties it is enough to prove

$$n \longrightarrow (l)^k_r$$

for large enough n.

Assume that n is "large enough" (which will be specified later). Let

$$\chi: \quad [n]^k \to \{1,\ldots,r\}$$

be an r-coloring. By induction assumption, let

$$t = R_{k-1}(l;r).$$

We choose arbitrary elements $a_1, \dots, a_{k-2} \in [n]$ and denote

$$S_{k-2} = [n] \setminus \{a_1, \dots, a_{k-2}\}.$$

Finally, we define points a_i and sets S_i as follows:

(i) If S_i is defined, then choose any $a_{i+1} \in S_i$;

(ii) If a_{i+1} is defined, devide $S_i \setminus \{a_{i+1}\}$ into equivalence classes by

$$x \equiv y \iff (\forall T \subseteq \{a_1, \dots, a_{i+1}\}, |T| = k-1) \quad \chi(T \cup \{x\}) = \chi(T \cup \{y\}).$$

And choose S_{i+1} equal to (one of) the maximal classes.

Claim: The number of these equivalence classes is at most $r^{\binom{i+1}{k-1}}$.

Indeed, there are $\binom{i+1}{k-1}$ $(k-1)$-element subsets T. Each r coloring of these subsets yields an equivalence class. Hence the claim follows.

It follows from the construction that

$$\begin{aligned} S_{i+1} &\subseteq S_i \setminus \{a_{i+1}\}, \text{ and} \\ |S_{i+1}| &\geq \frac{(|S_i| - 1)}{r^{\binom{i+1}{k-1}}}. \end{aligned} \tag{2}$$

Now, the requirement for n is: a_t must be defined, that is the S_i sets are not allowed to be empty. The existence of such an n is guaranteed, if we require that the recursion

$$u_{i+1} = \frac{(u_i - 1)}{r^{\binom{i+1}{k-1}}}, \quad u_{k-2} = n - (k-2)$$

yields $u_t \geq 1$. Surely

$$n = 2r^{\sum_{i=k-1}^{t-1} \binom{i+1}{k-1}}$$

suffices.

Next consider the sequence

$$a_1, a_2, \dots, a_t$$

thus defined. Assume that

$$1 \leq i_1 < i_2 < \cdots < i_{k-1} < s \leq t.$$

Then, by (i) and (2)

$$a_s \in S_{s-1} \subseteq S_{i_{k-1}+1}.$$

Further by the definition of the equivalence relation:

$$\chi(a_{i_1}, \ldots, a_{i_{k-1}}, a_s) = \chi(a_{i_1}, \ldots, a_{i_{k-1}}, x), \quad \forall x \in S_{i_{k-1}+1}. \tag{3}$$

In particular, (3) is true for $x = a_r$, where $i_{k-1} < r < t$. It follows that we can define the coloring χ^* of $(k-1)$-element subsets of $\{a_1, \ldots, a_t\}$ by the condition

$$\chi^*(a_{i_1}, \ldots, a_{i_{k-1}}) = \chi(a_{i_1}, \ldots, a_{i_{k-1}}, a_s), \quad \forall i_{k-1} < s \leq t. \tag{4}$$

Observe that in above $i_{k-1} < t$. If $i_{k-1} = t$ we can choose that value of χ^* arbitrarily!

We are almost done. By induction hypothesis and the choice of t, the sequence $a_1, \ldots, a_t$ has a subsequence $b_1, \ldots, b_l$, which is monochromatic under χ^*, that is to say:

- Each $(k-1)$-element subset of $\{b_1, \ldots, b_l\}$ has the same color, say red, under χ^*.

Then for all sequences of indices $1 \leq j_1 < \cdots < j_{k-1} < j_k \leq l$, we have

$$\chi(b_{j_1}, \ldots, b_{j_{k-1}}, b_{j_k}) = \chi^*(b_{j_1}, \ldots, b_{j_{k-1}}) = \text{red}.$$

This follows from (4).

So we have found an l-element subset of $[n]$ such that its all k-element subsets have the same color under our original coloring χ. This ends the proof. □

Let us consider two applications:

Example 8. Consider a totally ordered set A (e.g. each pair of elements are comparable, and no repetitions). We claim that each long enough sequence $a_1, \ldots, a_n$ contains a monotonic subsequence of length l.

Indeed, this follows directly from the existence of $R(l, l)$:
Define the coloring

$$\begin{aligned} \chi(i,j) &= \text{red if } a_i < a_j, \\ \chi(i,j) &= \text{blue if } a_i > a_j. \end{aligned}$$

By no means, $R(l,l)$ does not give an optimal bound for n. It can be shown that $(l-1)^2+1$ suffices.

Secondly, we give a solution to the "Budapest Problem" of Erdös.

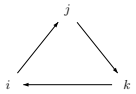
Example 9 (Budapest Problem). Given n there exists N such that out of any N points in the plane so that no three are on the same line, one can choose n points forming convex n-polygon.

We claim that we can choose

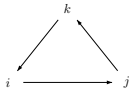
$$N \longrightarrow (n)^3,$$

that is the result follows from the existence of $R_3(n,n)$. Assume that we have N points ordered as $1,2,\ldots,N$. We define a coloring χ of 3-element subsets:

$\chi(i,j,k) = \text{red}$, if $i < j < k$ and the path $i \to j \to k \to i$ is clockwise:

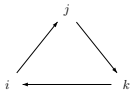


$\chi(i,j,k) = \text{blue}$, if $i < j < k$ and the path $i \to j \to k \to i$ is counterclockwise:



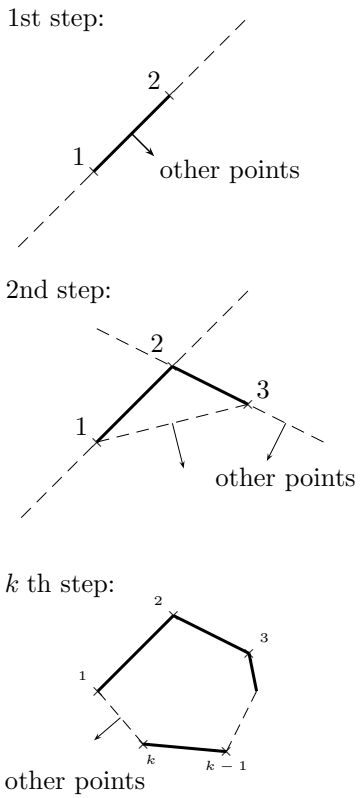
Now, by Theorem 3, there exists an n-element subset of N such that the above triangles have the same orientations. Consequently, by renaming the points we have the points $1,\ldots,n$ such that:

whenever $i < j < k$ then the orientation is:



The convex hull of these points is a required n-polygon. Indeed we can

construct this step by step:



So at each step we obtain a convex hull of the "first points" and the others are "outside".

Next result is an infinitary variant of Ramsey's Theorem.

Theorem 4. *For each finite coloring χ of $[N]^2$, there exists an infinite subset $A \subseteq N$ such that $[A]^2$ is monochromatic under χ.*

Proof. Very similar to the previous ones. In fact, even notationally clearer, since the numerical estimates are not needed. As earlier we define, for each i, sets S_i and elements $x_i \in S_i$ as follows:

(i) $S_1 = N$,

(ii) If S_i is defined, we choose $x_i \in S_i$,

(iii) If $x_i \in S_i$ is defined, set

$$T_j = \{u \in S_i \mid \chi(x_i, u) = j\} \quad \text{for colors } j.$$

These sets define a partition of $S_i \setminus \{x_i\}$. We choose S_{i+1} to be one of the infinite classes of this partition. Now, as in the proof of Theorem 1, the sequence $x_1, x_2, \ldots$ (by the conditions $x_j \in S_j \subseteq S_{i+1}$, $x_k \in S_k \subseteq S_{i+1}$ and $\chi(x_i, u) = \text{constant}$ for $u \in S_{i+1}$) satisfies:

$$\chi(x_i, x_j) = \chi(x_i, x_k), \quad \forall i < j, k.$$

Therefore we can define a coloring of $X = \{x_1, x_2, \ldots\}$ by the condition:

$$\chi^*(x_i) = \chi(x_i, x_j), \quad \forall i < j.$$

Now χ^* is a finite coloring of an infinite set so that there exists

$$X' = \{x_{i_1}, x_{i_2}, \ldots\} \subseteq X \text{ and } j$$

such that

$$|X'| = \infty$$

and

$$\chi^*(x_{i_s}) = j, \quad \forall s.$$

Then

$$\chi(x_{i_s}, x_{i_t}) = \chi^*(x_{i_s}) = j$$

so that $[X']^2$ is monochromatic under χ. □

The above proof extends directly to colorings of k-element sests.

Theorem 4 claims that "Every infinite set contains an infinite (and thus large) regular subset". A natural question is whether this infinitary variant implies a corresponding finite one: "Every large enough set contains large (in advance fixed size) regular subset".

The answer is: not directly, but relatively easily by so-called *compactness principle*. We formulate this for k-element sets.

Theorem 5 (Compactness principle). *Let $k \in N$ and $\mathcal{A}$ a family of finite subsets of N. Assume further that, for any finite coloring of $[N]^k$, there exists a subset $A \in \mathcal{A}$ such that $[A]^k$ is monochromatic. Then, for each number $r \in N$, there exists $n_0 = n_0(r)$ such that $\forall n \geq n_0$: For any r-coloring of $[n]^k$ there exists $A \in \mathcal{A}$ such that $[A]^k$ is monochromatic.*

Proof. Assume the contrary that no n_0 exists. Then there exists an infinite sequence of

$$r\text{-colored subsets of } [n]^k \tag{5}$$

such that none of those contains $[A]^k$, $A \in \mathcal{A}$, as monochromatic.

Fix the order of elements of $[N]^k$: $[N]^k = \{y_1, y_2, \ldots\}$. Next consider such sets of (5), where the color of y_1 is the same. Some of these sets is infinite. Now replace (5) by this new infinite set and continue with y_2, and so on. This procedure defines a coloring

$$\chi : \quad [N]^k \to \{1, \ldots, r\}.$$

Then, by the assumption of the theorem, there exists a monochromatic A under χ. Let $t \subseteq N$ be such that

$$A \subseteq \{y_1, \ldots, y_t\}$$

and

$$(\chi_0[n_0]^k) \text{ an element of (5) determined by } y_t.$$

It follows from the construction that

$$\chi(y_i) = \chi_0(y_i) \quad \text{for } i = 1, \ldots, t.$$

Consequently, A is monochromatic with respect to χ_0. Contradiction! □

The above compactness principle together with Theorem 4, when applied to

$$\mathcal{A} = \{A \mid l \leq A < \inf\}, \quad l \in N$$

yields

Corollary 1. Let $r, l \in N$. There exists $n_0 = n_0(l, r)$ such that

$$\forall n \geq n_0 : \quad n \longrightarrow (l)_r.$$

What we obtained is a mathematically simpler proof for Theorem 1 with arbitrary number of colors. However, this method *does not give* any bound for the number n, unlike direct combinatorial arguments.

We have considered so far different variants of just one theorem, Ramsey's Theorem and its proof. Actually there are quite a collection of *Ramsey-Type Theorems*. We list here a few such results, and will consider two of those in more details in the next sections.

I Ramsey's Theorem *For each triple (l, r, k), there exists n_0 such that whenever $n \geq n_0$ and $[n]^k$ is r-colored there exists a monochromatic $[l]^k$.*

II van der Waerden's Theorem *For each pair (l, r), there exists n_0 such that whenever $n \geq n_0$ and $[n]$ is r-colored there exists an arithmetic progression $\{a_0, a_0 + d, \ldots, a_0 + (l-1)d\} \subseteq [n]$ of length l which is monochromatic.*

III Schur's Theorem *For each number r there exists n_0 such that whenever $n \geq n_0$ and $[n]$ is r-colored there exists monochromatic x, y and z such that $x + y = z$.*

The above extends as follows. We say that a system of equations is *regular*, if for each number r there exists n_0 such that whenever $n \geq n_0$ and $[n]$ is r-colored the system has a monocromatic solution $x_1, \ldots, x_m \in [n]$.

IV Rado's Theorem *An equation*

$$c_1 x_1 + c_2 x_2 + \cdots + c_n x_n = 0$$

is regular if and only if some of repetition-free sums of the coefficients equals to zero.

V Hales–Jewett's Theorem *For each pair (r, k), there exists n_0 such that whenever $n \geq n_0$ and n-dimensional cube*

$$C_k^n = \{(x_1, \ldots, x_n) \mid x_i \in \{0, \ldots, k-1\},\ 1 \leq i \leq n\}$$

is r-colored, it contains a monochromatic line.

VI Graham–Leeb–Rothschild's Theorem *Let F be a finite field with k elements. Then, for each triple (k, l, r), there exists n_0 such that for all $n \geq n_0$ the following holds: Let V be n-dimensional vector space over F and χ an r-coloring of k-dimensional subspaces of V. Then there exists an l-dimensional subspace of V such that its k-dimensional subspaces are monochromatic.*

In Theorem V a *line* is a set of points $\{x_0, \ldots, x_{k-1}\}$, where $x_i = (x_{i1}, \ldots, x_{in})$, satisfying (possibly after reindexing): For each j $(1 \le j \le n)$

$$\begin{aligned} &\text{either} \quad x_{0j} = x_{1j} = \cdots = x_{k-1,j}, \\ &\text{or} \qquad x_{sj} = s \text{ for } 0 \le s < k \end{aligned}$$

and the second condition holds at least once. For example,

$$\{020, 121, 222, 323\} \quad \text{and} \quad \{031, 131, 231, 333\}$$

are lines.

III van der Waerden's Theorem

The original van der Waerden's Theorem, conjectured by I. Schur, is from the year 1927.

Claim A: *If the set N of natural numbers is divided into two parts, one part contains arbitrarily long arithmetic progressions.*

The above can be generalized by

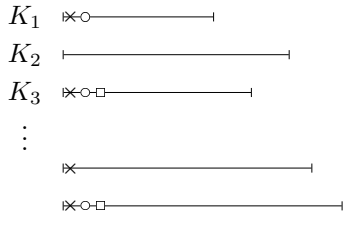
- using r parts, e.g. r-colorings of N;
- coloring only initial segments.

We obtain:

Claim B: *For each pair (k, r) of natural numbers there exists a number $W(k, r)$ such that if $\{1, \ldots, W(k, r)\}$ is divided into r parts, one part contains an arithmetic progression of length k.*

The Claims A and B are equivalent (assuming also in A r parts):
Implication $B \Rightarrow A$: Clear (since r-coloring of N is that of any initial part).
Implication $A \Rightarrow B$: Compactness argument:
Assume the contrary. Then there exists a pair (k, r) such that for all n there exists an r-coloring of $\{1, \ldots, n\}$ such that it does not contain a monochromatic arithmetic progression of length k. This means that there exists an infinite sequence $K_1, K_2, K_3, \ldots$ of such r-colored sets. From this we derive an r-coloring of N contradicting the claim A:

Since $(K_i)_{i \geq 1}$ is infinite and we use only finitely many colors there exists an infinite subsequence of $(K_i)_{i \geq 0}$ where the first elements have the same color. The procedure can be repeated. Thus we obtain an r-coloring of N. If Claim A holds for this coloring, then, by the construction, some K_i would contain an arithmetic progression of length k.



Next we consider a few first values of $W(k,r)$:

$W(2,r) = r+1$: clear

$W(3,2) = 325$: (To be read : 325 works here!)

We define

$$\begin{aligned}[1,325] &= [1,5] \cup [6,10] \cup \cdots \cup [321,325] \\ &= B_1 \cup B_2 \cup \cdots \cup B_{65}\end{aligned}$$

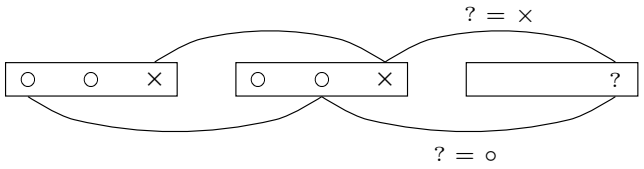
Since $r = 2$ and $|B_i| = 5$ there are $2^5 = 32$ different colors for B_i's. Consequently, there exists t, s with $s < t$ such that

- B_t and B_s have the same color,
- B_{t+t-s} is defined.

Now consider the three first elements of B_s. Two of those have the same color, say j and $j+d$ are those elements. Now $j+2d$ is in B_s. If this has the same color than j we are done, the progression is even in B_s. Otherwise we consider the element

$$j + 2d \text{ in } B_{t+t-s}$$

and we find independently of the color of $j+2d$ in B_{t+t-s}, the required progression:



$W(3,3) = 7 \cdot (2 \cdot 3^7 + 1) \cdot (2 \cdot 3^{7(2 \cdot 3^7+1)} + 1)$ $(> 3^{14 \cdot 3^7})$:

We generalize the above construction in three steps:

First, we divide $[1, W(3,3)]$ into blocks of size $7(2 \cdot 3^7 + 1)$:

$B_1, B_2, \ldots, B_t$, $t = 2 \cdot 3^{7(2 \cdot 3^7+1)} + 1$. Since $r = 3$ there are $3^{7(2 \cdot 3^7+1)}$ different colorings of these. It follows that there exist two equaly colored blocks

$$B_{i_1} \equiv B_{i_1+d_1}, \tag{6}$$

and moreover,

$$B_{i_1+2d} \text{ is defined.}$$

In the *second* step we divide each B_i to subblocks of length 7: $B_{i,1}, \ldots, B_{i,2\cdot 3^7+1}$. These subblocks have 3^7 different colorings, so that as in step one there exists two equally colored subblocks

$$B_{i_1,i_2} \equiv B_{i_1,i_2+d_2}, \tag{7}$$

and moreover,

$$B_{i_1,i_2+2d_2} \text{ is defined inside } B_{i_1}.$$

In the *third* step we note that out of four first positions of B_{i_1,i_2} two have the same color, say

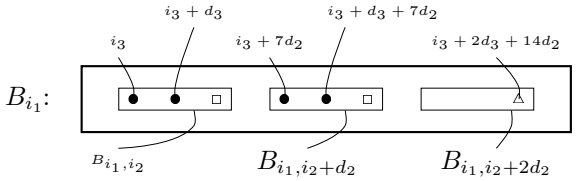
$$i_3 \equiv i_3 + d_3 = \text{red} = \bullet$$

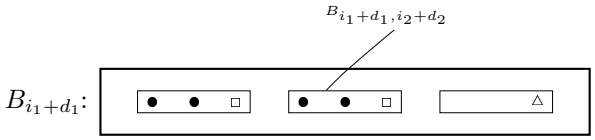
and moreover,

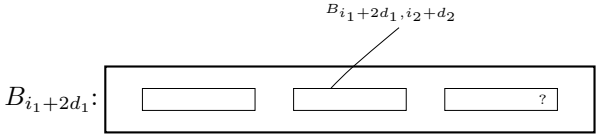
$$i_3 + 2d_3 \text{ is inside } B_{i_1,i_2}.$$

If this $i_3 + 2d_3$ would be red, we are done. so assume that this color is blue= $\square$.

We can illustrate our choices as follow: $|B_{i_1}| = 7 \cdot (2 \cdot 3^7 + 1)$, $|B_{i_1,i_2}| = 7$







Now we consider the block $B_{i_1,i_2+2d_2} \subseteq B_{i_1}$. By (7)

$$i_3 \equiv i_3 + 7d_2 = \text{red} = i_3 + d_3 + 7d_2 \equiv i_3 + d_3 = \bullet$$

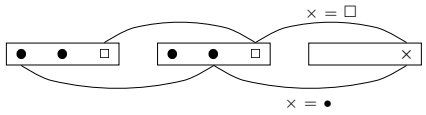
and

$$i_3 + 2d_3 \equiv i_3 + 2d_3 + 4d_2 = \text{blue} = \square.$$

Now

$$i_3 + 2d_3 + 14d_2 = \text{yellow} = \triangle$$

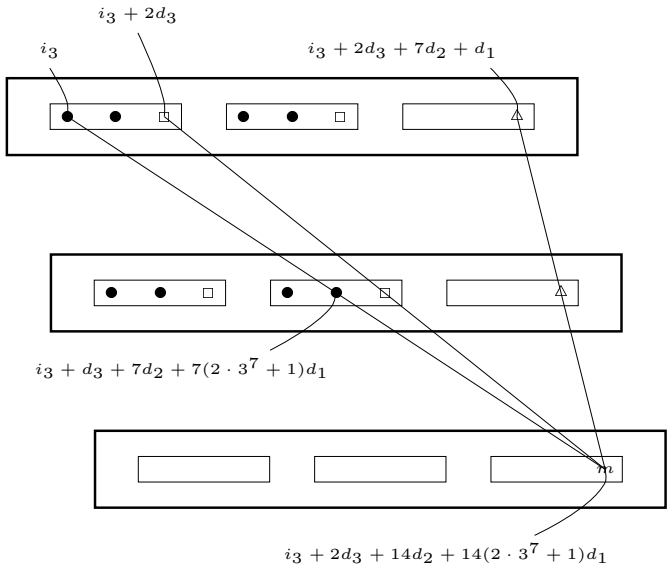
since otherwise we have:



that is a required arithmetical progression.
By (6),

$$B_{i_1} \equiv B_{i_1+d_1}$$

so that the colors in the figure below are as indicated:



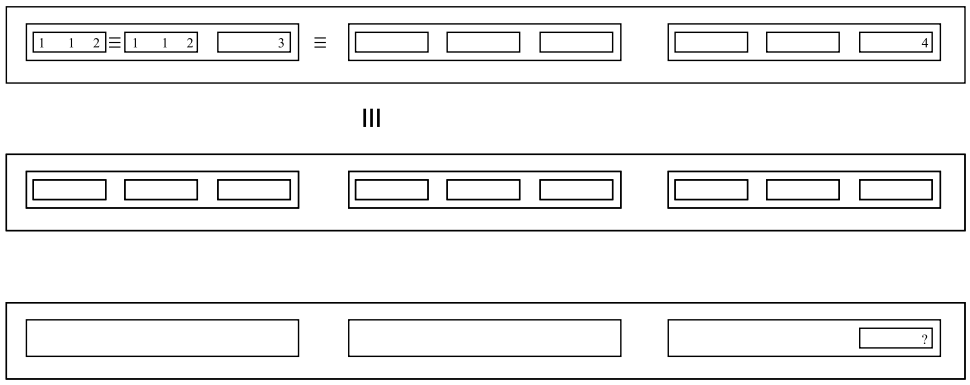
For example, the fourth red in $B_{i_1+d_1}$ is $i_3 + d_3 + 7d_2 + 7(2 \cdot 3^7 + 1)d_1$.
Now, let us consider the color of

$$m = i_3 + 2d_3 + 14d_2 + 14(2 \cdot 3^7 + 1)d_1.$$

Independently what it is we have an arithmetic progression of length 3 as shown by the lines of the figure.

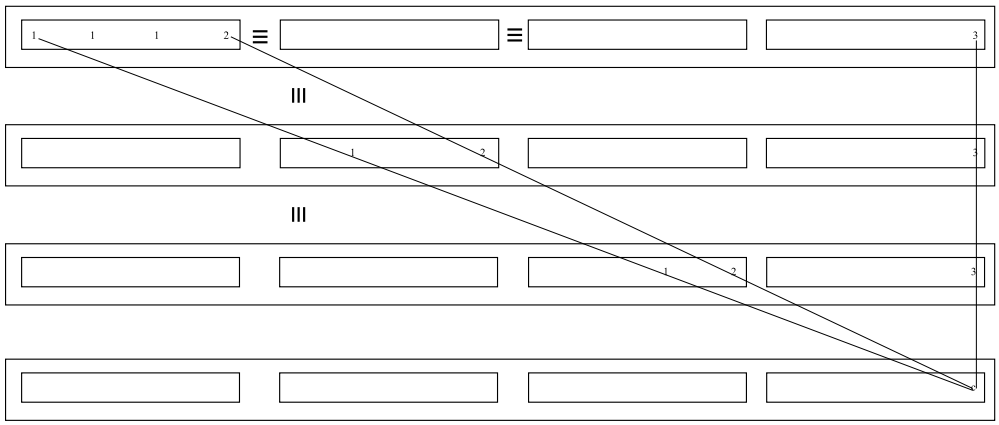
The above indicates already the general proof. However, let us consider it on intuitive level a bit more. More precisely consider the numbers $W(3,4)$ and $W(4,3)$.

$W(3,4)$: The method is that we *force a new color* (or the required progression is found). In this case this is illustrated as follows:



Indeed, 2 is forced by the first inmost block, 3 by the inmost equality, 4 by the middle level equality and ? to be a new color by the positions of the blocks.

$W(4,3)$: Follows from the figure:



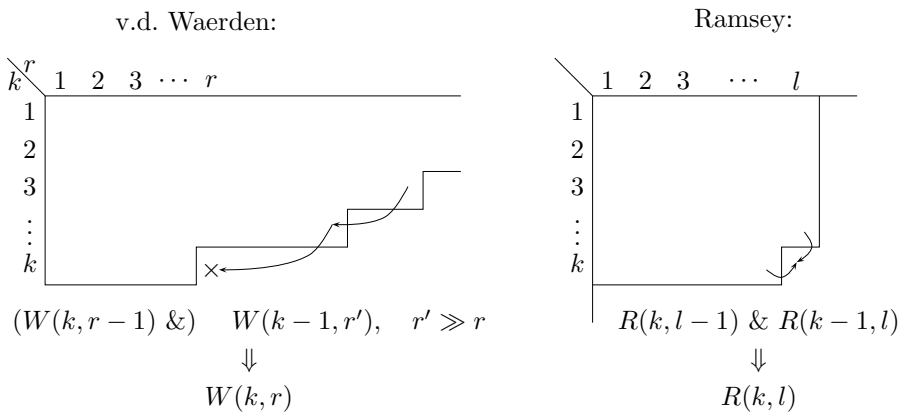
In general: $W(k,r)$ has to be chosen so large that

(i) r'-colored progression of length $k-1$ is guaranteed, and

(ii) the kth position of the above progression is defined.

Here r' is the number of all colorings of blocks of certain length, so r' is of form r^k.

More concretely the proof of the existence of the numbers $W(k,r)$ is by *double induction*. However, there are clear differences to the proof of the existence of Ramsey's numbers:



In the above approach the dificulty in the general proof is to find proper notation. This can be avoided by proving a *more general* result. At the same time the proof becomes *shorter but less intuitive*.

We proceed as follows. Consider the set $[0,l]^m$. We define $m+1$ disjoint subsets of it, so-called *l-equivalence classes* $X(i)$, $i = 0, \ldots, m$, as follows:

$$X(i) = \{(x_1, \ldots, x_m) \in [0,l]^m \mid x_1 = \cdots = x_i = l \ \& \ x_{i+1}, \ldots, x_m \neq l\}$$

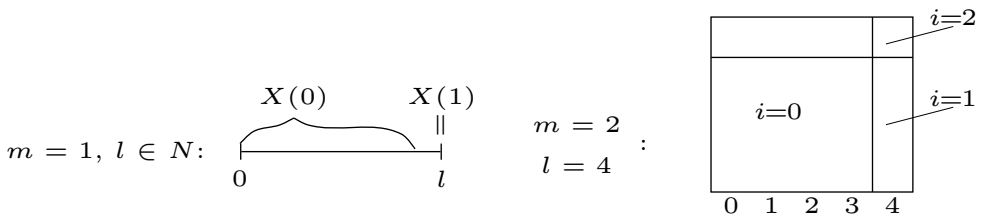
Clearly,

$$X(i) \cap X(i') = \emptyset \quad \text{if } i \neq i'$$

and

$$\bigcup_{i=0}^{m} X(i) \subsetneq [0,l]^m$$

Example 10. We have:



We shall prove:

Claim $S(l,m)$: For each number r there exists a number $N(l,m,r)$ such that, for any r-coloring,

$$\chi : [1, N(l,m,r)] \to [1,r]$$

there exist positive integers $a, d_1, d_2, \ldots, d_m$ such that

$$|\chi(a + \sum_{i=1}^{m} x_i d_i)| = 1 \text{ for } (x_1, \ldots, x_m) \in X(i)$$

for $i = 0, \ldots, m$.

Note that $N(l,1,r) = W(l,r)$ since then $m = 1$ and x_i goes from 0 to $l-1$, that is

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

Theorem 6. *$S(l,m)$ holds true for all l and m.*

Proof. Clearly $S(1,1)$ is obvious.
We show

(i) $S(l,m')$, $m' \leq m \implies S(l,m+1)$, and

(ii) $S(l,m)\ \forall m \implies S(l+1,1)$.

Now fix r arbitrarily.

Implication (i). Assume (i.h.) that

$$M = N(l,m,r) \text{ and } M' = N(l,1,r^M)$$

We claim that we can choose:

$$N(l,m+1,r) = M(M'+1).$$

So let

$$\chi : [1, (M'+1)M] \to [1,r]$$

be an r-coloring. This induces a coloring

$$\chi' : [1, M'] \to [1, r^M]$$

by

$$\chi'(k) = \chi'(k') \iff \chi(kM+j) = \chi(k'M+j) \text{ for } j \in (0, M].$$

Clearly, χ' is well defined.

By i.h. there exists a' and d' such that

$$\chi'(a' + xd') = \text{constant on } x \in [0, l-1] \text{ (or } x = l).$$

Now we can apply i.h. $S(l, m)$ to the interval $[a'M + 1, (a' + 1)M]$: There exist numbers $a, d_1, \ldots, d_m$ such that

$$a + \sum_{i=1}^{m} x_i d_i \in [a'M + 1, (a' + 1)M] \text{ for } x_i \in [0, l]$$

and

$$\chi(a + \sum_{i=1}^{m} x_i d_i) = \text{constant on } l\text{-equivalence classes } (m + 1).$$

We set

$$d'_i = d_i \quad \text{for } i \in [1, m]$$

and

$$d'_{m+1} = d'M.$$

Then we have to show that

$$\chi(a + \sum_{i=1}^{m+1} x_i d_i) = \text{constant on } l\text{-equivalence classes } (m + 2). \tag{8}$$

First in the class where $x_{m+1} = l$ there is just one element, namely $(l, \ldots, l)$, so that (8) holds. In the other classes x_{m+1} gets the values $0, \ldots, l-1$. Let us consider such a fixed class. By the choice of a' and d'

$$\chi'(a' + x_{m+1}d') = \chi'(a') \text{ for } x_{m+1} = 0, \ldots, l-1.$$

Consequently, by the definition of χ'

$$\chi((a'+x_{m+1}d')M+j) = \chi(a'M+j) \text{ for } j \in (0, M] \text{ and } x_{m+1} \in [0, l-1]. \tag{9}$$

But by the choice of the numbers a and d_i we have:

$$\begin{cases} a + \sum_{i=1}^{m} x_i d_i &= a'M + j \quad \text{for some } j = 1, \ldots, M \\ \chi(a + \sum_{i=1}^{m} x_i d_i) &= \text{constant in } l\text{-equivalence classes} \end{cases} \tag{10}$$

Therefore (8) follows from (9) and (10):

$$a + \sum_{i=1}^{m+1} x_i d'_i = x_{m+1} d' M + a + \sum_{i=1}^{m} x_i d_i = (a' + x_{m+1} d')M + j$$

Implication (ii). We claim that we can choose

$$N(l+1, 1, r) = N(l, r, r) + C.$$

So consider the coloring

$$\chi : [1, N(l, r, r) + C] \to [1, r].$$

By i.h., there exists numbers $a, d_1, \ldots, d_r$ such that

$$a + \sum_{i=1}^{r} x_i d_i \leq N(l, r, r) \quad \text{for } x_i \in [0, l]$$

and

$$\chi(a + \sum_{i=1}^{r} x_i d_i) = \text{constant in } l\text{-equivalence classes } (r+1). \tag{11}$$

Now, by PHP there exist numbers u and v, $0 \leq u < v \leq r$, such that

$$\chi(a + \sum_{i=1}^{u} l d_i) = \chi(a + \sum_{i=1}^{v} l d_i). \tag{12}$$

Denote

$$a' = a + \sum_{i=1}^{u} l d_i$$

and

$$d' = \sum_{i=u+1}^{v} d_i.$$

There exists just 2 $(l+1)$-equivalence classes (since $m = 1$): $[0, \ldots, l]$ and $\{l+1\}$. The latter is singleton so that for sure

$$\chi(a' + xd') \text{ is constant.} \tag{13}$$

Note that in order to make (13) defined we use C in the choice of $N(l+1, 1, r)$.

What remains to be shown is that

$$\chi(a' + xd') = \text{constant for } x = 0, \ldots, l. \tag{14}$$

Now, by (12), we have

$$\chi(a' + 0d') = \chi(a + \sum_{i=1}^{u} ld_i) = \chi(a + \sum_{i=1}^{v} ld_i) = \chi(a' + ld').$$

In the l-equivalence class

$$\{(x_1, \ldots, x_r) \mid x_i = l \text{ for } i \leq u \text{ and } x_j \in [0, l-1] \text{ otherwise}\}$$

(11) is valid. So for any $x \in [0, l-1]$, we have:

$$\begin{aligned} \chi(a' + 0d') &= \chi(a + \sum_{i=1}^{u} ld_i + \sum_{i=u+1}^{r} 0d_i) \\ &= \chi(a + \sum_{i=1}^{u} ld_i + \sum_{i=u+1}^{v} xd_i + \sum_{i=v+1}^{r} 0d_i) \\ &= \chi(a' + xd'). \end{aligned}$$

Hence (14) and the whole induction step has been proved. □

The following example is important to notice.

Example 11. van der Waerden's Theorem tells that if N is finitely colored it contains arbitrarily long monochromatic arithmetic progressions. A natural question is: Does it contain infinite monochromatic arithmetic progressions? In other words is "infinite monochromatic arithmetic progression" an unavoidable regularity?

The answer is "no". A simple argument for that is as follows. Each infinite arithmetic progression AL_∞ is specified by a pair (a, d), where a corresponds the threshold and d the period, that is

$$(a, d) \leftrightarrow \{u \in N \mid u = a + jd \text{ for some } j \in N\}.$$

It follows that the cardinality of all infinite arithmetic progressions is denumerable, and therefore they can be enumerated:

$$(a_1, d_1), (a_2, d_2), (a_3, d_3), \ldots$$

But the cardinality of all r-colorings, with $r \geq 2$, is nondenumerable. Hence, we can define an r-coloring $\chi_{diag} : N \to [1, \ldots, r]$ such that:

$$\text{for any } i, \text{ two points of } (a_i, d_i) \text{ have different color.}$$

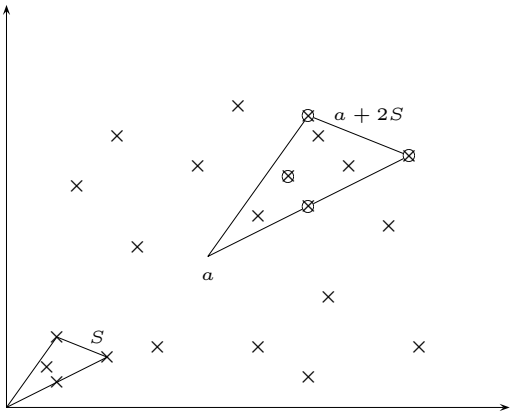
It follows, that no infinite arithmetic progression is monochromatic under χ_{diag}.

Van der Waerden's Theorem has the following extension, see Lothaire p. 41–43.

Theorem 6'. *Let $S \subseteq N^m$ with $m > 1$. Then , for any r-coloring of N^m, there exist a number d and a vector $a \in N^m$ such that $a + dS$ is monochromatic.*

By choosing $S = \{0, \ldots, l-1\}$ we obtain van der Waerden's Theorem.

Example 12. In the case $m = 2$ and $r = 2$ we have the following situation:



We continue by giving a completely different proof, a *topological one*, to van der Waerden's Theorem, see Lothaire (original proof is by Fürstenberg and Weiss, 1978). We shall prove:

Claim B *If N is divided into r classes one of those contains an arbitrarily long arithmetic progression.*

The proof goes in several steps. First let

$$\begin{aligned} \mathcal{C} &= \{C_1, \ldots, C_r\} \text{ be a partition of } N, \\ l &\quad \text{a natural number} > 1, \text{ and} \\ E &= \{u \mid u : \mathbb{Z} \to [1, \ldots, r]\}. \end{aligned}$$

We define in E a *distance* d by

$$d(u, u') = \inf\{\tfrac{1}{q+1} \mid u(n) = u'(n) \text{ for } |n| < q\}.$$

This makes E a *metric space*; for example $d(u,u')=0$ iff $u=u'$. E is also *compact*. (Each sequence has a converging subsequence).

We continue by defining $v:\mathbb{Z}\to[1,\dots,r]$:

$$v(n)=\begin{cases} t & \text{if } n\geq 0 \text{ and } n\in C_t \\ 1 & \text{if } n<0. \end{cases}$$

Our goal is to show that, for any $l>0$, there exist natural numbers n and m such that

$$(*) \qquad v(m)=v(m+n)=\cdots=v(m+ln),$$

that is $C_{v(m)}$ contains an arithmetic progression of length $l+1$.

Now, let $S:E\to E$ be a *shiftoperator*

$$(Su)(n)=u(n+1) \quad \forall n\in\mathbb{Z}.$$

Then S is continuous and bijective, that is a *homeomorphism*. We set

$$X=\lim\{S^n(v)\mid n\geq 0\}.$$

Since E is compact, $X\neq\emptyset$. Further by the definition of X it is closed, and hence compact.

We show

Lemma 1 (By Zorn's Lemma). *There exists $K\subseteq X$ such that $K\neq\emptyset$, K is closed, and $S(K)=K$ and, moreover, K is minimal (that is no proper subset of K satisfies these conditions).*

Proof. Let $\mathcal{E}$ be the family of sets satisfying the conditions of Lemma 1. It is nonempty ($X\in\mathcal{E}$) and *partially ordered* by inclusion relation. Let further

$$\mathcal{F}=\{F_i\mid i\in I\}$$

be a *totally ordered* subset of $\mathcal{E}$. We set

$$F=\bigcap_{i\in I}F_i.$$

Clearly, $F\subseteq X$ and $S(F)=F$. To show that $F\neq\emptyset$ we assume the contrary: $F=\emptyset$. Then $X\setminus F=X$, so that

$$\bigcup_{i\in I}(X\setminus F_i)$$

is an open cover of X. Hence, by the compactness of X, a finite subcover covers it. This, however, is a contradiction since $\mathcal{F}$ is totally ordered so that $\bigcap_{\text{finite}} F_i$ is nonempty.

It follows that F is a lower bound in $\mathcal{F}$, and hence by Zorn's lemma, $\mathcal{F}$ contains a minimal element. □

Now, assume that K is a minimal set guaranteed by Lemma 1. We prove the following core result:

Key Lemma *For each $\epsilon > 0$, there exist $z \in K$ and $n > 0$ such that*

$$d(S^n z, z) < \epsilon, d(S^{2n} z, z) < \epsilon, \ldots, d(S^{ln} z, z) < \epsilon.$$

Proof. By induction on l.
$l = 1$. Let $x \in K$. By the compactness, the sequence $(S^n x)_{n \geq 0}$ has an accumulation point in K. Consequently, if $\epsilon > 0$, there exist $i < j$ such that

$$\epsilon > d(S^j x, S^i x) = d(S^n z, z),$$

where $n = j - i$ and $z = S^i x$.

To conclude the induction step we need two lemmata.

Lemma 2. *For each $\epsilon > 0$ there exist integers $k_1, \ldots, k_N$ such that*

$$\forall a, b \in K : \min_{1 \leq i \leq N} d(S^{k_i} a, b) < \epsilon.$$

Proof. We make use of the minimality of K. This implies that the only subsets of K which are closed and *stable* (e.g. satisfy $S(Y) = Y$) are K and $\emptyset$. Since the notions of "open" and "closed" are complementary we may replace "closed" by "open" in our considerations.

So let ω be an open subset of K. Then

$$\bigcup_{n \in \mathbb{Z}} S^n \omega \tag{15}$$

is open (as a union of open sets), and clearly stable. Therefore it equals to K, and so by the compactness of K it has a finite subcover from (15).

Let now $\{\omega_1, \ldots, \omega_N\}$ be a cover of K by open balls of the radius $< \frac{\epsilon}{2}$. It follows from the above that, for each $i = 1, \ldots, N$, there exist integers r_i and $n_{i,j}$ such that

$$\{S^{n_{i,j}}\omega_i \mid 1 \leq j \leq r_i\}$$

is an open cover of K.

Now, take arbitrary elements $a, b \in K$. Then $b \in \omega_i$, for some i. Further

$$a \in S^{n_{i,j}}\omega_i \quad \text{for some } j.$$

Consequently,

$$S^{-n_{i,j}}a \in \omega_i$$

and so

$$d(S^{-n_{i,j}}a, b) < \epsilon.$$

Therefore Lemma 1 follows from the inequality

$$\min_{1\leq i\leq N}\left(\min_{1\leq j\leq r} d(S^{-n_{i,j}}a, b)\right) < \epsilon.$$

□

The other lemma is as follows:

Lemma 3. *For each $\epsilon > 0$ and $a \in K$, there exist $b \in K$ and $n > 0$ such that*

$$d(S^n b, a) < \epsilon, \ldots, d(S^{ln}b, a) < \epsilon.$$

Proof. By Lemma 2, there exist integers $k_1, \ldots, k_N$ such that for all $a, b \in K$:

$$\min_{1\leq i\leq N} d(S^{k_i}a, b) < \frac{\epsilon}{2}.$$

Since the functions S^{k_i} are continuous, and therefore also uniformly continuous in K, there exists $\eta > 0$ such that

$$d(a, a') < \eta \implies d(S^{k_i}a, S^{k_i}a') < \frac{\epsilon}{2} \quad \text{for all } i = 1, \ldots, N.$$

Now, by induction hypothesis, there exist $a_0 \in K$ and $n > 0$ such that

$$d(S^n a_0, a_0) < \eta, \ldots, d(S^{(l-1)n}a_0, a_0) < \eta.$$

By setting $b_0 = S^{-n}a_0$ we obtain

$$d(S^n b_0, a_0) < \eta, \ldots, d(S^{ln} b_0, a_0) < \eta,$$

so that

$$d(S^{n+k_i} b_0, S^{k_i} a_0) < \frac{\epsilon}{2}, \ldots, d(S^{ln+k_i} b_0, S^{k_i} a_0) < \frac{\epsilon}{2}$$

for $i = 1, \ldots, N$. Further for each $a \in K$ there exists an index j such that

$$d(S^{k_j} a_0, a) < \frac{\epsilon}{2}.$$

By choosing $b = S^{k_j} b_0$ we obtain

$$d(S^n b, a) < \epsilon, \ldots, d(S^{ln} b, a) < \epsilon,$$

as was to be shown. □

Proof of Key Lemma. Let $a_0 \in K$. We construct inductively

points: $a_1, \ldots, a_t \in K$;

nonnegative integers: $n_1, \ldots, n_t$;

positive numbers: $\epsilon_1, \ldots, \epsilon_t < \epsilon$

such that

$$d(S^{n_i} a_i, a_{i-1}) < \frac{\epsilon_i}{2}, \ldots, d(S^{ln_i} a_i, a_{i-1}) < \frac{\epsilon_i}{2}. \tag{16}$$

To start with we set $\epsilon_1 = \frac{\epsilon}{2}$. Then, by Lemma 3, there exist $a_1 \in K$ and $n_1 > 0$ such that (16) holds for $i = 1$. Now, assume that (16) holds for $0 \leq i \leq q$. Then we choose $\epsilon_{q+1} < \frac{\epsilon}{2}$ such that

$$d(a, a') < \epsilon_{q+1} \implies d(S^{n_q} a, S^{n_q} a') < \frac{\epsilon_q}{2}, \ldots, d(S^{ln_q} a, S^{ln_q} a') < \frac{\epsilon_q}{2}. \tag{17}$$

Then, by Lemma 3, there exist $a_{q+1} \in K$ and n_{q+1} such that (16) holds for $i = q + 1$.

As a matter of fact, we claim that, for $0 < i \leq j$, we have

$$d(S^{n_j + \cdots + n_i} a_j, a_{i-1}) < \epsilon_i, \ldots, d(S^{l(n_j + \cdots + n_i)} a_j, a_{i-1}) < \epsilon_i. \tag{18}$$

This is seen by induction on $j-i$:
$j-i=0$: This is in (16).
By induction hypothesis we have:

$$d(S^{n_j+\cdots+n_{i+1}}a_j, a_i) < \epsilon_{i+1}, \ldots, d(S^{l(n_j+\cdots+n_{i+1})}a_j, a_i) < \epsilon_{i+1}$$

Consequently, by (17),

$$d(S^{n_j+\cdots+n_i}a_j, S^{n_i}a_i) < \frac{\epsilon_i}{2}, \ldots, d(S^{l(n_j+\cdots+n_i)}a_j, S^{ln_i}a_i) < \frac{\epsilon_i}{2}.$$

From this and (16) the condition (18) follows.

To conclude, by the compactness of K, there exist indices $i<j$ such that

$$d(a_i, a_j) < \frac{\epsilon}{2},$$

and therefore by (18)

$$d(S^n a_j, a_j) < \epsilon, \ldots, d(S^{ln}a_j, a_j) < \epsilon,$$

where $n = n_{i+1}+\cdots+n_j$. Hence the Key Lemma is proved. □

Now, finally, we are ready for

Proof of Claim B. By the Key Lemma, there exist $z \in K$ and $n>0$ such that

$$d(S^n z, z) < \frac{1}{2}, \ldots, d(S^{ln}z, z) < \frac{1}{2}.$$

Hence, by the definiton of the distance function

$$z(0) = S^n z(0) = \cdots = S^{ln} z(0),$$

and so

$$z(0) = z(n) = \cdots = z(ln). \tag{19}$$

But by the construction

$$z \in K \subseteq \lim\{S^n v \mid n \geq 0\}.$$

Therefore for some $m>0$ we have

$$d(S^m v, z) < \frac{1}{ln+1}.$$

Consequently,

$$z(i) = S^m v(i) = v(m+i) \quad \text{for } 0 \leq i \leq ln.$$

Hence $(*)$ follows from (19). So the proof is completed $\square$

The above topological method can be used to prove also other Ramsey type theorems – also some which no other proof is known.

IV Numerical estimates

In this section we estimate van der Waerden numbers, as well as Ramsey numbers. In doing so we consider so-called *probabilistic argument* of Erdös to conclude the *existence* of certain type of graphs. Let us denote

$$W_l(r) = W(l, r),$$

the number obtained by our construction! As we observed

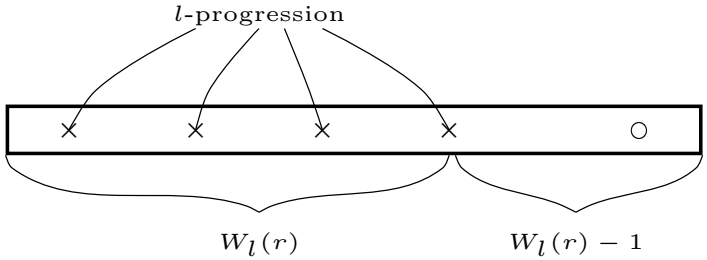
$$W_2(r) = r + 1.$$

The argument to compute $W_{l+1}(r)$ was that assuming $W_l(r)$, for all r, we forced a new color for a particular element, or $(l+1)$-progression was found. This process could be repeated at most r times. In each step the requirements were

- l-progression is guaranteed for the considered number of colors;
- $(l+1)$st element (block) is defined.

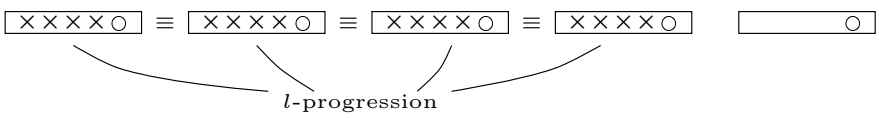
Here:

(i) The sequence of length $W_l(r)$ guarantees an l-progression, and that of $W_l(r) + W_l(r) - 1$ in addition that $(l+1)$st element is defined.



(ii) Now, we color blocks of length $2W_l(r) - 1$, so that there are $r^{2W_l(r)-1}$ colors. A new color (or $(l+1)$-progression) is forced, if the number of blocks equals to

$$2W_l(r^{2W_l(r)-1}) - 1$$



(iii) We proceed r times, when we cannot any more choose a new color so that

"$(l+1)$-progression is guaranteed".

We define

$$\begin{aligned} c_1 &= 2W_l(r) - 1, \\ c_2 &= 2W_l(r^{c_1}) - 1, \\ &\vdots \\ c_{i+1} &= 2W_l(r^{c_i}) - 1 \end{aligned}$$

Then the lengths of the initial parts of N are

c_1 (after (i)),
$c_2 \times$ the length of the blocks $= c_2 \times c_1$ (after (ii)),
$c_r \times$ the length of the blocks in the previous step (after (iii)).
$= c_r \times c_{r-1} \times \cdots c_1$

Hence, can be shown that

$$W_{l+1}(r) \geq \text{Ackermann}(l-1),$$

where

$$\text{Ackermann}(n) = f_\omega(n) = f_n(n),$$

where f_i's are defined as follows:

(20)

$$\begin{cases} f_1(x) &= 2x, \text{ and} \\ f_{i+1}(1) &= 2 \\ f_{i+1}(x+1) &= f_i(f_{i+1}(x)) \end{cases} \quad \text{or} \quad \begin{cases} f_1(x) &= 2x \\ f_{i+1}(x) &= f_i^{(x)}(1) \\ & \text{(apply previous } x \text{ times)} \end{cases}$$

(Here $f_i^{(x)}$ means iterating x times.)

So we have

$$\begin{aligned} f_1(x) &= \text{DOUBLE}(x) = 2x \\ f_2(x) &= \text{EXPONENT}(x) = 2^x \\ f_3(x) &= \text{TOWER}(x) = 2^{2^{2^{\cdots^2}}} \quad (x \text{ 2's in a tower!}) \\ f_4(x) &= \text{WOW!} \end{aligned}$$

In the table form we have:

	1	2	3	4	5	6	7	
Double	2	4	6	8	10	12	14	$\cdots$
Exponent	2	4	8	16	32	64	$\cdots$	
Tower	2	4	16	65536	2^{65536}	$\cdots$		
Wow	2	4	65536	Wow	$\cdots$			
f_5	2	4	Wow	$\cdots$				
Ackermann	2	4	16	Wow	$\cdots$			

One can show that the Ackermann function is not *primitively recursive*. Hence, our approach does not give primitively recursive upper bound for $W_{l+1}(r)$. On the other hand, Shelah proved that

$$W_l(r) \leq \text{Wow}(l+2).$$

The known exact values for van der Waerden numbers are

$$\begin{aligned} &W(3,2) = 9 \qquad W(3,3) = 27 \\ &W(4,2) = 35 \qquad W(3,4) = 76 \\ &W(5,2) = 178. \end{aligned}$$

Next we search lower bounds for Ramsey numbers $R(k,l)$. Not only because of the results, but even more before the techniques used. This method, so-called *probabilistic argument* (of Erdös), does not allow to construct a graph avoiding red K_k and blue K_l, but shows that such a graph exists!

Theorem 7. $R(k) > 2^{\frac{k}{2}} - 1$.

Proof. We show first that

$$\binom{n}{k} 2^{1-\binom{k}{2}} < 1 \implies R(k) > n. \tag{21}$$

In other words if the above inequality holds, then there exists 2-colored complete graph of size at least n not containing a complete monochromatic subgraph of size k (Here, of course, edges are colored).

Let us consider a random 2-coloring of K_n. So the color of an edge is defined by coin flipping, and the coloring of the graph by the sequence of length $\binom{n}{2}$ of such coin flippings. Therefore,

$$P[(i,j) = \text{red}] = \frac{1}{2}, \tag{22}$$

and the color of each edge is independent of the colors of other edges. So the number of colorings and its probability are

$$2^{\binom{n}{2}} \quad \text{and} \quad 2^{-\binom{n}{2}},$$

respectively. Now fix subgraph S with $|S| = k$. Further let

$$A_S \text{ be an event "}S \text{ is monochromatic"}.$$

Then

$$P(A_S) = 2 \cdot 2^{-\binom{k}{2}} = 2^{1-\binom{k}{2}}.$$

On the other hand, the event

$$\text{(23)} \qquad \text{"Some } k\text{-element } S \text{ is monochromatic"} \equiv \bigvee_{|S|=k} A_S$$

Therefore

$$P(\bigvee_{|S|=k} A_S) \leq \sum_{|S|=k} P(A_S) \leq \binom{n}{k} 2^{1-\binom{k}{2}} < 1$$

But this means that (23) is not always true, so that some 2-coloring of K_n does not contain a monochromatic subgraph of size k. Hence (21) is proved.

It remains to be proved that the left hand side of (21) is true when

$$n = 2^{\frac{k}{2}} - 1.$$

This is straightforward:

$$\begin{aligned} \binom{n}{k} &\leq \binom{2^{\frac{k}{2}}}{k} = \frac{2^{\frac{k}{2}}!}{k!(2^{\frac{k}{2}} - k)!} = \frac{2^{\frac{k}{2}} \cdots (2^{\frac{k}{2}} - k + 1)}{k!} \\ &\leq \frac{(2^{\frac{k}{2}})^k}{k!} = \frac{2^{\frac{k^2}{2}}}{k!} \leq \frac{2^{\frac{(k+2)}{2}}}{k!} 2^{\frac{(k^2-k)}{2}-1} \\ &= \frac{2 \cdot 2^{\frac{k}{2}}}{k!} 2^{\binom{k}{2}-1} \overset{k\geq 3}{\leq} 2^{\binom{k}{2}-1}. \end{aligned}$$

So if $k \geq 3$ and $n \leq 2^{\frac{k}{2}} - 1$, then

$$\binom{n}{k} < 2^{\binom{k}{2}-1}$$

proving the left hand side of (21). The case $k = 2$ is easy to verfy. □

In fact, one could (by improving the estimate of the left hand side of (21)) conclude that

$$R(k) > k \cdot 2^{\frac{k}{2}} \left[\frac{1}{e\sqrt{2}} + o(1) \right].$$

On the other hand, see Exc. 7/II, for some constant c,

$$R(k) < c \frac{4^k}{\sqrt{k}}.$$

It follows that

$$\sqrt{2} \leq \liminf_{k\to\infty} R(k)^{\frac{1}{k}} \leq \limsup_{k\to\infty} R(k)^{\frac{1}{k}} \leq 4.$$

It is not known wheather the limit $\lim_{k\to\infty} R(k)^{\frac{1}{k}}$ exists.

Theorem 7 does not require equal probabilities:

Theorem 8. *Let $p \in [0,1]$ and*

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l} (1-p)^{\binom{l}{2}} < 1.$$

Then $R(k,l) > n$.

Proof. In the above method we choose

$$P[(i,j) \text{ is red}] = p$$

Let S with $|S| = k$ and T with $|T| = l$ be fixed complete subgraphs of K_n. Further let A_S and A_T be the events

$$A_S : "S \text{ is red}"$$

$$A_T : "T \text{ is blue}".$$

Then the event

$$"K_n \text{ contains red } K_k \text{ or blue } K_l"$$

is

$$E = \bigvee_{|S|=k} A_S \vee \bigvee_{|T|=l} A_T$$

and its probability is

$$P(E) \leq \binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l} (1-p)^{\binom{l}{2}} < 1.$$

Therefore E is not always true, so the theorem follows. □

V Hales-Jewett Theorem

In this section we prove a generalization of van der Waerden's Theorem, so-called *Hales-Jewett Theorem*. It captures the combinatorial core of vdWT. In this theorem we color *arbitrary sequences* instead of sequences of numbers.

We need some terminology. The *n-cube over* $T = \{0, \ldots, t-1\}$ is

$$C_t^n = \{(x_1, \ldots, x_n) \,|\, x_i \in T\}.$$

So $|C_t^n| = t^n$. A *line* in the n-cube C_t^n is a set

$$\{\bar{x}_0, \ldots, \bar{x}_{t-1}\} \subseteq C_t^n \quad \text{with } \bar{x}_i = (x_{i1}, \ldots, x_{in})$$

such that for each $j \in \{1, \ldots, n\}$, one of the following holds

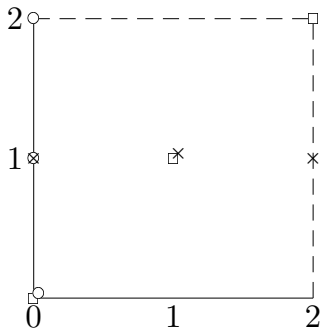
(i) $x_{0j} = x_{1j} = \cdots = x_{t-1,j}$ or

(ii) $x_{sj} = s$ for $x = 0, \ldots, t-1$,

and, moreover, (ii) holds at least for one value of j. Note that line is a set of vectors, and not sequence of those.

Example 13. In $C_3^2 = \{00, 01, 02, 10, 11, 12, 20, 21, 22\}$ the sets

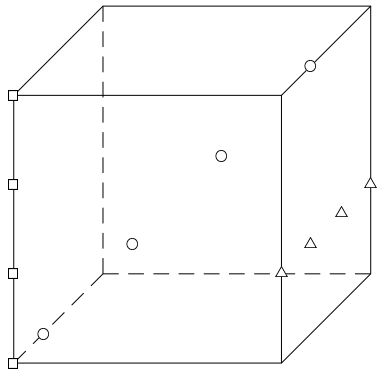
○: $\{00, 01, 02\}$,
×: $\{01, 11, 21\}$,
□: $\{00, 11, 22\}$
are lines, but
$l_g = \{02, 11, 20\}$
is not.



In the definition of a line: we want to have a *combinatorial* line and not a *geometric* one. Indeed, l_g is not accepted since if instead of $\{0, 1, 2\}$ we would use $\{0, 1, 3\}$ with normal metric l_g is not any more a geometric line.

In C_4^3 again sets

$l_\circ$: $\{020, 121, 222, 323\}$,
$l_\square$: $\{030, 031, 032, 033\}$,
$l_\triangle$: $\{301, 311, 321, 331\}$
are lines, while
$\{300, 211, 112, 003\}$
is not.



As an intuitive explanation a line is a set of vectors, where some components are fixed and the others go simultaneously from 0 to $t-1$. (suitably ordered) In particular our combinatorial line is independent of the underlying set T, so we fix

$$T = \{0, \ldots, t-1\}.$$

Let $1 \leq k \leq n$. We define *k-dimensional* subspace of C_t^n as follows. Let

$$\{1, \ldots, n\} = B_0 \dot\cup B_1 \dot\cup \cdots \dot\cup B_k, \quad B_i \neq \emptyset \quad i = 1, \ldots, k \tag{24}$$

and

$$f : B_0 \to \{0, \ldots, t-1\}.$$

Define

$$\hat{f} : C_t^k \to C_t^n$$

by

$$\hat{f}(y_1, \ldots, y_k) = (x_1, \ldots, x_n),$$

where

$$\begin{aligned} x_i &= f(i) \quad \text{if } i \in B_0, \\ x_i &= y_j \quad\;\; \text{if } i \in B_j. \end{aligned}$$

Now the k-dimensional subspace is the range of $\hat{f}$ for some function f and some partition (24).

Example 14. Let $t = 3$, $n = 7$ and $k = 2$. Further set

$$\{1, 2, \ldots, 7\} = B_0 \dot\cup B_1 \dot\cup B_2 = \{6, 7\} \dot\cup \{1, 2\} \dot\cup \{3, 4, 5\}$$

and

$$f(6) = 2 \quad \text{and} \quad f(7) = 0.$$

Now $\hat{f} : \{y_1, y_2\} \to \{x_1, \ldots, x_7\}$ is as follows:

$00 \mapsto 00\,000\,20$	$10 \mapsto 11\,000\,20$	$20 \mapsto 22\,000\,20$
$01 \mapsto 00\,111\,20$	$11 \mapsto 11\,111\,20$	$21 \mapsto 22\,111\,20$
$02 \mapsto 00\,222\,20$	$12 \mapsto 11\,222\,20$	$22 \mapsto 22\,222\,20$

So here the computed 9 vectors constitute 2-dimensional subspace S of C_3^7. Each row and column corresponds a line in C_3^7.

As we defined the subspaces of C_t^n correspond to *ordered partitions* of the set $\{1, \ldots, n\}$ together with the mappings f

$$B_0 \dot{\cup} B_1 \dot{\cup} \cdots \dot{\cup} B_t, \quad f : B_0 \to T.$$

Clearly, the actual values of T are meaningless. Note also that 1-dimensional subspaces of C_t^n are presicely the lines of C_t^n:

"f fixes the components determined by B_0 and B_1 makes the others to grow simultaneously."

In above example:

$$B_0 := B_0 \dot{\cup} B_2, \quad B_1 := B_1 \quad \text{and} \quad f(3) = f(4) = f(5) = 0$$

The considered subspace S corresponding to first row is isomorphic with C_3^2 where the isomorphism is given by

$$\varphi : S \to C_3^2, \quad \varphi(aabbb02) = ab.$$

We shall prove:

Theorem 9 (Hales-Jewett, 1963). *For each r, t, there exists a number $N' = HJ(r, t)$ such that, for $N \geq N'$, we have: If C_t^N is r-colored, then C_t^N contains a monochromatic line. (Here, of course, we color components of the vectors).*

Before the proof let us see how vdWT follows from Theorem 9. We search for an arithmetic progression of length t. Let a satisfy

$$1 \leq a < t^N$$

and write

$$a = \sum_{i=0}^{N-1} a_i t^i \qquad (t\text{-aric presentation of } a).$$

We associate this with N-dimensional vectors

$$a \leftrightarrow (a_0, \ldots, a_{N-1}), \quad a_i \in \{0, \ldots, t-1\}.$$

Clearly, this is a bijection.

Let now

$$\chi : [N] \to \{1, \ldots, r\}$$

be an r-coloring. This determines an r-coloring of the N-cube C_t^N. Now when N is large enough, this determines a monochromatic line of C_t^N, say

$$\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_t, \quad \begin{cases} \bar{x}_i = \{x_{i1}, \ldots, x_{it} \\ x_{ij} \in \{0, \ldots, t-1\}. \end{cases}$$

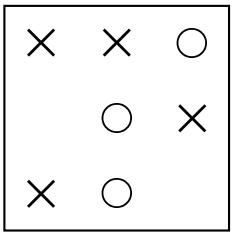
So in the vectors $\bar{x}_i$ some components are constant, and the others grow simultaneously. For example, if $N = 5$, $t = 4$ we might have

$\bar{x}_0$	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$
01020	11121	21222	31323

Then $\bar{x}_i$'s constitute the arithmetic progression, where the difference of the consecutive elements is

$$1 + t^2 + t^4.$$

Example 15. (Tic-tac-toe) This game with any number r of players on the large enough cube C_t^N never ends with a draw! This follows directly from HJT. Note that the above holds even for combinatorial lines and so for sure also for geometric lines. In C_3^2 with two players this, of course is not true.



Proof of Theorem 9. We first recall the equivalence classes of Theorem 6. For each $0 \leq i \leq n$ the ith equivalence class consists of the vectors

$$(x_1, \ldots, x_n) \in C^n_{t+1}, \text{ where } x_1, \ldots, x_{n-i} \neq t, \quad x_{n-i+1} = \cdots = x_n = t.$$

Consequently, the i rightmost components equal to t, and the others $\neq t$ (Earlier we used a different enumeration and instead of "rightmost" "leftmost"). We call a coloring of C^m_{t+1} *layered* if it is constant in the above equivalence classes. Further a k-dimensional subspace is *layered*, if its color is constant, when it is identified with canonical k-cube C^k_{t+1}. (When talking about layered subspaces we assume that the coloring is given). A line is layered if its t first components are colored with the same color.

In Example 14 $S \simeq C^2_{2+1}$ with $aabbb20 \mapsto ab$. So being layered means that the classes

b \ a	0	1
0	00 000 20	11 000 20
1	00 111 20	11 111 20

and

b \ a	0	1
2	00 222 20	11 222 20

are monochromatic.

Example 16. Let $t = 27$ and $\Sigma = \{A, B, \ldots, Z, \phi\}$. So $C^n_{27} = \Sigma^{27}$. We call $w \in \Sigma^{27}$ *left proper* if all ϕ components are at the end. Now the coloring of Σ^{27} is layered iff all left proper words of the same length are similarly colored. So a line

$$\{\alpha A \alpha \mid A \in \Sigma\}$$

is layered if the words

$$AAA, BAB, \ldots, ZAZ \quad \text{have the same color,}$$

but $\phi A \phi$ might have any color. Further if the 2-dimensional subspace

$$\{\alpha A \beta \beta \mid \alpha, \beta \in \Sigma\}$$

is layered, then the words

$$BALL, MASS \text{ and } PARR \text{ correspond to the same color,}$$

as well as the words

$$MA\phi\phi, PA\phi\phi \text{ and } LA\phi\phi.$$

What we are going to prove is the following statements depending on t:

$HJ(t)$: For each r, there exists $N' = HJ(r,t)$ such that if $N \geq N'$ and C_t^N is r-colored, then it contains a monochromatic line.

$LHJ(t)$: For each r and k, there exists $M' = LHJ(r,t,k)$ such that if $M \geq M'$ and C_{t+1}^M is r-colored, then it contains layered k-dimensional monochromatic subspace.

The proofs are by induction on t. More precisely we show

$$\begin{aligned} HJ(t) &\implies LHJ(t), \text{ and} \\ LHJ(t) &\implies HJ(t+1). \end{aligned}$$

$HJ(2)$ We can choose $HJ(r,2) = r$ (Exc).

$HJ(t) \implies LHJ(t)$. So we assume $HJ(t)$, and conclude $LHJ(t)$ by induction on k. This is done *simultaneously for all values of* r as in the proof of van der Waerden's Theorem.

$k = 1$ We set

$$M' = LHJ(r,t,1) = HJ(r,t).$$

Indeed, let

$$M \geq M' \quad \text{and} \quad C_{t+1}^M \quad r\text{-colored}.$$

Now, C_t^M consists of these points of C_{t+1}^M where no component assumes the value t. So, by our assumption and choice of M and M', C_t^M contains a monochromatic line, for example ($t = 5$, $N = 5$)

$$S_1 = \{11020, 11121, 11222, 11323, 11424\}.$$

Consequently, C_{t+1}^M contains a monochromatic layered line

$$S_1' = \{11020, 11121, 11222, 11323, 11424, 11525\}$$

In S_1' the color of 11525 may be different from that of others.

Induction step, $k \to k+1$. This is the core of the proof. We use the method of *induced colorings*. We set

$$\begin{aligned} m &= LHJ(r,t,k) \quad \text{(i.h.)}, \\ s &= r^{(t+1)^m} \quad (= \#\text{ of different colorings of } C_{t+1}^m) \end{aligned}$$

and

$$m' = LHJ(s,t,1) = HJ(s,t).$$

Here m might be "gigantic", but m' is still "unbelievably larger"!! We claim that we can choose

$$LHJ(r,t,k+1) = m + m'.$$

Let now

$$\chi : C_{t+1}^{m+m'} \to \{1,\ldots,r\} \quad (r\text{-coloring of } C_{t+1}^{m+m'}).$$

We can identify:

$$C_{t+1}^{m+m'} = C_{t+1}^{m} \times C_{t+1}^{m'}$$

That means the product (catenation) of vectors $x \in C_{t+1}^{m}$, $y \in C_{t+1}^{m'}$ equals to $xy \in C_{t+1}^{m+m'}$. (For example, $(2,6,5)(3,6) = 26536$). We define an r-coloring χ^* of $C_{t+1}^{m'}$ (so-called induced coloring) by setting

$$\chi^*(x) = \chi^*(x') \quad \text{iff} \quad \chi(xy) = \chi(x'y) \quad \forall y \in C_{t+1}^{m}. \tag{25}$$

Consequently, χ^* is $r^{(t+1)^m} = s$-coloring. Now, $C_{t+1}^{m'}$ is s-colored, and so by the choice of m' there exists a layered monochromatic line

$$\{x_0, x_1, \ldots, x_t\} \subseteq C_{t+1}^{m'}.$$

Therefore χ^* colors the vectors $x_0,\ldots,x_{t-1} \in C_{t+1}^{m'}$ with the same color. But these vectors do not contain t as any component, so that they are in fact in $C_t^{m'}$.

Now, we color C_{t+1}^{m} by setting

$$\chi^{**}(y) = \chi(x_i y) \quad \forall i \in \{0,\ldots,t-1\}. \quad \text{(OK by (25))}.$$

So C_{t+1}^{m} is r-colored by χ^{**}. Hence, by the choice of m, induction hypothesis applies: There exists k-dimensional layered (with respect to χ^{**}) subspace $S \subseteq C_{t+1}^{m}$. We set

$$T = \{x_i s \mid 0 \le i \le t,\ s \in S\} \subseteq C_{t+1}^{m'+m}.$$

Let the equivalence classes of S be $S_0, S_1, \ldots, S_k$. Then the equivalence classes of T are

$$T_j = \{x_i s \mid 0 \le i < t,\ s \in S_j\} \quad \text{for } j = 0,\ldots,k$$

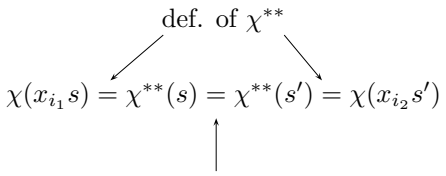
and

$$T_{k+1} = \{x_t s_k\}, \quad \text{where } S_k = \{s_k\}.$$

Consider now the elements

$$x_{i_1} s, x_{i_2} s' \in T_j, \quad j = 0, \ldots, k.$$

Then



S layered under χ^{**} and s and s' in the same class.

Consequently, T is $(k+1)$-dimensional layered subspace of $C_{t+1}^{m+m'}$ with respect to χ. Hence, the implication $HJ(t) \implies LHJ(t)$ has been proved.

Some intuition of the above explaining why the number n is so huge. Here the procedure is done in k steps. We try to choose

$$M = LHJ(r, k, t)$$

so huge that we can write,

$$M = m_1' + m_1,$$

where m_1 is "gigantic" and m_1' "even much more larger". Writing

$$C_{t+1}^{m_1'+m_1} = C_{t+1}^{m_1'} \times C_{t+1}^{m_1}$$

r-coloring of it induces s-coloring for $C_{t+1}^{m_1'}$. Here $s \gg m_1$, but $m_1' \gg s$. Assuming m_1' large enough $C_{t+1}^{m'}$ has a layered monochromatic line with respect to s-coloring:

$$L_1 = \{x_0^{(1)}, \ldots, x_t^{(1)}\}$$

In the set

$$L_1 \times C_{t+1}^{m_1}$$

the original color of the point $x_i^{(1)}$ is independent of i, if $i \neq t$. Next we color

$$y \in C_{t+1}^{m_1} \qquad \text{(new coloring)}$$

by the color of $x_i^{(1)}y$ when $i \neq t$. Since m_1 is "gigantic" we write

$$m_1 = m_2' + m_2,$$

where $m_2' \gg m_2$ and m_2 is still "gigantic". Now we have an r-coloring of the cube

$$C_{t+1}^{m_1} = C_{t+1}^{m_2'} \times C_{t+1}^{m_2}$$

inducing an s-coloring of $C_{t+1}^{m_2'}$ and hence, assuming m_2' large enough, a layered monochromatic line in $C_{t+1}^{m_2'}$:

$$L_2 = \{x_0^{(2)}, \ldots, x_t^{(2)}\}.$$

In the set

$$L_1 \times L_2 \times C_{t+1}^{m_2}$$

the color of the point

$$x_i^{(1)} x_j^{(2)} y$$

is independent of i, if $i \neq t$, and independent of i and j, if $i \neq t$ and $j \neq t$. Continuing the procedure k times a k-dimensional layered subspace is found for large enough value of M.

What remains is the implication:
$LHJ(t) \implies HJ(t+1)$.
Claim. Any layered k-dimensional subspace colored with at most k colors contains a monochromatic line.

Proof of Claim. All ordered k-dimensional subspaces are isomorphic. So it is enough to look at the space C_{t+1}^k. For ilustration let $k = 2$ and $t = 4$. The equivalence classes are:

○	○	○	○	$*$
×	×	×	×	?
×	×	×	×	?
×	×	×	×	?
×	×	×	×	?

So if $\times = \circ$, there exists a monochromatic column,
$\circ = *$, there exists a monochromatic row,
$\times = *$, there exists a monochromatic diagonal,
Hence, in any case a line since the number of colors $\leq k = 2$.

In general, the argument goes as follows. Let C^k_{t+1} be layered subspace and the points x_i, $0 \le i \le k$, defined as follows:

$$x_i = (x_{i1}, \ldots, x_{ik}) \quad \text{with} \quad x_{ij} = \begin{cases} t & \text{if } j \le i \\ 0 & \text{if } i < j \end{cases}.$$

For example, in C^2_5: $\{00, 40, 44\}$. By PHP, there exists $u < v$ such that x_u and x_v have the same color, say red. Then the line

$$\{y_0, \ldots, y_t\},$$

where

$$y_s = (y_{s1}, \ldots, y_{sk}) \quad \text{with} \quad y_{si} = \begin{cases} t & \text{if } i \le u \\ s & \text{if } u < i \le v \\ 0 & \text{if } v < i \end{cases}$$

is red. In the above example in C^2_5 if 00 and 44 are red, so is $\{00, 11, 22, 33, 44\}$.

Now the required implication follows from the Claim:
Let r be given. We Choose $N' = LHJ(r, t, r)$ (by assumption), and claim that N' such chosen works for $HJ(r, t+1)$. Consider the cube

$$C^N_{t+1} \quad \text{with } N \ge N'$$

and its r-coloring. Then, by the choice of N, C^N_{t+1} contains r-dimensional layered subspace, and therefore, by Claim, a monochromatic line.

This ends the proof of Hales-Jewett Theorem. □

Actually, via the isomorphism

$$C^{ns}_t \simeq C^s_{t^n}$$

Hales-Jewett Theorem extends straightforwardly (we do not go into the details) to n-dimensional subspaces:

Theorem 10 (n-dimensional Hales-Jewett theorem). *For each numbers r, t and n there exists a number $N' = N'(r, t, n)$ such that if C^N_t, with $N \ge N'$, is r-colored, then it contains monochromatic n-dimensional subspace.*

VI Shirshov's Theorem

In this section we prove another "unavoidable regularity" result. It is formulated for words, e.g. for sequences of symbols taken from a finite alphabet. It has a lot of applications, for example in semigroup theory and in automata theory. Intuitively it says that any long enough word is either

(i) "periodic" in the sense that it contains as a factor a high power, that is a word of the form u^n; or

(ii) "minimal" in a certain precise sense.

We fix some terminology.
Let Σ be a finite *alphabet*. A *word* over Σ is any sequence of symbols including the empty sequence, called *empty word* denoted by 1. The set of all words Σ^* is a monoid under the operation of *product* or *catenation*:

$$u \cdot v = uv.$$

It is *freely* generated by the alphabet Σ, that is each $w \in \Sigma^*$ has the unique representation as the product of letters (elements of Σ). Similarly $\Sigma^+ = \Sigma^* \setminus \{1\}$ is a *free semigroup*. The *length* of w is denoted by $|w|$.

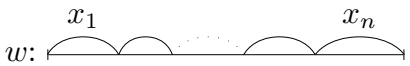
Subsets of Σ^* are called *languages*. For a language $X \subseteq \Sigma^*$ we can define

$$X^* = \{x_1 \cdots x_n \mid n \geq 0, x_i \in X\}$$

and

$$X^+ = \{x_1 \cdots x_n \mid n \geq 1, x_i \in X\}.$$

Note that $1 \in X^*$ always. An *X-factorization* of $w \in X^*$ is a sequences of words $x_1, \ldots x_n \in X$ such that $w = x_1 \cdots x_n$, depicted as



A language X is a *code* if each $w \in X^*$ possesses the unique X-factorization, in other words, X^* is *freely* generated by X. A word u is a *factor* of w if there exist words p and s such that $w = pus$. In the case $p = 1$ (resp. $s = 1$)

u is a *prefix* (resp. *suffix*) of w. The prefix of length k of a word w is denoted by $\mathrm{pref}_k(w)$ (In the case $|w| < k$, we set $\mathrm{pref}_k(w) = w$). Similarly we define $\mathrm{suf}_k(w)$. Let $\mathrm{pref}(w)$ denote all prefixes of w.

A word w is a *nth power* if it is of the form $w = u^n$ for some word u. In particular, *squares* are words of a form u^2. Note that powers can be defined also for rational numbers, e.g. $ababa = (ab)^{2\frac{1}{2}}$, and even for irrational numbers: w is a ρth power if it is a qth power for some rational $q > \rho$. If $w = u^n$ we say that u is of *repetition order* n. Finally, a word w is *k-free* if it does not contain as a factor any word of repetition order $\geq k$. In particular square-free words do not contain repetitions of any word. e.g. a factor uu.

For Shirshov's Theorem it is important to order the set of all words. Let Σ be *totally* ordered by $<$. We extend this to Σ^* as follows:

$$u < v \quad \text{iff} \quad \begin{cases} v = ux & \text{with } x \neq 1, \text{ or} \\ u = pas \text{ and } v = pbs' & \text{with } a < b, \text{ where} \\ a, b \in \Sigma,\ p, s, s' \in \Sigma^*. & \end{cases}$$

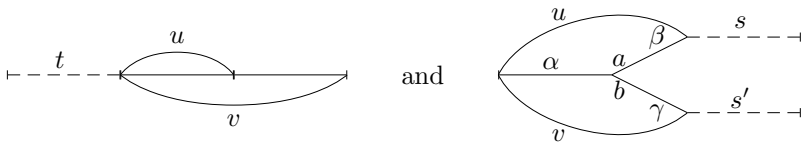
Clearly, $<$ is a total order in Σ^*, so-called *alphabetic order*.

Facts: *The above order $<$ satisfies:*

(i) if $u < v$, then $tu < tv$ for all $t \in \Sigma^*$;

(ii) if $u < v$ and $u \notin \mathrm{pref}(v)$, then $us < vs'$ for all $s, s' \in \Sigma^*$.

The illustrations of these cases are as follows:



Words of particular form play an important role in our considerations. We define

$$X = a^+(\Sigma \setminus a)^+ \subseteq \Sigma^+.$$

Then

$$\begin{aligned} X^* &= 1 \cup a\Sigma^*(\Sigma \setminus a)^+ \\ &= 1 \cup \{w \in \Sigma^+ \mid \mathrm{pref}_1(w) = a, \mathrm{suf}_1(w) \neq a\}. \end{aligned}$$

Moreover, each $w \in X^*$ possesses the unique X-factorization:

$$\underbrace{aabcabcaaabbcab}$$

This means that X is a code.

Now, the order $<_\Sigma$ of Σ^* induces an order for X which can be extended to a total order $<_X$ of X^* (or, in fact, $1 \cup X \cup X \times X \cup \cdots$). But X is a code, so that $<_X$ can be viewed as total order of subset X^* of Σ^*. Indeed, we have:

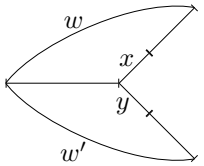
Lemma 4. *For each $w, w' \in X^*$ we have*

$$w <_\Sigma w' \iff w <_X w'.$$

Proof. The implication $\Leftarrow$ is crucial. Assume that $w <_X w'$. Then either

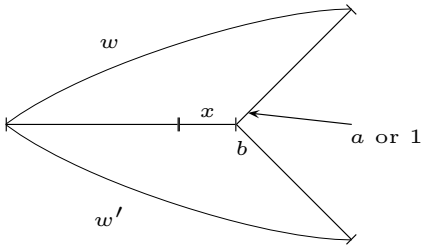
(i) $w' = ww''$ with $w'' \neq 1$, and hence $w <_\Sigma w'$, or

(ii)



$$\text{e.g.} \begin{cases} w = w_1 x w_2 \\ w' = w_1 y w_2' \end{cases} \text{ with } x <_X y, \text{ where}$$
$$x, y \in X, w_1, w_2, w_2' \in X^*.$$

If $x = uax'$ and $y = uby'$ with $a < b$, then $w <_\Sigma w'$.
If, on the other hand, $y = xu$, then $b = \mathrm{pref}_1(u) \neq a$, so that we have in Σ^*:



So also in this case $w <_\Sigma w'$
This concludes the proof of $\Leftarrow$.

Implication $\Rightarrow$. Clearly, since $<_X$ is a total order, we have

$$w <_\Sigma w' \implies \begin{cases} w =_X w' & \text{The first case does not hold.} \\ w >_X w' & \text{Neither does the second by above.} \\ w <_X w' & \text{Hence the third holds.} \end{cases}$$

Now we are ready to define the notions needed. We say that $w \in \Sigma^*$ is *n-divided* or (*n-Σ-divided*) if

$$w = w_1 \cdots w_n$$

and, for each permutation $\sigma \in G_n, \sigma \neq id$, we have

$$w < w_{\sigma(1)} \cdots w_{\sigma(n)}.$$

We can extend this to n-X-divided words in a natural way.

Example 17. Let $x, y \in X^*$ and $xy <_X yx$. Then xy is 2-X-divided. We claim that the word $xya \in \Sigma^*$ can be 3-Σ-divided as follows:

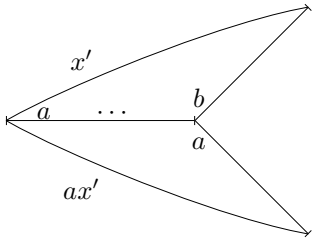
$$xya = ax'ay'a = a.x'a.y'a$$

So we have to show, in what ever way we permute the three factors (shown in the last form), we obtain a word larger than xya: For example

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad : \qquad ay'ax'a = yxa >_\Sigma xya$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad : \qquad x'ay'aa = x'yaa \overset{(*)}{>} ax'yaa > xya$$

$(*)$ since $ax' < x'$:



□

Actually, the above example extends to:

Lemma 5. *If $w \in X^*$ and is $(n-1)$-X-divided, then wa is n-Σ-divided.*

Proof. Let w have the $(n-1)$-X-division

$$w = w_1 \cdots w_{n-1}.$$

Consequently,

$$w < w_{\sigma(1)} \cdots w_{\sigma(n-1)} \qquad \forall \sigma \in G_{n-1}, \sigma \neq id.$$

Since $w_i \in X$ we can write $w_i = aw_i'$, and further

$$\begin{aligned} wa &= aw_1'a \cdots aw_{n-1}'a \\ &= \underbrace{a}\underbrace{w_1'a} \cdots \underbrace{w_{n-1}'a} = u_1 u_2 \cdots u_n. \end{aligned}$$

We claim that this is the n-Σ-division of wa.

Let $\sigma \in G_n$ and

$$\bar{\sigma} = u_{\sigma(1)} \cdots u_{\sigma(n)}.$$

Clearly, we can decompose σ as

$$\sigma = \alpha \circ \tau,$$

where

$$\alpha(1) = 1 \text{ and } \tau \text{ is a cycle } (\overbrace{1, \ldots, r}).$$

This holds since: if $r = \sigma^{-1}(1)$, $\tau = (1, \ldots, r)$ and $\alpha = \sigma \circ \tau^{-1}$, then $\sigma = \alpha \circ \tau$ and $\alpha(1) = \sigma \circ \tau^{-1}(1) = \sigma(r) = 1$. From our requirement $\sigma \neq id$ it follows that

$$\alpha \neq id \text{ or } \tau \neq id.$$

We have two cases to be considered:

(i) $\tau = id$. In this case $\sigma = \alpha \neq id$. Then

$$\begin{aligned} \bar{\alpha} &= au_{\alpha(2)} \cdots u_{\alpha(n)} = aw'_{\alpha(2)-1}a \cdots w'_{\alpha(n)-1}a \\ &= w_{\beta(1)} \cdots w_{\beta(n-1)}a, \end{aligned}$$

where

$$\beta(i) = \alpha(i+1) - 1.$$

Now $\beta \in G_{n-1}$ and $\beta \neq id$. Hence, by our assumption,

$$w <_X w_{\beta(1)} \cdots w_{\beta(n-1)}.$$

Therefore, by Lemma 4,

$$w <_\Sigma w_{\beta(1)} \cdots w_{\beta(n-1)},$$

and also

$$wa <_\Sigma w_{\beta(1)} \cdots w_{\beta(n-1)} a = \bar{\alpha} = \bar{\sigma}.$$

(ii) $\tau \neq id$, e.g. $r \geq 2$. Then

$$\begin{aligned} \bar{\sigma} &= u_{\sigma(1)} u_{\sigma(2)} \cdots u_{\sigma(n)} \\ &= u_{\alpha(2)} \cdots u_{\alpha(r)} u_{\alpha(1)} u_{\alpha(r+1)} \cdots u_{\alpha(n)} \\ &= u_{\alpha(2)} \cdots u_{\alpha(r)} a u_{\alpha(r+1)} \cdots u_{\alpha(n)}. \end{aligned}$$

Now, $\alpha(2) \neq 1$, so that

$$u_{\alpha(2)} = a^k bv \qquad \text{with } k \geq 0, b \neq a, v \in \Sigma^*.$$

Then

$$\begin{aligned} \bar{\sigma} &= a^k bv', \text{ and} \\ \bar{\alpha} &= u_{\alpha(1)} u_{\alpha(2)} \cdots = a^{k+1} bv'', \end{aligned}$$

and therefore

$$(*) \qquad \bar{\alpha} <_\Sigma \bar{\sigma}$$

Now, if $\alpha = id$, then $wa = \bar{\alpha}$, and hence $wa < \bar{\sigma}$. If, on the other hand, $\alpha \neq id$, then, by (i), $wa <_\Sigma \bar{\alpha}$, and hence, by $(*)$, also now $wa < \bar{\sigma}$. This completes the proof of Lemma 5. □

Now, we are ready to prove the Shirshov's Theorem. Here the intuitive "minimality" is formalized by "n-divisions". As in all Ramsey type results we have considered this is also proved by a double induction, and hence the bounds are extremely large.

Theorem 11 (Shirshov, 1957). *Let k, p and n be natural numbers and Σ a totally ordered k-letter alphabet. There exists a number $N(k,p,n)$ such that, for any $w \in \Sigma^*$ with $|w| \geq N(k,p,n)$, either*

(i) w contains as a factor a pth power, or

(ii) w contains as a factor an n-Σ-divided word.

Proof. By double induction on n and k. So we fix p to be a constant.
Clearly,

$$N(k,p,1) = 1$$

and

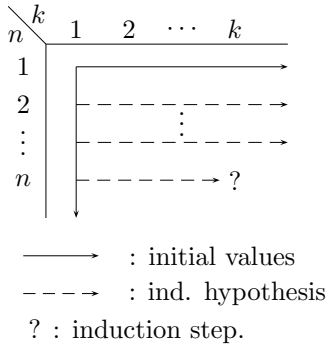
$$N(1,p,n) = p.$$

We assume

$$N(k,p,n-1) \quad \text{for all } k$$

and

$$N(j,p,n) \quad \text{for } j < k$$



In order to prove the induction step we show:
Claim: $N = N(k,p,n) = (p + N(k-1,p,n))(N(k^{N(k-1,p,n)+p}, p, n-1) + 1)$

Let $|w| \geq N$.
We note first that

(i) if $w = uw'$, with $u \in (\Sigma \setminus a)^{N(k-1,p,n)}$, then, by i.h., w contains an n-Σ-divided word or a pth power;

(ii) if $w = w''a^{p+t}$, then w contains a pth power.

It follows that w contains a factor w_1 such that

$$|w_1| \geq (p + N(k-1,p,n)N(k^{N(k-1,p,n)+p}, p, n-1)$$

and

$$w_1 \in a\Sigma^*(\Sigma \setminus a).$$

Then we can write

$$w_1 = x_1 \cdots x_r \quad \text{with } x_i \in X = a^+(\Sigma \setminus a)^+,$$

and moreover

$$x_i = a^q s \quad \text{with } s \in (\Sigma \setminus a)^+.$$

As in (i) and (ii) we may assume that

$$q < p \text{ and } |s| < N(k-1, p, n).$$

In particular,

$$r > N(k^{N(k-1,p,n)+p}, p, n-1).$$

We apply i.h. to the alphabet

$$\{x \in X \mid |x| < p + N(k-1, p, n)\}.$$

Therefore the word $x_1 \cdots x_{r-1}$ contains either a pth power or an $(n-1)$-X-divided factor. Hence, by Lemma 5, also the word $x_1 \cdots x_{r-1}a$ contains either

- a pth power, or
- an n-Σ-divided factor.

This concludes the proof □

Again the numbers $N(k, p, n)$ grow extremely fast:

$$N(k+1, p, n) = n_{k+1} \gg (k+1)^{n_k} \gg \cdots \gg (k+1)^{k^{\cdot^{\cdot^{\cdot^{2}}}}}.$$

In some special cases the length of the pth power can be bounded. In order to prove that we recall that a word ρ is *primitive* if it is not a proper integer power of any other word. So *aba* is primitive while *abab* is not.

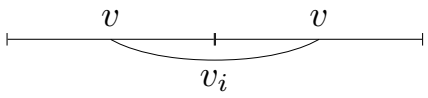
We also recall that u and v are *conjugates* if there exist words p and q such that $u = pq$ and $v = qp$, that is v is obtained from u by moving its prefix to the end. Of course, each word u has at most $|u|$ different conjugates. And it is not difficult to see that a primitive word has exactly that many different conjugates.

Corollary 2. If $p \geq 2n$ in Theorem 11, then the pth power can be chosen to be u^p with $|u| < n$.

Proof. Assume that w contains a pth power v^p and $|v| \geq n$. We can assume that v is primitive. Then v has at least n different conjugates, say

$$v_1 < v_2 < \cdots < v_n.$$

Since $p \geq 2n$, the word v^p contains as a factor $(v^2)^n$. Further each v^2 contains each of the v_i's:



Consequently, v^p (and hence also w) contains

$$f = (v_1 v_1')(v_2 v_2') \cdots (v_n v_n').$$

We claim that this is a required n-division.
Let $\sigma \in G_n \setminus \{id\}$ and i the smallest index for which $\sigma(i) \neq i$. Then

$$\begin{aligned} \sigma(t) &= t \quad \text{if } t = 1, \ldots, i-1, \\ \sigma(i) &= j > i, \end{aligned}$$

and so

$$\sigma(f) = (v_1 v_1') \cdots (v_{i-1} v_{i-1})(v_j v_j') g.$$

But

$$f < \sigma(f)$$

proving the Corollary. □

An application to Burnside Problem

Burnside Problem *asks whether finitely generated semigroup, all elements of which generate a finite subsemigroup, is finite. Or formally:*

$$\left.\begin{array}{l} S = \langle \Sigma \rangle = \{a_{i_1} \cdots a_{i_t} \mid a_{i_j} \in \{a_1, \ldots, a_n\}\} \\ \operatorname{card} \langle a \rangle = \operatorname{card}\{a^i \mid i \in N\} < \infty \quad \forall a \in S \end{array}\right\} \stackrel{?}{\Longrightarrow} \operatorname{card}(S) < \infty$$

This is indeed a crucial problem in algebra. In what follows we give a few examples of the answers.

Example 18. Answer is "no" in general. The solution comes from the existence of repetition free words (cf. Example 19). Indeed consider the semigroup

$$S = \{w \in \{a, b, c\}^* \mid w \text{ is cube-free}\} \cup \{0\}$$

with the operation

$$\begin{aligned} u \cdot v &= uv \qquad \text{if } uv \text{ is cube-free,} \\ u \cdot v &= 0 \qquad \text{if } uv \text{ contains a cube,} \\ u \cdot 0 &= 0 \cdot u = 0. \end{aligned}$$

This, clearly, gives a negative answer to Burnside Problem due to the fact there exist infinitely many cube-free words over $\{a, b, c\}$.

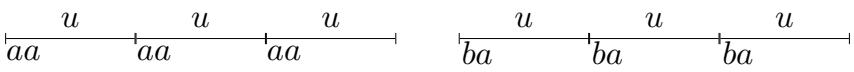
Example 19. Consider the morphism

$$h : \begin{array}{ccc} a & \mapsto & aba \\ b & \mapsto & abb \end{array}.$$

Then a is a prefix of $h(a)$, and inductively $h^i(a)$ is a prefix of $h^{i+1}(a)$ so that the limit

$$\gamma = \lim_{i \to \infty} h^i(a)$$

exists. Now note that aab is of the form $uauaub$ and h preserves words of this form. So for any $\epsilon > 0$, γ contains a repetition of order larger than $3 - \epsilon$. On the other hand, γ does not contain any cubes. This is seen as follows: First, γ does not contain aaa or bbb. Second, if it contains a cube uuu then it contains also a shorter cube. Indeed we have one of the following cases (two others are symmetric to the first case):



In the first case aa can be covered by $h(a)$ and $h(b)$ just in the unique way: $\overbrace{aa}$ Consequently all u's are covered in the same way so that $h^{-1}(\gamma)$ contains a shorter cube. In the second case the ba can be covered in two different ways: $\overbrace{ba}$. Now, if the two first occurences are covered in the same way then, by the length argument, so is the third one, and we are in

the above case. If these two are covered differently then again, by the length argument, the third one is covered still differently, that is by $\widehat{ba}$. However, this is impossible. □

Example 20. For Abelian semigroups the answer is trivially "yes".

Example 21. For *free idempotent semigroups* the answer is also "yes". This semigroup is defined as follows: We define a relation *idempotently equal* in Σ^* as follows

$$u \sim_i v \qquad \text{iff} \qquad \begin{array}{l}\text{we can transform } u \text{ to } v \text{ by applying rules}\\ x \to xx \text{ or } xx \to x \text{ for factors finitely many}\\ \text{times.}\end{array}$$

Note that $\sim_i$ is congruence (that is an equivalence relation satisfying: $u \sim_i v \implies sut \sim_i svt$ for all s, t). Now, the free idempotent semigroup over Σ is the semigroup of these congruence classes, more formally

$$S = {}^{\Sigma^*}\!/_{\sim_i}. \qquad \square$$

Example 22. As one of the major results in (combinatorial) group theory we state that the Burnside Problem has a negative answer for groups, as was shown by Adian and Novikov.

We conclude with another affirmative answer. We call a semigroup S *permutable*, if there exists an n such that, for each $s_1, \ldots, s_n \in S$, there exists a permutation $\sigma \in G_n \setminus \{id\}$ such that

$$s_1 \cdots s_n = s_{\sigma(1)} \cdots s_{\sigma(n)}.$$

We shall prove:

Theorem 12. (A. Restivo, Ch. Reutenauer, J. Alg 84). *Each finitely generated permutable semigroup S such that its all elements generate a finite semigroup is finite.*

Proof. Let $S = \langle \Sigma \rangle$, $\mathrm{card}(\langle s \rangle) < \infty$ for all $s \in S$ and $k = \mathrm{card}(\Sigma)$. Further let S be permutable with the value n. Define the morphism

$$\varphi : \Sigma^* \to S, \qquad \varphi(w) = w \in S \quad \text{(the canonical morphism)},$$

and let $N(p)$ be the number given by the Corollary above. We choose $p \geq 2n$ satisfying:

$$w \in \Sigma^*, |w| < n \implies \exists p' : \varphi(w)^p = \varphi(w)^{p'} \quad \text{with } p' < p.$$

This is possible, since $\text{card}(\Sigma) < \infty$ and $\text{card}(\langle s \rangle) < \infty$ for all $s \in S$. Now, for any $s \in S$, we define

$$\begin{aligned} \text{w(s)} &= \text{the largest with respect to} < \text{of the words of} \\ &\quad \text{minimal length in } \varphi^{-1}(s). \end{aligned}$$

Claim. *$|w(s)| < N(p)$ for all $s \in S$.*

Proof. Assume the contrary: $|w(s)| \geq N(p)$.
Then, by the choice of $N(p)$, $w(s)$ contains either

(i) an n-divided word, or

(ii) a pth power x^p with $0 < |x| < n$.

We show that both of these statements are contradictory.

First assume that $w(s) = ux_1 \cdots x_n v$, where $x = x_1 \cdots x_n$ is an n-division. Since S is permutable with value n, there exists $\sigma \neq id$ such that the

$$\begin{aligned} s = \varphi(w(s)) &= \varphi(u)\varphi(x_1)\cdots\varphi(x_n)\varphi(v) = \varphi(u)\varphi(x_{\sigma(1)})\cdots\varphi(x_{\sigma(n)})\varphi(v) \\ &= \varphi(ux_{\sigma(1)}\cdots x_{\sigma(n)}v). \end{aligned}$$

Consequently, by the choice of $w(s)$,

$$w(s) > ux_{\sigma(1)}\cdots x_{\sigma(n)}v,$$

and therefore

$$x_1 \cdots x_n > x_{\sigma(1)} \cdots x_{\sigma(n)}.$$

This is a contradiction since $x = x_1 \cdots x_n$ is an n-division.

Second assume that $w(s) = ux^p v$ with $|x| < n$. Then

$$\begin{aligned} \varphi(w(s)) &= \varphi(ux^p v) = \varphi(u)\varphi(x)^p\varphi(v) = \varphi(u)\varphi(x)^{p'}\varphi(v) \\ &= \varphi(ux^{p'}v) \end{aligned}$$

with $p' < p$. But this contradicts with the minimality of $|w(s)|$. So the claim, and hence also the theorem is proved. □